Non-linear Dimensionality Reduction on Extracellular Waveforms Reveals Cell Type 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 are thought to contain a large number of cell types that coordinate to produce behavior. Current *in vivo* methods rely on clustering of specified features of extracellular waveforms to identify putative cell types, but these capture only 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 revealing previously unknown diversity within these subtypes. The eight clusters exhibited distinct laminar distributions, characteristic firing rate patterns, and decision-related dynamics. Such insights were weaker when using feature-based approaches. *WaveMAP* therefore provides a more nuanced understanding of the dynamics of cell types in cortical circuits.

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

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

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 (Chandrasekaran et al., 2017; Finn et al., 2019), cell types (Pinto and Dan, 2015; Estebanez et al., 2017; Lui et al., 2021; Kvitsiani et al., 2013) and brain areas (Gold and Shadlen, 2007; Cisek, 2012). A large body of research has described differences in decision-related dynamics across brain areas (Ding and Gold, 2012; Thura and Cisek, 2014; Roitman and Shadlen, 2002; Hanks et al., 2015) and a smaller set of studies has provided insight into layer-dependent dynamics during decision-making (Chandrasekaran et al., 2017; Finn et al., 2019; Chandrasekaran et al., 2019; Bastos et al., 2018). However, we currently do not understand how decision-related dynamics emerge across putative cell types. Here, we address this open question by developing a new method, *WaveMAP*, that combines non-linear dimensionality reduction and graph-based clustering. We apply *WaveMAP* 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 (Pinto and Dan, 2015; Lui et al., 2021), whereas *in vivo* studies in primates are largely restricted to using features of the extracellular action potential (EAP) such as trough to peak duration, spike width, and cell firing rate (FR). Early *in vivo* monkey work (Mountcastle et al., 1969) 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 (McCormick et al., 1985), cat (Azouz et al., 1997), and the rat (Simons, 1978; Barthó et al., 2004) then helped establish the idea that these broad- and narrow-spiking extracellular waveform shapes mostly corresponded to excitatory and inhibitory cells respectively. These results have been used as the basis for identifying cell types in primate recordings (Johnston et al., 2009; Merchant et al., 2008, 2012). 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 (Hodge et al., 2019; Krienen et al., 2020). Furthermore, recent observations in the monkey defy this simple classification of broad- and narrow-spiking cells as corresponding to

excitatory and inhibitory cells respectively. Three such examples in the primate that have resisted this principle are narrow-spiking pyramidal tract neurons in deep layers of M1 (Vigneswaran et al., 2011; Soares et al., 2017, Betz cells), narrow and broad spike widths among excitatory pyramidal tract neurons of premotor cortex (Lemon et al., 2021), and narrow-spiking excitatory cells in layer III of V1, V2, and MT (Constantinople et al., 2009; Amatrudo et al., 2012; Onorato et al., 2019; Kelly et al., 2018).

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 (Ardid et al., 2015), repolarization time (Trainito et al., 2019; Boroujeni et al., 2021), and triphasic waveform shape (Barry, 2015; Robbins et al., 2013). Although these user-specified methods are amenable to human intuition, they are insufficient to distinguish between previously identified cell types (Krimer et al., 2005; Vigneswaran et al., 2011; Merchant et al., 2012). It is also unclear how to choose these user-specified features in a principled manner (i.e., one set 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 (Trainito et al., 2019; Viskontas et al., 2007; Katai et al., 2010; Sun et al., 2021). 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 (Weir et al., 2015).

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 (Tasic et al., 2018; Becht et al., 2018), in analyzing models of biological neural networks (Maheswaranathan et al., 2019; Kleinman et al., 2019), the identification of behavior (Bala et al., 2020; Hsu and Yttri, 2020; Dolensek et al., 2020), and in electrophysiology (Jia et al., 2019; Gouwens et al., 2020; Klempíř et al., 2020; Markanday et al., 2020; Dimitriadis et al., 2018).

Here, in a novel technique that we term *WaveMAP*, we combine a non-linear dimensionality reduction method (Universal Manifold Approximation and Projection [UMAP] (McInnes et al., 2018)) with graph community detection (Louvain community detection (Blondel et al., 2008); we colloquially call "clustering") to understand the physiological properties, decision-related dynamics, and laminar distribution of candidate 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 constructed spike features 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 significantly 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 (Chandrasekaran et al., 2017). 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 randomized 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. After another short randomized target viewing period, 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 signed color coherence and color coherence. The signed color coherence (SC) provides an estimate of whether there are more red or green squares in the checkerboard. Positive SC indicates the presence of more red squares while negative SC indicates more green squares. SC close to zero (positive or negative) indicates an almost even number of red or green squares (Fig. 1B, bottom). The coherence (C) provides an estimate of the difficulty of a stimulus. Higher coherence indicates that there is more of one color than the other (an easy trial) whereas a lower coherence indicates that the two colors are more equal in number (a difficult trial).

Our monkeys demonstrated the range of behaviors typically observed in decision-making tasks: monkeys made more errors and were slower for lower coherence checkerboards compared to higher coherence checkerboards (Fig. 1C, D). We used coherence, choice, 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 (FHC electrodes) or linear multi-contact electrodes (Plexon U-Probes, 625 neurons, 490 U-probe waveforms; Fig. 1E, right) and a Cerebus Acquisition System (Blackrock Microsystems). In this study, we analyzed the average 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

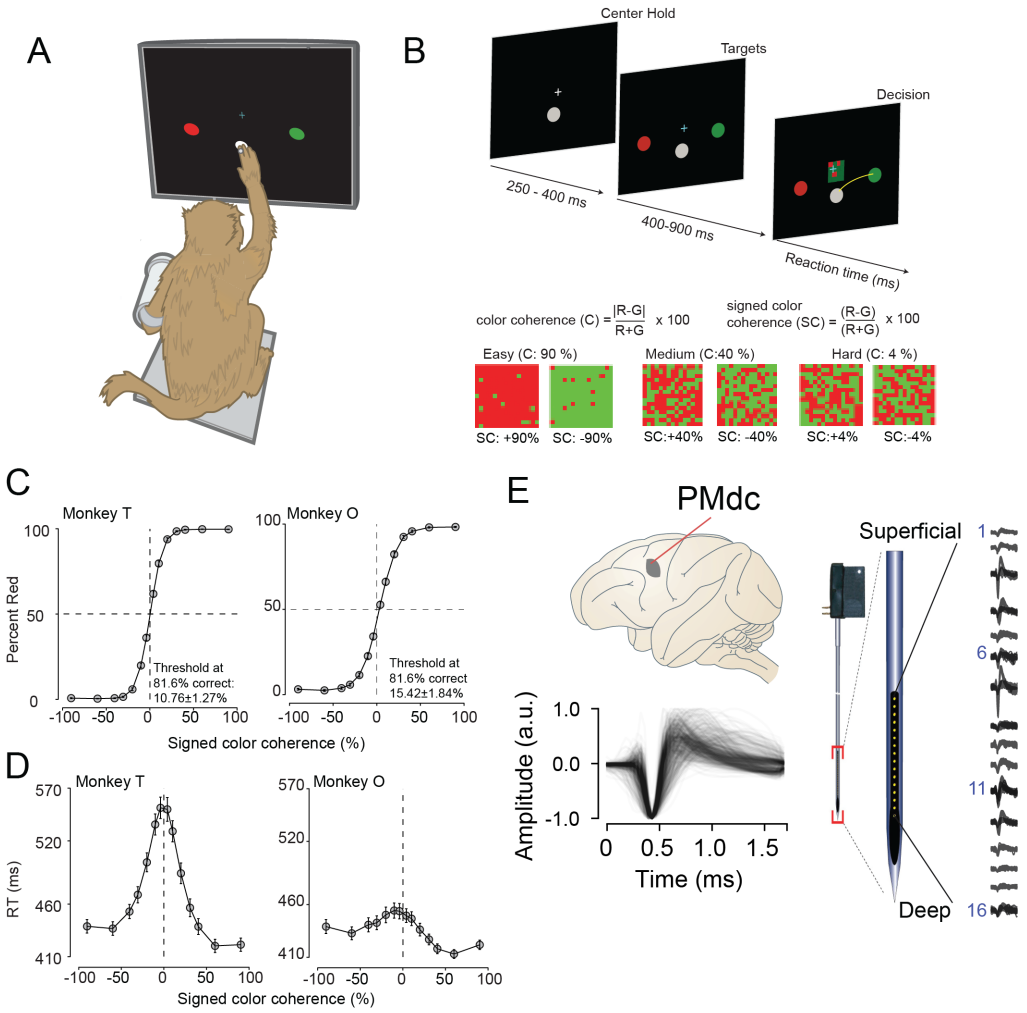

Figure 1: Recording locations, waveform shapes, techniques, task, and discrimination behavior (A) An illustration of the behavioral setup in the discrimination task. The monkey was seated with one arm free and one arm gently restrained in a plastic tube via a cloth sling. An infrared-reflecting (IR) bead was taped to the forefinger of the free hand and was used in tracking arm movements. This gave us a 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 (SC). **(C)** Average discrimination performance and **(D)** Reaction time (RT) over sessions of the two monkeys as a function of the SC 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 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-channel Plexon U-probe (right) used during the behavioral experiment.

each spike and used in these analyses, a duration longer than many studies of waveform shape (Merchant et al., 2012).

We restricted our analysis to well-isolated single neurons identified through a combination of careful online isolation combined with offline spike sorting (see Methods section: *Identification of single neurons during recordings*). Extracellular waveforms were isolated as single neurons by only accepting waveforms with minimal ISI violations ($1.5\% < 1.5$ ms). This combination of online vigilance, combined with offline analysis, provides us the confidence to label these waveforms as single neurons.

We used previously reported approaches to align, average, and normalize spikes (Kaufman et al., 2013; Snyder et al., 2016). Spikes were aligned in time via their depolarization trough and normalized between -1 and 1. "Positive spiking" units with large positive amplitude pre-hyperpolarization spikes were dropped from the analysis due to their association with dendrites and axons (Gold et al., 2009; Barry, 2015; Sun et al., 2021). 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* (Fig. 2), we use a three step strategy for the analysis of extracellular waveforms: We first passed the normalized and trough-aligned waveforms (Fig. 2A-i) into UMAP to obtain a high-dimensional graph (Fig. 2A-ii) (McInnes et al., 2018). Second, we used this graph (Fig. 2B-iii) and passed it into Louvain clustering (Fig. 2B-iv, (Blondel et al., 2008)), to delineate high-dimensional clusters. Third, we used UMAP to project the high-dimensional graph into two dimensions (Fig. 2B-v). We colored the data points in this projected space according to their Louvain cluster membership found in step two to arrive at our final *WaveMAP* clusters (Fig. 2B-vi). We also analyzed the *WaveMAP* clusters using interpretable machine learning (Fig. 2B-vii) and also an inverse transform of UMAP (Fig. 2B-viii). A detailed explanation of the steps associated with *WaveMAP* is available in the methods, and further mathematical details of *WaveMAP* are available in the supplementary information.

Fig. 3A shows how *WaveMAP* provides a clear organization without the need for prior specification of important features. For expository reasons, and to link to prior literature (McCormick et al., 1985; Connors et al., 1982), we use the trough to peak duration to loosely subdivide these eight clusters into "narrow-spiking" and "broad-spiking" cluster sets. The broad-spiking clusters had a trough to peak duration of 0.74 ± 0.24 ms (mean $\pm$ S.D.) and the narrow-spiking clusters had a trough to peak duration of 0.36 ± 0.07 ms (mean $\pm$ S.D.). The narrow-spiking neurons are shown in warm colors (including green) at right in Fig. 3A and the broad-spiking neurons are shown in cool colors at left in the same figure. The narrow-spiking set 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) respectively of the total waveforms, for ∼68% of total waveforms. The broad-spiking set 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.

The number of clusters identified by *WaveMAP* is dependent on the resolution parameter for Louvain clustering. A principled way to choose this resolution parameter is to use the modularity score (a measure of how tightly interconnected the members of a cluster are) as the objective function to maximize. We chose a resolution parameter of 1.5 that maximized modularity score

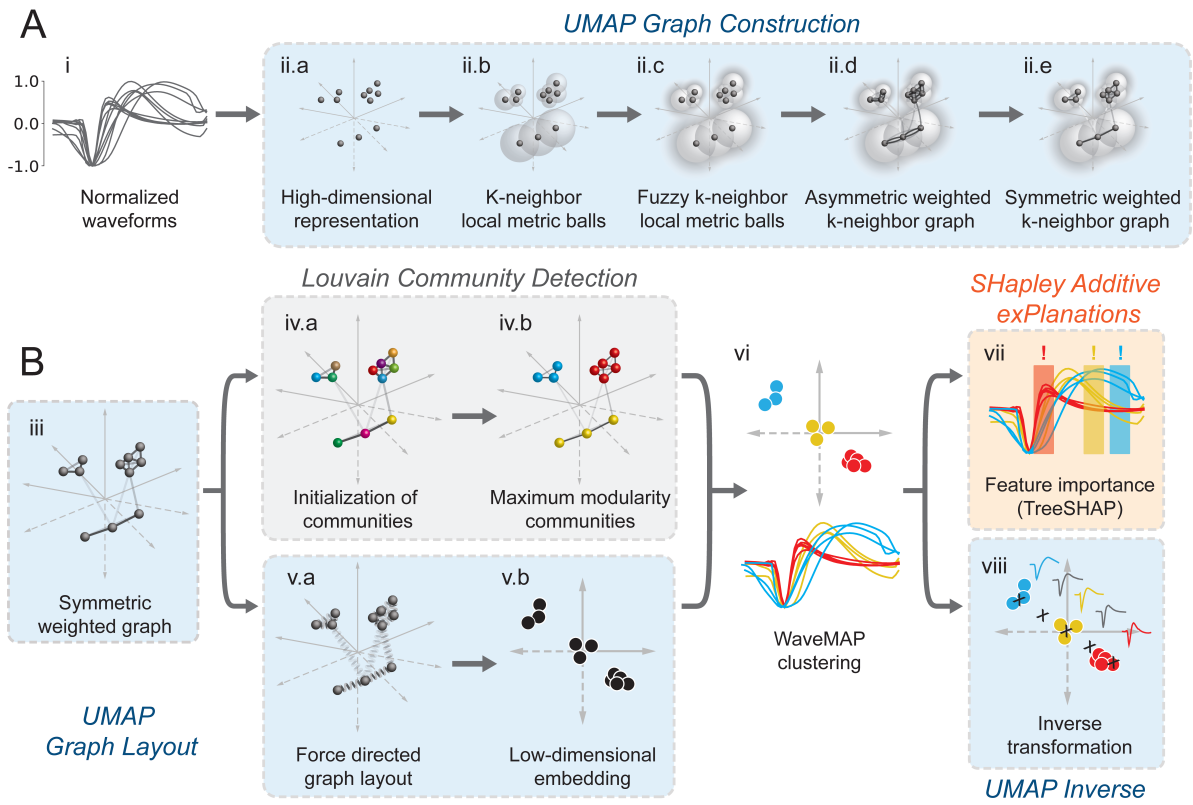

Figure 2: Schematic of *WaveMAP* (A) *WaveMAP* begins with UMAP which projects high-dimensional data into lower dimension while preserving local and global relationships (see Fig. S1A for an intuitive diagram). Normalized average waveforms from single units (**i**) are passed to UMAP (McInnes et al., 2018) which begins with the construction of a high-dimensional graph (**ii**). In the high-dimensional space (**ii.a**), UMAP constructs a distance metric local to each data point (**ii.b**). The unit ball (ball with radius of one) of each local metric stretches to the 1st-nearest neighbor. Beyond this unit ball, local distances decrease (**ii.c**) according to an exponential distribution that is scaled by the local density. This local metric is used to construct a weighted graph with asymmetric edges (**ii.d**). The 1-nearest neighbors are connected by en edge of weight 1.0. For the next $k-1$-nearest neighbors, this weight then falls off according to the exponential local distance metric (in this diagram k $= 4$ with some low weight connections omitted for clarity). These edges, a and b, are made symmetric according to $a + b - a \cdot b$ (**ii.e**). **(B)** The high-dimensional graph (**iii**) captures latent structure in the high-dimensional space. We can use this graph in Louvain community detection (Louvain, **iv**) (Blondel et al., 2008) to find clusters (see Fig. S1B for an intuitive diagram). In Louvain, each data point is first initialized as belonging to its own "community" (**iv.a**, analogous to a cluster in a metric space). Then, in an iterative procedure, each data point joins neighboring communities until a measure called "modularity" is maximized (**iv.b**, see Supplemental Information for a definition of modularity). Next, data points in the same final community are aggregated to a single node and the process repeats until the maximal modularity is found on this newly aggregated graph. This process then keeps repeating until the maximal modularity graph is found and the final community memberships are passed back to the original data points. We can also use this graph to find a low-dimensional representation through a graph layout procedure (**v**). The graph layout proceeds by finding a "low energy" configuration that balances attractive (shown as springs in **v.a**) and repulsive (not shown) forces between pairs of points as a function of edge weight or lack thereof. This procedure iteratively minimizes the cross-entropy between the graph in low dimension and the high-dimensional graph (**v.b**). The communities found through Louvain are then combined with the graph layout procedure to arrive at a set of clusters in a low-dimensional embedded space (**vi**). These clusters (**vi, top**) can be used to classify the original waveforms (**vi, bottom**). To investigate "why" these data points became clusters, each cluster is examined for locally (within-cluster) important features (SHAP (Lundberg and Lee, 2017), **vii**) and globally important trends (UMAP inverse transform, **viii**). Not shown is the classifier SHAP values are calculated from. The diagrams for the graph construction and layout are based on UMAP documentation and the diagram for Louvain community detection is based on Blondel et al. (2008).

while ensuring that we did not overly fractionate the dataset ($n < 20$ within a cluster; Fig. 3A, B, and columns of Fig. S2A).Additional details are available in the "Parameter Choice" section of the Supplementary Information.

Louvain clustering with this resolution parameter of 1.5 identified eight clusters in total (Fig. 3A). Note, using a slightly higher resolution parameter (2.0), a suboptimal solution in terms of modularity, led to seven clusters (Fig. S2A). The advantage of Louvain clustering is that it is hierarchical and choosing a slightly larger resolution parameter will only merge clusters rather than generating entirely new cluster solutions. Here, we found that the higher resolution parameter merged two of the broad-spiking clusters ⑥ and ⑦ while keeping the rest of the clusters largely intact and more importantly, did not lead to material changes in the conclusions of analyses of physiology, decision-related dynamics, or laminar distribution described below. Finally, an alternative ensembled version of the Louvain clustering algorithm (ensemble clustering for graphs [ECG] (Poulin and Théberge, 2018)), which requires setting no resolution parameter, produced a clustering almost exactly the same as our results (Fig. S2C).

To validate that *WaveMAP* finds a "real" representation of the data, we examined if a very different method could learn the same representation. We trained a gradient boosted decision tree classifier (with a softmax multi-class objective) on the exact same waveform data (vectors of 48 time points, 1.6 ms time length) passed to *WaveMAP* and used a test-train split with k-fold cross validation applied to the training data. Hyperparameters were tuned with a five-fold cross-validated grid search on the training data and final parameters shown in Table S1. After training, the classification was evaluated against the held-out test set (which was never seen in model training/tuning) and the accuracy, averaged over clusters, was 91%. Fig. 3C 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 levels of accuracy was only possible because there were "generalizable" clusterings of similar waveform shapes in the high-dimensional space revealed by UMAP.

We find that cluster memberships found by *WaveMAP* are stable with respect to random seed when `resolution parameter` and `n_neighbors` parameter are fixed. This stability of *WaveMAP* clusters with respect to random seed is because much of the variability in UMAP layout is the result of the projection process (Fig. 2B-v.a). Louvain clustering operates before this step on the high-dimensional graph generated by UMAP which is far less sensitive to the random seed. Thus, the actual layout of the projected clusters might differ subtly according to random seed, but the cluster memberships largely do not (see Supplementary Information and columns of Fig. S2A). Here, we fix the random seed purely for visual reproducibility purposes in the figure. Thus, across different random seeds and constant resolution, the clusters found by *WaveMAP* did not change because the graph construction was consistent across random seed at least on our dataset (Fig. S2A).

We also found that *WaveMAP* was robust to data subsetting (randomly sampled subsets of the full dataset, see Supplementary Information (Tibshirani and Walther, 2012)), unlike other clustering approaches (Fig. S2B, green, Fig. S3). We applied *WaveMAP* to 100 random subsets each from 10% to 90% of the full dataset and compared this to a "reference" clustering produced by the procedure on the full dataset. *WaveMAP* was consistent in both cluster number (Fig. S2B, red) and cluster membership (which waveforms were frequently "co-members" of the same cluster;

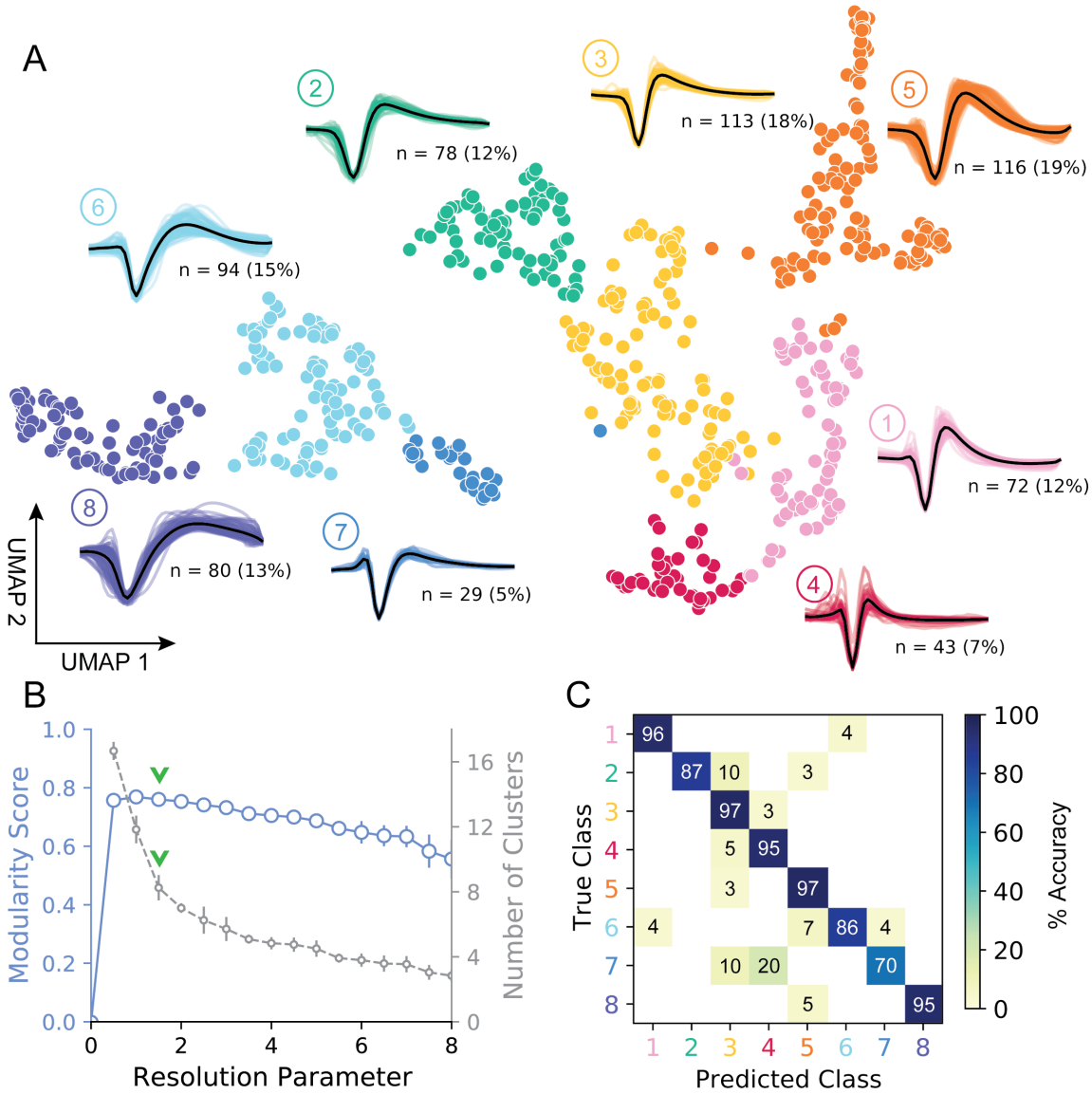

Figure 3: UMAP and Louvain clustering reveal a robust diversity of averaged single-unit waveform shapes (A) Scatter plot of normalized EAP waveforms in UMAP space colored by Louvain cluster membership. Adjacent to each numbered cluster (① through ⑧) is shown all member waveforms and the average waveform 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, axis at left) and the number of clusters (communities) found (in gray, axis at right). This was averaged over 25 runs for *WaveMAP* using 25 random samples and seeds of 80% of the full dataset at each resolution parameter from 0 to 8 in 0.5 unit increments (a subset of the data was used to obtain error bars). 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 gradient boosted decision tree classifier with five-fold cross-validation and hyperparameter optimization. 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 across all clusters was 91%.

Fig. S2B, green).

Finally, our results were also robust to another standard approach to normalizing spike waveforms: normalization to trough depth. This method exhibited the same stability in cluster number (Fig. S4C), and also showed no differences in downstream analyses (Fig. S4D). Without amplitude normalization, interesting structure was lost (Fig. S4E) because UMAP likely attempts to explain both waveform amplitude and shape (shown as a smooth gradient in the trough to peak height difference Fig. S4F). In addition, common recommendations to apply PCA before non-linear dimensionality reduction were not as important for our waveform dataset, which was fairly low-dimensional (first three PC's explained 94% variance). Projecting waveforms into a 3-dimensional PC-space before *WaveMAP* produced a clustering very similar to data without this step (Fig. S5).

Traditional clustering methods with specified features sub-optimally capture waveform diversity

Our unsupervised approach (Fig. 3) generates a stable clustering of waveforms. However, is our method better than the traditional approach of using specified features (Snyder et al., 2016; Trainito et al., 2019; Kaufman et al., 2010, 2013; Mitchell et al., 2007; Barthó et al., 2004; Merchant et al., 2008, 2012; Song and McPeek, 2010; Simons, 1978; Johnston et al., 2009)? 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 commonly used waveform features. In accordance with previous work, the features we chose (Fig. 4A) were action potential (AP) width of the spike (width in milliseconds of the full-width half minimum of the depolarization trough (Vigneswaran et al., 2011)); the peak ratio (the ratio of pre-hyperpolarization peak (A1) to the post-hyperpolarization peak (A2) (Barry, 2015)); 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 analyses of extracellular recordings (Snyder et al., 2016; Kaufman et al., 2013; Merchant et al., 2012).

The result of the GMM applied to these three measures is shown in Fig. 4B. This method identified four waveform clusters that roughly separated into broad-spiking (BS, $\sim$33%, n = 208), narrow-spiking (NS, $\sim$43%, n = 269), broad-spiking triphasic (BST, $\sim$9%, n = 55), and narrow-spiking triphasic (NST, $\sim$15%, n = 93) (Fig. 4B). Triphasic waveforms, thought to be neurons with myelinated axons or neurites (Barry, 2015; Robbins et al., 2013; Deligkaris et al., 2016; Bakkum et al., 2013; Sun et al., 2021), contain an initial positive spike before the trough and can be identified by large peak ratios (Fig. 4A). These GMM clusters are similar to those obtained from other clusterings of EAP's in macaque cortex (Gur et al., 1999; Trainito et al., 2019). We selected four clusters by examining the Bayesian information citerion (BIC) statistic as a function of the number of clusters and identified the cluster number at the elbow (green chevron in Fig. 4C).

To compare the generalizability of this representation with the representation provided by UMAP, we trained the same decision tree classifier on the waveform data (after separate hyperparameter tuning, Table S1) but this time using the four GMM classes as target labels. After training, the accuracy across all four classes averaged $\sim$78% with no classification accuracy over 95% and misclassifications between every class (Fig. 4D). 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

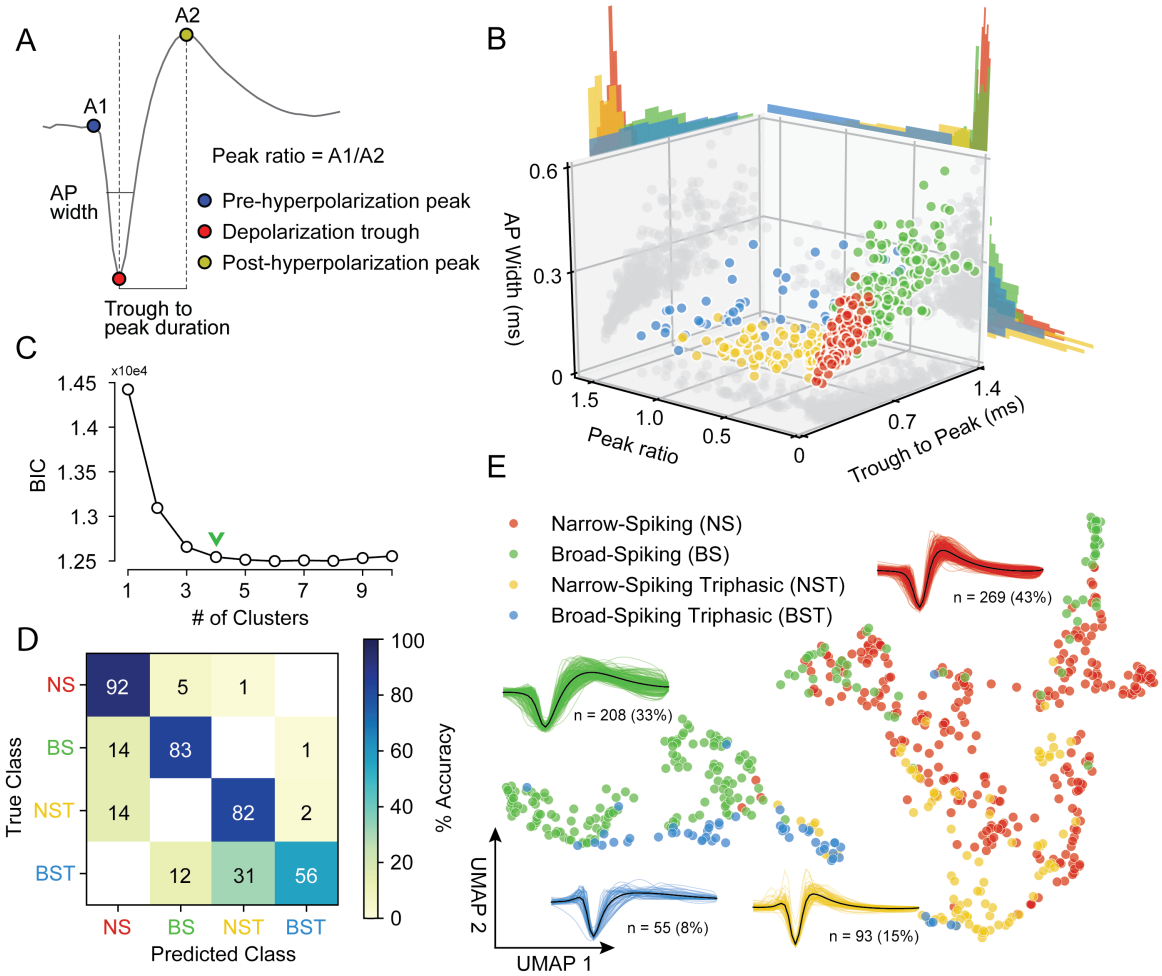

Figure 4: Gaussian mixture model clustering on specified features fails to capture the breadth of waveform diversity (A) The three EAP waveform landmarks used to generate the specified features passed to the GMM on a sample waveform. ● is the pre-hyperpolarization peak (A1); ● is the depolarization trough; 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 (Stuttgen, 2019). **(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) (Trainito et al., 2019). 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 decision tree classifier with five-fold cross-validation with hyperparameter optimization. The main diagonal contains the classification accuracy percentages across the four GMM clusters and the off-diagonal contains the misclassification rates. The average accuracy across classes was 78%. **(E)** The same scatter plot of normalized EAP waveforms in UMAP space as in Fig. 3A but now colored by GMM category.

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 latent structure of the data. If we examine the gray data point shadows (Fig. 4B), 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. 4B, none of the features induce a clear separability between the clusters alone or in conjunction. Furthermore, the reproducible clusters found by *WaveMAP* are linearly inseparable in the feature space of the three GMM features (Fig. S6A). 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. S6B), a classifier is still unable to generalize this clustering with even modest accuracy (average of 56% across clusters; Fig. S6C) even if the waveforms shapes found by the GMM with eight clusters seem somewhat sensible (Fig. S6D). In fact, across all cluster numbers (`n_components` from 2 to 16), a classifier tuned *for the GMM* performed more poorly on the GMM labels than a *WaveMAP* projection with the same number of clusters (Fig. S6E, in red). Tuning *WaveMAP* parameters that induce different cluster numbers, whether `n_neighbors` (in dark blue) or `resolution` (in light blue), had little effect on classifier performance (Fig. S6E, in blues). *WaveMAP* yielded mappings that were more generalizable than a GMM on features across every number of clusters and both parameters investigated. Thus, it is a deficit of the GMM on constructed feature-based approach to capture the full diversity of waveforms, especially at high cluster number, and not a peculiarity of the model parameters chosen or number of clusters induced.

We also investigated the representation of specified features in the projected UMAP space. We color coded the waveforms in UMAP, in Fig. S7, according to each point's feature values using the same features as in Fig. 4 (Fig. S7A): AP width (Fig. S7B), trough to peak duration (Fig. S7C), and peak ratio (Fig. S7D). We find that *WaveMAP* implicitly captures each of these specified features shown as a smooth gradient of values. Our method also exposes the correlation between certain specified features: the gradient between trough to peak duration and AP width points point roughly in the same direction so thus both features are highly correlated. This correlation between features exposes their redundancy and is another reason why traditional approaches fail to capture the full diversity of waveform shapes.

To obtain a clearer picture of how *WaveMAP* captures latent structure missed by specified features, we color the points in UMAP space by their GMM cluster identity in Fig. 4E. Here, *WaveMAP* is able to recapitulate the same structure observed by specified features as a gradient from broad- to narrow-spiking along the UMAP-1 direction. Our technique also captures the transition from triphasic to biphasic along the UMAP-2 direction. *WaveMAP* is also able to find clusters that occupy an intermediate identity between GMM classes. For instance, *WaveMAP* cluster ② (Fig. 3A) is nearly equal parts broad- and narrow-spiking in the GMM clustering (Fig. 4E). 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 non-linear methods is that they are uninterpretable. Here using an interpretable machine learning approach, we show that *WaveMAP* produces sensible results (Molnar, 2020; Azodi et al., 2020). 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 leverage the gradient boosted decision tree classifier in Fig. 3C and used a decision tree implementation (path-dependent TreeSHAP (Lundberg et al., 2018)) of SHapley Additive exPlanations (SHAP values (Lundberg and Lee, 2017; Lundberg et al., 2020)) to reveal what waveform features are implicitly used to differentiate clusters.

To quantify the differences between Louvain clusters, we applied a grid of "test points" to the UMAP projected space (Fig. 5A, 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. 5A, bottom). On the bottom of Fig. 5A 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 rate (proportional to trough to peak width) and hyperpolarization slope inflection (proportional to repolarization time) are separate but highly informative properties (Trainito et al., 2019; Ardid et al., 2015). 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. Repolarization rates are goverened by potassium channel dynamics and may play an important in waveform shape (Soares et al., 2017). Thus, *WaveMAP* not only finds an interpretable and smoothly varying low-dimensional space it also offers biological insights; in this case how cell types might differ according to channel protein expression and dynamics.

In Fig. 5B, we made use of SHAP values to identify which aspects of waveform shape the gradient boosted decision tree classifier utilizes in assigning what waveform to which cluster (Lundberg and Lee, 2017; Lundberg et al., 2020). SHAP values build off of the game theoretic quantity of Shapley values (Shapley, 1988; Štrumbelj and Kononenko, 2014) 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 (Lundberg and Lee, 2017). Fig. 5B shows the top-10 time points in terms of mean absolute SHAP value (colloquially called "SHAP value") 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(Fig. 5C).

In the inset is shown all waveforms along with each of the top ten time points (in gold) 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

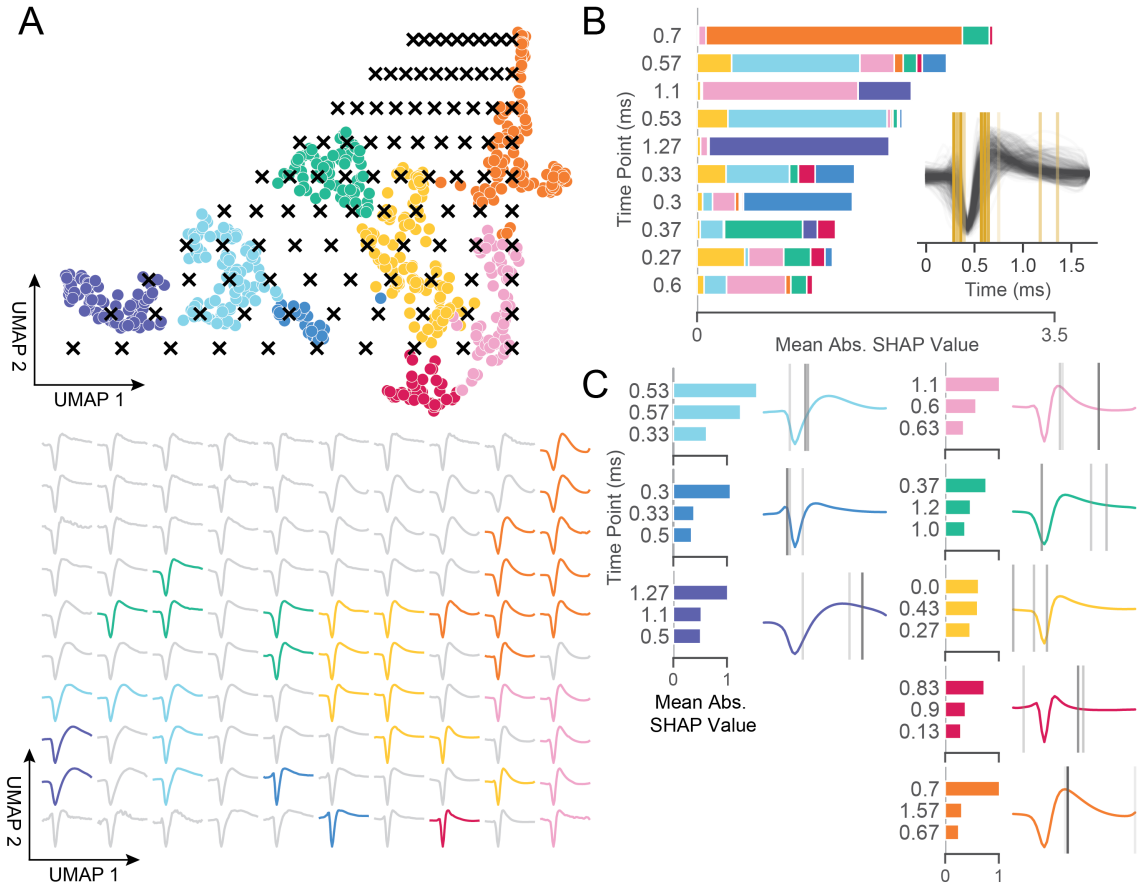

Figure 5: *WaveMAP* provides interpretable representations that both validate and extend known and unknown features importances (A) *WaveMAP* applied to the EAP's as in Fig. 3A but overlaid with a grid of test points (black **x**'s, top) spanning the embedded space. At bottom, the inverse UMAP transform is used to show the predicted waveform at each test point. For each **x** above, the predicted waveform is shown, 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 embedded space. **(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 applying path-dependent TreeSHAP to a gradient boosted decision tree classifier trained on the waveforms with the *WaveMAP* clusters as labels. In the inset, all waveforms are shown and in gold 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 SHAP value across *WaveMAP* clusters; the least opaque, the smallest SHAP value. **(C)** Each averaged *WaveMAP* 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 also shown as a bar graph at left.

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. (Vigneswaran et al., 2011)). 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 (Barry, 2015). Importance is also placed on the location at 0.6 ms corresponding to the inflection point which is similarly noted as being informative (Trainito et al., 2019; Boroujeni et al., 2021). 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. 5C, 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 SHAP 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 could give credence that *WaveMAP* clusters connect to cell types.

WaveMAP clusters have distinct physiological properties

A defining aspect of cell types is that they vary in their physiology and especially firing rate properties (Mountcastle et al., 1969; McCormick et al., 1985; Connors and Gutnick, 1990; Nowak et al., 2003; Connors et al., 1982; Contreras, 2004). However, these neuronal characterizations via waveform *ex vivo* are not always conserved when the same waveform types are observed *in vivo* during behavior (Steriade, 2004; Steriade et al., 1998). To connect our waveform clusters to physiological cell types *in vivo* 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. 6A) and narrow-spiking (Fig. 6B) 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 FR, max FR, and FR range.

Baseline FR: Cell types are thought to demonstrate different baseline FRs. We estimated baseline FR (Fig. 6C) 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. 6A),

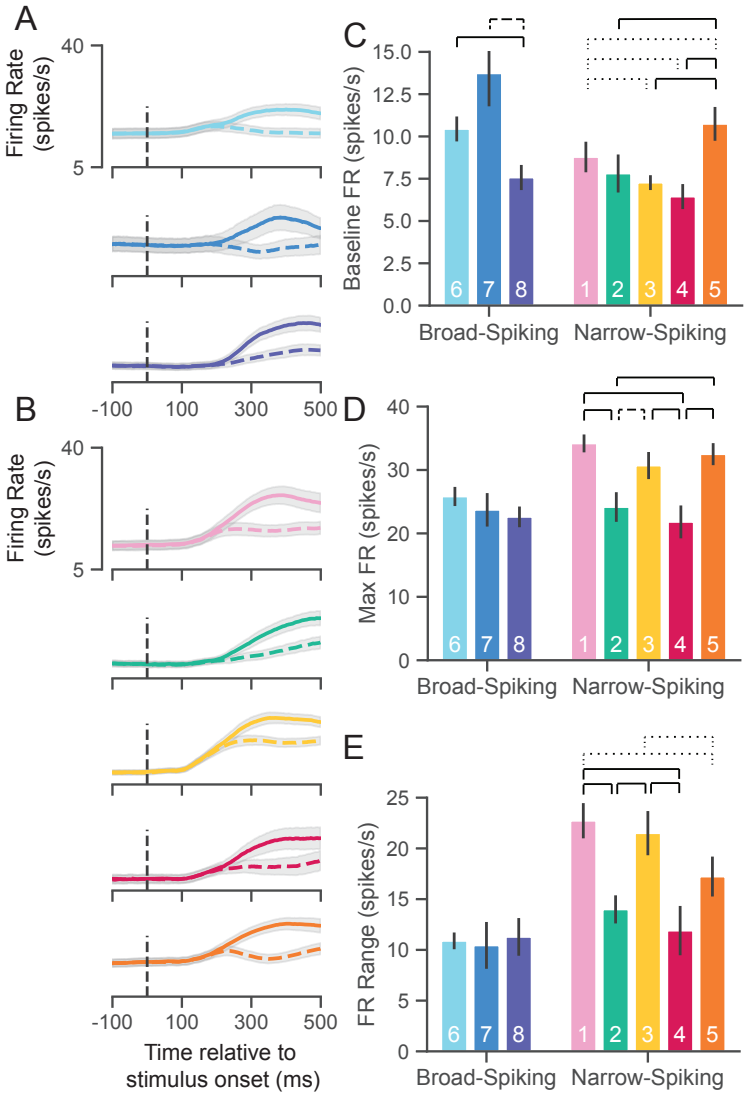

Figure 6: UMAP clusters exhibit distinct physiological properties (A) Stimulus-aligned trial-averaged firing rate (FR; spikes/s) activity in PMd for broad-spiking *WaveMAP* clusters. The traces shown are separated into trials for PREF direction reaches (solid traces) and NONPREF direction reaches (dashed traces) and across the corresponding *WaveMAP* clusters. Shaded regions correspond to bootstrapped standard error of the mean. Dashed vertical line is stimulus-onset time. **(B)** The same plots as in (A) but for narrow-spiking *WaveMAP* clusters. **(C)** Baseline median FR $\pm$ S.E.M. for the neurons in the eight different classes. Baselines were taken as the average FR from 200 ms of recording before checkerboard stimulus onset. **(D)** Median maximum FR $\pm$ S.E.M. for the neurons in the eight different clusters. This was calculated by taking the median of the maximum FR for each neuron across the entire trial. **(E)** Median FR range $\pm$ 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.

and narrow-spiking clusters (Fig. 6B). The broad-spiking clusters showed significant differences in baseline FR when compared against the 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 (Kaufman et al., 2010). We also found that not all broad-spiking neurons had low baseline FR and not all narrow-spiking neurons had high baseline FR.

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 ± bootstrap S.E.]; $p = 0.0052$, $p = 0.0029$ respectively, Mann-Whitney *U* test, FDR adjusted). The narrow-spiking clusters (Fig. 6B, 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 ± bootstrap S.E.) and were not significantly different from one another but all were significantly different from ① and ⑤ ($p = 0.04$, $p = 2.8e\text{-}4$, $p = 2.8e\text{-}7$, $p = 4.9e\text{-}5$ respectively, Mann-Whitney *U* test, FDR adjusted; see Fig. 6C).

Maximum FR: A second important property of cell types is their maximum FR (Mountcastle et al., 1969; McCormick et al., 1985; Connors and Gutnick, 1990). 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. 6D). In addition to significant differences in baseline FR, broad- vs. narrow-spiking cl clusters showed a significant difference in max FR ($p = 1.60e\text{-}5$, Mann-Whitney *U* test). Broad-spiking clusters were fairly homogeneous with low median max FR (24.3 ± 1.0, median ± 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 ± bootstrap S.E.) while two others (② and ④) were uniformly lower in max FR (23.0 ± 1.4, median ± 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 FR and also span a wide range during behavior (Kaufman et al., 2010, 2013; Chandrasekaran et al., 2017; Johnston et al., 2009; Hussar and Pasternak, 2009). 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 broad- and 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 showed more variation in their FR range—similar to the pattern observed for max FR. ①, ③, and ⑤ had a large FR range (20.3 ± 1.1 spikes/s, median ± bootstrap S.E.) and the clusters ② and ④ had a relatively smaller FR range (13.4 ± 1.3 spikes/s, median ± bootstrap S.E.). These results demonstrate that some narrow-spiking clusters, in addition to having high baseline FR, highly modulated their FR over the course of a behavioral trial.

Such physiological heterogeneity in narrow-spiking cells has been noted before (Ardid et al., 2015; Boroujeni et al., 2021; Quirk et al., 2009) and in some cases attributed to different subclasses of a single inhibitory cell type (Povysheva et al., 2013; Zaitsev et al., 2009). Other work also strongly suggests that narrow-spiking cells contain excitatory neurons with distinct FR properties contributing to this diversity (Vigneswaran et al., 2011; Onorato et al., 2019).

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. 6 but with the GMM clusters (Fig. S8). 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 fewer groups are compared (Fig. S8B, C, and D). We also quantitatively estimate 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 FR (0.070 vs. 0.013), maximum FR (0.035 vs. 0.011), and FR range (0.055 vs. 0.034).

WaveMAP 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 argues that neuronal cell types have distinct functional roles during cortical computation with precise timing. For instance, studies of macaque premotor (Song and McPeek, 2010), inferotemporal (IT) (Mruczek and Sheinberg, 2012), and frontal eye field (FEF) (Ding and Gold, 2012) 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 (Pinto and Dan, 2015; Estebanez et al., 2017). Here we examine the functional properties of each cluster based on two inferred statistics: choice-related dynamics and discrimination time.

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 (Chandrasekaran et al., 2017; Ding and Gold, 2012; Roitman and Shadlen, 2002). 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. 7A and B, we show average choice-selectivity signals across the seven color coherence levels (Fig. 7A, legend) for an example broad- (⑥) and narrow-spiking cluster (①). For ⑥ (Fig. 7A), easier stimuli (higher coherence) only led to modest increases in the rate at which the choice selectivity signal increases. In contrast, ① (Fig. 7B) 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. 7A,B). This rate of rise for the choice-selective signal (spikes/s/s) vs. coherence is shown for broad- (Fig. 7C) and narrow-spiking (Fig. 7D) 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.

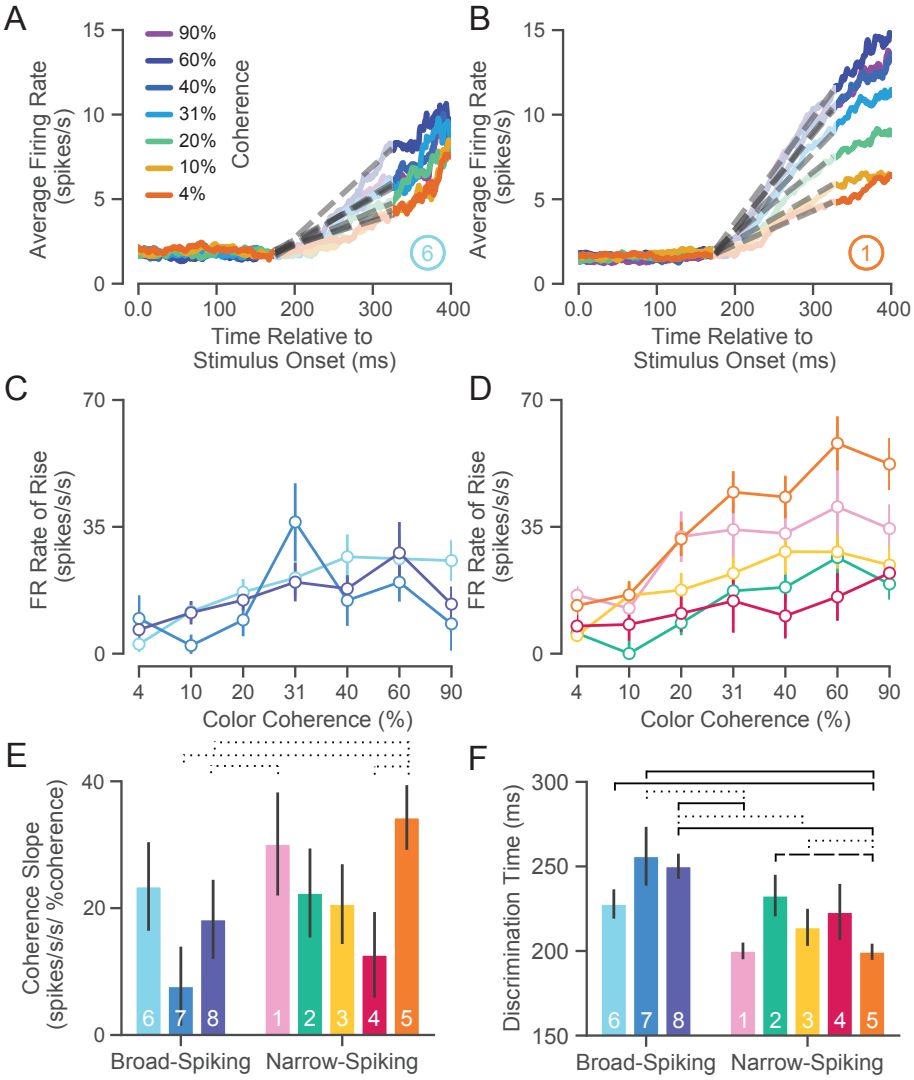

Figure 7: UMAP clusters exhibit distinct functional properties (A) Average firing rate (FR) over time for ⑥ (used as a sample broad-spiking cluster) across trials of different color coherences. The gray dashed lines indicate the linear regression lines used to calculate the FR rate of rise. **(B)** Average FR over time for ① (used as a sample narrow-spiking cluster) across different color coherences. **(C)** FR rate of rise vs. color coherence for broad- and **(D)** narrow-spiking clusters. Error bars correspond to standard error of the mean across trials. **(E)** Bootstrapped median color 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. 7C and D. **(F)** Median bootstrapped discrimination time for each cluster with error bars as the bootstrapped standard error of the median. Discrimination time was calculated as the the amount of time after checkerboard appearance at which the choice-selective signal could be differentiated from the baseline FR (Chandrasekaran et al., 2017). ····· $p < 0.05$; - - - - $p < 0.01$; —— $p < 0.005$; Mann-Whitney U test, FDR adjusted.

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. color coherence for each cluster (Fig. 7E). 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. 7C and 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 (Chandrasekaran et al., 2017)). 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 (Hodge et al., 2019; Tosches et al., 2018). 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 approximately 2.5 mm in depth from the pial surface to white matter (Arikuni et al., 1988). 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. 8A and B are the laminar distributions for all waveforms in the associated set of clusters (in gray); below these are the laminar distributions for each cluster set's constituent clusters. On the right (Fig. 8C), 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 cluster set of neurons (⑥, ⑦, and ⑧) are generally thought to contain cortical excitatory pyramidal neurons enriched in middle to deep layers (Nandy et al., 2017; McCormick et al., 1985). Consistent with this view, we

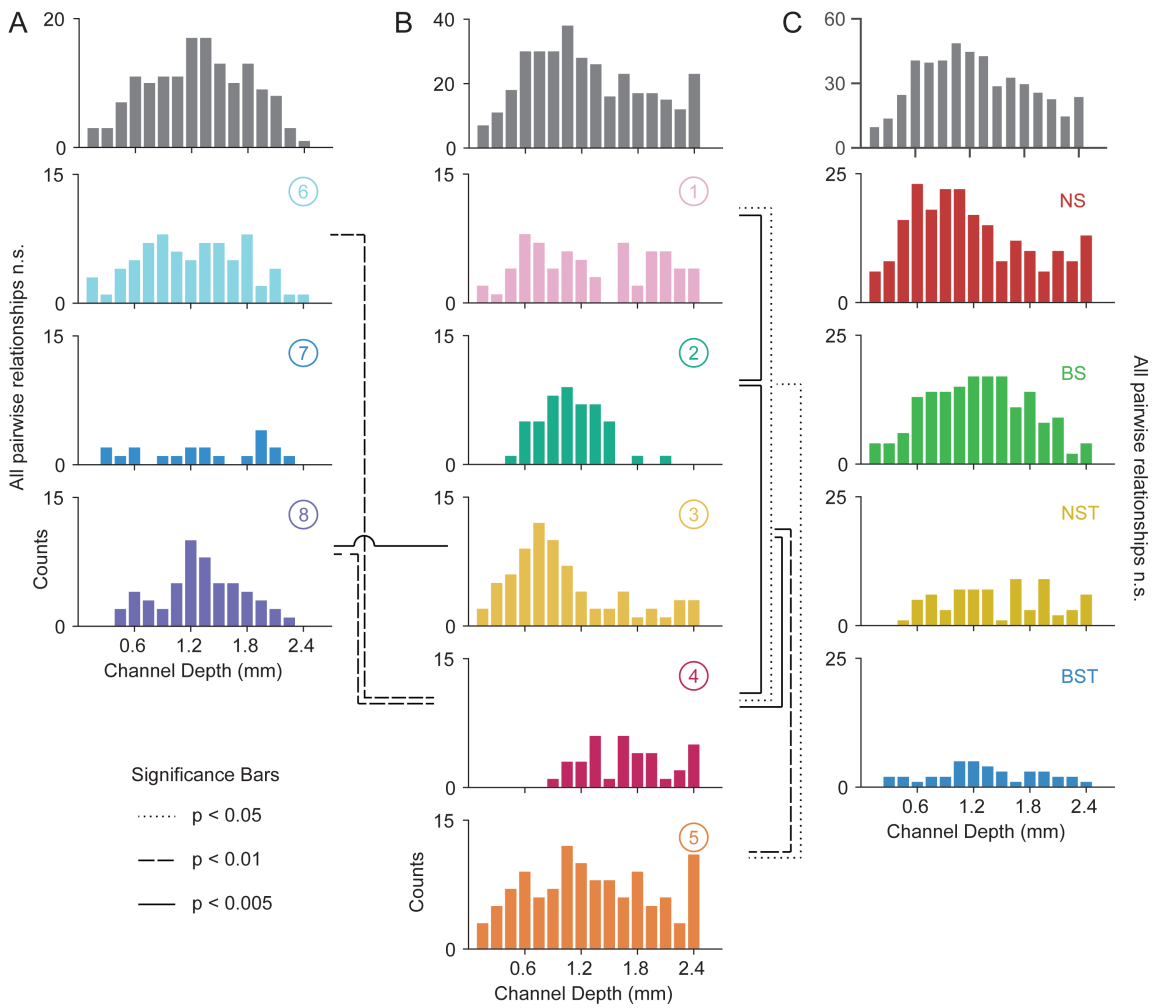

Figure 8: Laminar distribution of *WaveMAP* waveform clusters (A, B) The overall histogram for the broad- and narrow-spiking waveform clusters are shown at top across cortical depths on the left and right respectively (in gray); below are shown histograms for their constituent *WaveMAP* clusters. These waveforms are shown sorted by the cortical depth at which they were recorded from the (0.0 mm (presumptive 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.

found these broad-spiking clusters (Fig. 8A) 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. 8B) 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.).

Cluster ④ was 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 (Barry, 2015) which are especially dense in this caudal region of PMd (Barbas and Pandya, 1987).

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 (Arikuni et al., 1988).

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. 8C; 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 understanding of cell types using histology (discussed in the next section).

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) (DeFelipe, 1997). 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. 9A and B) (Schmitz et al., 2014) and compared it to the distributions in Fig. 8.

Both CB^+ and CR^+ cells (Fig. 9C 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 (Zaitsev et al., 2005). 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. 9C and D) and similar physiology (Zaitsev et al., 2005).

In contrast, *WaveMAP* cluster ①, had laminar properties consistent with PV^+ neurons (Fig. 9B): both were concentrated superficially but proliferated into middle layers (Fig. 9E). In addition, there were striking physiological and functional similarities between ① and PV^+ cells. In particular, both ① and PV^+ cells have low baseline FR, early responses to stimuli and robust modulation of FR similar to PV^+ cells in mouse M1 (Estebanez et al., 2017). Cluster ⑤ also had similar properties to ① and could also correspond to PV^+ cells.

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

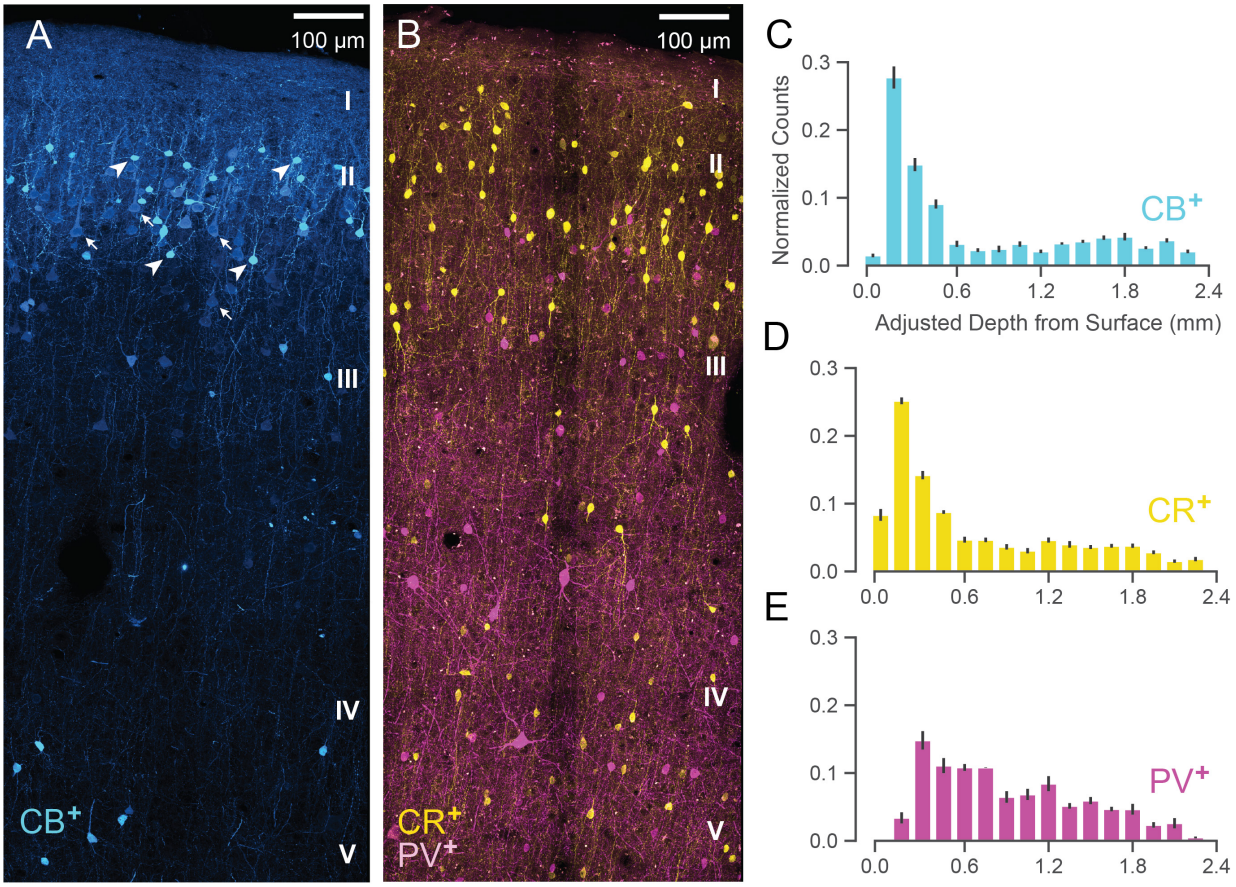

Figure 9: Anatomical labeling of three inhibitory interneuron types in PMd: **(A)** Sample maximum intensity projection of immunohistological (IHC) staining of rostral PMd calbindin-positive (CB^+) interneurons in blue. Note the many weakly-positive excitatory pyramidal neurons (arrows) in contrast to 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 the 2.4 mm (approximately the top of white matter). Layer IV exists as a thin layer in this area. Layer divisions were estimated based on depth and referencing Arikuni et al. (1988). **(B)** Sample maximum intensity projection of IHC staining of PMd calretinin-positive (CR^+) and parvalbumin-positive (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 (Schmitz et al., 2014) (mean $\pm$ S.D.) of CB^+, CR^+, PV^+ cells in PMd respectively. Counts were collected from six specimens, each with all three IHC stains, and with counts normalized to each sample.

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 over and above previous methods (Chandrasekaran et al., 2017). Heterogeneity in decision-related activity can emerge from cortical depth, different cell types within each layer, or both. To quantify the relative contributions of *WaveMAP* clusters and cortical depth, we regressed discrimination time on both separately and together and examined the change in variance explained (adjusted R^2). We then compared this against the GMM clusters with cortical depth 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 (Chandrasekaran et al., 2017). 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 also explained 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.

In contrast, 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* clustering explains a much larger variance relative to cortical depth alone. This demonstrates 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, *WaveMAP* clusters 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 us better understand decision-making circuits.

Discussion

Our goal in this study was to further understand the relationship between waveform shape and the physiology, function , and laminar distribution 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 that 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. Finally, we show that the candidate cell classes identified by *WaveMAP* have distinct physiological properties, decision-related dynamics, and laminar distribution.

WaveMAP combines UMAP with high-dimensional graph clustering and interpretable machine learning to better identify candidate cell classes. Our approach might also be useful in other domains that employ non-linear dimensionality reduction such as computational ethology (Ali et al., 2019; Hsu and Yttri, 2020; Bala et al., 2020), analysis of multi-scale population structure (Diaz-Papkovich et al., 2019), and metascientific analyses of the literature (Noichl, 2019). We also note that while traditional uses of non-linear dimensionality reduction and UMAP has been to data lacking autoregressive properties, such as transcriptomic expression (Becht et al., 2018), this doesn't seem to be an issue for *WaveMAP*. Even though our waveforms have temporal autocorrelation, our method still is able to pick out interesting structure. Other work has found

similar success in analyzing time series data with non-linear dimensionality reduction (Sedaghat-Nejad et al., 2021; Dimitriadis et al., 2018; Jia et al., 2019; Gouwens et al., 2020; Ali et al., 2019).

Advantages of WaveMAP over traditional methods

At the core of *WaveMAP* is UMAP which has some advantages over other non-linear dimensionality reduction methods that have been applied in this context. Although most algorithms offer fast implementations that scale well to large input dimensionalities and volumes of data (Linderman et al., 2019; Nolet et al., 2020), UMAP also projects efficiently into arbitrary *output* dimensionalities while also returning an invertible transform. That is, we can efficiently project new data into any arbitrary dimensional projected space without having to recompute the mapping.

These properties provide three advantages over other non-linear dimensionality reduction approaches: First, our method is stable in the sense that it produces a consistent number of clusters and each cluster has the same members across random subsamples (Fig. S2B). Clustering in the high-dimensional space rather than the projected space lends stability to our approach. Second, it allows exploration of any region of the projected space no matter the intuited latent dimensionality—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. Third, it enables cross-validation of a classifier trained on cluster labels, impossible with methods that don't return an invertible transform. To cross-validate unsupervised methods, unprocessed test data must be passed into a transform computed *only* on training data and evaluated with some loss function (Moscovich and Rosset, 2019). This is only possible if an invertible 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 some other methods do not (Sainburg et al., 2020). 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") (Bellet et al., 2013; Yang and Jin, 2006). Experimentalists could theoretically pass biological ground truths to *WaveMAP* as training labels and "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 (Roux et al., 2014; Deubner et al., 2019; Jia et al., 2019; Cohen et al., 2012; Hangya et al., 2015), 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 also be useful 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 (Trautmann et al., 2019; Jun

et al., 2017)). We believe this is a rich area that can be explored in future work.

WaveMAP uses 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 interpretable machine learning (Shapley, 1988; Lundberg and Lee, 2017), 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 recombined 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 (Shenoy et al., 2013; Churchland et al., 2012; Mante et al., 2013; Remington et al., 2018; Wang et al., 2017). 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 (Georgopoulos et al., 1986) and visual cortex (Hubel and Wiesel, 1959). 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 (Bruce and Goldberg, 1985) and premotor cortex (Chandrasekaran et al., 2017). 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 (Shenoy et al., 2013; Churchland et al., 2012; Cunningham and Byron, 2014), tensor component analysis (Williams et al., 2018), demixed principal component analysis (Kobak et al., 2016), targeted dimensionality reduction (Mante et al., 2013), and autoencoder neural networks (Pandarinath et al., 2018). 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 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 from user-defined features to data-driven methods follows similar trends in the field of machine learning.

Waveform cluster shapes are unlikely to arise from electrode placement

It is a possibility 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 respect to the distance between the neuron and the electrode. This is unlikely because both *in vitro* studies (Deligkaris et al., 2016) and computational simulations (Gold et al., 2006) show distance from the soma mostly induces changes in amplitude. There is a small widening in waveform width but this occurs at distances in which the amplitude has attenuated below even very low spike thresholds (Gold et al., 2006). We controlled for this cell type-irrelevant variation in amplitude by normalizing spike troughs/peaks during preprocessing to be between -1 and +1. It should also be noted that without any normalization, all structure was lost in the UMAP projection which instead yielded one large point cloud (Fig. S4E). Intuitively, this can be understood as UMAP allocating most of

the projected space to explaining amplitude differences rather than shape variation. This can be visualized by coloring each point by the log of the amplitude of each spike (log of difference in maximum vs. minimum values) and observing that it forms a smooth gradient in the projected space (Fig. S4F).

It is possible that differences that we observe in waveform shape could be due to recording from different morphological structures (dendrites, soma, or axons) rather than different cell types. However, we believe that most of our waveforms are from the soma. While it is true that there are some cell structures associated with different waveform shapes (such as triphasic waveforms near neurites, especially axons (Barry, 2015; Deligkaris et al., 2016; Robbins et al., 2013; Sun et al., 2021)), highly-controlled *in vitro* studies show that a large majority of EAP's are from somata (86%) (Deligkaris et al., 2016). In concordance with these results, we only observed one cluster (④, 6% of all EAP's) with a triphasic shape and 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. Finally, we observed distinct physiological properties (Fig. 6), decision-related dynamics (Fig. 7), and laminar distribution (Fig. 8) for each *WaveMAP* cluster. This would not be the case if the waveforms were just obtained from different compartments of the same neurons.

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; Fig. 5A). 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 (Kaczmarek and Zhang, 2017). These channels are necessary for sustained high-frequency firing (Ding et al., 2011). 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 (Erisir et al., 1999; Bean, 2007). Both of these changes in intracellular waveform shape likely have a strong effect on the shape of extracellular waveforms (Henze et al., 2000).

Reliance on waveform features might obscure cell type diversity

Our results show a greater proportion of narrow- (putatively inhibitory) vs. broad-spiking (putatively excitatory) neurons (69% vs. 31% respectively); this is inconsistent with anatomical studies (Dombrowski et al., 2001; Zaitsev et al., 2009; Povysheva et al., 2013). 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: Onorato and colleagues (Onorato et al., 2019) 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 shape and cell type (Lemon et al., 2021). A resolution to this discrepancy is to rethink equating narrow-spiking to inhibitory cells and broad-spiking to excitatory cells. 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 (Constantinople et al., 2009; Kelly et al., 2018; Kelly and Hawken, 2020; Ichinohe et al., 2004; Lemon et al., 2021). Furthermore, researchers have used antidromic stimulation to show that narrow-spiking neurons can be excitatory in motor and premotor cortex (Vigneswaran et al., 2011; Lemon et al., 2021)

We therefore believe prior studies have underexplored the diversity of classes accessed by their physiological recordings. Evidence of this is that histograms of peak width (and other specified features) across literature are often not cleanly bimodal (Krimer et al., 2005; Zhu et al., 2020) especially in premotor cortices (Merchant et al., 2012). In addition, the relative proportions of narrow vs. broad is often dependent on the cutoff chosen which widely varies across studies (Vigneswaran et al., 2011; Merchant et al., 2012). Analyses like ours which look at entire waveforms—rather than a few specified features—extract this diversity from extracellular recordings whereas specified features mix waveform classes.

Better parcellation of waveform variability leads to biological insight

We find that many narrow-spiking subtypes in PMd signal choice earlier than broad-spiking neurons in our decision-making task (Fig. 7F). These observations are consistent with another study of PMd in monkeys in reach target selection and movement production (Song and McPeek, 2010). 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) (Ding and Gold, 2012) and inferior temporal area (IT) (Mruczek and Sheinberg, 2012) 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. 7F). This lends further evidence to our theory that ② and ④ are likely narrow-spiking excitatory cells. In contrast, ③ and ① while both narrow-spiking, had distributions that more aligned with histologically-verified inhibitory types. In addition, ③ and ① had physiological properties more in line with inhibitory cell types.

WaveMAP suggests that narrow-spiking waveforms encompass many cell classes with distinct shape and laminar distribution. One of our narrow-spiking clusters (cluster ②) was restricted to more superficial layers (Fig. 8B) and had certain functional properties—low baseline firing rate and longer discrimination times—which are thought to be more closely aligned to properties of excitatory neurons (Song and McPeek, 2010). Another narrow-spiking cluster, ④, exhibited physiological and functional properties similar to ② (all comparisons not significant in Fig. 6C, D, and E or Fig. 7E and F) but with a distinct laminar distribution (Fig. 8B) and highly triphasic waveform shape (Fig. S9B). In contrast to ②, which was concentrated in layer III, ④ was restricted to deep layers. These tri-phasic neurons could either be large corticospinal excitatory pyramidal cells (Ichinohe et al., 2004; Soares et al., 2017; Vigneswaran et al., 2011), or axons (Barry, 2015; Robbins et al., 2013; Sun et al., 2021).

High-density probes and optogenetics can provide better insight into cell classes in the primate

Our recordings here were performed with 16 channel U-probes which 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 (Jun et al., 2017; Dimitriadis et al., 2020). High-density recordings would allow us to perform *WaveMAP* in an additional dimension (across multiple electrodes) to increase confidence in identified cell classes (Mosher et al., 2020) and localization of signal to somata (Jia et al., 2019). Sensitive electrodes providing spatial access to neural activity (Jun et al., 2017) can also improve our understanding of how these cell classes are organized both parallel and perpendicular to cortical surface (Saleh et al., 2019; Mosher et al., 2020) and across areas (Dimitriadis et al., 2020). Access to cell types with high-density recordings would also allow for the identification of "me-types" through electromorphology (Gouwens et al., 2020; Tasic et al., 2018). This information could also help inform detailed models of cortical circuits that incorporate cell type information (Gouwens et al., 2018; Billeh et al., 2020; Reimann et al., 2013).

Another powerful tool that has been leveraged in the study of cell types during behavior is optogenetics (Pinto and Dan, 2015; Lui et al., 2021; Kvitsiani et al., 2013). Although in its infancy relative to its use in the mouse, optogenetics in monkeys offers direct interrogation of cell types. Future studies will allow us to more precisely link putative cell classes *in vivo* to function (Courtin et al., 2014). NHP optogenetics is slowly advancing and efforts in many research groups around the world are producing new methods for *in vivo* optogenetics (Tremblay et al., 2020). We expect future experiments using the promising new mDlx (De et al., 2020) and h56d (Mehta et al., 2019) promoter sequences to selectively opto-tag inhibitory neurons or PV^+ neurons directly (Vormstein-Schneider et al., 2020) will greatly benefit validation of these derived cell classes. Finally, *WaveMAP*'s ability to find clusters of putative biological relevance using waveform shape alone encourages its application in settings where ground truth evaluation is particularly difficult to obtain such as in the human brain (Paulk et al., 2021).

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)

Code and Data Availability

All figures and supplemental figures can be generated from the code and data included with the manuscript. Figure 1 is generated using MATLAB whereas all other figures were generated using the Jupyter/Google Colab notebook available with this manuscript. Please see the Readme.md file included in the zip file for further instructions. Note, we have not included the raw firing rates across conditions for the neurons because of the large size of the data. This data can be made available by emailing the authors. Further information about *WaveMAP* and updated notebooks can also be obtained from the Chandrasekaran lab website at Boston University (http://www.chandlab.org).

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. This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All of the procedures were approved by the Stanford Administrative Panel on Laboratory Animal Care (**APLAC, Protocol Number 8856 entitled "Cortical Processing of Arm Movements"**). Surgical procedures were performed under anesthesia, and every effort was made to minimize suffering. Appropriate analgesia, pain relief, and antibiotics were administered to the animals when needed after surgical approval.

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 their 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 (Kleiner et al., 2007)). 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 made by ISCAN along with associated software (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 their 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 predominant color of 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 hold period, we stopped providing the small amounts of juice for holding but maintained the juice reward for correct reaches. After the animal learned to stay still during the target viewing period, we introduced more difficult checkerboard cues (decreased color coherences) 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

To guide the stereotaxic coordinates for our eletrophysiological recordings we used known response-to-muscle palpation properties of PMd and M1. Our chambers were placed normal to the surface of cortex and aligned with the skull of the monkey. 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 (UEWLGCSEEN1E, 110 mm long and 250 µm thick electrodes with a standard blunt tip and profile, epoxylite insulation, and an impedance of 5 - 7 MΩ) Every effort was made to isolate single units during the recordings with FHC electrodes by online monitoring and seeking out well-isolated signals (see next section below).

We performed linear multi-contact electrode (U-probe) recordings in the same manner as single electrode recordings with some minor modifications. We used 180 μm thick 16-electrode U-probes (15 μm Pt/Ir electrode site diameter, 150 µm spacing, circular shape, polyimide insulation, and secured in medical-grade epoxy. Electrode contacts were $\sim$100 KΩ in impedance). We used a slightly sharpened guide tube to provide more purchase on dura. We also periodically scraped away, under ketamine-dexmetotomidine anesthesia, any overlying tissue on the dura. Both these modifications greatly facilitated penetration of the U-probe. We typically penetrated the brain at very slow rates ($\sim$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 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 disappearance of spiking activity 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. 8.

Identification of single neurons during recordings

Our procedure for identifying well-isolated single neurons was as follows: In the case of the single FHC tungsten electrode recordings, we moved the electrode and conservatively adjusted the threshold until we identified a well-demarcated set of waveforms. We took extra care to separate these waveforms from the noise and other smaller neurons. Our ability to isolate neurons was helped by the fact that these electrodes have a very small exposed area (hence high impedance) allowing for excellent isolation. Once a stable set of waveforms was identified, hoops from the Central software (Blackrock Microsystems) were used to demarcate the waveforms from noise and other potential single neurons. The electrode was allowed to settle for at least 15 minutes to ensure that the recording was stable. Once stability was confirmed, we began data collection. If we found that the recording was unstable, we discarded the neuron and moved the electrode to a newly isolated neuron and repeated the procedure. For a stable recording, we stored the waveform snippet and the time of the spike. Offline, we first visualized the waveforms in MATLAB by performing PCA. If we found that our online identification of the waveforms was inadequate, we either discarded the recording, identified it as a multi-unit (not used in this paper), or exported the data to Plexon Offline Sorter and redrew the cluster boundaries. We also took advantage of Plexon Offline Sorter's ability to visualize how PCA features changed with time to ensure the quality and stability of our isolation. Finally, after redrawing cluster boundaries, we exported the data back to our analysis pipeline.

For our 16-channel Plexon U-Probe recordings, we again lowered the electrode until we found a stable set of waveforms. The small contact area of these electrode sites again ensured excellent identification and low levels of background noise (~10-20 μV). We then waited at least 45 minutes until the recordings were very stable. Such an approach ensured that we minimized electrode drift. In our experience, the U-probes also have less drift than Neuropixel or other recording methods. We then again repeated the conservative thresholding and identification procedure outlined for the FHC electrodes. For U-probes, we did not move the electrodes once they had settled. Instead, we constantly monitored the recordings and any changes in a particular electrode over time led to the units from that electrode being discarded and not included in further analysis. Finally, the same offline procedures used for FHC electrodes were repeated for the U-probe recordings.

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% (Chandrasekaran et al., 2017). 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 (Kaufman et al., 2013). 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 was less than 0.4 (Kaufman et al., 2013). 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 trough occurring first. The remaining waveforms, unless stated otherwise here, were removed from the analysis. We excluded positive-spiking waveforms because of their association with axons (Sun et al., 2021). Finally, we normalized the waveforms by dividing the extreme value of the amplitude such that the maximum deviation is ± 1 unit (Snyder et al., 2016).

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 (Kaufman et al., 2010), 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. S4A and C). Furthermore, both normalizations picked out the same structure (Fig. S4B 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.

A step-by-step guide to UMAP and Louvain clustering in WaveMAP

To provide the reader with an intuitive overview of *WaveMAP*, we provide a step-by-step exposition of the different stages in the workflow shown in Fig. 2. UMAP is a non-linear method that enables the capture of latent structures in high-dimensional data as a graph. This graph can then be used to visualize the latent structure in a low-dimensional embedding (McInnes et al., 2018). For a detailed description of the methods, please refer to the Supplemental Information or respective references for UMAP (McInnes et al., 2018) and Louvain community detection (Blondel et al., 2008).

Fig. 2A-i: We pass 625 normalized single-unit waveforms into UMAP. This normalization is crucial for exposing interesting structure in downstream analysis although the particular normalization is

less important (Fig. S4). UMAP uses a five step (**ii.a** to **ii.e** in Fig. 2A) procedure to construct a weighted high-dimensional graph.

Fig. 2A-ii.a: In the first step, the data for each waveform is viewed in its original (sometimes called "ambient") 48-dimensional space with each dimension corresponding to one of 48 time points along the waveform recording.

Fig. 2A-ii.b: A local metric is then assigned to each data point such that a unit ball (distance of one) surrounding it extends to the 1st-nearest neighbor. This ensures that every point is connected to at least one other point.

Fig. 2A-ii.c: Beyond this first connection, the distances to the next $(k-1)$-nearest neighbors increases according to an exponential distribution scaled by the local density. This is shown as a "glow" around each of the unit balls in Fig. 2A-ii.c.

Fig. 2A-ii.d: The distances from the local point to the $k-1$ data points beyond the unit ball are made to be probabilistic ("fuzzy") according to their distance (k $=$ 4 in Fig. 2A-ii.d with some low weight connections omitted for clarity). This also means that the metric around each data point has a different notion of "how far" their neighbors are. Distances are shorter in dense regions (with respect to the ambient space) than are distances in sparser regions leading to a graph with asymmetric edge weights. If the notion of a probabilistic connection is confusing, this construction can just be understood as an asymmetric directed graph with edge weights between zero and one. One way to understand this is through the following real life example: to someone living in a dense city, traveling several miles may seem very far, while for a rural resident, this distance might be trivial even if the absolute distance is the same.

Fig. 2A-ii.e: The edge weights between any two data points, a and b, are "averaged together" according to the formula $a + b - a \cdot b$ known as probabilistic sum. This results in a graph that now has symmetric edge weights.

Fig. 2B-iv: However, before we project this graph into lower dimension, we first apply clustering to the high-dimensional graph with a method known as Louvain community detection. This method proceeds in two steps per pass: modularity optimization and community aggregation (Fig. S1B).

Fig. 2B-iv.a: In the first step of Louvain, each node in the graph is assigned to its own "community" which can be interpreted as its own cluster. Next, each node will join a neighboring node's community such as to maximize an objective function known as "modularity score" (see Supplemental Information for the exact equation). This score is maximized when the sum of the weighted edges within a community is maximal relative to the sum of the weighted edges incident on the community from nodes outside the community. This procedure operates on all nodes until modularity can no longer be increased across the network; this concludes the modularity optimization step. In the next step, community aggregation, all nodes in a community are collapsed into a single node to create a new network. This completes the first pass of Louvain which then repeats modularity optimization and aggregation on this new graph until modularity is once again maximized. This continues until modularity no longer increases across hierarchies of graphs. Note that the resolution parameter is set to one in the work of Blondel et al. (2008) but we use an implementation of Louvain that allows for changing of this parameter according to the definition of modularity given in Lambiotte (2007). The resolution parameter gives the user some

ability to specify how large of a community they expect as might be related to a phenomenon of interest and should be chosen empirically.

Fig. 2B-iv.b: The final Louvain community memberships are propagated back to the original nodes and assigned as labels for the associated data points which completes the classification step.

Fig. 2B-v: In the second step of UMAP, graph layout, the high-dimensional graph with symmetric edge weights from the previous step is projected down into some lower dimension (here it is two dimensions). To initialize this process, the graph is first passed through a Laplacian eigenmap (Belkin and Niyogi, 2002) which helps regularize the initial embedding of the graph in low dimension (Kobak and Linderman, 2019).

Fig. 2B-v.a: We know the graph in high dimension but we haven't yet found this graph in low dimension. From here a force directed graph layout procedure is used to align the initialized low-dimensional graph with the one found in high dimension. The force directed graph layout procedure minimizes the difference in distances in the instantiated graph compared to the high-dimensional one by alternatingly applying an attractive and a repulsive force according to the edge weights. These forces are chosen to minimize the cross-entropy between the graphs. This process ensures that points close in high dimension but far in low dimension are brought together (attraction) and those that are far in high dimension but close in low dimension are pushed apart (repulsion).

Fig. 2B-v.b: A final embedding of the data is found using stochastic gradient descent but by fixing the seed in our procedure, we enforce standard gradient descent. Although this requires more memory and is less performant, it guarantees that embeddings will look the same (even if this doesn't affect clustering).

Fig. 2B-vi: In the final step of our method, we combine the labels found through Louvain clustering with a low-dimensional embedding to arrive at our *WaveMAP* solution.

WaveMAP parameter selection and validation

The normalized extracellular waveforms were passed into the Python package `umap 0.4.0rc3` (McInnes et al., 2018) with the parameters shown in Table 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. 3A. 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 (Porter et al., 2009; Lambiotte et al., 2008). To select the resolution parameter t, we chose a value that best maximized 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 statistically analyzable number of clusters (n $>$ 20). We selected a resolution parameter (green marker on Fig. 3B) that maximized modularity score of Louvain clustering while still returning clusters of $n > 20$ to allow for downstream

statistical analyses. We note that this was also very close to the "elbow" in terms of number of clusters; this verifies that we have reached near-optimality in a second sense of obtaining stable cluster number. These scores were calculated over 25 random UMAP instantiations of 80% of the full dataset in 5% intervals. 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 dataset (Fig. S2B) (Tibshirani and Walther, 2012). We obtained 100 random samples from 10% to 90% of the full data set in 10% increments while simultaneously choosing a different random seed for the UMAP algorithm each time. We calculated both the number of Louvain clusters and the adjusted mutual information score (AMI) across these random samples and plot it on the same graph Fig. S2B, red and green. The AMI is a measure of how much "information" is shared between a pair of clusterings with information specifically as Shannon entropy. The Shannon entropy of a given clustering (often called a "partitioning"), $H(X)$, is defined as,

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$

with the probability $P(x_i)$ pertaining to the probability that a certain data point x_i will belong to a certain cluster. The clustering entropy can be understood as how "unpredictable" the results of a random variable are (in this case the random variable is the particular clustering solution). For instance, a fair coin is less predictable (greater entropy) than a loaded coin (lower entropy).

Intuitively, AMI is how much information we receive about one variable given an observation of another and vice versa (Timme and Lapish, 2018) (or how much knowing one clustering allows us to predict another) corrected for random chance. Thus, it's bounded between 0 and 1 with 0 for two completely independent variables and 1 for completely identical variables. Formally, un-adjusted mutual information, $I(X, Y)$, is defined as,

$$\begin{aligned} I(X, Y) &= H(X) - H(X|Y) \\ &= D_{KL}(P_{X,Y}(x, y) \parallel P_X(x) \cdot P_Y(y)) \end{aligned}$$

where $H(X|Y)$ is the conditional Shannon entropy, $P_{X,Y}(x, y)$ is the joint distribution of X and Y, and $P_X(x)$ is a marginal distribution.

The value at 100% is omitted because it has the same cluster number as our dataset and zero variance since *WaveMAP* is invariant to random seed selection. Thus the variation in cluster number due to sampling and random seed are compounded and shown together in Fig. S2B.

Ensemble clustering for graphs (ECG) (Poulin and Théberge, 2018, 2019) was used to validate the clusters found in Fig. 3A (see Fig. S2C). We added the algorithm (https://github.com/ftheberge/Ensemble-Clustering-for-Graphs) into the `python-igraph` package (Csardi and Nepusz, 2006) 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 level-1 graphs (called level-1 partitions since one round of Louvain was performed) 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 section *Ensemble clustering for graphs (ECG)*) yields the final ECG graph.

Gradient boosted decision tree classifier

We then trained a gradient boosted decision tree classifier in `xgboost 1.0.2` (Chen and Guestrin, 2016). A 30% - 70% test-train split was used with the test set never seen by model training or hyperparameter tuning. A five-fold cross-validation was applied to the training data and optimal hyperparameters were obtained after a grid search on the folds using `scikit-learn`'s `GridSearchCV` function with final parameters in Table S1. The default multi-class objective function `multi:softmax` was also used. The percent accuracy for each cluster against all others is plotted as a confusion matrix in Fig. 3C by applying the final classifier model to the unseen test set.

The same procedure was used when training on the GMM labels found in Fig. 4D and for the eight cluster GMM labels in Fig. S6B. 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 by looking at the entire dataset (both via the algorithm itself and our tuning of parameters). This results in data leakage (Moscovich and Rosset, 2019; Kaufman et al., 2011) which potentially hurts out-of-dataset performance. Thus, classifier performance is only used here to demonstrate UMAP's ability to sensibly separate waveforms within-dataset relative to traditional GMM methods (Fig. 4D). Fig. 3C is not leveraged to provide firm insight into how such a classifier would perform out-of-dataset. It is also important to note that none of the parameters for *WaveMAP* (`n_neighbors` or `resolution`) were tuned to optimize for classifier performance and thus the direction of bias is not necessarily deleterious.

Specified waveform shape features

To compute specified features for each normalized waveforms (Fig. 4A), 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 (Stuttgen, 2019). 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. 4B). 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 a "goodness of fit" term obtained from the negative log likelihood function, i.e., the conditional probability of observing the sample X given a vector of parameters θ. In the particular case of GMM, the function $P(X|\theta)$ is the probability of observing the clusters given an underlying particular sum of multivariate Gaussians (the likelihood). The second term $K\ln(n)$ is a penalty on the number of parameters n which approximates model complexity. Penalizing the number of model parameters (number of clusters K) scaled by the number of data points n captures the idea that "simplicity is better". This criterion ultimately constrains the number of Gaussians used to fit the data.

Assuming we have N_f features and N_c clusters we can calculate K using the following framework: For each Gaussian mixture model, 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 for a GMM 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 minimized (at the "elbow" in Fig. 4C).

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 projected space Fig. 5A. 100 evenly-spaced test coordinates were generated spanning a portion of the embedded space 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, (Lundberg et al., 2020)), SHAP values were calculated for the classifier trained on *WaveMAP* identified clusters. The trained `XGBoost` model was passed directly into the tree model-specific `shap.TreeExplainer` (Lundberg et al., 2018) which then calculated the mean absolute SHAP values (the average impact on model performance, postive or negative) 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. 5C). The SHAP values for each time point, across classes, was summed to generate the overall SHAP values for each time point (Fig. 5B).

Choice-selective signal

We use an approach developed in Meister et al. (2013) to estimate the choice-selective signal. We chose such an approach because decision-related activity of PMd neurons does not simply increase or decrease 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.

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. 7A,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 color coherence provided us with an estimate of the rate of change of the choice selectivity signal (η) for all the coherences in spikes/s/s. Averaging over neurons for each cluster provided us with the graphs in Fig. 7C,D. We then estimated the dependence of η on color coherence by regressing η and color coherence to estimate how strongly choice-selectivity signals in a particular cluster were modulated by the stimulus input. This modulation is summarized in Fig. 7E and measured as "coherence slope" in units of spikes/s/s/% color coherence.

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.

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 perfusion protocol as described (Medalla and Luebke, 2015). 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 (Rosene et al., 1986). 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 (Estrada et al., 2016).

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 (DeFelipe, 1997). 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 $\times$ 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 $\times$ 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 (Barbas and Pandya, 1987; Morecraft et al., 2004, 2019) and area F7 in several functional maps (Matelli and Luppino, 1996; Rizzolatti et al., 1998). 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 (Medalla and Luebke, 2015).

Stereological cell counting

Due to its demonstrated ability in producing minimally-biased results, 3D stereologic cell counting (Schmitz et al., 2014) was utilized to count parvalbumin- (PV^+), calretinin- (CR^+) and calbindin-positive (CB^+) cells. Using the CellCounter plugin in `Fiji` (Shindelin et al., 2012) 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` (Virtanen et al., 2020). Multiple comparisons were corrected for using false detection-rate adjusted p-values (Benjamini-Hochberg); this was done using `scipy.stats.multitest` and `scikit-posthocs` (Terpilowski, 2019). Ordinary least squares regressions were conducted in the package `statsmodels` (Seabold and Perktold, 2010). 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

Aggarwal CC, Hinneburg A, Keim DA (2001) On the surprising behavior of distance metrics in high dimensional space In Van den Bussche J, Vianu V, editors, *Database Theory — ICDT 2001*, pp. 420–434, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ali M, Jones MW, Xie X, Williams M (2019) TimeCluster: dimension reduction applied to temporal data for visual analytics. *The Visual Computer* 35:1013–1026.

Amatrudo JM, Weaver CM, Crimins JL, Hof PR, Rosene DL, Luebke JI (2012) Influence of Highly Distinctive Structural Properties on the Excitability of Pyramidal Neurons in Monkey Visual and Prefrontal Cortices. *The Journal of Neuroscience* 32:13644–13660.

Ardid S, Vinck M, Kaping D, Marquez S, Everling S, Womelsdorf T (2015) Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. *The Journal of Neuroscience* 35:2975–2991.

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* 277:21–40.

Azodi CB, Tang J, Shiu SH (2020) Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends in Genetics* 36:442–455.

Azouz R, Gray CM, Nowak LG, McCormick DA (1997) Physiological properties of inhibitory interneurons in cat striate cortex. *Cerebral Cortex* 7:534–545.

Bakkum DJ, Frey U, Radivojevic M, Russell TL, Müller J, Fiscella M, Takahashi H, Hierlemann A (2013) Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nature Communications* 4:2181.

Bala PC, Eisenreich BR, Yoo SBM, Hayden BY, Park HS, Zimmermann J (2020) OpenMonkeyStudio: Automated Markerless Pose Estimation in Freely Moving Macaques. *bioRxiv* p. 2020.01.31.928861.

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* 256:211–228.

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.

Barthó P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsáki G (2004) Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. *Journal of Neurophysiology* 92:600–608.

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* 115:1117–1122.

Bean BP (2007) The action potential in mammalian central neurons. *Nature Reviews Neuroscience* 8:451–465.

Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW (2018) Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 37:38–44.

Belkin M, Niyogi P (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering In Dietterich TG, Becker S, Ghahramani Z, editors, *Advances in Neural Information Processing Systems 14*, pp. 585–591. MIT Press.

Bellet A, Habrard A, Sebban M (2013) A Survey on Metric Learning for Feature Vectors and Structured Data. *arXiv* .

Bengio Y, Vincent P, Paiement J (2003) Spectral clustering and kernel pca are learning eigenfunctions.

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, p. 217–235, Berlin, Heidelberg. Springer-Verlag.

Billeh YN, Cai B, Gratiy SL, Dai K, Iyer R, Gouwens NW, Abbasi-Asl R, Jia X, Siegle JH, Olsen SR, Koch C, Mihalas S, Arkhipov A (2020) Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex. *Neuron* .

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008:P10008.

Boroujeni KB, Tiesinga P, Womelsdorf T (2021) Interneuron-specific gamma synchronization indexes cue uncertainty and prediction errors in lateral prefrontal and anterior cingulate cortex. *eLife* 10:e69111.

Bruce CJ, Goldberg ME (1985) Primate frontal eye fields. i. single neurons discharging before saccades. *Journal of neurophysiology* 53:603–635.

Chah E, Hok V, Della-Chiesa A, Miller JJH, O'Mara SM, Reilly RB (2011) Automated spike sorting algorithm based on Laplacian eigenmaps and k-means clustering. *Journal of Neural Engineering* 8:016006.

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.

Chandrasekaran C, Peixoto D, Newsome WT, Shenoy KV (2017) Laminar differences in decision-related neural activity in dorsal premotor cortex. *Nature Communications* 8:614.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. arxiv 2016. *arXiv preprint arXiv:1603.02754* .

Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. *Nature* 487:51–56.

Cisek P (2012) Making decisions through a distributed consensus. *Current opinion in neurobiology* 22:927–936.

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* 482:85–88.

Connors BW, Gutnick MJ, Prince DA (1982) Electrophysiological properties of neocortical neurons in vitro. *Journal of Neurophysiology* 48:1302–1320.

Connors BW, Gutnick MJ (1990) Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences* 13:99–104.

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* 516:291–311.

Contreras D (2004) Electrophysiological classes of neocortical neurons. *Neural Networks* 17:633–646.

Courtin J, Chaudun F, Rozeske RR, Karalis N, Gonzalez-Campo C, Wurtz H, Abdi A, Baufreton J, Bienvenu TC, Herry C (2014) Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature* 505:92–96.

Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal* Complex Systems:1695.

Cunningham JP, Byron MY (2014) Dimensionality reduction for large-scale neural recordings. *Nature neuroscience* 17:1500–1509.

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.

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* 14:1–19.

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.

Deubner J, Coulon P, Diester I (2019) Optogenetic approaches to study the mammalian brain. *Current Opinion in Structural Biology* 57:157–163.

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:e1008432.

Dimitriadis G, Neto JP, Aarts A, Alexandru A, Ballini M, Battaglia F, Calcaterra L, Chen S, David F, Fiáth R, Frazão J, Geerts JP, Gentet LJ, Helleputte NV, Holzhammer T, Hoof Cv, Horváth D, Lopes G, Lopez CM, Maris E, Marques-Smith A, Márton G, McNaughton BL, Meszéna D, Mitra S, Musa S, Neves H, Nogueira J, Orban GA, Pothof F, Putzeys J, Raducanu BC, Ruther P, Schroeder T, Singer W, Steinmetz NA, Tiesinga P, Ulbert I, Wang S, Welkenhuysen M, Kampff AR (2020) Why not record from every electrode with a CMOS scanning probe? *bioRxiv* p. 275818.

Dimitriadis G, Neto JP, Kampff AR (2018) t-SNE Visualization of Large-Scale Neural Recordings. *Neural Computation* 30:1750–1774.

Ding L, Gold JI (2012) Neural correlates of perceptual decision making before, during, and after

decision commitment in monkey frontal eye field. *Cerebral Cortex* 22:1052–1067.
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* 105:554–570.
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.)* 368:89–94.
Dombrowski S, Hilgetag C, Barbas H (2001) Quantitative Architecture Distinguishes Prefrontal Cortical Systems in the Rhesus Monkey. *Cerebral Cortex* 11:975–988.
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.
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* 82:2476–2489.
Estebanez L, Hoffmann D, Voigt BC, Poulet JFA (2017) Parvalbumin-Expressing GABAergic Neurons in Primary Motor Cortex Signal Reaching. *Cell Reports* 20:308–318.
Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, p. 226–231. AAAI Press.
Estrada LI, Robinson AA, Amaral AC, Giannaris EL, Heyworth NC, Mortazavi F, Ngwenya LB, Roberts DE, Cabral HJ, Killiany RJ, Rosene DL (2016) Evaluation of Long-Term Cryostorage of Brain Tissue Sections for Quantitative Histochemistry. *Journal of Histochemistry & Cytochemistry* 65:153–171.
Finn ES, Huber L, Jangraw DC, Molfese PJ, Bandettini PA (2019) Layer-dependent activity in human prefrontal cortex during working memory. *Nature Neuroscience* 22:1687–1695.
Georgopoulos A, Schwartz A, Kettner R (1986) Neuronal population coding of movement direction. *Science* 233:1416–1419.
Gold C, Girardin CC, Martin KAC, Koch C (2009) High-Amplitude Positive Spikes Recorded Extracellularly in Cat Visual Cortex. *Journal of Neurophysiology* 102:3340–3351.
Gold C, Henze DA, Koch C, Buzsáki G (2006) On the Origin of the Extracellular Action Potential Waveform: A Modeling Study. *Journal of Neurophysiology* 95:3113–3128.
Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annual Review of Neuroscience* 30:535–574 PMID: 17600525.
Gouwens NW, Berg J, Feng D, Sorensen SA, Zeng H, Hawrylycz MJ, Koch C, Arkhipov A (2018) Systematic generation of biophysically detailed models for diverse cortical neuron types. *Nature Communications* 9:710.
Gouwens NW, Sorensen SA, Baftizadeh F, Budzillo A, Lee BR, Jarsky T, Alfiler L, Baker K, Barkan E, Berry K, Bertagnolli D, Bickley K, Bomben J, Braun T, Brouner K, Casper T, Crichton K, Daigle TL, Dalley R, de Frates RA, Dee N, Desta T, Lee SD, Dotson N, Egdorf T, Ellingwood L, Enstrom R, Esposito L, Farrell C, Feng D, Fong O, Gala R, Gamlin C, Gary A, Glandon A, Goldy J, Gorham M, Graybuck L, Gu H, Hadley K, Hawrylycz MJ, Henry AM, Hill D, Hupp M, Kebede S, Kim TK, Kim L, Kroll M, Lee C, Link KE, Mallory M, Mann R, Maxwell M, McGraw M, McMillen D, Mukora A, Ng L, Ng L, Ngo K, Nicovich PR, Oldre A, Park D, Peng H, Penn O, Pham T, Pom A, Popović Z, Potekhina L, Rajanbabu R, Ransford S, Reid D, Rimorin C, Robertson M, Ronellenfitch K, Ruiz A, Sandman D, Smith K, Sulc J, Sunkin SM, Szafer A, Tieu M, Torkelson A, Trinh J, Tung H, Wakeman W, Ward K, Williams

G, Zhou Z, Ting JT, Arkhipov A, Sümbül U, Lein ES, Koch C, Yao Z, Tasic B, Berg J, Murphy GJ, Zeng H (2020) Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* 183:935–953.e19.

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* 82:1451–1464.

Hangya B, Ranade SP, Lorenc M, Kepecs A (2015) Central Cholinergic Neurons Are Rapidly Recruited by Reinforcement Feedback. *Cell* 162:1155–1168.

Hanks TD, Kopec CD, Brunton BW, Duan CA, Erlich JC, Brody CD (2015) Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520:220–223.

Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G (2000) Accuracy of Tetrode Spike Separation as Determined by Simultaneous Intracellular and Extracellular Measurements. *Journal of Neurophysiology* 84:401–414.

Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G (2000) Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo. *Journal of Neurophysiology* 84:390–400.

Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, Close JL, Long B, Johansen N, Penn O et al. (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573:61–68.

Hsu AI, Yttri EA (2020) B-soid: An open source unsupervised algorithm for discovery of spontaneous behaviors. *bioRxiv* .

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* 148:574.

Hussar CR, Pasternak T (2009) Flexibility of Sensory Representations in Prefrontal Cortex Depends on Cell Type. *Neuron* 64:730–743.

Ichinohe N, Watakabe A, Miyashita T, Yamamori T, Hashikawa T, Rockland K (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* 129:179–185.

Jia X, Siegle JH, Bennett C, Gale SD, Denman DJ, Koch C, Olsen SR (2019) High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification. *Journal of Neurophysiology* 121:1831–1847.

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* 29:5516–5524.

Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Ç et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551:232–236.

Kaczmarek LK, Zhang Y (2017) Kv3 Channels: Enablers of Rapid Firing, Neurotransmitter Release, and Neuronal Endurance. *Physiological Reviews* 97:1431–1468.

Katai S, Kato K, Unno S, Kang Y, Saruwatari M, Ishikawa N, Inoue M, Mikami A (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* 31:1322–1338.

Kaufman MT, Churchland MM, Santhanam G, Yu BM, Afshar A, Ryu SI, Shenoy KV (2010) Roles of monkey premotor neuron classes in movement preparation and execution. *Journal of*

neurophysiology 104:799–810.
Kaufman MT, Churchland MM, Shenoy KV (2013) The roles of monkey M1 neuron classes in movement preparation and execution. *Journal of neurophysiology* 110:817–25.
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.
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* 29:1921–1937.
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* 225:1135–1152.
Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in psychtoolbox-3. *Perception* 36:1–16.
Kleinman M, Chandrasekaran C, Kao JC (2019) Recurrent neural network models of multi-area computation underlying decision-making. *bioRxiv* .
Klempíř O, Krupička R, Kršek J, Dittert I, Petráková V, Petrák V, Taylor A (2020) Application of spike sorting algorithm to neuronal signals originated from boron doped diamond micro-electrode arrays. *Physiological Research* 69.
Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi XL, Romo R, Uchida N, Machens CK (2016) Demixed principal component analysis of neural population data. *Elife* 5:e10989.
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.
Krienen FM, Goldman M, Zhang Q, Del Rosario RC, Florio M, Machold R, Saunders A, Levandowski K, Zaniewski H, Schuman B et al. (2020) Innovations present in the primate interneuron repertoire. *Nature* 586:262–269.
Krimer LS, Zaitsev AV, Czanner G, Kröner S, González-Burgos G, Povysheva NV, Iyengar S, Barrionuevo G, Lewis DA (2005) Cluster Analysis–Based Physiological Classification and Morphological Properties of Inhibitory Neurons in Layers 2–3 of Monkey Dorsolateral Prefrontal Cortex. *Journal of Neurophysiology* 94:3009–3022.
Kvitsiani D, Ranade S, Hangya B, Taniguchi H, Huang JZ, Kepecs A (2013) Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498:363–366.
Lambiotte R (2007) Finding communities at different resolutions in large networks .
Lambiotte R, Delvenne JC, Barahona M (2008) Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770* .
Lee JA, Verleysen M, editors (2007) *Nonlinear Dimensionality Reduction* Springer New York.
Lemon RN, Baker SN, Kraskov A (2021) Classification of Cortical Neurons by Spike Shape and the Identification of Pyramidal Neurons. *Cerebral Cortex* pp. bhab147–.
Linderman GC, Rachh M, Hoskins JG, Steinerberger S, Kluger Y (2019) Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nature Methods* 16:243–245.
Lui JH, Nguyen ND, Grutzner SM, Darmanis S, Peixoto D, Wagner MJ, Allen WE, Kebschull JM, Richman EB, Ren J, Newsome WT, Quake SR, Luo L (2021) Differential encoding in prefrontal cortex projection neuron classes across cognitive tasks. *Cell* 184:489–506.e26.
Lundberg S, Lee SI (2017) A Unified Approach to Interpreting Model Predictions. *arXiv* .

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI (2020) From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2:56–67.

Lundberg SM, Erion GG, Lee SI (2018) Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* .

Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of machine learning research* 9:2579–2605.

Mahallati S, Bezdek JC, Popovic MR, Valiante TA (2019) Cluster tendency assessment in neuronal spike data. *PLOS ONE* 14:e0224547.

Maheswaranathan N, Williams AH, Golub MD, Ganguli S, Sussillo D (2019) Universality and individuality in neural dynamics across large populations of recurrent networks .

Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84.

Markanday A, Bellet J, Bellet ME, Inoue J, Hafed ZM, Thier P (2020) Using deep neural networks to detect complex spikes of cerebellar Purkinje cells. *Journal of Neurophysiology* 123:2217–2234.

Matelli M, Luppino G (1996) Thalamic input to mesial and superior area 6 in the macaque monkey. *Journal of Comparative Neurology* 372:59–87.

McCormick DA, Connors BW, Lighthall JW, Prince DA (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology* 54:782–806.

McInnes L (2019) Umap uniform manifold approximation and projection for dimension reduction.

McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* .

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* 35:112–127.

Mehta P, Kreeger L, Wylie DC, Pattadkal JJ, Lusignan T, Davis MJ, Turi GF, Li WK, Whitmire MP, Chen Y, Kajs BL, Seidemann E, Priebe NJ, Losonczy A, Zemelman BV (2019) Functional access to neuron subclasses in rodent and primate forebrain. *Cell Reports* 26:2818 – 2832.e8.

Meister MLR, Hennig JA, Huk AC (2013) Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *Journal of Neuroscience* 33:2254–2267.

Merchant H, Naselaris T, Georgopoulos AP (2008) Dynamic Sculpting of Directional Tuning in the Primate Motor Cortex during Three-Dimensional Reaching. *Journal of Neuroscience* 28:9164–9172.

Merchant H, Lafuente Vd, Peña-Ortega F, Larriva-Sahd J (2012) Functional impact of interneuronal inhibition in the cerebral cortex of behaving animals. *Progress in Neurobiology* 99:163–178.

Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area v4. *Neuron* 55:131–141.

Molnar C (2020) *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* LULU COM.

Morecraft R, Cipolloni P, Stilwell-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* 469:37–69.

Morecraft RJ, Ge J, Stilwell [U+2010] Morecraft KS, Rotella DL, Pizzimenti MA, Darling WG (2019) Terminal organization of the corticospinal projection from the lateral premotor cortex to the cervical enlargement (C5–T1) in rhesus monkey. *Journal of Comparative Neurol-*

ogy 527:2761–2789.
Moscovich A, Rosset S (2019) On the cross-validation bias due to unsupervised pre-processing. *arXiv* .
Mosher CP, Wei Y, Kaminski J, Nandi A, Mamelak AN, Anastassiou CA, Rutishauser U (2020) Cellular classes in the human brain revealed in-vivo by heartbeat-related modulation of the extracellular action potential waveform. *Cell Reports* 30:3536 – 3551.e6.
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* 32:452–484.
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* 108:2725–2736.
Nandy AS, Nassi JJ, Reynolds JH (2017) Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* 93:235–246.
Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys. Rev. E* 69:026113.
Noichl M (2019) Modeling the structure of recent philosophy. *Synthese* pp. 1–12.
Nolet CJ, Lafargue V, Raff E, Nanditale T, Oates T, Zedlewski J, Patterson J (2020) Bringing UMAP Closer to the Speed of Light with GPU Acceleration. *arXiv* .
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* 89:1541–1566.
Onorato I, Neuenschwander S, Hoy J, Lima B, Rocha KS, Broggini AC, Uran C, Spyropoulos G, Klon-Lipok J, Womelsdorf T, Fries P, Niell C, Singer W, Vinck M (2019) A Distinct Class of Bursting Neurons with Strong Gamma Synchronization and Stimulus Selectivity in Monkey V1. *Neuron* .
Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR et al. (2018) Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods* 15:805–815.
Paulk AC, Kfir Y, Khanna A, Mustroph M, Trautmann EM, Soper DJ, Stavisky SD, Welkenhuysen M, Dutta B, Shenoy KV, Hochberg LR, Richardson M, Williams ZM, Cash SS (2021) Large-scale neural recordings with single-cell resolution in human cortex using high-density neuropixels probes. *bioRxiv* .
Pinto L, Dan Y (2015) Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron* 87:437–450.
Porter MA, Onnela JP, Mucha PJ (2009) Communities in Networks. *arXiv* .
Poulin V, Théberge F (2018) Ensemble Clustering for Graphs. *arXiv* .
Poulin V, Théberge F (2019) Ensemble clustering for graphs: comparisons and applications. *Applied Network Science* 4:51.
Povysheva NV, Zaitsev AV, Gonzalez-Burgos G, Lewis DA (2013) Electrophysiological Heterogeneity of Fast-Spiking Interneurons: Chandelier versus Basket Cells. *PLoS ONE* 8:e70553.
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.
Reimann MW, Anastassiou CA, Perin R, Hill SL, Markram H, Koch C (2013) A Biophysically

Detailed Model of Neocortical Local Field Potentials Predicts the Critical Role of Active Membrane Currents. *Neuron* 79:375–390.

Remington ED, Narain D, Hosseini EA, Jazayeri M (2018) Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* 98:1005 – 1019.e5.

Rizzolatti G, Luppino G, Matelli M (1998) The organization of the cortical motor system: new concepts. *Electroencephalography and Clinical Neurophysiology* 106:283–296.

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.

Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience* 22:9475–9489.

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* 34:1301–1315.

Roux L, Stark E, Sjulson L, Buzsáki G (2014) In vivo optogenetic identification and manipulation of GABAergic interneuron subtypes. *Current opinion in neurobiology* 26:88–95.

Sainburg T, McInnes L, Gentner TQ (2020) Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning. *arXiv* .

Saleh MS, Ritchie SM, Nicholas MA, Bezbaruah R, Reddy JW, Chamanzar M, Yttri EA, Panat RP (2019) Cmu array: A 3d nano-printed, customizable ultra-high-density microelectrode array platform. *bioRxiv* .

Schmitz C, Eastwood BS, Tappan SJ, Glaser JR, Peterson DA, Hof PR (2014) Current automated 3D cell detection methods are not a suitable replacement for manual stereologic cell counting. *Frontiers in Neuroanatomy* 8:1–14.

Seabold S, Perktold J (2010) statsmodels: Econometric and statistical modeling with python In *9th Python in Science Conference*.

Sedaghat-Nejad E, Fakharian MA, Pi J, Hage P, Kojima Y, Soetedjo R, Ohmae S, Medina JF, Shadmehr R (2021) P-sort: an open-source software for cerebellar neurophysiology. *bioRxiv* .

Shapley LS (1988) *A value for n-person games*, p. 31–40 Cambridge University Press.

Shenoy KV, Sahani M, Churchland MM (2013) Cortical control of arm movements: a dynamical systems perspective. *Annual review of neuroscience* 36.

Shindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Freibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: An Open-source Platform for Biological-image Analysis. *Nature Methods* 9:676—-682.

Simons DJ (1978) Response properties of vibrissa units in rat SI somatosensory neocortex. *Journal of Neurophysiology* 41:798–820.

Snyder AC, Morais MJ, Smith MA (2016) Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *Journal of Neurophysiology* 116:1807–1820.

Soares D, Goldrick I, Lemon RN, Kraskov A, Greensmith L, Kalmar B (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* 525:2164–2174.

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* 103:2124–2138.

Spivak DI (2009) Metric realization of fuzzy simplicial sets.

Steriade M (2004) Neocortical cell classes are flexible entities. *Nature Reviews Neuroscience* 5:121–134.

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* 79:483–490.

Stuttgen M (2019) Mlib - toolbox for analyzing spike data.

Sun SH, Almasi A, Yunzab M, Zehra S, Hicks DG, Kameneva T, Ibbotson MR, Meffin H (2021) Analysis of extracellular spike waveforms and associated receptive fields of neurons in cat primary visual cortex. *The Journal of Physiology* .

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller J, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirschstein D, Pendergraft J, Sullivan HA, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H (2018) Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563:72–78.

Tenenbaum JB, Silva Vd, Langford JC (2000) A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290:2319–2323.

Terpilowski M (2019) scikit-posthocs: Pairwise multiple comparison tests in python. *The Journal of Open Source Software* 4:1169.

Thura D, Cisek P (2014) Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron* 81:1401–1416.

Tibshirani R, Walther G (2012) Cluster Validation by Prediction Strength. *Journal of Computational and Graphical Statistics* 14:511–528.

Timme NM, Lapish C (2018) A Tutorial for Information Theory in Neuroscience. *eNeuro* 5:ENEURO.0052–18.2018.

Tosches MA, Yamawaki TM, Naumann RK, Jacobi AA, Tushev G, Laurent G (2018) Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* 360:881–888.

Trainito C, Nicolai Cv, Miller EK, Siegel M (2019) Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex. *Current Biology* .

Trautmann EM, Stavisky SD, Lahiri S, Ames KC, Kaufman MT, O'Shea DJ, Vyas S, Sun X, Ryu SI, Ganguli S, Shenoy KV (2019) Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron* .

Tremblay S, Acker L, Afraz A, Albaugh DL, Amita H, Andrei AR, Angelucci A, Aschner A, Balan PF, Basso MA, Benvenuti G, Bohlen MO, Caiola MJ, Calcedo R, Cavanaugh J, Chen Y, Chen S, Chernov MM, Clark AM, Dai J, Debes SR, Deisseroth K, Desimone R, Dragoi V, Egger SW, Eldridge MA, El-Nahal HG, Fabbrini F, Federer F, Fetsch CR, Fortuna MG, Friedman RM, Fujii N, Gail A, Galvan A, Ghosh S, Gieselmann MA, Gulli RA, Hikosaka O, Hosseini EA, Hu X, Hüer J, Inoue Ki, Janz R, Jazayeri M, Jiang R, Ju N, Kar K, Klein C, Kohn A, Komatsu M, Maeda K, Martinez-Trujillo JC, Matsumoto M, Maunsell JH, Mendoza-Halliday D, Monosov IE, Muers RS, Nurminen L, Ortiz-Rios M, O'Shea DJ, Palfi S, Petkov CI, Pojoga S, Rajalingham R, Ramakrishnan C, Remington ED, Revsine C, Roe AW, Sabes PN, Saunders RC, Scherberger H, Schmid MC, Schultz W, Seidemann E, Senova YS, Shadlen MN, Sheinberg

DL, Siu C, Smith Y, Solomon SS, Sommer MA, Spudich JL, Stauffer WR, Takada M, Tang S, Thiele A, Treue S, Vanduffel W, Vogels R, Whitmire MP, Wichmann T, Wurtz RH, Xu H, Yazdan-Shahmorad A, Shenoy KV, DiCarlo JJ, Platt ML (2020) An Open Resource for Non-human Primate Optogenetics. *Neuron* .

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.

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* 31:14235–14242.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Jarrod Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey C, Polat İ, Feng Y, Moore EW, Vand erPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Contributors S (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17:261–272.

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* 17:49–57.

Vormstein-Schneider D, Lin JD, Pelkey KA, Chittajallu R, Guo B, Arias-Garcia MA, Allaway K, Sakopoulos S, Schneider G, Stevenson O, Vergara J, Sharma J, Zhang Q, Franken TP, Smith J, Ibrahim LA, astro KJM, Sabri E, Huang S, Favuzzi E, Burbridge T, Xu Q, Guo L, Vogel I, Sanchez V, Saldi GA, Gorissen BL, Yuan X, Zaghloul KA, Devinsky O, Sabatini BL, Batista-Brito R, Reynolds J, Feng G, Fu Z, McBain CJ, Fishell G, Dimidschstein J (2020) Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. *Nature Neuroscience* pp. 1–8.

Štrumbelj E, Kononenko I (2014) Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems* 41:647–665.

Wang J, Narain D, Hosseini EA, Jazayeri M (2017) Flexible timing by temporal scaling of cortical responses. *Nature Neuroscience* 21:102–110.

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.

Williams AH, Kim T, Wang F, Vyas S, Ryu SI, Shenoy KV, Schnitzer M, Kolda TG, Ganguli S (2018) Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron* 98:1099–1115.e8.

Yang L, Jin R (2006) Distance metric learning: A comprehensive survey.

Zaitsev AV, Povysheva NV, Gonzalez-Burgos G, Rotaru D, Fish KN, Krimer LS, Lewis DA (2009) Interneuron Diversity in Layers 2–3 of Monkey Prefrontal Cortex. *Cerebral Cortex* 19:1597–1615.

Zaitsev A, Gonzalez-Burgos G, Povysheva N, Kröner S, Lewis D, Krimer L (2005) Localization of Calcium-binding Proteins in Physiologically and Morphologically Characterized Interneurons of Monkey Dorsolateral Prefrontal Cortex. *Cerebral Cortex* 15:1178–1186.

Zhu S, Xia R, Chen X, Moore T (2020) Heterogeneity of Neuronal Populations Within Columns of Primate V1 Revealed by High-Density Recordings. *bioRxiv* p. 2020.12.22.424048.

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 for clustering (Aggarwal et al., 2001). Specifically, as dimensionality increases, the difference between distances of randomly chosen points all converge to the same constant distance (Verleysen et al., 2003; Beyer et al., 1999). UMAP counters this by using graph distances on a nearest neighbor graph which is not susceptible to the concentration of measure phenomenon (McInnes, 2019). A common strategy is to cluster in some low-dimensional projected space with a dimensionality reduction method that preserves important latent structure. These use clustering methods like a Gaussian mixture model (GMM) or k-Means but clusters in low-dimension are not necessarily Gaussian in shape (an assumption of both the GMM and k-Means). This violation is induced by the perturbations introduced by non-linear projections even if the distributions were Gaussian in the original high-dimensional space.

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 (Tenenbaum et al., 2000) and t-SNE (Maaten and Hinton, 2008) (see (Lee and Verleysen, 2007) 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 embedded within the high-dimensional space. It is also assumed that the underlying manifold is locally connected (i.e. doesn't have any breaks or isolated points) which is satisfied by the local distance metric unit ball extending to of the 1st-nearest neighbor (Fig. 2A-ii.b). 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 (McInnes et al., 2018):

> **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 (using the manifold's distance metric) 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-nearest neighbor and the assumption of uniform sampling on the manifold is enforced.

To compensate for the impact of certain nearest neighbors in the ball lying much further than those closer by (as in very sparse regions in ambient space), the normalizing distances are transformed

by the exponential function (Fig. 2A-ii.c),

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

To unite these disparate metric spaces (each data point has a unique local notion of distance), category theory is used to show that an equivalent representation can be made from a fuzzy simplicial set via an adjunction that will not be defined here (see **Section 3** of McInnes et al. 2018 (McInnes et al., 2018) and the Nerve Theorem) In this way, the topological structure of the data can be represented as a metric space or as the union of a set of fuzzy simplices (Fig. 2A-ii.e). 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 (McInnes, 2019). 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 and implemented with a force directed graph layout procedure (Fig. 2B-v.a). Given two fuzzy sets with the same members A and separate membership strength functions μ and ν (of Spivak's characteristic form (Spivak, 2009)), 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 short edges in low-dimension or vice versa.

From a computational perspective, this whole UMAP process proceeds in two steps: construction of a k-nearest neighbor graph (Fig. 2A) and layout of the graph (Fig. 2B) into a low-dimensional manifold. Note that after the first step, the k-nearest neighbor graph is passed to Louvain community detection (Fig. 2B-iv) 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 layout parameters such as `min_dist`. The embedding *is* however used for visualization and interpretability and for consistency, the random seed for UMAP's layout procedure is set (UMAP will use gradient descent vs. stochastic gradient descent to compute the force directed graph layout).

> **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 (Dong et al., 2011). 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,
>
> $$\rho_i = \min\{d(x_i, x_{i_j}) | 1 \leq j \leq k, d(x_i, x_{i_j}) > 0\}$$

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,

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

. 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 (Belkin and Niyogi, 2002; Bengio et al., 2003; Kobak and Linderman, 2019). 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,

$$Ly = \lambda D y.$$

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,

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

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

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

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) (Blondel et al., 2008) operates on a weighted network and locates highly-interconnected nodes called a community. This "connectedness" is measured by their modularity Q (taking real values between -1 and 1 inclusive) with added resolution parameter t (Newman and Girvan, 2004; Lambiotte, 2007) 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 (Lambiotte et al., 2008). The larger the resolution parameter, the fewer the number of communities and the larger their size (Fig. S2A). 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 (Blondel et al., 2008). $A_{i,j}$ 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) which proceeds in two steps: modularity optimization and community aggregation.

Modularity Optimization: Each node is assigned to its own singleton community (in the initialization step for only the first pass) and then each node is moved into a community with a random neighbor and the change in modularity, ΔQ_t, is calculated. The equation for this change in modularity is,

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

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 caclulated 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 nodes, 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 summed to form a new graph. The process is then repeated from the previous step. This process repeats until the graph with maximum modularity is found and each original node is assigned to a final cluster membership.

The graph produced by UMAP was passed into this Louvain clustering algorithm (Blondel et al., 2008) 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. 3B. Each waveform was then plotted in UMAP space and color-coded to its associated cluster label found by Louvain Clustering Fig. 3A.

Ensemble clustering for graphs (ECG)

ECG is a consensus clustering method for graphs and was used to validate the Louvain clustering algorithm (Poulin and Théberge, 2018, 2019). 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 level-1 refers to only computing the first pass of Louvain). The randomization comes from the randomization of vertices in the initialization 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,
>
> $$W_\mathcal{P}(u,v) \triangleq \begin{cases} w_* + (1 - w_*) \cdot \left(\frac{\sum_{i=1}^{k} v_{\mathcal{P}_\rangle}(u,v)}{k} \right) & \text{, if } (u,v) \text{ is in the 2-core of } G \\ w_* & \text{, otherwise} \end{cases}$$
>
> 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 (Shapley, 1988) 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 (Lundberg and Lee, 2017) 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 ith one in ordering R and M is the number of input features.

Extending SHAP values to tree classifiers, the authors create `shap.TreeExplainer` (Lundberg et al., 2020) to calculate SHAP values by using path-dependent feature perturbation to yield the plots in Fig. 5B and C.

WaveMAP is stable over parameter choice with respect to random seed and data bootstrap

Parameter Choice: Louvain clustering (Blondel et al., 2008) requires the specification of a resolution parameter, 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 (Lambiotte, 2007).

We selected a resolution parameter based on two factors. The most important factor was modularity score. Modularity (the "connectedness" of a community, see Methods) is a community-wise measure 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 with high amounts of intra-connectivity and low out-connectivity. The second factor we considered was the sizes of the resulting clusters after choosing a resolution parameter. We did not want clusters with too few members (n $<$ 20 which would be statistically difficult to interpret). The regions with the highest modularity score were at a resolution parameter of 1 and decreased from there onwards. However, choosing a resolution parameter of 1 led to a large number of clusters which often had very few members (Fig. S2A, leftmost column) making downstream statistical comparisons underpowered. We therefore chose t to be 1.5 which resulted in the next best 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 data permutations (Fig. 3B). In this manner, we found a set of waveform clusters that balanced the diversity found by UMAP and statistical interpretability.

Random seed: To show this hierarchy of clustering resolutions along the curve in Fig. 3C 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, in red). As the data portion increased, we found that the number of clusters increased then tapered off to around eight clusters at $\sim$60% of the full dataset. In the same manner, we also calculated the adjusted mutual information score (AMI) (Timme and Lapish, 2018) between a subset of the data and the full dataset (Fig. S2B, in green). This is a measure of how closely two sets of clusterings agree with each other intuitively interpreted as how "informative" knowing one potential clustering is for predicting another. Just as with the number of clusters, AMI increases steeply until around 40% at a score of 0.8 of the dataset and then slowly increased. This analysis is reassuring and demonstrates that we have an adequate number of waveforms to yield consistent clustering with which 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 (Poulin and Théberge, 2019). To test whether these eight clusters consistently contained the same constituent data points run-to-run, we used ensemble clustering for graphs (ECG) (Poulin and Théberge, 2018, 2019). 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 $\pm$ 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. 3A). 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.

WaveMAP's robustness comes from clustering in high-dimension rather than the projected space

Another common approach to clustering high-dimensional data has been to cluster in the embedded space, that is, clustering in the lower-dimensional projected space of a dimensionality reduction method. While this approach is successful for similar applications such as spike sorting (Dimitriadis et al., 2018; Mahallati et al., 2019; Harris et al., 2000; Chah et al., 2011), we found it to be more unstable when used for our use case of classifying waveform types (Fig. S3). Using the robustness analysis in Fig. S2B, we compared *WaveMAP* against both DBSCAN (Ester et al., 1996) on the embedded 2-dimensional t-SNE (Maaten and Hinton, 2008) space (DBSCAN on t-SNE) and a GMM on the projected space formed by the first three principal components of the waveform data (94% variance explained). As before, we first applied each method to the full dataset to construct a "reference" to compare against. Thus each method is compared only to its own reference as to facilitate a fair head-to-head comparison.

DBSCAN on t-SNE and the GMM on PCA formed what seemed to be reasonable clusterings of the data with DBSCAN on t-SNE forming clusters similar to *WaveMAP* (Fig. S3A, top) and GMM on PCA forming four clusters just as in Fig. 4B albeit in PCA- and not the feature-space (Fig. S3A, bottom). When analyzing the AMI of clusterings across data bootstraps, *WaveMAP*'s AMI was moderate ($\sim$0.5) even at low percentages of the full dataset and quickly increased to high values (> 0.8) from about 50% of the dataset onwards (Fig. S3B, in blue). However, when we conducted our robustness analysis on the other methods, both of them proved to be less stable (Fig. S3B, green and red). With DBSCAN on t-SNE, although it was able to reach high AMI values when the data fraction increased past 75%, it was low to moderate before this point. In this regime of high data fraction, many of the points between random samples were shared and so much of this stability may be due to similarity in the data (when the fraction of the full dataset is high, most of the data is the same between random samples). Ideally, a method should be able to pick up on structure predictive of out-of-sample data at low data fractions i.e. results should generalize. With the GMM on PCA approach, AMI started out about as high as *WaveMAP* but failed to increase as more data was included. Examining the individual samples, we see that at 40% of the full dataset (Fig. S3C, left), a continuum of AMI values are occupied from low to high. At 90% of the full dataset (Fig. S3C, right), the AMI scores cluster around different locations and overall occupying the same range as at 40%. This suggests that the GMM on PCA, having nearly the full dataset, is converging to one of several solutions (local minima) with some of them being highly suboptimal.

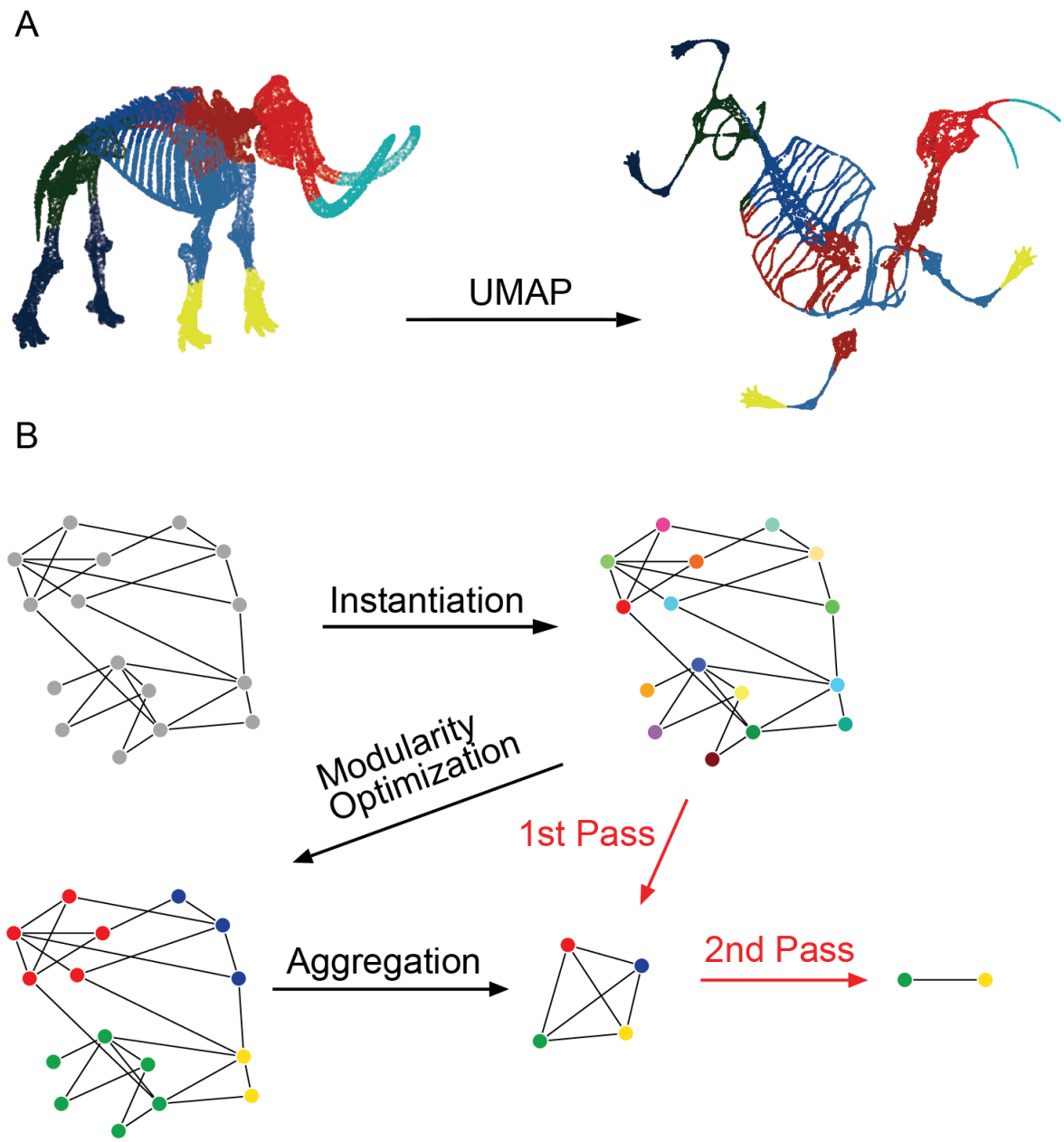

Figure S1: Diagrams of UMAP and Louvain Community Detection (supplement to Fig. 2): **(A)** A demonstration of UMAP projection on a 3D point cloud skeleton of a wooly mammoth. Local and global structures are incorporated and projected into lower dimension. This preservation of information is evident in the maintained structure of individual bone shapes and sensible spatial relationships between the body parts. Idea from M. 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 symmetric 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 modularity no longer increases. The final cluster memberships are then passed back to the data points on the original graph.

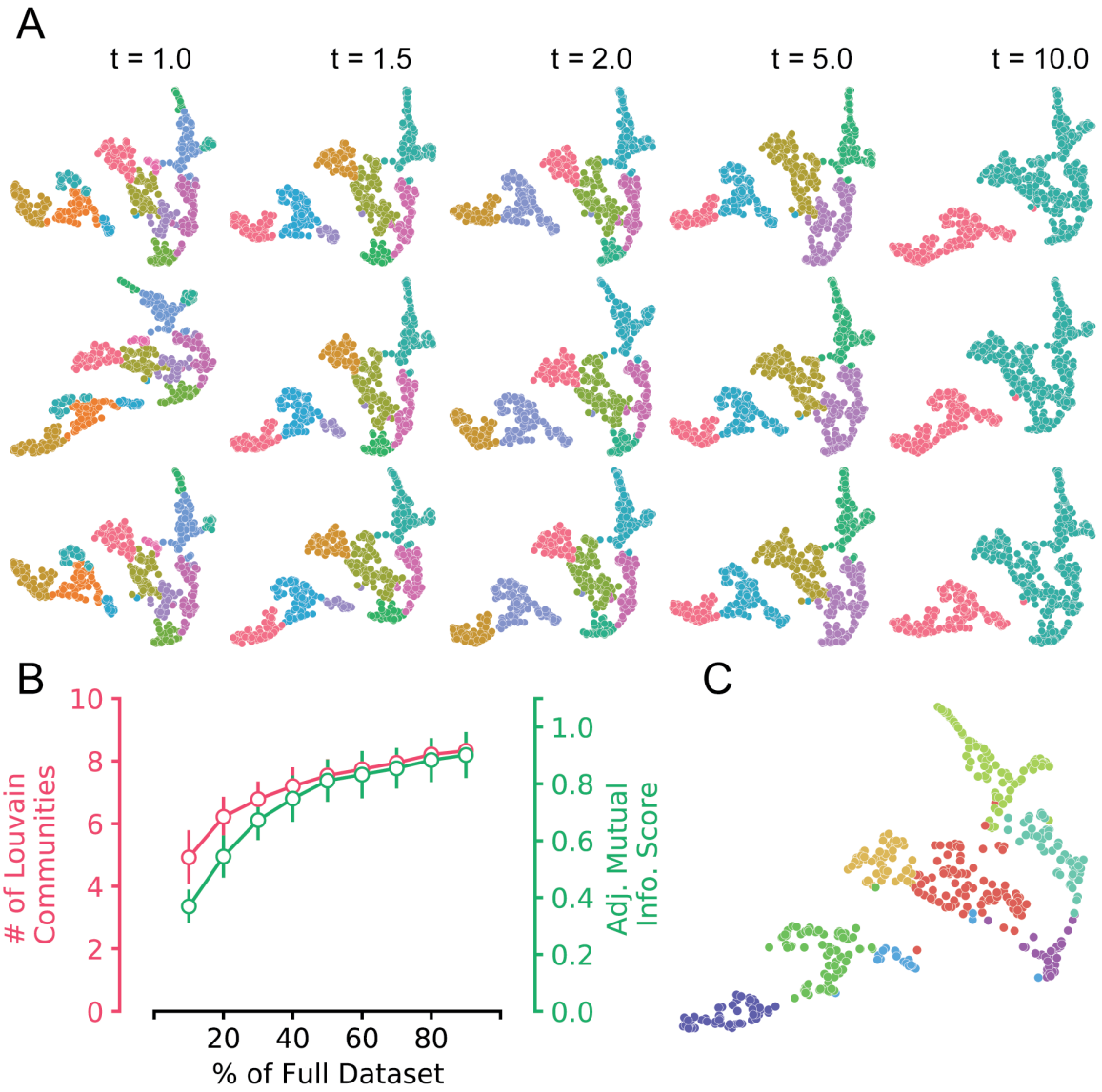

Figure S2: Stability analysis of *WaveMAP* (supplement to Fig. 3): **(A)** *WaveMAP* instantiated with three different UMAP random seeds (each row is a different seed) and Louvain `resolution` parameters. **(B)** The mean ± S.D. number of clusters (Louvain communities; in red) produced by *WaveMAP* and adjusted mutual information score (in green) across 100 random samples at various proportions of the full dataset. Number of clusters and adjusted mutual information (AMI) are omitted for 100% of the data because the number of clusters is equal to our result and thus AMI is 1.0 by definition of it being itself. **(C)** Ensemble clustering on graphs (ECG) is also applied to the UMAP graph. This also produced eight clusters.

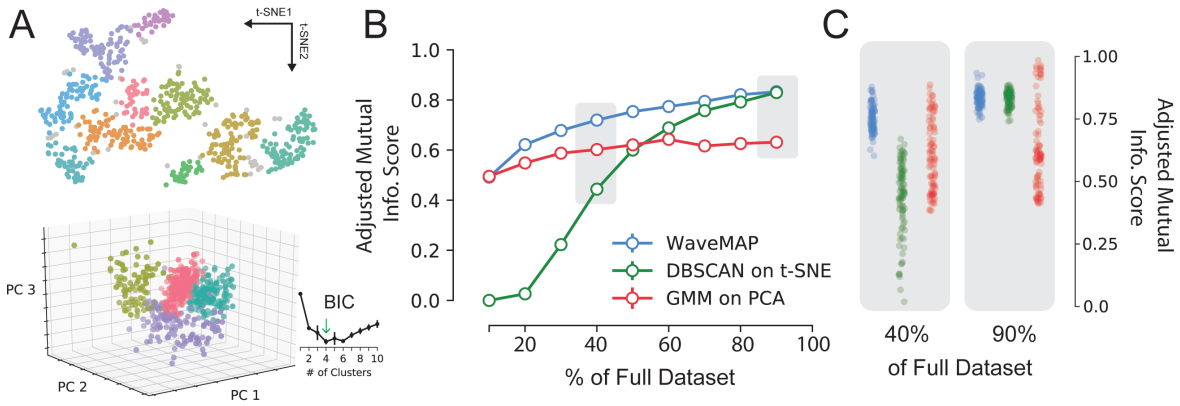

Figure S3: *WaveMAP* clusterings are more consistent than either DBSCAN on t-SNE or a GMM on PCA (supplement to Fig. 4): **(A)** At top, DBSCAN (`eps = 3`, `n_neighbors = 15`) was applied to t-SNE's projected space (`perplexity = 30`) over the full dataset producing ten clusters. Parameters were chosen to produce similar structure to *WaveMAP* to facilitate comparison. At bottom, a Gaussian mixture model (`n_components = 4`, `replicates = 25`) is applied to the 3-dimensional projected space (94% variance explained) produced by the first three principal components of the full dataset (GMM on PCA). The number of clusters chosen was by selecting the value at the elbow (green arrow) of a BIC $\pm$ S.D. vs. number of clusters plot shown in the inset. **(B)** Adjusted mutual information scores (AMI; mean $\pm$ S.E.M.) for 100 random sample clusterings of both *WaveMAP*, DBSCAN on t-SNE, and GMM on PCA. Standard error of the mean bars are smaller than marker size. The AMI was calculated by constructing a "reference" clustering by applying the respective method to the full dataset and then compared to the clusterings produced by random subsamples. **(C)** A jittered strip plot showing the AMI for 100 random subsets using each method. Subsets were randomly sampled from 40% (left) and 90% (right) of the full dataset. Gray boxes correspond to the data points surrounded by gray boxes in (B).

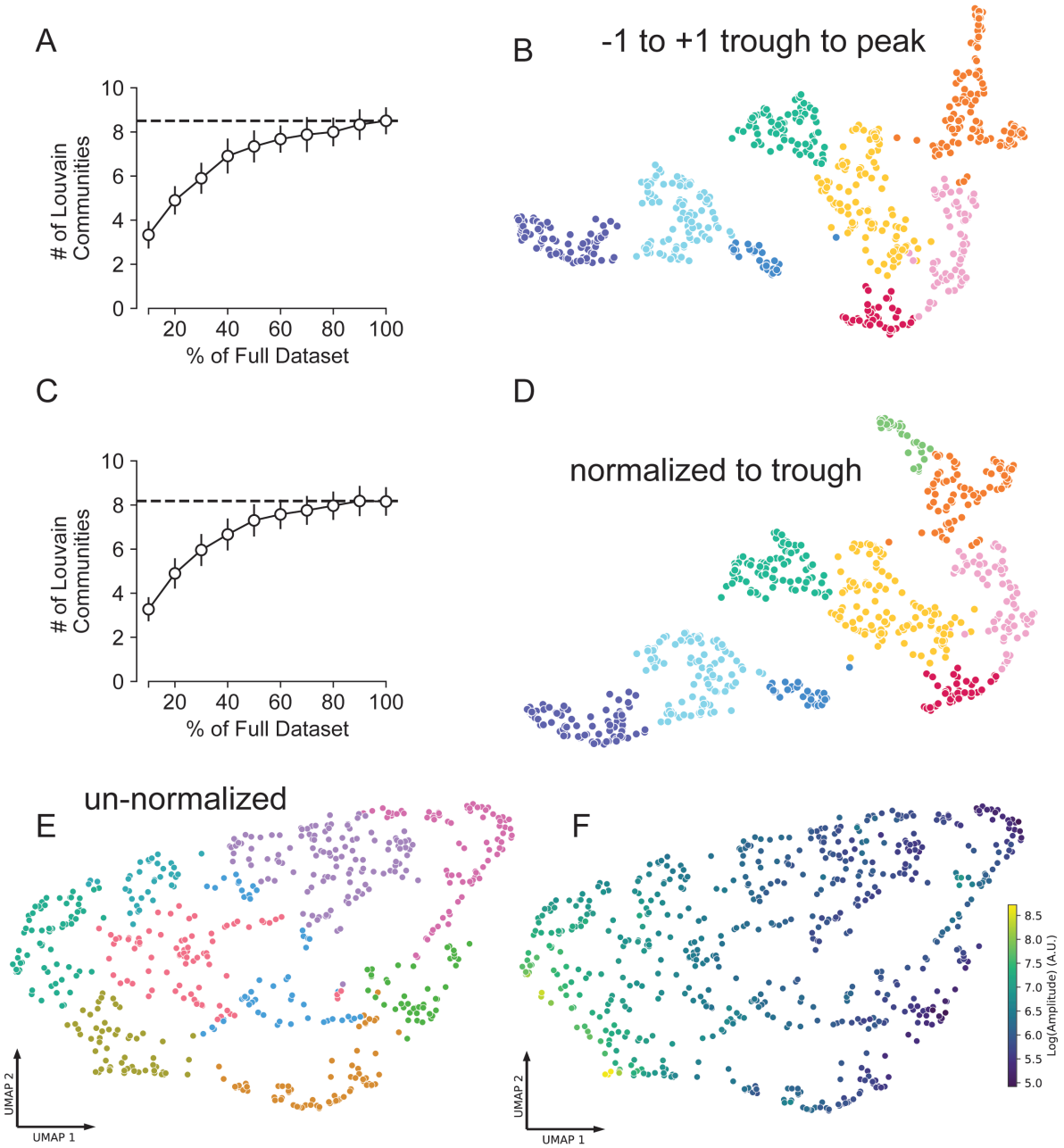

Figure S4: Different amplitude normalizations have similar effect but are essential for meaningful *WaveMAP* structure (supplement to Fig. 3): **(A)** As in Fig. S2B, the number of Louvain communities found across various random subsets and random seeds. The mean number of clusters shown on the full dataset with a dashed line. **(B)** the *WaveMAP* clusters on waveforms with ± 1 trough to peak normalization (used in the paper). **(C)** The same random subsetting and random seed strategy in (A) applied to waveform data normalized to trough depth. **(D)** *WaveMAP* clusters applied to waveform data normalized to trough depth. **(E)** Un-normalized waveforms were passed through *WaveMAP* with the same parameters used previously. **(F)** Each waveform in the projected UMAP space found in (E) is colored according to the amplitude (log of the difference between maximum and minimum values).

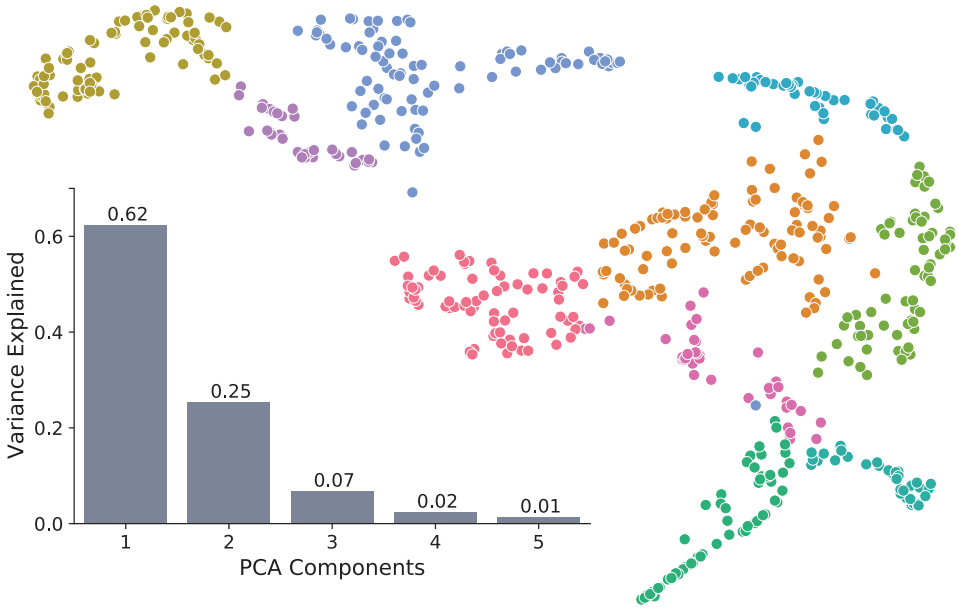

Figure S5: Pre-processing with PCA does not alter *WaveMAP* structure (supplement to Fig. 3): the full dataset was pre-processed with principal component analysis (PCA) and projected into the space of the first three principal components. The scree plot, with explained variance above each bar, shows that the dataset is low-dimensional with 94% of the variance explained in three components. The clustering on the right was produced by applying *WaveMAP* to the embedded dataset and is very similar to those produced by *WaveMAP* on data without pre-processing.

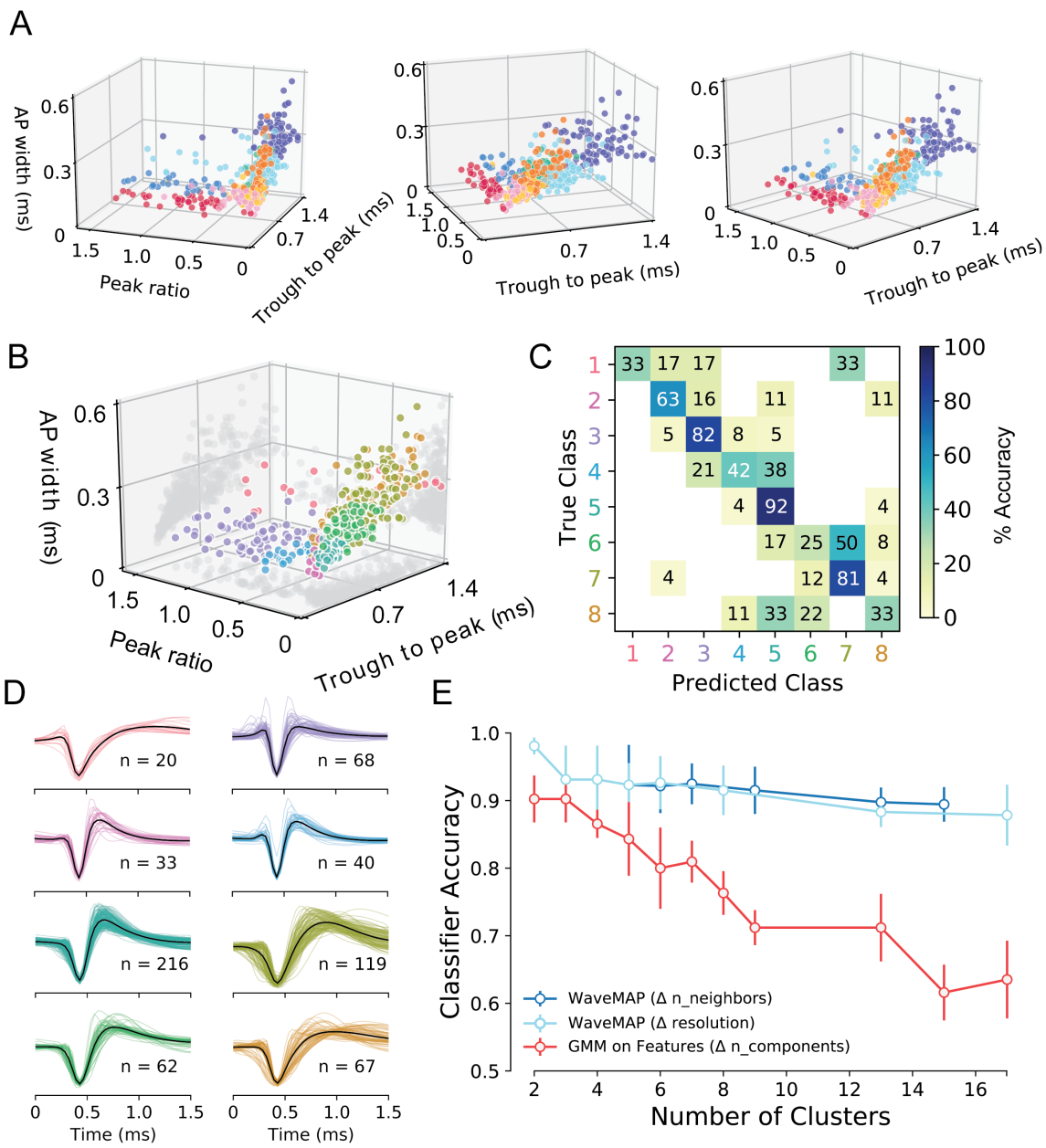

Figure S6: Comparison of GMM and UMAP in the constructed feature space (supplement to Fig. 4): **(A)** Three views of the eight *WaveMAP* clusters shown in the constructed feature space. The clusters maintain some structure but are largely mixed and linearly inseparable. **(B)** A GMM instantiated with eight clusters in the constructed feature space of Fig. 4B. **(C)** Confusion matrix for a gradient boosted decision tree classifier with the same hyperparameters as the one trained on four GMM classes (see hyperparameters in Table 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 with average waveforms in black. **(E)** Both *WaveMAP* and a GMM on features were used on the full dataset to generate results of various cluster number and a gradient boosted decision tree (hyperparameters optimized to the 4-class GMM) was trained on each. Shown is the classifier accuracy (mean $\pm$ S.D.) across stratified k-folds (`k = 5`) and various cluster number from 2 to 16. In dark blue, we generated *WaveMAP* mappings of different cluster number by changing the `n_neighbors` parameter associated with UMAP; in light blue we generated *WaveMAP* mappings of various cluster number by changing `resolution` associated with Louvain. In red, we changed the `n_components` for the Gaussian mixture model.

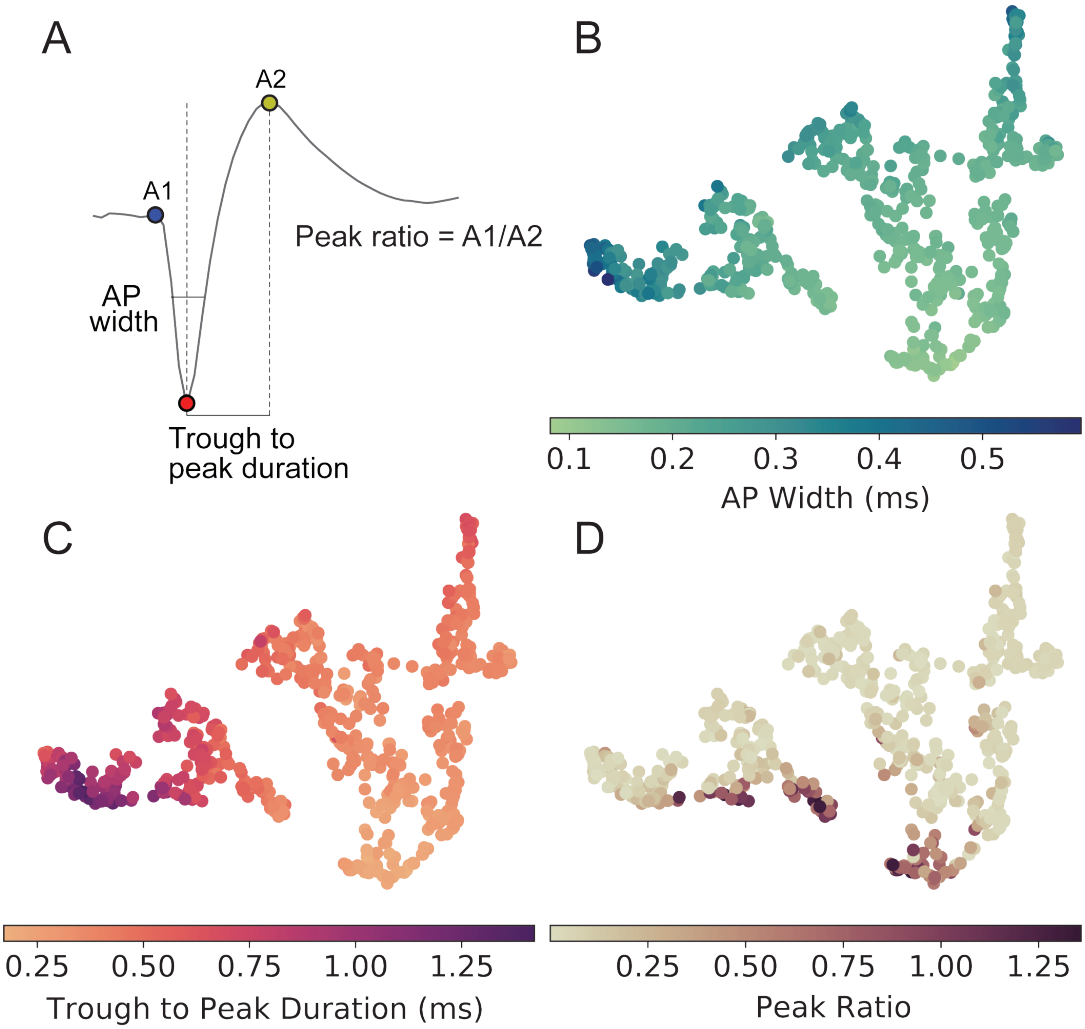

Figure S7: *WaveMAP* implicitly captures waveform features without the need for specification (supplement to Fig. 5): **(A)** Three waveform shape features used in traditional clustering approaches. The three EAP waveform landmarks used to generate the specified features passed to the GMM on a sample 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. 4A 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.

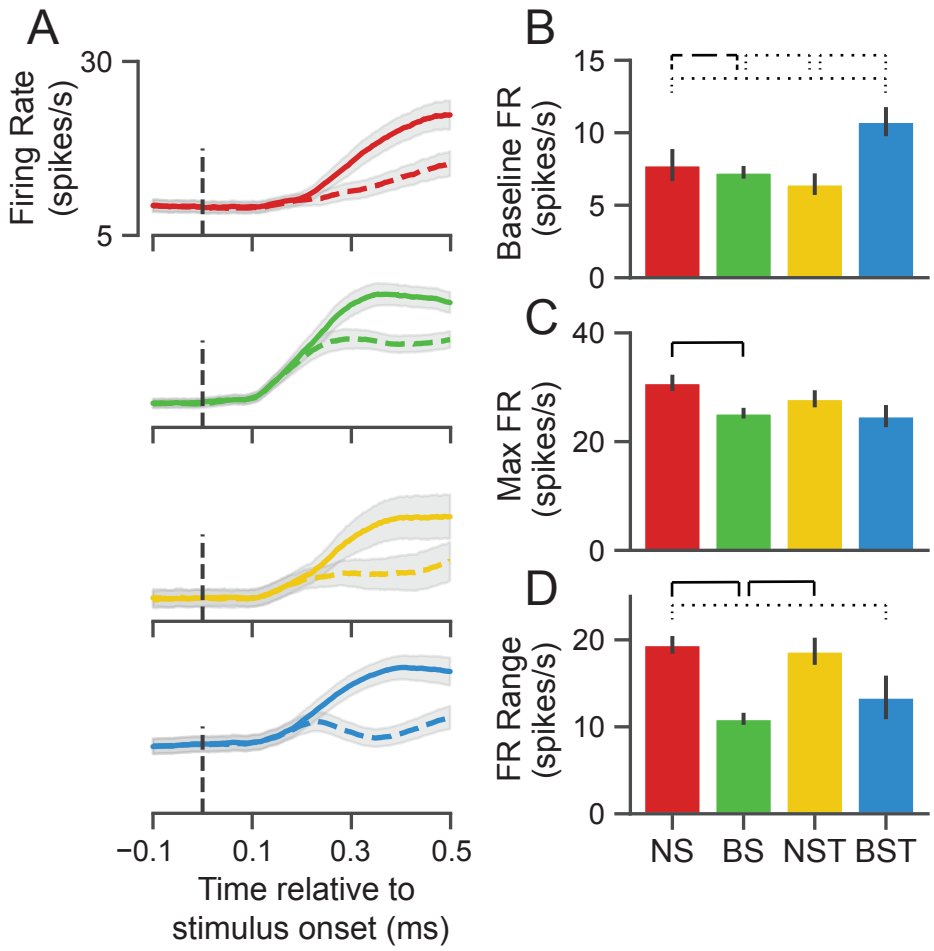

Figure S8: GMM clusters are less physiologically distinguishable than *WaveMAP* clusters (supplement to Fig. 6): **(A)** Stimulus-aligned trial-averaged firing rate activity in PMd for GMM clusters. As in Fig. 6, the traces are separated into PREF and NONPREF trials with solid and dashed lines respectively. Shaded regions correspond to bootstrapped standard error of the mean (S.E.M.). The dashed vertical line denotes the stimulus-onset time. **(B)** Baseline median firing rates (FR) $\pm$ S.E.M. for the four GMM clusters. Baselines were calculated as the average firing rate during the first 200 ms of the trial. (C) Median maximum FRs $\pm$ S.E.M. for the neurons in the four GMM clusters. This was caculated by taking the median of the maximum FR for each neuron across the entire trial. (D) Median FR range $\pm$ 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.

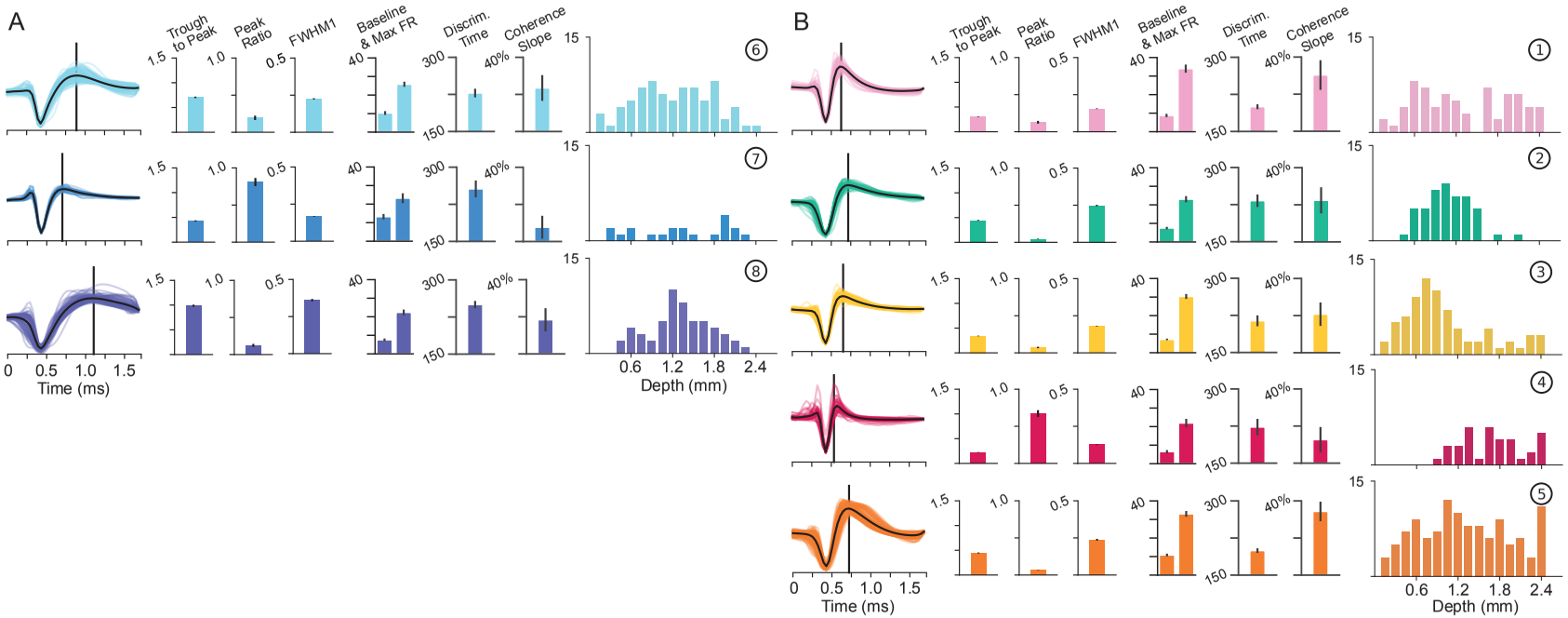

Figure S9: Detailed summary of each UMAP cluster and features (supplement to Fig. 6, Fig. 7, Fig. 8): **(A, B)** A detailed summary of broad- (A) and narrow-spiking (B) cluster waveform shapes, physiological measures, and laminar distribution. Each waveform shape is shown at left with the average waveform 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. 4A) 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`	20
		`min_dist`	0.1
		`random_state`	42
		`metric`	'euclidean'
Louvain Clustering (Python)	`cylouvain.best_partition`	`resolution`	1.5
UMAP Gradient Boosted Decision Tree (Python)	`xgboost.XGBClassifier`	`max_depth`	4
		`min_child_weight`	2.5
		`n_estimators`	100
		`learning_rate`	0.3
		`objective`	'multi:softmax'
		`rand_state`	42
GMM Gradient Boosted Decision Tree (Python)	`xgboost.XGBClassifier`	`max_depth`	10
		`min_child_weight`	2.5
		`n_estimators`	110
		`learning_rate`	0.05
		`objective`	'multi:softmax'
		`seed`	42
8-Class GMM Gradient Boosted Decision Tree (Python)	`xgboost.XGBClassifier`	`max_depth`	2
		`min_child_weight`	1.5
		`n_estimators`	100
		`learning_rate`	0.3
		`objective`	'multi:softmax'
		`seed`	42
Gaussian Mixture Model (MATLAB)	`fitgmdist`	k	4
		`start`	'randsample'
		`replicates`	50
		`statset('MaxIter')`	200
DBSCAN (Python)	`sklearn.cluster.DBSCAN`	eps	3
		min_samples	15

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