Article

Current Biology

Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex

Highlights

- 2,488 single neurons recorded in macaque dlPFC, FEF, and LIP
- Identification of four distinct cell classes from extracellular spike waveforms
- Cell classes differ in firing statistics, response dynamics, and information coding
- Cell classes are robust across cortical regions

Authors

Caterina Trainito, Constantin von Nicolai, Earl K. Miller, Markus Siegel

Correspondence

markus.siegel@uni-tuebingen.de

In Brief

Trainito et al. use a data-driven approach to robustly identify four cell classes from extracellular spike waveforms recorded in three cortical regions of macaque monkeys. The four cell classes are functionally distinct in terms of firing statistics, response dynamics, and information coding.

Trainito et al., 2019, Current Biology *29*, 1–10
September 23, 2019
https://doi.org/10.1016/j.cub.2019.07.051

CellPress

Please cite this article in press as: Trainito et al., Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.07.051

Current Biology
Article

Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex

Caterina Trainito,[1,2,3,4] Constantin von Nicolai,[1,2,3] Earl K. Miller,[5,6] and Markus Siegel[1,2,3,6,7,*]
[1]Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Strasse 25, 72076 Tübingen, Germany
[2]Hertie Institute for Clinical Brain Research, University of Tübingen, Otfried-Müller-Strasse 27, 72076 Tübingen, Germany
[3]MEG Center, University of Tübingen, Otfried-Müller-Strasse 47, 72076 Tübingen, Germany
[4]IMPRS for Cognitive and Systems Neuroscience, University of Tübingen, Österbergstrasse 3, 72074 Tübingen, Germany
[5]The Picower Institute for Learning and Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[6]Senior author
[7]Lead Contact
*Correspondence: markus.siegel@uni-tuebingen.de
https://doi.org/10.1016/j.cub.2019.07.051

SUMMARY

Understanding the function of different neuronal cell types is key to understanding brain function. However, cell-type diversity is typically overlooked in electrophysiological studies in awake behaving animals. Here, we show that four functionally distinct cell classes can be robustly identified from extracellular recordings in several cortical regions of awake behaving monkeys. We recorded extracellular spiking activity from dorsolateral prefrontal cortex (dlPFC), the frontal eye field (FEF), and the lateral intraparietal area of macaque monkeys during a visuomotor decision-making task. We employed unsupervised clustering of spike waveforms, which robustly dissociated four distinct cell classes across all three brain regions. The four cell classes were functionally distinct. They showed different baseline firing statistics, visual response dynamics, and coding of visual information. Although cell-class-specific baseline statistics were consistent across brain regions, response dynamics and information coding were regionally specific. Our results identify four functionally distinct spike-waveform-based cell classes in primate cortex. This opens a new window to dissect and study the cell-type-specific function of cortical circuits.

INTRODUCTION

Neuronal cell types are central to brain function. The unique physiology, morphology, and connectivity of different cortical interneurons and pyramidal cells shape their functional role in local and large-scale circuit operations [1–3]. Cell-type-specific neuronal properties shape characteristic circuit oscillations associated with various computational and cognitive processes [4–6]. Thus, knowledge about cell types and their role in cortical circuits is key to understanding brain function.

The assessment of cell types ideally relies on morphological, molecular, or genetic markers [7, 8]. Although these markers are often not available for extracellular electrophysiology studies, firing patterns and action-potential shape also provide some handle on cell-type diversity. *In vitro* studies first demonstrated that morphologically identified pyramidal cells and GABAergic interneurons differ in firing patterns and action-potential shape. Pyramidal cells show regular, low-rate firing patterns and have broad spike waveforms (“broad-spiking” units), whereas inhibitory cells fire at sustained high frequencies with characteristically thin spike waveforms (“narrow-spiking” units) [9–11]. In principle, these intracellular features map onto extracellular features recorded *in vivo* [12].

Based on these findings, several studies have inferred putative cell types from extracellular single-unit activity. In primate prefrontal cortex (PFC) [13–16], frontal eye field (FEF) [17, 18], inferior temporal (IT) cortex [19, 20], and V4 [21, 22], spike-waveform width is bimodally distributed, indicative of the known separation between excitatory cells and inhibitory interneurons. The proportion of narrow-spiking units in these studies (around 15%–25%) is consistent with anatomical estimates of the proportion of GABAergic cells in the cortex [23] (note laminar variability [24, 25]). Firing properties, selectivity, and task-related modulations differ between broad- and narrow-spiking units, further supporting the physiological interpretation of distinct cell types [16, 26]. In sum, so far waveform width has been shown to be informative about cell-type diversity in the primate brain, allowing to dissociate two broad classes of putative cell types (excitatory versus inhibitory). However, in order to better understand cell-type-specific mechanisms and functions, more cell types need to be identified. Furthermore, cell-type classification needs to be compared across different cortical regions.

To address this, we characterized putative cortical cell types based on spike waveforms in a large dataset of extracellular recordings from three different cortical regions (FEF, dorsolateral prefrontal cortex [dlPFC], and lateral intraparietal area [LIP]) in two macaque monkeys [27]. In contrast to the typically reported dichotomy between broad-spiking and narrow-spiking units, we were able to distinguish four cell classes based on waveform shape. These four distinct cell classes were

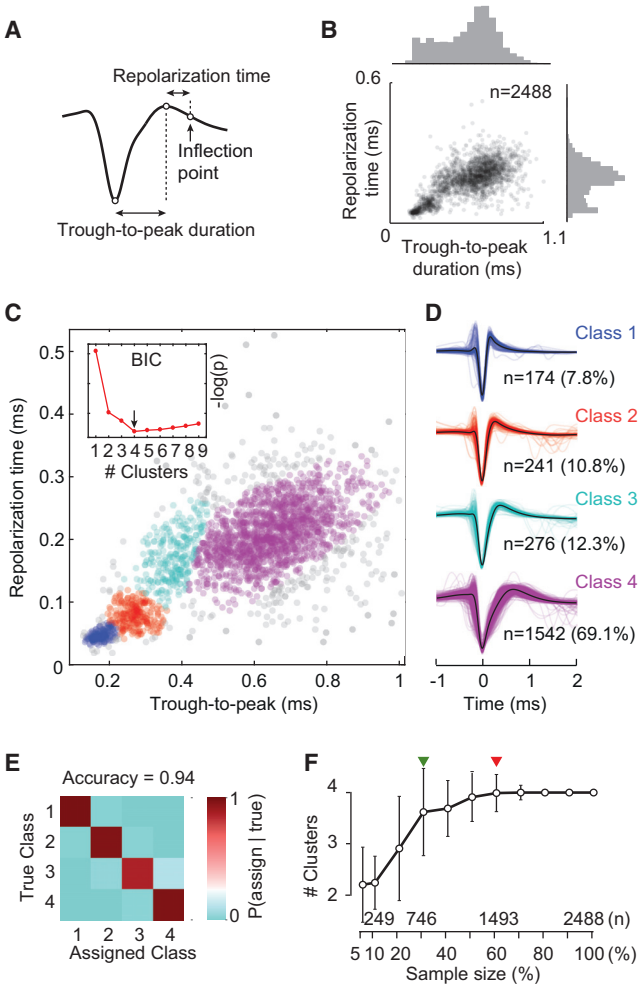

Figure 1. Cluster Analysis of Extracellular Spike Waveforms

(A) Illustration of the two spike-waveform features used for classification.

(B) 2D feature space and marginal distributions of waveforms for all recorded single units.

(C) Clusters of spike waveforms obtained from the Gaussian mixture model. Single units are assigned to the cluster with the highest posterior probability. Gray data points are excluded outliers: an initial fifth high-variance cluster and outliers of the Gaussian mixture distribution (n = 281). Inset: the negative log likelihood of the BIC as a function of the number of clusters after outlier removal is shown.

(D) All waveforms by cell class (average waveforms are in black).

(E) Class separation. To quantify the separation of the four clusters, 10^4 data points were randomly generated from the fitted Gaussian mixture distribution, and their true cluster was compared with their assigned cluster. The classification outcome is shown by the confusion matrix of marginal probabilities. Accuracy is the mean of the four diagonal probabilities.

(F) Mean and SD of the number of identified clusters across 100 random subsamples of the original data for different sub-sample sizes. The number of identified clusters drops for smaller sample sizes. Green arrowhead: for 30% of the original sample size (746 units), 4 clusters are identified for half of the sub-samples. Red arrowhead: for 60% of the original sample size (1,493 units), 4 clusters are identified for 95% of the sub-samples.

See also Figure S1.

confirmed by cell-class-specific firing patterns, response dynamics, and information coding. Although the four cell classes were consistently found across all cortical regions, their functional profiles differed between areas. These findings open a new window into cell-type-specific functions in awake behaving animals.

RESULTS

Cell-Class Separation Based on Spike Waveform

We analyzed data from 2,488 single units recorded in the FEF (793), dlPFC (1,050), and LIP (645) of two macaque monkeys (Figure 1). In a first step, we identified different cell classes in a purely data-driven fashion based on spike waveform. To increase statistical power, we pooled the data across all cortical regions and, for each unit, quantified two parameters of the spike waveform that contribute to the overall spike width: trough-to-peak duration and repolarization time (Figure 1A). Trough-to-peak duration is the interval between the global minimum of the curve and the following local maximum. Repolarization time is the interval between the late positive peak and the inflection point of the following falling flank of the curve. Although correlated, these two measures capture different aspects of the intracellular action potential—the speed of depolarization and of the subsequent after-hyperpolarization [12]—that are both distinguishing features of neuronal cell types [28]. All 2,488 waveforms were scored on the two measures to obtain a two-dimensional feature space for classification (Figure 1B).

To identify different cell classes in an unsupervised way, we performed a two-dimensional cluster analysis of the waveform parameters (Gaussian mixture model). We used the Bayesian information criterion (BIC) to select the number of Gaussian components in the model. The BIC showed a global minimum for four components indicating four distinct waveform classes (Figure 1C). Ranging from narrow to wide waveforms, the four classes comprised 7.8%, 10.8%, 12.3%, and 69.1% of the sample, respectively (Figure 1D). Thus, most units were attributed to the widest waveform class (class 4). We quantified cluster separation by calculating the probability of correctly classifying each cell class based on the Gaussian mixture model underlying the clustering (Figure 1E). The average classification accuracy across all four classes was 94%, indicating well-separated clusters.

To assess the effect of the large sample size on the number of identified clusters, we sub-sampled the data at various sub-sample sizes (100 random sub-samples for each size) and repeated the cluster analysis (Figure 1F). As expected, the number of identified clusters dropped for smaller sample sizes. 30% (746 units) and 60% (1,493 units) of the original sample were required to identify 4 clusters in at least 50% and 95% of the sub-samples, respectively (Figure 1F, green and red arrowheads).

To compare the present result to previous approaches separating waveforms into only two classes (narrow versus broad) [13–15, 18, 20–22, 29, 30], we performed a 2-class Gaussian mixture model clustering on the trough-to-peak duration only (Figure S1). This revealed that a 2-class separation would have split the intermediate class 3, assigning it to both narrow- and broad-waveform categories.

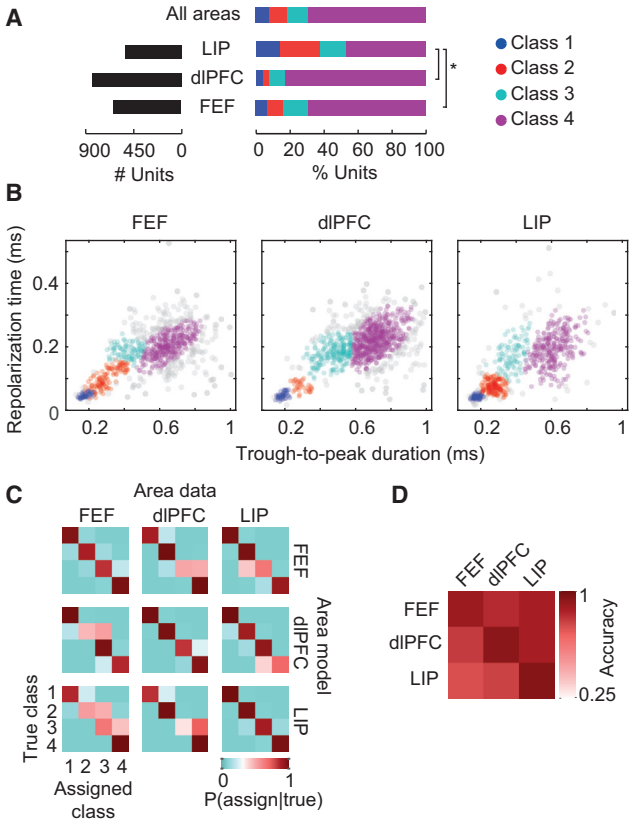

Figure 2. Reliability of Waveform Clustering across Cortical Regions

(A) Distribution of units across cortical regions and cell classes. Brackets indicate significant post hoc χ^2 tests for different cell-class distributions across areas ($p < 0.05$, Bonferroni corrected).

(B) Waveform feature spaces with clustering run separately for the FEF, dlPFC, and LIP.

(C) Cluster separation and similarity of cell classes across areas. Confusion matrices on the diagonal show separation of the four clusters for each area's own Gaussian mixture model (as in Figure 1E). For all other area pairs, confusion matrices measure the similarity between the same cell-class clusters in the two areas. Cluster similarity is estimated by randomly generating 10^4 data points from one area's Gaussian mixture distribution ("Area data") and classifying them based on the Gaussian mixture distribution of the other area ("Area model").

(D) Mean diagonal probabilities of confusion matrices in (C).

Cell Classes across Cortical Regions

We next investigated whether the waveform-based cell classes were robust across different cortical areas (Figure 2). Splitting the data by areas revealed that the four classes were unequally distributed across cortical regions (χ^2 omnibus test, $p < 0.001$; Figure 2A). Thus, we asked whether the four waveform clusters were consistently identified within each region. Indeed, clustering run separately on each area consistently returned four classes with the same overall structure (Figure 2B).

To estimate the cluster separability within each area, we quantified the probability of correctly classifying each class based on the Gaussian mixture model within each region (Figure 2C, diagonal plots). Furthermore, to estimate the waveform class similarity across brain regions, we quantified cross-classification accuracy between different regions, i.e., we trained and tested the classifier on different regions (Figure 2C, off-diagonal plots). For both cases and across all brain regions, classification accuracy was above 75% (Figure 2D). This indicates both a consistently high separation between the four clusters within each region and a high overlap of each cluster across regions. In sum, the four waveform-based cell classes were robustly and similarly observed across the three cortical regions.

Firing Statistics of Cell Classes

What are the functional properties of the four putative identified cell types? If the four spike-waveform clusters reflect distinct physiological cell types, the corresponding units should show different functional characteristics. We started by examining firing statistics during the 500-ms blank fixation baseline before stimulus onset of a flexible visual decision-making task (see Figure 4A for task timing). For each neuron, we computed four statistics during this trial period: mean firing rate (FR) across trials, Fano factor (variance over mean of spike counts across trials; FF), coefficient of variation of the inter-spike interval distribution (CV_{ISI}), and burst index (BI). Both Fano factor and CV_{ISI} are mean-standardized measures of dispersion that reflect firing variability, with an expected value of 1 for Poisson firing and values below 1 indicating more regular firing [31]. Burst index was defined as the ratio between the observed proportion of bursts (inter-spike intervals < 5 ms) and the proportion of bursts expected for a Poisson process with equal mean rate [13]. To rule out a confound due to the region-specific distribution of cell classes, we stratified the proportion of cells per cell class across regions (STAR Methods). One-way ANOVAs showed significant cell-class separation on all four measures (all $p < 0.05$) (Figure 3A). Firing rate was highest for class 1 units (narrow waveforms), followed by the two intermediate-waveform classes 2 and 3 (not significantly different from each other), and lowest for class 4 (broad-spiking units). Fano factor showed a similar pattern: class 4 had the lowest Fano factor and therefore more regular firing, also confirmed by the low CV_{ISI}. These results agree with the classical designation of narrow-waveform neurons as fast spiking (FS) and broad-waveform neurons as regular spiking (RS) [9, 10, 28]. On the other hand, the intermediate-waveform class 3 was more likely to fire in bursts than any other class.

Firing Statistics Validate Four Cell Classes

The significant differences of firing statistics between cell classes support the conclusion that the four waveform-defined cell classes reflect distinct physiological cell types. To further validate this conclusion, we employed a machine-learning approach: assuming the waveform-based classes as ground truth, we trained a multivariate classifier (SVM; support vector machine) to decode these four cell classes from all four firing statistics. Again, if the four waveform clusters reflect distinct cell types, class membership should be predictable from functional cell properties. Indeed, we were able to significantly predict all four cell classes with high classification accuracy (Figure 3B; classifier accuracy, 0.53 ± 0.02; mean ± SD over 50 area-stratified sub-samples, all $p < 0.05$, false discovery rate [FDR] corrected, binomial test).

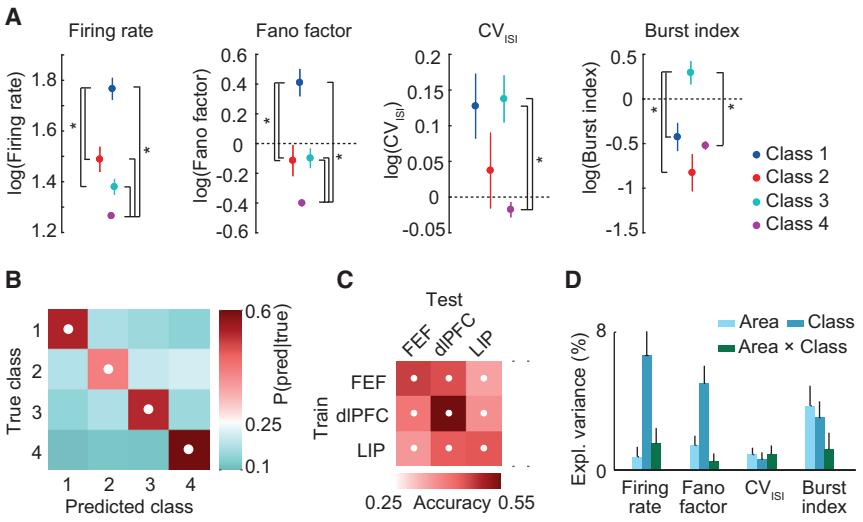

Figure 3. Cell-Class-Specific Baseline Firing Statistics

(A) Firing statistics by cell class. All measures were computed during a 500-ms blank fixation period at the beginning of each trial. CV_{ISI}: coefficient of variation of the inter-spike interval distribution. Burst index: proportion of inter-spike intervals <5 ms over the proportion expected from a Poisson neuron. Brackets indicate significant post hoc pairwise differences ($p < 0.05$, FDR corrected). Error bars denote SEM. Means and SEMs were area-stratified.

(B) Confusion matrix for supervised classification of cell classes using the four baseline firing statistics as features. White dots indicate overall performance significantly above chance (binomial test, $p < 0.05$, FDR corrected).

(C) Mean diagonal probabilities ("Accuracy") for cross-area classification. Classifiers trained on data from one area ("Train") were used to predict class labels of the other area ("Test"). White dots indicate significant class prediction (permutation test, $p < 0.05$, FDR corrected).

(D) Variance explained by cell class, cortical area, and their interaction in a two-way ANOVA performed on each firing statistic by area. All main and interaction effects were significant ($p < 0.05$). Error bars denote SEM.

See also Figure S2.

The classification approach allowed us to test whether class-specific functional properties were stable across cortical regions. We trained classifiers on data from one cortical area and tested them on data from a different area (cross-classification). Classification performance always exceeded chance level (permutation test, all $p < 0.05$, FDR corrected; Figure 3C), suggesting that cell classes maintain their functional profiles across regions. This was also indicated by the univariate measures broken down by area (Figure S2). Although there was a significant interaction between cell class and area (two-way ANOVAs, all measures $p < 0.05$), the cell-class effect was significant for all statistics and, for firing rate and Fano factor, most of the variance was independently explained by cell class (Figure 3D). In sum, firing statistics differed between, and thus validated, four waveform-based cell classes that were robust across cortical regions.

Cell-Class-Specific Firing Dynamics

Next, we investigated whether the four cell classes differed in their firing dynamics in response to a sensory stimulus (Figure 4). We characterized each neuron's response to presentation of a visual cue that indicated the upcoming task condition for each trial of the perceptual decision-making task (see the schematic of behavioral paradigm in Figure 4A). We computed peristimulus time histograms (PSTHs) in a window including the baseline fixation period (0.5 s) and the subsequent cue period (1 s).

The cell-class-specific PSTHs pooled across regions suggested differences of the response dynamics between cell classes (Figure 4B). For example, class 4 showed less transient responses to cue onset as compared to the other cell classes (see also Figure S3 for individual regions). To statistically assess this in an efficient way, we captured the firing dynamics in a low-dimensional space. We performed a principal component analysis (PCA) of the PSTHs of all neurons pooled across regions. We then used a cross-validation procedure to estimate the effective rank of the dataset, i.e., the number of underlying orthogonal dynamical features or principal components. We found four significant components (Figure 4C), explaining 63% of the response dynamics' variance and capturing different response modulations.

We projected each neuron's PSTH onto the four significant PCs and used the resulting low-dimensional representation of the response dynamics as features for multivariate decoding. Importantly, we normalized PSTHs on the mean spike rate of the baseline period. This ensured that the decoder was not classifying merely based on differences in overall activity levels but that it specifically probed cue-related responses. Again, to rule out a region-confound, we stratified the proportion of cells per class across regions. We found that cell classes 3 and 4 could be significantly decoded, confirming cell-class-specific PSTH dynamics (Figure 4D; binomial test on confusion matrix probabilities, both $p < 0.05$, FDR corrected for multiple comparisons; classifier accuracy, 0.32 ± 0.03; mean ± SD). Classes 1 and 2 showed similar decodability, though they did not reach statistical significance, likely due to the smaller number of units for these two classes.

Averaging the PSTHs of all units within each brain region revealed different response dynamics across regions (Figure 4E). Thus, we hypothesized that cell-class-specific response dynamics would be area specific. To test this, we performed cross-area decoding of cell classes based on response dynamics. We performed a PCA and trained the classifier on a training area and applied the PCA and classifier to another test area. Indeed, we found that, unlike for baseline firing statistics, cell-class decoding based on response dynamics did not significantly generalize across areas (Figure 4F). Accordingly, the cell classes' response dynamics showed dissimilar patterns in the three areas (Figure S3). For example, whereas class 4 showed the least transient responses in all areas, class 1 showed the most transient responses only in LIP (compare Figures 4B and S3).

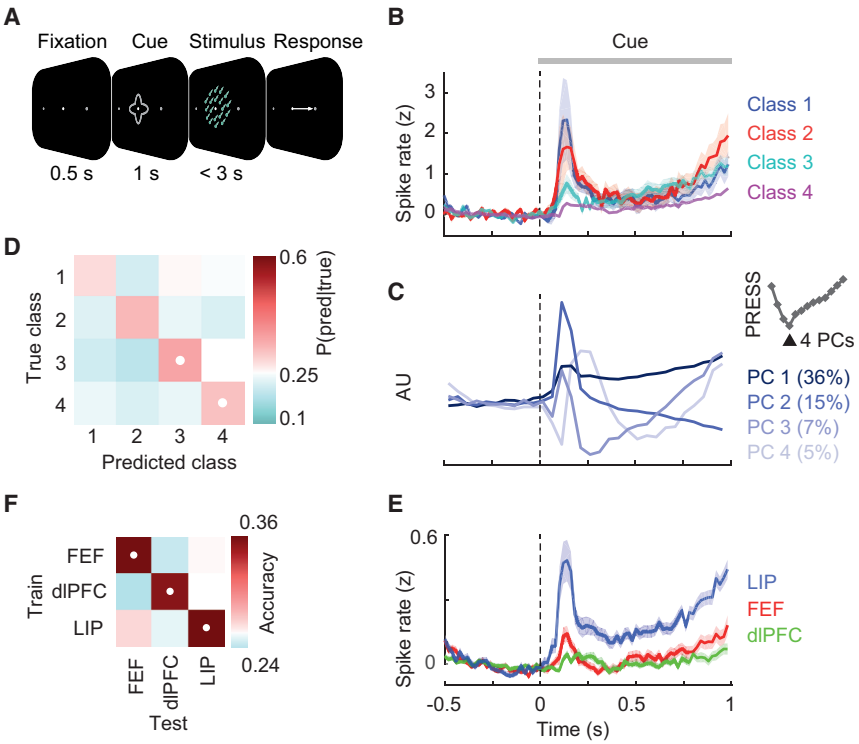

Figure 4. Cell-Class-Specific Response Dynamics

(A) Schematic of the behavioral task.

(B) Average PSTH for each cell class. PSTHs were *Z* scored on the mean and SD of the baseline period across trials. PSTH means and their SEs (shaded regions) were calculated after stratifying cell-class proportions across areas. The SE of class 4 is overlaid by the mean trace.

(C) Four significant principal components (PCs) explaining the PSTH variance across cell classes. Percentages denote the variance explained by each PC. Top right inset: the reconstruction error (PRESS; prediction residual error square sum) as a function of the number of principal components is shown. PRESS is minimal for 4 components.

(D) Confusion matrix for supervised classification of cell classes using PCA projections of the PSTHs. White dots indicate significant class prediction (binomial test, $p < 0.05$, FDR corrected).

(E) Average PSTHs for the units recorded within each of the three brain areas. Error bars denote SEM across units.

(F) Mean diagonal probabilities (Accuracy) for cross-area classification. Classifiers trained on data from one area (Train) were used to predict class labels of the other area (Test). The PCA transformation was estimated on the training area and applied to data of the test area. White dots indicate significant class prediction (permutation test, $p < 0.05$, FDR corrected).

See also Figure S3.

Cell-Class-Specific Information Coding

If cell classes vary in their cue-evoked response dynamics, do they also differentially code for specific cues? To address this question, for each neuron, we quantified the amount of cue information encoded by its firing rate, by measuring the amount of firing-rate variance across trials explained by cue identity (Figure 5A; ANOVA, 4 cues). We then trained a classifier to decode cell classes based on cue information. Again, we controlled for a potential confound of area by stratifying cell classes across brain regions. Furthermore, to control for confounds due to firing statistics, before classification, we regressed out linear dependencies of cue information on baseline firing statistics (firing rate, Fano factor, coefficient of variation of the ISI distribution, burst index). We found that cell classes 2 and 4 could be significantly decoded from cue information (Figure 5B). We next performed cross-area classification to assess the region specificity of class-specific information. We found that cross-area classification performance was low (Figure 5C). Thus, although neurons of classes 2 and 4 on the whole carried different cue-related information, the pattern of cue information across cell classes was area specific. This was confirmed by plotting average information for each cell class and region (Figure 5D), suggesting that, e.g., cell class 4 was more cue informative than classes 2 and 3 in dlPFC but less informative than classes 2 and 3 in the FEF.

Specificity of Functional Properties

Having established that the four cell classes differ in baseline activity, response dynamics, and information coding, we pooled together all three feature sets to construct an "omnibus" decoder that could predict all cell classes well (Figure 6A; mean accuracy, 0.49). To assess each feature's relative contribution to classification, we recast the problem in a linear framework (linear discriminant analysis; LDA) and used the univariate class effects, normalized to a common scale, as a proxy for feature importance. We computed feature importance for each of the six pairwise cell classifications (Figure 6C) and then averaged to show the overall weightings (Figure 6D). Furthermore, we compared cell-class classification accuracy (Figure 6E) and area specificity (Figure 6F) for each individual feature set and all combined sets.

These analyses showed that cell classes were most strongly separable by the four baseline firing statistics. This separation was most consistent across cortical regions (compare Figure 3C), suggesting that cell types maintain their basic firing properties even when embedded in functionally diverse areas. Although also showing class effects, cue-related response dynamics and information coding were less cell-class specific, and to a greater extent reflected area-specific process. Furthermore, pairwise feature importance (Figure 6C) showed that cell-class separation differed for distinct response dynamics depending on which two classes were being compared.

Finally, we performed two control analyses to rule out potential confounds (Figure 7). First, we ruled out that the observed effects were driven by a systematic difference in single-unit sorting quality between the four cell classes. To this end, we employed two measures of sorting quality: a subjective quality index (QI) that was manually specified for each unit during sorting, and the Mahalanobis distance of each unit's waveform to the unsorted noise-waveform cluster of the corresponding electrode and recording. Indeed, both measures showed a significant cell-class effect (QI: $p < 0.001$, χ^2(degrees of freedom

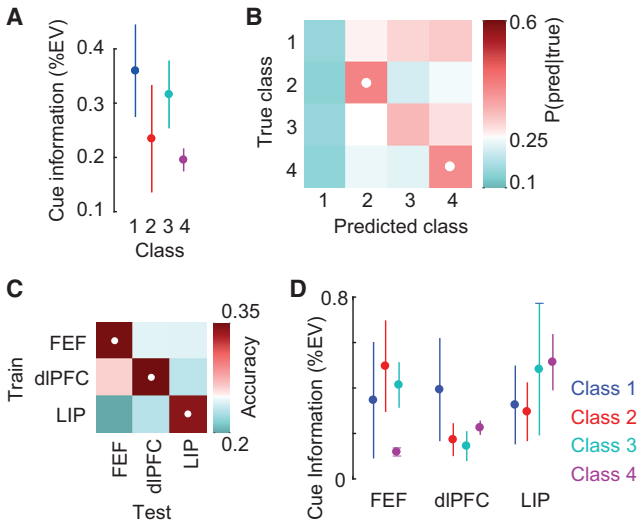

Figure 5. Decoding of Cell Classes from Cue-Related Information

(A) Cue information by cell class. Cue information was quantified as spike-rate variance (ω^2) in the late cue period (500–1,000 ms from stimulus onset) across trials explained by cue identity. The four baseline firing statistics (firing rate, Fano factor, CV_{ISI}, burst index) were regressed out. Mean information and SEMs (error bars) were calculated after stratifying cell-class proportions across areas. Cue information significantly differed between cell classes (one-way ANOVA, $p < 0.05$).

(B) Confusion matrix for supervised classification of cell classes from cue information. White dots indicate significant classification performance (binomial test, $p < 0.05$, FDR corrected).

(C) Mean diagonal probabilities (Accuracy) for cross-area classification using cue information. Classifiers trained on data from one area (Train) were used to predict class labels of the other area (Test). White dots indicate significant classification (permutation test, $p < 0.05$, FDR corrected).

(D) Cue information by cell class and brain region. Error bars denote SEM.

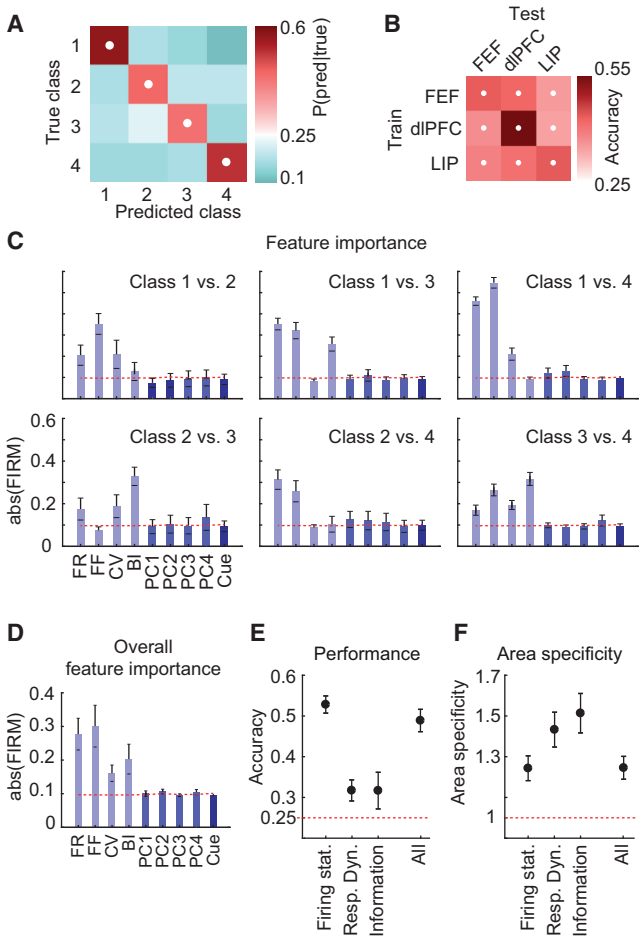

Figure 6. Decoding of Cell Classes from All Combined Functional Measures

(A) Confusion matrix for classification of cell classes using all functional measures as features (4 baseline firing statistics, 4 PCA projections of PSTH, cue information). White dots indicate significance (binomial test, $p < 0.05$, FDR corrected).

(B) Mean diagonal probabilities for cross-area classification. For PSTH features, the PCA transformation was estimated on the training area and applied to the test area. White dots indicate significance (permutation test, $p < 0.05$, FDR corrected).

(C) Feature importance for all features derived from pairwise linear classifiers quantified as the magnitude of FIRM (feature importance ranking measure). Error bars are SD across 50 area-stratified sub-sampled datasets. Red lines show reference FIRM values for "null" classifiers using shuffled class labels (FR, mean firing rate; FF, Fano factor; CV, coefficient of variation of the ISI distribution; BI, burst index; PC1–PC4, PSTH PCA projections; Cue, cue information.

(D) Feature importance for all features, averaged across the six pairwise binary classifiers. Error bars show the SEM across binary classifiers. The red line shows the reference FIRM value for shuffled class labels.

(E) Accuracy across cell classes for all four classifiers. Accuracy is the mean diagonal probability of the confusion matrix. Error bars show the SD across 50 area-stratified datasets. The red dashed line indicates chance-level accuracy (0.25).

(F) Area specificity for all four classifiers computed as the ratio between average within-area and cross-area classification accuracy. The red dashed line indicates the value expected for perfect generalizability across areas. Error bars show the SD across 50 area-stratified datasets.

[df], 3) = 32.3; Mahalanobis distance: $p < 0.001$, one-way ANOVA). Thus, for both measures, we stratified the dataset to equate sorting quality across the four cell classes and repeated the cell-class decoding using all functional measures as features (four baseline firing statistics, PCA projections of PSTH, cue information) (Figure 7A). Both stratifications had hardly any effect on the result. All cell classes remained significantly and similarly decodable from the functional measures (mean accuracy QI stratified, 0.49; mean accuracy Mahalanobis distance stratified, 0.46; compare Figures 7A and 6A). Thus, the reported effects were not driven by a sorting-quality confound.

Second, we ruled out that the results merely reflected different spike waveforms or functional cell properties for the two monkeys rather than distinct cell classes. To this end, we independently repeated the cell-class decoding for each of the two animals using all functional measures (Figure 7B). This revealed very similar independent results for both animals (Figure 7B; mean accuracy monkey P, 0.48; mean accuracy monkey R, 0.44).

DISCUSSION

We employed a large dataset of electrophysiological recordings in awake behaving monkeys to distinguish cortical cell types based on extracellular spike waveform. Across dlPFC, FEF,

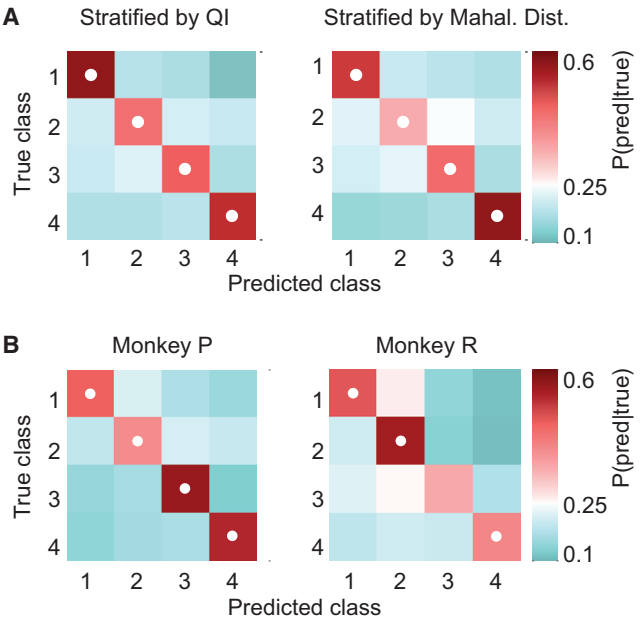

Figure 7. Control Analyses

(A) Confusion matrix for supervised classification of cell classes using all functional measures as features (four baseline firing statistics, PCA projections of PSTH, cue information) after stratification of units to equate sorting quality across cell classes. Left: stratification based on the sorting quality index (QI). Right: stratification based on the Mahalanobis distance of each cell waveform from the unsorted noise waveforms of the same electrode and recording.

(B) Confusion matrix for supervised classification of cell classes using all functional measures for both individual animals.

and LIP, we robustly identified four distinct cell types that showed distinct functional properties in terms of baseline firing statistics, sensory response dynamics, and information coding.

Four Waveform-Based Cell Classes

Our results go beyond previous studies that dissociated only two cell classes (narrow-spiking putative interneurons versus broad-spiking putative pyramidal cells) based on extracellular spike waveform in monkeys [13–15, 18, 20–22, 29, 30]. An important factor for this advance is likely that we employed a two-dimensional feature space for waveform classification. We considered two highly informative waveform measures that have a known physiological relationship to cell-type-specific action-potential dynamics [28]. Most previous work using trough-to-peak duration as a single-waveform feature found a clear bimodal distribution, which justified a two-class scheme. In our data, only the repolarization time showed clearly two distinct modes (see marginal histograms in Figure 1B), whereas trough-to-peak duration likely consisted of even more latent components. Together, these measures allowed for defining four bivariate clusters that were less discriminable when projected only onto one dimension (see also [32, 33]). Future studies may investigate whether additional features, such as, e.g., waveform amplitude or spectral features, can further enhance waveform classification.

It will also be important to assess the effect of the specific band-pass filtering applied to the recorded extracellular voltage traces before spike extraction. The choice of band-pass filtering certainly affects waveform shape. The filtering that we employed (0.5–6 kHz) was similar to that of previous studies [18, 19, 21]. Broader filtering may reveal additional waveform features useful for cell-class separation [16], but it may also enhance waveform noise. Future studies are required to systematically investigate and optimize band-pass-filter choices.

Importantly, owing to the high statistical power of our large dataset, we were able to use unsupervised methods to discover waveform clusters in the data. We performed classification without *a priori* definition of the number of clusters. We also determined class assignments by purely statistical criteria, instead of using prespecified thresholds (i.e., specific values of spike width). This avoided potential confounds due to *a priori* parameter selection. A sub-sampling analysis confirmed that the large size of the dataset was key for this approach.

A cross-classification analysis revealed that waveform clustering was robust across cortical regions. This has two important implications. First, while increasing statistical power, pooling of single units across the FEF, dlPFC, and LIP meant that clustering outcomes could be biased by cortical area. For example, if there were only two true classes that occupied slightly different regions of the 2D feature space depending on the recording area, then the whole sample would spuriously appear to contain multiple latent classes. This was not the case, as we ascertained by rerunning the unsupervised cluster analysis independently on data from the three areas, which reliably revealed four cell classes with comparable statistical structure in each area (Figure 2). Second, this finding supports the notion of cell types as stable physiological entities at the level of cortical microcircuits and columns, yet with specific functional roles across different cortical regions [34]. However, it should be noted that research on area specificity of cell types is still in its infancy [3] and that excitatory cells indeed show distinct transcription profiles across cortical regions [8].

Waveform Width as a Cell-Class Marker

Our results add to a growing body of evidence suggesting action-potential width as a versatile cell-class marker. In monkey dlPFC *in vitro*, a morphologically confirmed "adapting non-pyramidal" cell class shows a distinct intermediate spike waveform, significantly different in width from that of both regular-spiking and fast-spiking cells [11]. Among 12 intracellularly measured physiological parameters, action-potential duration had the largest effect size [35]. The discriminating power of spike width has been systematically tested in an analysis of electrophysiologically defined cell types ("e types") in rat S1 [36]. Here, spike width was ranked as the best-discriminating feature out of 38 electrophysiological measures. Taken together, these and our present results suggest that spike waveform is a sufficiently sensitive and specific marker to dissociate more than two cell classes from extracellular recordings.

Functional Dissociation between Cell Classes

We found significant differences of functional properties between waveform-based cell classes, in terms of firing statistics, response dynamics, and information coding. For the present data, no ground truth on cell-class membership was available. Thus, functional differences provide an important independent validation of the waveform-based cell classes. In accordance

with the distinct functional roles of the FEF, dlPFC, and LIP, cell-class-specific response dynamics and information coding varied substantially across areas [27]. In contrast, baseline firing statistics were consistently cell-class specific across brain regions. This confirms the cell-class specificity of baseline firing statistics reported in previous extracellular [13, 16, 19, 21, 22, 26] and intracellular [11, 28, 36–38] studies. Furthermore, functionally dissociating four waveform-based cell classes critically extends previous studies that dissociated only two cell classes based on extracellular recordings (narrow and broad spiking) [13, 14, 18, 21, 22]. This provides a powerful new window to study cortical circuit function in awake behaving animals.

The present results set the stage for future studies of the functional characteristics of the four identified waveform-based cell classes. On the one hand, this may entail assessing other measures of neuronal activity, such as more sophisticated burst-firing statistics, spectral properties of spiking, and the coupling of spiking to local and remote neuronal activity. On the other hand, it will be interesting to investigate how the four identified cell classes match on other functionally defined categories such as, e.g., visual, motor, and visuomotor neurons [39].

Physiological Correlates of Cell Classes

What are the physiological correlates of the four identified cell classes? With more than two classes, we need to consider subtypes within the excitatory and inhibitory groups. Histological analyses of monkey dlPFC [35], which examined three electrophysiological classes and verified their morphology, showed that broad-spiking RS cells were mostly of the pyramidal type and narrow-spiking FS cells were to a majority GABAergic basket and chandelier cells, as classically described (e.g., [10, 28]). A third intermediate-waveform class consisted exclusively of inhibitory interneurons, with a major proportion of "non-fast-spiking" subtypes (neurogliaform and vertically oriented cells), which is in line with studies in mice using optogenetic labeling of interneuron subtypes [37, 38, 40]. The fast-spiking, narrow-waveform profile is typical of parvalbumin-expressing (PV+) interneurons, which morphologically are basket cells. Non-PV+ interneuron types, such as somatostatin-expressing (SOM+) cells, show higher variance in spike width and firing rate, with some overlap with the FS profile. Thus, cell class 4 in the present data (broad waveform, regular low-rate spiking) likely corresponds to pyramidal cells and cell class 1 (narrowest waveform, high firing rates, low bursting) likely corresponds to PV+ fast-spiking interneurons. Class 1 units in LIP also showed phasic visual-evoked responses (Figure S3), consistent with the short timescale of FS units [38] and stronger stimulus modulation described for FS cells (in V4 [21]; in the FEF [18, 26]). Non-FS interneurons are likely captured in cell class 2, which shows relatively narrow but more dispersed waveform widths than class 1. The "intermediate" firing rate of class 2 is also in agreement with studies showing differences in firing between FS and non-FS neurons in mice [37, 38]. The broad-waveform class 4 fits the classical description of RS pyramidal cells, being numerically most abundant in cortex and having low-rate, regular activity. It is not clear whether class 3 is also part of the excitatory population. A possible clue is given by the relatively strong burstiness specifically of class 3. We can thus speculate that this class comprises intrinsically bursting (IB) neurons, an electrophysiologically defined subtype of pyramidal cells that, despite not exhibiting distinct morphology, has often been distinguished from the RS majority based on its atypical firing mode [9, 16, 26, 28].

The proposed correspondence between the four present classes and physiological cell types is likely to entail some degree of misclassification. For example, some excitatory corticospinal neurons in macaque motor and premotor cortex have FS-like narrow waveforms, with the biggest cells (inferred from axonal conduction velocity) having the thinnest spikes [41]. It is not known whether this finding applies to other frontal or parietal areas and to what extent this may bias classification. Another case of potential ambiguity between excitatory and inhibitory classes is constituted by "chattering cells," a class of narrow-spiking pyramidal neurons first described in superficial layers of cat visual cortex that can fire high-frequency repetitive bursts in response to stimulation [42]. Although there is some evidence of this cell type in the primate ([26, 43]; but see [44]), its presence is hard to verify, especially outside of V1 with potentially sub-optimal stimuli as employed in the present study [42]. Complementary morphological, molecular, or genetic information [3, 8, 45] is needed to unequivocally identify the different physiological cell types underlying the four cell classes established here.

Conclusions

In sum, we show that four functionally distinct neuronal cell classes can be robustly identified from the spike waveform of extracellular recordings across several cortical regions of awake behaving monkeys. These results open a powerful new window to dissect and study the function of cortical micro- and macrocircuits.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Electrophysiological recordings
  - Behavioral task
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Waveform preprocessing
  - Waveform clustering
  - Analysis of firing statistics
  - Multivariate decoding
  - Linear Discriminant Analysis for feature importance estimation
  - Principal component decomposition of PSTH
  - Cue information
  - Sorting quality
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.07.051.

ACKNOWLEDGMENTS

This work was supported by NIMH grant R37MH087027 (E.K.M.), European Research Council (ERC) StG335880 (M.S.), Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) project 276693517 (SFB 1233) (M.S.) and grant SI1332-3/1 (M.S.), and the Centre for Integrative Neuroscience (DFG, EXC 307) (M.S.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.S., C.T., and E.K.M.; Methodology, C.T., M.S., and C.v.N.; Investigation, M.S.; Formal Analysis, C.T., M.S., and C.v.N.; Writing – Original Draft, C.T. and M.S.; Writing – Review & Editing, C.v.N.; Funding Acquisition, M.S. and E.K.M.; Resources, M.S. and E.K.M.; Supervision, M.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 14, 2019
Revised: June 21, 2019
Accepted: July 17, 2019
Published: August 22, 2019

REFERENCES

1. Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic interneurons in the neocortex: from cellular properties to circuits. Neuron *91*, 260–292.
2. Lee, J.H., Koch, C., and Mihalas, S. (2017). A computational analysis of the function of three inhibitory cell types in contextual visual processing. Front. Comput. Neurosci. *11*, 28.
3. Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat. Rev. Neurosci. *18*, 530–546.
4. Womelsdorf, T., Valiante, T.A., Sahin, N.T., Miller, K.J., and Tiesinga, P. (2014). Dynamic circuit motifs underlying rhythmic gain control, gating and integration. Nat. Neurosci. *17*, 1031–1039.
5. Veit, J., Hakim, R., Jadi, M.P., Sejnowski, T.J., and Adesnik, H. (2017). Cortical gamma band synchronization through somatostatin interneurons. Nat. Neurosci. *20*, 951–959.
6. Siegel, M., Donner, T.H., and Engel, A.K. (2012). Spectral fingerprints of large-scale neuronal interactions. Nat. Rev. Neurosci. *13*, 121–134.
7. Petilla Interneuron Nomenclature Group, Ascoli, G.A., Alonso-Nanclares, L., Anderson, S.A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsáki, G., Cauli, B., Defelipe, J., Fairén, A., et al. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. Nat. Rev. Neurosci. *9*, 557–568.
8. Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature *563*, 72–78.
9. Connors, B.W., and Gutnick, M.J. (1990). Intrinsic firing patterns of diverse neocortical neurons. Trends Neurosci. *13*, 99–104.
10. McCormick, D.A., Connors, B.W., Lighthall, J.W., and Prince, D.A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. J. Neurophysiol. *54*, 782–806.
11. González-Burgos, G., Krimer, L.S., Urban, N.N., Barrionuevo, G., and Lewis, D.A. (2004). Synaptic efficacy during repetitive activation of excitatory inputs in primate dorsolateral prefrontal cortex. Cereb. Cortex *14*, 530–542.
12. Henze, D.A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K.D., and Buzsáki, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. J. Neurophysiol. *84*, 390–400.
13. Constantinidis, C., and Goldman-Rakic, P.S. (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. J. Neurophysiol. *88*, 3487–3497.
14. Diester, I., and Nieder, A. (2008). Complementary contributions of prefrontal neuron classes in abstract numerical categorization. J. Neurosci. *28*, 7737–7747.
15. Johnston, K., DeSouza, J.F.X., and Everling, S. (2009). Monkey prefrontal cortical pyramidal and putative interneurons exhibit differential patterns of activity between prosaccade and antisaccade tasks. J. Neurosci. *29*, 5516–5524.
16. Ardid, S., Vinck, M., Kaping, D., Marquez, S., Everling, S., and Womelsdorf, T. (2015). Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. J. Neurosci. *35*, 2975–2991.
17. Cohen, J.Y., Pouget, P., Heitz, R.P., Woodman, G.F., and Schall, J.D. (2009). Biophysical support for functionally distinct cell types in the frontal eye field. J. Neurophysiol. *101*, 912–916.
18. Thiele, A., Brandt, C., Dasilva, M., Gotthardt, S., Chicharro, D., Panzeri, S., and Distler, C. (2016). Attention induced gain stabilization in broad and narrow-spiking cells in the frontal eye-field of macaque monkeys. J. Neurosci. *36*, 7601–7612.
19. Tamura, H., Kaneko, H., Kawasaki, K., and Fujita, I. (2004). Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons. J. Neurophysiol. *91*, 2782–2796.
20. Mruczek, R.E.B., and Sheinberg, D.L. (2012). Stimulus selectivity and response latency in putative inhibitory and excitatory neurons of the primate inferior temporal cortex. J. Neurophysiol. *108*, 2725–2736.
21. Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron *55*, 131–141.
22. Vinck, M., Womelsdorf, T., Buffalo, E.A., Desimone, R., and Fries, P. (2013). Attentional modulation of cell-class-specific gamma-band synchronization in awake monkey area v4. Neuron *80*, 1077–1089.
23. Hendry, S.H., Schwark, H.D., Jones, E.G., and Yan, J. (1987). Numbers and proportions of GABA-immunoreactive neurons in different areas of monkey cerebral cortex. J. Neurosci. *7*, 1503–1519.
24. Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Dev. Neurobiol. *71*, 45–61.
25. Gabbott, P.L.A., and Bacon, S.J. (1996). Local circuit neurons in the medial prefrontal cortex (areas 24a,b,c, 25 and 32) in the monkey: II. Quantitative areal and laminar distributions. J. Comp. Neurol. *364*, 609–636.
26. Katai, S., Kato, K., Unno, S., Kang, Y., Saruwatari, M., Ishikawa, N., Inoue, M., and 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. Eur. J. Neurosci. *31*, 1322–1338.
27. Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355.
28. Nowak, L.G., Azouz, R., Sanchez-Vives, M.V., Gray, C.M., and McCormick, D.A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. J. Neurophysiol. *89*, 1541–1566.
29. Oemisch, M., Westendorff, S., Everling, S., and Womelsdorf, T. (2015). Interareal spike-train correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts. J. Neurosci. *35*, 13076–13089.
30. Hussar, C.R., and Pasternak, T. (2009). Flexibility of sensory representations in prefrontal cortex depends on cell type. Neuron *64*, 730–743.
31. Shinomoto, S., Shima, K., and Tanji, J. (2003). Differences in spiking patterns among cortical neurons. Neural Comput. *15*, 2823–2842.
32. Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K.D., and Buzsáki, G. (2004). Characterization of neocortical principal cells and

interneurons by network interactions and extracellular features. J. Neurophysiol. *92*, 600–608.

33. Insel, N., and Barnes, C.A. (2015). Differential activation of fast-spiking and regular-firing neuron populations during movement and reward in the dorsal medial frontal cortex. Cereb. Cortex *25*, 2631–2647.
34. Harris, K.D., and Shepherd, G.M.G. (2015). The neocortical circuit: themes and variations. Nat. Neurosci. *18*, 170–181.
35. Krimer, L.S., Zaitsev, A.V., Czanner, G., Kröner, S., González-Burgos, G., Povysheva, N.V., Iyengar, S., Barrionuevo, G., and Lewis, D.A. (2005). Cluster analysis-based physiological classification and morphological properties of inhibitory neurons in layers 2-3 of monkey dorsolateral prefrontal cortex. J. Neurophysiol. *94*, 3009–3022.
36. Druckmann, S., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2013). A hierarchical structure of cortical interneuron electrical diversity revealed by automated statistical analysis. Cereb. Cortex *23*, 2994–3006.
37. Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C.C.H. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. J. Neurophysiol. *107*, 3116–3134.
38. Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J.Z., and Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature *498*, 363–366.
39. Bruce, C.J., and Goldberg, M.E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. J. Neurophysiol. *53*, 603–635.
40. Muñoz, W., Tremblay, R., and Rudy, B. (2014). Channelrhodopsin-assisted patching: *in vivo* recording of genetically and morphologically identified neurons throughout the brain. Cell Rep. *9*, 2304–2316.
41. Vigneswaran, G., Kraskov, A., and Lemon, R.N. (2011). Large identified pyramidal cells in macaque motor and premotor cortex exhibit "thin spikes": implications for cell type classification. J. Neurosci. *31*, 14235–14242.
42. Gray, C.M., and McCormick, D.A. (1996). Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. Science *274*, 109–113.
43. Friedman-Hill, S., Maldonado, P.E., and Gray, C.M. (2000). Dynamics of striate cortical activity in the alert macaque: I. Incidence and stimulus-dependence of gamma-band neuronal oscillations. Cereb. Cortex *10*, 1105–1116.
44. Chen, D., and Fetz, E.E. (2005). Characteristic membrane potential trajectories in primate sensorimotor cortex neurons recorded in vivo. J. Neurophysiol. *94*, 2713–2725.
45. Buzsáki, G., Stark, E., Berényi, A., Khodagholy, D., Kipke, D.R., Yoon, E., and Wise, K.D. (2015). Tools for probing local circuits: high-density silicon probes combined with optogenetics. Neuron *86*, 92–105.
46. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.
47. Schwarz, G. (1978). Estimating the dimension of a model. Ann. Stat. *6*, 461–464.
48. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. *57*, 289–300.
49. Zien, A., Krämer, N., Sonnenburg, S., and Rätsch, G. (2009). The feature importance ranking measure. In Machine Learning and Knowledge Discovery in Databases, W. Buntine, M. Grobelnik, D. Mladenić, and J. Shawe-Taylor, eds. (Springer), pp. 694–709.
50. Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage *87*, 96–110.
51. Bro, R., Kjeldahl, K., Smilde, A.K., and Kiers, H.A.L. (2008). Cross-validation of component models: a critical look at current methods. Anal. Bioanal. Chem. *390*, 1241–1251.
52. Olejnik, S., and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychol. Methods *8*, 434–447.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Rhesus Macaque (Macaca Mulatta)	Covance Research Products	N/A
Software and Algorithms		
MATLAB	The Mathworks	RRID: SCR_001622
Fieldtrip	[46]	RRID: SCR_004849
Offline Sorter	Plexon	RRID: SCR_000012

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Markus Siegel (markus.siegel@uni-tuebingen.de). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Two, adult monkeys (*Macaca mulatta*) were used in this study: one male (11 years old), and one female (10 years old) weighing about 9 kg and 5 kg, respectively. They were both experimentally naive, pair-housed, on a 12-hr day/night cycle, and in a temperature-controlled environment (80°F). Experiments were performed in a dedicated laboratory around the middle of their light cycle. Each monkey was surgically implanted with a titanium post for head restraint and three cylindrical 20-mm diameter titanium recording chambers. The sterile surgery was performed under general anesthesia. Post-surgical pain was controlled with an opiate analgesic. Chambers were stereotaxically placed based on coordinates from structural MRI scans in each monkey.

During training and experimental testing, the animals were allowed to obtain water through the behavioral task to the point of satiety each day. Satiety was indicated when the monkey will no longer perform the behavioral task. Animals that failed to obtain their normal amount of water on any given day were supplemented. Good health was ensured by keeping daily records of weight, carefully monitoring the monkey's physical state, supplementing the diet with fresh and dried fruit to ensure adequate nutrition, and providing regular intervals of free access to water.

The animals were handled in accord with National Institutes of Health guidelines and approved by the Massachusetts Institute of Technology Committee on Animal Care. MIT veterinary staff continuously assessed the welfare of the animals prior to, during, and after the experiment. No adverse events occurred, and no procedural modifications were necessary.

METHOD DETAILS

Electrophysiological recordings

We briefly review the electrophysiological recording methods here. Further details on the electrophysiological recordings can be found in [27]. Extracellular signals were recorded in 70 recording sessions in two rhesus monkeys using Tungsten microelectrodes simultaneously inserted in FEF, dorso-lateral PFC, and LIP. Electrodes were lowered in pairs (1 mm spacing) or triplets (0.7 mm triangular spacing) using custom microdrive assemblies. Electrodes were inserted without targeting of a specific cortical depth, were acutely inserted into the brain and removed at the end of each daily experiment. Broad-band extracellular signals were recorded at a sampling rate of 40 kHz and then bandpass-filtered between 0.5–6 kHz to extract spiking activity. The dataset partially overlaps with the multiunit data analyzed in [27].

Behavioral task

During the recordings, monkeys performed a flexible visuomotor decision-making task. Each trial started with a 'baseline' period lasting 0.5 s during which the monkey maintained central fixation. This was followed by a 1 s 'cue' period in which a visual cue stimulus was shown to indicate the condition of the upcoming task. Cue stimuli were four different shapes, two of which cued a motion discrimination task and two a color discrimination task. Depending on the cue, the task consisted in judging either the motion direction (up versus down) or color (green versus red) of a random dot stimulus presented after the cue. The monkeys responded with a leftward or rightward saccade within 3 s after stimulus onset.

QUANTIFICATION AND STATISTICAL ANALYSIS

Waveform preprocessing

To obtain spike waveforms, we extracted segments of the filtered voltage traces in a window of 3 ms around each noise threshold-crossing (4 SD; 1 ms before crossing) aligned on the main trough of the waveform. The noise level (SD) was robustly estimated as 0.6745 times the median of the absolute of the filtered data. The minimum duration between 2 threshold-crossings that triggered a waveform was 1 ms. Spike waveforms were trough aligned after spline-based up-sampling. Spike sorting was performed manually offline using Plexon Offline Sorter. Single-unit isolation was assessed by an expert user (CvN) and judged according to a quality (QI) index with 4 scales (1: very-well isolated single unit, 2: well isolated single unit, 3: potential multi-unit; 4: clear multi-unit). Only units with quality index 1 and 2 were included in the analysis.

We used principal components (PC) 1 and 2 of the spike waveform as well as the nonlinear energy function of the spike as axes in 3D sorting space. The manual 3D clustering was separately and dynamically performed in a sliding window throughout each recording session and for each electrode. Thus, clusters could move across time. We carefully looked for elongated waveform-clusters that often reflect bursting and paid attention not to artificially split these into multiple units. A putative single unit had to exhibit clear separability of its cluster in this 3D feature space, was only defined as long as it could be separated from other waveforms or clusters across time and had to show a clean stack of individual waveforms in its overlay plot.

We analyzed the average spike waveform of each well-isolated single unit. Waveforms were up-sampled and normalized on their amplitude. To exclude axonal spikes and temporally overlapping spike as well as to ensure that the two employed waveform-classification metrics (trough-to-peak and time-to-repolarization; see below) were valid and robust, we excluded waveforms that satisfied any of three criteria for atypical shape: (1) the amplitude of the main trough was smaller than the subsequent positive peak (n = 41), (2) the trace was noisy, defined as > = 6 local maxima of magnitude > = 0.01 (n = 38), (3) there was one or more local maxima in the period between the main trough and the subsequent peak (n = 35).

To assess the effect of sorting quality, in addition to the above subjective sorting quality index QI, we quantified for each well-isolated single unit the Mahalanobis distance of its average waveform to the cluster of all unsorted noise-waveforms of the same electrode and recording. The Mahalanobis distance was computed in 2 dimensions based on the same first 2 PCs of the spike waveforms that were employed for spike sorting.

Waveform clustering

As features for cell class classification, we computed two measures of waveform shape: trough-to-peak duration and time for repolarization. Trough-to-peak duration is the distance between the global minimum of the curve and the following local maximum. Time for repolarization is the distance between the late positive peak and the inflection point of the falling branch of the curve [12, 28].

All preprocessed waveforms (n = 2488) were scored on the two measures to obtain a two-dimensional feature space for classification. To identify clusters in the data in an unsupervised way, we used the expectation-maximization (EM) algorithm for Gaussian mixture model (GMM) clustering. We modeled the data as a weighted sum of multivariate Gaussians:

$$P(x) = \sum_k \pi_k \, N(x \mid \mu_k, \Sigma_k)$$

with k components parametrized by mean μ, covariance Σ and mixing coefficient π. The EM algorithm fits this model by iteration of a two-step process: it first estimates posterior probabilities of the data given the current set of parameters (E step), and then updates the parameters to maximize the log-likelihood function of the model given the current estimates (M step). The steps are repeated until convergence. We initialized the process with random parameters for 50 repetitions and chose the fit with the largest log-likelihood among the replicates.

To select the number of Gaussian components in the model we used the Bayesian information criterion (BIC) [47]:

$$BIC = -2 \ln P(x \mid \theta) + K \ln(n)$$

where $P(x \mid \theta)$is the maximized likelihood for the estimated model, K is the number of parameters, and n is the sample size. By including a penalty term that grows with the number of parameters, the BIC cost function effectively favors simpler models and reduces overfitting. The optimal number of clusters was chosen as the value that minimized the BIC computed between 2 and 10 components.

After fitting the model, we determined cluster memberships by 'hard' assignment: each unit was assigned to the class associated with the highest posterior probability.

Given the initial clustering outcome, we excluded one high-variance cluster ('noise cluster') that captured the waveforms dispersed around the high-density axis of the data cloud (n = 212). We also excluded units that were outliers of the whole data cloud, defined as having Mahalanobis distance large than 5 from the centroid of the Gaussian cluster they were assigned to (n = 69). After outlier rejection, we re-ran the clustering (including the BIC analysis for choosing number of components) to obtain the final cell class classification.

To assess the degree of cluster separation, we calculated the overlap between GMM components using a Monte Carlo approach. We randomly generated 10000 data points from the fitted GM distribution and compared, for each data point, the true cluster from which the observation was drawn with the class to which it was assigned. The outcome of this comparison can be represented by a

confusion matrix, where the off-diagonal terms provide empirical estimates of the area of overlap between the GMM components. Overall class separation was quantified as the mean of the diagonal probabilities of the confusion matrix.

This method was extended to assess the similarity of the clustering schemes obtained for individual cortical areas. For each pair of areas A and B, 10000 data points were randomly drawn from the GM distribution of area A and assigned to classes defined on the GM distribution of area B. The confusion matrix shows, for area A's data, the true generating cluster against the assigned class label. For all area pairs, overall class similarity was quantified as the mean of the diagonal probabilities of the corresponding confusion matrix.

To assess how to the number of identified clusters depends on the number of neurons, we repeated the clustering analysis after randomly sub-sampling the original data. For each sub-sample size, we repeated the analysis 100 times. For each sub-sample, we performed the same 2-step approach as for the complete data. Given the initial clustering outcome, we excluded a 'noise-cluster' with the highest ratio of 2D-variance over proportion of assigned cells if there were more than 2 initial clusters and if not most cells were assigned to the putative noise cluster. We also excluded outliers with a Mahalanobis distance large than 5 from their corresponding cluster-centroid. Then, we re-ran the clustering and performed the BIC analysis for choosing the final number of clusters.

Analysis of firing statistics

To characterize spontaneous activity, we analyzed spiking activity during the baseline fixation period. We averaged across baseline periods of all trials. We computed four firing statistics: mean firing rate across trials (FR), Fano factor (variance over mean of spike counts across trials, FF), coefficient of variation of the inter-spike interval distribution (CV_{ISI}) and burst index (BI). Both Fano factor and CV_{ISI} are mean-standardized measures of dispersion that reflect firing regularity, with an expected value of 1 for Poisson firing and values below 1 indicating more regular firing [31]. The burst index was defined as the ratio between the observed proportion of bursts, defined as inter-spike intervals < 5 ms, and the proportion of bursts expected from a Poisson process with equal mean rate [13]. This measure quantifies the tendency to fire in bursts unbiased by firing rate. To avoid under-sampling, we only computed the burst index for units with more than 50 inter-spike intervals pooled across all trials ($n = 1388$ units), which excluded neurons with very low firing rates.

We tested for significant differences in activity between cell classes using a one-way ANOVA on each firing statistic. All measures were log-transformed to optimize normality. To control for the unequal regional distribution of cell classes, we randomly subsampled the data to have equal cell class proportions in the three areas ('area-stratified datasets'). The ANOVA F-statistic was computed as the ratio of mean square between and mean squared error. Both numerator and denominator were calculated on 1000 area-stratified subsamples and then averaged across subsamples, so that the F-ratio was obtained from the two averaged quantities. For post hoc comparisons, we computed pairwise t tests using the average difference of means and the average pooled standard error across 1000 area-stratified subsamples. We corrected for multiple tests using False Discovery Rate (FDR) [48] correction.

Multivariate decoding

We performed cell class decoding using Support Vector Machine (SVM) classification on several different feature sets. For all feature sets, decoding was performed on 50 area-stratified subsamples (where cell class proportions were matched across cortical regions by random subsampling) and averaged across subsamples to obtain the final classification estimate.

Classification procedure

To reduce the multiclass problem to binary classification, we independently trained and tested six binary SVMs for each pair of cell classes. The six sets of predicted labels were combined by majority vote ('one-versus-one' classification); in case of ties, one of the two winning classes was chosen at random. The SVM algorithm employed a Gaussian radial basis function kernel with a scaling factor of 1.

Each binary classifier was evaluated using 10-fold cross-validation. Within each classifier, we randomly subsampled the data such that both cell classes had N equal to the minimum sample size across all cell classes. Equal cell class proportions were preserved in each cross-validation fold's training and test set, so that chance-level classification performance for a given test set was always 0.5. By using the global minimum of sample sizes across cell classes, we also ensured that all pairwise classifiers had comparable signal-to-noise ratio. This stratification procedure was repeated 100 times and the estimates were combined by majority vote.

Classification outcome was summarized in a confusion matrix of probabilities based on the average counts over 50 area-stratified datasets. Counts were divided by true class total counts to obtain the probability of predicting each class given the true label. Each class was considered to be decodable if its true positive rate was significantly greater than chance level of 0.25 in a binomial test, corrected for four tests using FDR correction. As a summary measure of the confusion matrix, we quantified classifier accuracy as the average true positive rate across cell classes (mean of the diagonal of the confusion matrix).

Cross-area classification

To assess area specificity of cell class decoding, we trained classifiers on data from one cortical area and used them to predict data from other areas. We matched areas' signal-to-noise ratio by creating 50 randomly subsampled datasets for which all areas had the same number of observations, equal to the minimum N across the three areas. A separate classification instance was run for each subsample, and the resulting 50 confusion matrices of counts were averaged.

To test for significance of cross-area classification performance, we used a permutation test that compared the observed accuracy with an empirical null distribution. The null distribution was constructed by training classifiers with randomly permuted class labels.

We ran 1000 instances of null classifiers, each resulting from the average across 20 area-stratified subsamples. The observed accuracy was declared significant if it exceeded the accuracy values of 1000 label-shuffled classifiers at $p < 0.05$, FDR-corrected for 9 tests (3*3 cross-area classifiers).

Area specificity was computed from the 3-by-3 matrix of cross-area classification accuracy as the ratio between the average of the diagonal and the average of the non-diagonal values.

Linear Discriminant Analysis for feature importance estimation

For the dataset comprising all features (four baseline firing statistics, PSTH, cue information) we also trained a Linear Discriminant Analysis (LDA) classifier using the same procedure described above for the SVMs. To quantify each feature's contribution to the decoding, we computed FIRM (Feature Importance Ranking Measure) using the simplified formulation for unregularized OLS regression [49, 50]:

$$FIRM = D^{-1}\, cov(X, y)$$

where D is a diagonal matrix of standard deviations of the features, X is the training data matrix, and y is the vector of true class labels of the training data (recoded as +1 and −1). As discussed in [50], FIRM for linear decoding can be approximated by the covariance between the data and the labels, effectively reducing to a univariate analysis on each feature. The normalization by D ensures that FIRM is invariant to feature scaling. As our measure for feature importance we considered the magnitude (absolute value) of FIRM.

Principal component decomposition of PSTH

Peristimulus time histograms (PSTH) of single-unit spike counts were computed using 50 ms bins, within a 1.5 s trial window comprising the 0.5 s baseline fixation period and the 1 s cue period. Each unit's PSTH was z-scored on the mean and standard deviation of the baseline period across trials.

To test if the response dynamics captured by the PSTHs predicted, i.e., differed between, cell classes, we extracted low-dimensional features from the PSTHs. This primarily served to enhance statistical power, rather than to characterize the response dynamics themselves. These response dynamics are directly shown as the averages of PSTHs for each cell class (Figure 4B) and region (Figure 4E). We extracted low-dimensional PSTH features using Principal Component Analysis (PCA). The number of significant principal components was determined using cross-validation [51]. For each number of PCs k, we fit a PCA to one portion of the data ('training' set) and, using the first k components, we reconstructed the left-out data ('test' set). Importantly, to obtain reconstructions that were truly independent from the data, we excluded in turn each variable from the fitting on the training set, and that variable only was predicted for the test set observations. Using 10-fold cross-validation, a predicted value was calculated for each data point. The overall prediction error was calculated as the sum of squared differences between true and reconstructed data points (PRESS, Prediction Residual Error Square Sum) as a function of number of components:

$$PRESS(k) = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_j^{(i)} - \left(P^{(-i)}\left(P_{-j}^{(-i)}\right)^T x_{-j}^{(i)}\right)_j\right)^2$$

with k number of PCs, data samples x_{ij} having number of observations I and number of variables J, and P being the $J \times J$ matrix of PCA coefficients, with columns corresponding to principal components. The number of significant components was chosen as the global minimum of the PRESS curve.

PCA feature extraction, including selection of number of PCs, was performed independently on each of the 50 area-stratified datasets for cell class classification. For cross-area classification, the PCA transformation was estimated on the training area only and then applied to data of the test area.

Cue information

Cue information was quantified as the effect size of the cue-factor in a 7-way ANOVA computed on the late cue period of the trial (500–1000 ms after cue onset) [27]. Effect size was quantified by the ω^2 statistic, an unbiased estimator of the variance explained by an ANOVA factor independently from all others factors in the model [52]:

$$\omega^2 = \frac{SS_{between} - df * MSE}{SS_{total} + MSE}$$

where $SS_{between}$ is the between-groups sum of squares, df is the degrees of freedom, SS_{total} is the total sum of squares, and MSE is the mean squared error. The first three factors of the ANOVA corresponded to the cue of each trial grouped into two levels according to all three possible pairwise pairings of the four task cues. We computed cue information as the average explained variance of all first three factors [27]. Thus, the cue information assessed here, was only about the four cue identities and did not allow for inferences on information about the motion or color task. The remaining 4 factors of the ANOVA were the motion direction of the random-dot stimulus, the color of the random-dot stimulus, the response of the animal on the current trial, and the response on the previous trial.

To control for linear dependencies between cue information and activity measures, we took the residuals of cue information after regressing out the four baseline firing statistics (firing rate, Fano factor, coefficient of variation of the ISI distribution, burst index).

Cell class differences in cue information were assessed with a one-way ANOVA using the same procedure employed for the firing statistics: the data was subsampled 1000 times to match cell class proportions across cortical areas, and the F-ratio was calculated from the averages of the numerator and denominator across subsamples. Post hoc pairwise comparisons between classes were similarly evaluated taking the average numerator and denominator of the t-statistic.

Sorting quality

To rule out a potential confound due to a difference in sorting quality between cell classes, we repeated the cell-class decoding after equating sorting quality across cell classes. To this end, we independently stratified the data to equate either the sorting quality index QI or the Mahalanobis-distance distribution between cell classes. For both measures, we repeated the stratification 20 times and averaged the results. We stratified QI by randomly removing units from each class such that the proportion of cells with QI = 1 and QI = 2 cells was the same for all cell classes. QI stratification excluded 306 units (13.7% of sample). We stratified the Mahalanobis-distance by randomly removing units from each class such that the Mahalanobis-distance distributions were matched between all classes. Mahalanobis-distance stratification excluded on average 1577 units (70.6% of sample).

DATA AND CODE AVAILABILITY

The data and code supporting the current study have not been deposited in a public repository because of ongoing projects but are available on request by contacting Earl K. Miller (ekmiller@mit.edu).