ACKNOWLEDGEMENTS

We thank Ali Williford and Jennifer Luviano for help with surgical procedures, Gregg Heller for help running experiments, Shiella Caldejon for performing intrinsic signal imaging, Timothy Cox for help with tissue clearing and OPT imaging, Will Allen for advice on brain clearing techniques, Marius Pachitariu for help with the spike sorting algorithm. We wish to thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement, and support.

METHODS

Data acquisition and preprocessing:

Animal preparation

All experimental procedures were approved by the Allen Institute for Brain Science Institutional Animal Care and Use Committee. For recordings in visual cortex and hippocampus, a metal headframe with a 10 mm circular opening was attached to the skull with Metabond. In the same procedure, a 5 mm diameter craniotomy was drilled over left visual cortex, and sealed with a circular piece of PDMS silicone, ~0.3 mm thick (Heo et al., 2016). Following a 2-week recovery period, a visual area map was obtained through intrinsic signal imaging (Juavinett et al., 2016). On the day of the experiment, the mouse was placed under light isoflurane anesthesia for approximately 40 minutes to remove the silicone window. A ground wire was secured to the skull, and the exposed brain was covered with a layer of 4% agar in ACSF. Following recovery from anesthesia, the mouse was head-fixed on the experimental rig. Three or more Neuropixels probes coated in CM-DiI were independently lowered vertically into visual cortex at a rate of 100 μm/min using a piezo-driven microstage (New Scale Technologies). When the probes reached their final depths of 1200-1500 μm, the tip of each probe extended through visual cortex into hippocampus.

For cerebellar recordings, skin and muscle were resected from above the posterior skull to expose the skull above the cerebellum. Animals were fitted with an aluminum headplate with a 5 mm circular opening above the exposed skull. On the day of recording, the animal was anesthetized and burr holes were made in several locations above the cerebellar cortex. The animal was then head-fixed in the recording apparatus and allowed to recover from anesthesia. For each insertion (n=3 in one mouse), a Neuropixels probe coated in DiI was lowered through a burr hole to a final depth of 3.4 - 3.6 mm from the pia at a fixed rate of 100 μm/min. The probe was fixed at a roughly 15º angle relative to the dorsal plane of the skull, resulting in an 6º to 19º angle relative to the cerebellum surface for each insertion. Recordings extended through multiple ganglionic layers and into the reticular nuclei (Supplementary Figure 1). The probe was allowed to rest in place for at least 15 minutes following insertion before data were recorded.

For LP, LGN, and SC recordings, a metal headframe was attached to the skull with Metabond. One week after surgery, mice were handled (3-5 days) and habituated to head-fixation (~2 weeks). On the day of recording, the animal was anesthetized with isoflurane and a small burr hole (~200 μm diameter) was drilled according to stereotactic coordinates (in mm relative to lambda, LP/LGN: 1.5-2.2 anterior, 1.9-1.5 lateral; SC: 0.25 anterior, 0.5 lateral). Mice were given two hours to recover before being head-fixed in the recording apparatus. A Neuropixels probe was coated in DiI and lowered through the burr hole at a rate of 200 μm/min to a final depth of 3-3.3 (for LP and LGN) or 1.3-1.6 (SC) mm from the brain surface. The probe was allowed to settle for 30 minutes before recording began. For most mice, recordings were made on two consecutive days.

Data acquisition system

In vivo recordings were performed in awake, head-fixed mice allowed to run freely on a rotating disk. During the recordings the mice either passively viewed visual stimuli or remained in the dark. For recordings in visual cortex (VIS) and hippocampus (HP), data were collected from 11 mice (25 probe

insertions). On average, we recorded 64±6 units in cortex per probe insertion. For recordings in LP, LGN and SC, data were collected from 31 mice (n= 9 in SC, 4 in LGN, and 18 in LP). For recordings in cerebellum (Ce), data were collected from 1 mouse with 3 different penetrations. All spike data were acquired with Neuropixels probes (Jun et al., 2017) with 30 kHz sampling rate, and recorded with the Open Ephys GUI (Siegle et al., 2017). A 300 Hz highpass filter was present in the Neuropixels probe, and another 300 Hz highpass filter (3rd order Butterworth) was applied offline prior to spike sorting.

Histology

For recordings in visual cortex and hippocampus, the probe location was confirmed by clearing brains with dichloromethane and dibenzylether (https://idisco.info/idisco-protocol/) and imaging with optical projection tomography (OPT; Supplementary Figure 1A). OPT showed most recordings from hippocampus are from CA1 region given our probe insertion location and depth. To assign probes to specific visual areas, we overlaid an image of the brain surface obtained during the recording session on images obtained from intrinsic signal imaging, using the vasculature for registration (Supplementary Figure 1B). For recordings in other brain areas, recording location was subsequently verified by identifying the DiI fluorescence in sectioned brain tissue (Supplementary Figure 1C-F).

Data pre-processing

In all experiments, spike times and waveforms were automatically extracted from the raw data using KiloSort (Pachitariu et al., 2016). KiloSort is a spike sorting algorithm developed for electrophysiological data recorded by hundreds of channels simultaneously. It implements an integrated template matching framework for detecting and clustering spikes, rather than clustering based on spike features, which is commonly used by other spike sorting techniques. The outputs of Kilosort were loaded into Phy (Rossant et al., 2016b) for manual refinement, which consisted of merging and splitting clusters, as well as marking non-neural clusters as "noise." Noise units were identified by their abnormal waveform shape, as well as distinct cyclical patterns in the autocorrelogram. Merges were made when the cross-correlogram between two units dropped near zero at a 1-2 ms temporal offset, indicating the presence of a refractory period. Splits were made when the spikes for one unit displayed multiple readily separable clusters when projected in PCA space. In this case, a boundary was manually drawn around one of the clusters, thereby splitting the unit in two. If the resulting cross-correlogram showed no evidence of a refractory period, the split was maintained; otherwise, the split was reverted. We then used a set of heuristic rules based on the features of waveforms remove abnormal waveforms (the parameters used for this purpose were PTratio<0.99 and recovery slope<0). No further constraints were imposed.

Waveforms for each unit were extracted from the raw data by slicing around the trough time (pre-trough points = 20 samples, total waveform size = 82 samples, with 30 kHz sampling rate). For each unit, the mean waveform was calculated from bootstrapped waveforms (number_of_spikes=100; number_of_repititions=100) from all spikes. If the number of spikes for a given unit was smaller than 100, then all the waveforms were used to calculate the mean waveform. Mean waveforms for experiments with optotagging were calculated only on waveforms prior to the light stimulation period, to avoid artifacts in waveform traces caused by light.

Optotagging:

Optotagging was performed in a subset of the visual cortex experiments described above, using Pvalb-Cre x Ai32 (ChR2 reporter) mice. In each experiment, a 200 micron diameter bare fiber optic cable (Thorlabs) connected to a 465 nm LED (Plexon) was aligned with the center of the cranial window, such that it illuminated a surface area of approximately 20 mm^2. Stimulus trains were delivered with a Cyclops LED Driver (Newman et al., 2015) and consisted of either 2.5 ms square-wave pulses at 10 Hz, individual square-wave pulses lasting 5 ms or 10 ms, or 1 s raised cosine ramps. Peak light power varied from 0.1 mW to 10 mW on a given trial. Each stimulus condition (pulse type x power) was repeated 120 times. Light artifacts were visible on all channels, but were readily separable from actual spikes based on timing relative to the stimulus and waveform shape.

Analysis:

Waveform feature extraction

To classify cell types, we first extracted features from the extracellular waveform. With high density Neuropixels probes, we can record extracellular waveforms of a single unit from multiple sites. We define the recording site with highest amplitude (absolute difference between trough and peak of an extracellular waveform) as the site closest to neuron soma and the extracellular waveform recorded here is our single channel (1-ch) waveform. To take advantage of signals detected by multiple sites, we consider the profile of extracellular waveforms of a single sorted unit recorded from multiple adjacent recording sites as a multi-channel waveform. The probe is inserted along the dorsal-ventral axis of the brain. Since the Neuropixels probe has two recording sites at each depth, we used only the side of the probe with the highest amplitude 1-channel waveforms to calculate the multi-channel waveform. The distance between sites is approximated by their vertical spacing (20 μm). The horizontal spacing is ignored here, but it could potentially contribute to differences between adjacent sites.

For 1-ch waveform, we extracted five features: amplitude, duration, peak-to-trough ratio, repolarization slope, and recovery slope (Figure 2B). Waveform peak is defined by the maximum point of extracellular waveform. Trough is defined by the minimum point. Amplitude is defined by the absolute difference between peak and trough. Duration is defined by the time between waveform trough and peak. This feature is commonly used to separate fast spiking (FS) neurons from regular spiking (RS) neurons (McCormick et al., 1985; Mitchell et al., 2007; Niell and Stryker, 2008; Swadlow, 2003). Peak-to-trough ratio is determined by the absolute amplitude of peak divided by absolute amplitude of trough relative to 0. The repolarization slope is defined by fitting a regression line to the first 30 μs from trough to peak. The recovery slope is defined by fitting a regression line to the first 30 μs from peak to tail.

For multi-channel waveform (Figure 3A), we extracted three additional features in the space domain for classification: spread along probe, inverse of propagation velocity above (1/v_above) and below soma (1/v_below) along the probe. The spread of a unit is defined by the distribution of its waveform amplitude. If we plot amplitude against recording site distance relative to soma, we get a curve with peak at 0 (Figure 3C). We define the range with amplitude above 12% of the maximum amplitude as the spread of a unit along the probe. The multi-channel waveform has information in both time and space dimension for signal propagation velocity estimation. Because the time difference between the trough of adjacent sites could be 0, to avoid infinite numbers, we calculated the inverse of velocity (ms/mm) instead by fitting a regression line to the time of waveform trough at different sites against the distance of the sites relative to soma (Buzsáki and Kandel, 1998).

All waveforms and features are available at https://github.com/jiaxx/waveform_classification.

Random forest classification

To classify the brain structure each unit belongs to with extracellular waveforms, we used random forest classification. This supervised learning algorithm provides the contribution of each feature to classification accuracy. In addition, because the accuracy of random forest classification is averaged across many estimators, it is less likely to overfit the data than a decision tree. The two hyperparameters for our random forest classifier, the number of estimators and the depth of the decision trees, were estimated via grid search implemented in Scikit-Learn using 5-fold cross-validation for different feature sets. Because classifier performances for different feature sets plateaued above certain hyperparameter values (Supplemental Figure 4), we chose a fixed set of hyperparameters that reached plateau performance (maximum tree depth=14 and number of estimator = 80) for all feature sets rather than fine-tuning hyperparameters for individual set of features.

All classifications were performed with a 5-fold cross-validation where the classifier is trained on a subset of the data (80%), and then the classifier's performance is evaluated on the held-out test data (20%). Classification accuracy is determined by the out-of-bag (OOB) score, which is estimated based on the prediction accuracy of data left out in each fit of decision tree (estimator) on bootstrapped sub-samples. Since we have significantly more units in the visual cortex than other brain regions, we sub-sampled 77 units (determined by n=78 units in cerebellum) randomly from all regions to balance the size of dataset from

different brain regions to minimize the influence from underlying class distribution on accuracy. The confusion matrix is computed by comparing predicted classes to true classes for 100 sub-sampled datasets under 100 random initial states. The trend of classification accuracy compared across different feature sets is not different between matched-sample classification and unbalanced-sample classification.

K-means clustering

We applied k-means algorithm to determine cell clusters within visual cortex. The k-means method is a widely-used clustering technique that seeks to find centroids that minimize the average Euclidian distance between points in the same cluster to the centroid. However, one of its drawbacks is the requirement for the number of clusters, K, to be specified before the algorithm is applied. We applied two methods in estimating number of clusters in visual cortex (see Supplementary Figure 5). One method is the standard elbow method, which estimates the percentage of variance explained for a given number of K. The number of K is estimated at the point when the curve turns to plateau. Another method estimates the data distribution for a given K, calculated by a density function f(K) (Pham et al., 2005). The value of f(K) is the ratio of the real distortion to the estimated distortion. When the data is uniformly distributed, the value f(K) is 1. When there are areas of concentration in the data distribution, the value of f(K) decreases. Therefore, the number of K cluster is determined by finding the minimum of f(K). Combining estimation of K using the above two methods, we decide on K and apply k-means to data with appropriate number of K for 1000 times with random initial values.

t-SNE

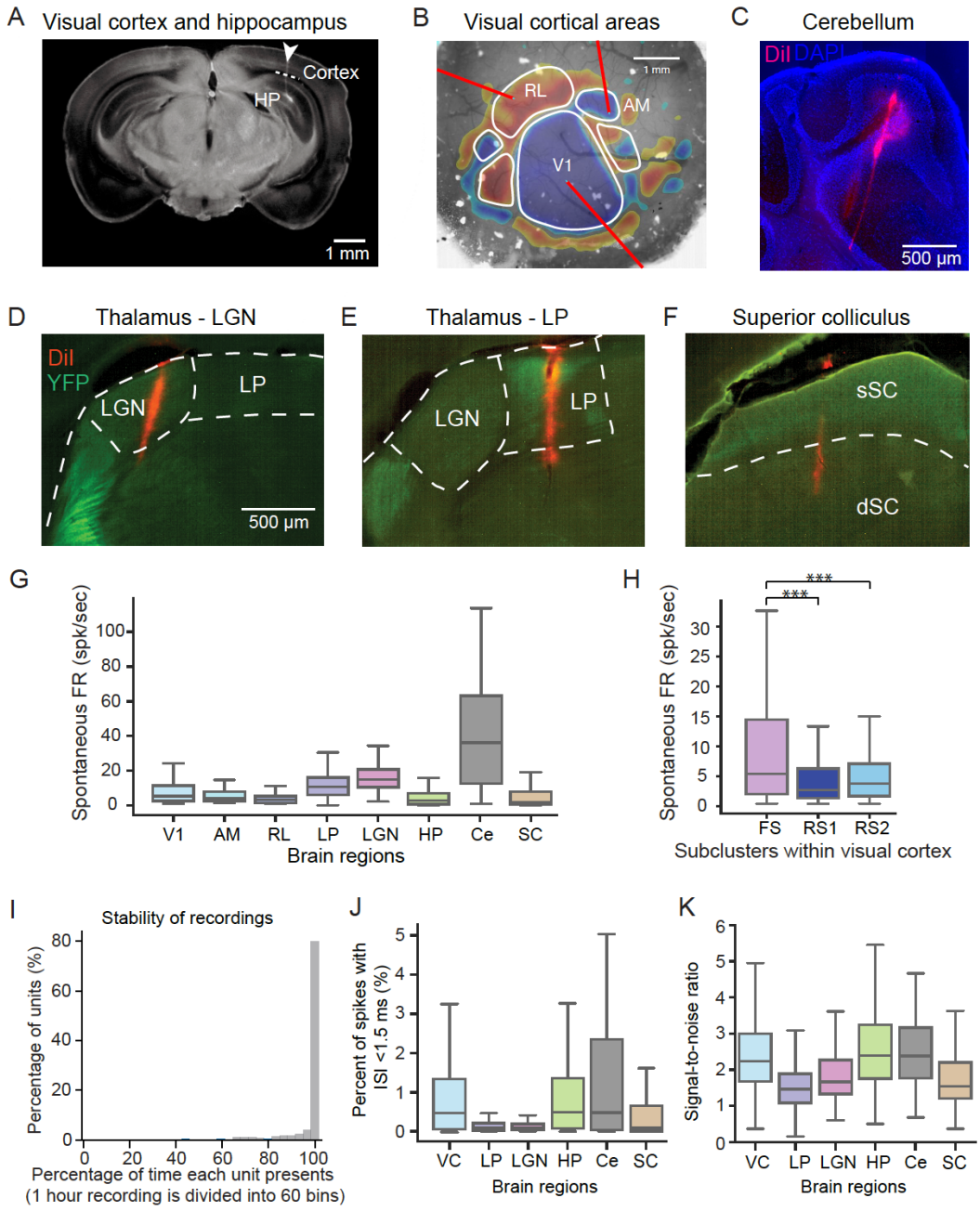
t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction for the visualization of high-dimensional datasets. We used Laures van der Maaten's method (Van Der Maaten and Hinton, 2008) to visualize all units in 2-dimentional space with different features. The purpose is to visualize whether units from same structure are mapped to similar regions with a given feature set (Figure 4A). We can also visualize unlabeled data to check whether there are any clusters in units with a given feature set.

Classification of opto-tagged neurons

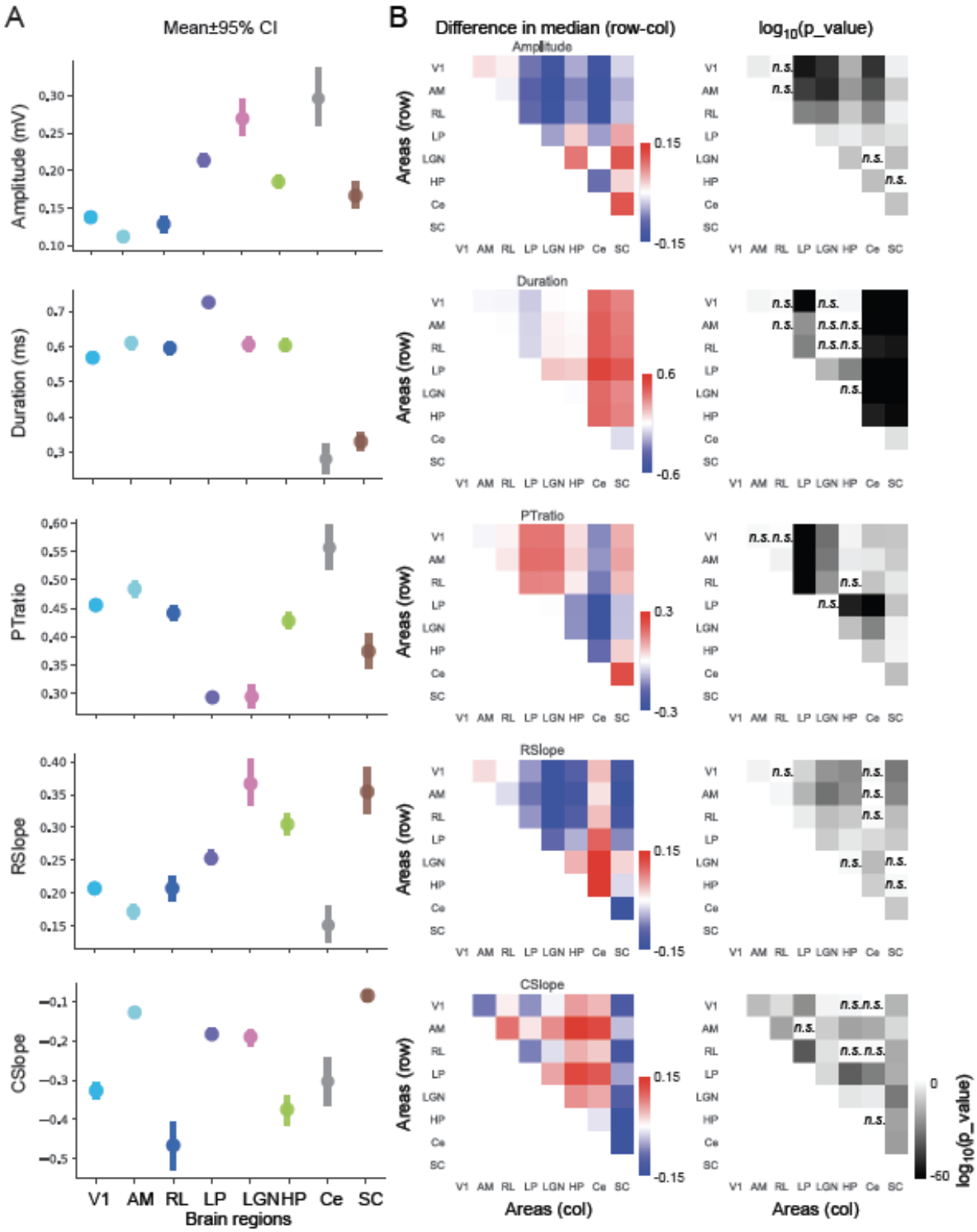
We developed a new method to determine opto-tagged cells based on their response to different light stimulation pattern. Response post-stimulus time histograms (bin size=1 ms) to different light patterns (individual square-wave pulses lasting 5 ms or 10 ms, or 1 s raised cosine ramps) are concatenated to form a response vector for each unit, which result in an n_unit (neurons) by m_feature (time) matrix. We then normalized the matrix and applied PCA to reduce the dimensionality while keeping 90% variance. K-means was applied to the normalized data for 1000 times with random initial value. This process is repeated 100 times to generate a probability matrix with each pixel value indicating the probability of a pair of units belonging to the same cluster. Hierarchical clustering is applied to the probability matrix to determine different clusters. The cluster of units with responses that tightly follow the light pattern is defined as opto-tagged cells. The rest are determined as non-optotagged units. These results were confirmed with labeling based on changes in FR and the significance of those changes (Hangya et al., 2015).

CSD

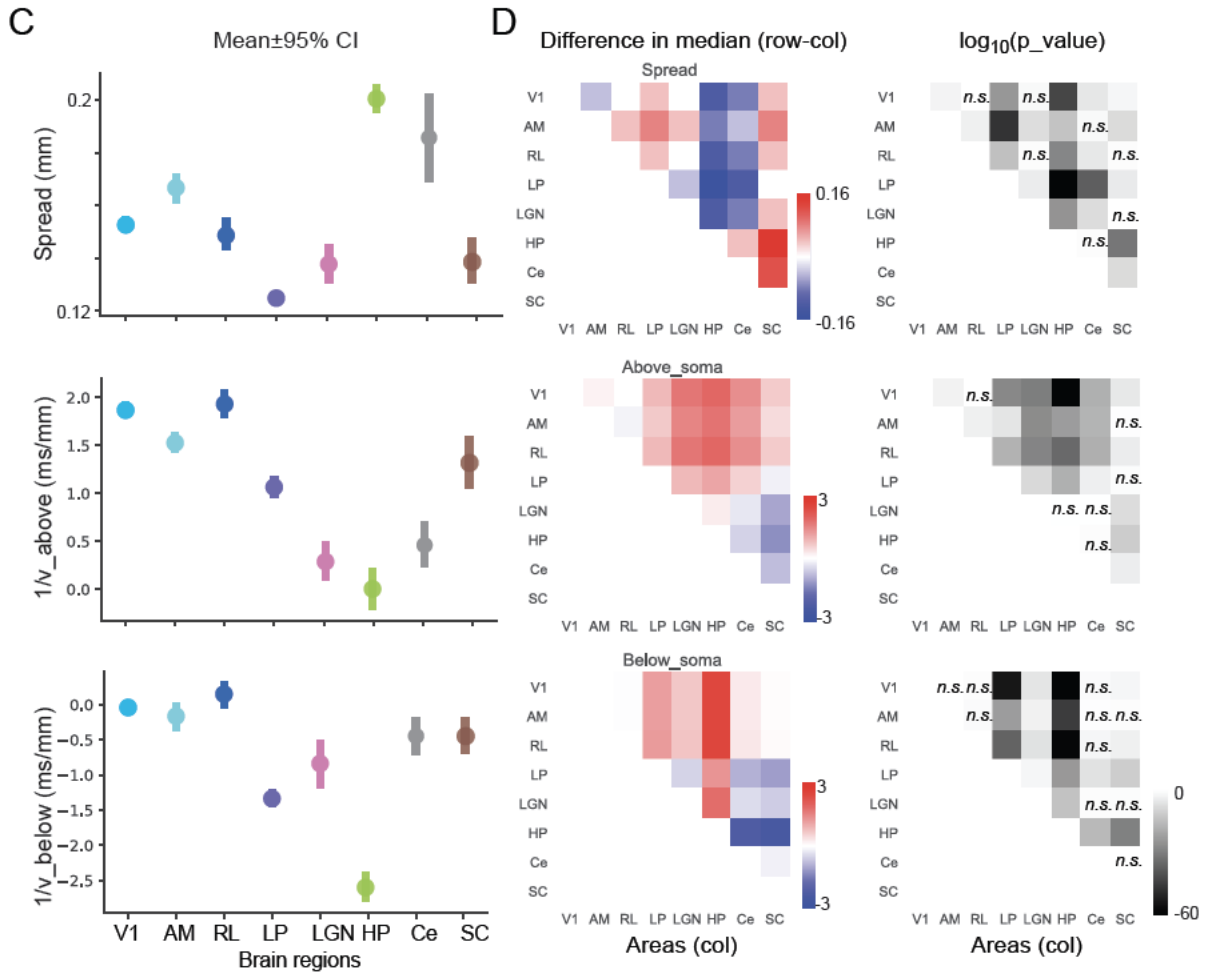
For estimation of unit depth within the cortex, we performed current source density (CSD) analysis by computing the average evoked (stimulus-locked) LFP at each site, smoothing these signals across sites, and then calculating the second spatial derivative (Smith et al., 2013; Stoelzel et al., 2009). For spike triggered CSD, we applied the same metric to the average multi-channel waveform across units (Bereshpolova et al., 2007).



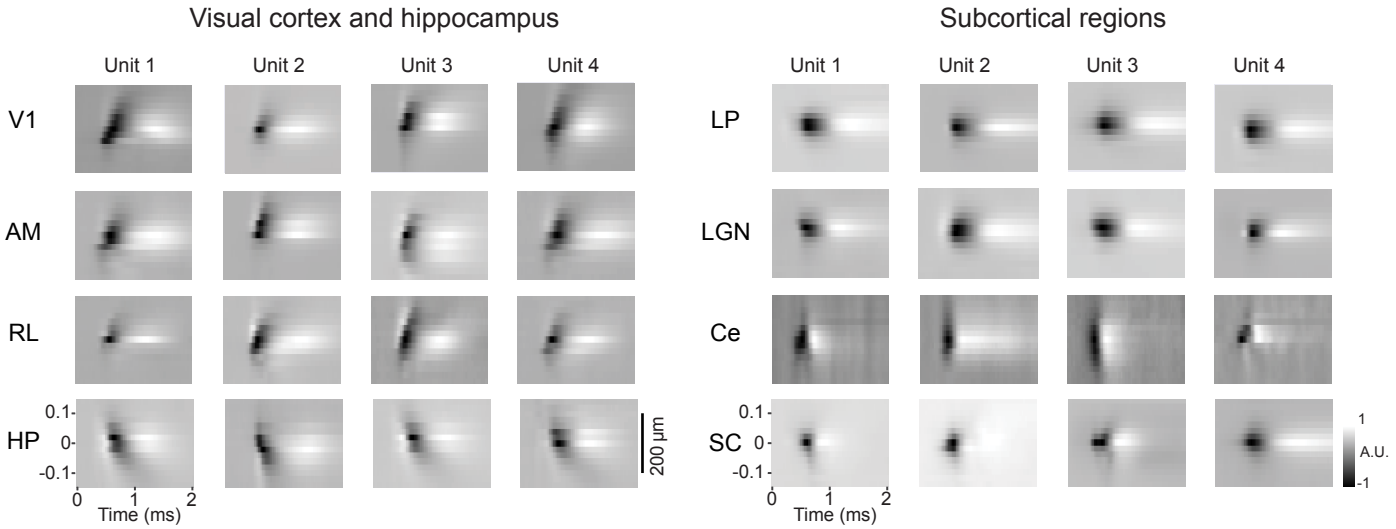
Supplementary Figure 1. Histological verification of recording location from different brain regions. **A)** Example probe track imaged and reconstructed using optical project tomography (indicated by white arrow). The white matter separates the visual cortex and hippocampus (illustrated with white dashed line). **B)** Overlay of retinotopic sign map and vasculature image to define subareas in visual cortex to guide probe insertion (red lines indicate probes). **C)** Example probe track in cerebellum (red). DAPI (blue) labels cell bodies and shows cytoarchitecture of cerebellum. **D)** Probe track labeled with DiI (red) from example LGN recording in a VGAT-ChR2-EYFP mouse (green labeling shows EYFP in inhibitory neurons of the thalamic reticular nucleus). **E,F)** Probe track verification for LP and SC recordings in NTSR1-GN209 x Ai32 mice. sSC: superficial layers of SC, dSC: deep layers of SC. **G)** Distribution of spontaneous firing rate in different cortical and subcortical areas. **H)** Spontaneous firing rate for subclasses of units in visual cortex (*** p_value<0.001). **I)** Stability of recordings for 1 hour sessions quantified as percent of time each unit is present during the session. Overall, 86±12% units are present for more than 90% of the total time. **J)** Refractory period violations of sorted units. We use a conservative value for refractory period of 1.5 ms consistently across areas, which may overestimate the true violation rate. Nonetheless the average percent of refractory period violations across all units is very low (0.88±0.03%). **K)** Signal-to-noise ratio of spike waveforms from different brain areas (Suner et al., 2005).



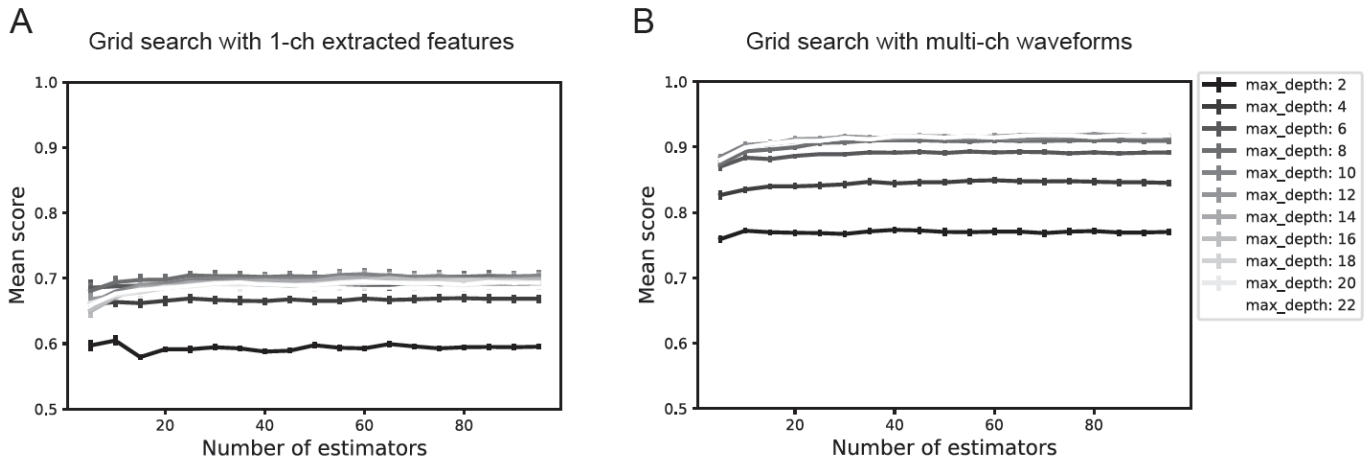
Supplementary Figure 2. Statistics of extracted waveform features. A) Mean and confidence interval (95% confidence interval) of features extracted from 1-ch waveforms from different brain regions. **B)** Pairwise statistics of extracted waveform features. Left panel: Difference of median of the distributions (row-column) for features extracted from 1-channel waveform. Right panel: paired t-test p_value shown in log10 scale, with darker color corresponds to smaller p-value. Significant comparisons were determined by p_value > 0.05 (Bonferroni corrected n=28). Non-significant comparisons are indicated by n.s.



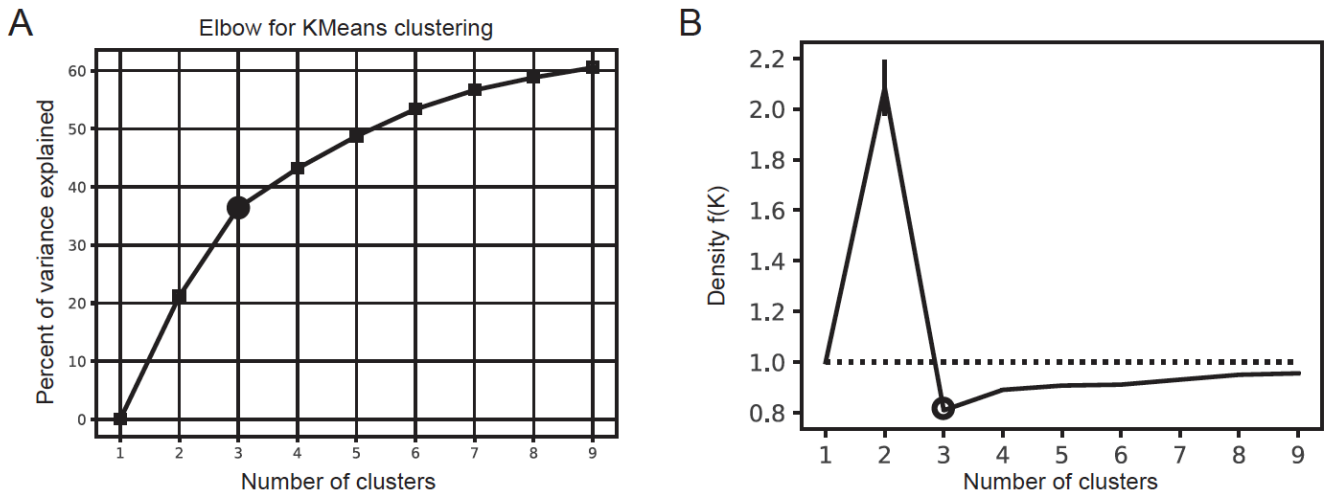
Supplementary Figure 2 (continued). Statistics of extracted waveform features. C) Mean and confidence intervals of features extracted from multi-channel waveforms from different brain areas. **D)** Pairwise comparisons for features extracted from multi-channel waveforms.



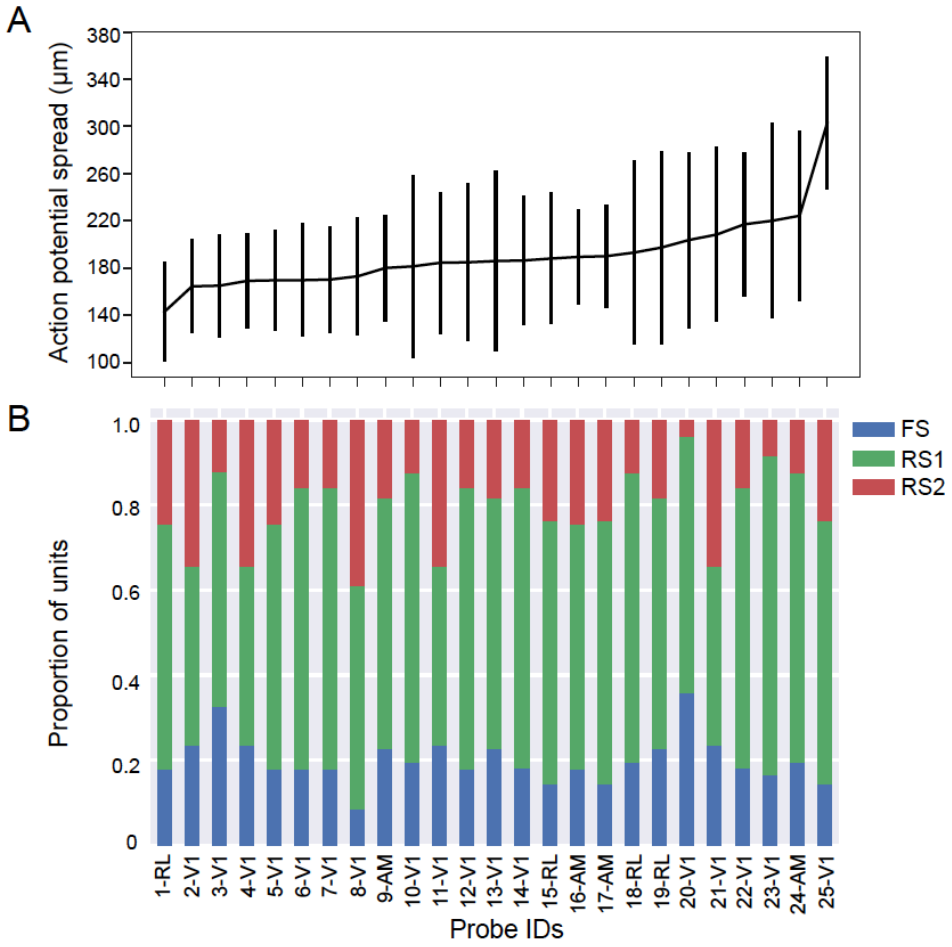
Supplementary Figure 3. Additional examples of multi-channel waveforms from different brain areas. Four example single unit waveforms from each of eight different brain regions. Each heat map shows amplitude of spike over time, measured on channels above and below the soma.



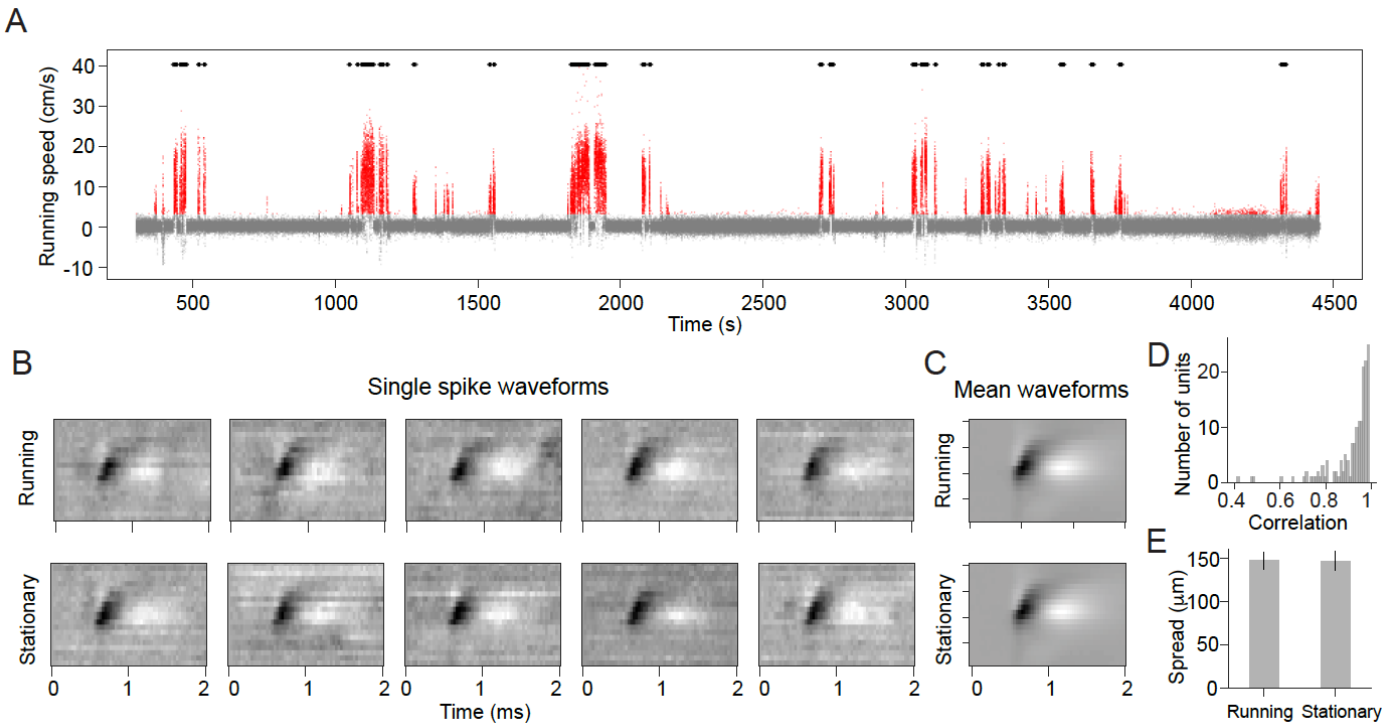
Supplementary Figure 4. Grid search with different hyperparameters for random forest classification. **A)** 5-fold cross-validation with features extracted from 1-channel waveforms (n_feature=4) as a function of number of estimators (decision trees). Standard error calculated across 5 repetitions. Color intensity of the lines corresponds to different tree depths (see legend). **B)** 5-fold cross-validation with multi-channel waveforms (n_feature=1800) as a function of number of estimators. This result indicates that performance of random forest classification is more sensitive to the maximum tree depth rather than the number of estimators for our dataset. Since performance plateaus above certain tree depth and number of estimators for different feature sets, there is no need to fine tune hyperparameters individually for different feature sets.



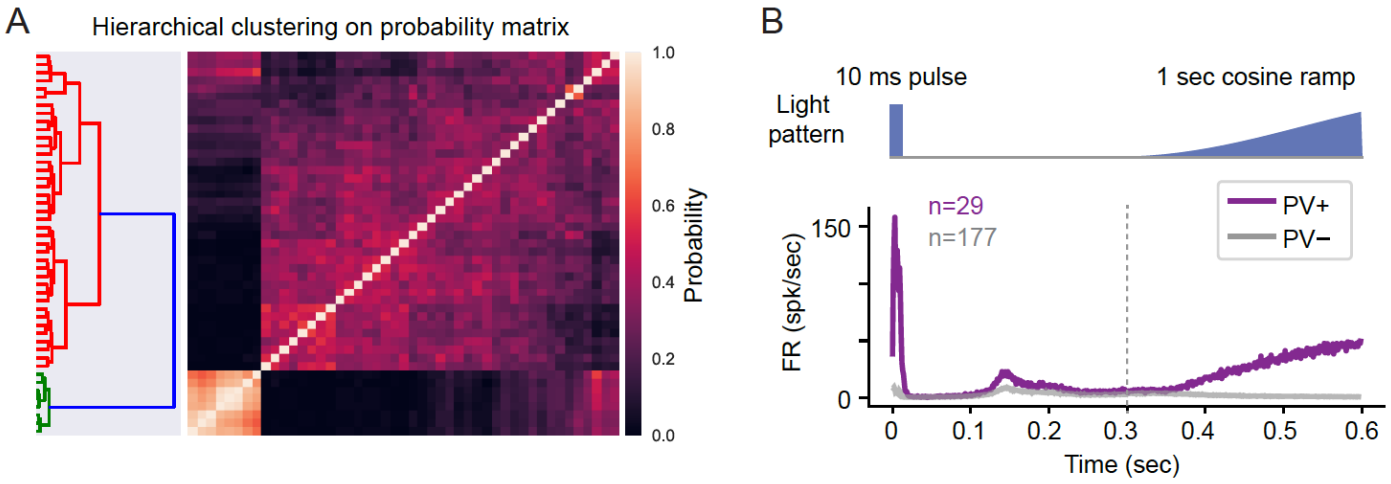
Supplementary Figure 5. Determining number of K clusters in visual cortex. A) Elbow method for finding number of K clusters. Percent of variance explained is calculated as a function of number of K. The turning point is highlighted by a black circle. **B)** Density function f(K) as a function of number of k (see Methods). Values lower than 1 indicate concentration of data given the value of K. The number of K clusters that gives rise to the minimum f(K) is determined as the optimal number of clusters (highlighted by a red circle). Both methods indicate K = 3 clusters.



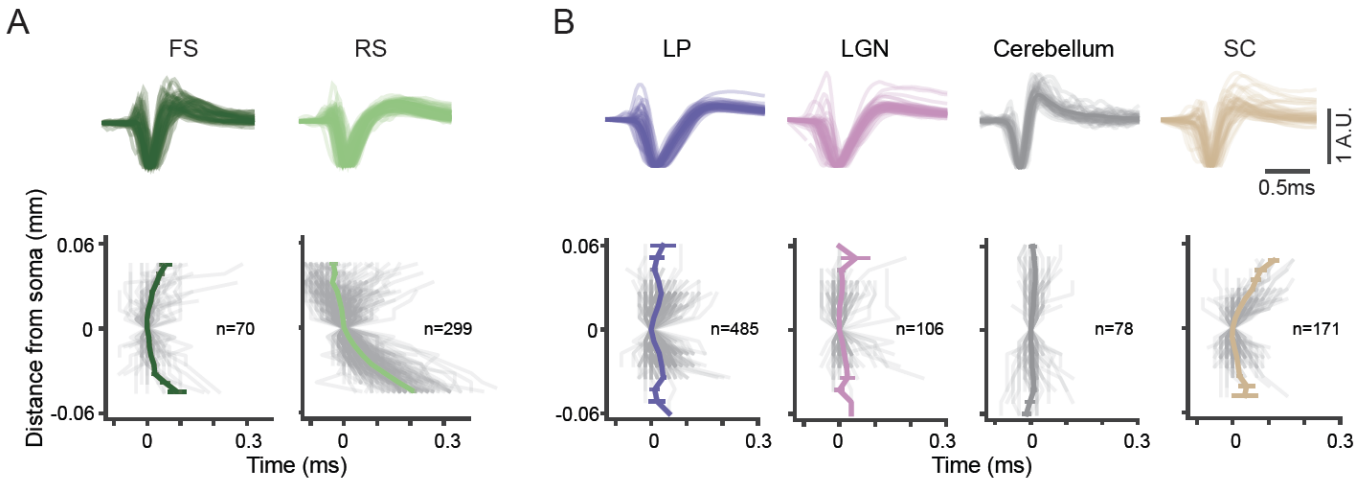
Supplemental Figure 6. Proportion of neuronal clusters in visual cortex for different probes. **A**) Sorted distribution of spread of RS1 units on each probe (n=25 probes from 11 mice). The spread of RS1 units show unidirectional BAPs, which could indicate the alignment of probes with the apical dendrites of individual neurons. **B**) Proportion of FS, RS1 and RS2 units on each probe arrange by the averaged spread of RS1 units of each probe. This result showed that there is no clear bias of RS1/RS2 distribution as a function of individual probes, cortical areas or the spread of BAPs.



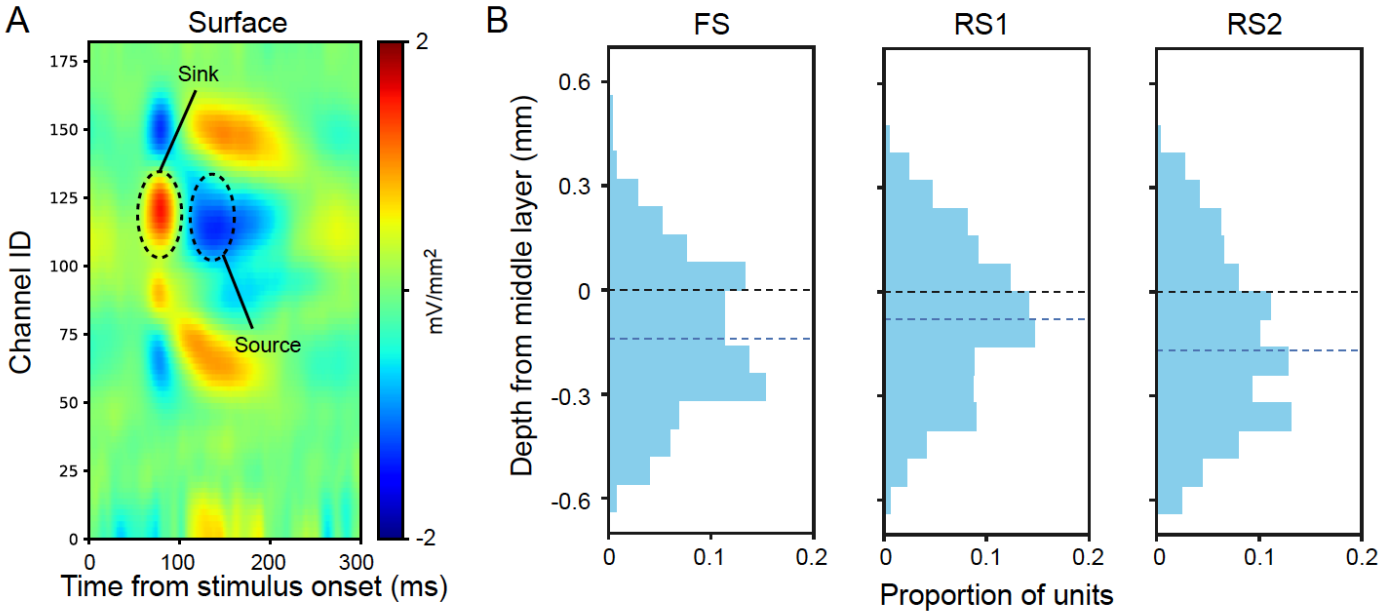
Supplementary Figure 7. Stability of backpropagating action potentials during running and stationary states. A) Running speed of an example mouse during a complete recording session. Gray dots indicate speed smaller than 3 cm/s, and red dots indicate speed larger than 3 cm/s. Running epochs are indicated by black bars at the top, which is defined by continuous running faster than 10cm/s for 2 seconds. **B**) Example single spike waveforms of cortical unit with BAPs during different epochs of running or stationary. **C**) Averaged waveform of an example unit during running and stationary states. **D**) Distribution of correlation coefficient between the averaged multi-channel waveforms during running and stationary states for all units in the example mouse (0.92±0.09, n=131 units). **E**) Averaged spread of BAPs during running and stationary states (t-test $p>0.05$). Error bar indicates standard deviation.



Supplementary Figure 8. Hierarchical clustering of neuronal response to photostimulation. **A)** Post-stimulus time histograms (bin size=1 ms) describing response to photostimulation is used as a feature vector for each unit and K-means clustering is applied 1000 times to the PCA components of the response matrix to construct a co-clustering matrix that reflects the probability of pairs of units belonging to the same cluster. Hierarchical clustering is applied to the probability matrix to find different clusters. This example shows the probability matrix (right) and the hierarchical clustering dendrogram (left) for an example mouse (PV-IRES-Cre; Ai32 (ChR2)) for units in visual cortex. The cluster of units with responses that tightly follow the light pattern is defined as the opto-tagged subpopulation. **B)** Summary plot of mean response to light pattern for all PV+ neurons (n=29 units) from two mice and mean response of the non-opto-tagged neurons (n=177 units).



Supplementary Figure 9. Spike propagation profiles for subcortical regions. **A)** Top: Example waveforms from FS and RS units in hippocampus. Bottom: waveform trough propagation trajectories for all hippocampal FS and RS units. **B)** Top: Example waveforms from LP, LGN, SC and cerebellum. Bottom: propagation trajectories for each region. Gray lines indicate individual units and the colored lines indicate the mean±sem.



Supplementary Figure 10. Laminar distribution of visual cortex waveform clusters. A) An example CSD from one insertion. Layer 4 is estimated by the structure of current sinks and sources. Center of mass of the first sink is defined as 0, indicating middle of layer 4. Data from 8 mice (n=1285 units) is aligned to middle layer of visual cortex. **B)** Proportion of units as a function of laminar depth. Dashed black line indicates depth 0. Dashed blue line indicate median of distribution. The distributions of RS1 and RS2 are significantly different (t-test, p=1.44E-5).