

A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex

Saskia E. J. de Vries^{1,5*}, Jerome A. Lecoq^{1,5*}, Michael A. Buice^{1,5*}, Peter A. Groblewski¹, Gabriel K. Ocker¹, Michael Oliver¹, David Feng¹, Nicholas Cain¹, Peter Ledochowitsch¹, Daniel Millman¹, Kate Roll¹, Marina Garrett¹, Tom Keenan¹, Leonard Kuan¹, Stefan Mihalas¹, Shawn Olsen¹, Carol Thompson¹, Wayne Wakeman¹, Jack Waters¹, Derric Williams¹, Chris Barber¹, Nathan Berbesque¹, Brandon Blanchard¹, Nicholas Bowles¹, Shiella D. Caldejon¹, Linzy Casal¹, Andrew Cho¹, Sissy Cross¹, Chinh Dang¹, Tim Dolbeare¹, Melise Edwards¹, John Galbraith¹, Nathalie Gaudreault¹, Terri L. Gilbert¹, Fiona Griffin¹, Perry Hargrave¹, Robert Howard¹, Lawrence Huang¹, Sean Jewell², Nika Keller¹, Ulf Knoblich¹, Josh D. Larkin¹, Rachael Larsen¹, Chris Lau¹, Eric Lee¹, Felix Lee¹, Arielle Leon¹, Lu Li¹, Fuhui Long¹, Jennifer Luviano¹, Kyla Mace¹, Thuyanh Nguyen¹, Jed Perkins¹, Miranda Robertson¹, Sam Seid¹, Eric Shea-Brown^{1,3}, Jianghong Shi³, Nathan Sjoquist¹, Cliff Slaughterbeck¹, David Sullivan¹, Ryan Valenza¹, Casey White¹, Ali Williford¹, Daniela M. Witten^{2,4}, Jun Zhuang¹, Hongkui Zeng¹, Colin Farrell¹, Lydia Ng¹, Amy Bernard¹, John W. Phillips¹, R. Clay Reid¹ and Christof Koch¹

To understand how the brain processes sensory information to guide behavior, we must know how stimulus representations are transformed throughout the visual cortex. Here we report an open, large-scale physiological survey of activity in the awake mouse visual cortex: the Allen Brain Observatory Visual Coding dataset. This publicly available dataset includes the cortical activity of nearly 60,000 neurons from six visual areas, four layers, and 12 transgenic mouse lines in a total of 243 adult mice, in response to a systematic set of visual stimuli. We classify neurons on the basis of joint reliabilities to multiple stimuli and validate this functional classification with models of visual responses. While most classes are characterized by responses to specific subsets of the stimuli, the largest class is not reliably responsive to any of the stimuli and becomes progressively larger in higher visual areas. These classes reveal a functional organization wherein putative dorsal areas show specialization for visual motion signals.

Traditional understanding, based on decades of research, is that visual cortical activity can be largely characterized by responses to a specific set of local features (modeled with linear filters followed by nonlinearities) and that these features become more selective and specialized in higher cortical areas^{1–4}. However, it remains unclear to what extent this understanding can account for the whole of V1 (refs. ^{5–7}), let alone the rest of the visual cortex. A key challenge results from the fact that this understanding is based on many small studies, recording responses from different stages in the circuit and using different stimuli and analyses⁵. The inherent experimental selection biases and lack of standardization of this approach introduce additional obstacles to creating a cohesive understanding of cortical function. On the basis of these issues, influential reviews have questioned the validity of this standard model^{5–7} and have argued that “what would be most helpful is

to accumulate a database of single-unit or multi-unit data (stimuli and neural responses) that would allow modelers to test their best theory under ecological conditions” (ref. ⁵). To address these issues, we conducted a survey of visual responses across multiple layers and areas in the awake mouse visual cortex, while using a diverse set of visual stimuli. This survey was executed in pipeline fashion, with standardized equipment and protocols and with strict quality-control measures not dependent upon stimulus-driven activity (Methods).

Previous work in mouse has revealed functional differences among cortical areas in layer 2/3 in terms of the spatial and temporal frequency tuning of neurons in each area^{8,9}. However, it is not clear how these differences extend across layers and across diverse neuron populations. Here we expand such functional studies to include 12 transgenically defined neuron populations, including Cre driver

¹Allen Institute for Brain Science, Seattle, WA, USA. ²Department of Statistics, University of Washington, Seattle, WA, USA. ³Department of Applied Mathematics, University of Washington, Seattle, WA, USA. ⁴Department of Biostatistics, University of Washington, Seattle, WA, USA. ⁵These authors contributed equally: Saskia E. J. de Vries, Jerome A. Lecoq, Michael A. Buice. *e-mail: saskiad@alleninstitute.org; jerome@alleninstitute.org; michaelbu@alleninstitute.org

lines for excitatory populations across four cortical layers (from layer 2/3 to layer 6) and for two inhibitory populations (defined by Vip and Sst expression). Further, it is known that stimulus statistics affect visual responses, such that responses to natural scenes cannot be well predicted by responses to noise or grating stimuli^{10–15}. To examine the extent of this discrepancy and its variation across areas and layers, we designed a stimulus set that included both artificial (gratings and noise) and natural (scenes and movies) stimuli. While artificial stimuli can be easily parameterized and interpreted, natural stimuli are closer to what is ethologically relevant. Finally, as recording modalities have enabled recordings from larger populations of neurons, it has become clear that populations might code visual and behavioral activity in a way that is not apparent by considering single neurons alone¹⁶. Here we imaged populations of neurons (173 ± 115 neurons for excitatory populations and 19 ± 11 neurons for inhibitory populations, mean \pm s.d.) to explore both single-neuron and population coding properties.

We find that 77% of neurons in the mouse visual cortex respond to at least one of these visual stimuli, with many showing classical tuning properties, such as orientation- and direction-selective responses to gratings. These tuning properties exhibit differences across cortical areas and Cre lines. While subtle differences do exist between the excitatory Cre lines, these populations of excitatory neurons are largely similar; the more marked differences are among the inhibitory interneurons. The responses to all stimuli are highly sparse and variable. We find that the variability of responses is not strongly correlated across stimuli, in general, but this variability provides evidence for functional response classes. We validate these functional response classes with a model of neural activity that contains most of the basic features found in visual neurophysiological modeling (for example, ‘simple’ and ‘complex’ components) as well as the running speed of the mouse. For one class of neurons, these models perform quite well, predicting responses to both artificial and natural stimuli equally well. However, for many neurons, the models provide a poor description, particularly for those in our largest single class of neurons, those that respond reliably to none of our visual stimuli. The representation of these response classes across areas reveals a separation of motion processing from spatial computations. These results demonstrate the importance of a large, unbiased survey for understanding neural computation.

Results

By using adult C57BL/6 mice (age 108 ± 16 d, mean \pm s.d.) that expressed a genetically encoded calcium sensor (GCaMP6f) under the control of specific Cre driver lines (ten excitatory lines and two inhibitory lines), we imaged the activity of neurons in response to a battery of diverse visual stimuli. Data were collected from six different cortical visual areas (V1, LM, AL, PM, AM and RL) and four different cortical layers. Visual responses of neurons at the retinotopic center of gaze were recorded in response to drifting gratings (DG), flashed static gratings (SG), locally sparse noise, natural scenes (NS), and natural movies (NM) (Fig. 1f), while the mouse was awake and free to run on a rotating disc. In total, 59,610 neurons were imaged from 432 experiments (Table 1). Each experiment consisted of three 1-h imaging sessions, with 33.6% of neurons matched across all three sessions; the rest were present in either one or two sessions (Methods).

To systematically collect physiological data on this scale, we built data collection and processing pipelines (Fig. 1). The data collection workflow progressed from surgical headpost implantation and craniotomy to retinotopic mapping of cortical areas using intrinsic signal imaging, *in vivo* two-photon calcium imaging of neuronal activity, brain fixation and histology using serial two-photon tomography (Fig. 1a–c). To maximize data standardization across experiments, we developed multiple hardware and software tools (Fig. 1d). One of the key components was the development of

a registered coordinate system that allowed an animal to move from one data collection step to the next, on different experimental platforms, and maintain the same experimental and brain coordinate geometry (Methods). In addition to such hardware instrumentation, formalized standard operating procedures and quality-control metrics were crucial for the collection of these data over several years (Fig. 1e).

Following data collection, fluorescence movies were processed using automated algorithms to identify somatic regions of interest (ROIs) (Methods). Segmented ROIs were matched across imaging sessions. For each ROI, events were detected from $\Delta F/F$ by using an L0-regularized algorithm¹⁷ (Methods). The median average event magnitude during spontaneous activity was 0.0004 (arbitrary units, AU; event magnitude has the same units as $\Delta F/F$) and showed some dependence on depth and on transgenic Cre line (Extended Data Fig. 1).

For each neuron, we computed the mean response to each stimulus condition using the detected events, and we parameterized the neuron’s tuning properties. Many neurons showed robust responses, exhibiting orientation-selective responses to gratings, localized spatial receptive fields, and reliable responses to natural scenes and movies (Fig. 2a–f and Extended Data Fig. 2). For each neuron and each categorical stimulus (that is, drifting gratings, static gratings, and natural scenes), the preferred stimulus condition was identified as the one which evoked the largest mean response for that stimulus (for example, the orientation and temporal frequency with the largest mean response for drifting gratings). For each trial, the activity of a neuron was compared to a distribution of activity for that neuron taken during the epoch of spontaneous activity and a *P* value was computed. If at least 25% of the trials of the neuron’s preferred condition had a significant difference from the distribution of spontaneous activities (*P* < 0.05), the neuron was labeled responsive to that stimulus (the Methods describes the responsiveness criteria for locally sparse noise and natural movies). Neurons meeting this criterion showed a change in activity with some degree of reproducibility across trials. The maximum evoked responses were an order of magnitude larger than the spontaneous activity (Extended Data Fig. 1; median of 0.006 (AU) for neurons responsive to drifting gratings).

In total, 77% of neurons were responsive to at least one of the visual stimuli presented (Fig. 2g). The percentage of responsive neurons depended on area and stimulus, such that V1 and LM showed the highest number of visually responsive neurons. This proportion dropped in other higher visual areas and was lowest in RL, where only 33% of neurons responded to any of the stimuli. Natural movies elicited responses from the most neurons, while static gratings elicited responses from the fewest (Fig. 2h). In addition to varying by area, the percentage of responsive neurons also varied by Cre line and layer, suggesting functional differences across these dimensions (Extended Data Figs. 3–7).

For responsive neurons, visual responses were parameterized by computing several metrics, including preferred spatial frequency, preferred temporal frequency, direction selectivity, receptive field size, and lifetime sparseness (Methods). We mapped these properties across cortical areas, layers, and Cre lines to examine the functional differences across these dimensions (Fig. 3 and Supplementary Figs. 1 and 2).

Comparisons across areas and layers revealed that direction selectivity was highest in layer 4 of V1 (Fig. 3b). While previous literature has found higher direction selectivity in layer 4 within V1 (ref. ¹⁸), we found here that this result was significant across all layer 4-specific Cre lines and extended to the higher visual areas as well. Comparisons across the higher visual areas revealed that, in superficial layers, the lateral higher visual areas (LM and AL) showed significantly higher direction selectivity than the medial ones (PM and AM), but this difference was not significant in the deeper layers.

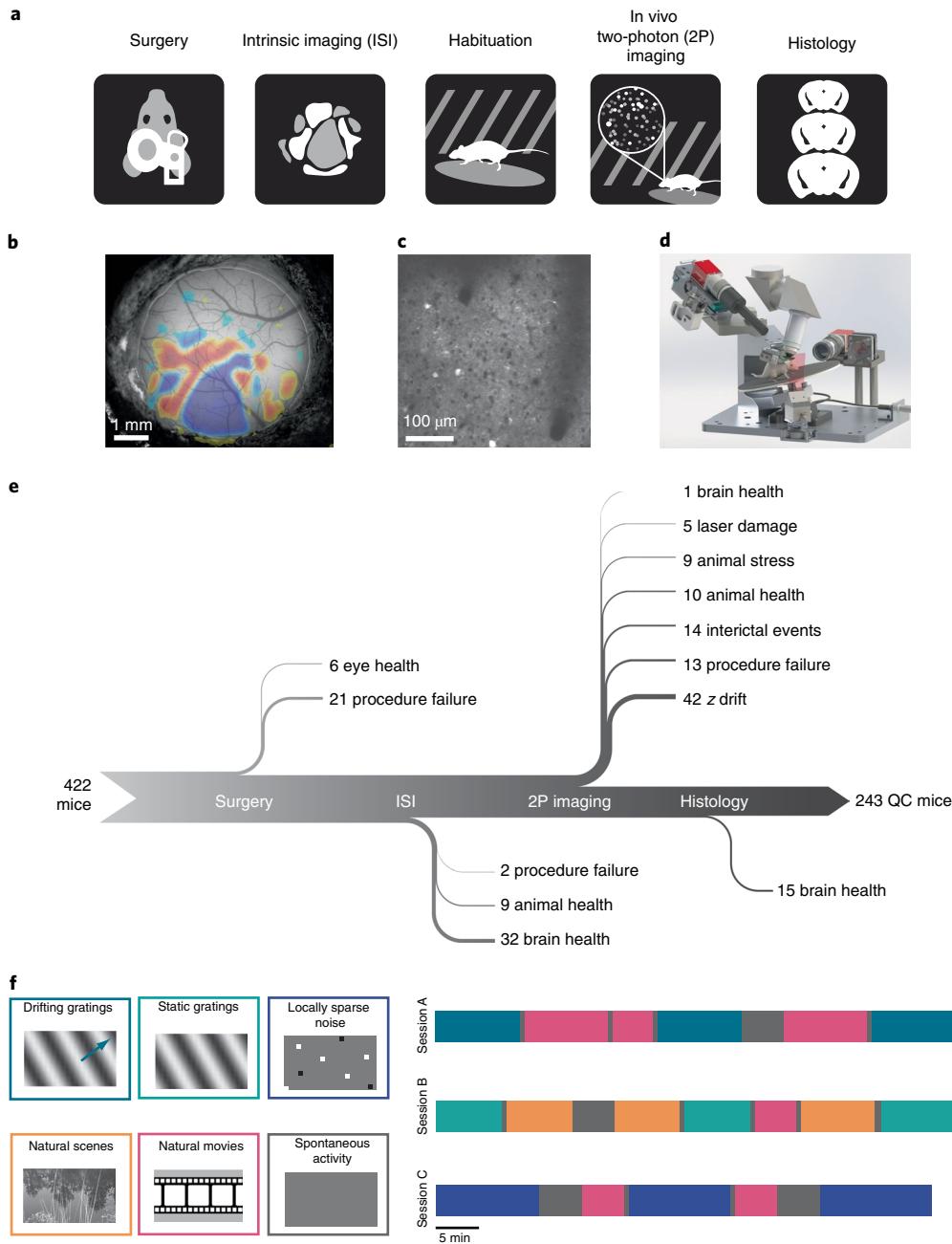


Fig. 1 | A standardized systems neuroscience data pipeline to map visual responses. **a**, Schematic describing the workflow followed by each mouse going through our large-scale data pipeline. **b**, Example intrinsic imaging map labeling individual visual brain areas. **c**, Example averaged two-photon imaging field of view ($400\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$) showcasing neurons labeled with GCaMP6f. **d**, Custom-designed apparatus to standardize the handling of mice in two-photon imaging. We engineered all steps of the pipeline to co-register data and tools, enabling reproducible data collection (Supplementary Figs. 13–16). **e**, Number of mice passing quality-control (QC) criteria established by standardized operating procedures at each step of the data collection pipeline, with the recorded reasons for failure. The data collection pipeline is closely monitored to maintain consistently high data quality. **f**, Standardized experimental design of sensory visual stimuli to map the response properties of neurons across the visual cortex. Six blocks of different stimuli were presented to mice (left) and were distributed into three separate imaging sessions called session A, session B, and session C across different days (right).

This erosion of the differences between higher visual areas in deeper layers was found for all metrics reported here, with the population differences less pronounced and often not significant in layers 5 and 6 (Fig. 3c–e and Supplementary Fig. 2).

Across all areas, layers, and stimuli, visual responses in mouse cortex were highly sparse (Fig. 3f). When considering the responses to natural scenes, we found that most neurons responded to very few scenes (examples in Fig. 2d). The sparseness of individual neurons was measured by using lifetime sparseness, which captures

the selectivity of a neuron's mean response to different stimulus conditions^{19,20} (Methods). A neuron that responds strongly to only a few scenes will have a lifetime sparseness close to 1 (Supplementary Fig. 3), whereas a neuron that responds broadly to many scenes will have a lower lifetime sparseness. Excitatory neurons had a median lifetime sparseness of 0.77 in response to natural scenes. While Sst neurons were comparable to excitatory neurons (median 0.77), Vip neurons exhibited low lifetime sparseness (median 0.36). Outside of layer 2/3, there was lower lifetime sparseness in areas RL, AM,

Table 1 | Visual coding dataset

Cre line	Layers	E/I	n (M/F)	Age range (d)	V1	LM	AL	PM	AM	RL
Emx1-IRES-Cre; Camk2a-tTA;Ai93	2/3, 4, 5	E	18 (13/5)	73-156	3,073 (10)	2,098 (8)	1,787 (7)	835 (4)	457 (3)	2,152 (9)
Slc17a7-IRES2-Cre; Camk2a-tTA;Ai93	2/3, 4, 5	E	31 (20/11)	80-149	4,840 (17)	3,230 (16)	374 (2)	1,970 (15)	235 (2)	137 (2)
Cux2-CreERT2; Camk2-tTA;Ai93	2/3, 4	E	38 (26/12)	79-155	5,081 (16)	2,792 (11)	3,103 (13)	2,361 (13)	1,616 (11)	1,578 (12)
Rorb-IRES2-Cre; Camk2a-tTA;Ai93	4	E	24 (14/10)	77-141	2,218 (8)	1,191 (6)	1,242 (6)	764 (7)	735 (8)	1,126 (5)
Scnn1a-Tg3-Cre; Camk2a-tTA;Ai93	4	E	7 (3/4)	75-133	1,873 (9)					
Nr5a1-Cre;Camk2a-tTA; Ai93	4	E	23 (15/8)	78-168	578 (8)	421 (6)	220 (6)	331 (7)	171 (6)	1,354 (6)
Rbp4-Cre_KL100; Camk2a-tTA;Ai93	5	E	23 (11/12)	68-144	458 (7)	485 (7)	441 (6)	509 (6)	355 (8)	93 (4)
Fezf2-CreER; Ai148 (corticofugal)	5	E	8 (4/4)	88-134	407 (4)	981 (5)				
Tlx3-Cre_PL56; Ai148 (cortico-cortical)	5	E	7 (5/2)	74-136	1181 (6)	946 (3)				
Ntsr1-Cre_GN220;Ai148	6	E	10 (5/5)	79-134	573 (6)	719 (7)		581 (5)		
Sst-IRES-Cre;Ai148	4, 5	I	30 (20/10)	67-154	266 (17)	301 (15)	24 (1)	247 (14)		46 (2)
Vip-IRES-Cre;Ai148	2/3, 4	I	24 (7/17)	81-148	352 (17)	315 (17)		387 (16)		

The numbers of cells (and experiments) imaged for each Cre line in each cortical visual area are indicated. In total, 59,610 cells imaged in 432 experiments in 243 mice are included in this dataset. E, excitatory; I, inhibitory; F, female; M, male.

and PM than in areas V1, LM, and AL. Lifetime sparseness did not increase outside of V1; responses did not become more selective in the higher visual areas (Fig. 3f and Supplementary Fig. 3).

The pattern in single-neuron direction selectivity was reflected in our ability to decode the visual stimulus from single-trial population vector responses, by using all neurons, responsive and unresponsive (Fig. 4a and Supplementary Fig. 4). We used a k -nearest-neighbors classifier to predict the grating direction. When matching the tuning properties, areas V1, AL, and LM showed higher decoding performance than areas AM, PM, and RL, and these differences were more pronounced in superficial layers than in deeper layers. Similarly, the population sparseness (Supplementary Fig. 3), a measure of the selectivity of each scene (that is, how many neurons respond on a given trial), largely mirrored the high average lifetime sparseness of the underlying populations (Fig. 4b). Such high sparseness suggests that neurons are active at different times and thus their activities are weakly correlated. The noise correlations of the populations reflected the results on population sparsity, where excitatory populations showed weak correlations (median 0.02) while inhibitory populations showed somewhat higher correlations (Sst neurons, median 0.06; Vip neurons, median 0.15) (Fig. 4c). The structure of the correlations in each population may serve to either help or hinder information processing^{16,21}. To test this, we measured the decoding performance when stimulus trials were shuffled to break trial-wise correlations. This had variable effects on decoding performance with little pattern across areas or Cre lines (Fig. 4d). While the decoding performance for excitatory populations in V1 was aided by removing correlations, in line with previous literature²², this effect was not consistent across other areas. The decoding performance for Sst populations, on the other hand, was more consistently hurt by removing correlations, suggesting that the high correlations among Sst neurons were informative about the drifting grating stimulus.

For all stimuli, the visually evoked responses throughout the cortex showed large trial-to-trial variability. Even when removing

the neurons deemed unresponsive, the percentage of responsive trials for most responsive neurons at their preferred conditions was low, with the median less than 50% (Fig. 5a and Supplementary Fig. 5). This means that the majority of neurons in the mouse visual cortex do not usually respond to individual trials, even when presented with the stimulus condition that elicits their largest average response. This was true throughout the visual cortex, although V1 showed slightly more reliable responses than higher visual areas and Sst interneurons, in particular, showed very reliable responses. The variability of responses was reflected in the high coefficient of variation, with median values for excitatory neurons above 2, indicating that these neurons are super-Poisson (Fig. 5b). We sought to capture this variability with a simple categorical model for drifting grating responses that attempts to predict the trial response (the integrated event magnitude during each trial) from the stimulus condition (the direction and temporal frequency of the grating or whether the trial was a blank sweep). This regression quantifies how well the average tuning curve predicts the response for each trial. Comparing the trial responses to the mean tuning curve showed a degree of variability even when the model was fairly successful (Fig. 5c). In line with this variability in visual responses, this model did a poor job of predicting responses to drifting gratings for most neurons (Fig. 5d). Few neurons were well predicted by their average tuning curve alone (21% of responsive neurons had $r > 0.5$, which became 11% when considering all neurons, where r is the cross-validated correlation between model prediction and actual response). As expected, the ability to predict responses was correlated with the measured variability ($r = 0.8$, Pearson correlation).

One possible source of trial-to-trial variability is the locomotor activity of the mouse. Previous studies have shown that, not only is some neural activity in the mouse visual cortex associated with running, but visual responses are also modulated by running²²⁻²⁶. The mice in our experiments were free to run on a disc, and animals showed a range of running behaviors (Supplementary Fig. 6). When ignoring the stimulus, we found that the activities of some neurons

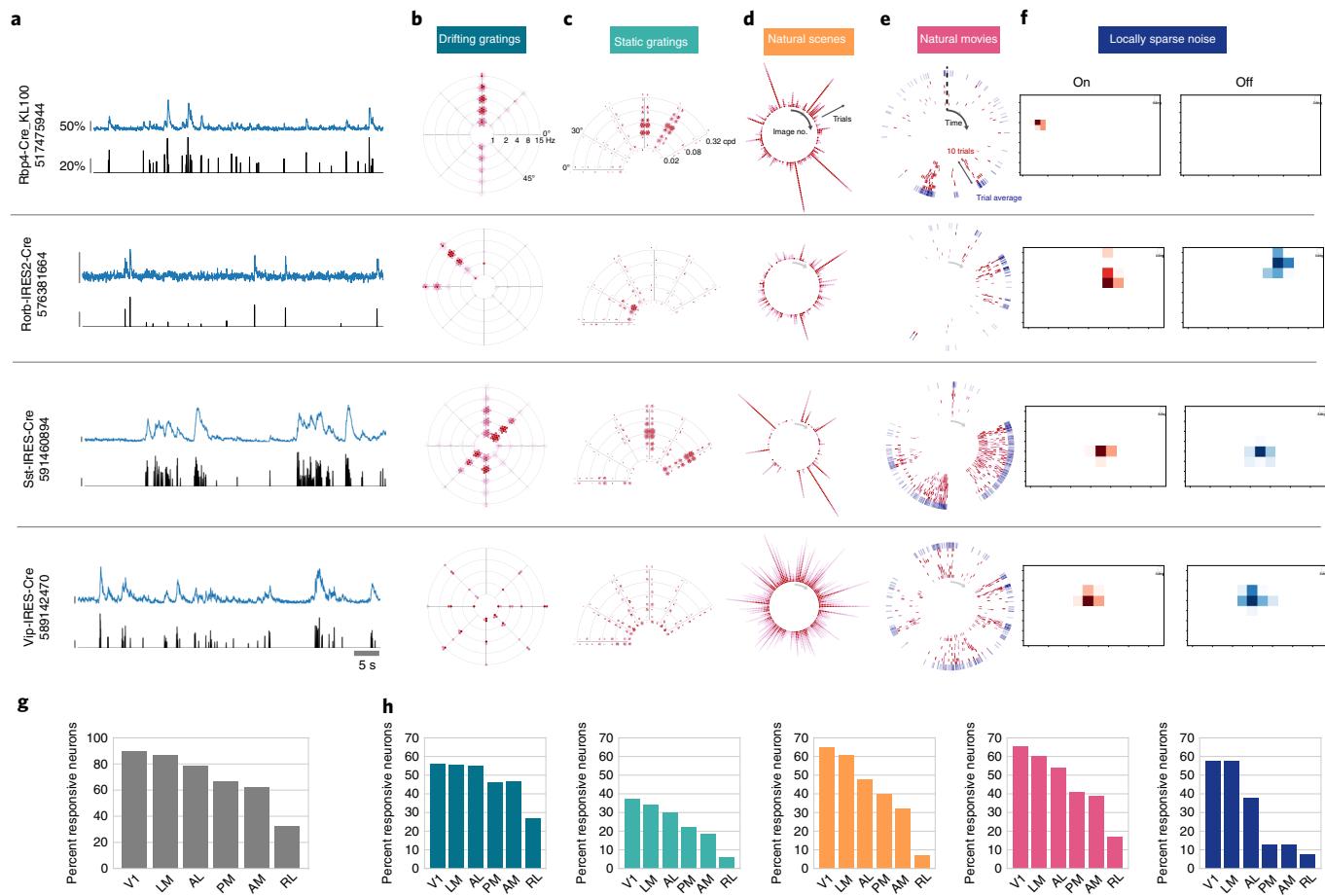


Fig. 2 | Neurons exhibit diverse responses to visual stimuli. **a**, Activity for four example neurons, including two excitatory neurons (Rorb, layer 4; Rbp4, layer 5) and two inhibitory neurons (Sst, layer 4; Vip, layer 2/3). $\Delta F/F$ (top, blue) and extracted events (bottom, black) are shown for each neuron. **b**, Star plots summarizing orientation and temporal frequency tuning for responses to the drifting grating stimulus. The arms of the star represent the different grating directions, and the rings represent the different temporal frequencies. At each intersection, the color of the circles corresponds to the strength of the response during a single trial of that direction and temporal frequency. (For details on response visualizations, see Extended Data Fig. 2.) **c**, Fan plots summarizing orientation and spatial frequency tuning for responses to static gratings. The arms of the fan represent the different orientations, and the arcs represent the spatial frequencies. For each condition, four phases of gratings were presented. cpd, cycles per degree. **d**, Corona plots summarizing responses to natural scenes. Each arm represents the response to an image, with individual trials represented by circles whose color corresponds to the strength of the response on that trial. Ten trials are represented in red, along with the mean peristimulus time histogram (PSTH) in the outer ring in blue. **e**, Track plots summarizing responses to natural movies. The response is represented as a raster plot moving clockwise around the circle. **f**, Receptive field subfields mapped using locally sparse noise. **g**, Percentage of neurons that responded to at least one stimulus across cortical areas. **h**, Percentage of neurons that responded to each stimulus across cortical areas. Colors correspond to the labels at the top of the figure. See Extended Data Figs. 3–7 for sample sizes.

were correlated with the running speed (Fig. 5e). While layer 5 showed strong correlations in all visual areas, in the other layers, V1 had stronger correlations than the higher visual areas, with some visual areas showing median negative correlations. Within V1, the inhibitory interneurons showed the strongest correlation with running, most notably Vip neurons in layer 2/3 (median 0.25), while the excitatory neurons showed weaker correlations (median 0.03).

For experiments with sufficient numbers of stimulus trials for a neuron's preferred condition when the mouse was both stationary and running ($>10\%$ trials for each), we compared the responses in these two states. In line with other reports, many neurons showed modulated responses, but the effect was modest (Fig. 5f). The majority of neurons showed enhanced responses. When considering the entire population, there was a 1.9-fold increase in the median evoked response with running. The effect on individual neurons, however, was varied such that only 13% of neurons showed significant modulation in these conditions ($P < 0.05$, Kolmogorov–Smirnov test).

To test whether running accounted for the variability in trial-wise responses to visual stimuli, we included a binary running state as a condition-dependent gain into the categorical regression (that is, computing separate tuning curves for the running and stationary conditions; Fig. 5g). This did not consistently and significantly improve the response prediction. When comparing the model performance when the running state was included to that of the stimulus-only model, we found that the distribution was largely centered along the diagonal, with a slight asymmetry in favor of the running-dependent model for the better-performing models (Fig. 5h; 28% of responsive neurons had $r > 0.5$ for stimulus \times running state, which was 21% when considering all neurons). This was further corroborated by a simpler model that predicts neural response on the basis of the running speed (rather than a binary condition and without stimulus information) (Supplementary Fig. 7). However, when considering only the 13% of neurons that showed significant modulation of evoked responses (Fig. 5f), the inclusion of running in the categorical model provided a clear predictive advantage (Fig. 5i; the

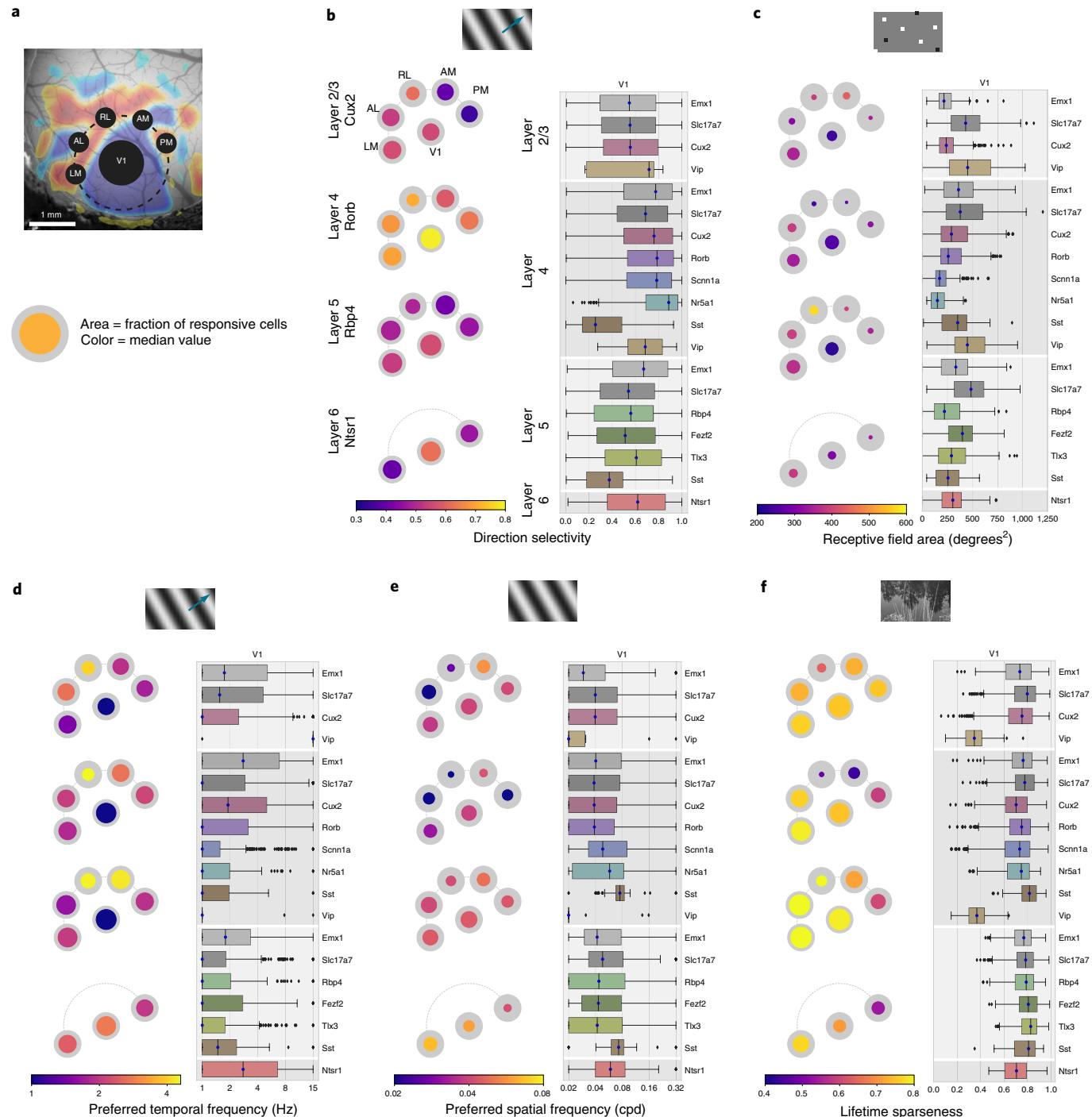


Fig. 3 | Tuning properties reveal functional differences across areas and Cre lines. **a**, Pawplot visualization summarizing the median value of a tuning metric across visual areas. Top: each visual area is represented as a circle, with V1 in the center and the higher visual areas surrounding it according to their location on the surface of the cortex. Bottom: each paw-pad (visual area) has two concentric circles. The area of the inner, colored circle, relative to the outer circle, represents the proportion of responsive cells for that layer and area. The color of the inner circle reflects the median value of the metric for the responsive cells. For a metric's summary plot, four pawplots are shown, one for each layer. Only data from one Cre line are shown for each layer. For each panel, a pawplot is paired with a box plot or a strip plot (for single-cell and population metrics, respectively) showing the full distribution for each Cre line and layer in V1. Data were assigned to cortical layers on the basis of both the Cre line and the imaging depth. Data collected less than 250 μ m from the surface were considered to be in layer 2/3, data collected between 250 and 365 μ m were considered to be in layer 4, data collected between 375 and 500 μ m were considered to be in layer 5, and data collected at 550 μ m were considered to be in layer 6. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range; points outside this range are shown as outliers. For other cortical areas, see Supplementary Fig. 1. **b**, Pawplots and box plots summarizing direction selectivity. See Extended Data Fig. 3 for sample sizes. **c**, Pawplots and box plots summarizing receptive field area. See Extended Data Fig. 5 for sample sizes. **d**, Pawplots and box plots summarizing preferred temporal frequencies. See Extended Data Fig. 3 for sample sizes. **e**, Pawplots and box plots summarizing preferred spatial frequencies. See Extended Data Fig. 4 for sample sizes. **f**, Pawplots and box plots summarizing lifetime sparseness of responses to natural scenes. See Extended Data Fig. 6 for sample sizes.

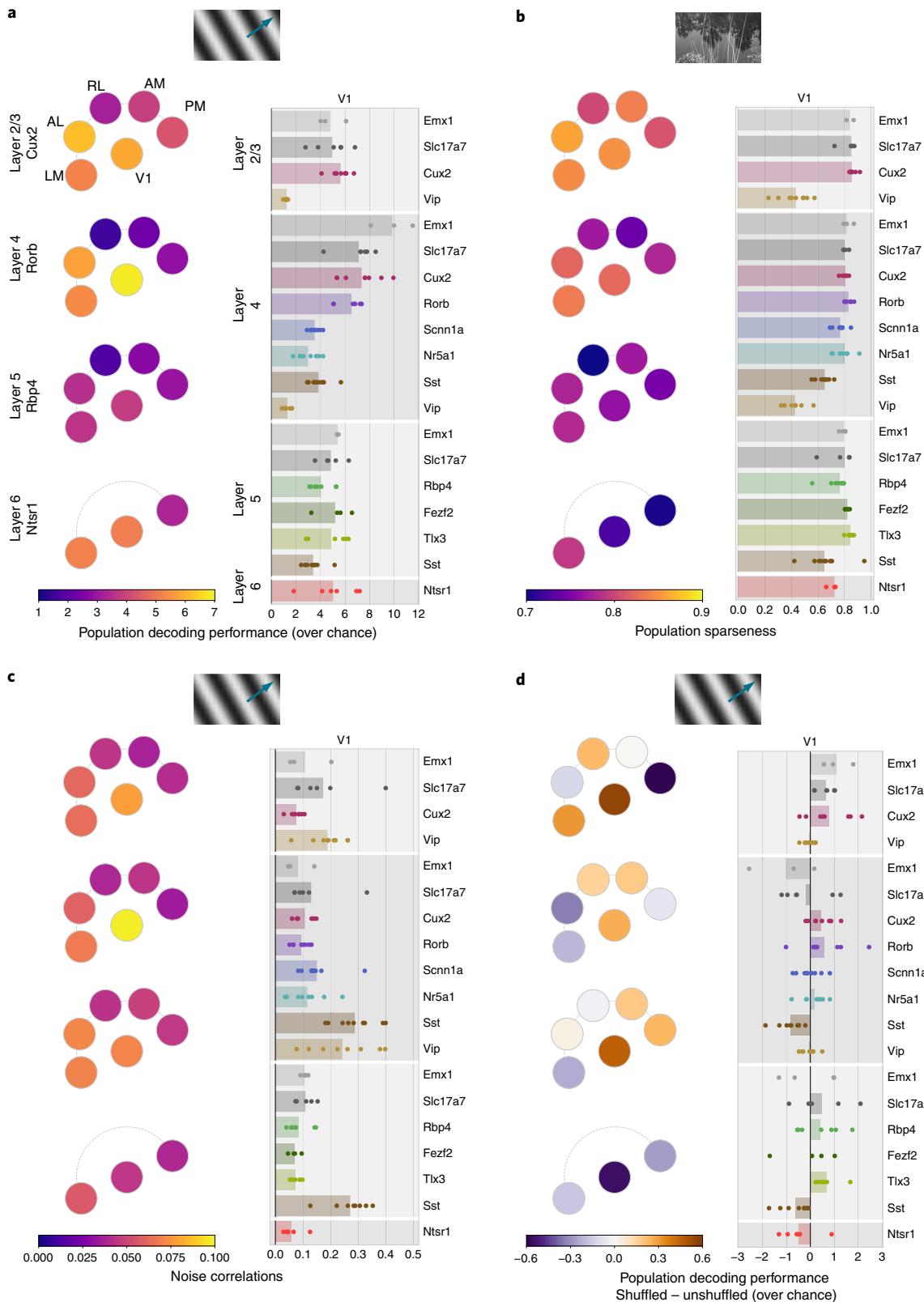


Fig. 4 | Population correlations have heterogeneous impact on decoding performance. **a**, Pawplots and strip plots summarizing decoding performance for drifting grating direction using k -nearest-neighbors analysis. Each dot represents the mean fivefold cross-validated decoding performance from a single experiment, with the median performance for a Cre line and layer represented by the bar. See Extended Data Fig. 3 for sample sizes (column ‘expts’). For other cortical areas, see Supplementary Fig. 4. **b**, Pawplots and strip plots summarizing the population sparseness of responses to natural scenes. See Extended Data Fig. 6 for sample sizes (column ‘expts’). For other cortical areas, see Supplementary Fig. 3. **c**, Pawplots and strip plots summarizing noise correlations in the responses to drifting gratings. See Extended Data Fig. 3 for sample sizes (column ‘expts’). **d**, Pawplots and strip plots summarizing the impact of shuffling on decoding performance for drifting grating direction. See Extended Data Fig. 3 for sample sizes (column ‘expts’). Note the diverging colorscale representing both negative and positive values.

mean r for stimulus only was 0.35 and for stimulus \times running was 0.44, whereas the mean r for non-modulated neurons for stimulus only was 0.21 and for stimulus \times running was 0.20).

One of the unique aspects of this dataset is the broad range of stimuli, allowing for a comparison of response characteristics and model predictions across stimuli. Surprisingly, knowing whether a neuron responded to one stimulus type (for example, natural scenes, drifting gratings, etc.) was largely uninformative of whether it responded to another stimulus type. Unlike the examples shown in Fig. 2, which were chosen to highlight responses to all stimuli, most neurons were responsive to only a subset of the stimuli (Fig. 6a). To explore the relationships between neural responses to different types of stimuli, we computed the correlation between the percentage of responsive trials for each stimulus. This comparison removes the threshold of ‘responsiveness’ and examines underlying patterns of activity. We found that most stimulus combinations were weakly correlated (Fig. 6b), demonstrating that knowing that a neuron responds reliably to drifting gratings, for example, carries little to no information about how reliably that neuron responds to one of the natural movies. There was a higher correlation between the reliability of the responses to the natural movie that was repeated across all three sessions (natural movies 1A, 1B and 1C), providing an estimate of the variability introduced by imaging across days and thus a ceiling for the overall correlations across stimuli. Very few of the cross-stimulus correlations approached this ceiling, with the exception of the correlation between static gratings and natural scenes.

We characterized the variability by clustering the reliability, defined by the percentage of significant responses to repeated stimuli. We used a Gaussian mixture model to cluster the 25,958 neurons that were imaged in both sessions A and B (Fig. 1f) and excluded the locally sparse noise stimulus owing to the lack of a comparable definition of reliability. Using neurons imaged in all three sessions did not qualitatively change the results (Supplementary Fig. 8). The clusters are described by the mean percentage of responsive trials for each stimulus for each cluster (Fig. 6c). Note that there was only a weak relationship between the percentage of responsive trials to one stimulus and the percentage of responsive trials to any other stimulus. We grouped the clusters into ‘classes’ by first defining a threshold for responsiveness by identifying the cluster with the lowest mean percentage of responsive trials across stimuli and then setting the threshold equal to the maximum value across stimuli plus 1 s.d. for that cluster. This allowed us to identify each cluster as responsive (or not) to each of the stimuli. Clusters with the same profile (for example, responsive to drifting gratings and natural

movies, but not to static gratings or natural scenes) were grouped into one of 16 possible classes.

The clustering was performed 100 times with different initial conditions to evaluate robustness. The optimal number of clusters, evaluated with model comparison, and the class definition threshold were consistent across runs (Supplementary Fig. 8). By far, the largest single class revealed by this analysis was that of neurons that were largely unresponsive to all stimuli, termed ‘none’, which contained $34 \pm 2\%$ of the neurons (Fig. 6d). Other large classes included neurons that responded to drifting gratings and natural movies (DG-NM; $14 \pm 3\%$), to natural scenes and natural movies (NS-NM; $14 \pm 2\%$) and to all stimuli (DG-SG-NS-NM; $10 \pm 1\%$).

Notably, we did not observe all 16 possible stimulus response combinations. For instance, very few neurons were classified as responding to one stimulus alone, with the most prominent exception being neurons that responded uniquely to natural movies. Thus, while the pairwise correlations between most stimuli were relatively weak, there was meaningful structure in the patterns of responses. Nevertheless, within each class there remains a great deal of heterogeneity. For example, within the class that responded to all stimuli, there was a cluster in which the neurons responded with roughly equal reliability to all four stimuli (cluster 27 in Fig. 6c) as well as clusters in which the neurons responded reliably to drifting and static gratings and only weakly to natural scenes and natural movies (clusters 25 and 28). This heterogeneity underlies the inability to predict whether a neuron responds to one stimulus given that it responds to another.

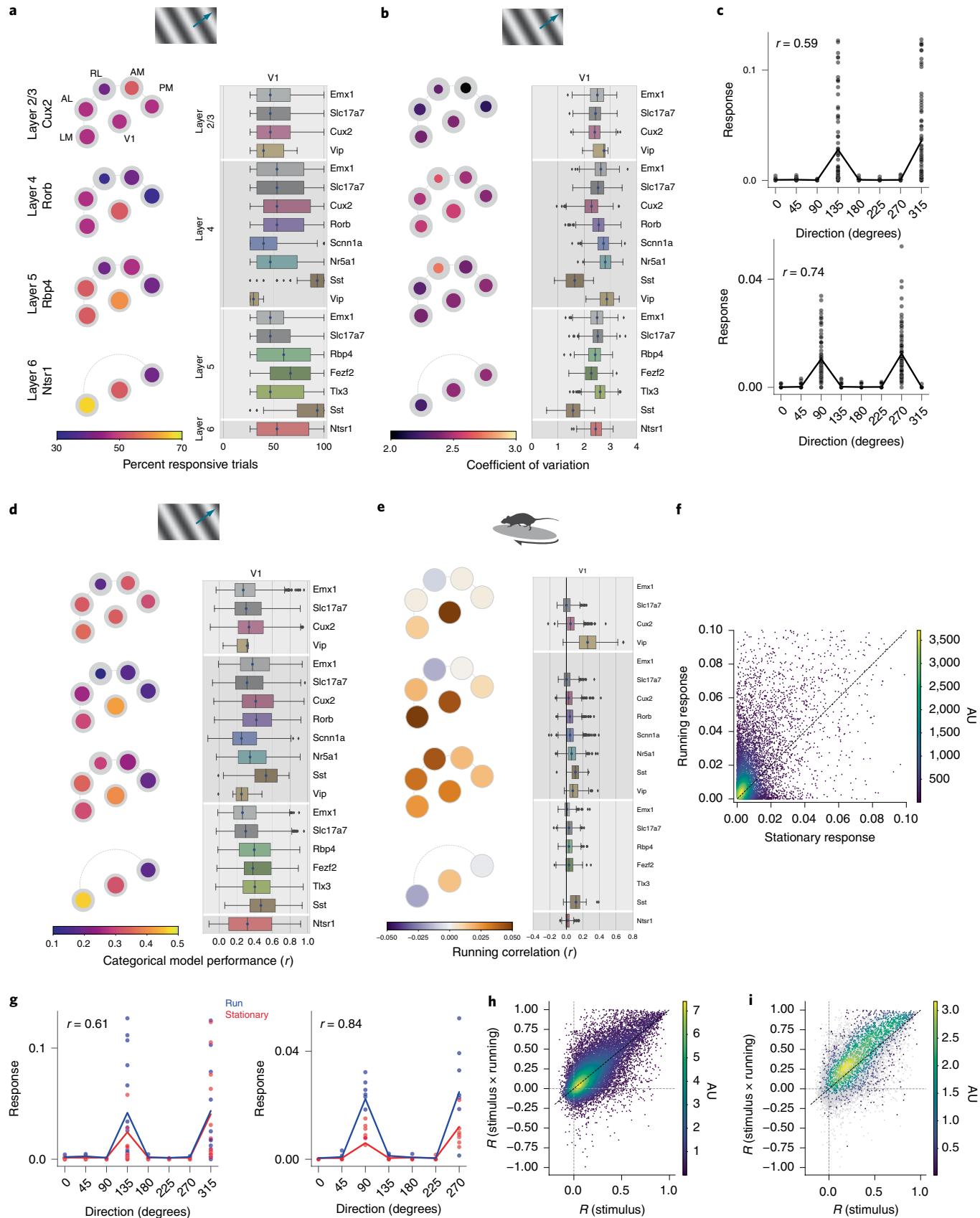
Classes were not equally represented in all visual areas (Fig. 6e). The ‘none’ class was larger in the higher visual areas than in V1 and was largest in RL (Fig. 2g). Classes related to moving stimuli, including NM, DG-NM and DG, had relatively flat distributions across the visual areas, excluding RL. The classes responsive to natural stimuli, including NS-NM, DG-NS-NM, SG-NS-NM, and DG-SG-NS-NM, were most numerous in V1 and LM, with lower representation in the other visual areas. This divergence in representation of the motion stimulus response classes and natural stimulus response classes in areas AL, PM, and AM is consistent with the putative dorsal and ventral stream segregation in the visual cortex³².

In addition to differential representation across cortical areas, the response classes were also differentially represented among the Cre lines (Fig. 6f). Notably, Sst interneurons in V1 had the fewest neurons labeled as ‘none’ and the most DG-SG-NS-NM neurons. Meanwhile, the plurality of Vip interneurons were in the classes responsive to natural stimuli, specifically natural movies.

Fig. 5 | Neural activity is extremely variable, and this variability is not accounted for by running behavior. **a**, Pawplots and box plots summarizing the percentage of responsive trials that had a significant response for each neuron’s preferred drifting grating condition. The responsiveness criterion was that a neuron responded in 25% of the trials; hence, the values in the box plots are bounded at 25%. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range; points outside this range are shown as outliers. For box plots for other cortical areas, see Supplementary Fig. 5. See Extended Data Fig. 3 for sample sizes. **b**, Pawplots and box plots summarizing the coefficient of variation for each neuron’s response to its preferred drifting grating condition. See Extended Data Fig. 3 for sample sizes. **c**, Two example neurons showing individual trial responses along with the mean tuning curve, where r is the Pearson correlation coefficient between the measured and predicted values. $n=45$ trials per stimulus condition. **d**, Pawplots and box plots summarizing the categorical regression, where r is the cross-validated Pearson correlation between model prediction and actual response. Only neurons that were responsive to drifting gratings when using our criterion are included. See Extended Data Fig. 3 for sample sizes. **e**, Pawplots and box plots summarizing the Pearson correlation of neural activity with running speed. Only neurons in imaging sessions where the running fraction was between 20% and 80% are included (Supplementary Fig. 6). See Extended Data Fig. 8 for sample sizes. For neurons present in multiple sessions that met the running criterion, the mean of their running correlations across these sessions was used here. Note the diverging colorscale representing both negative and positive values. **f**, Density plot of the evoked response to a neuron’s preferred drifting grating condition when the mouse is running ($\text{running speed} > 1 \text{ cm s}^{-1}$) compared to when it is stationary ($\text{running speed} < 1 \text{ cm s}^{-1}$). Only neurons that were responsive to drifting gratings and had a sufficient number of running and stationary trials for their preferred condition are included; $n=10,440$. **g**, Categorical model for two example neurons (as in **c**) in which the running (blue) and stationary (red) trials have been segregated, where r is the Pearson correlation coefficient between the measured and predicted values. $n=14$ (left) and 7 (right) trials per condition. **h**, Density plot of r for the categorical regression for drifting gratings using only the stimulus condition (horizontal axis) and the stimulus condition \times running state (vertical axis). Only neurons that were responsive to drifting gratings and had a sufficient number of running and stationary trials across stimulus conditions are included. $n=11,799$. **i**, As in **h**, except that only neurons that were significantly modulated by running are shown in the density ($n=2,791$); other neurons are in gray.

Having characterized neurons by their joint reliabilities to multiple stimuli, we next asked to what extent we can predict neural responses, not on a trial-by-trial basis but including the temporal

response dynamics, given the stimulus and knowledge of the animal's running condition. We used a model class that remains in widespread use for predicting visual physiological responses and



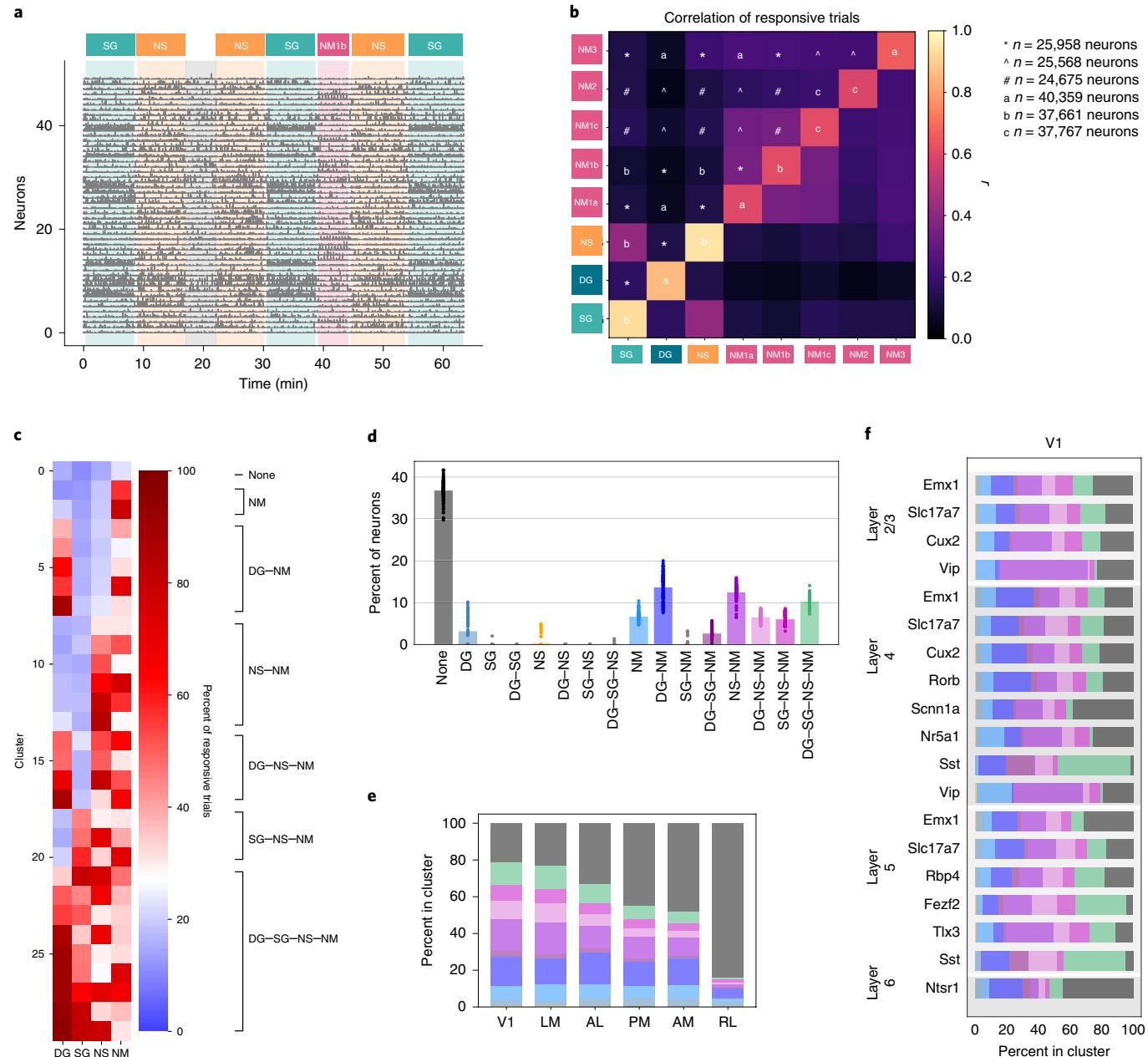


Fig. 6 | Correlated response variability reveals functional classes of neurons. **a**, Responses of 50 neurons during one imaging session (Cux2 neurons from layer 2/3 in V1), with stimulus epochs shaded according to the stimulus colors from Fig. 1f. **b**, Heat map of the Pearson correlation of the percentage of responsive trials for neurons' responses to each pair of stimuli. The diagonal is the mean correlation between bootstrapped samples of the percentage of responsive trials for the given stimulus. **c**, Mean percentage of responsive trials for each cluster per stimulus for one example clustering from the Gaussian mixture model ($n=25,958$ neurons). On the right, classes are identified according to the response profile of each cluster. **d**, Strip plots representing the percentage of neurons belonging to each class predicted by the model over 100 repeats. The mean across all repeats is indicated by the height of the bar. Clustering was performed on 25,958 neurons imaged in sessions A and B. **e**, The percentages of neurons belonging to each class per cortical area. Colors correspond to those in **d**. **f**, The percentage of neurons belonging to each class for each transgenic Cre line within V1. Colors correspond to those in **d**. For other cortical areas, see Supplementary Fig. 8.

that captures both simple and complex cell behaviors. The model structure uses a dense wavelet basis (sufficiently dense to capture spatial and temporal features at the level of mouse visual acuity and temporal response) and computes from this both linear and quadratic features, each of which are summed, along with the binary running trace convolved with a learned temporal filter, and sent through a soft rectification (Fig. 7a). We trained these models on either the collective natural stimuli or the artificial stimuli to predict the extracted event trace. Whereas we found example neurons for

which this model worked extremely well (Fig. 7b and Supplementary Figs. 9–11), across the population only 2% of neurons were well fit by this model ($r>0.5$; 2% of neurons for natural stimuli and 1% of neurons for artificial stimuli; Fig. 7c), with median r values of ~0.2 (natural stimuli). Model performance was slightly higher in V1 than in the higher visual areas and showed little difference across Cre lines. It is also worth noting that there is a great deal of visually responsive activity that is not captured by these models (Supplementary Fig. 5). When comparing the models' performances across stimulus categories,

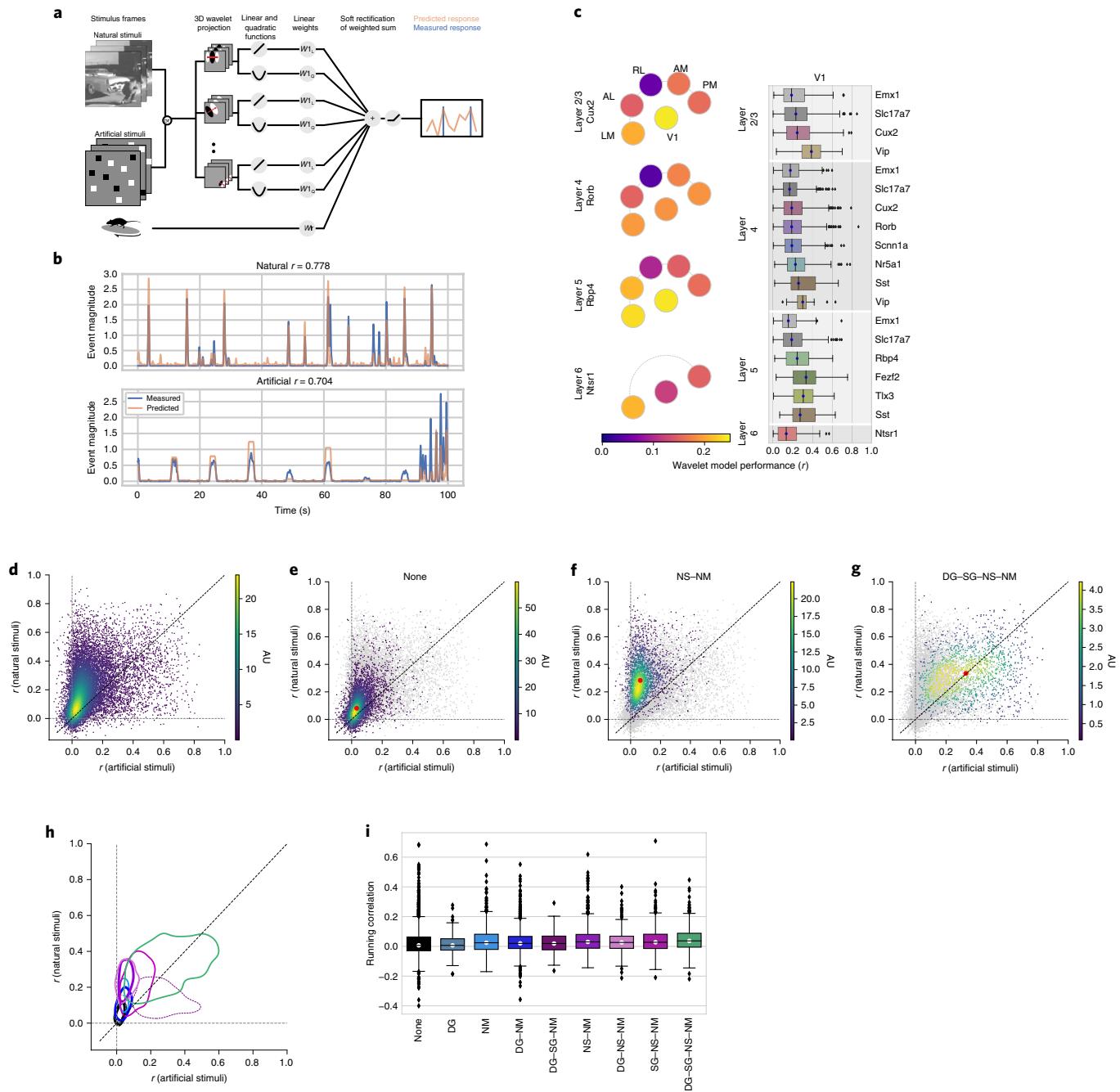


Fig. 7 | Validation of class labels by model performance. **a**, Schematic for the wavelet models. **b**, Example model performance for one neuron for both natural (top) and artificial (bottom) stimuli, where r is the Pearson correlation coefficient between the measured and predicted values. **c**, Pawplots and box plots of model performance, r , for wavelet models trained on natural stimuli. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range; points outside this range are shown as outliers. Only neurons imaged in all three sessions are included; $n=15,921$. See Extended Data Fig. 9 for sample sizes. For other cortical areas, see Supplementary Fig. 10. **d**, Density plot comparing r values for a model trained and tested on natural stimuli to r values for a model trained and tested on artificial stimuli; $n=15,921$. r is the Pearson correlation coefficient between the measured and predicted values. **e**, As in **d**, except that only neurons in the ‘none’ class are shown in the density ($n=5,566$); all other neurons are in gray. The red dot marks the median model performance for neurons in this class. **f**, As in **e** for the NS-NM class ($n=2,412$). **g**, As in **e** for the DG-SG-NS-NM class ($n=1,451$). **h**, Contours for the density of model performance, as in **e–g**, for all classes ($n=15,921$). The contours mark the boundaries of each class within which 66% of data points lie. Line widths reflect the number of neurons in each class, as provided in Fig. 6d. r is the Pearson correlation coefficient between the measured and predicted values. **i**, Box plots of running correlation for each class.

we found that the overall distribution of performance for models trained and tested with natural stimuli was higher than for the corresponding models for artificial stimuli (Fig. 7d), in line with previous reports^{10–15}. The running speed of the mouse did not add significant predictive power to the models, as most regression weights were

near zero, with the exception of Vip neurons in V1 (Supplementary Fig. 10). Similarly, incorporating pupil area and position had little effect, as was also the case when the quadratic weights were removed at the population level (Supplementary Fig. 10). Well-fit models tended to have sparser weights (Supplementary Fig. 11).

When comparing the model performance for the neurons in each of the classes defined through the clustering analysis, we found that these classes occupied spaces of model performance consistent with their definitions (Fig. 7e–h). The neurons classified as ‘none’ formed a relatively tight cluster and constituted the bulk of the density close to the origin (Fig. 7e). By definition, these neurons had the least response reliability for all stimuli (Fig. 6c) and were likewise the least predictable. Neurons in the NS–NM class showed high model performance for natural stimuli and low performance for artificial stimuli (Fig. 7f). Finally, neurons that reliably responded to all stimuli (DG–SG–NS–NM) showed a broad distribution of model performance, with the highest median performance, equally predicted by both artificial and natural stimuli (Fig. 7g). As running has been shown to influence neural activity in these data independently of visual stimuli (Fig. 5e), one might expect that the ‘none’ class would be composed largely of neurons that are strongly driven by running activity rather than visual stimuli. Instead, we found that the ‘none’ class had one of the smallest median correlations, overall, with the running speed of the mouse, whereas the DG–SG–NS–NM class had the largest correlation (Fig. 7i).

Discussion

Historically, visual physiology has been dominated by single-neuron electrophysiological recordings in which neurons were identified by responding to a test stimulus. The stimulus was then hand-tuned to elicit the strongest reliable response from that neuron, and the experiment proceeded using manipulations around this condition. Such studies discovered many characteristic response properties, namely, that visual responses can be characterized by combinations of linear filters with nonlinearities such as half-wave rectification, squaring, and response normalization⁷ or that neurons (in V1 at least) largely cluster into ‘simple’ and ‘complex’ cells. But these studies may have failed to capture the variability of responses, the breadth of features that will elicit a neural response, and the breadth of features that do not elicit a response. This results in systematic bias in the measurement of neurons and a confirmation bias regarding model assumptions. Recently, calcium imaging and denser electrophysiological recordings²⁷ have enabled large populations of neurons to be recorded simultaneously. Here we scaled calcium imaging, combining standard operating procedures with integrated engineering tools, to address some of the challenges of this difficult technique, as a means to create an unprecedented survey of 59,610 neurons in mouse visual cortex across 243 mice while using a standard and well-studied but diverse set of visual stimuli. This pipeline reduced critical experimental biases by separating quality control of data collection from response characterization. Such a survey is crucial for assessing the successes and shortcomings of contemporary models of the visual cortex.

By using standard noise and grating stimuli, we find many of the standard visual response features, including orientation selectivity, direction selectivity, and spatial receptive fields with opponent on and off subfields (Figs. 2 and 3). On the basis of responses to these stimuli, we observed functional differences in visual responses across cortical areas, layers, and transgenic Cre lines. In a new analysis of overall reliabilities to both artificial and naturalistic stimuli, we find classes of neurons responsive to different constellations of stimuli (Fig. 6). The different classes are largely intermingled and found in all of the cortical areas recorded here, suggesting a largely parallel organization²⁸. At the same time, the over-representation of classes responsive to natural movies and motion stimuli in areas AL, PM, and AM relative to the other classes (which are more responsive to spatial stimuli) is consistent with the assignment of these areas to the putative ‘dorsal’ or ‘motion’ stream²⁹. The lack of an inverse relationship, wherein spatial information is over-represented relative to motion in a putative ventral stream, likely reflects the fact that we were unable to image the putative ventral areas LI, POR,

and P within our cranial window. Area LM has previously been loosely associated with the ventral stream, but with evidence that it is more similar to V1 than other higher-order ventral areas^{9,29}, and our results appear consistent with the latter. Area RL has the largest proportion of neurons in the ‘none’ class, over 85%, in line with the very low percentage of responsive neurons (Fig. 2). It is possible that neurons in this area are specialized for visual features not probed here or that they show a greater degree of multimodality than in the other visual areas, integrating somatosensory and visual features³⁰.

One of the unique features of this dataset is that it includes a large number of different transgenic Cre lines for characterization that label specific populations of excitatory and inhibitory neurons. On a coarse scale, excitatory populations behave similarly; however, closer examination reveals distinct functional properties across Cre lines. For instance, neurons expressing Rorb, Scnn1a-Tg3, and Nr5a1, which label distinct layer 4 populations in V1, exhibit distinct spatial and temporal tuning properties (Fig. 3 and Supplementary Figs. 1 and 2), different degrees of running correlation (Fig. 5) and subtle differences in their class distribution (Fig. 6). These differences suggest that there are separate channels of feed-forward information. Similar differences between Fezf2 and Ntsr1 in V1, which label two distinct populations of corticofugal neurons found in layers 5 and 6, respectively, indicate distinct feedback channels from V1.

The Brain Observatory data also provide a broad survey of visually evoked responses for both Vip and Sst inhibitory Cre lines. Sst neurons are strongly driven by all visual stimuli used here, with the plurality belonging to the DG–SG–NS–NM class (Fig. 5f). Their responses to drifting gratings are particularly robust in that 94% of Sst neurons in V1 are responsive to drifting gratings and respond quite reliably across trials, far more than is seen in the other Cre lines (Extended Data Fig. 3 and Supplementary Fig. 5). Vip neurons, on the other hand, are largely unresponsive to, and even suppressed by, drifting gratings, with only 9% of Vip neurons in V1 labeled responsive. This extreme difference between these two populations is consistent with previous literature examining the size tuning of these interneurons and supports the disinhibitory circuit between them^{23,31,32}. Vip neurons, however, are very responsive to both natural scenes and natural movies, with the majority falling in the NS–NM class (Fig. 6f, Extended Data Figs. 6 and 7, and Supplementary Fig. 8), but show little selectivity to these stimuli, as their median lifetime sparseness is lower than that for both the Sst and excitatory neurons (Fig. 3f). Interestingly, receptive field mapping using locally sparse noise revealed that Vip neurons in V1 have remarkably large receptive field areas, larger than those of both Sst and excitatory neurons (Fig. 3f), in contrast to the smaller summation area for Vip neurons previously measured using windowed drifting gratings^{23,33}. This suggests that Vip neurons respond to small features over a large region of space. Further, both populations show strong running modulation: responses for both correlate more strongly with the mouse’s running speed than is seen for the excitatory populations (Fig. 5e), and a model based solely on the mouse’s running speed does a better job at predicting their activity than for the excitatory populations (Supplementary Fig. 7).

The true test of a model is its ability to predict arbitrary novel responses, in addition to responses from stimuli used for characterization. Even with the inclusion of running, our models predict responses in a minority of neurons (Fig. 7).

Neurons in the DG–SG–NS–NM class were well predicted, with values comparable to those found in primates^{7,11,34,35}, for both natural and artificial stimuli (Fig. 7g). On the basis of the way we chose our stimulus parameters, we expected that neurons with a strong ‘classical receptive field’ would be most likely to appear in this class. However, this class constitutes only 10% of the mouse visual cortex (Fig. 6d). Neurons in the NS–NM class show equally high prediction for natural stimuli, but poor prediction for artificial stimuli

(Fig. 7f). It is possible that these neurons could be classical neurons as well but are tuned for spatial or temporal frequencies that were not included in our stimulus set. As our stimulus parameters were chosen to match previous measurements of mouse acuity, this could suggest that the acuity in mice has been underestimated³⁶.

Remarkably, the largest class of neurons was the ‘none’ class, constituting neurons that did not respond reliably to any of the stimuli (34% of neurons). These neurons are the least likely to be described by classical receptive fields, as evidenced by their poor model performance for all stimuli (Fig. 7f). What, then, do these neurons do? It is possible that these neurons are visually driven but are responsive to highly sparse and specific natural features that may arise through hierarchical processing³⁷. Indeed, the field has a growing body of evidence that the rodent visual system exhibits sophisticated computations. For instance, neurons as early as V1 show visual responses to complex stimulus patterns³⁸. Alternatively, these neurons could be involved in non-visual computation, including behavioral responses such as reward timing and sequence learning³⁹, as well as modulation by multimodal sensory stimuli^{39,40} and motor signals^{24,26,41–43}. While we found little evidence that these neurons were correlated with a mouse’s running, recent work has found running to be among the least predictive of such motor signals⁴³.

We believe that the openly available Allen Brain Observatory provides an important foundational resource for the community. In addition to providing an experimental benchmark, these data serve as a testbed for theories and models. Already, these data have been used by other researchers to develop image processing methods^{44,45}, to examine stimulus encoding and decoding^{46–49} and to test models of cortical computations⁵⁰. Ultimately, we expect that these data will seed as many questions as they answer, fueling others to pursue both new analyses and further experiments to unravel how cortical circuits represent and transform sensory information.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-019-0550-9>.

Received: 23 May 2019; Accepted: 28 October 2019;

Published online: 16 December 2019

References

- Hubel, D. & Wiesel, T. Receptive fields of single neurones in the cat’s striate cortex. *J. Physiol.* **148**, 574–591 (1959).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiol.* **160**, 106–154 (1962).
- Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
- DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
- Olshausen, B. & Field, D. What is the other 85% of V1 doing? in *23 Problems in Systems Neuroscience* (eds. van Hemmen, J. & Sejnowski, T.) 182–211 (Oxford University Press, 2006).
- Masland, R. H. & Martin, P. R. The unsolved mystery of vision. *Curr. Biol.* **17**, R577–R582 (2007).
- Carandini, M. et al. Do we know what the early visual system does? *J. Neurosci.* **25**, 10577–10597 (2005).
- Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional specialization of mouse higher visual cortical areas. *Neuron* **72**, 1025–1039 (2011).
- Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040–1054 (2011).
- Fournier, J., Monier, C., Pananceau, M. & Frégnac, Y. Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. *Nat. Neurosci.* **14**, 1053–1060 (2011).
- David, S., Vinje, W. & Gallant, J. L. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.* **24**, 6991–7006 (2004).
- Talebi, V. & Baker, C. L. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. *J. Neurosci.* **32**, 1560–1576 (2012).
- Yeh, C.-I., Xing, D., Williams, P. & Shapley, R. Stimulus ensemble and cortical layer determine V1 spatial receptive fields. *Proc. Natl Acad. Sci. USA* **106**, 14652–14657 (2009).
- Sharpee, T. O. et al. Adaptive filtering enhances information transmission in visual cortex. *Nature* **439**, 936–942 (2006).
- Felsen, G., Touryan, J., Han, F. & Dan, Y. Cortical sensitivity to visual features in natural scenes. *PLoS Biol.* **3**, e342 (2005).
- Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
- Jewell, S., Hocking, T. D., Fearnhead, P. & Witten, D. Fast nonconvex deconvolution of calcium imaging data. *Biostatistics* <https://doi.org/10.1093/biostatistics/kxy083> (2019).
- Sun, W., Tan, Z., Mensh, B. D. & Ji, N. Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs. *Nat. Neurosci.* **19**, 308–315 (2016).
- Rolls, E. T. & Tovee, M. J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **73**, 713–726 (1995).
- Vinje, W. E. & Gallant, J. L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
- Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. Correlations and neuronal population information. *Annu. Rev. Neurosci.* **39**, 237–256 (2016).
- Dadarlat, M. C. & Stryker, M. P. Locomotion enhances neural encoding of visual stimuli in mouse V1. *J. Neurosci.* **37**, 3764–3775 (2017).
- Dipoppa, M. et al. Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron* **98**, 602–615 (2018).
- Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
- Polack, P. O., Friedman, J. & Golshani, P. Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nat. Neurosci.* **16**, 1331–1339 (2013).
- Saleem, A., Ayaz, A., Jeffery, K., Harris, K. & Carandini, M. Integration of visual motion and locomotion in mouse visual cortex. *Nat. Neurosci.* **16**, 1864–1869 (2013).
- Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232 (2017).
- Han, Y. et al. The logic of single-cell projections from visual cortex. *Nature* **556**, 51–56 (2018).
- Wang, Q., Sporns, O. & Burkhalter, A. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J. Neurosci.* **32**, 4386–4399 (2012).
- Olcese, U., Iurilli, G. & Medini, P. Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron* **79**, 579–593 (2013).
- Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013).
- Fu, Y. et al. A cortical circuit for gain control by behavioral state. *Cell* **156**, 1139–1152 (2014).
- Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J. & Scanziani, M. A neural circuit for spatial summation in visual cortex. *Nature* **490**, 226–231 (2012).
- McFarland, J. M., Cumming, B. G. & Butts, D. A. Variability and correlations in primary visual cortical neurons driven by fixational eye movements. *J. Neurosci.* **36**, 6225–6241 (2016).
- Vintch, B., Movshon, J. A. & Simoncelli, E. P. A convolutional subunit model for neuronal responses in macaque V1. *J. Neurosci.* **35**, 14829–14841 (2015).
- Dyballa, L., Hoseini, M. S., Dadarlat, M. C., Zucker, S. W. & Stryker, M. P. Flow stimuli reveal ecologically appropriate responses in mouse visual cortex. *Proc. Natl Acad. Sci. USA* **115**, 11304–11309 (2018).
- Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. & Lipson, H. Understanding neural networks through deep visualization. *Deep Learning Workshop, 31st International Conference on Machine Learning*, Lille, France. Preprint at arXiv <https://arxiv.org/abs/1506.06579> (2015).
- Palagin, G., Meyer, J. F. & Smirnakis, S. M. Complex visual motion representation in mouse area V1. *J. Neurosci.* **37**, 164–183 (2017).
- Bieler, M. et al. Rate and temporal coding convey multisensory information in primary sensory cortices. *eNeuro* **4**, ENEURO.0037-17.2017 (2017).
- Ibrahim, L. A. et al. Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. *Neuron* **89**, 1031–1045 (2016).
- Keller, G. B., Bonhoeffer, T. & Hübener, M. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* **74**, 809–815 (2012).
- Stringer, C. et al. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, eaav7893 (2019).

43. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
44. Petersen, A., Simon, N. & Witten, D. SCALPEL: extracting neurons from calcium imaging data. *Ann. Appl. Stat.* **12**, 2430–2456 (2018).
45. Sheintuch, L. et al. Tracking the same neurons across multiple days in Ca²⁺ imaging data. *Cell Rep.* **21**, 1102–1115 (2017).
46. Ellis, R. J. et al. High-accuracy decoding of complex visual scenes from neuronal calcium responses. Preprint at *bioRxiv* <https://doi.org/10.1101/271296> (2018).
47. Cai, L., Wu, B. & Ji, S. Neuronal activities in the mouse visual cortex predict patterns of sensory stimuli. *Neuroinformatics* **16**, 473–488 (2018).
48. Zylberberg, J. Untuned but not irrelevant: a role for untuned neurons in sensory information coding. Preprint at *bioRxiv* <https://doi.org/10.1101/134379> (2017).
49. Christensen, A. J. & Pillow, J. W. Running reduces firing but improves coding in rodent higher-order visual cortex. Preprint at *bioRxiv* <https://doi.org/10.1101/214007> (2017).
50. Sweeney, Y. & Clopath, C. Population coupling predicts the plasticity of stimulus responses in cortical circuits. Preprint at *bioRxiv* <https://doi.org/10.1101/265041> (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Transgenic mice. All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at the Allen Institute for Brain Science in compliance with NIH guidelines. Transgenic mouse lines were generated by using conventional and bacterial artificial chromosome (BAC) transgenic or knock-in strategies as previously described^{51,52}. External sources included Cre lines generated as part of the NIH Neuroscience Blueprint Cre Driver Network (<http://www.credrivermice.org>) and the GENSAT project (<http://gensat.org/>), as well as individual labs. In transgenic lines with regulatable versions of Cre, young adult tamoxifen-inducible mice (CreERT2) were treated with ~200 µl of tamoxifen solution (0.2 mg g⁻¹ body weight) via oral gavage once per day for five consecutive days to activate Cre recombinase.

We used the transgenic mouse line Ai93, in which GCaMP6f expression is dependent on the activity of both Cre recombinase and the tetracycline-controlled transactivator protein (tTA)⁵¹. Ai93 mice were first crossed with Camk2a-tTA mice, and the double-transgenic mice were then crossed with a Cre driver line. For some Cre drivers, we alternatively leveraged the TIGRE2.0 transgenic platform that combines the expression of tTA and GCaMP6f in a single reporter line (Ai148(TIT2L-GC6f-ICL-tTA2))⁵³.

In Cux2-CreERT2;Camk2a-tTA;Ai93(TITL-GCaMP6f) mice, expression is regulated by the tamoxifen-inducible *Cux2* promoter, the induction of which results in Cre-mediated expression of GCaMP6f predominantly in superficial cortical layers 2, 3 and 4 (ref. ⁵⁴, Supplementary Fig. 12 and Supplementary Table 1). Both Emx1-IRES-Cre;Camk2a-tTA;Ai93 and Slc17a7-IRES2-Cre;Camk2a-tTA;Ai93 are pan-excitatory lines and show expression throughout all cortical layers^{55,56}. Sst-IRES-Cre;Ai148 mice exhibit GCaMP6f in somatostatin-expressing neurons⁵⁷. Vip-IRES-Cre;Ai148 mice exhibit GCaMP6f in Vip-expressing cells by the endogenous promoter/enhancer elements of the vasoactive intestinal polypeptide locus⁵⁷. Rorb-IRES2-Cre;Camk2a-tTA;Ai93 mice exhibit GCaMP6f in excitatory neurons in cortical layer 4 (dense patches) and layers 5 and 6 (sparse)⁵⁵. Scnn1a-Tg3-Cre;Camk2a-tTA;Ai93 mice exhibit GCaMP6f in excitatory neurons in cortical layer 4 and in restricted areas within the cortex, in particular in primary sensory cortices. Nr5a1-Cre;Camk2a-tTA;Ai93 mice exhibit GCaMP6f in excitatory neurons in cortical layer 4 (ref. ⁵⁸). Rbp4-Cre;Camk2a-tTA;Ai93 mice exhibit GCaMP6f in excitatory neurons in cortical layer 5 (ref. ⁵⁹). Fezf2-CreER;Ai148 mice exhibit GCaMP6f in subcerebral projection neurons in layers 5 and 6 (ref. ⁶⁰). Tlx3-Cre_PL56;Ai148 mice exhibit GCaMP6f primarily restricted to IT corticostrial in layer 5 (ref. ⁵⁹). Ntsr1-Cre_GN220;Ai148 mice exhibit GCaMP6f in excitatory corticothalamic neurons in cortical layer 6 (ref. ⁶¹).

We maintained all mice on a reverse 12-h light cycle after surgery and throughout the duration of the experiment and performed all experiments during the dark cycle.

Cross-platform registration. To register data acquired between instruments and repeatedly target and record neurons in brain areas identified with intrinsic imaging, we developed a system for cross-platform registration (Supplementary Fig. 13).

Surgery. Transgenic mice expressing GCaMP6f were weaned and genotyped at approximately postnatal day (P) 21, and surgery was performed between P37 and P63. Surgical eligibility criteria included (1) weight ≥ 19.5 g (males) or ≥ 16.7 g (females); (2) normal behavior and activity; and (3) healthy appearance and posture. A preoperative injection of dexamethasone (3.2 mg kg⁻¹, subcutaneously) was administered 3 h before surgery. Mice were initially anesthetized with 5% isoflurane (1–3 min) and placed in a stereotaxic frame (model 1900, Kopf), and isoflurane levels were maintained at 1.5–2.5% for the duration of the surgery. An injection of carprofen (5–10 mg kg⁻¹, subcutaneously) was administered, an incision was made to remove skin and the exposed skull was leveled with respect to pitch (bregma-lambda level), roll and yaw (Supplementary Fig. 14).

Intrinsic imaging. A retinotopic map was created by using ISI to define visual area boundaries and target *in vivo* two-photon calcium imaging experiments to consistent retinotopic locations⁶². Mice were lightly anesthetized with 1–1.4% isoflurane administered with a somnosuite (model 715, Kent Scientific). Vital signs were monitored with a Physiosuite (model PS-MSTAT-RT, Kent Scientific). Eye drops (Laci-Lube Lubricant Eye Ointment, Refresh) were applied to maintain hydration and clarity of eye during anesthesia. Mice were headfixed for imaging normal to the cranial window.

The brain surface was illuminated with two independent LED lights, one green (peak λ = 527 nm and full width at half-maximum (FWHM) = 50 nm; Cree, C503B-GCN-CY0C0791) and one red (peak λ = 635 nm and FWHM = 20 nm; Avago Technologies, HLMP-EG08-Y2000), mounted on the optical lens. A pair of Nikon lenses (Nikon Nikkor 105mm f/2.8, Nikon Nikkor 35mm f/1.4) provided $\times 3.0$ magnification ($M = 105/35$) onto an Andor Zyla 5.5 10tap sCMOS camera. A bandpass filter (Semrock, FF01-630/92nm) was used to only record reflected red light onto the brain.

A 24-inch monitor was positioned 10 cm from the right eye. The monitor was rotated 30° relative to the animal's dorsoventral axis and tilted 70° off the horizon to ensure that the stimulus was perpendicular to the optic axis of the eye. The visual stimulus displayed comprised a 20° × 155° drifting bar containing a

checkerboard pattern, with individual square sizes measuring 25°, that alternated black and white as it moved across a mean-luminance gray background. The bar moved in each of the four cardinal directions ten times. The stimulus was warped spatially so that a spherical representation could be displayed on a flat monitor⁶³.

After defocusing from the surface vasculature (between 500 µm and 1,500 µm along the optical axis), up to ten independent ISI time series were acquired and used to measure the hemodynamic response to the visual stimulus. Averaged sign maps were produced from a minimum of three time series images for a combined minimum average of 30 stimulus sweeps in each direction⁶³.

The resulting ISI maps were automatically segmented by comparing the sign, location, size, and spatial relationships of the segmented areas against those compiled in an ISI-derived atlas of visual areas. Manual correction and editing of the segmentation were applied to correct errors. Finally, target maps were created to guide *in vivo* two-photon imaging location by using the retinotopic map for each visual area, restricted to within 10° of the center of gaze (Supplementary Fig. 15).

Habituation. After successful ISI mapping, mice spent 2 weeks being habituated to head fixation and visual stimulation. During the first week, mice were handled and headfixed for progressively longer durations, ranging from 5 to 10 min. During the second week, mice were headfixed and presented with visual stimuli, starting for 10 min and progressing to 50 min of visual stimuli by the end of the week, including all of the stimuli used during data collection. Mice received a single 60-min habituation session on the two-photon microscope with visual stimuli.

Two-photon *in vivo* calcium imaging. Calcium imaging was performed with a two-photon imaging instrument (either a Scientifica Vivoscope or a Nikon A1R MP+; the Nikon system was adapted to provide space to accommodate the running disc). Laser excitation was provided by a Ti:Sapphire laser (Chameleon Vision—Coherent) at 910 nm. Precompensation was set at ~10,000 fs². Movies were recorded at 30 Hz with resonant scanners over a 400-µm field of view (FOV). Temporal synchronization of all datastreams (calcium imaging, visual stimulation, body and eye tracking cameras) was achieved by recording all experimental clocks on a single NI PCI-6612 digital IO board at 100 kHz.

Mice were headfixed on top of a rotating disc and were free to walk at will. The disc was covered with a layer of removable foam (Super-Resilient Foam, 86375K242, McMaster). Data were initially obtained with the mouse eye centered both laterally and vertically on the stimulus monitor and positioned 15 cm from the monitor, with the monitor parallel to the mouse's body. Later, the monitor was moved to better fill the visual field. The normal distance of the monitor from the eye remained 15 cm, but the monitor center moved to a position 118.6 mm lateral, 86.2 mm anterior and 31.6 mm dorsal to the right eye.

An experiment container consisted of three 1-h imaging sessions at a given FOV during which mice passively observed three different stimuli. One imaging session was performed per mouse per day, for a maximum of 16 sessions per mouse.

On the first day of imaging at a new FOV, the ISI targeting map was used to select spatial coordinates. A comparison of superficial vessel patterns was used to verify the appropriate location by imaging over an FOV of ~800 µm using epifluorescence and blue light illumination. Once a region was selected, the objective was shielded from stray light coming from the stimulus monitor with opaque black tape. In two-photon imaging mode, the desired depth of imaging was set to record from a specific cortical depth. On subsequent imaging days, we returned to the same location by matching (1) the pattern of vessels in epifluorescence with (2) the pattern of vessels in two-photon imaging and (3) the pattern of cellular labeling in two-photon imaging at the previously recorded location.

Once a depth location was stabilized, a combination of PMT gain and laser power was selected to maximize laser power (on the basis of a look-up table against depth) and dynamic range while avoiding pixel saturation. The stimulus monitor was clamped in position, and the experiment began. Two-photon movies (512 × 512 pixels, 30 Hz), eye tracking (30 Hz) and a side-view full-body camera (30 Hz) were recorded. Recording sessions were interrupted and/or failed if any of the following was observed: (1) mouse stress, as shown by excessive secretion around the eye, nose bulge and/or abnormal posture; (2) excessive pixel saturation (>1,000 pixels), as reported in a continuously updated histogram; (3) loss of baseline intensity in excess of 20% caused by bleaching and/or loss of immersion water; and (4) hardware failures causing a loss of data integrity. Immersion water was occasionally supplemented while imaging by using a micropipette taped to the objective (Microfil, MF28G67-5WPI) and connected to a 5-ml syringe via extension tubing. At the end of each session, a z stack of images (±30 µm around the imaging site, 0.1-µm step) was collected to evaluate cortical anatomy and z drift during the course of the experiment. Experiments with z drift above 10 µm over the course of the entire session were excluded. In addition, for each FOV, a full-depth cortical z stack (~700-µm total depth, 5-µm step) was collected to document the imaging site location (Supplementary Figs. 16 and 17).

Detection of epileptic mice for exclusion. Before two-photon imaging, each mouse was screened for the presence of interictal events in two ways. First, on the habituation day on the two-photon rig, we collected a 5-min-long video on the surface of S1 by using the epifluorescence light path of the two-photon rig. For

each of these videos, we detected all calcium events present across the entire FOV and counted the number of events with a prominence superior to 10% $\Delta F/F$ and a width between 100 and 300 ms (ref. ⁶⁴). Second, a similar analysis was performed for all two-photon calcium videos collected. Except for inhibitory lines, any mouse that showed the presence of these large and fast events was reviewed and excluded from the pipeline. Inhibitory lines were excluded from this analysis as the neuronal labeling was too sparse to reliably distinguish these events from normal spontaneous activity.

Visual stimulation. Visual stimuli were generated with custom scripts written in PsychoPy^{65,66} (Peirce, 2007, 2008) and were displayed on an ASUS PA248Q LCD monitor, with $1,920 \times 1,200$ pixels. Stimuli were presented monocularly, and the monitor was positioned 15 cm from the eye and spanned $120^\circ \times 95^\circ$ of visual space. Each monitor was gamma corrected and had a mean luminance of 50 cd m^{-2} . To account for the close viewing angle, spherical warping was applied to all stimuli to ensure that the apparent size, speed, and spatial frequency were constant across the monitor as seen from the mouse's perspective.

Visual stimuli included drifting gratings, static gratings, locally sparse noise, natural scenes, and natural movies. These stimuli were distributed across three ~60-min imaging sessions (Fig. 1f). Session A included drifting gratings and natural movies 1 and 3. Session B included static gratings, natural scenes and natural movie 1. Session C included locally sparse noise and natural movies 1 and 2. The different stimuli were presented in segments of 5–13 min and interleaved. At least 5 min of spontaneous activity was recorded in each session.

The drifting gratings stimulus consisted of a full-field drifting sinusoidal grating at a single spatial frequency (0.04 cycles per degree) and contrast (80%). The grating was presented at eight different directions (separated by 45°) and at five temporal frequencies (1, 2, 4, 8 and 15 Hz). Each grating was presented for 2 s, followed by 1 s of mean-luminance gray. Each grating condition was presented 15 times. Trials were randomized, with blank sweeps (that is, mean-luminance gray instead of grating) presented approximately once every 20 trials.

The static gratings stimulus consisted of a full-field static sinusoidal grating at a single contrast (80%). The grating was presented at six different orientations (separated by 30°), five spatial frequencies (0.02, 0.04, 0.08, 0.16 and 0.32 cycles per degree), and four phases (0, 0.25, 0.5 and 0.75). The grating was presented for 0.25 s, with no inter-grating gray period. Each grating condition was presented ~50 times. Trials were randomized, with blank sweeps presented approximately once every 25 trials.

The natural scenes stimulus consisted of 118 natural images. Images were taken from the Berkeley Segmentation Dataset⁶⁷, the van Hateren Natural Image Dataset⁶⁸, and the McGill Calibrated Colour Image Database⁶⁹. The images were presented in grayscale and were contrast normalized and resized to $1,174 \times 918$ pixels. The images were presented for 0.25 s each with no inter-image gray period. Each image was presented ~50 times. Trials were randomized, with blank sweeps presented approximately once every 100 images.

Three natural movie clips were used from the opening scene of the movie *Touch of Evil* (Orson Wells, Universal Pictures, 1958). Natural movies 1 and 2 were both 30-s clips, while natural movie 3 was a 120-s clip. All clips had been contrast-normalized and were presented in grayscale at 30 frames per second. Each movie was presented ten times with no inter-trial gray period. Natural movie 1 was presented in each imaging session.

The locally sparse noise stimulus consisted of white and dark spots on a mean-luminance gray background. Each spot was square, with 4.65° on a side. Each frame had ~11 spots on the monitor, with no two spots within 23° of each other, and was presented for 0.25 s. Each of the 16×28 spot locations was occupied by white and black spots a variable number of times (mean = 115). For most of the collected data, this stimulus was adapted such that half of it used 4.65° spots while the other half used 9.3° spots, with an exclusion zone of 46.5° .

Serial two-photon tomography. Serial two-photon tomography was used to obtain a three-dimensional (3D) image volume of coronal brain images for each specimen. This 3D volume enables spatial registration of each specimen's associated ISI and optical physiology data to the Allen Mouse Common Coordinate Framework (CCF). Methods for this procedure have been described in detail in whitepapers associated with the Allen Mouse Brain Connectivity Atlas and in Oh et al.⁷⁰.

Post-mortem assessment of brain structure. Morphological and structural analysis of each brain was performed following collection of the two-photon serial imaging (TissueCyté) dataset (Supplementary Fig. 18).

The following characteristics warranted an automatic failure of all associated data: (1) an abnormal GCaMP6 expression pattern; (2) necrotic brain tissue; (3) compression of the contralateral cortex that resulted in disruption to the cortical laminar structure; and (4) compression of the ipsilateral cortex or adjacent to the cranial window.

The following characteristics were further reviewed and may have resulted in failure of the associated data: (1) compression of the contralateral cortex due to a skull growth; (2) excessive compression of the cortex underneath the cranial window; and (3) abnormal or enlarged ventricles.

Image processing. For each two-photon imaging session, the image processing pipeline performed: (1) spatial or temporal calibration specific to a particular microscope, (2) motion correction, (3) image normalization to minimize confounding random variations between sessions, (4) segmentation of connected shapes and (5) classification of soma-like shapes from remaining clutter (Supplementary Figs. 19 and 20).

The motion correction algorithm relied on phase correlation and only corrected for rigid translational errors. It performed the following steps. Each movie was partitioned into 400 consecutive frame blocks, representing 13.3 s of video. Each block was registered iteratively to its own average three times (Supplementary Fig. 20a,b). A second stage of registration integrated the periodic average frames themselves into a single global average frame through six additional iterations (Supplementary Fig. 20c). The global average frame served as the reference image for the final resampling of every raw frame in the video (Supplementary Fig. 20d).

Each 13.3-s block was used to generate normalized periodic averages with the following steps. First, we subtracted the mean from the maximum projection to retain pixels from active cells (Supplementary Fig. 20e–g). To select objects of the right size during segmentation, we convolved all periodic normalized averages with a 3×3 median filter and a 47×47 high-pass mean filter. We then normalized the histogram of all resulting frames (Supplementary Fig. 20g,h).

All normalized periodic averages were then segmented with an adaptive threshold filter to create an initial estimate of binarized ROI masks of unconnected components (Supplementary Fig. 20i). Given the lower expression of GCaMP6f in cell nuclei, good detections from somata tended to show bright outlines and dark interiors. We then performed a succession of morphological operations to fill closed holes and concave shapes (Supplementary Fig. 20j,k).

These initial ROI masks included shapes from multiple periods that were actually from a single cell. To further reduce the number of masks to putative individual cell somas, we computed a feature vector from each mask that included morphological attributes such as location, area, perimeter, and compactness, among others (Supplementary Fig. 20l). A battery of heuristic decisions applied on these attributes allowed us to combine, eliminate, or maintain an ROI (Supplementary Fig. 20l,m). A final discrimination step, using a binary relevance classifier fed by experimental metadata (for example, Cre line and imaging depth) along with the previous morphological features, further filtered the global masks into the final ROIs used for trace extraction.

Targeting refinement for putative RL neurons. In all experiments, the center of the two-photon FOV was aimed close to the retinotopic center of the targeted visual region, as mapped by ISI. Retinotopic mapping of RL commonly yielded retinotopic centers close to the boundary between RL and somatosensory cortex. Consequently, for some RL experiments, the FOV spanned the boundary between the visual and somatosensory cortex. All RL experiments were reviewed with a semiautomated process (Supplementary Fig. 21), and ROIs that were deemed to lie outside putative visual cortex boundaries (~25%) were excluded from further analysis.

Neuropil subtraction. To correct for contamination of the ROI calcium traces by surrounding neuropil, we modeled the measured fluorescence trace of each cell as $F_M = F_C + rF_N$, where F_M is the measured fluorescence trace, F_C is the unknown true ROI fluorescence trace, F_N is the fluorescence of the surrounding neuropil and r is the contamination ratio. To estimate the contamination ratio for each ROI, we selected the value of r that minimized the cross-validated error

$$E = \sum_t |F_C - F_M + rF_N|^2$$

over four folds. We computed the error over each fold with a fixed value of r , for a range of r values. For each fold, F_C was computed by minimizing

$$C = \sum_t |F_C - F_M + rF_N|^2 + \lambda |LF_C|^2$$

where L is the discrete first derivative (to enforce smoothness of F_C) and λ is a parameter set to 0.05. After determining r , we computed the true trace as $F_C = F_M - rF_N$, which was used in all subsequent analysis (Supplementary Fig. 22).

Demixing traces from overlapping ROIs. We demixed the activity of all recorded ROIs by using a model where every ROI had a trace distributed in some spatially heterogeneous, time-dependent fashion

$$F_{it} = \sum_k W_{kit} T_{kt}$$

where W is a tensor containing time-dependent weighted masks: W_{kit} measures how much of neuron k 's fluorescence is contained in pixel i at time t . T_{kt} is the fluorescence trace of neuron k at time t —this is what we want to estimate. F_{it} is the recorded fluorescence in pixel i at time t .

This model applied to all ROIs before filtering for somas. We filtered out duplicates (defined as two ROIs with >70% overlap) and ROIs that were the union

of others (any ROI where the union of any other two ROIs accounted for 70% of its area) before demixing and applied the remaining filtering criteria afterwards. Projecting the movie F onto the binary masks, A , reduced the dimensionality of the problem from 512×512 pixels to the number of ROIs

$$\sum_i A_{ki} F_{it} = \sum_{k,i} A_{ki} W_{kit} T_{kt}$$

where A_{ki} is 1 if pixel i is in ROI k and 0 otherwise—these are the masks from segmentation, after filtering. At time point t , this yields the linear regression

$$AF(t) = (AW^T(t))T(t)$$

where we estimated the weighted masks W by projection of the recorded fluorescence F onto the binary masks A . On every frame t , we computed the linear least-squares solution \hat{T} to extract each ROI's trace value.

It was possible for ROIs to have negative or zero demixed traces \hat{T} . This occurred if there were unions (one ROI composed of two neurons) or duplicates (two ROIs in the same location with approximately the same shape) that our initial detection missed. If this occurred, those ROIs and any that overlapped them were removed from the experiment. This led to the loss of ~1% of ROIs (Supplementary Fig. 22).

ROI matching. The FOV for each session and the segmented ROI masks were registered to each other with an affine transformation. To map cells, a bipartite graph matching algorithm was used to find the correspondence of cells between sessions A and B, A and C, and B and C. The algorithm took cells in the pairwise experiments as nodes and the degree of spatial overlapping and closeness between cells as edge weight. By maximizing the summed weights of edges, the bipartite matching algorithm found the best matching between cells. Finally, a label combination process was applied to the matching results of sessions A and B, A and C, and B and C, producing a unified label for all three experiments.

ΔF/F. To calculate the $\Delta F/F$ for each fluorescence trace, we first calculated baseline fluorescence by using a median filter of width 5,401 samples (180 s). We then calculated the change in fluorescence relative to baseline fluorescence (ΔF), divided by baseline fluorescence (F). To prevent very small or negative baseline fluorescence, we set the baseline as the maximum of the median filter-estimated baseline and the s.d. of the estimated noise of the fluorescence trace.

L0-penalized event detection. We used the L0-penalized method of Jewell et al. for event detection^{17,71}. We refer to this as ‘event’ detection because low firing rate activity is difficult to detect. For each $\Delta F/F$ trace, we removed slow timescale shifts in the fluorescence with a median filter of width 101 samples (3.3 s). We then applied the L0-penalized algorithm to the corrected $\Delta F/F$ trace. The L0 algorithm has two hyperparameters: gamma and lambda. Gamma corresponds to the decay constant of the calcium indicator. We set gamma to be the decay constant obtained from jointly recorded optical and electrophysiology with the same genetic background and calcium indicator. Time constants can be found at https://github.com/AllenInstitute/visual_coding_2p_analysis/blob/master/visual_coding_2p_analysis/l0_analysis.py. Supplementary Fig. 23 shows the extracted linear kernels for Emx1-Ai93 and Cux2-Ai93 from which gamma has been extracted by fitting the fluorescence decay with a single exponential. The rise time, amplitude and shape of the extracted linear kernels are mainly a function of the genetically encoded calcium indicator (GCaMP6f) and appear to be largely independent of the specific promoter driving expression.

To estimate lambda, which controls the strength of the L0 penalty, we estimated the s.d. of the trace. We set lambda by using bisection to minimize the number of events smaller than 2 s.d. of the noise, while retaining at least one recovered event. We chose 2 s.d. by maximizing the hit–miss rate on eight hand-annotated traces during locally sparse noise stimulation. These traces were uniformly sampled from a distribution of signal-to-noise ratio for $\Delta F/F$ traces. The noise level was computed as the robust s.d. (1.4826 times the median absolute deviation), and the signal level was the median $\Delta F/F$ after thresholding at the robust s.d.

To assess how the events detected with the above procedure related to actual spikes, we performed event detection on the fluorescence of cells that were imaged simultaneously with loose patch recordings. Because the true spike train is known for these data, we computed the expected probability of detecting an event, as well as the expected event magnitude, as a function of the number of spikes observed in a set of detection windows relevant to the pipeline data analyses (for example, static gratings, natural scenes and locally sparse noise templates are presented for 0.25 s each) (Supplementary Fig. 23).

Analysis. All analysis was performed with custom Python scripts by using NumPy⁷², SciPy⁷³, Pandas⁷⁴ and Matplotlib⁷⁵.

Direction selectivity was computed from mean responses to drifting gratings, at the cell's preferred temporal frequency,

$$DSI = \frac{R_{pref} - R_{null}}{R_{pref} + R_{null}}$$

where R_{pref} is a cell's mean response in its preferred direction and R_{null} is its mean response to the opposite direction.

The temporal frequency tuning, at the preferred direction, was fit using either an exponential curve (for the highest and lowest peak temporal frequency) or a Gaussian curve (other values). The reported preferred temporal frequency was taken from these fits. The same was done for spatial frequency tuning, fit at the cell's preferred orientation and phase in response to the static gratings. In both cases, if a fit could not converge, a preferred frequency was not reported.

Spatial receptive fields were computed from locally sparse noise, in two stages. First, we determined whether a cell had a receptive field by a statistical test, described in the “Statistics” section below. Second, we computed the receptive field itself. A second statistical test, described in the “Statistics” section below, was used to determine inclusion of each spot in the receptive field. Determining statistical significance was a less common but important step necessary because of the size of the dataset.

If a neuron was found to have a receptive field, the spots that were identified for receptive field membership were fit with a two-dimensional Gaussian distribution, with orientation, azimuth/elevation and x/y s.d. serving as degrees of freedom for the optimization. On and Off subregions (that is, white and black spots) were fit separately. Subregion area was defined as the 1.5 s.d. ellipse under this fit Gaussian, measured in units of squared visual degrees. Up to two On and Off subregions were fit. The total area of the receptive field was computed as the sum of all subregion areas, correcting for overlap.

Lifetime sparseness was computed by using the definition in Vinje and Gallant²⁰

$$S_L = \frac{1 - \frac{1}{N} \left(\sum_i r_i \right)^2}{1 - \frac{1}{N}}$$

where N is the number of stimulus conditions and r_i is the response of the neuron to stimulus condition i averaged across trials. Population sparseness was computed with the same metric, but where N was the number of neurons and r_i was the average response vector of neuron i to all stimulus conditions.

For each stimulus, we computed CC_{max} , the expected correlation between the sample trial-averaged response and the true (unmeasured) mean response. It provides an upper bound on the expected performance of any model that predicts response from the given stimulus trial structure. We followed the computation from Schoppe et al.⁷⁶

$$\frac{1}{CC_{max}} = \sqrt{1 - \frac{1}{N} \left(\frac{(1 - \frac{1}{N}) \times \sum_{n=1}^N Var(R_n)}{Var(\sum_{n=1}^N R_n) - \sum_{n=1}^N Var(R_n)} \right)}$$

where N is the number of trials and R_n is the time series of the response on the n th trial. For R_n , we used the trace of extracted event magnitudes at 30 Hz, smoothed with a Gaussian window of width 0.25 s.

We computed ‘noise’ and ‘signal’ correlations in the population responses. Signal correlations were computed as the Pearson correlation between the trial-averaged stimulus responses of pairs of neurons. To prevent trial-by-trial fluctuations from contaminating our signal correlation estimates, we separated the trials for each stimulus into two subsets and calculated the correlation between the trial-averaged responses with each subset of trials. We averaged the signal correlations over 100 random splits of the trials. Noise correlations were computed as the Pearson correlation of the single-trial stimulus responses for a pair of neurons and a given stimulus and then averaged over stimuli. For natural movies, we computed the noise and signal correlations of the binned event counts in non-overlapping ten-frame windows. We computed ‘spontaneous correlations’ as the Pearson correlation of the detected event trains during the periods of spontaneous activity recording.

Decoding. We used k -nearest-neighbors classifiers to decode the visual stimulus identity (for example, the natural scene number, within the natural scenes responses) from the population vector of single-trial responses, by using the correlation distance between response vectors. We report the performance on the heldout data from fivefold cross-validation. On each cross-validation fold, we performed an inner-round of twofold cross-validation to choose the number of neighbors from eight logarithmically spaced options (1, 2, 4, 7, 14 and 27).

Categorical regression model for trial responses to drifting gratings. We fit linear ridge regression models for the trial-averaged responses (events summed during each stimulus presentation). The response for trial t , R_t , is governed by the following equation

$$R_t = \sum_s w_s I_s^t$$

where I_s^t is the characteristic function for stimulus condition s during trial t . I_s^t is equal to 1 when the stimulus condition is equal to s during trial t and 0 otherwise, and w_s is the weight for stimulus condition s (it gives the response of the neuron to stimulus s).

We fit two separate models, one for which the stimulus conditions enumerate the different values of the drifting grating (that is, orientations and temporal frequencies, including the blank sweep; 41 total conditions) and another for which each stimulus condition occurred in pairs, one during running and one when the animal was stationary. On each stimulus trial, we classified the locomotion as running or stationary by using a Gaussian mixture model with a Dirichlet process prior for the number of components. Stationary trials were identified by the component with the smallest variance among those with mean speed $< 1 \text{ cm s}^{-1}$ (if any existed). We used stimulus conditions with at least five repetitions in each behavioral state and used the same number of trials for each stimulus condition in each behavioral state.

We regressed the summed trial against the combination of stimulus condition and behavioral state (for example, 180 degrees, 4 Hz and running). The regularization weight was chosen by leave-one-out cross-validation on the training data. We also regressed against just the locomotion state, binning the activity and running speed into pseudo-trials of the same length as the drifting grating trials. We measured model performance by the correlation of the predictions and data on heldout trials (fivefold cross-validation).

Regression models for mouse running speed. We performed a polynomial regression of each neuron's activity against running speed. To do this, we rank-sorted the running speed and binned it into 900-point bins. All speeds between -1 cm s^{-1} and 1 cm s^{-1} were labeled stationary. We summed each neuron's events in the same speed bins to compute the speed tuning. We then fit a polynomial regression for the speed tuning with fivefold cross-validation. On each training fold, we performed an inner twofold cross-validation to select the polynomial degree between 1 and 4. We used ridge regression with leave-one-out cross-validation to choose the regularization parameter between 0.5 and 100.

3D Gabor wavelet model for temporal responses. Each neuron was modeled as a sparse linear combination of linear and quadratic basis functions, similarly to other approaches^{77–79}. We used a pyramid of 3D Gabor wavelet filters that tiled the stimulus at multiple scales, directions, and temporal frequencies (Fig. 6a). The filters are defined by

$$f(x, y, t : \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + y'^2 + \gamma t^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi t\right)\right)$$

where

$$x' = x \cos\theta + y \sin\theta$$

$$y' = -x \sin\theta + y \cos\theta$$

λ controls spatial frequency, θ orientation, ψ temporal frequency, σ the Gaussian envelope, and γ the Gaussian envelope in time. This linear basis forms a reasonably tight frame. The parameters that generate the set of filters were adapted and scaled to the tuning properties of mouse visual cortex. We estimated weights for ten time lags for each basis function to enable fitting of the temporal kernel. The weighted sum of the basis functions was passed to a parameterized soft-plus nonlinearity. The filters were temporally convolved with the stimulus. The output of each filter was z scored before fitting with threshold gradient descent.

$$H_i(t) = \sum_{x,y} \sum_{\tau} f_i(x, y, \tau) S(x, y, t - \tau)$$

The model is technically a generalized linear model (where the linear model is built by considering linear combinations of the features $H_i(\tau)$ and $H_i^2(\tau)$, along with a temporal filter for the running signal of the animal, $r(t)$, with a parameterized soft-plus output). The weights w_i were fit to the data by using threshold gradient descent and the Poisson negative log-likelihood cost function, with rate

$$\hat{R}(t) = \log\left(1 + \exp\left(k \sum_i \sum_{\tau} H_i(t - \tau) w_i^l(\tau) + H_i^2(t - \tau) w_i^q(\tau) + r(t - \tau) w_i^r(\tau)\right)\right)$$

This model, with a quadratic dependence on the stimulus $H_i^2(\tau)$, is akin to a regularized STA/STC analysis, adapted to fit the full spatiotemporal receptive field by using stimuli from the dataset.

We estimated a sparse combination of basis functions for each neuron by using a variant of threshold gradient descent⁸⁰. In threshold gradient descent, only basis functions whose gradients have magnitudes larger than some threshold, t , of the largest gradient magnitude have their weights updated. All weights start at 0, and the descent is terminated by using early stopping. The threshold parameter, which can range from 0 to 1, controls the sparsity of the solution. We used a threshold value of 0.8.

We modified the threshold gradient descent algorithm in three ways. First, we updated the weights at all time lags for any basis function over the threshold, allowing the temporal kernel to be smooth. Second, at each iteration, any basis

function whose gradient exceeded the threshold had its weight added to the 'active set', which was maintained over the optimization, and then all weights in the active set were updated, preventing oscillations. Third, we used an adaptive step size. The step size increased by a factor of 1.2 at each iteration if generalization to the stopping set improved and decreased by a factor of 0.5 if generalization worsened⁸¹.

We used a nested sixfold cross-validation framework. We split the data into six sets each containing many 50-sample-long continuous blocks from throughout the dataset. A model was trained by starting with five separate models, each trained on a different combination of four of the five training sets, with the remaining set functioning as the stopping set. The five models were averaged together to make predictions on the test set. Reported model performance is the average on the test set across the six folds. Separate models were fit for the natural stimuli and the artificial stimuli. The weights for these models were sparse, and for all models fewer than 20% of the basis functions had non-zero weight values (Supplementary Fig. 11). The number of parameters for the model was 517,451.

Pupil position and area were measured (Supplementary Fig. 24), and these were incorporated into some of the models. These corrections had little effect on model performance (Supplementary Fig. 10).

We show examples of this model on four neurons in Supplementary Fig. 9.

Clustering of reliabilities. We performed a clustering analysis using the reliabilities by stimulus for each cell (defined as the percentage of responsive trials to the cell's preferred stimulus condition). We did not include locally sparse noise in this analysis. We combined the reliabilities for natural movies by taking the maximum reliability over the different natural movie stimuli. We performed this analysis with two different inclusion criteria. For criterion 1, we included all cells that appeared in both sessions A and B. For criterion 2, we included all cells that appeared in all three sessions, A, B and C. This resulted in a set of four reliabilities for each cell (for the drifting gratings, static gratings, natural movies, and natural scenes). We performed a Gaussian mixture model clustering on these reliabilities for cluster numbers 1 to 50, by using the average Bayesian information criterion on heldout data with fourfold cross-validation to select the optimal number of clusters. Once the optimal model was selected, we defined a threshold for responsiveness by selecting the cluster with the lowest mean reliability over all stimuli. We set the threshold to be the maximum reliability plus 1 s.d. over the reliabilities for this cluster. By using this threshold, we identified each cluster according to its profile of responsiveness (that is, whether it responded to drifting gratings, etc.), defining these profiles as 'classes'. For each cell, we predicted cluster membership by using the optimal model and then the class membership by using the threshold. We repeated this process 100 times to estimate the robustness of the clustering and derive uncertainties for the number of cells belonging to each class.

Statistics. No statistical methods were used to predetermine sample sizes per location, but our sample sizes are similar to those reported in previous publications^{8,9}. Data collection and analyses were not performed with blinding to the conditions of the experiments, as there was a single experimental condition for all acquired data. Within each transgenic Cre line, mice were randomly assigned to data collection to sample different areas and imaging depths. Stimulus conditions for gratings and natural scenes were presented in a randomized order within each epoch, as described. No other randomization was used as there were fixed experimental conditions for all other aspects of the dataset. Additional research design information can be found in the Nature Research Reporting Summary accompanying this study.

Test for significance of a receptive field map. We performed a chi-squared test to assess whether there was a significant response at each location to the locally sparse noise stimulus. For each location, we considered a 7×7 grid of locally sparse noise pixels centered on that location. The null model for this test was defined by assuming that a neuron lacking a receptive field has equal probability of producing a response regardless of the location and luminance (that is, black or white) of the spots displayed on the screen on any given trial. A neuron has a receptive field if there is a deviation beyond chance based on the null distribution. Chi-squared tests for independence were performed for each neuron and for each location by using the number of responses to quantify the dependence of responsive trials on the stimulus.

An assumption of the chi-squared test is that the response of the neuron on a given trial can only be attributed to a single spot; that is, only a single stimulus spot is presented on each trial. Although multiple non-gray spots appeared on the screen during each trial, the exclusion region of the locally sparse noise stimulus prevented two non-gray pixels within a 23° radius (for the 4.65° spot size) or 46° radius (for the 9.3° spot size) of each other from being presented on the same trial. Leveraging this structure in the stimulus, chi-squared tests were performed on patches in a visual space small enough to ensure that two or more non-gray pixels were rarely presented on the same trial, but large enough to ensure that the patch completely contained the receptive field for the test to detect the dependence of neuron responses on spot locations. We chose $32.2^\circ \times 32.2^\circ$ patches for 4.65° spots and $64.4^\circ \times 64.4^\circ$ patches for the 9.3° spot locally sparse noise (that is, a 7×7 grid of spot locations in each case). For each neuron, multiple chi-squared tests were performed on such patches to tile the entire stimulus monitor and the P values

from these tests were then corrected with the Šidák method to account for multiple comparisons. If the *P* value for any patch on the stimulus monitor was significant (*P*<0.05) after multiple-comparison correction, the neuron was considered to have a receptive field.

Test for inclusion of locally sparse noise spots in a receptive field. The receptive field was computed by using an event-triggered average. Because more than one stimulus spot was present during a given trial, it is not possible to infer the stimulus–response relationship between spot locations and responses on a per-trial basis. Therefore, a statistically significant co-occurrence of spot presentation and responses across trials defined the inclusion criteria for membership of a stimulus spot in the receptive field. To begin, the stimulus was convolved with a spatial Gaussian (4.65° per sigma), to allow pooling of contributions to responses from nearby spots. A *P* value was computed for each spot (black and white separately) by constructing a null distribution for the number of trials that a spot was present during responsive trials. This per-spot null distribution was estimated by shuffling the identity of the responsive trials (*n*=10,000 shuffles). Statistical outliers were identified by computing a *P* value for each spot relative to its null distribution. These *P* values were corrected for false discoveries by using the Šidák multiple-comparisons correction and thresholded at *P*=0.05 to identify receptive field membership.

Comparison of single-cell response metric distributions. To compare the distributions of single-cell response metrics across areas, layers and Cre lines, we used a Kolmogorov–Smirnov test with a Bonferroni correction for the number of comparisons, defined as the number of other distributions to which we were comparing, for example, for area-wise comparison of the Cux2 line, there were six areas in total and thus five comparisons for each area (first row of Supplementary Fig. 2). The Kolmogorov–Smirnov test was chosen as it does not assume a normal distribution nor equal variance.

Data product. The Allen Brain Observatory Visual Coding dataset is publicly available and accessible via a dedicated web portal (<http://observatory.brain-map.org/visualcoding/>), with a custom Python-based application programming interface (API), the AllenSDK (<https://github.com/AllenInstitute/AllenSDK>). Data from each imaging session are contained within a NWB (Neurodata Without Borders) file⁸².

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

This is an openly available dataset, accessible via a dedicated web portal (<http://observatory.brain-map.org/visualcoding>), and a Python-based API, the AllenSDK (<https://alleninstitute.github.io/AllenSDK>).

Code availability

Code for analyses presented in this paper is available at https://github.com/alleninstitute/visual_coding_2p_analysis.

References

51. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
52. Madisen, L. et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
53. Daigle, T. L. et al. A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480 (2018).
54. Franco, S. J. et al. Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science* **337**, 746–749 (2012).
55. Harris, J. A. et al. Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front. Neural Circuits* **8**, 1–16 (2014).
56. Gorski, J. A. et al. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
57. Taniguchi, H. et al. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
58. Dhillon, H. et al. Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron* **49**, 191–203 (2006).
59. Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC Cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**, 1368–1383 (2013).
60. Guo, C. et al. Fezf2 expression identifies a multipotent progenitor for neocortical projection neurons, astrocytes, and oligodendrocytes. *Neuron* **80**, 1167–1174 (2013).
61. Gong, S. et al. Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
62. Kalatsky, V. A. & Stryker, M. P. New paradigm for optical imaging: temporally encoded maps of intrinsic signal. *Neuron* **38**, 529–545 (2003).
63. Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and areal organization of mouse visual cortex. *J. Neurosci.* **34**, 12587–12600 (2014).
64. Steinmetz, N. A. et al. Aberrant cortical activity in multiple GCaMP6-expressing transgenic mouse lines. *eNeuro* **4**, ENEURO.0207-17.2017 (2017).
65. Peirce, J. W. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* **2**, 10 (2008).
66. Peirce, J. W. PsychoPy-Psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
67. Martin, D., Fowlkes, C., Tal, D. & Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proc. Eighth IEEE Int. Conf. Comput. Vision 2001* **2**, 416–423 (2001).
68. van Hateren, J. H. & Van Der Schaaf, A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci.* **265**, 359–366 (1998).
69. Olmos, A. & Kingdom, F. A. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* **33**, 1463–1473 (2004).
70. Oh, S. W. et al. A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
71. Jewell, S. & Witten, D. Exact spike train inference via ℓ_0 optimization. *Ann. Appl. Stat.* **12**, 2457–2482 (2018).
72. Oiphant, T. E. *A Guide to NumPy* (Trelgol Publishing, 2006).
73. Jones, E., Oliphant, T., Peterson, P., et al. SciPy: Open source scientific tools for Python. <http://www.scipy.org/> (2001).
74. McKinney, W. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.), 51–56 (2010).
75. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* <https://doi.org/10.1109/MCSE.2007.55> (2017).
76. Schoppe, O., Harper, N. S., Willmore, B. D. B., King, A. J. & Schnupp, J. W. H. Measuring the performance of neural models. *Front. Comput. Neurosci.* **10**, 1–11 (2016).
77. Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).
78. Nishimoto, S. et al. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
79. Willmore, B. D. B., Prenger, R. J. & Gallant, J. L. Neural representation of natural images in visual area V2. *J. Neurosci.* **30**, 2102–2114 (2010).
80. Friedman, J. H. & Popescu, B. E. Predictive learning via rule ensembles. *Ann. Appl. Stat.* **2**, 916–954 (2008).
81. Riedmiller, M. & Braun, H. RPROP—a fast adaptive learning algorithm. *Proceedings of the International Symposium on Computer and Information Science VII* (1992).
82. Teeters, J. L. et al. Neurodata Without Borders: creating a common data format for neurophysiology. *Neuron* **88**, 629–634 (2015).

Acknowledgements

We thank the Animal Care, Transgenic Colony Management, and Lab Animal Services for mouse husbandry. We thank Z. J. Huang of Cold Spring Harbor Laboratory for use of the Fezf2-CreER line. We thank D. Denman, J. Siegle, Y. Billeh, and A. Arkhipov for critical feedback on the manuscript. This work was supported by the Allen Institute and in part by NSF DMS-1514743 (E.S.B.), the Falconwood Foundation (C.K.), the Center for Brains, Minds & Machines funded by NSF Science and Technology Center Award CCF-1231216 (C.K.), the Natural Sciences and Engineering Research Council of Canada (S.J.), NIH grant DP5OD009145 (D.W.), NSF CAREER Award DMS-1252624 (D.W.), Simons Investigator Award in Mathematical Modeling of Living Systems (D.W.), and NIH grant 1R01EB026908-01 (M.A.B., D.W.). We thank A. Jones for providing the critical environment that enabled our large-scale team effort. We thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.

Author contributions

S.E.J.d.V., M.A.B., K.R., M.G., T.K., S.M., S.O., J.W., H.Z., C.D., L.N., A.B., J.W.P., R.C.R. and C.K. conceived of and designed the experiment. J.A.L., T.K., P.H., A.L., C.S., D.S. and C.F. built and maintained the hardware. S.E.J.d.V., J.A.L., M.A.B., G.K.O., D.F., N.C., L.K., W.W., D.W., R.V., C.B., B.B., T.D., J.G., T.G., S.J., N.K., C.L., F. Lee, F. Long, J.P., N.S., D.M.W., J.Z. and L.N. developed algorithms and software, including the SDK and website. K.R., N. Berbesque, N. Bowles, S.C., L.C., A.C., S.D.C., M.E., N.G., F.G., R.H., L.H., U.K., J.L., J.D.L., R.L., E.L., L.L., J.L., K.M., T.N., M.R., S.S., C.W. and A.W. collected data. J.A.L., P.A.G., S.E.J.d.V. and M.A.B. supervised the work. S.E.J.d.V., J.A.L., M.A.B., G.K.O., M.O., N.C., P.L., D.M., J.S., E.S.B. and R.V. analyzed data. C.T. and W.W. provided project administration. S.E.J.d.V., J.A.L. and M.A.B. wrote the paper with input from P.A.G., G.K.O., M.O., N.C., P.L., D.M., R.C.R., M.G. and C.K.

Competing interests

The authors declare no competing interests.

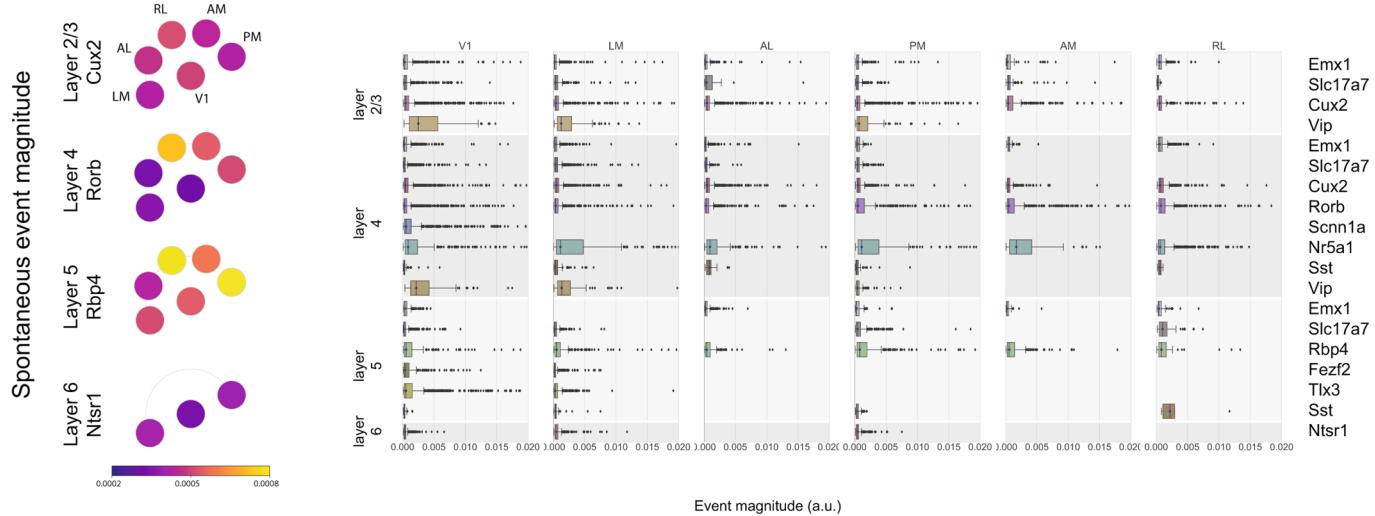
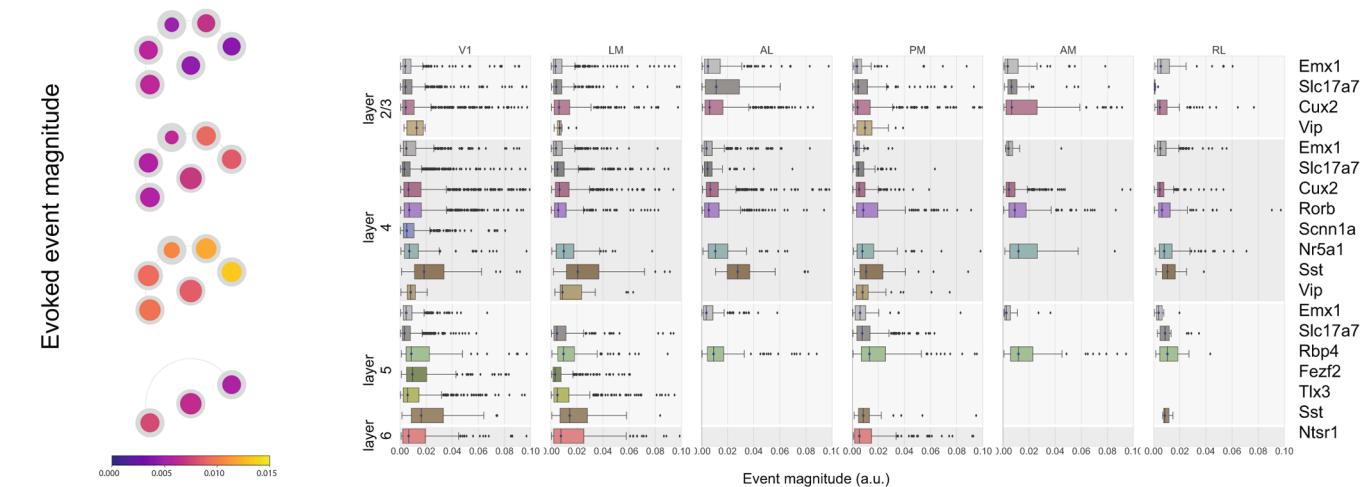
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41593-019-0550-9>.

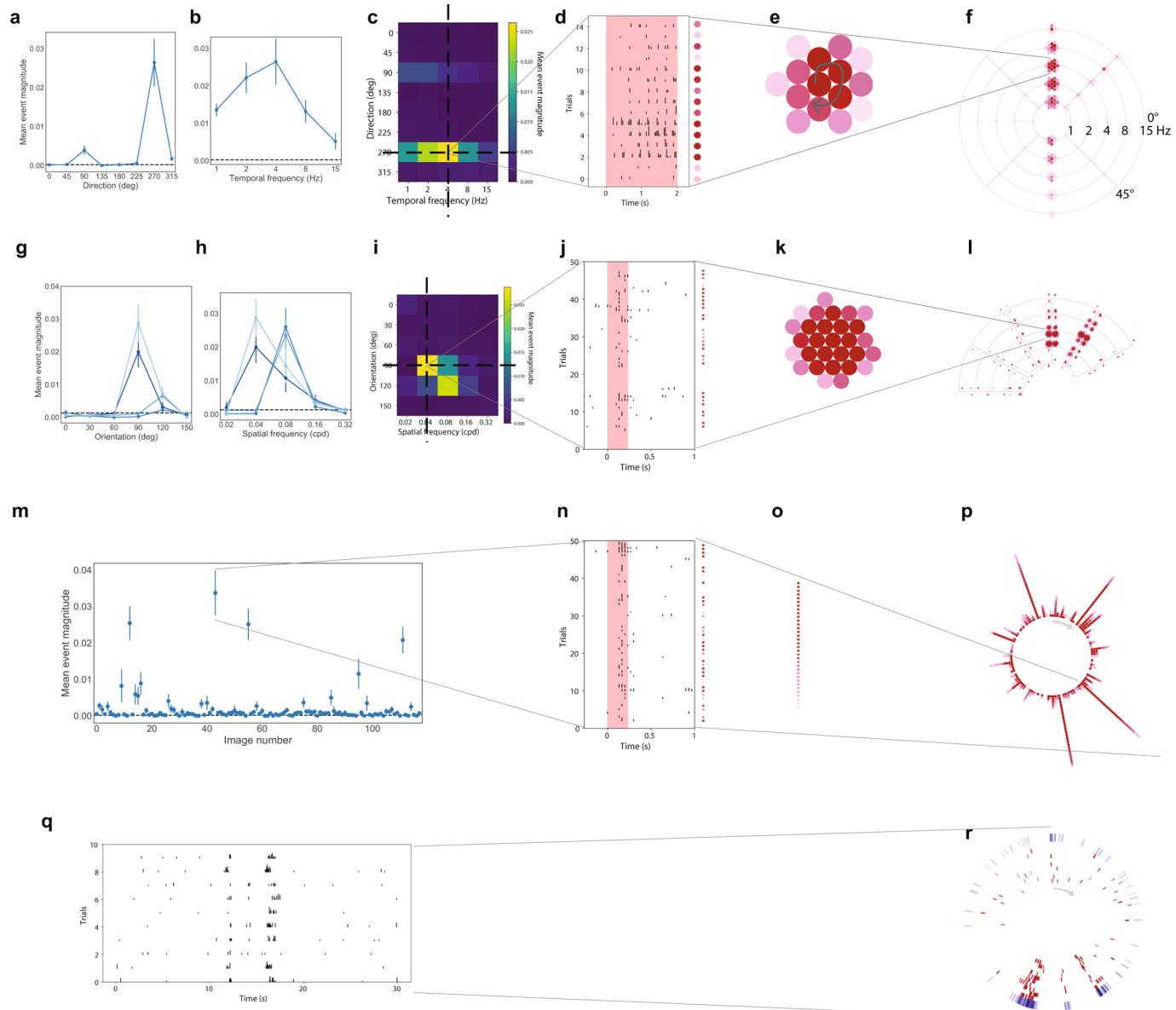
Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-019-0550-9>.

Correspondence and requests for materials should be addressed to S.E.J.d., J.A.L. or M.A.B.

Reprints and permissions information is available at www.nature.com/reprints.

a**b**

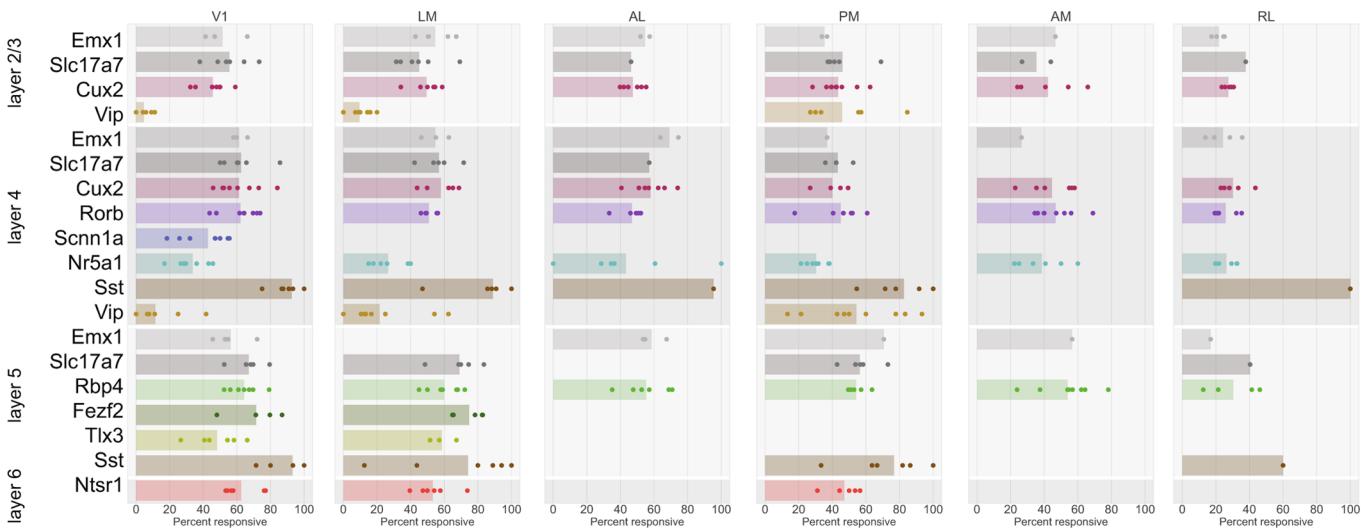
Extended Data Fig. 1 | Spontaneous and evoked event magnitude. **a**, Pawplot and box plots summarizing the mean event magnitude for neurons during the 5 minute spontaneous activity (mean luminance gray) stimulus. For a description of the visualization see Fig. 3. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range. Points outside this range are shown as outliers. See Extended Data Figure 3 for sample sizes. **b**, Pawplot and box plots summarizing the maximum evoked event magnitude for neurons' responses to drifting gratings. See Extended Data Figure 3 for sample sizes.



Extended Data Fig. 2 | Response visualizations. Conventional tuning curves for drifting grating responses for one neuron. **a**, Direction tuning plotted at the preferred temporal frequency (4 Hz) (mean \pm sem across 15 trials). Dotted line represents the mean response to the blank sweep. **b**, Temporal frequency tuning plotted at the preferred grating direction (270°) (mean \pm sem). **c**, Heatmap of the direction and temporal frequency responses for cell, showing any possible interaction of direction and temporal frequency. **d**, All 15 trials at the preferred direction and temporal frequency, 2 second grating presentation is indicated by pink shading. The mean event magnitude is represented by intensity of the dot to the right of the trial. **e**, All trials are clustered, with the strongest response in the center and weaker responses on the outside. **f**, Clusters are plotted on a "Star plot". Arms indicated the direction of grating motion, arcs indicate the temporal frequency of the grating, with the lowest in the center and the highest at the outside. Clusters of red dots are located at the intersection and arms and arcs, representing the trial responses at that condition. Tuning curves for static gratings for one neuron. **g**, Orientation tuning plotted at the preferred spatial frequency (0.04 cpd) for each of the four phases. (mean \pm sem across 50 trials) Dotted line represents the mean response to the blank sweep. **h**, Spatial frequency tuning plotted at the preferred orientation (90°) for each of the four phases (mean \pm sem). **i**, Heatmap of the orientation and spatial frequency at the preferred phase **j**, All trials at the preferred orientation, spatial frequency and phase, the 250 ms grating presentation is indicated by pink shading. The mean event magnitude is represented by the intensity of the dot to the right of the trial. **k**, All trials are clustered, with the strongest response in the center and weaker responses on the outside. **l**, Clusters are placed on a "Fan plot". Arms represent the orientation and arcs represent the spatial frequency of the grating. At each intersection, there are four lobes of clustered dots, one for each phase at that grating condition. Responses to natural scenes for one neuron. **m**, Responses to each image presented (mean \pm sem across 50 trials). Dotted line represents the mean response to the blank sweep. **n**, All trials of the image which elicited the largest mean response, the 250ms image presentation is indicated by pink shading. The mean event magnitude is represented by the intensity of the dot to the right of the trial. Trials are sorted **o**, and are plotted on a "Corona plot" **p**, Each ray represents the response to one image, with the strongest response on the inside and weaker responses at the outside. Responses to natural movies for one neuron. **q**, Responses of one neuron's response to each of 10 trials of the natural movie. **r**, Responses are plotted on a "Track plot". Each red ring represents the activity of the cell to one trial, proceed clockwise from the top of the track. The outer blue track represents the mean response across all ten trials.

Drifting gratings

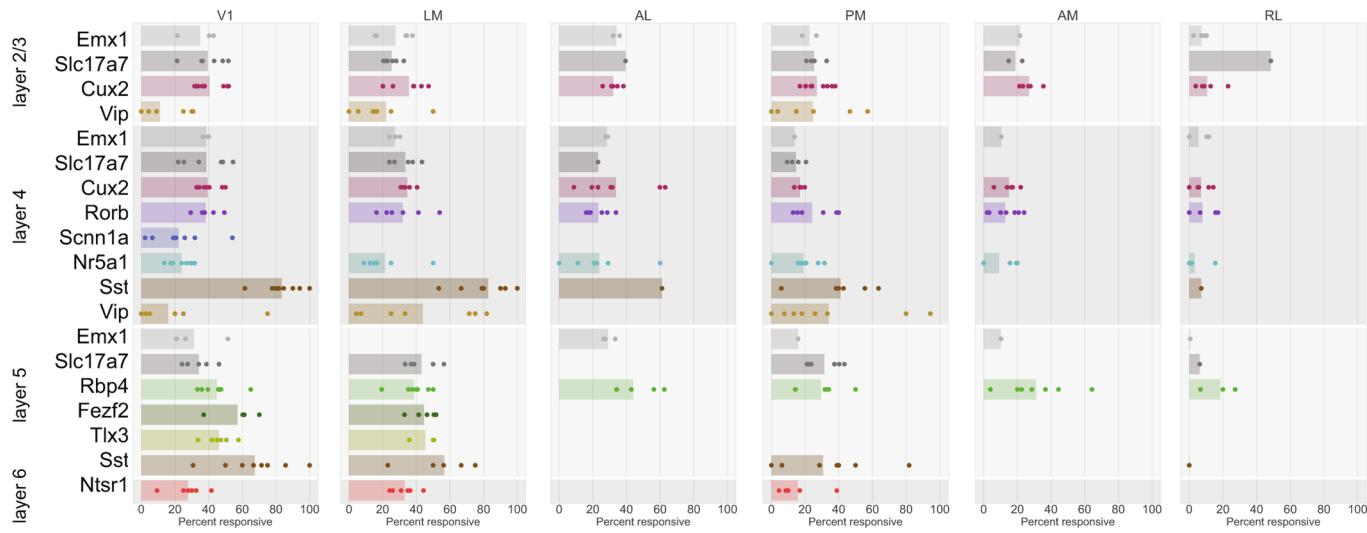
Cre	V1				LM				AL				PM				AM				RL				
	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	
Emx1 - layer 2/3	3	628	341	54%	5	769	453	59%	2	328	180	55%	2	293	104	35%	1	105	49	47%	4	448	100	22%	
Emx1 - layer 4	3	826	509	62%	3	677	359	53%	2	591	413	70%	1	168	62	37%	1	83	22	27%	4	823	209	25%	
Emx1 - layer 5	4	745	414	56%					3	316	187	59%	1	75	53	71%	1	53	30	57%	1	130	22	17%	
Slc17a7 - layer 2/3	6	955	539	56%	6	685	342	50%	1	41	19	46%	5	342	155	45%	2	138	51	37%	1	45	17	38%	
Slc17a7 - layer 4	6	1538	993	65%	5	1177	687	58%	1	182	104	57%	4	363	156	43%									
Slc17a7 - layer 5	5	898	586	65%	5	322	231	72%					6	540	299	55%					1	52	21	40%	
Cux2 - layer 2/3	8	1619	732	45%	6	794	399	50%	6	799	371	46%	9	948	404	43%	5	362	155	42%	6	376	107	28%	
Cux2 - layer 4	8	1939	1181	61%	5	1026	576	56%	7	1420	815	57%	4	676	258	38%	6	678	339	48%	6	705	194	28%	
Rorb	8	1617	1013	63%	6	767	391	51%	6	794	389	49%	7	511	255	50%	8	516	252	49%	5	743	188	25%	
Scnn1a	9	1200	517	43%																					
Nr5a1	8	441	146	33%	6	256	67	26%	6	178	73	41%	7	203	65	32%	6	110	39	35%	6	875	249	28%	
Rbp4	7	320	210	66%	7	333	199	60%	6	267	155	58%	6	375	198	53%	8	244	136	56%	4	64	22	34%	
Fzef2	4	285	211	74%	5	550	406	74%																	
Tlx3	6	840	401	48%	3	684	406	59%																	
Ntsr1	6	406	242	60%	7	436	229	53%					5	397	187	47%									
Sst - layer 4	9	101	94	93%	8	127	114	90%	1	22	21	95%	6	70	58	83%					1	15	15	100%	
Sst - layer 5	8	77	73	95%	7	82	56	68%					8	81	62	77%					1	5	3	60%	
Vip - layer 2/3	9	116	5	4%	8	137	13	9%					7	108	48	44%									
Vip - layer 4	8	83	12	14%	9	105	26	25%					9	120	62	52%									
Total	125	14634	8219	56%	101	8927	4954	55%	41	4938	2727	55%	87	5270	2426	46%	38	2309	1073	46%	40	4281	1147	26.8%	



Extended Data Fig. 3 | Responsiveness to drifting gratings. **a**, Table summarizing the numbers of experiments (expts) and neurons imaged for each Cre line, layer, area combination in response to drifting grating stimulus and the number, and percent, of neurons that were responsive to the drifting grating stimulus. **b**, Strip plots of the percent of neurons responsive to the drifting grating stimulus for each experiment.

Static gratings

Cre	V1				LM				AL				PM				AM				RL				
	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	
Emx1 - layer 2/3	3	579	209	36%	5	721	208	29%	2	340	117	34%	2	251	58	23%	1	88	19	22%	4	463	31	7%	
Emx1 - layer 4	3	631	243	39%	3	634	168	26%	2	543	154	28%	1	188	26	14%	1	134	14	10%	4	780	41	5%	
Emx1 - layer 5	4	684	204	30%					3	233	68	29%	1	44	7	16%	1	39	4	10%	1	162	1	1%	
Slc17a7 - layer 2/3	6	866	346	40%	6	645	168	26%	1	38	15	39%	5	338	84	25%	2	137	27	20%	1	33	16	48%	
Slc17a7 - layer 4	6	1369	549	40%	5	1163	376	32%	1	242	56	23%	4	288	42	15%									
Slc17a7 - layer 5	5	847	305	36%	5	302	130	43%					6	524	163	31%					1	48	3	6%	
Cux2 - layer 2/3	8	1691	671	40%	6	775	271	35%	6	753	240	32%	9	810	211	26%	5	385	108	28%	6	361	38	11%	
Cux2 - layer 4	8	1733	680	39%	5	1069	371	35%	7	1156	374	32%	4	522	92	18%	6	534	82	15%	6	673	31	5%	
Rorb	8	1404	529	38%	6	753	219	29%	6	791	184	23%	7	431	90	21%	8	388	49	13%	5	777	58	7%	
Scnn1a	9	1114	260	23%																					
Nr5a1	8	392	98	25%	6	240	37	15%	6	115	25	22%	7	228	45	20%	6	123	17	145	6	893	31	3%	
Rbp4	7	262	115	44%	7	308	121	39%	6	256	103	40%	6	350	87	25%	8	212	62	29%	4	61	11	18%	
Fezf2	4	291	172	59%	5	582	264	45%																	
Tlx3	6	829	379	46%	3	612	280	46%																	
Ntsr1	6	356	78	22%	7	452	139	31%					5	349	40	11%									
Sst - layer 4	9	111	93	84%	8	98	81	83%	1	18	11	61%	6	67	25	37%					1	14	1	7%	
Sst - layer 5	8	71	47	66%	7	70	36	51%					8	92	26	28%					1	7	0	0%	
Vip - layer 2/3	9	113	12	11%	8	118	23	19%					7	145	27	19%									
Vip - layer 4	8	104	14	13%	9	104	42	40%					9	144	44	31%									
Total	125	13447	5004	37%	101	8646	2934	34%	41	4485	1347	30%	87	4771	1067	32%	38	2040	382	19%	40	4272	262	6%	

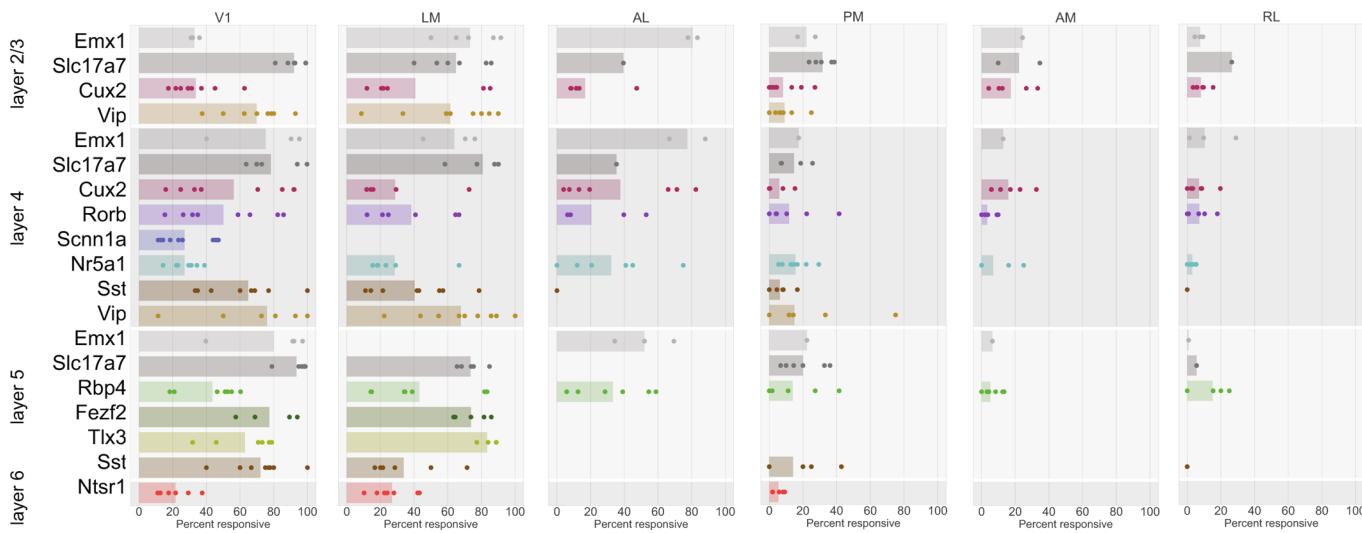


Extended Data Fig. 4 | Responsiveness to static gratings. a, Table summarizing the numbers of experiments and neurons imaged for each Cre line, layer, area combination in response to static grating stimulus and the number, and percent, of neurons that were responsive to the static grating stimulus.

b, Strip plots of the percent of neurons responsive to the static grating stimulus for each experiment.

Locally sparse noise

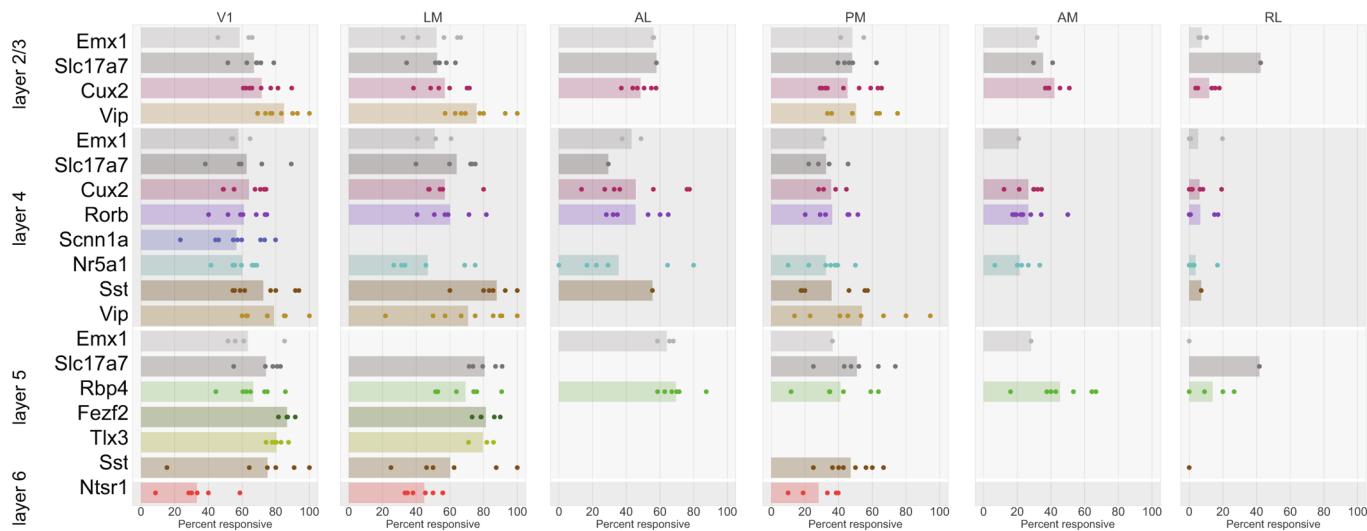
Cre	V1				LM				AL				PM				AM				RL				
	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	
Emx1 - layer 2/3	3	549	179	33%	5	654	518	79%	2	316	257	81%	2	270	60	22%	1	78	19	24%	4	500	39	8%	
Emx1 - layer 4	3	705	551	78%	3	584	407	70%	2	557	422	76%	1	206	36	17%	1	117	15	13%	4	881	103	12%	
Emx1 - layer 5	4	640	498	78%					3	275	150	55%	1	76	17	22%	1	31	2	6%	1	120	1	1%	
Slc17a7 - layer 2/3	6	962	898	93%	6	666	483	73%	1	38	15	39%	5	344	109	32%	2	149	34	23%	1	34	9	26%	
Slc17a7 - layer 4	6	1394	1150	82%	5	1180	991	84%	1	229	81	35%	4	296	43	14%									
Slc17a7 - layer 5	5	899	850	95%	5	294	220	75%					6	578	125	22%					1	36	2	6%	
Cux2 - layer 2/3	8	1363	427	31%	6	817	301	37%	6	742	93	13%	9	762	52	7%	5	354	59	17%	6	371	32	9%	
Cux2 - layer 4	8	1775	894	50%	5	951	229	24%	7	1144	413	36%	4	537	18	3%	6	517	87	17%	6	734	52	7%	
Rorb	8	1414	685	48%	6	783	298	38%	6	753	163	22%	7	484	48	10%	8	404	18	4%	5	829	52	7%	
Scnn1a	9	1176	321	27%																					
Nr5a1	8	378	101	27%	6	267	58	22%	6	106	31	29%	7	234	38	16%	6	115	10	9%	6	811	22	2.7	
Rbp4	7	313	124	40%	7	338	141	42%	6	283	68	24%	6	336	32	10%	8	237	14	6%	4	64	13	20%	
Fezf2	4	266	214	80%	5	604	448	74%																	
Tlx3	6	874	541	62%	3	629	520	83%																	
Ntsr1	6	371	67	18%	7	395	96	24%					5	353	20	6%									
Sst - layer 4	9	107	68	64%	8	104	46	44%	1	14	0	0%	6	93	5	5%					1	25	0	0%	
Sst - layer 5	8	57	40	70%	7	71	20	28%					6	65	9	14%					1	7	0	0%	
Vip - layer 2/3	9	120	82	68%	8	118	65	55%					7	105	8	8%									
Vip - layer 4	8	122	64	52%	9	94	60	64%					9	121	10	8%									
Total	125	14634	7754	53%	101	8927	4901	55%	41	4938	1693	34%	87	5270	630	12%	38	2309	258	11%	40	4281	325	8%	



Extended Data Fig. 5 | Responsiveness to locally sparse noise. **a**, Table summarizing the numbers of experiments (expts) and neurons imaged for each Cre line, layer, area combination in response to locally sparse noise stimulus and the number, and percent, of neurons that were responsive to the locally sparse noise stimulus. **b**, Strip plots of the percent of neurons responsive to the locally sparse noise stimulus for each experiment.

Natural scenes

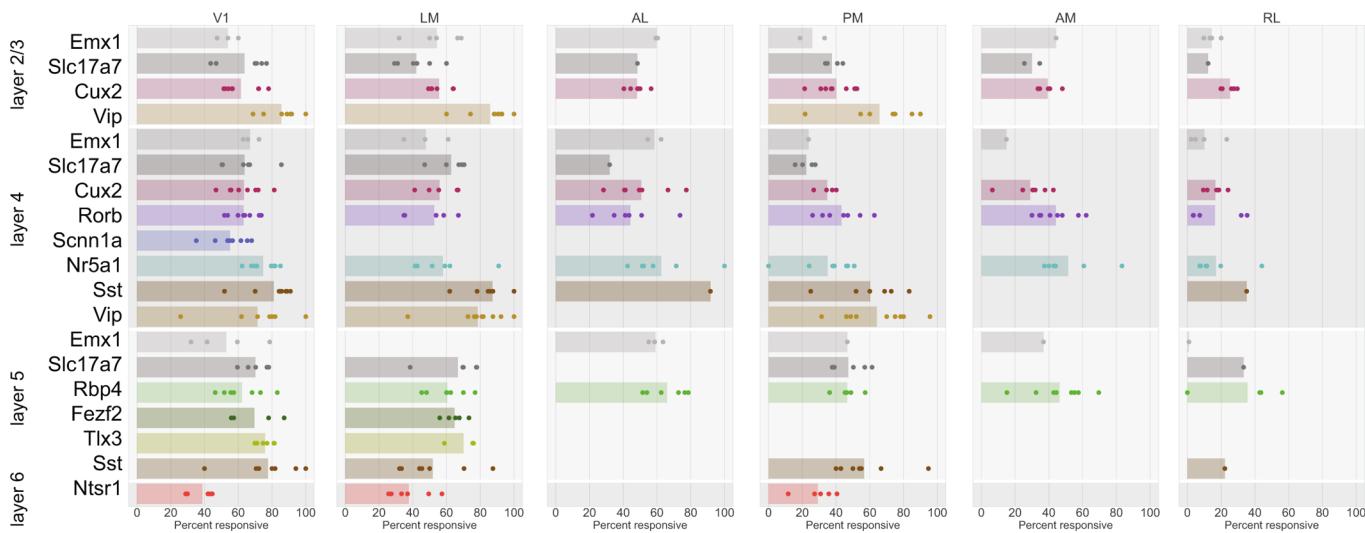
Cre	V1				LM				AL				PM				AM				RL				
	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	
Emx1 - layer 2/3	3	579	346	60%	5	721	394	55%	2	340	191	56%	2	251	123	49%	1	88	28	32%	4	463	33	7%	
Emx1 - layer 4	3	631	364	58%	3	634	307	48%	2	543	231	43%	1	188	59	31%	1	134	28	21%	4	780	53	7%	
Emx1 - layer 5	4	684	426	62%					3	233	151	65%	1	44	16	36%	1	39	11	28%	1	162	0	0%	
Slc17a7 - layer 2/3	6	866	585	68%	6	645	349	54%	1	38	22	58%	5	338	161	48%	2	137	50	37%	1	33	14	42%	
Slc17a7 - layer 4	6	1369	889	65%	5	1163	743	64%	1	242	71	29%	4	288	92	32%									
Slc17a7 - layer 5	5	847	631	74%	5	302	246	81%					6	524	280	53%					1	48	20	42%	
Cux2 - layer 2/3	8	1691	1198	71%	6	775	432	56%	6	753	369	49%	9	810	353	44%	5	385	171	44%	6	361	51	14%	
Cux2 - layer 4	8	1733	1090	63%	5	1069	578	54%	7	1156	511	44%	4	522	182	35%	6	534	155	29%	6	673	30	4%	
Rorb	8	1404	846	60%	6	753	443	59%	6	791	360	46%	7	431	145	34%	8	388	86	22%	5	777	42	5%	
Scnn1a	9	1114	649	58%																					
Nr5a1	8	392	230	59%	6	240	91	38%	6	115	42	37%	7	228	80	35%	6	123	28	23%	6	893	34	4%	
Rbp4	7	262	181	69%	7	308	208	68%	6	256	169	66%	6	350	123	35%	8	212	94	44%	4	61	12	20%	
Fezf2	4	291	252	87%	5	582	474	81%																	
Tlx3	6	829	669	81%	3	612	490	80%																	
Ntsr1	6	356	94	26%	7	452	189	42%					5	349	97	28%									
Sst - layer 4	9	111	81	73%	8	98	87	89%	1	18	10	56%	6	67	22	33%					1	14	1	7%	
Sst - layer 5	8	71	50	70%	7	70	38	54%					8	92	43	47%					1	7	0	0%	
Vip - layer 2/3	9	113	94	83%	8	118	88	75%					7	145	66	46%									
Vip - layer 4	8	104	78	75%	9	104	69	66%					9	144	70	49%									
Total	125	13447	8753	65%	101	8646	5226	60%	41	4485	2127	47%	87	4771	1912	40%	38	2040	651	32%	40	4272	290	7%	



Extended Data Fig. 6 | Responsiveness to natural scenes. **a**, Table summarizing the numbers of experiments (expts) and neurons imaged for each Cre line, layer, area combination in response to locally sparse noise stimulus and the number, and percent, of neurons that were responsive to the locally sparse noise stimulus. **b**, Strip plots of the percent of neurons responsive to the locally sparse noise stimulus for each experiment.

Natural movies

Cre	V1				LM				AL				PM				AM				RL				
	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	expts	total	responsive	percent	
Emx1 - layer 2/3	3	931	512	55%	5	1122	625	56%	2	512	307	60%	2	422	112	27%	1	167	74	44%	4	712	99	14%	
Emx1 - layer 4	3	1043	700	67%	3	976	470	48%	2	838	484	58%	1	308	73	24%	1	195	29	15%	4	1223	129	11%	
Emx1 - layer 5	4	1099	535	49%					3	437	258	59%	1	105	49	47%	1	95	35	37%	1	217	2	1%	
Slc17a7 - layer 2/3	6	1433	931	65%	6	1115	482	43%	1	64	31	48%	5	581	222	38%	2	235	72	31%	1	65	8	12%	
Slc17a7 - layer 4	6	2122	1409	66%	5	1645	1031	63%	1	310	99	32%	4	551	123	22%									
Slc17a7 - layer 5	5	1285	900	70%	5	470	329	70%					6	838	402	48%					1	72	24	33%	
Cux2 - layer 2/3	8	2419	1479	61%	6	1208	677	56%	6	1226	605	49%	9	1433	556	39%	5	619	237	38%	6	558	144	26%	
Cux2 - layer 4	8	2662	1641	62%	5	1584	872	55%	7	1877	928	49%	4	928	315	34%	6	997	311	31%	6	1020	144	14%	
Rorb	8	2218	1383	62%	6	1191	648	54%	6	1242	543	44%	7	764	305	40%	8	735	313	43%	5	1126	168	15%	
Scnn1a	9	1873	1058	58%																					
Nr5a1	8	578	424	73%	6	421	213	51%	6	220	116	53%	7	331	125	38%	6	171	77	45%	6	1354	268	20%	
Rbp4	7	458	304	66%	7	485	293	60%	6	441	267	61%	6	509	241	47%	8	355	170	48%	4	93	38	41%	
Fezf2	4	407	292	72%	5	981	642	65%																	
Tlx3	6	1181	893	76%	3	946	665	70%																	
Ntsr1	6	573	196	34%	7	719	262	36%					5	581	170	29%									
Sst - layer 4	9	159	125	79%	8	164	141	86%	1	24	22	92%	6	113	67	59%					1	37	13	35%	
Sst - layer 5	8	107	81	76%	7	137	67	49%					8	134	78	58%					1	9	2	22%	
Vip - layer 2/3	9	157	131	83%	8	161	134	83%					7	182	98	54%									
Vip - layer 4	8	195	105	54%	9	154	113	73%					9	205	121	59%									
Total	125	20900	13661	65%	101	13479	8115	60%	41	7191	3860	54%	87	7985	3252	41%	38	3569	1380	39%	40	5486	1085	17%	



Extended Data Fig. 7 | Responsiveness to natural movies. **a**, Table summarizing the numbers of experiments (expts) and neurons imaged for each Cre line, layer, area combination in response to any of the natural movie stimuli and the number, and percent, of neurons that were responsive to the natural movie stimuli. **b**, Strip plots of the percent of neurons responsive to the natural movie stimuli for each experiment.

Running correlations

Cre	V1		LM		AL		PM		AM		RL	
	expts	neurons										
Emx1 - layer 2/3			2	254					1	143	2	327
Emx1 - layer 4			1	148	1	254			1	117	2	322
Emx1 - layer 5	1	334			2	298			1	53	1	162
Slc17a7 - layer 2/3	3	656	4	559	1	41	4	370	2	216		
Slc17a7 - layer 4	4	873	4	1061			1	154				
Slc17a7 - layer 5	4	806	3	201			3	342				
Cux2 - layer 2/3	7	1928	5	875	6	1127	7	770	4	504	5	379
Cux2 - layer 4	7	1954	3	916	3	704	3	614	4	789	3	214
Rorb	4	984	2	513	5	896	6	605	5	475	4	807
Scnn1a	5	881										
Nr5a1	6	350	5	421	4	177	5	182	3	80	4	942
Rbp4	6	325	4	272	4	306	6	397	6	254	2	47
Fezf2	3	290	2	257								
Tlx3			1	303								
Ntsr1	3	230	4	457			4	355				
Sst - layer 4	4	49	6	128	1	24	5	94			1	24
Sst - layer 5	6	56	5	65			3	58			1	9
Vip - layer 2/3	4	77	5	88			5	133				
Vip - layer 4	6	115	4	46			7	136				

Extended Data Fig. 8 | Populations for running correlation analysis. Table summarizing the number of experiments and neurons, for each Cre line, layer, area combination, included in the running correlation analysis. These are from sessions in which the mouse was running between 20–80% of the time.

Wavelet models

Cre	V1		LM		AL		PM		AM		RL	
	expts	neurons										
Emx1 - layer 2/3	3	296	4	217	2	160	2	127	1	28	3	178
Emx1 - layer 4	2	247	2	152			1	86	1	36	2	168
Emx1 - layer 5	4	324			3	139	1	31	1	5	1	59
Slc17a7 - layer 2/3	5	366	6	274	1	14	5	136	2	66	1	14
Slc17a7 - layer 4	4	446	2	207	1	123	4	124				
Slc17a7 - layer 5	3	200	5	168			6	290			1	23
Cux2 - layer 2/3	8	769	6	431	6	362	9	350	5	160	6	204
Cux2 - layer 4	6	639	4	359	7	663	4	270	6	236	4	180
Rorb	7	652	6	398	6	378	7	242	8	191	2	114
Scnn1a	9	531										
Nr5a1	8	238	6	122	6	59	7	123	6	68	5	401
Rbp4	7	151	7	177	6	123	6	209	8	129	4	38
Fezf2	4	177	5	237								
Tlx3	5	400	2	209								
Ntsr1	6	206	7	186			5	182				
Sst - layer 4	9	64	8	65	1	10	6	49			1	5
Sst - layer 5	7	34	6	30			8	35			1	4
Vip - layer 2/3	9	80	8	91			7	68				
Vip - layer 4	7	34	9	59			9	74				

Extended Data Fig. 9 | Populations for wavelet model analysis. Table summarizing the number of experiments and neurons for each Cre line, layer, area combination for which wavelet models were fit. The neurons had to be present in all three imaging sessions to be included.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Two photon imaging data was collected using either Nikon Elements (v4.3) or Sciscan (v12.0), with custom Python 2.7 scripts to run the workflow and interface with the eye-tracking and behavior cameras. Intrinsic signal imaging data was collected using custom scripts written in Python.

Data analysis

All analyses were performed using custom scripts written in Python 2.7, using NumPy, SciPy, Pandas, Matplotlib, Seaborn, Keras, and Tensorflow, or MATLAB. Analysis code is available at <http://alleninstitute.github.io/AllenSDK/> and https://github.com/alleninstitute/visual_coding_2p_analysis. Event extraction was performed using FastLZeroSpikeInference available at <https://github.com/jewellsean/FastLZeroSpikeInference>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

This is an openly available dataset, accessible via a dedicated web portal (<http://observatory.brain-map.org/visualcoding>), with a custom Python-based Application Programming Interface (API), the AllenSDK (<http://alleninstitute.github.io/AllenSDK/>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined qualitatively to balance repeated experiments for each area/layer/Cre line combination and the preserve the breadth of survey. Our sample size matched, or exceeded, those found in previous publications.
Data exclusions	Mice were excluded for evidence of epileptiform activity, and individual imaging sessions were failed if there were signs of bleaching, saturation, excessive z-drift, or animal stress, among other factors, as described in our Methods.
Replication	Data acquired from multiple mice from multiple litters per transgenic line surveyed. Extensive acquisition metadata as well as detailed white papers are reported as part of the Allen Cell Brain Observatory (observatory.brain-map.org); these additional details are intended to aid other laboratories if they seek to replicate the results presented in this study.
Randomization	Within each transgenic Cre line, mice were randomly assigned to data collection in order to sample different areas and imaging depths. Stimulus conditions for gratings and natural scenes were presented in a randomized order within each epoch, as described. No other randomization was used as there were fixed experimental condition for all other aspects of the data set.
Blinding	Blinding was not relevant to this study as there was a single experimental condition for all data collected.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Mus musculus, male and female, mean age 108 ± 17 (st. dev.) days Wild animals
Wild animals	This study did not use wild animals.
Field-collected samples	This study did not use samples collected from the field
Ethics oversight	Experiments involving mice were approved by the Institutional Animal Care and Use Committees of the Allen Institute for Brain Science in accordance with NIH guidelines.

Note that full information on the approval of the study protocol must also be provided in the manuscript.