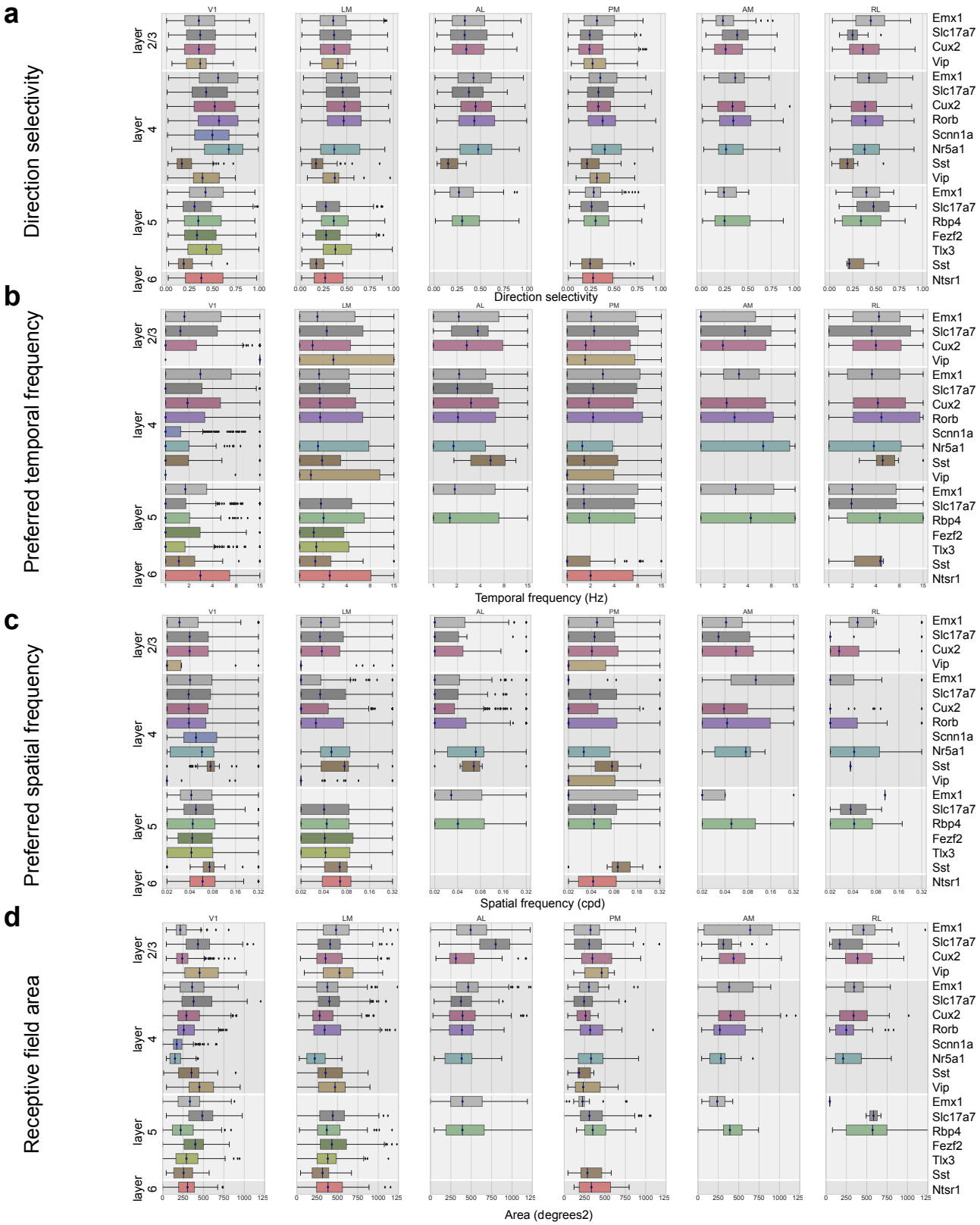
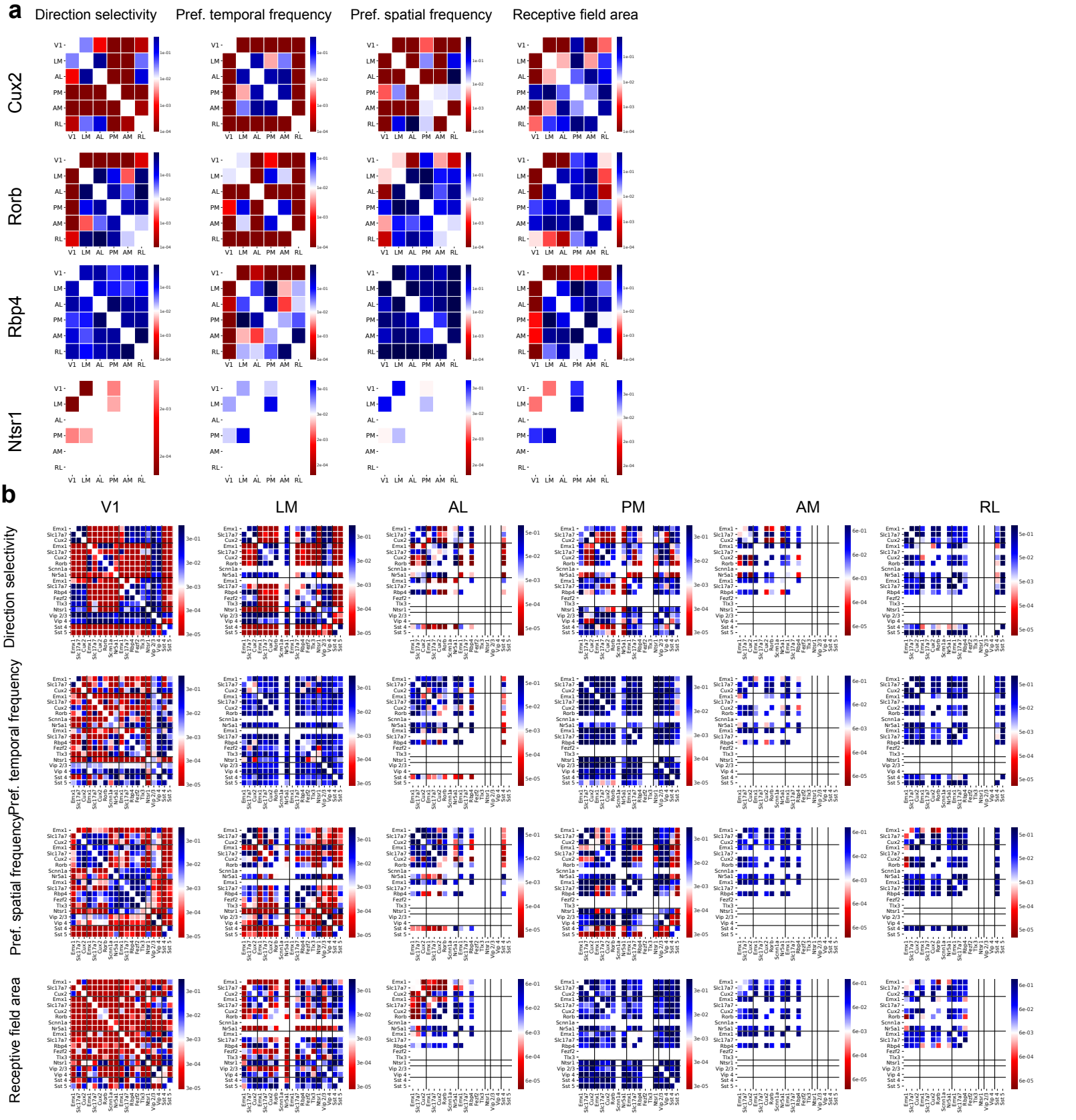


Supplementary Figure 1

Response characterizations across all areas

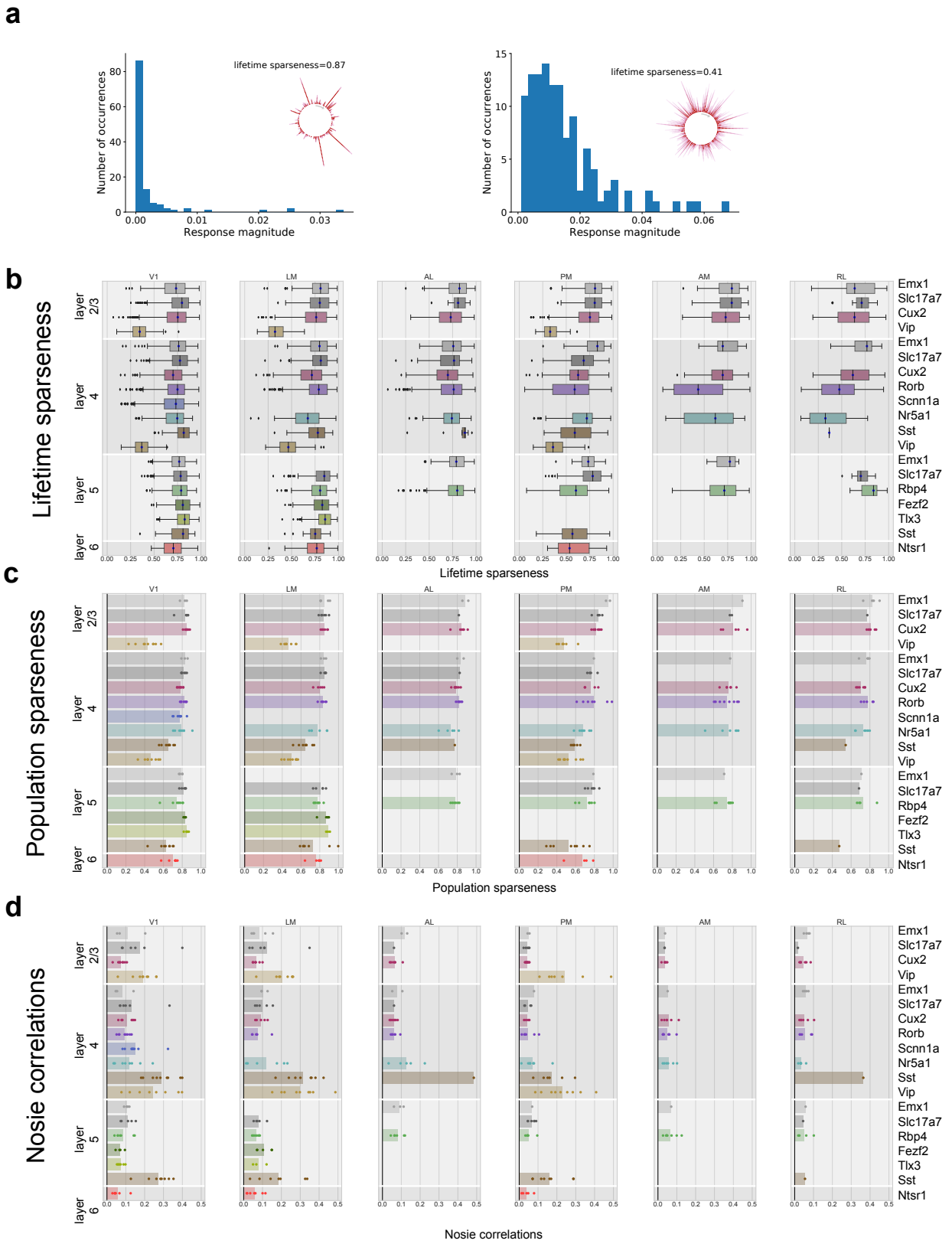
Box plots representing the distribution of values for (a) direction selectivity (See Extended Data Figure 3 for sample sizes), (b) preferred temporal frequency (See Extended Data Figure 3 for sample sizes), (c) preferred spatial frequency (See Extended Data Figure 4 for sample sizes), and (d) receptive field area (See Extended Data Figure 5 for sample sizes) for each Cre line and layer across all six visual areas. 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.



Supplementary Figure 2

Statistical tests for single cell response metrics

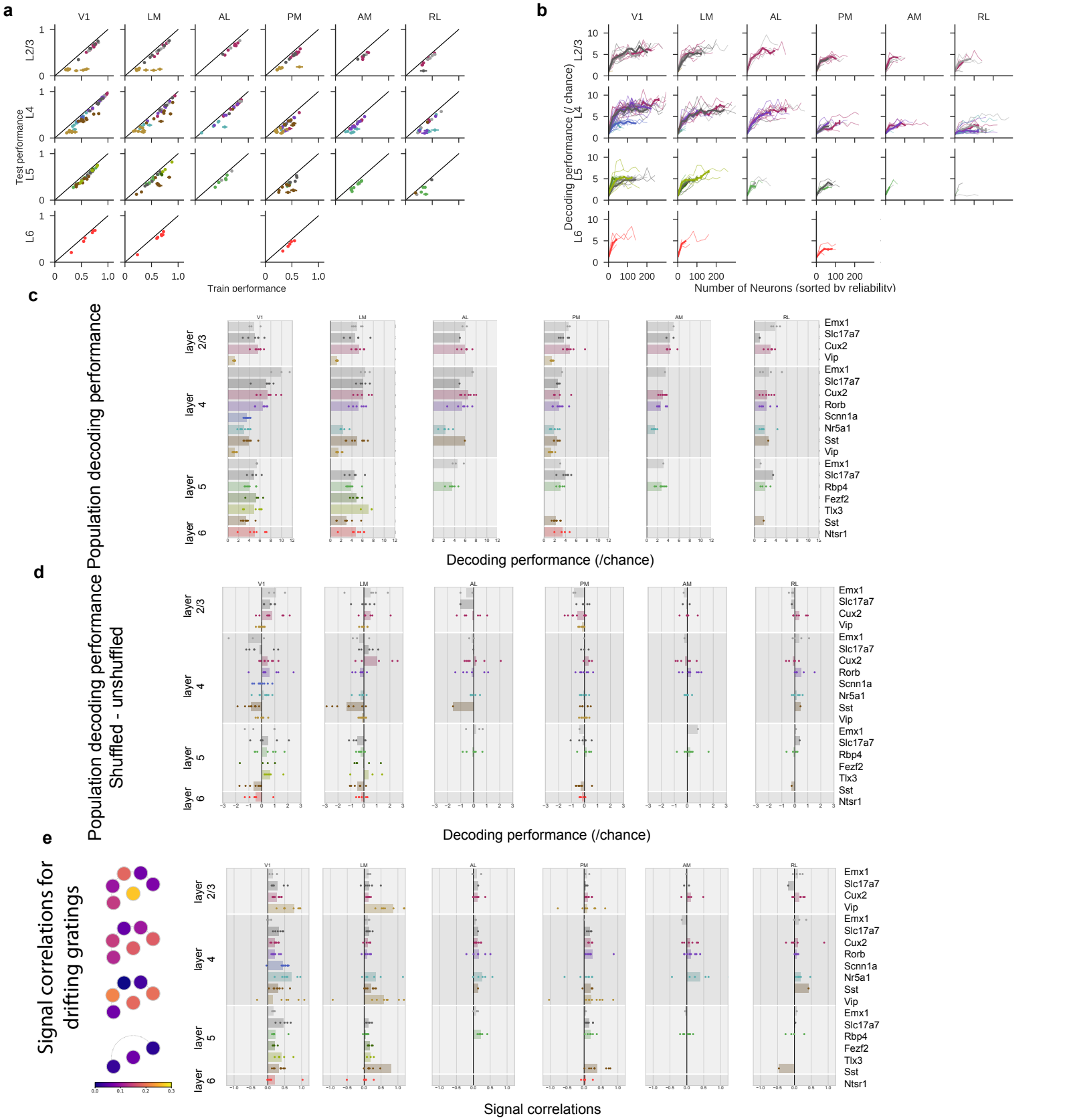
(a) Heatmaps of p-values of pairwise comparisons for each Cre line across areas using the (two-sided) Kolmogorov-Smirnov (KS) test with a Bonferroni correction for the number of comparisons. The heatmap is centered on the significance criteria. E.g. for comparisons across all six visual areas, p<0.01 (=0.05/5) is significant. For comparisons across only 3 visual areas (eg. for Ntsr1) p<0.025 is significant. (b) Heatmaps of p-values for pairwise comparison across each Cre-line and layer combination within each visual area, using KS test with a Bonferroni correction for the number of comparisons. The colormap for the p-values is centered at the significance criteria. For sample sizes see Extended Data Figure 3 (direction selectivity & temporal frequency), Extended Data Figure 4 (spatial frequency), and Extended Data Figure 5 (receptive field area).



Supplementary Figure 3

Sparsity characterization across all areas

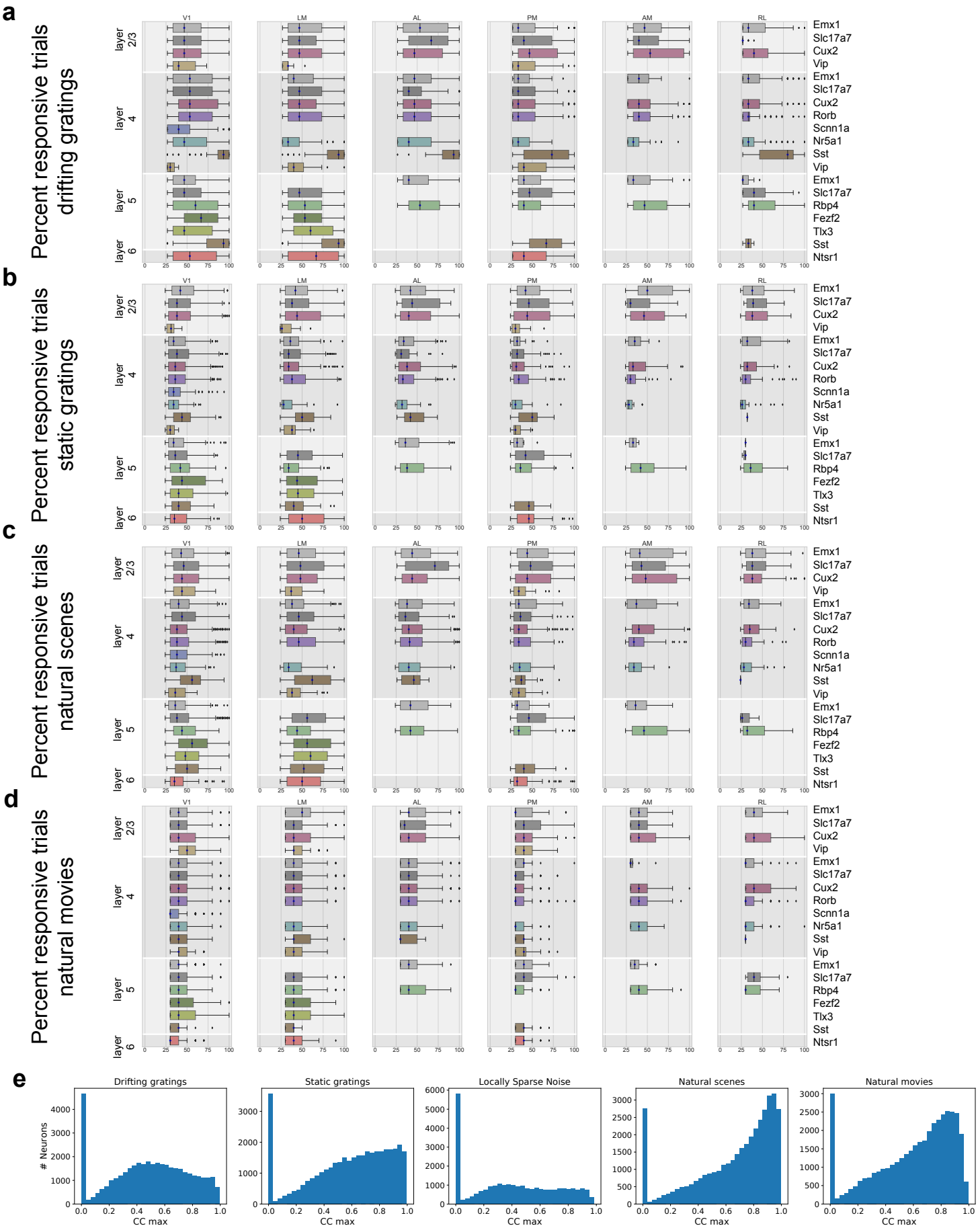
(a) Distribution of evoked responses for two example cells showing high (left) or low (right) lifetime sparseness. The corona plot for each cell is inset. These neurons are the top and bottom examples in Figure 2d. (b) Box plots representing the distribution of values for lifetime sparseness for each Cre line and layer across all six visual areas. 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 sample sizes see Extended Data Figure 6. (c) Strip plot representing the population sparseness for each Cre line and layer across all six visual areas. Each dot represents the mean population sparseness for a single experiment, with the median for a Cre-line/layer represented by the bar. For sample sizes see Extended Data Figure 6 (expts) (d) Strip plot representing the noise correlations for each Cre line and layer across all six visual areas. For sample sizes see Extended Data Figure 3 (expts).



Supplementary Figure 4

Population decoding

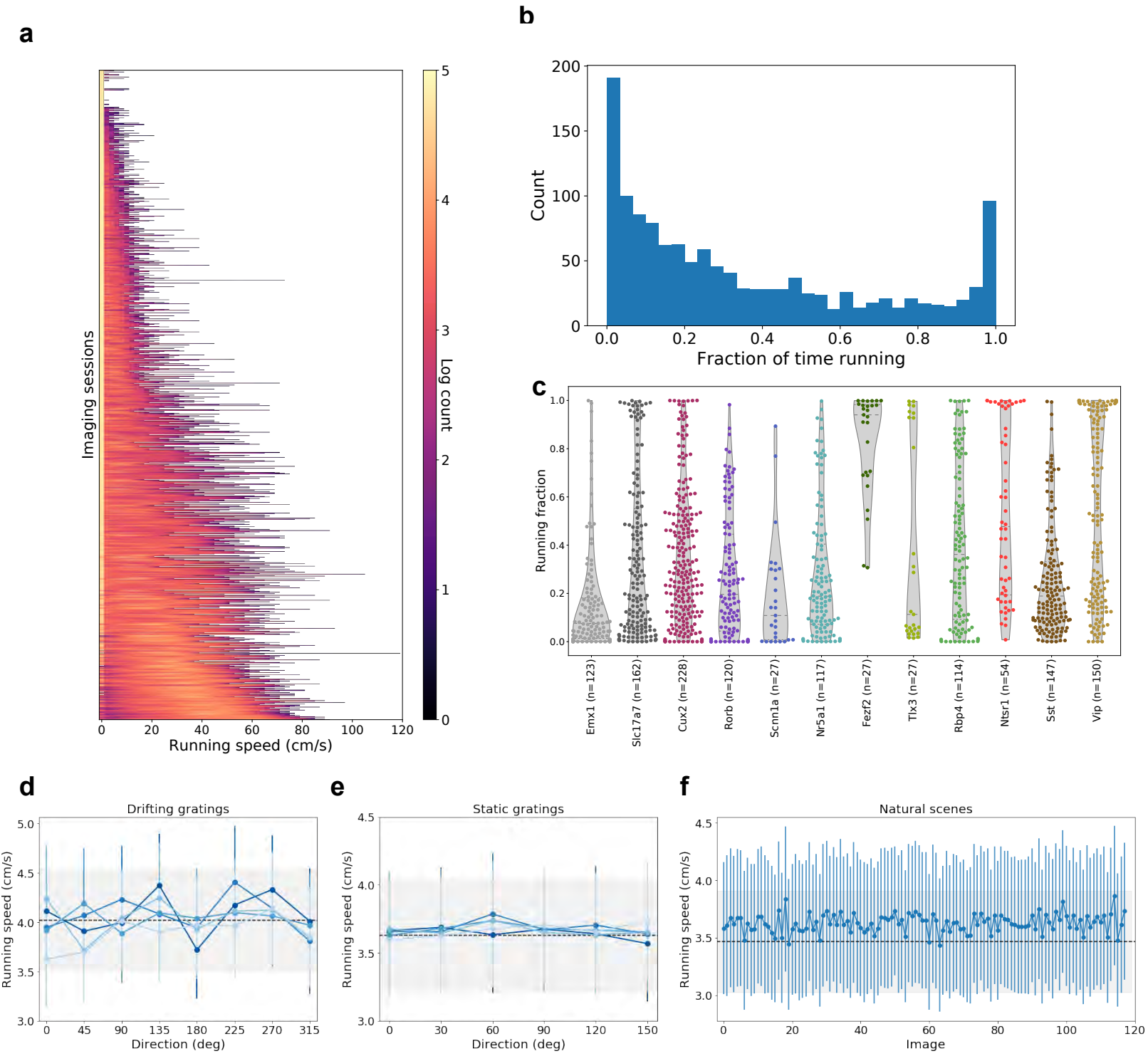
a) Test vs train performance for K-Nearest Neighbor decoding of drifting grating direction for each experiment, across all areas, layers and Cre lines. For all experiments, other than Vip, the experiments are close to unity, reflecting good generalization. Error bars represent the standard error of the mean across five-fold cross-validation. For sample sizes see Extended Data Figure 3 (expt). (b) Dependence of decoding performance on the number of cells included in the decoding, ordered by decreasing reliability, for all areas and Cre lines. (c) Strip plots summarizing decoding performance across all Cre lines and areas. Each dot represents the mean five-fold cross-validated decoding performance of a single experiment, with the median performance for a Cre-line/layer represented by the bar. For sample sizes see Extended Data Figure 3 (expt). (d) Strip plots of the difference between the decoding performance when trials are shuffled, destroying noise correlations, and trials are kept intact. Positive values indicate that noise c orrelations hurt decoding performance, while negative values indicate that noise correlations improve decoding performance. For sample sizes see Extended Data Figure 3 (expt). (e) Pawplot and strip plot of the median signal correlation in each experiment. For sample sizes see Extended Data Figure 3 (expt).



Supplementary Figure 5

Response Variability

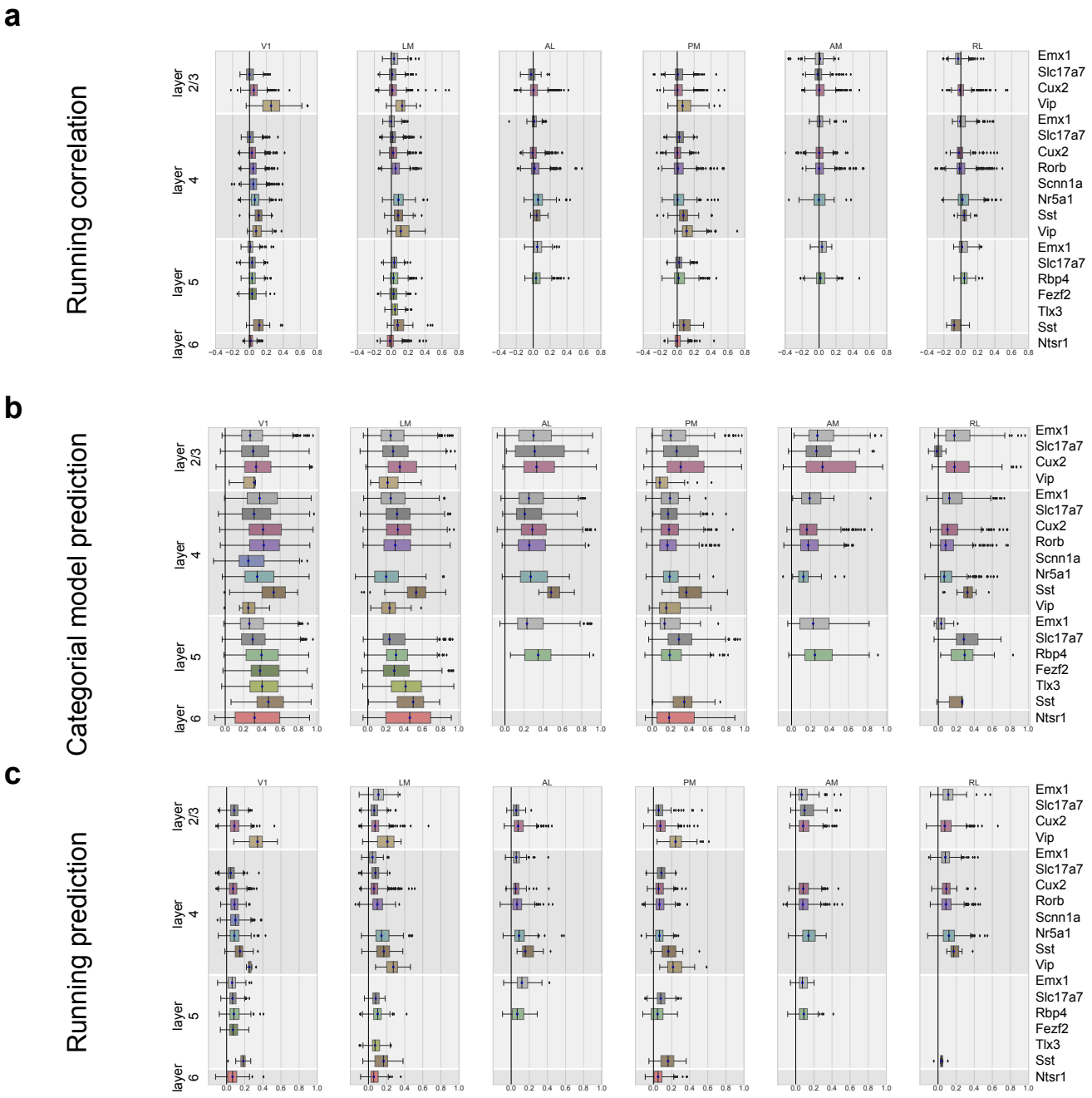
Box plots representing the distribution of values for percent responsive trials for responsive neurons for (a) drifting gratings, (b) static gratings, (c) natural scenes, and (d) natural movie three for all Cre lines, layers, and areas. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range. Points outside this range are shown as outliers. For sample sizes see Extended Data Figure 3 (drifting gratings), 4 (static gratings), 6 (natural scenes), and 7 (natural movies).
(e) Histograms of the CC max for all stimuli (see Methods).



Supplementary Figure 6

Characterization of running activity

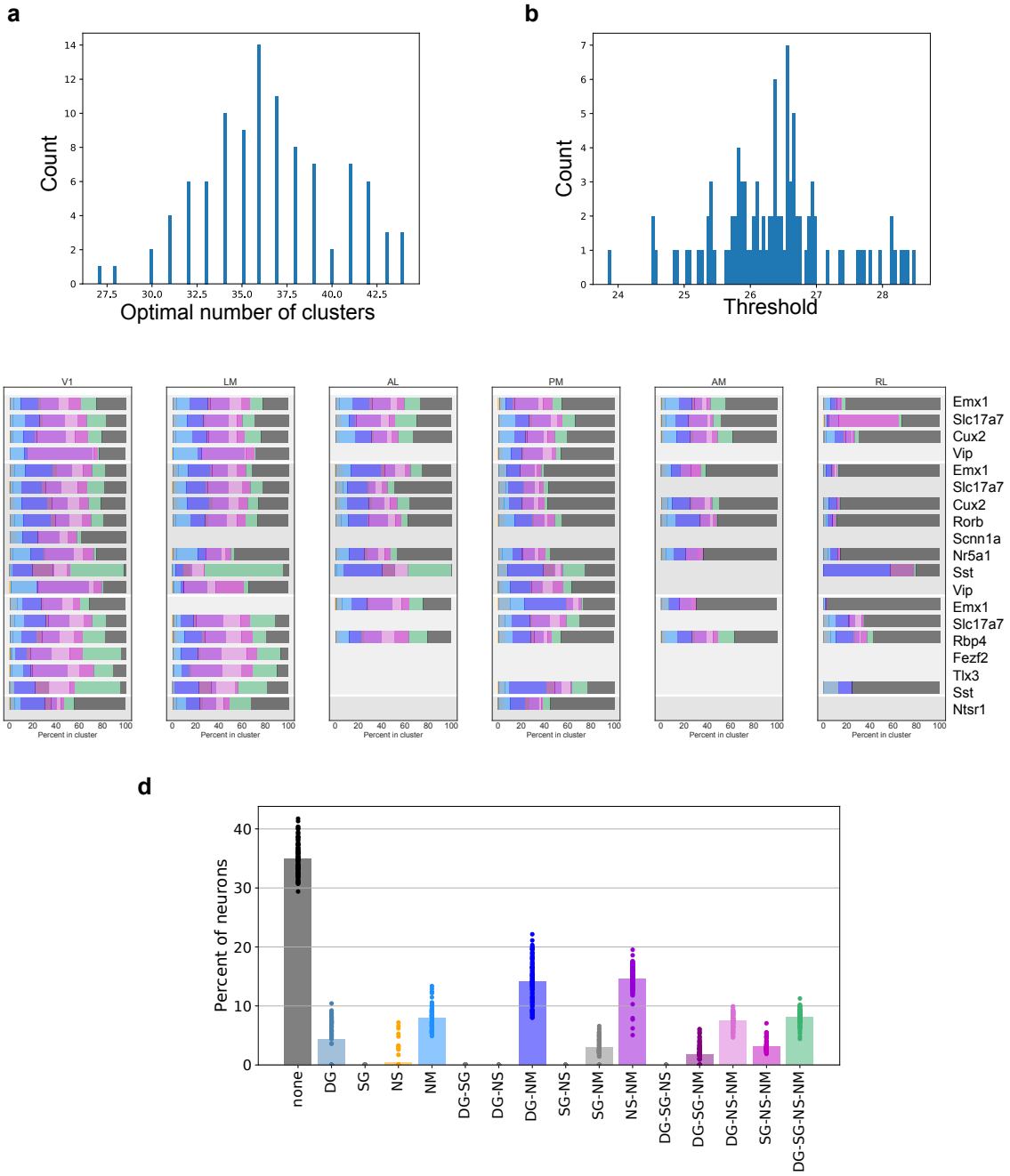
(a) Heatmap of running speed distributions for all imaging sessions, ordered by the mean running speed. (b) Histogram of the fraction of time the mouse is running (>1 cm/s) for all imaging sessions. (c) Distribution of the fraction of time the mouse is running for each Cre line. The violinplot uses Scott's rule for the kernel bandwidth and is scaled so each violin has the same area. Mice from some Cre lines exhibit more running (eg. Fezf2). (d) Mean (± s.e.m.) running speed for grating direction for drifting gratings at each temporal frequency, with the mean running speed during the blank sweep indicated by the dashed line ± s.e.m. (gray shading) across all 243 mice. (e) Mean (± s.e.m.) running speed for grating orientation for static gratings at each spatial frequency, with the mean running speed during the blank sweep indicated by dashed line ± s.e.m. (gray shading) across all 243 mice. (f) Mean (± s.e.m.) running speed for natural scenes, with the mean running speed during the blank sweep indicated by the dashed line ± s.e.m. (gray shading) across all 243 mice.



Supplementary Figure 7

Categorical model and running correlations

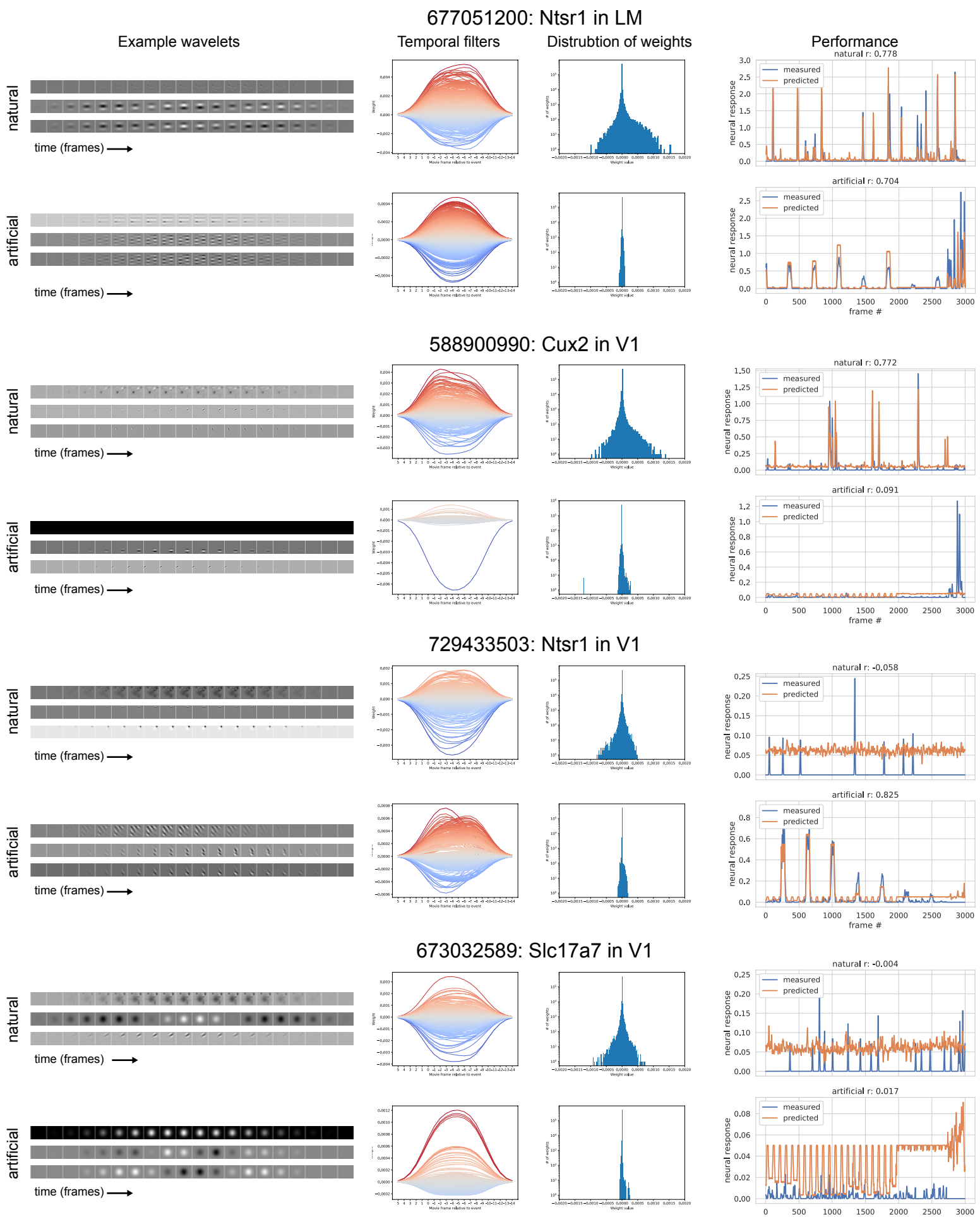
(a) Box plots of running (Pearson) correlation for all Cre lines and areas. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range. Points outside this range are shown as outliers. Only data from sessions in which the mouse ran between 20-80% of the session were included. For sample sizes see Extended Data Figure 8. (b) Box plots of categorial model prediction (Pearson correlation between predicted and measured activity) for all Cre lines and areas. For sample sizes see Extended Data Figure 3. (c) Box plots of the prediction of activity (Pearson correlation between the predicted and measured activity) based on the polynomial regression on running speed (see Methods) across all Cre lines and areas. For sample sizes see Extended Data Figure 3.



Supplementary Figure 8

Clustering of responsiveness to stimuli

(a) Distribution of the optimal number of clusters determined through model comparison for each of the 100 runs of the clustering analysis. (b) Distribution of the responsiveness threshold computed for each run of the clustering analysis. (c) The mean percent of cells belonging to each class for each transgenic Cre line within each visual area. (d) Percent of neurons belonging to each class predicted by the model, mean and st. dev. over 100 repeats, when clustering was performed on neurons imaged in all three sessions A, B, and C, n=20,024 neurons.

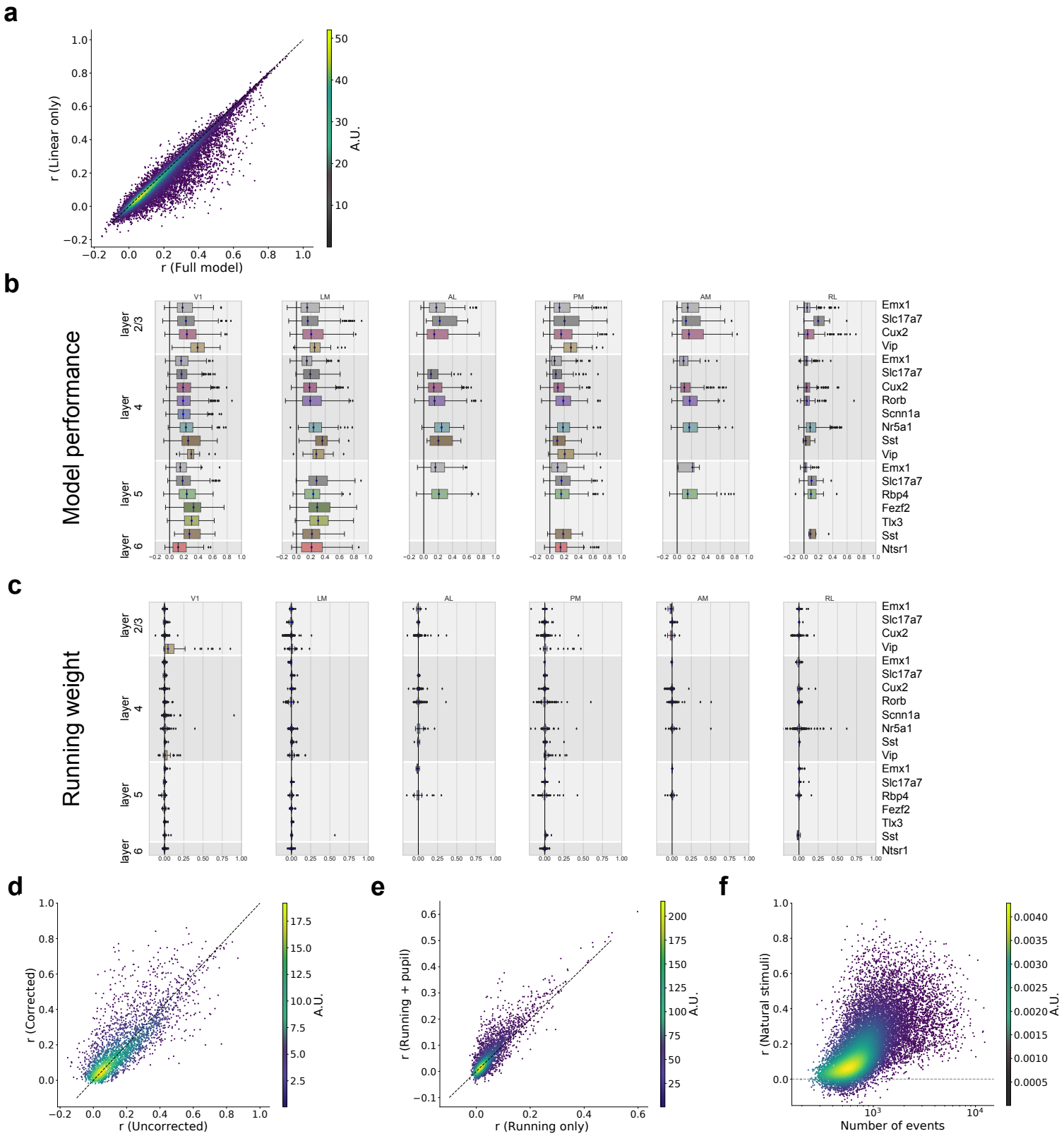


Supplementary Figure 9

Example wavelet models

Four example neurons showing wavelets, temporal filters, distributions of non-zero weights, and performance comparison for both natural and artificial stimuli. We chose exemplar neurons that 1) performed well on both natural and artificial

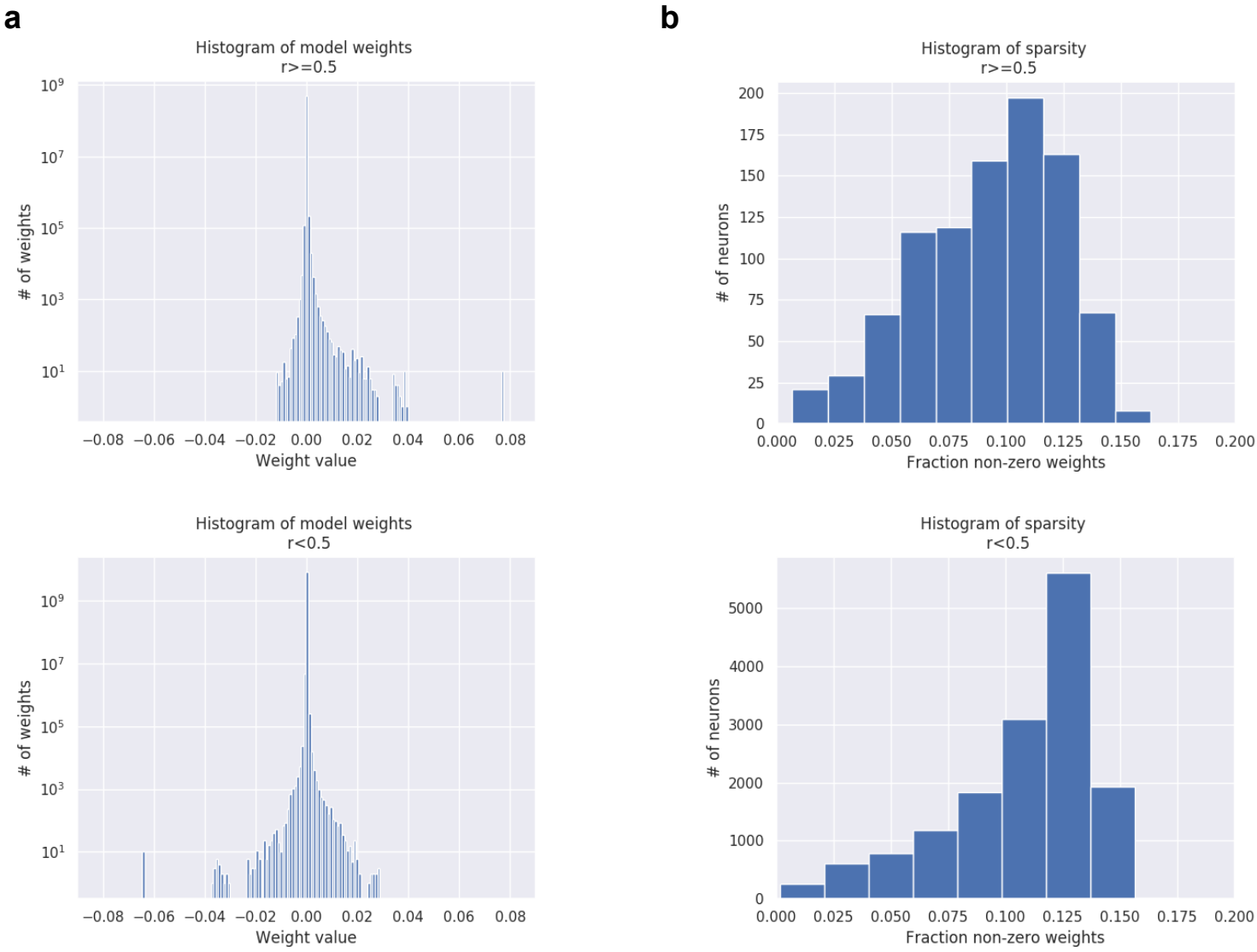
stimuli, 2) performed well on the natural, but not the artificial, stimuli, 3) performed well on the artificial, but not natural, stimuli, 4) performed well on neither. The linear component of each model can be visualized directly. To visualize the quadratic component, the weighted quadratic basis functions must first be converted into an equivalent matrix of second order terms. This matrix is then analyzed by eigenvector decomposition, similar to STC. (far left) example filters showing spatial profiles over time. The top row for each example is the linear filter and the middle and bottom rows show two components of the quadratic part of the filter. (second from left) The envelope of the temporal part of the filters for each wavelet. These are computed by convolving the Gaussian envelope of each wavelet with the temporal weights w(t) (51745 curves; see Methods). Color is a gradient from highest to lowest amplitude. (third from left) The distribution of weights for each wavelet (517450 weights; see Methods). (fourth from left) Examples of predictions compared with neural activity. For each neuron, the top row shows the model fit to natural stimuli and the bottom row shows the model fit to artificial stimuli. The r value is the Pearson's correlation of the predicted response to the measured response.



Supplementary Figure 10
Wavelet model

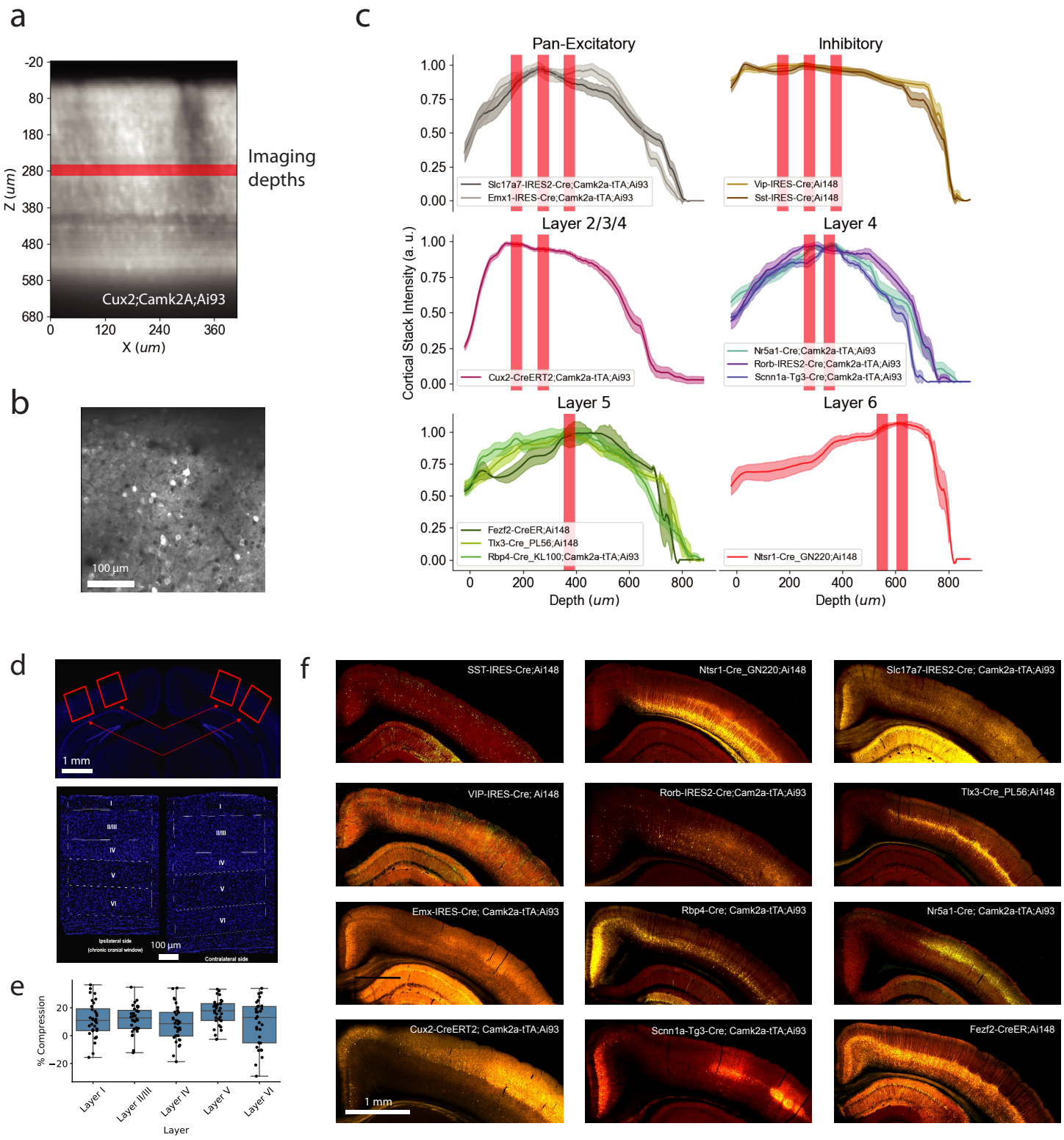
(a) Density plot comparing model performance using only the linear component with the full model performance, including both linear and quadratic components where the r value is the Pearson's correlation of the predicted response to the measured response While some neurons are better predicted with the inclusion of quadratic components, the bulk lies along unity. n=16,092. (b) Box plots of model performance across all areas for the 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. For sample sizes see Extended Data Figure 9. (c) Box plots of the running weights across all areas. For sample sizes see Extended Data Figure 9.

(d) Density plot comparing the model performance when the stimulus was corrected for pupil position with when it was uncorrected. Eye movements tend to be small and have a negligible effect of the model performance. n=2617. (e) Density plot comparing model performance including both running and pupil area as regressors compared to model performance with only running as a regressor. For a small number of neurons, the pupil area improves model performance, but the bulk of neurons lie close to unity. n=3001. (f) Density plot showing the relationship between the number of extracted events and model performance for natural stimuli. n=16,266.



Supplementary Figure 11
Weights for wavelet model

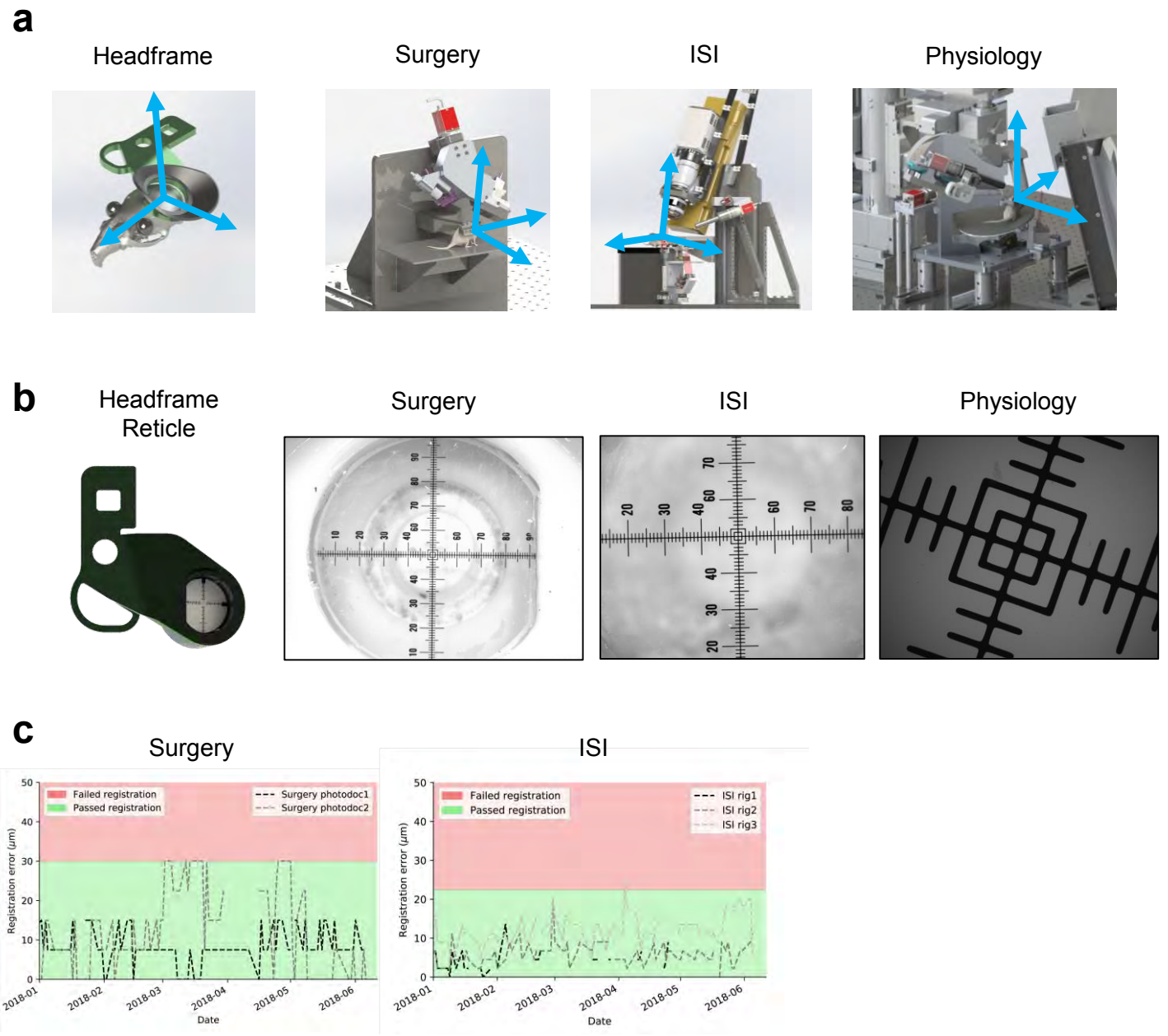
(a) Distributions of weights for the wavelet basis functions for good models (top, r>0.5, where the r value is the Pearson's correlation of the predicted response to the measured response; n=945) and bad models (bottom, r<0.5; n=15321). (b) Distribution of the non-zero weights for the for good models (top, r>0.5; n=945) and bad models (bottom, r<0.5; n=15321).



Supplementary Figure 12

Quantification of cortical layer expression levels across all transgenic mice for mice implanted with cortical glass windows

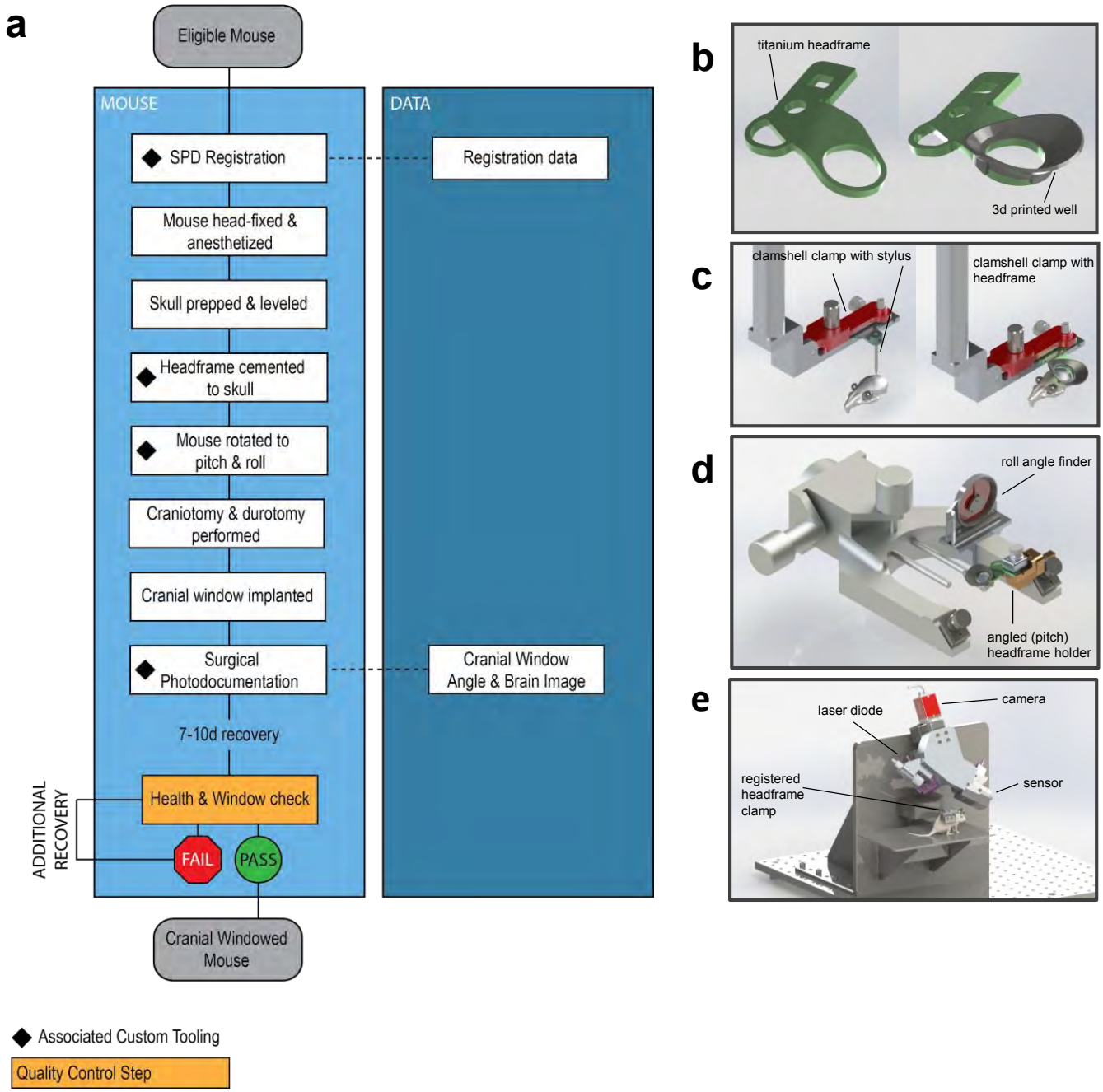
(a) Side view projection (XZ) of a single cortical stack acquired using two photon imaging. All experimental sessions were associated with such a cortical stack. Red bar denotes imaging depth shown in (b). (c) Distribution of labelling intensity across all cortical stacks acquired in V1 for all imaged cre-lines. Red bars denote imaging depths used in each line. N=10 (Slc17a7), 18 (Emx1), 20 (Vip), 20 (Sst), 21 (Cux2), 17 (Nr5a1), 11 (Rorb), 9 (Scnn1a), 2 (Fezf2), 7 (Tlx3), 18 (Rbp4), and 6 (Ntsr1). (d) Brain compression was quantified using DAPI staining (ThermoFisher, D1306) and confocal imaging of entire coronal sections. Cortical layers were annotated and their thickness measured for both the ipsi-lateral and contra-lateral side. (e) Distribution of cortical thickness ratio for all 6 cortical layers between the ipsi-lateral and contra-lateral side. (n=4 measurements of 8 mice). (f) Example expression distribution of all cre-lines as shown on coronal sections imaged with serial two photon imaging.



Supplementary Figure 13

Cross-platform registration for pipeline data collection

(a) All experimental hardware components were custom-designed and assembled in house so that we could register the underlying geometry to a common coordinate system. (b) A registration artifact (a headframe with reticle glued inside the well) defined the imaging coordinate system and all experimental systems (including surgical photo-documentation, ISI, and 2P optical physiology microscopes) were calibrated so they produce data that is referenced within the imaging coordinate system. All imaging datasets were registered to this common coordinate system using shared reticles clamped throughout the data collection pipeline. (c) Registration integrity was monitored by routine measurement of a registration artifact (6 months of data shown). If the registration parameters exceeded tight control limits the system was re-calibrated. We maintained this coordinate system by monitoring the stability of the reticles on a weekly basis across all steps of the data collection pipeline. Any excessive deviations were flagged for further inspection.

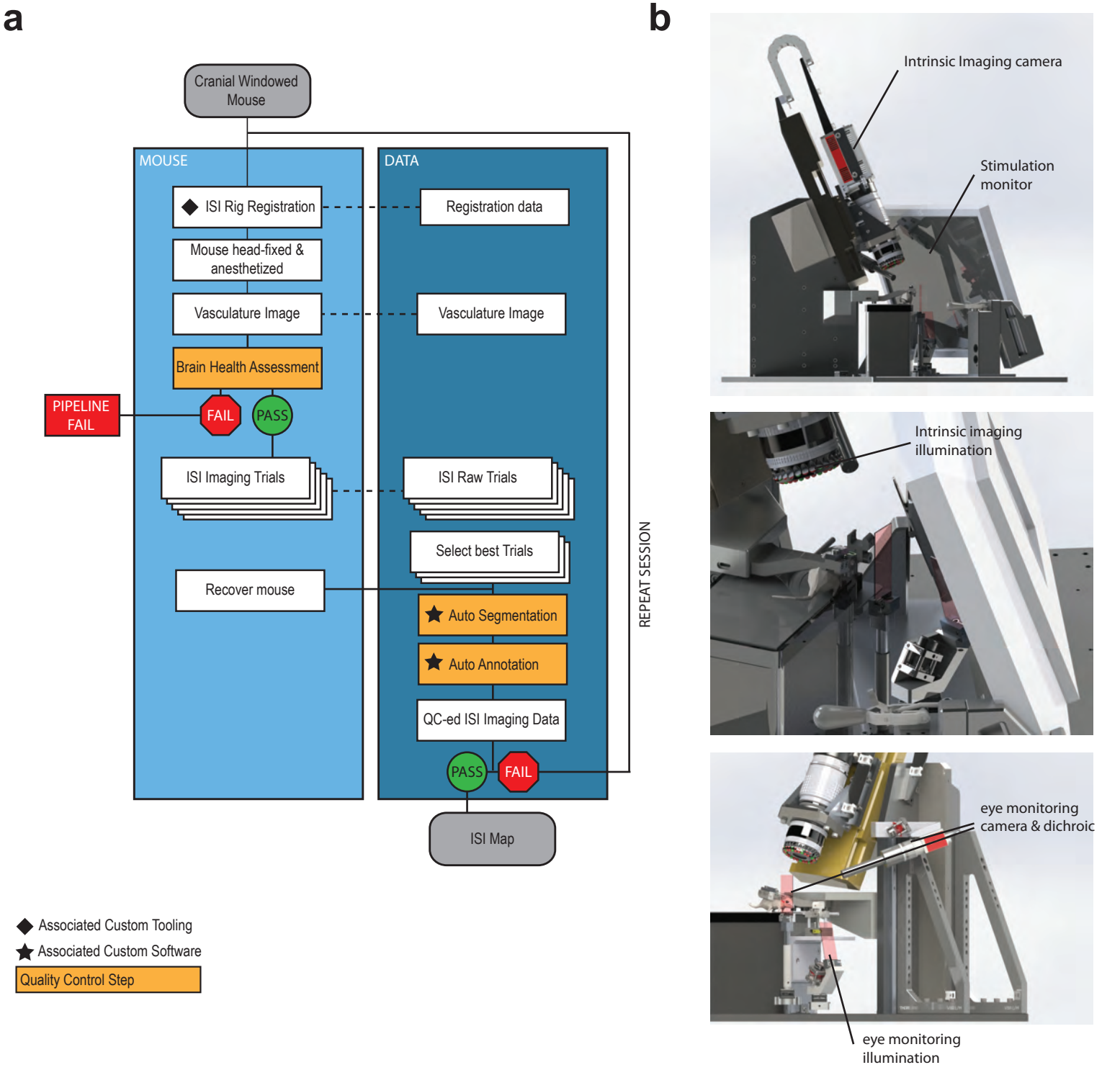


Supplementary Figure 14

A standardized workflow for headframe and cranial window placement

(a) The cranial window surgery consisted of procedural and data collection steps. (b) The surgery was standardized across mice through precise placement of a registerable titanium headframe with objective well. Prior to implantation, a 3D printed acrylic photopolymer microscope well was glued to the titanium headframe with Loctite 406 using a jig to ensure uniform offset between the center of the well and the reference surfaces of the clamp. (c) The stereotax was zeroed on lambda using a custom headframe holder equipped with a stylus affixed to a clamp-plate. The stylus was then replaced with the headframe, which was lowered and affixed to the skull with Metabond. (d) Once dried, the mouse was placed in a custom clamp to position the skull at a rotated angle of 23 and pitch angle of 6, such that visual cortex was horizontal to facilitate the craniotomy. The craniotomy was centered at X = -2.8mm and Y = 1.3mm

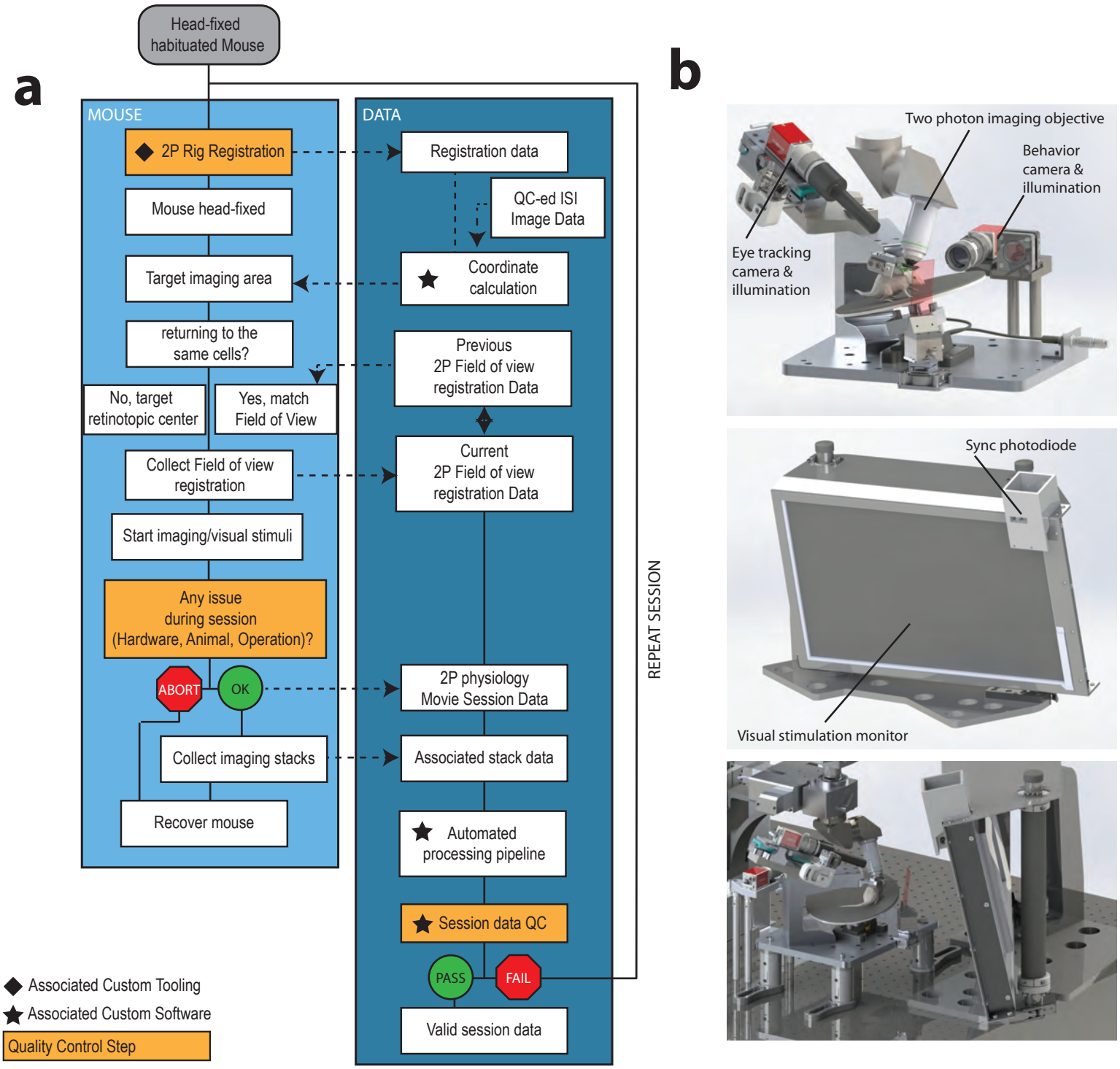
with respect to lambda. A circular piece of skull 5 mm in diameter was removed, and a durotomy was performed. A coverslip stack (two 5mm and one 7mm glass coverslip adhered together) was cemented in place with Vetbond (Goldey, G. J. et al. Removable cranial windows for long-term imaging in awake mice. Nat. Protoc. 9, 2515–2538 (2014). Metabond cement was applied around the cranial window inside the well to secure the glass window. (e) Post-surgical brain health was documented using a custom photo-documentation system to acquire a spatially registered image of the cranial window. One, two, and seven days following surgery, animals were assessed for overall health, cranial window clarity, and brain health.



Supplementary Figure 15

A standardized intrinsic imaging experimental workflow to map the visual cortex at scale

(a) The intrinsic imaging consists of procedural and data collection steps. Data is standardized using custom tools and data quality metrics and control procedure. (b) The brain surface was illuminated via custom LED rings positioned around the imaging objective (middle) and the fluctuation in light reflection was imaged using an imaging camera (top). Periodic stimuli were presented on the stimulation screen to create retinotopic maps for cortical visual areas. (bottom) A camera monitored eye position in anesthetized mice.



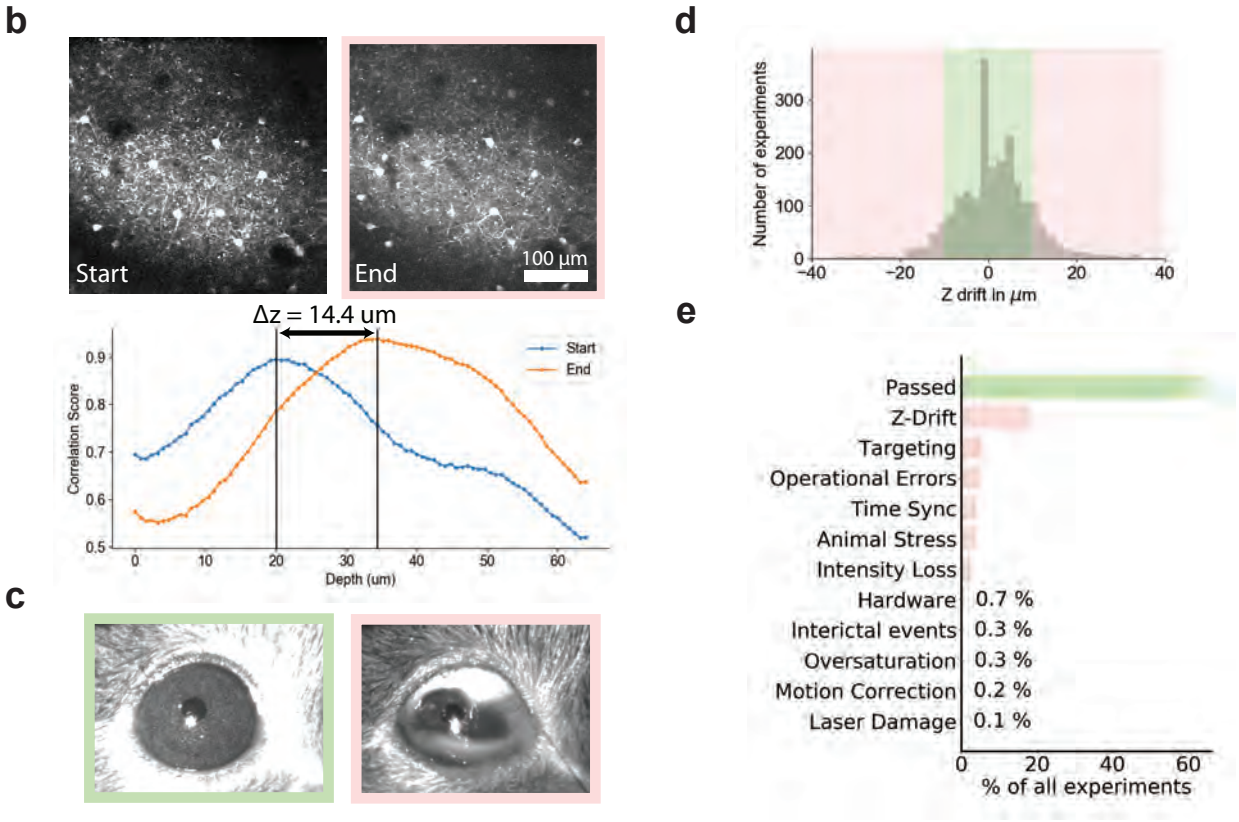
Supplementary Figure 16

A standardized 2-photon calcium imaging experimental workflow

(a) Two photon calcium imaging consists of procedural and data collection steps standardized using a formal experimental workflow. (b-top) Mice were head-fixed using custom behavioral hardware and monitored with behavior and eye tracking cameras. (b-bottom) Visual stimuli were presented on a stimulation screen positioned reliably from session to session. (b-middle) Visual stimulation timing is monitored using a photodiode positioned on top of the screen for every experiment. The experimental workflow integrated tightly controlled experimental procedures with QC metrics and any experiments that did not meet standardized criteria (see Supplemental Figure 8) were re-attempted.

a

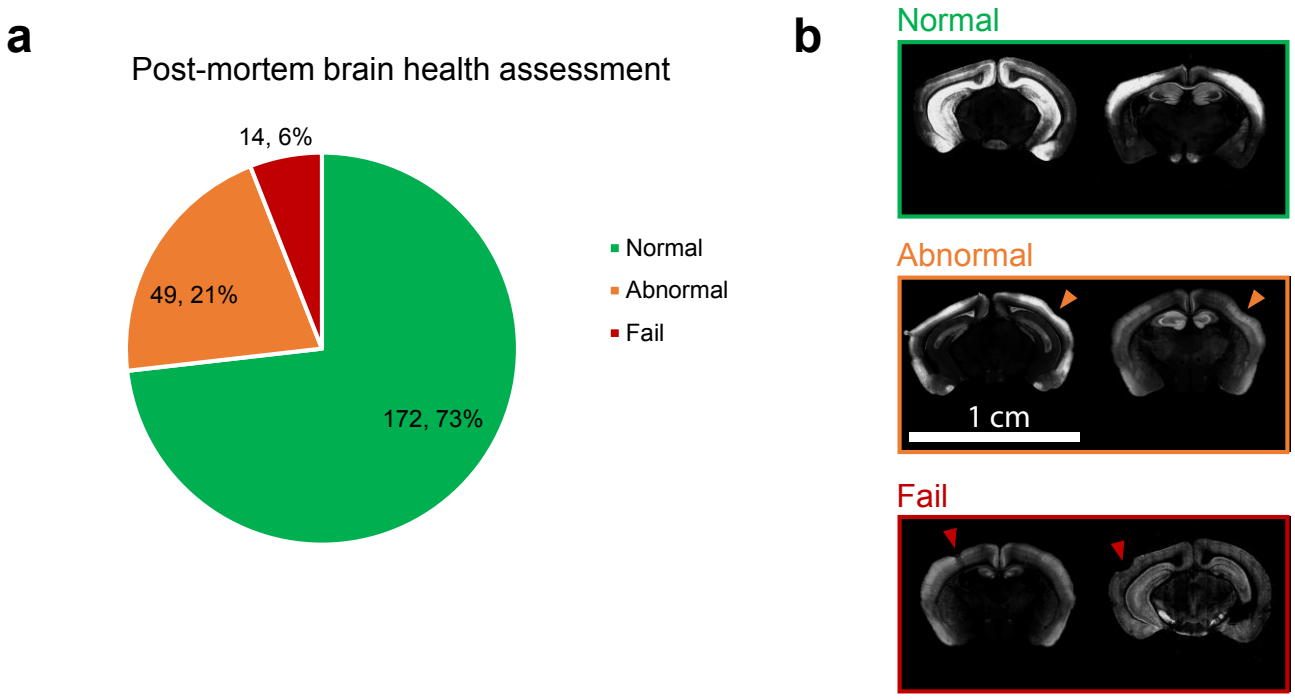
QC evaluation	Metric(s)	Threshold(s)	Notes
z-drift	Z distance between start and end of physiological movie	10 um	Measured using cross-correlation between movie plane and a local z-stack
Animal stress	Presence of foam over the eye, general animal stress	Foam covering pupil	Evaluated using eye tracking movie, body posture movie and general handling
Cell matching	Subjective evaluation of matched field of view between consecutive session	~ more than 50 % cells matching	
Wheel rotation failure	Wheel rotation uncorrelated with behavior movie; Presence of discontinuous spikes in wheel data		
Interictal events	Width and size of whole field calcium events	size larger than 30 % ΔF/F, width between 100 and 300 ms	
Laser damage	Number of discontinuous cellular compartments appearing between sessions	>0	
Bleaching and loss of signal	Average fluorescence from start to end of session	20% drop in baseline intensity	
Abnormal gcamp expression	Depth profile of gcamp	Any abnormal pattern	Evaluated by comparing to reference z stack data
Calcium physiology syncing	Number of dropped frames	1 for Nikon,60 for Scientifica	
Eye tracking syncing	Number of dropped frames	100 frames	
Visual stimulus syncing	Number of long frames	60 frames	Frame duration is plotted over time and number of abnormal frames are tracked
Saturation	Number of satured pixels	1000 pixels	
Laser alignement stability	Power at objective (mW)	>10% change	Trigger maintenance or evaluation of rig



Supplementary Figure 17

2-photon quality control metrics used for Quality Control

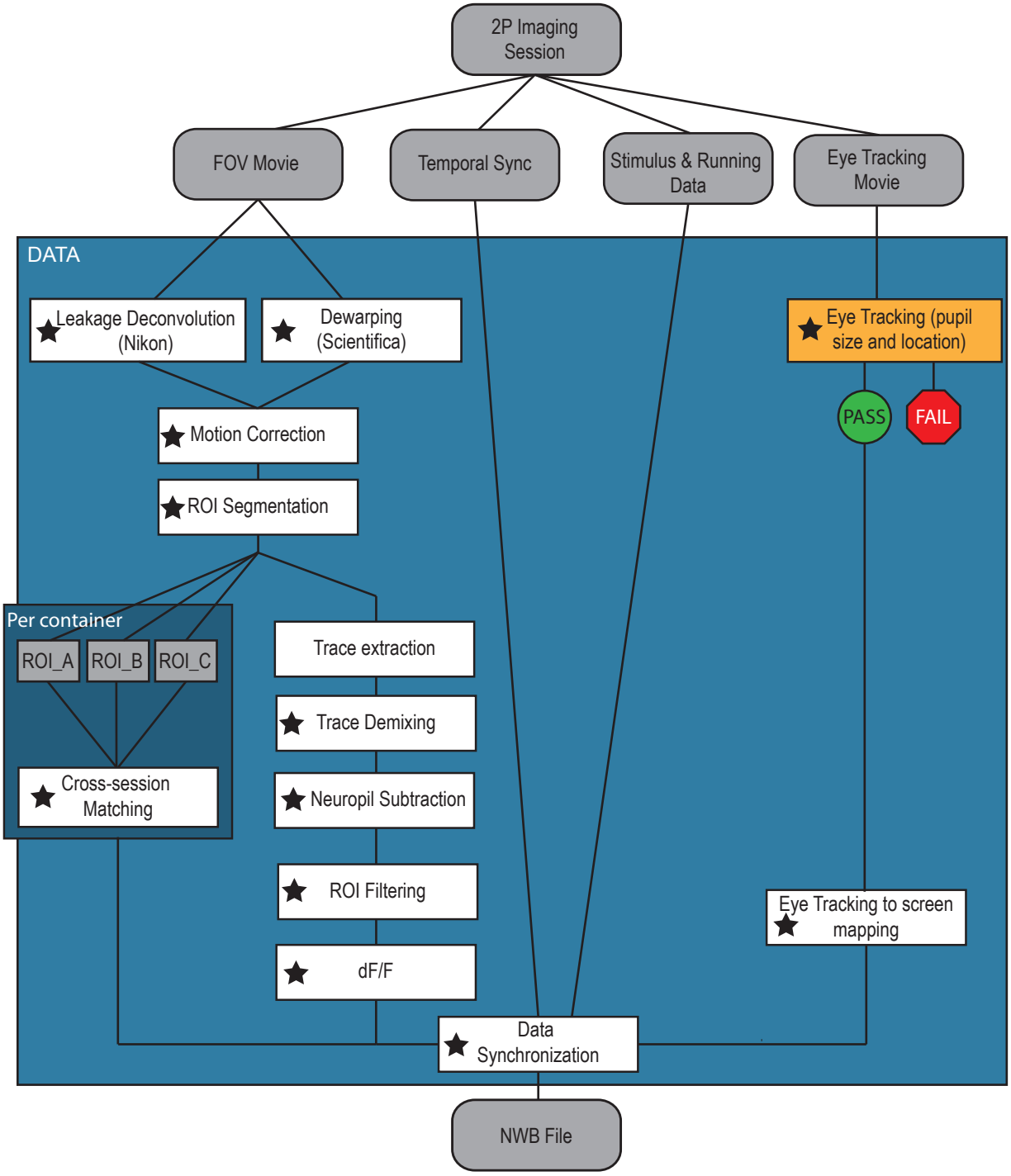
(a) Table of metrics used to validate experiment collected in two photon imaging. (b) Example images at the onset and end of a session showcasing excessive z-drift (>10 µm). Bottom panel in (b) shows the correlation score of all images in a local 60um z-stack with the onset and end image shown at the top. The peak was used as a measure of the actual z-position. (c) Example eye tracking data showing passing (green) and failing (red) signs of stress. (d) Distribution of z-drift across both passed and failed experiments. Green and red areas indicate failure threshold. (e) Distribution of all experimental failures for all individual sessions collected on the 2 photon imaging pipeline.



Supplementary Figure 18

Brain health assessment

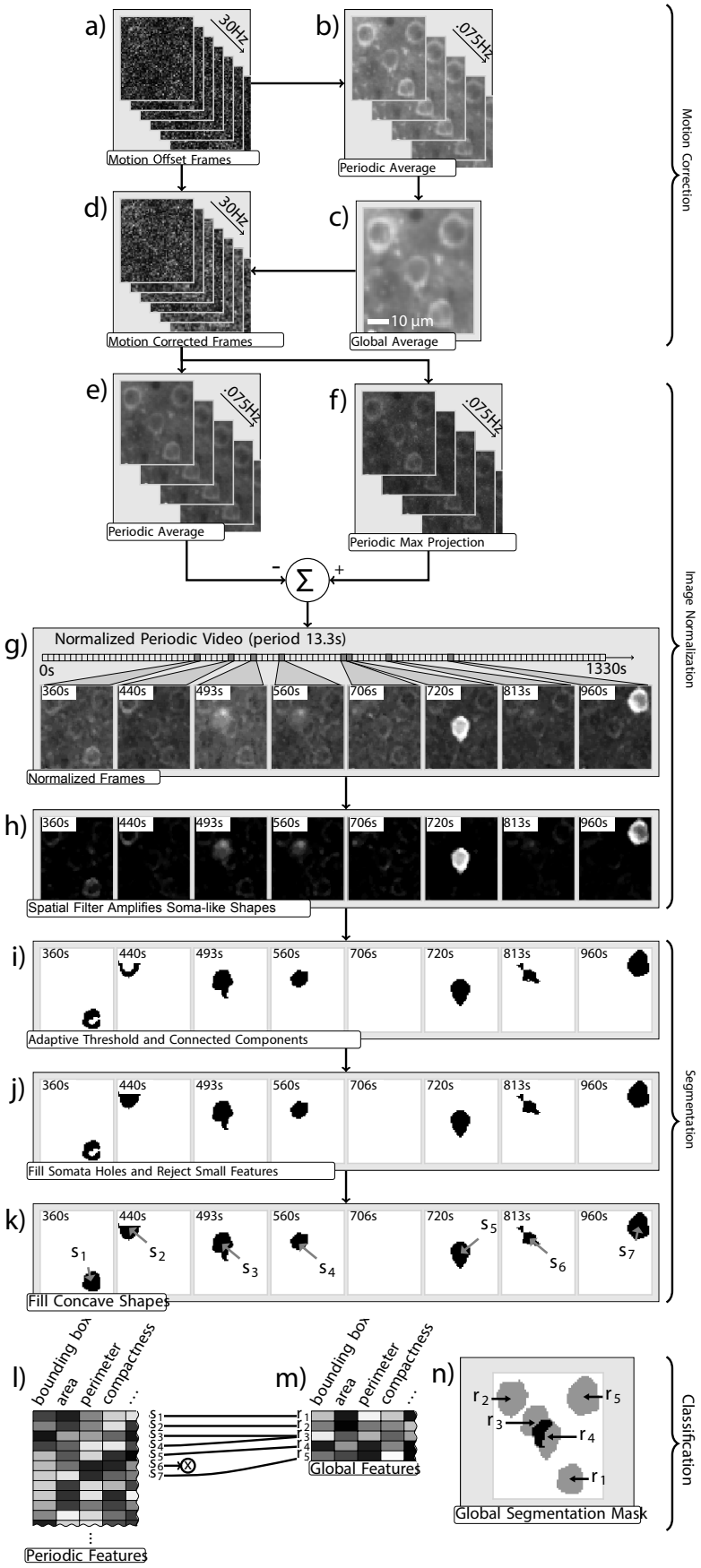
(a) We observed brain abnormalities in approximately 27% of the mice that completed the experimental workflow. Abnormalities were identified during post-mortem histological analysis of the tissue (examples shown in (b)) and had no observable behavioral manifestation during the life of the mouse. Additionally, analysis of a random subset of datasets obtained from mice with abnormalities revealed no differences in physiological responses. In most cases the abnormalities either occurred in the contralateral hemisphere or did not result in observable disruption to the cortical laminar structure (Abnormal). In rarer cases, the abnormalities 1) occurred under or adjacent to the cranial window, 2) disrupted the cortical layers, or 3) resulted in necrosis. In these cases, experimental datasets from these mice were excluded from analysis (Fail). (b) Examples of Normal, Abnormal and Failed brains.



Supplementary Figure 19

A standardized workflow for image processing.

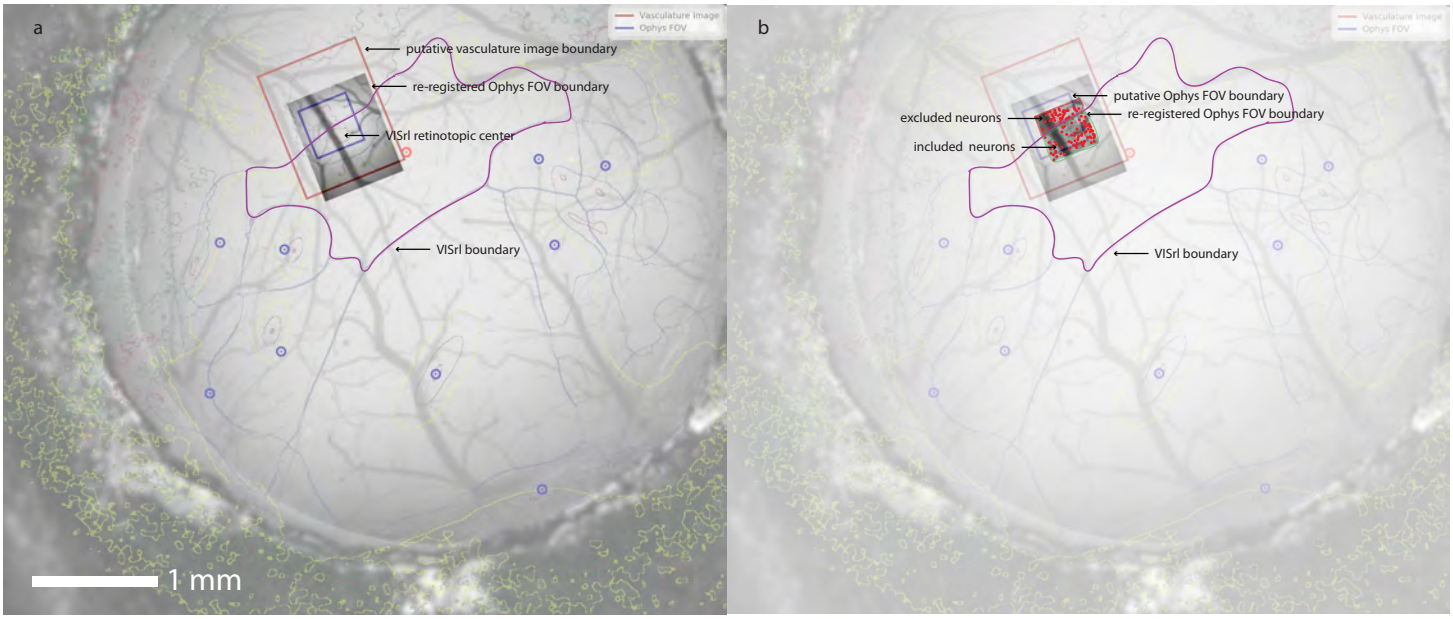
In vivo 2-photon imaging data is processed using a standardized pipeline. Calcium movies were motion corrected and segmented using an automated segmentation algorithm. Within each session, the traces were extracted from the identified ROI, and overlapping ROIs demixed. Across all three imaging sessions in a single experiment, the segmented ROIs were matched across sessions. Following the matching step, ROIs were filtered to select only somatic masks, neuropil contamination was subtracted, and $\Delta F/F$ was computed within each session. These traces were aligned to the stimulus, running, and eye tracking data using the temporal sync that was recorded during the experiment. The eye tracking movie was processed (**Supplemental Figure 24**) and aligned to visual space.



Supplementary Figure 20

Image processing pipeline

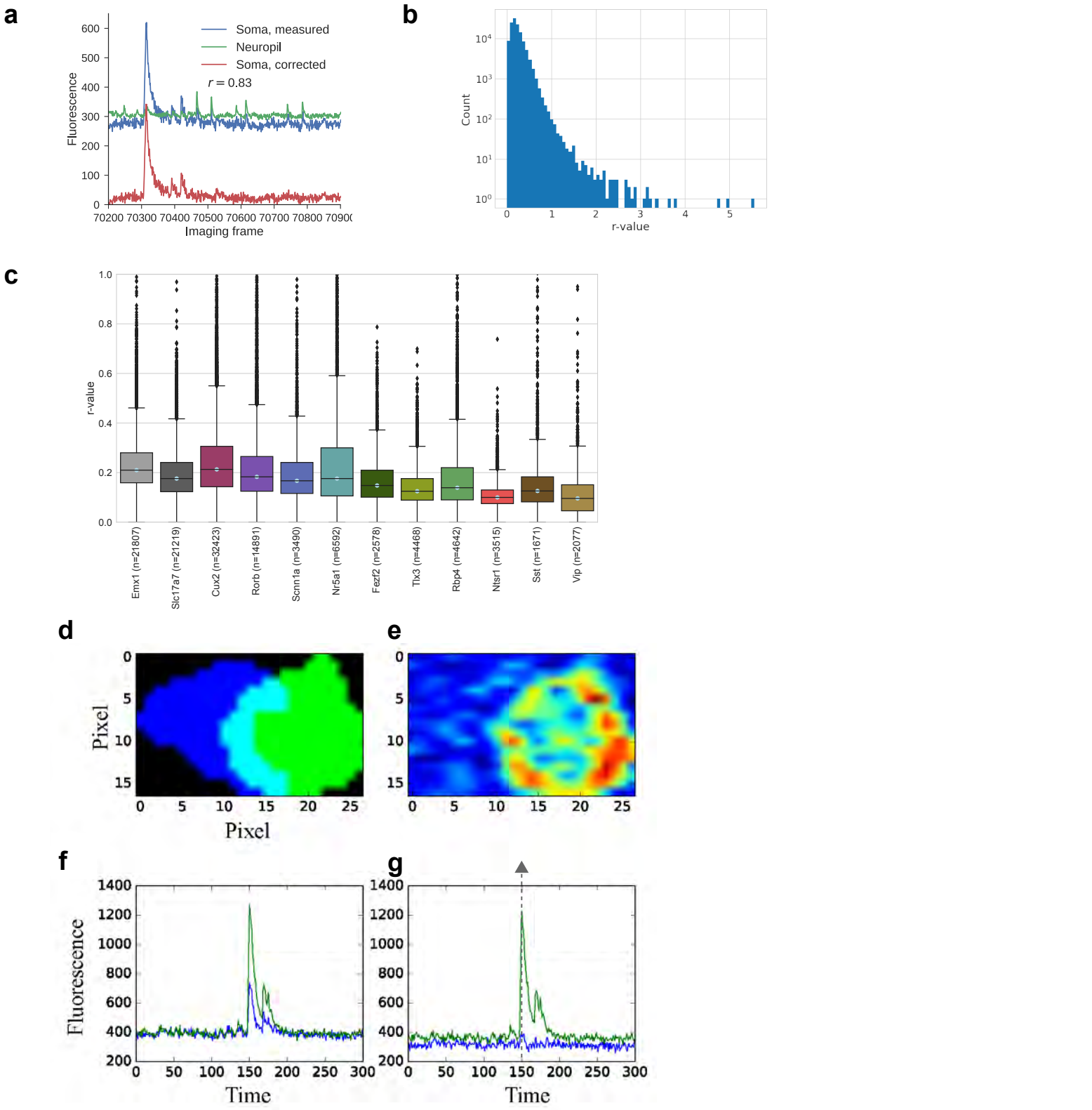
(a) Uncorrected movie. (b) Periodic average of 400-frame set. (c) Global average frame computed by registering the periodic averages together. (d) Motion corrected movie, aligned to the global average, used for downstream processing. (e) Periodic average frame from full motion corrected movie subtracted from (f) periodic maximum projection frame creates (g) normalized periodic frames. Periodic frames after (h) spatial filtering, (i) adaptive thresholding, and morphological operations (j, k). Features of each ROI mask identified in each period are computed (l). Heuristic decisions combine duplicate masks and label non-somatic masks for filtering (m), yielding the final ROI masks (n).



Supplementary Figure 21
Targetting refinement for RL

(a) We confirmed the registration between the ISI map and the FOV in two steps: A vasculature overview micrograph, acquired with the bright field camera of the 2P microscope just before imaging, was registered to the ISI map (retinotopic center located within red boundary). (b) A surface 2P image was registered to the vasculature overview micrograph. Neurons with medians located on the somato-sensory side of the RL boundary were excluded from further analysis. Note how the density of ROIs (red) post registration refinement is consistent with shadowing effects from the surface vasculature. All registrations were performed by fitting similarity transforms to 10-20 operator-selected control points (using MATLAB's cpselect function).

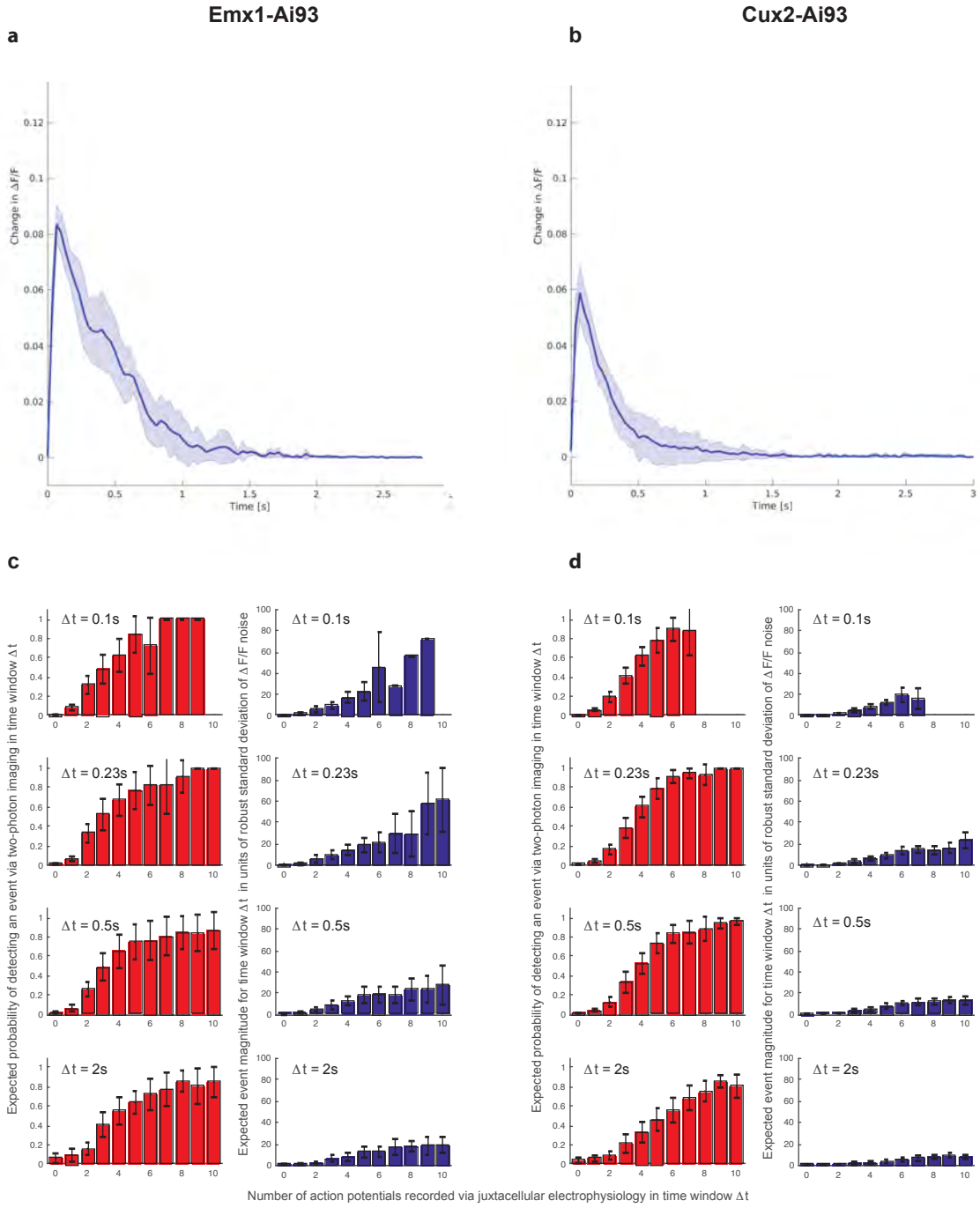
Supplemental Figure 10



Supplementary Figure 22

Neuropil subtraction and demixing

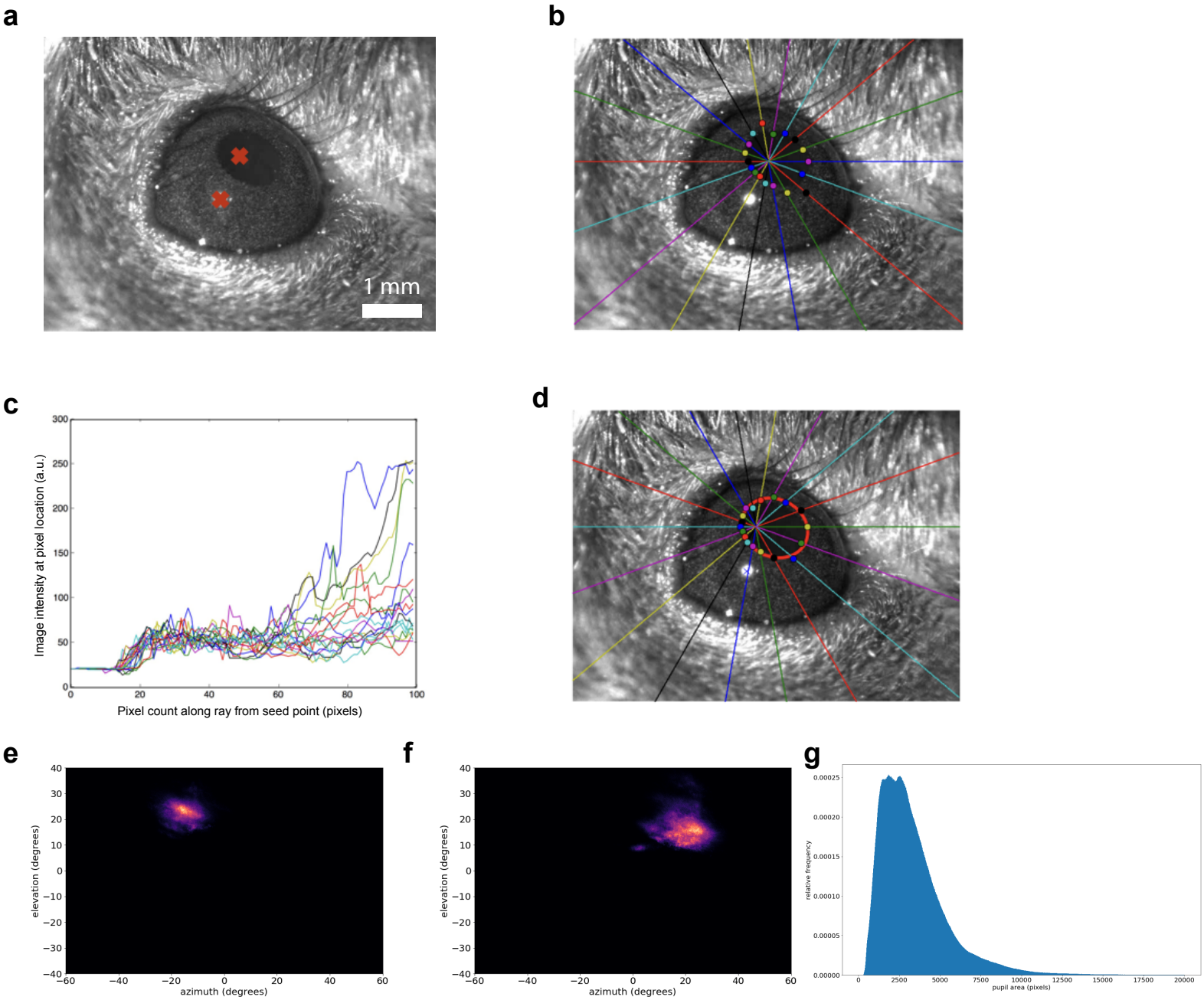
(a) Example of neuropil subtraction showing somatic fluorescence, neuropil fluorescence, and the corrected trace. r is the contamination ratio (see Methods). (b) Histogram of computed r-values for all cells in the dataset. (c) Distributions of r-values for each Cre line. The box shows the quartiles of the data, and the whiskers extend to 1.5 times the interquartile range. Points outside this range are shown as outliers. (d) Example of two overlapping ROIs (blue and green, with overlapping pixels in cyan). (e) Non-uniform distribution of fluorescence across a ROI, used for demixing. (f) Raw fluorescence traces from each ROI, centered on the time point in e. (g) Demixed traces for each ROI, centered on the time point in e.



Supplementary Figure 23

Evaluation of event detection

(a) Single-spike calcium response extracted from calcium imaging data acquired simultaneously with juxtacellular electrophysiology in vivo (15 cells in Emx1-Ai93 mice; error regions are 2x sem over 3 folds). (b) Single-spike calcium response extracted from calcium imaging data acquired simultaneously with juxtacellular electrophysiology in vivo (20 cells in Cux2-Ai93 mice; error regions are 2x sem over 3 folds). (c) The red bar graphs indicate the expected probability of detecting an event via L0 event detection (y-axis) as a function of the number of action potentials juxtacellularly recorded (x-axis) in a given time window (subplot title) based on 15 cells in Emx1-Ai93 mice. The blue bar graphs indicate the expected event magnitude reported by the L0 event detection (y-axis) as a function of the number of action potentials juxtacellularly recorded (x-axis) in a given time window (subplot title) based on 15 cells in Emx1-Ai93 mice. (d) As in c but based on 20 cells in Cux2-Ai93 mice. The error bars in the graphs represent twice the standard error of the mean across cells. The false positive rate, as measured by the probability of detecting an event given no spike in the simultaneously recorded membrane potential in a window of given length, is generally low but increases with detection window length. Bursts within short time windows correspond to high instantaneous firing rates and result in detected events with high likelihood (e.g. >5 spikes within 100 ms result in events detected with > 80% chance). Bursts within longer windows can have lower instantaneous firing and, thus, for a given spike count, the average detection probability decreases with increasing detection window length. Expected event magnitudes tend to decrease for longer event detection windows. The size of the error bars decreases with increasing window length because, for longer windows, more instances of any given spike count were observed.



Supplementary Figure 24

Eye tracking and characterization

Videos of the ipsilateral eye (relative to the monitor) were used to extract pupil location and size. Coordinates for eye position were extracted independently for each frame of the eye position movie. A variant of the star-burst algorithm was used (Li, Winfield and Parkhurst, 2005; Zoccolan, Graham and Cox, 2010). This algorithm fits an ellipse to the pupil or corneal reflection (CR) area. (a) The initial seed points for the pupil and corneal reflection were were identified via a convolution with a black square (pupil) or a bright square against a black background (CR). (b) 18 rays were drawn starting at the seed point, spaced 20 degrees apart. For each ray, the initial 10 pixels along that ray were averaged to define a_0. (c) We identify a candidate boundary point for the pupil by choosing the pixel along the ray that first exceeds the threshold F*a_0, where F is an input parameter to the algorithm (~1.5, varies by experiment). For the CR, we look for the first point below this threshold. (d)A RANSAC algorithm is used to fit the ellipse from the candidate boundary points using linear regression with a conic section constraint. The fit parameters from the regression are converted into the five ellipse parameters: x,y location of the center, the major and minor axis sizes, and the angle of rotation with respect to the x-axis.

The pupil and CR ellipse fit parameters are converted into coordinates for the location of the pupil in a coordinate system centered in the mouse eye, which is assumed spherical (an approximation) and thus acts as a spherical mirror. The location of the CR in the mouse eye coordinate system is thus a vector of length r_0/2 in the direction of the LED (where r_0 is the radius of the mouse eye). This is used along with the camera and monitor positions to reconstruct the pupil position elevation and azimuth relative to the center of the monitor.

The pupil area is reported as the area of the ellipse fit to the pupil region: pi*major_axis^2. In some cases the algorithm did not find appropriate parameters for the ellipses describing the pupil or corneal reflection. In other cases, outlying values were detected (based on overall area or discontinuous jumps in the location, for example) that are clearly not physical values of the pupil or corneal reflection. In these cases, the pupil position is reported as NaN.
(e) Eye position for the first monitor position. (f) Eye position for the second monitor position. (g) Histogram of pupil area.