

**Neuron, Volume 84**

**Supplemental Information**

**Pupil Fluctuations Track Fast Switching  
of Cortical States during Quiet Wakefulness**

Jacob Reimer, Emmanouil Froudarakis, Cathryn R. Cadwell,  
Dimitri Yatsenko, George H. Denfield, and Andreas S. Tolias

## Supplemental Data

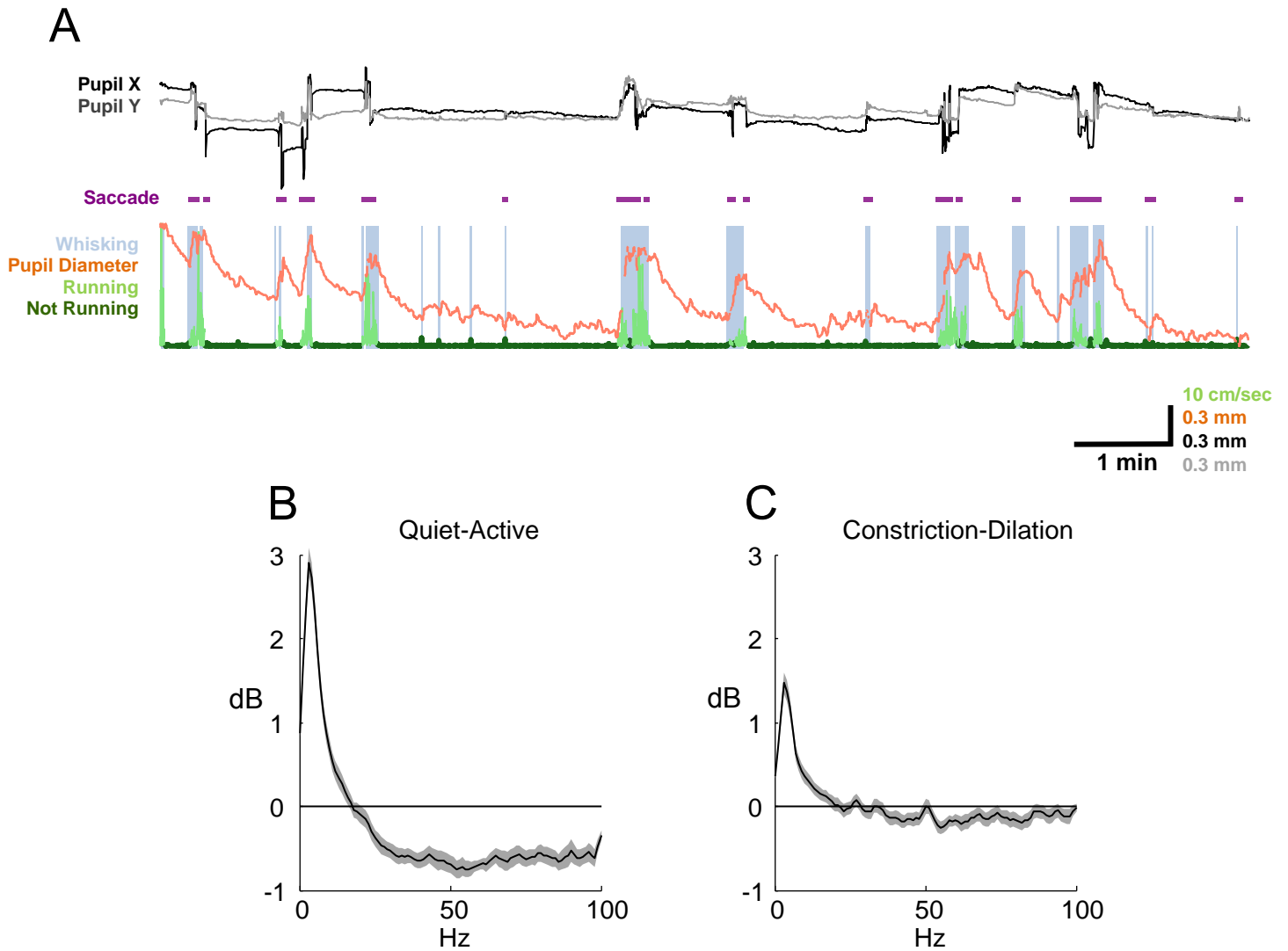
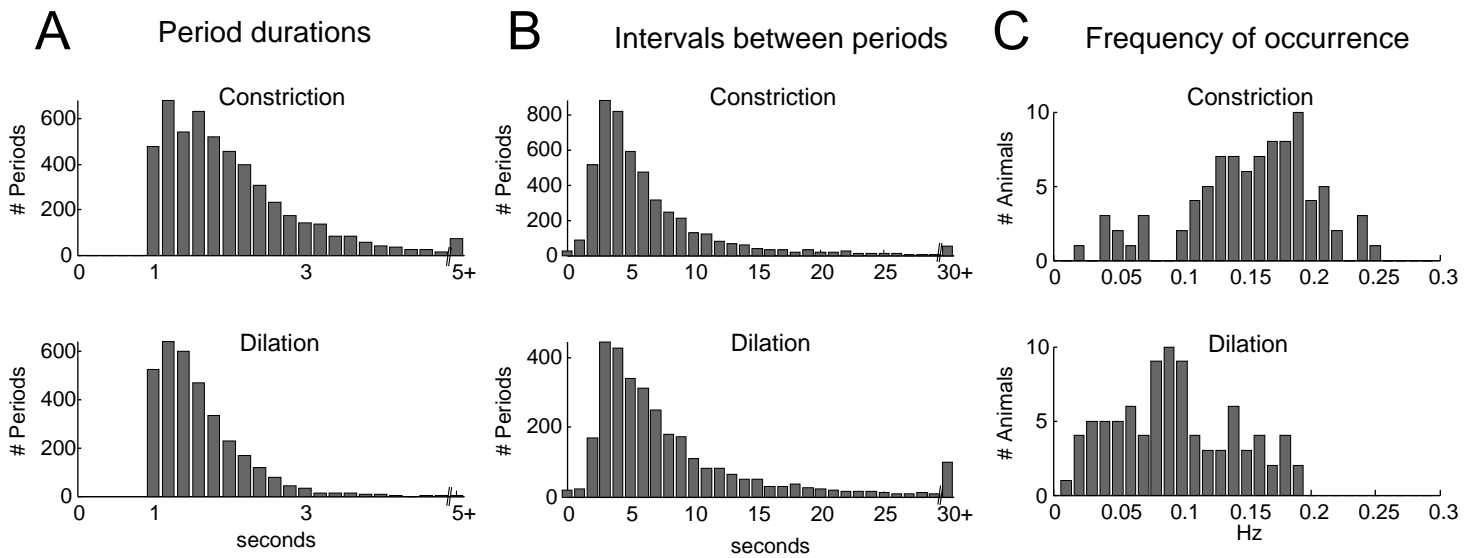


Figure S1, related to Figure 1. Saccades and power spectra during running and quiet wakefulness.

(A) Example traces of simultaneously recorded treadmill speed (running in light green, quiet in dark green), pupil size (orange), whisking periods (blue rectangles) and horizontal (x, black) and vertical (y, gray) eye movements. Detected saccades are indicated in purple below the

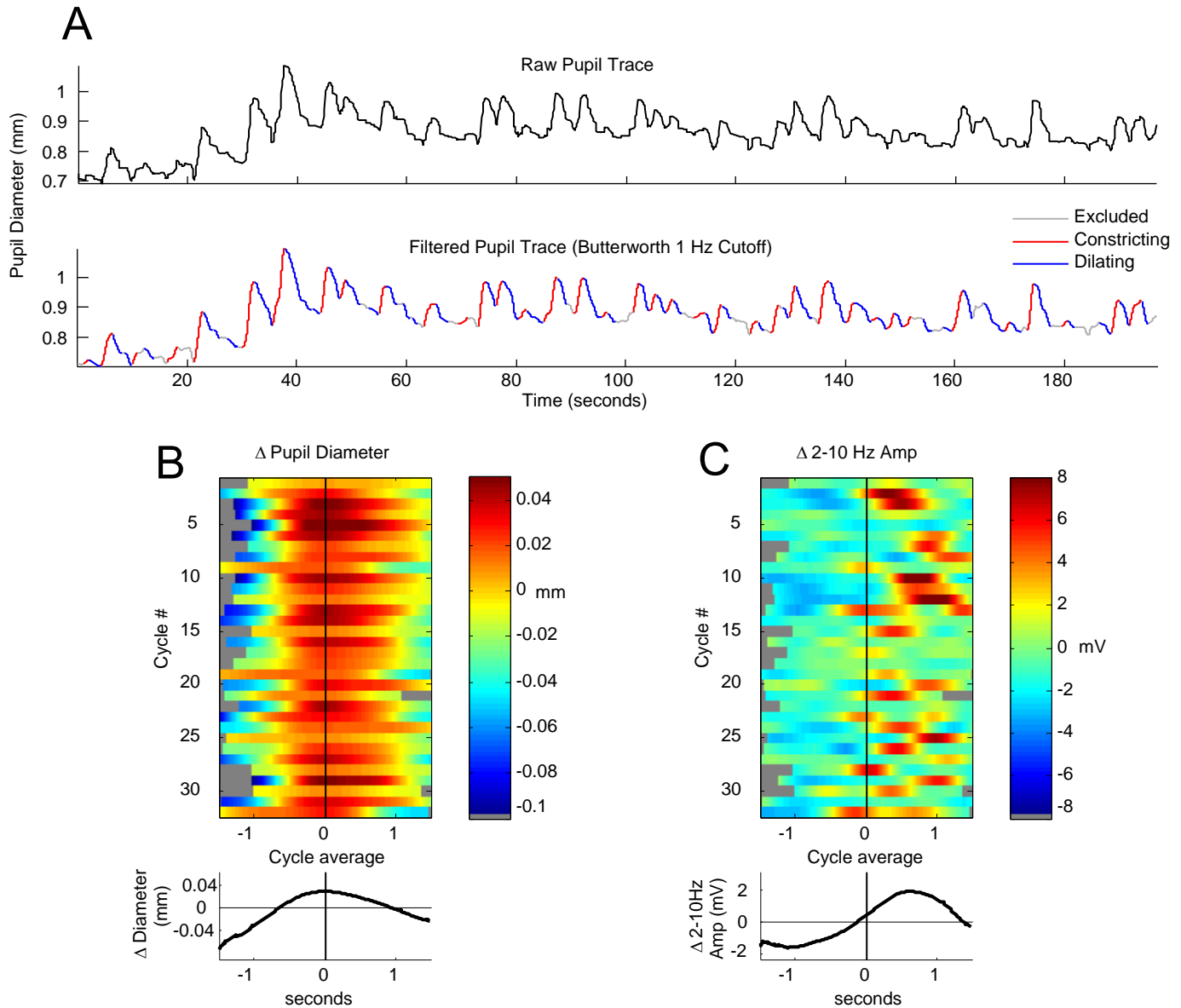
position traces. Saccades are much more frequent during running. Brief periods of missing data in the pupil diameter is due to blinks or frames where the exact diameter and position of the pupil could not be defined due to blurring during a saccade. **(B)** Ratio of mean  $V_m$  power spectra between quiet wakefulness and active periods of whisking and running. Lower-frequency oscillations are enhanced during quiet wakefulness, especially in the 2-10 Hz range; and there is a wide-band increase in gamma frequencies during active behavior. ( $V_m$  power spectra for each condition estimated using Welch's method; mean  $\pm$  SEM over  $n=52$  cells from both S1 and V1). **(C)** Ratio of mean power spectra during constriction and dilation in quiet wakefulness. Power in the lower frequencies in the 2-10 Hz range is higher during constriction, and there is a modest wide-band power increase in higher gamma frequencies during dilation (mean  $\pm$  SEM,  $n=54$  cells).



**Figure S2, related to Figure 2. Temporal characteristics of dilating and constricting periods during quiet wakefulness.**

Dilation and constriction periods were defined by extrema in the low-pass (4<sup>th</sup> order Butterworth, 1 Hz cutoff) filtered pupil trace. Only periods lasting more than one second with an absolute mean dilation or constriction rate greater than 10  $\mu\text{m/s}$  are included in this figure and all our analyses, as our ability to reliably detect fluctuations smaller than this was limited by the shape and sharpness of the pupil boundary and the camera's resolution and frame rate. These periods comprised approximately 55% of our recordings during quiet wakefulness, and removing this threshold did not qualitatively change our findings. **(A)** Histogram of detected period durations. In general, constricting periods (top) were longer in duration (mean: 2.0 s, median 1.8 s) than dilating periods (bottom; mean: 1.6 s, median 1.5 s). **(B)** Histogram of intervals between sequential detected constricting (top) and dilating (bottom) periods (mean interval between constriction periods: 6.3 s, median: 4.6 s; mean interval between dilation periods: 8.4 s, median: 5.8 s). **(C)** Histogram of average frequency of detected constricting (top)

and dilating (bottom) periods for  $n=84$  mice, including some animals not used elsewhere in this study. On average during quiet wakefulness, constricting periods were detected at approximately 0.15 Hz while dilating periods were detected at approximately 0.09 Hz.



**Figure S3, related to Figure 2. Single-cell example of raw data showing state change over multiple cycles of dilation and constriction.**

(A) Unfiltered continuous trace of pupil diameter over time (top), and filtered trace (bottom). Dilating (red), constricting (blue), and excluded (grey) periods are indicated on filtered trace. (B, top) Change in pupil diameter for consecutive individual dilation and constriction cycles from trace in (A). (B, bottom) Mean change in pupil diameter averaged over all cycles. (C, top)

Corresponding segments of 2-10 Hz Hilbert amplitude. Entire Hilbert amplitude trace is band-pass filtered between 0.1 Hz and 1 Hz, and segments are collected corresponding to each cycle in **(B)**. **(C, bottom)** Mean 2-10 Hz amplitude change averaged over all cycles. Note the similarity in the mean change in 2-10 Hz amplitude for this single cell example with the phase-binned average in Figure 2D.

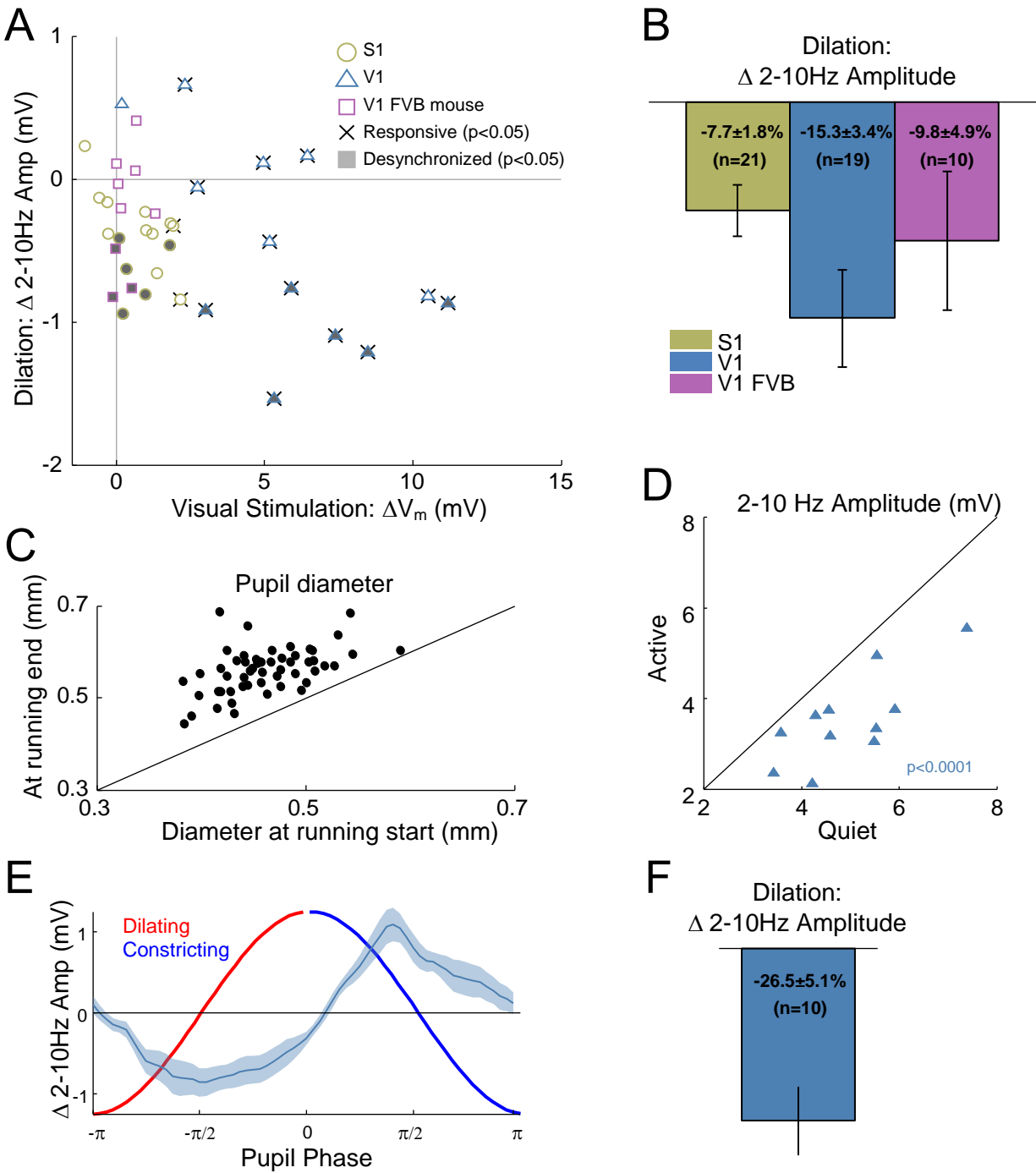
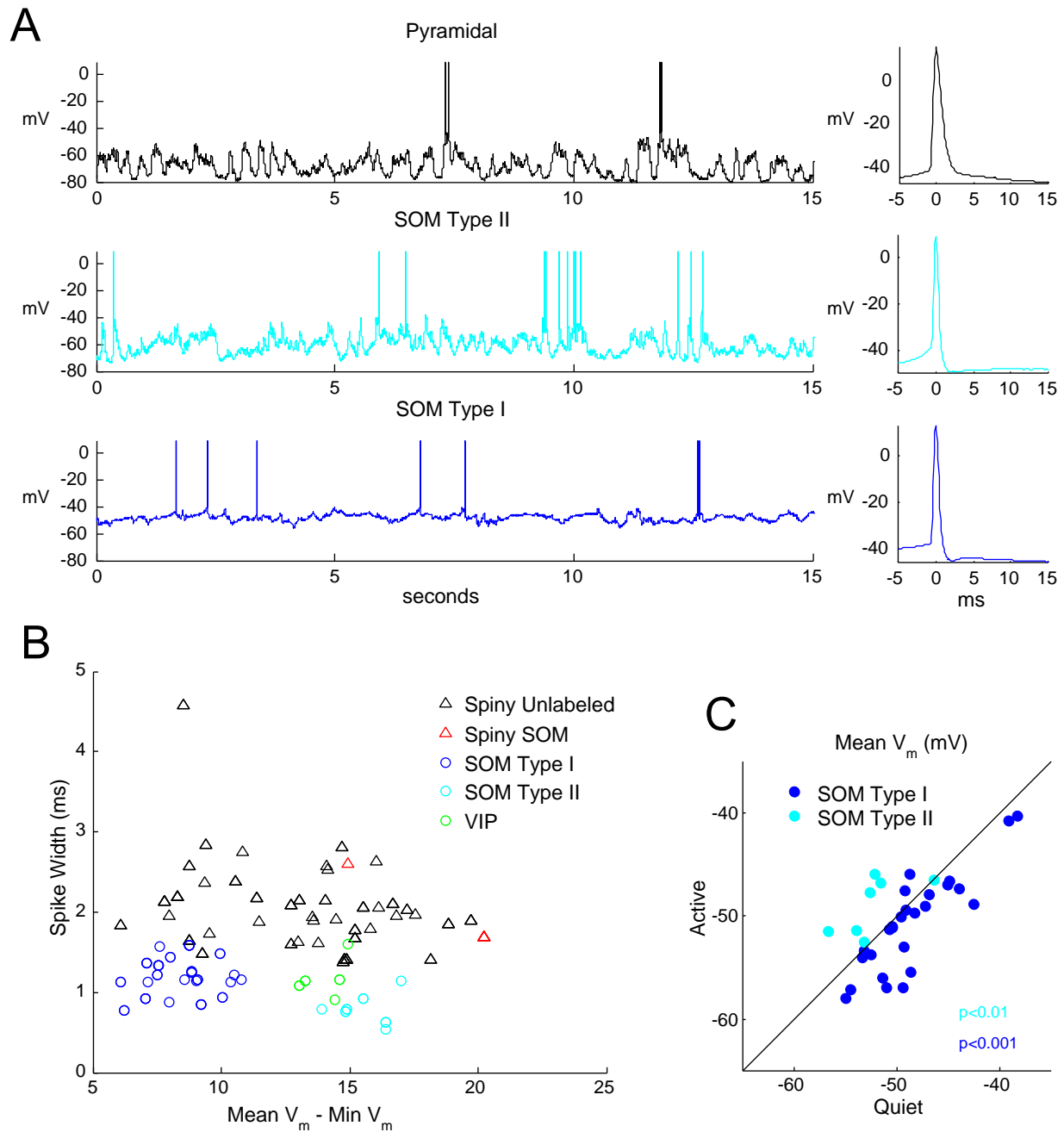


Figure S4, related to Figures 1 and 2. Relationship between pupil dilation and  $V_m$  dynamics persists when small whisker or treadmill movements are excluded, and also in mice habituated to the treadmill.



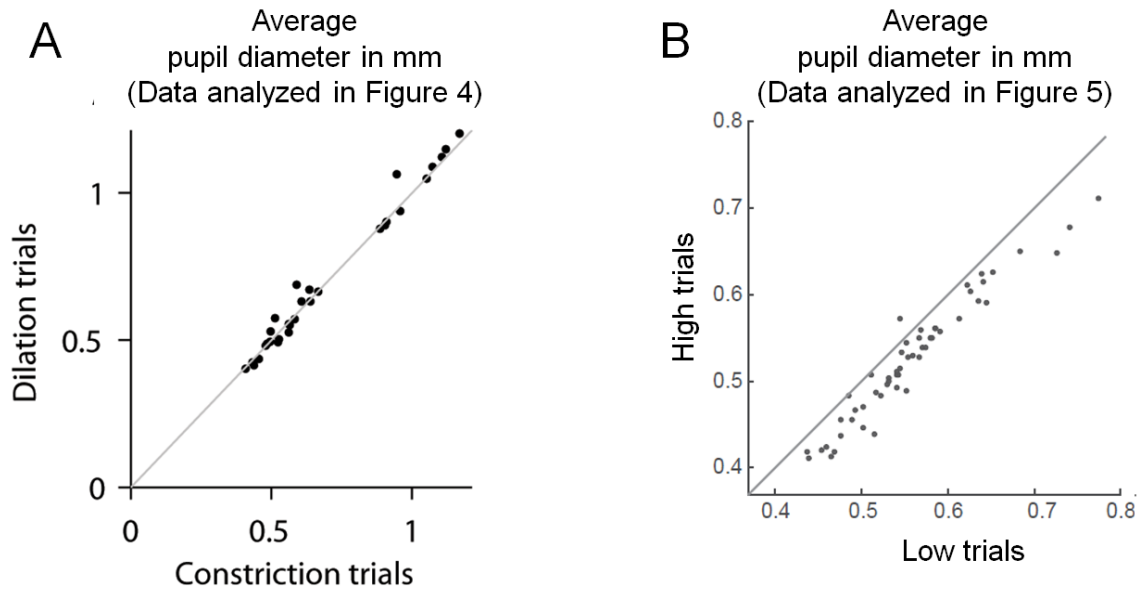
(A) Same plot as Figure 2E, after removing all periods of detectable whisker or treadmill movements. Visual responsiveness (X-axis) plotted against desynchronization during dilation (Y-axis) for cells from S1, V1, and V1 in FVB mice. Because some data have been excluded, fewer cells show significant individual effects. (B) Overall, cells are still robustly desynchronized during dilation in S1 ( $7.7 \pm 1.8\%$  reduction in 2-10 Hz Hilbert amplitude during dilation, mean  $\pm$  SEM,  $P < .01$ ) and V1 ( $15.3 \pm 3.4\%$  reduction,  $P < .01$ , t-test). The effect of dilation was in the same direction for FVB cells ( $9.8 \pm 4.9\%$  reduction), but was no longer significant. (C) Pupil diameter increases significantly during running in mice habituated to head restraint on the treadmill for five days for 30 minutes per day ( $n=113$  running epochs in 3 mice;  $102 \pm 5 \mu\text{m}$ , mean change  $\pm$  SEM,  $P < .0001$ , paired t-test). (D) Similar to naïve mice, in habituated mice the amplitude of 2-10 Hz oscillations is reduced during active periods of whisking and running compared to quiet wakefulness ( $-1.5 \pm 2 \text{ mV}$ ,  $P < .001$ ; paired t-test;  $n=10$  cells in 3 mice; experiments performed after habituation on the treadmill for 30 minutes per day  $\times$  5 days). (E,F) There is a strong relationship between pupil dilation/constriction and low-frequency oscillations in habituated mice, similar to naïve mice (2-10 Hz Hilbert amplitude reduced during dilation by  $26.5 \pm 5.1\%$ ,  $p < .01$ , t-test).



**Figure S5, related to Figure 3. Two types of SOM+ labeled cells based on electrophysiological properties have different functional properties.**

(A) Example traces (left) and average spike waveform (right) for a pyramidal cell (top, black), a SOM+ labeled “Type II” cell (middle, cyan), and a SOM+ labeled “Type I” cell (bottom, blue).

The Type II cell displays large fluctuations in  $V_m$  similar to a pyramidal cell, but has a narrow spike waveform. The Type I cell displays comparatively much lower membrane potential variance, and also has a narrow spike waveform. **(B)** Spike width plotted against the difference between the mean and minimum spontaneous  $V_m$  for cells of multiple types. Type I (blue) and Type II (cyan) SOM+ cells form two distinct clusters. Pyramidal cells (black) and VIP+ cells (green) are included for reference. Two SOM+ labeled cells with a pyramidal (spiny) morphology likely due to leakiness in the Cre line are indicated in red. **(C)** SOM Type I and Type II cells respond differently during active behavior (running and whisking). Type I cells are inhibited by running ( $P < .001$ ;  $n = 27$  cells), while Type II cells are depolarized ( $P < .01$ ;  $n = 7$  cells; paired t-test).



**Figure S6, related to Figures 4 and 5.** Pupil diameter is not larger during dilating or high trials, when visual responses are enhanced. **(A)** Pupil diameter is not significantly different for dilating trials compared to constricting trials for analysis in Figure 4. Each dot represents an individual imaging site (n=30 sites). **(B)** For analysis in Figure 5, pupil diameter is smaller during High trials (defined by dilation/constriction rate for each site), when visual responses are enhanced. Each dot is average over movie bins for each site (n=53 sites,  $P < 0.001$ ).

**Movie S1, related to Figure 1.** Pupil fluctuations during quiet and active periods. Fifty seconds of synchronized video from pupil and whisker cameras played at actual speed. Dilating and constricting periods are indicated by a red or blue outline, respectively. The movie begins during a period of quiet wakefulness. In the second half of the recording there are two active periods with locomotion (“RUNNING”) and whisking (“WHISKING”); these and similar periods are excluded from all analysis of quiet wakefulness. Smaller detected whisker movements and treadmill movements labeled (“EXCLUDE”) are excluded from the analysis in Figure S4, above.

**Movie S2, related to Figure 2.**  $V_m$  transitions between synchronized and desynchronized states during quiet wakefulness. Forty seconds of video of the eye, with the simultaneous filtered 2-10Hz Hilbert amplitude of  $V_m$ . Dilations and constrictions are indicated by a blue or red outline, respectively, and red or blue segments of the amplitude trace. The low-frequency amplitude tends to peak during pupil constriction. Black vertical line indicates time of current movie frame in  $V_m$  trace.

## Supplemental Experimental Procedures

### Animals and Surgery

All procedures were carried out in accordance with the ethical guidelines of the National Institutes of Health and were approved by the Institutional Animal Care and Use Committee (IACUC) of Baylor College of Medicine. In this study we use data from a total of 51 mice, aged 6 weeks to 6 months. For the patching experiments in Figures 1-3, we used 38 total mice. These included 34 C57Bl/6 mice (16 males, 22 females). Several of these mice were SST-Cre/Ai9 (n = 16), VIP-Cre/Ai9 (n = 6), or PV-Cre/Ai9 (n = 8) crosses on a C57Bl/6 background, where these interneuron subtypes were labeled with the fluorescent marker tdTomato. We also used 4 female FVB mice (two 8-week old and two 4-month old mice) for patching experiments. For the imaging experiments in Figures 4 and 5, we used an additional 13 male C57/Bl6 mice (7 mice for Figure 5, and 6 mice for Figure 4 including 1 PV-Cre/Ai9 mouse and 1 PV-Cre/Ai27 mouse; a small number of labeled cells in these recordings were not analyzed differently. Cre lines, reporters, and FVB mice are all from Jackson Labs as follows:

- SOM-Cre: <http://jaxmice.jax.org/strain/013044.html>
- VIP-Cre: <http://jaxmice.jax.org/strain/010908.html>
- PV-Cre: <http://jaxmice.jax.org/strain/008069.html>
- Ai9 reporter: <http://jaxmice.jax.org/strain/007909.html>
- Ai27 reporter: <http://jaxmice.jax.org/strain/012567.html>
- FVB/NJ mice: <http://jaxmice.jax.org/strain/001800.html>

Anesthesia was induced with 3% isoflurane and maintained with 1.5% to 2% isoflurane during the surgical procedure. Mice were injected with 5-10 mg/kg ketoprofen subcutaneously at the start of the surgery. Anesthetized mice were placed in a stereotaxic head holder (Kopf Instruments) and body temperature was maintained at 37C throughout the surgery using a

homeothermic blanket system (Harvard Instruments). After shaving the scalp, bupivacaine (0.05 cc, 0.5%, Marcaine) was applied subcutaneously under the scalp, and Ketoprofen (5–10 mg/kg, s.c.) was given for systemic analgesia. After 10-20 minutes an approximately 1 cm<sup>2</sup> area of skin was removed above the skull and the underlying fascia was scraped and removed. The wound margins were sealed with a thin layer of surgical glue (VetBond, 3M), and a headbar was attached with dental cement (Dentsply Grip Cement). At this point, the mouse was removed from the stereotax and the skull was held stationary on a small platform by means of the newly attached headbar. Using a surgical drill and HP 1/2 burr, a ~3 mm craniotomy was made centered on the area of interest (V1: 2.7mm lateral of the midline, contacting the lambda suture; S1: 3.3mm lateral and 1.5mm posterior of bregma), and the exposed cortex was washed with ACSF (125mM NaCl, 5mM KCl, 10mM Glucose, 10mM HEPES, 2mM CaCl<sub>2</sub>, 2mM MgSO<sub>4</sub>). In mice used for calcium imaging, one or two injections of Oregon Green 488 BAPTA-1 AM (OGB-1, Invitrogen; with 2% Pluronic F-127 in DMSO) were made in V1 under two-photon guidance (2-5 μm pipette tip, 40-200 mbar pressure, 0.5-1 μl total volume, targeted to 350 μm below the pia). The cortical window was then sealed with a 3 mm coverslip (Warner Instruments), using VetBond. In mice used for patching, the same 3 mm craniotomy was sealed with a coverslip with a ~500 μm hole that had been previously drilled with a diamond-tipped burr (Coltene/Whaledent). In this case, the hole in the coverslip was positioned to allow access the cortical area of interest with the patch pipette(s).

### **Behavioral recordings**

After surgery, the mouse's head was restrained above the spherical or cylindrical treadmill and the mouse was allowed to recover on a heating pad placed on top of the treadmill. Once awake, the mouse was given 30-60 minutes to fully recover from the effects of anesthesia and acclimate to head fixation. In some experiments the animals were habituated to the recording apparatus by head-restraining them on the treadmill for thirty minutes per day for five consecutive days before performing the craniotomy and recording. The treadmill rotated on a single axis so that treadmill movement was restricted to the rostro-caudal axis of the mouse. In some cases, movements were detected with an optical mouse and digitized at 200 Hz, while in others treadmill motion was measured using a rotary optical encoder with a resolution of 8000 counts/revolution. Running periods were defined as periods of one second or more with treadmill speeds above 1 cm/sec. Intervening periods of less than one second between running epochs were also included as running.

The mean increase in pupil size during running in Figures 1D and S2C was calculated after excluding running periods that occurred less than 10 seconds after a previous running period, in order to avoid cases where the pupil was already dilated by running. With this restriction, we found that 92% of running epochs in figure 1D were associated with an overall increase in pupil size. For about two-thirds of these epochs (67%), the increase in pupil size was greater than three times the standard deviation of diameter during the pre-running baseline (taken as the 10 seconds preceding run onset).

For C57Bl/6 mice, the eye was illuminated with a 680 nm or greater wavelength LED array and pupil position and size were recorded at 1280×1024 at 10 Hz on setup A (DCC1545M



camera, Thorlabs, with TML-HP 1x Telecentric lens, Edmund Optics) , and at 60 Hz on setup B (CR-GM00-H6400, Teledyne Dalsa). Because of the lack of pigmentation, red light did not provide sufficient contrast to illuminate the pupils of retinally-blind FVB mice, and oblique illumination with 490 nm light was used instead. The eye was not illuminated at all while testing for visual responsiveness in these mice. Whisking was recorded at 640×480 or 320×240 at 250 Hz with a separate camera (CR-GM00-H6400, Teledyne Dalsa) on setup A, or with the same camera used to record the eye on setup B. All behavioral data was acquired with custom software written in LabView (National Instruments).

Post-hoc pupil segmentation was performed with custom software written in MATLAB (Mathworks) and manually supervised. Reflection of the light source and visual stimulus on the eye were both masked from each frame of the image, which was smoothed using a 2D Gaussian kernel ( $\sigma=5-10$  pixels) and thresholded to isolate the darker pixels in the pupil. A circle was least-squares fit to the points along the edge of this region (Image Processing Toolbox, MATLAB; fitcircle, Richard Brown, Mathworks File Exchange). The detection threshold, smoothing widths, visual stimulus and specular reflection rejection were adjusted manually by visual inspection of the detection fit of every 10th frame (setup A) or every 20th frame (setup B).

Whisking was detected by extracting the optic flow from a manually-selected ROI of whisking movies (OpticalFlow Object, Computer Vision System Toolbox, MATLAB). Whisking and noise thresholds were set manually based on visual inspection of the optic flow trace and raw video. This analysis allowed sensitive detection of even small whisker movements, but did not allow full reconstruction of the whisker trajectory. We did not have high-speed video for

some of the recordings with patch data and all of the recordings with calcium imaging on Setup A (grating stimulus). In these cases we used movements of the face and whisker pad captured by the eye imaging camera to identify whisking periods. Comparing these data to the high-speed video, we found that this measurement captured more than 85% of the whisking periods where high speed video was available, as well as many of the smaller subthreshold movements associated with postural adjustments on the treadmill. We thus used whisker pad movement as a proxy for direct visualization of the whisker in the recordings without high-speed video. Whisking periods were defined as periods of suprathreshold activity of one second or more, and intervening periods of less than one second between whisking epochs were also included as whisking.

### **Visual stimuli**

Grating stimuli were presented on a 7" LCD monitor (Lilliput 665GL-70NP/HO/Y monitor; 60 Hz scan rate; setup A) positioned 10 cm away from the eye, covering approximately 88° (azimuth) by 72° (elevation) of the contralateral visual field. Natural movie stimuli were presented on a 30" LCD monitor (U3011t, Dell, 60Hz refresh rate; setup B) positioned ~30 cm away from the animal's eye, covering 94° (azimuth) by 67° (elevation) of the contralateral visual field. For the grating stimulus, full-field square wave gratings (spatial frequency 0.04 Hz, temporal frequency 2 Hz) were presented at eight orientations (100 blocks containing all orientations in random order). Gratings were presented for 500 ms and separated by a 1-second luminance-matched gray background. Stimulus presentation routines were written in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A photodiode attached to the

screen allowed exact time-stamping of each frame of the stimulus presentation. Natural movies were acquired as previously described (Froudarakis et al., 2014). Briefly, recordings were made from a head-mounted webcam while the mouse ran freely in an enriched and spacious (~0.5 m<sup>2</sup>) environment. We used 10 s clips of these movies, which contained periods of locomotion, looking around and resting.

### **Calcium imaging**

Two-photon imaging of OGB fluorescence was performed on modified Sutter microscopes. Imaging on setup A was performed with a fast resonant scanning system (ThorLabs, typically at 30 Hz). Imaging on setup B was performed with a galvo-galvo scanning system (Sutter Instruments, typically at 12 Hz). Excitation on both setups was via a Ti-Sapphire laser (Chameleon Vision, Coherent) tuned to either 800 nm or 1000 nm and either a 20× (1.0 NA, Olympus) or 25× (1.1 NA, Nikon) objective. Power out of the objective was controlled by calibrated rotations of a half-wave attenuator and depended on the magnification of the scan but was always kept below 50 mW.

Avoiding neuropil contamination of calcium signals during functional imaging is important and often underappreciated. Using high NA objectives to restrict the size of the point-spread function is critical. Typical empirical PSFs measured with our objectives were approximately 300 nm (Nikon) or 400 nm (Olympus) in both X and Y, and approximately 2 μm in Z (both objectives; full width at half-height). Characterizing and reducing movement of the prep is also important. Motion in the X-Y plane was corrected with post-hoc raster correction and sub-pixel motion correction prior to extracting calcium traces. We could not correct

movement in the Z-plane, but with our chronic window preparations typically motion in Z was very small. For about half the recordings used in Figure 4, we estimated the magnitude of Z motion by taking a high-resolution (500 nm step) Z-stack following each scan and registering each frame in the recording to this stack. Mean Z motion was slightly higher during running than quiet wakefulness, measured as the absolute displacement from the mean Z-position during each scan (running: mean 1.2  $\mu\text{m}$ , s.d. 0.6 $\mu\text{m}$ ; quiet: mean 0.88  $\mu\text{m}$ , s.d. 0.46  $\mu\text{m}$ ).

### **Whole-cell patching**

Patch pipettes were pulled from borosilicate glass (1.5 mm OD  $\times$  0.86 mm ID, Sutter Instruments) to an impedance of 6-10 M $\Omega$ . Pipettes were filled with standard internal solution (Jiang et al., 2013) (125mM Potassium Gluconate, 10mM HEPES, 4mM Mg-ATP, 0.3mM Na-GTP, 10mM Na<sub>2</sub>-Phosphocreatine, 13 mM biocytin, pH 7.25, osmolarity normalized to extracellular solution) and Alexa 488 or 598 was added (50  $\mu\text{M}$ ) to allow visualization of the pipette and extracellular space. A manometer (Fisher Scientific 06-664-19) and custom-built pressure manifold allowed fast switching between high pressures while entering the bath and penetrating the dura (~150 mbar), and low pressures (~20-50 mbar) while advancing the pipette through the cortex under two-photon guidance, which helped to reduce the overall volume of intracellular solution ejected from the pipette. Bias currents were zeroed once the pipette was placed in the bath. Gigaseals were allowed to stabilize for 3-5 minutes before break-in. Compensating for tissue distortion by retracting the pipette ~10  $\mu\text{m}$  during this time resulted in improved access and more stable recordings.  $V_m$  was not adjusted for the liquid junction potential.

## **Preprocessing of calcium imaging data**

For data from the galvanometric scanning system (setup B), we performed motion compensation in the horizontal plane, and cell detection was automated by detecting circular regions based on minimum diameter, cell contrast, edge sharpness and maximum number of cells per site, and the results were manually supervised and corrected if necessary. For resonant scanning data (setup A) cell detection was performed entirely manually. The data processing chain for this and subsequent analysis relied on the DataJoint library for MATLAB (<http://datajoint.github.com/datajoint-matlab/>). Calcium traces averaged from all pixels in segmented cells were normalized to  $\Delta F/F$ , down-sampled to 10-20 Hz and then filtered by using a nonnegative deconvolution algorithm (Vogelstein et al., 2010) to infer instantaneous firing rates. The parameters of the algorithm were optimized to yield optimal reconstruction ( $\lambda = 0.3$ ) as previously described (Cotton et al., 2013; Froudarakis et al., 2014).

## **Analysis of patching data**

After removing detected spikes, the amplitude of low-frequency oscillations was measured by filtering raw  $V_m$  from 2-10 Hz, taking the Hilbert amplitude of this oscillation, and filtering the resulting trace between 0.1 and 1 Hz (pupil phase plots in Figure 2D,3E), or low-pass filtering below 1Hz (all other figures). For registration to a standard dilation/constriction cycle in the phase plots, whisking, running and saccades were excluded from analysis, and the filtered  $V_m$  and Hilbert amplitude at each time point were binned by the Hilbert phase of the filtered (0.1 to 1 Hz) pupil trace (64 bins from  $-\pi$  to  $\pi$ ). In the phase plot analysis, the data were restricted to the same periods of pupil dilation and constriction used in Figure 2E-H (duration greater than

one second, with absolute dilation or constriction rates greater than 10  $\mu\text{m/s}$ ) Means of binned data in Figure 2D, Figure 3E, and Figure S3 were smoothed by averaging with adjacent bins.

#### **Analysis of responses to moving gratings (Figure 4)**

Orientation tuning was computed by linear regression of the deconvolved calcium trace to the stimulus conditions (drifting directions). Mean tuning curves in Figure 4C,D are smoothed with the Hamming filter with 40-degree support. In figure 4C, we restricted the analysis to recordings with at least one minute of total running time (n=516 cells in 19 sites), in order to correctly estimate the tuning of cells during both conditions. The reliability of responses across visual stimuli was calculated as the fraction of variance explained by the average response binned in 150 ms bins limited to stimulus conditions with at least 16 repeats in each state. Significance was calculated using the number of sites as the degrees of freedom rather than the number of cells in order to avoid the potential confounding effects of correlations between cells. The preferred direction was found by fitting the two-peaked von Mises function to all data irrespective of condition. OSI was then calculated for each cell under each condition as the difference between the mean response to this preferred direction and the mean response to the orthogonal direction, normalized by their average. Trials with any overlap with running or whisking periods were treated as “active,” and these trials were excluded from further analysis in Figure 4. Trials with saccades or blinks were also excluded from analysis. Trials were classified as “constricting” or “dilating” based on the difference in pupil size from the start to the end of the trial, and trials with absolute dilation or constriction rates less than 10  $\mu\text{m/s}$  were excluded. Noise correlations were calculated as correlations between residual signals after

subtracting the mean stimulus responses binned in 150 ms intervals and normalized by the standard deviation of responses in each bin. Signal correlations were calculated as correlations between mean stimulus responses binned in 150 ms intervals.

### **Analysis of responses to natural movies (Figure 5)**

To account for changes in pupil diameter due to luminance changes in the natural movies, we first binned the movies, neural responses, and the rate of change of the pupil diameter in 150 ms bins. Each bin contained a few frames of the movie (here we refer to these bins as “scenes”). For each scene, we sorted all repetitions of that scene (trials) by the rate of change of the pupil diameter, and for each cell we compared the neural responses in the 25% of trials with the lowest rates of pupil change (“low”) to the neural responses to the same scene in the 25% of trials with the highest rates of pupil change (“high”). Signal correlations were calculated as the mean Pearson correlation between average response traces across all pairs of cells in each imaging site. To compute noise correlations we first computed the z-scores across trials for each scene. Noise correlations were calculated as the Pearson correlation between the computed z-scores averaged across each imaging site. The reliability of each cell was defined as the variance of the average response across movie repetitions divided by the total variance over the entire recording. To measure discriminability, we used an optimal linear classifier assuming isotropic covariance matrices and leave-one-out cross-validation to estimate the decoding error between the neural representations of pairs of scenes. Each scene was represented as an N-dimensional vector of neural activity for each trial. We converted the decoding error to discriminability, the mutual information (measured in bits) between the true class label  $c$  and its estimate  $\hat{c}$  by

computing

$$MI(c, \hat{c}) = \sum_{i \in \{1,2\}} \sum_{j \in \{1,2\}} p_{ij} \log_2 \frac{p_{ij}}{p_i \cdot p_j},$$

where  $p_{ij}$  is the probability of observing true class  $i$  and predicted class  $j$ , and  $p_i$  and  $p_j$  denote the respective marginal probabilities.

To compute the change in the mean firing rate for different scenes (Figure 5D,E), we sorted the scenes by response magnitude of each cell (not taking pupil dilation/constriction into account), ordering them from least-preferred (smallest response) to most-preferred (largest response). We then grouped the sorted scenes into quartiles: 0%-25% (least-preferred scenes), 25%-50%, 50-75%, and 75%-100% (most-preferred scenes) and examined the difference in single-cell responses between the high and low conditions for each of these quartiles. Percent change was defined as  $(\text{response}_{\text{high}} - \text{response}_{\text{low}}) / \text{response}_{\text{high}}$ . Error bars were computed as 95% bootstrap confidence intervals with 5000 resamples. For this analysis we only included the cells that were reliably driven by the stimulus (>30% of the response variance explained in either the low or high condition;  $n=467/1249$  neurons).



## Supplemental References

- Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial vision* 10, 433-436.
- Cotton, R.J., Froudarakis, E., Storer, P., Saggau, P., and Tolias, A.S. (2013). Three-dimensional mapping of microcircuit correlation structure. *Front Neural Circuits* 7, 151.
- Froudarakis, E., Berens, P., Ecker, A.S., Cotton, R.J., Sinz, F.H., Yatsenko, D., Saggau, P., Bethge, M., and Tolias, A.S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat Neurosci* 17, 851-857.
- Jiang, X., Wang, G., Lee, A.J., Stornetta, R.L., and Zhu, J.J. (2013). The organization of two new cortical interneuronal circuits. *Nat Neurosci* 16, 210-218.
- Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision* 10, 437-442.
- Vogelstein, J.T., Packer, A.M., Machado, T.A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J Neurophysiol* 104, 3691-3704.