

SUPPLEMENTAL DATA**Automated analysis of cellular signals from large-scale calcium imaging data****Eran A. Mukamel^{1,3,*}, Axel Nimmerjahn¹, Mark J. Schnitzer^{1,2,*}**¹James H. Clark Center for Biomedical Engineering & Sciences²Howard Hughes Medical Institute

Stanford University, Stanford CA 94305

³Present address: Center for Brain Science, Harvard University, Cambridge, MA 02138*Correspondence: emukamel@fas.harvard.edu (E.A.M.), mschnitz@stanford.edu (M.J.S.)**TABLE OF CONTENTS**

SUPPLEMENTAL FIGURES	3
Figure S1. Automated Separation of Signal and Noise Components	3
Figure S2. Simultaneous Ca ²⁺ Imaging and Extracellular Recording Quantifies the Discriminability of Complex Spike Trains.....	5
Figure S3. Combining Spatial and Temporal Information Can Improve Cell Sorting.	6
Figure S4. Automated Cell Sorting in Somatosensory Cortex Removes Cross Talk from Neuropil and Overlapping Cells in Simulated and Experimental Data.	8
Figure S5. Automated Cell Sorting of Electrically Evoked Activity.	10
Figure S6. Automated Cell Sorting Separates Movement Artifacts from Cellular Signals.....	11
SUPPLEMENTAL TABLES	12
Table 1. CellSort 1.0: MATLAB Toolboxes for Automated Cell Sorting.....	12
SUPPLEMENTAL EXPERIMENTAL PROCEDURES.....	16
Analysis Software	16
Image Registration and Normalization	16
Dimensional Reduction	16
Cell Sorting by Independent Component Analysis (ICA)	18
Spatio-Temporal ICA	20
Temporal Deconvolution and Spike Detection.....	21
Image Segmentation.....	22
Artificial Cerebellar Movie Data Sets for Validation of Cell Sorting Procedures	22
Artificial Neocortical Movie Data Sets.....	25
Definitions of Signal Fidelity and Cross Talk	26
Linear Regression Analysis.....	26
Region of Interest (ROI) Analysis.....	27
Animals and Animal Surgery	27
Fluorescence Labeling	28

Two-Photon Imaging in Behaving Mice	28
Behavioral Assessment	29
Electrophysiology	29
Electrical Stimulation of Limbs	30
Receiver Operating Characteristic (ROC) Analysis	30
Identification of Periods of Active Locomotion	30
Detection of Cerebellar Microzones	31
SUPPLEMENTAL REFERENCES	31

SUPPLEMENTAL FIGURES

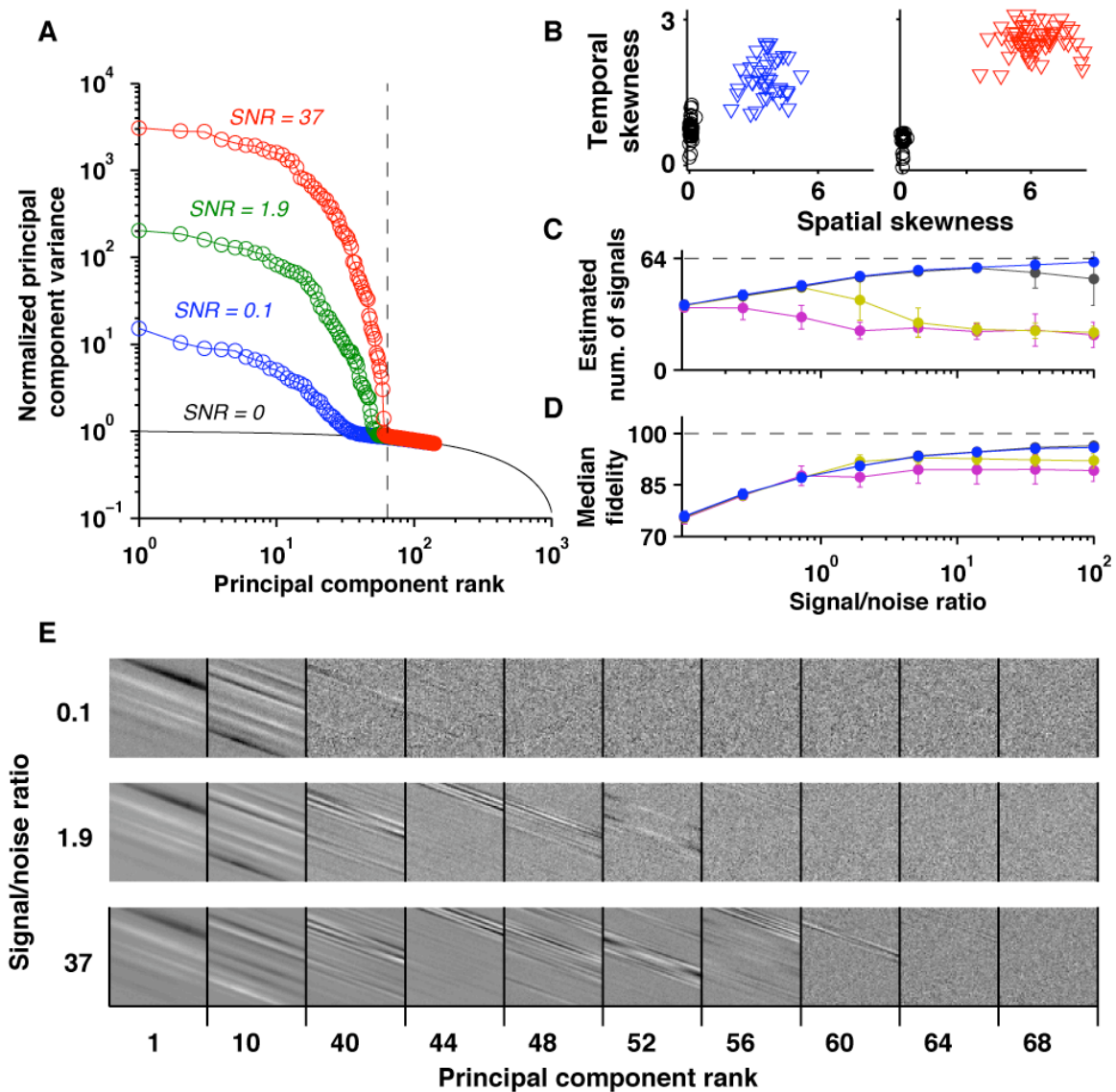


Figure S1. Automated Separation of Signal and Noise Components

(A) Sorted and normalized principal components' variance values for artificial data sets containing 64 cells with the signal/noise ratio (SNR) ranging from 0 (black solid curve) to 37 (red). The components with rank >64 (to the right of the dashed vertical line) approach the Gaussian noise floor (black curve), but the variances for the lower ranked components depend on the SNR.

(B) k-means clustering ($k=2$) of spatial and temporal skewness values automates the separation of independent components representing signals (triangles) from those that represent noise (circles). The blue and red data sets have SNR values of 0.1 and 37, respectively, as in (A).

(C) The estimated number of cellular signals (mean \pm s.d.) is the number of components in the high-skewness cluster. We compared the performance of automated separation of signal and noise components using fewer principal components (PCs) than true signals (45 PCs, magenta; 54 PCs,

yellow); the same number of PCs as signals (64, black); or more PCs than signals (91, blue). If the number of retained PCs is greater than the number of cells, the estimated number of signals approaches the true value in the high SNR limit.

(D) Fidelity (mean \pm s.d.), defined as the average over all cells that were extracted by automated sorting, improves with increasing SNR and number of PCs retained. The curves' colors denote the same numbers of retained PCs as in (C).

(E) Examples of PC spatial filters showing the transition from structured cellular signals to featureless noise. Movies with high signal/noise ratio (bottom row) have a relatively sharp transition occurring near the true number of signals (64), whereas noisier data sets show a gradual degradation of signals that increases with the rank of the PC.

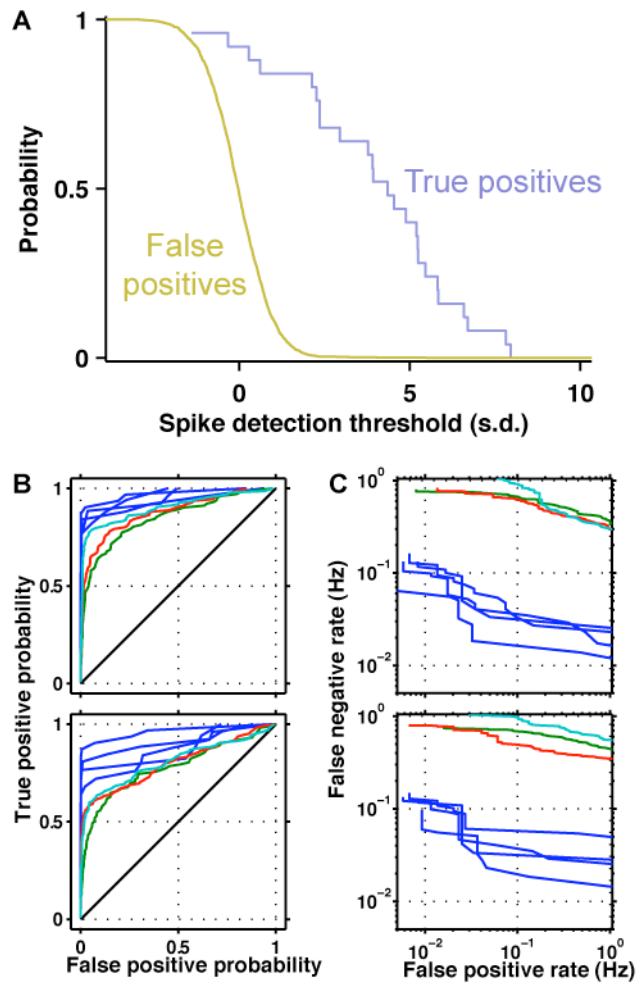


Figure S2. Simultaneous Ca^{2+} Imaging and Extracellular Recording Quantifies the Discriminability of Complex Spike Trains.

(A) Cumulative distribution functions for the deconvolved fluorescence signal (normalized to unit variance) during all movie time frames coinciding with an electrically recorded complex spike (purple) or during a background frame when no spike was present (yellow). The detection threshold sets a particular probability of false positive errors and corresponding probability of correctly identifying true spikes.

(B) Receiver operating characteristic (ROC) analysis of the discriminability of complex spikes using independent component time courses (top) or signals resulting from image segmentation (bottom). Each curve illustrates the tradeoff between false positives and true spike detection for one of the cells studied by simultaneous electrical and optical recording. Different mice are indicated by the colors.

(C) The same data as in (B) are displayed in terms of the rate (events per s) of false positive and false negative detection. The lower rate of complex spikes in one experiment (blue curves) improved discriminability in that subject.

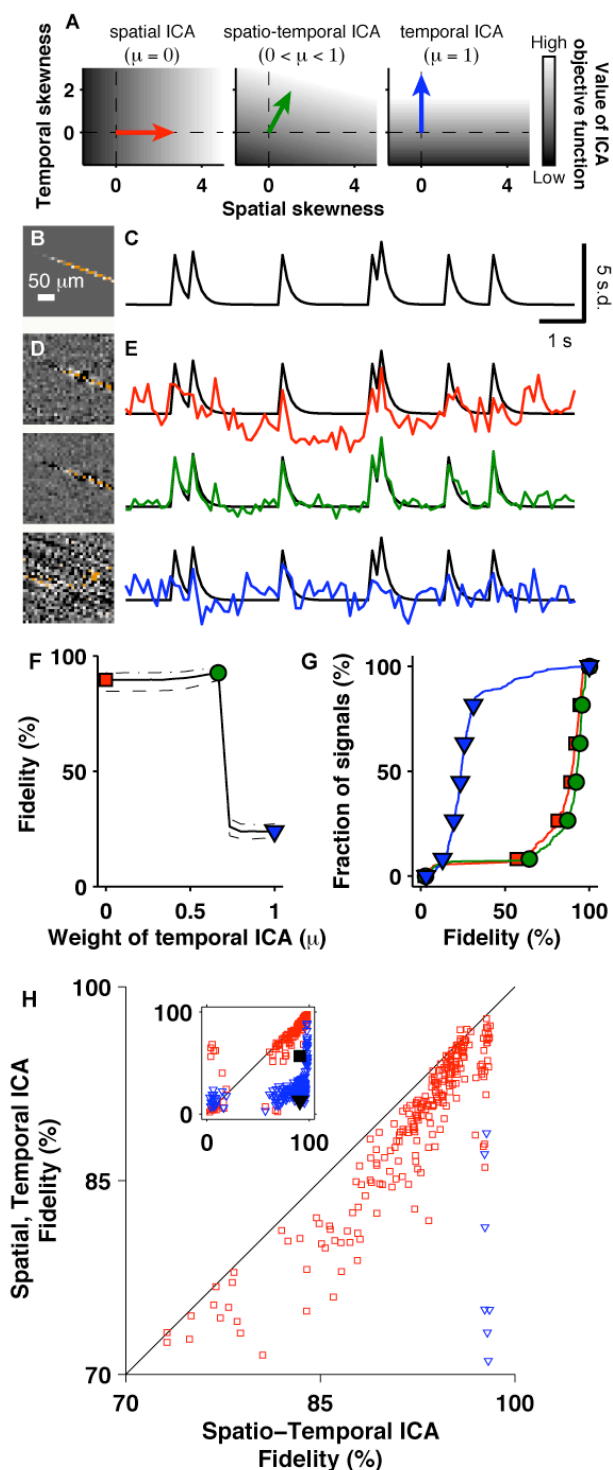


Figure S3. Combining Spatial and Temporal Information Can Improve Cell Sorting.

(A) Schematic illustration of the objective function (grayscale) maximized by spatial (left), temporal (right) and spatio-temporal (middle) ICA as a function of spatial and temporal signal skewness (compare Figure 2E). Arrows show the direction of the objective function's gradient.

(B,C) Example spatial filter (B) and signal time course (C) from an artificial data set mimicking Ca^{2+} -imaging data with large numbers of Purkinje cell dendritic trees.

(D,E) Estimates of artificial signals extracted by spatial (top row, red trace), temporal (bottom, blue) or spatio-temporal ICA (middle, green). Of the three methods, spatio-temporal ICA extracts a signal more closely correlating with the underlying intracellular Ca^{2+} dynamics

(F) Median fidelity (solid line) of the signals recovered by ICA, as a function of the relative weight of temporal information (μ). Dashed and dotted lines mark ± 1 s.d. from the median.

(G) Cumulative distribution of the fidelity of recovered signals for the three values of μ indicated by color in (F). Spatio-temporal ICA (green curve) yields the largest percentage of signal traces with high fidelity.

(H) Scatter plot showing the fidelity for each simulated cell's signal as estimated by spatio-temporal ICA ($\mu=0.67$) vs. spatial (red squares, $\mu=0$) and temporal (blue triangles, $\mu=1$) ICA. The black line indicates equal fidelity, and points to the right of the line correspond to cells for which spatio-temporal ICA improved the fidelity as compared with spatial or temporal ICA. Inset shows the full range of fidelity values. Black symbols mark the fidelity of the example cell shown in panels (D,E).

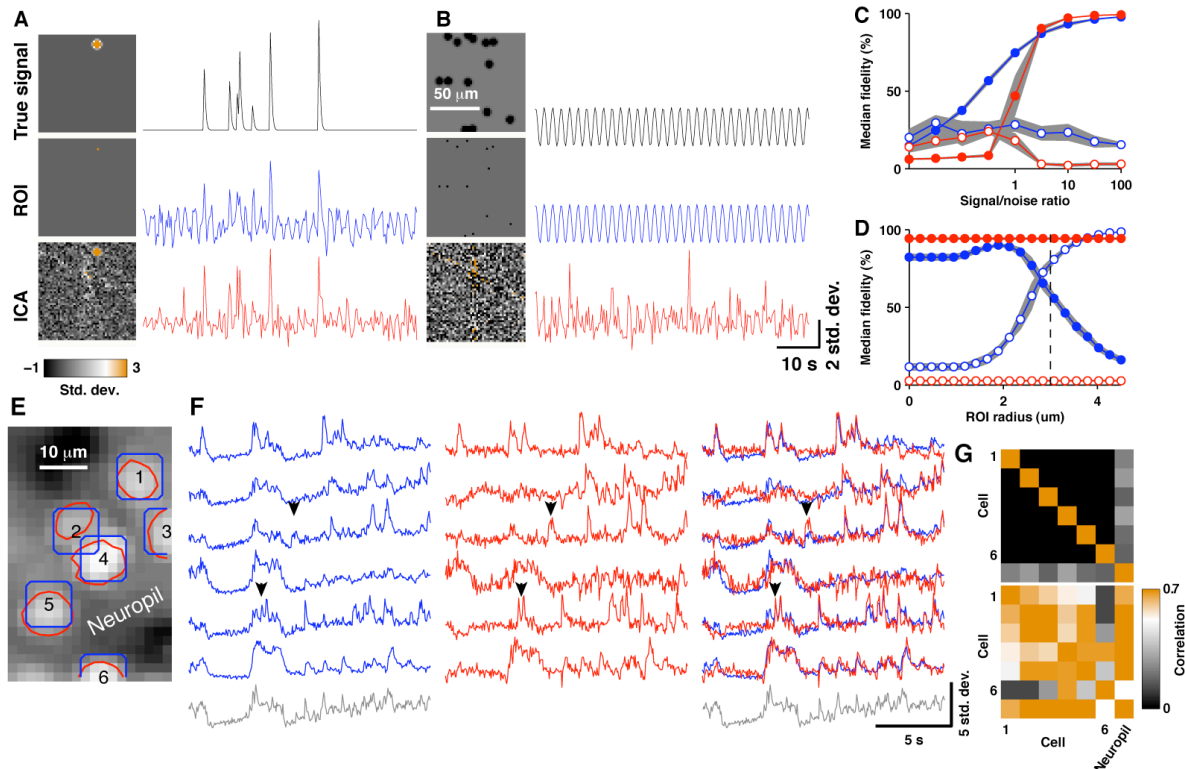


Figure S4. Automated Cell Sorting in Somatosensory Cortex Removes Cross Talk from Neuropil and Overlapping Cells in Simulated and Experimental Data.

(A,B) Simulations of neocortical data sets include both somatic signals (A) and low frequency (0.5 Hz) oscillatory neuropil activity (B) in the surrounding regions, which contaminates signals recorded from the cellular regions due to the finite imaging resolution. Top rows show examples of the signal sources' true spatial filters (left) and dynamics (right). Middle and lower rows show the signals estimated by ROI analysis and ICA, respectively. The simulated field of view was 100 μm square, containing 13 somata.

(C) Median fidelity of the extracted signals as a function of the signal/noise ratio, for ICA (red closed circles) and ROI analysis (blue closed circles). Median cross talk values between the cellular signals and the neuropil are plotted with red and blue open circles for ICA and ROI analysis, respectively. Gray regions here and in (D) indicate one standard deviation. The fidelity of the extracted signals rises with the signal/noise ratio, but ICA achieves higher asymptotic fidelity and lower cross talk.

(D) At high signal/noise ratio (SNR=5), the fidelity of signals extracted by ICA (red closed circles) is superior to those found by ROI analysis (blue closed circles), for all choices of ROI radius. Cross talk is lower for ICA (red open circles) than for ROI analysis (blue open circles). Dashed line shows the actual radius of the simulated somata.

(E) Contours showing six actual cells identified by ROI analysis (blue) or by spatio-temporal ICA (red; $\mu=0.1$) from experimental data recorded by two-photon microscopy in the somatosensory cortex of an awake mouse. Neuropil signal was estimated using pixels outside the union of all ROIs.

(F) Time courses of dynamics from each of the six cells of (E), as extracted by ROI analysis (blue traces, left column) or ICA (red traces, middle column). The neuropil signal trace is shown in gray. Black

arrowheads indicate events for which there is significant disparity between the two sets of estimates, which are overlaid in the right column.

(G) Matrices of correlation coefficients for the six cells and the neuropil in (E) reveal much higher levels of cross talk for the traces extracted by ROI analysis (bottom matrix) than for those extracted by ICA (top matrix). The last row and last column of each matrix show the correlation coefficients of the individual signals and the neuropil signal.

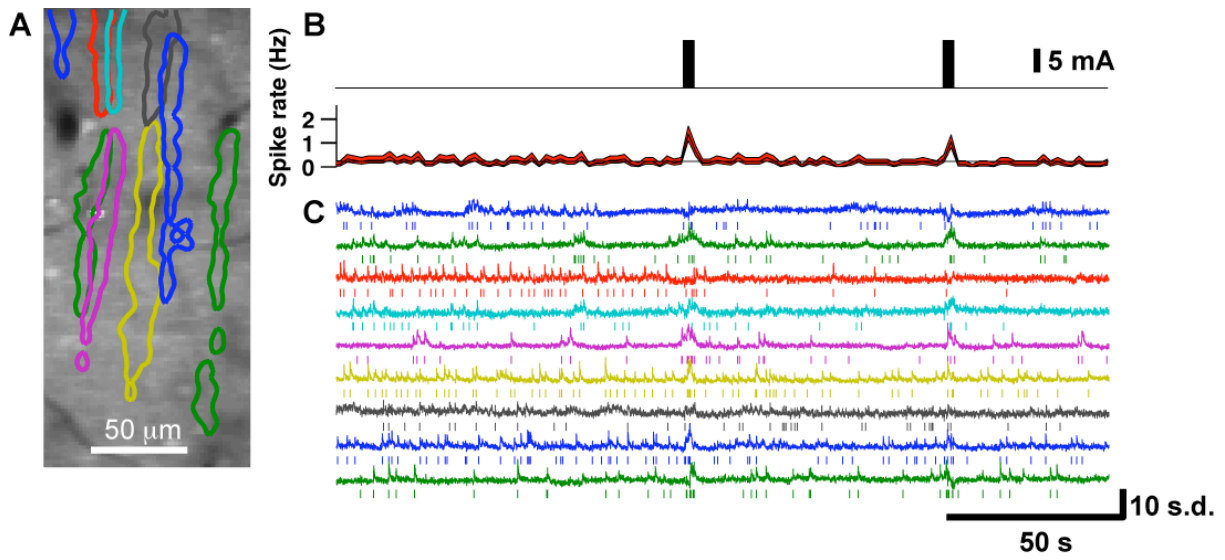


Figure S5. Automated Cell Sorting of Electrically Evoked Activity.

- (A) Contours from spatial filters identified by automated cell sorting in an anesthetized mouse.
- (B) Electrical stimulation (top black trace) and mean detected Ca^{2+} spike rate (bottom, mean \pm 1 s.d.), showing evoked increases in spike frequency during two periods of stimulation. Peripheral electrical stimulation applied to the ipsilateral forepaw consisted of trains of 60 square pulses, each of 20 ms duration and 10 mA amplitude, delivered at 20 Hz.
- (C) Extracted cellular signals and corresponding spike rasters for each of the nine cells shown in (A).

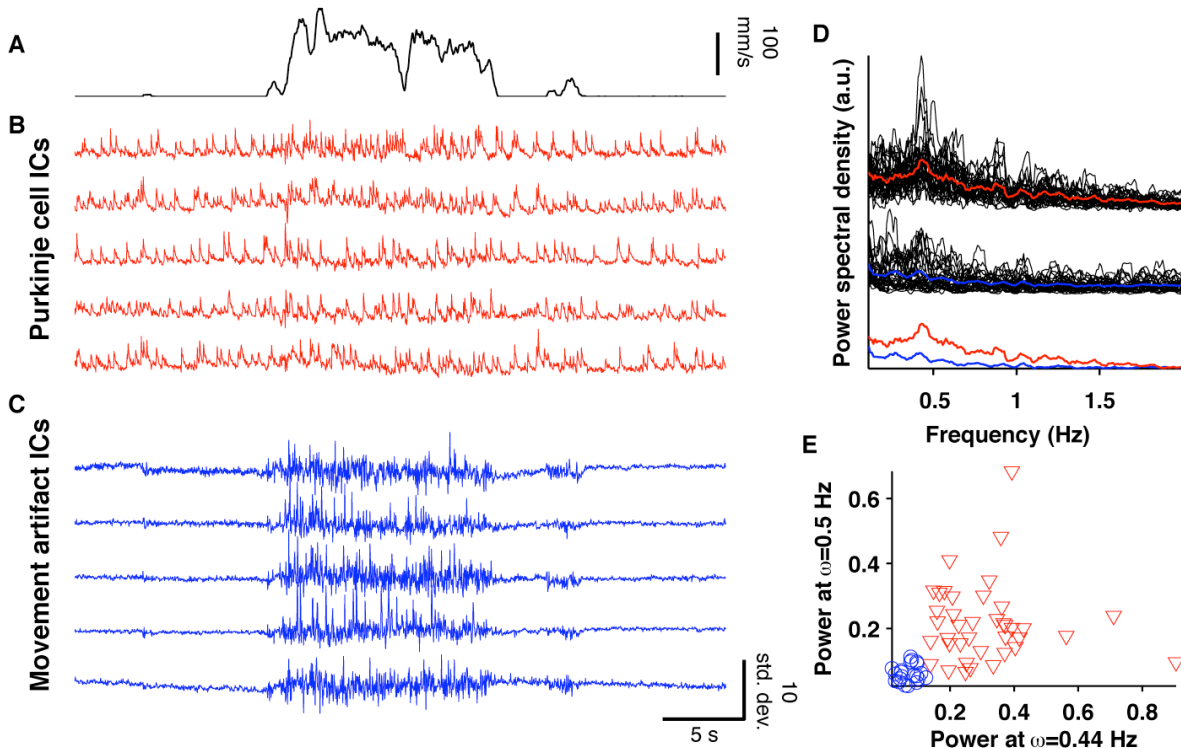


Figure S6. Automated Cell Sorting Separates Movement Artifacts from Cellular Signals.

This figure shows data recorded in the cerebellum of an awake, actively running mouse. By comparison to the data of Figure 2, motion artifacts in this data set were larger, to a degree that the spatial and temporal skewness values were similar for the independent components of cellular and artifactual origin. Thus, skewness cannot be used to distinguish the two classes of components. However, spectral analysis provides another means:

(A) Mouse running speed on the ball.

(B) Selected time courses of extracted independent components identified as neuronal signals.

(C) Time courses of independent components identified as movement artifacts. Note that the signals are largely high-frequency noise and are inconsistent with the known kinetics of the Ca^{2+} indicator.

(D) Power spectral density for signals with neuron-like dynamics (top; average over all cells in red), and for the remaining signals that seem to arise from movement artifacts (middle; average in blue). The bottom row shows an overlay of the two average power spectral densities. The spectral peak in neuronal signals around 0.5 Hz is absent from the artifactual signals, and thus spectral analysis can be used to automatically separate the two classes of components in a manner that coincides with the manual separation relying on the presence of neuronal signals in the time traces.

(E) Plot of power in two frequency bands (± 0.03 Hz bandwidth) for each of 60 independent components found in this experiment. Components manually identified as neurons (red triangles) are separated from those identified as artifacts (blue circles) and can be automatically separated by using a fixed threshold in this 2D space.

SUPPLEMENTAL TABLES

Table 1. CellSort 1.0: MATLAB Toolboxes for Automated Cell Sorting

File name and function specification		
[Outputs] = FunctionName(Inputs)		
Description	Inputs	Outputs
CellsortPCA <pre>[mixedsig, mixedfilters, CovEvals, covtrace, movm, movtm] = CellsortPCA(fn, flims, nPCs, dsamp, outputdir, badframes)</pre>		
Pre-process movie data to find the principal component (PC) spatial filters and time courses.	<ol style="list-style-type: none"> 1. fn - movie file name. Must be in TIFF format. 2. flims - 2-element vector specifying the endpoints of the range of frames to be analyzed. If empty, default is to analyze all movie frames. 3. nPCs - number of principal components to be returned 4. dsamp - optional downsampling factor. If scalar, specifies temporal downsampling factor. If two-element vector, entries specify temporal and spatial downsampling, respectively. 5. outputdir - directory in which to store output .mat files 6. badframes - optional list of indices of movie frames to be excluded from analysis 	<ol style="list-style-type: none"> 1. mixedsig - N x T matrix of N temporal signal mixtures sampled at T points. 2. mixedfilters - N x X matrix of N spatial signal mixtures sampled at X spatial points. 3. CovEvals - largest eigenvalues of the covariance matrix 4. covtrace - trace of covariance matrix, corresponding to the sum of all eigenvalues (not just the largest few) 5. movm - average of all movie time frames at each pixel 6. movtm - average of all movie pixels at each time frame, after normalizing each pixel ($\Delta F/F$)

CellsortChoosePCs <code>[PCuse] = CellsortChoosePCs(fn, mixedfilters)</code>		
Allow the user to select which principal components will be kept following dimensional reduction.	<ol style="list-style-type: none"> 1. fn - movie file name 2. mixedfilters - N x X matrix of N spatial signal mixtures sampled at X spatial points. 	<ol style="list-style-type: none"> 1. PCuse - vector of indices of the PCs to be kept for dimensional reduction
CellsortPlotPCspectrum <code>CellsortPlotPCspectrum(fn, CovEvals, pcuse)</code>		
Plot the principal component (PC) spectrum and compare with the corresponding random-matrix noise floor.	<ol style="list-style-type: none"> 1. fn - movie file name 2. CovEvals - eigenvalues of the covariance matrix 3. pcuse - [optional] - indices of PCs included in dimensionally reduced data set 	
CellsortICA <code>[ica_sig, ica_filters, ica_A, numiter] = CellsortICA(mixedsig, mixedfilters, CovEvals, PCuse, mu, nIC, ica_A_guess, termtol, maxrounds)</code>		
Perform spatio-temporal independent component analysis (ICA) using the PC spatial filters and time courses	<ol style="list-style-type: none"> 1. mixedsig - N x T matrix of N temporal signal mixtures sampled at T points. 2. mixedfilters - N x X matrix of N spatial signal mixtures sampled at X spatial points. 3. CovEvals - eigenvalues of the covariance matrix 4. PCuse - vector of indices of the components to be included. If empty, use all the components 5. mu - parameter (between 0 and 1) specifying weight of temporal information in spatio-temporal ICA 6. nIC - number of ICs to derive 7. ica_A_guess - (optional) Initial guess for the ICA unmixing matrix; dimensions should be (Number of PCs) x nIC. If empty, generate a random initial guess. 8. termtol - termination tolerance; fractional change in output at which to end iteration of the fixed point algorithm. 9. maxrounds - maximum number of rounds of iterations 	<ol style="list-style-type: none"> 1. ica_sig - nIC x T matrix of ICA temporal signals 2. ica_filters - nIC x X matrix of ICA spatial filters 3. ica_A - nIC x N orthogonal unmixing matrix to convert the input to output signals 4. numiter - number of rounds of iteration before termination

CellsortICApplot

```
CellsortICApplot(mode, ica_filters, ica_sig, f0, tlims, dt, ratebin, plottype, ICuse, spt, spc)
```

Display the results of ICA analysis in the form of paired spatial filters and signal time courses

1. mode - 'series' shows each spatial filter separately; 'contour' displays a single plot with contours for all spatial filters
2. ica_filters - nIC x X matrix of ICA spatial filters
3. ica_sig - nIC x T matrix of ICA temporal signals
4. f0 - mean fluorescence image
5. tlims - 2-element vector specifying the range of times to be displayed
6. dt - time step corresponding to individual movie time frames
7. ratebin - size of time bins for spike rate computation
8. plottype - type of spike plot to use:
 - plottype = 1: plot cellular signals
 - plottype = 2: plot cellular signals together with spikes
 - plottype = 3: plot spikes only
 - plottype = 4: plot spike rate over time
9. ICuse - vector of indices of cells to be plotted
10. spt - vector of spike times (needed for plottype > 1)
11. spc - vector of spike indices (which cell) (needed for plottype > 1)

CellsortSegmentation

```
[ica_segments, segmentlabel, segcentroid] =  
  CellsortSegmentation(ica_filters, smwidth, thresh, arealims, plotting)
```

Segment spatial filters derived by ICA.

- | | |
|---|--|
| <ol style="list-style-type: none"> 1. ica_filters - nIC x X matrix of ICA spatial filters 2. smwidth - standard deviation of Gaussian smoothing kernel (pixels) 3. thresh - threshold for spatial filters (standard deviations) 4. arealims - 2-element vector specifying the minimum and maximum area (in pixels) of segments to be retained; if only one element is specified, use this as the minimum area 5. plotting - [0,1] whether or not to show filters | <ol style="list-style-type: none"> 1. ica_segments - segmented spatial filters 2. segmentlabel - indices of the ICA filters from which each segment was derived 3. segcentroid - X,Y centroid, in pixels, of each segment |
|---|--|

CellsortFindspikes

```
[spmat, spt, spc] = CellsortFindspikes(ica_sig, thresh, dt, deconvtau, normalization)
```

Deconvolve
signal and find
spikes using a
threshold

1. ica_sig - nIC x T matrix of ICA temporal signals
2. thresh - threshold for spike detection
3. dt - time step
4. deconvtau - time constant for temporal deconvolution (Butterworth filter); if deconvtau is 0 or empty, no deconvolution is performed
5. normalization - type of normalization to apply to ica_sig; 0 - no normalization; 1 - divide by standard deviation and subtract mean

1. spmat - nIC x T sparse binary matrix, containing 1 at the time frame of each spike
2. spt - list of all spike times
3. spc - list of the indices of cells for each spike

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Analysis Software

Data analysis was performed using custom software written in MATLAB or using *ImageJ* plugins. Some routines were modified from the publicly available *FastICA* package (Hyvarinen and Oja, 2000). See the Supplemental Toolbox of our MATLAB routines, bundled into a zip file. Periodic updates may be available online at the Matlab File Exchange: <http://www.mathworks.com/matlabcentral/fileexchange>.

Image Registration and Normalization

We first corrected any lateral movement artifact within each movie frame, as assessed relative to a reference frame, by using the image registration algorithm implemented in the *ImageJ* plugin *TurboReg* (Thevenaz et al., 1998). We cropped the movies to exclude any pixels that moved outside the field of view at any time during the recording. The resulting usable field of view was typically 1-14 pixels smaller in each spatial dimension than the original movie. We then normalized the signal in each pixel by dividing by its mean value over all movie time frames and subtracting 1. We next subtracted from each time frame the mean fluorescence averaged over all pixels. In cases when we compared data from the same animal and brain area under different physiological conditions (e.g. anesthetized versus awake), we spatially aligned the movies to a common reference frame, usually the first frame of the anesthetized recording.

Dimensional Reduction

We used principal component analysis (PCA) to reduce the dimensionality of the data and remove much of the noise that in each pixel causes intensity fluctuations uncorrelated to signals in other pixels (Mitra and Pesaran, 1999). Each movie was represented as a matrix, \mathbf{M} , of size $N_x \times N_t$, where N_x is the total number of pixels and N_t is the total number of time frames. The entries, \mathbf{M}_{xt} , contain the normalized fluorescence intensity in pixel x at time t . By using PCA we reduced the data to a lower-dimensional, approximate representation in terms of a set of K (where $K < \min(N_x, N_t)$) principal components (PCs). In practice, K ranged from 40 to 120 and was chosen separately for each data set (see below). Each PC comprises a spatial filter and a signal time course. The PC spatial filters, U_k , can be arranged in a $N_x \times K$ matrix with orthogonal columns, which we denote \mathbf{U} . The PC time courses, V_k , similarly define a $N_t \times K$

matrix with orthogonal columns, which we denote \mathbf{V} . After dimensional reduction, the new data representation approximates the original as a linear combination of the PCs,

$$\mathbf{M}_{xt} \approx \bar{\mathbf{M}}_{xt} \equiv \sum_k S_k U_{kx} V_{kt} = [\mathbf{USV}^T]_{xt},$$

where \mathbf{S} is a $K \times K$ diagonal matrix whose k 'th diagonal entry squared, S_k^2 , is the variance of the data after projecting along the k 'th PC. PCA is defined by choosing the matrices, \mathbf{U} , \mathbf{V} and \mathbf{S} , that minimize the mean square difference between the exact and approximate descriptions of the data,

$$\varepsilon\{\mathbf{U}, \mathbf{S}, \mathbf{V}\} = \sum_{x,t} (\mathbf{M}_{xt} - \bar{\mathbf{M}}_{xt})^2, \quad \{\mathbf{U}, \mathbf{S}, \mathbf{V}\} = \arg \min \varepsilon.$$

We solved this optimization problem by computing the $N_t \times N_t$ temporal covariance matrix, $\mathbf{C} = \mathbf{M}^T \mathbf{M}$. (Recall that the mean time course was previously subtracted from the data). The PC time courses, $\{V_k\}$, are the K eigenvectors of \mathbf{C} corresponding to the K largest eigenvalues, denoted S_k^2 . In practice we sometimes excluded some of the PCs with the largest eigenvalues, sacrificing some of the accuracy of the reduced description of the data for the sake of rejecting some high-variance noise sources. We thus typically retained all components with index $k \in [K_{low}, K_{high}]$, so that the dimensionality of the reduced data was $K = K_{high} - K_{low} + 1$. Since we were only interested in the reduced data set, we needed to compute at least $\sim K_{high}$ eigenvectors and eigenvalues, not the full set of T components. We usually computed 200 PCs for all data sets in a batch pre-processing step, and then chose the two numbers $\{K_{low}, K_{high}\}$ for each data set in a subsequent semi-automated fashion (see below). Once we determined \mathbf{V} and \mathbf{S} in this way, we computed \mathbf{U} using: $\mathbf{U} = \mathbf{MVS}^{-1}$.

This procedure for PCA, in which we computed the temporal rather than the spatial covariance matrix, allowed us to analyze efficiently data sets with a very large number of pixels. This was because \mathbf{C} and \mathbf{U} were computed by summing over all movie pixels, which was performed by loading the data from only a small number of movie pixels at each stage. There was therefore no need to store the entire data matrix, \mathbf{M} , in memory at any time. The limiting resource for this procedure is the amount of memory required to store the matrix \mathbf{C} , which scales as N_t^2 .

An important step in performing PCA is the choice of the two values, $\{K_{high}, K_{low}\}$, that specify the range of PCs retained in the reduced data set. To set these parameters we

calculated the set of variances, $\{S_k^2\}$, of the data principal components. We used the analytic formula for singular values of a Gaussian random matrix to find a reference set of variances for a movie of the same dimensions containing independent, Gaussian distributed random noise in each pixel and time frame (Sengupta and Mitra, 1999). By choosing the power of the noise in the reference movie so that the highest-order principal component variances match those of the data, we could compare its variance spectrum with the data to determine the data principal components whose variances were well above the noise floor. We chose the cutoff, K_{high} , to be roughly the point at which the slopes of the signal and noise spectra converged. All data dimensions whose covariance rank was greater than K_{high} were eliminated from further analysis. To reduce artifacts due to motion and galvanometer jitter, which may have large variance, we visually examined the spatial filters (the rows of \mathbf{U} , reshaped and displayed as 2D images) and chose a lower cutoff, K_{low} , to eliminate modes with large variance that did not appear to contain biological signals. In practice, we usually set $K_{low} \leq 15$. With experience, setting $\{K_{high}, K_{low}\}$ can be done quickly and requires only a few minutes.

Following dimensional reduction, all subsequent analysis made use of a normalized, or whitened, representation of the data with equal covariance along every PC: $\mathbf{M}_{white} = \mathbf{U}\mathbf{V}^T$ (Hyvarinen and Oja, 2000). Working with this representation simplified the process of extracting cellular signals as orthogonal transformations of the PCs.

Cell Sorting by Independent Component Analysis (ICA)

The problem of cell sorting, as schematized in Figure 1A, can be described by the generative model for signal mixtures that is at the heart of the theoretical framework for ICA. According to this model, independent biological signals mix linearly to form the data matrix:

$$[\mathbf{M}_{white}]_{x,t} = \sum_j s_j(x) a_j(t).$$

Here the fluorescence at pixel x at time t is determined by the instantaneous amplitudes of the independent signal time courses, $a_j(t)$, and by a set of corresponding spatial filters, $s_j(x)$. In temporal ICA, $s_j(x)$ is regarded as a matrix of mixing coefficients for the independent signal sources, $a_j(t)$, that combine to form the mixtures observed at each pixel. In spatial ICA, these

roles are reversed, with the $a_j(t)$ serving as mixing coefficients for the spatial sources defined by $s_j(x)$.

ICA seeks to estimate $a_j(t)$ and $s_j(x)$ by finding an unmixing matrix, \mathbf{W} :

$$\begin{aligned}\hat{a}_j(t) &= \sum_k \mathbf{W}_{jk} \mathbf{V}_{tk}, \\ \hat{s}_j(x) &= \sum_k \mathbf{W}_{jk} \mathbf{U}_{xk},\end{aligned}$$

where the circumflex indicates the estimate based on the movie data. We find the unmixing matrix by maximizing statistical sparseness, as quantified by the skewness of the spatial and temporal signals. The skewness of a random variable, z , is related to its third order statistical moment:

$$Skew[z] = E[(z - E[z])^3] / E[(z - E[z])^2]^{3/2} = E[\tilde{z}^3]$$

where E denotes an expectation value over the set of observations of z , and $\tilde{z} = (z - E[z]) / \sigma_z$ has been centered and normalized to have zero mean and unit variance. We performed ICA using the *FastICA* algorithm (Hyvarinen and Oja, 2000), which uses an approximation of the Newton-Raphson method to maximize the skewness objective function (Hyvarinen, 1999). This objective function for joint analysis of a set of signals, $\{z^{(j)}\}$, is the total skewness,

$\sum_j Skew[z^{(j)}]$. The algorithm starts from a random initial guess for the mixing matrix and iteratively updates it to bring it closer to the maximum, by using an estimate of the local slope and curvature of the objective function at each iteration. Assuming the data have been whitened during pre-processing (see above) and the mean of each signal has been subtracted, the rule for updating \mathbf{W} is

$$\mathbf{W}_{ij}^{(n+1)} = \sum_k (z_k^{(i)})^2 y_k^{(j)}$$

where $\mathbf{W}^{(n)}$ is the estimated mixing matrix after the n 'th iteration, $y^{(j)}$ is the mixed signal from PCA (e.g. $y^{(j)} = V_j$ or $y^{(j)} = U_j$; see below), and $z^{(i)} = \sum_j \mathbf{W}_{i,j}^{(n)} y^{(j)}$ is the unmixed signal whose skewness is to be maximized. After each update, the unmixing matrix is orthogonalized via $\mathbf{W} = (\mathbf{W}\mathbf{W}^T)^{-1/2} \mathbf{W}$, which ensures that the independent components will be uncorrelated with each other (Hyvarinen, 1999). To complete the definition of the ICA algorithm used for cell sorting, we must specify the identity of the unmixed signals, $\{y^{(j)}\}$, which may include the principal component spatial filters and/or time courses.

In its most basic form, ICA can be applied to cell sorting by using either spatial or temporal statistics to assess sparseness. Temporal ICA relies on the assumption that the signals of interest are temporally sparse, *i.e.* with each cell being activated in a relatively small proportion of the movie frames and with a skewed amplitude distribution. In this case, the mixed signals are the principal component time courses, $y^{(j)} = V_j$, and the objective function is

$F_{time} = \sum_j Skew[\hat{a}_j]$, where the expectation value is an average over movie time frames. Imaging

data from populations of individual cells are generally also sparse in the spatial domain, with each cell contributing to only a small fraction of the pixels in the field of view. *FastICA* can exploit this spatial information simply by reversing the roles of space and time in the original data matrix, leading to spatial ICA and an objective function, $F_{space} = \sum_j Skew[\hat{s}_j]$. Here,

$y^{(j)} = U_j$ and the expectation values are averages over all pixels.

Cellular signals in our data sets had typically greater spatial than temporal skewness (Figure 2E). Each Purkinje cell dendritic tree occupied $1,760 \pm 589 \mu\text{m}^2$ (mean \pm s.d.; $n=199$ dendrites in 5 mice), or $\sim 3\%$ of the area of our typical $60,000 \mu\text{m}^2$ field of view. By comparison, the Purkinje cell Ca^{2+} spike rate was 0.76 ± 0.15 Hz ($n=199$) in awake, resting animals and each spike lasted 124 ± 63 ms ($n=150$ cells in 5 mice), meaning that in $\sim 14\%$ of time frames each cell was active. Consistent with this disparity, spatial ICA typically achieved higher fidelity signal extractions from our simulated data than temporal ICA (Figure S3F) and also extracted more components from experimental data that resembled cellular signals. Note, however, that sparse activation in time is not the only means of achieving a highly skewed distribution of signal amplitudes. For example, cells with high rates of activity that are strongly modulated in time, such as during bursts, will also exhibit skewed distributions of signal amplitudes.

Compared to PCA, the ICA stage of our analysis is more efficient computationally. PCA requires the raw data, and with computation of K principal components the number of operations rises as $KN_x N_t^2$ (assuming $N_x > N_t$). For the ICA stage this cost rises only as $K^2 (N_x + N_t)$. In the analysis below of our cerebellar recordings, for which $N_x \sim 10^5$, $N_t \sim 10^4$, and $K \sim 10^2$, PCA took ~ 10 -30 min of processing time, but ICA generally took < 1 min.

Spatio-Temporal ICA

To exploit sparseness in both the spatial and temporal domains, we used spatio-temporal ICA (Stone et al., 2002). In this case, a set of signals formed by concatenating time courses and spatial filters from each principal component plays the role of $\{y^{(j)}\}$. We used a single

parameter, μ , where $1 \geq \mu \geq 0$, to control the relative contribution of spatial and temporal information:

$$y_i^{(k)} = \begin{cases} \mu U_{ki}, & i \leq N_x \\ (1-\mu)V_{ki}, & N_x < i \leq (N_x + N_t) \end{cases}.$$

Note that $E[y^{(k)}] = (\mu \sum_x U_{kx} + (1-\mu) \sum_t V_{kt}) / (N_x + N_t) = 0$. As before, we set

$$z_i^{(k)} = \sum_j \mathbf{w}_{i,j}^{(n)} y_j^{(j)} = \begin{cases} \mu \hat{s}_i, & i \leq N_x \\ (1-\mu) \hat{a}_i, & N_x < i \leq (N_x + N_t) \end{cases}.$$

The objective function is therefore $F_\mu[C] = \sum_j \text{Skew}[y^{(j)}] \propto \sum_j \mu^3 \text{Skew}[\hat{s}_j] + (1-\mu)^3 \text{Skew}[\hat{a}_j]$, a weighted sum of both the spatial and temporal skewness of the extracted signals.

Spatio-temporal ICA may be viewed as a version of temporal ICA that combines a temporal skewness objective with the *a priori* assumption, weighted by a factor $\mu^3 / (1-\mu)^3$, that the spatial filters should be sparse (Karthikesh, 2000). The algorithm treats space and time symmetrically, so it can be viewed alternatively as maximizing spatial skewness under the constraint that the signal time courses should also be skewed; in this case, the prior weight is $(1-\mu)^3 / \mu^3$. Nonetheless, there remains a need for further theoretical investigation of the statistical interpretation of spatio-temporal ICA in terms of a generative model with prior assumptions on the filter weights (Karthikesh, 2000; Stone et al., 2002; Zhang and Chan, 2006).

Temporal Deconvolution and Spike Detection

We estimated the stereotyped waveform of a neuronal Ca^{2+} spike by averaging the time courses of spikes simultaneously recorded optically and by extracellular single unit electrical recording in anesthetized mice. We fit the average time course of these spikes with an exponentially decaying function of time constant $\tau = 182 \pm 97$ ms (mean \pm s.d., $n=155$ spikes in 4 experiments). Because this time course was brief, we could improve the detection of Ca^{2+} transients in neuronal signals by high-pass filtering the signals to first remove slow variations due to sources such as photobleaching. We filtered signals by first estimating the slow variations using a moving average with 2 s window size, then subtracting this smooth signal from the original. This approach was not used with glial Ca^{2+} signals, which evolve over longer time scales of several seconds.

For neuronal signals we corrected for the slow decay of fluorescence by using temporal deconvolution (Yaksi and Friedrich, 2006). The deconvolved signal was defined by

$\hat{s}_{deconv}(t) = \hat{s}(t) / \tau_{deconv} + [\hat{s}(t + dt) - \hat{s}(t)] / dt$, where $\tau_{deconv} = 150$ ms and dt is the time between movie frames. The deconvolved signal for Purkinje cell Ca^{2+} traces contained one large spike for each complex spike, followed by an immediate return to baseline in the next time frame. We defined spike times as local maxima in the deconvolved signal that exceeded a threshold defined as 2 standard deviations above the mean signal. Applying this spike detection procedure in our awake data sets, we found that the average time constant of decay following each detected spike was $\sim 124 \pm 63$ ms (mean \pm s.d., $n=150$ cells in 5 mice), within the range observed in our anesthetized recordings with paired electrophysiology.

Image Segmentation

Spatio-temporal ICA sometimes identified individual spatial filters that appeared to comprise multiple, spatially separated Purkinje cells. We therefore segmented these filters into distinct sub-components, each representing an individual Purkinje cell. To do this, the original filter was smoothed by a convolution with a Gaussian kernel of 1.5 pixels ($2.9 \mu\text{m}$) standard deviation. We next transformed the filter into a binary mask by applying a threshold 1.5 standard deviations above the mean intensity of all pixels. We used the MATLAB function `bwlabel`, which gives distinct labels to all spatially connected components within the binary mask (Haralick and Shapiro, 1992). In some cases we identified small image segments that were cut off by the edges of the field of view, or small pieces of dendrites separated from the main processes by a blood vessel. In such cases the signal amplitude relative to background fluctuations was often smaller than for dendritic segments with larger area, due to the smaller number of photons contributing. To avoid excessively noisy signals, we excluded any image segments covering an area < 50 pixels. (For comparison, Purkinje cell dendrites typically occupy 481 ± 161 pixels (mean \pm s.d.) in our recordings). Finally, for each image segment we created a spatial filter filled by setting to zero the weights of all pixels outside of the connected region. The filter weights inside the connected region were equal to those of the corresponding pixels in the original filter. The result was a set of spatial filters that had non-overlapping support and retained the fine spatial structure of the original filter.

Artificial Cerebellar Movie Data Sets for Validation of Cell Sorting Procedures

We created artificial movies designed to mimic Ca^{2+} imaging data acquired in the cerebellar molecular layer (Figures 3, S1, and S3). The artificial data were constructed by simulating

cellular signals with temporal dynamics and spatial characteristics, such as cell sizes, shape, and density, conforming to our expectations of intracellular Ca^{2+} signals from Purkinje cell dendrites and Bergmann glial processes. Each movie covered a square field of view either 32×32 (Figure S3), 50×50 or 64×64 (Figures 3, S1) pixels in size. The number of Purkinje cells and total rate of Bergmann glial Ca^{2+} transients increased in proportion to the field of view area, so as to maintain a constant density of ~ 1000 Purkinje cells per mm^2 and ~ 100 glial Ca^{2+} transients per mm^2 per second. This density of Purkinje cells was somewhat higher than the density of cellular signals we extracted from experimental data sets (663 ± 45 per mm^2), reflecting the presence of blood vessels occluding parts of the field of view in the experimental data as well as the possibility that our analysis missed some of the cells that were present.

We distributed the centroids of the Purkinje cell dendritic trees, $\{x_j, y_j\}$, randomly within the middle 64% of the field of view; we chose not to include cells in the periphery of the field of view so that all of the simulated cells would cover roughly the same number of pixels. We modeled Purkinje cell signals by assuming that the pixels near the centroid of the dendrite contribute the most weight to the signal. The relative weight of each pixel in a cell's spatial filter was given by an elongated Gaussian function:

$$s_j(x, y) \propto \exp[-(x - x_j)^2 / 2\sigma_x^2 - (y - y_j)^2 / 2\sigma_y^2]$$

where $\sigma_x = 110 \mu\text{m}$, $\sigma_y = 3.5 \mu\text{m}$ defined the major and minor axis standard deviations. We imposed a sharp spatial cutoff, setting s_j to zero for all pixels whose weight would have been 0.2% that of the center pixel. These simulated Purkinje cell dendrite shapes were roughly consistent with the total length ($90 \pm 40 \mu\text{m}$) and breadth ($7.0 \pm 2.3 \mu\text{m}$) of Purkinje cell dendrite image segments observed in our experimental data. Bergmann glial spatial filters were defined identically but had isotropic width parameters, $\sigma_x = \sigma_y = 30 \mu\text{m}$. We rotated all spatial filters relative to the square borders of the field of view by 20° . Finally, we normalized each spatial filter so that the maximum intensity of any pixel within the field of view was, on average, unity.

The time courses of Ca^{2+} -dependent fluorescent signals, $a_j(t)$, were modeled by first generating underlying spike trains by using independent, biased random coin flips for each Purkinje cell and for each time frame. The probability of a spike in each time frame was chosen to give a mean spike rate between 0.6 and 0.8 Hz, within the range reported for non-anesthetized mice (Flusberg et al., 2008; Schiffmann et al., 1999; Servais et al., 2004). We convolved each cell's spike train, represented as a vector of 1's and 0's, with an exponentially decaying function with time constant 150 ms to account for the slow decline of Ca^{2+}

fluorescence following a spike. The spike waveform was normalized so that its integral over time was unity.

For simplicity and to avoid multiple simultaneous glial Ca^{2+} transients, which we rarely observed in the experimental data from awake, resting mice, we chose onset times for glial events, t_i , that were distributed at regularly spaced intervals over the duration of the simulated movies. The time course of each glial Ca^{2+} transient was modeled by the function $(t - t_i)e^{-(t-t_i)/\tau}$, where $\tau = 1.6$ s.

Artificial movies were 100 s (Figures 3 and S1) or 200 s (Figure S3) in duration and the frame rate was 10 Hz. Cellular signals were added to a static fluorescence background, $F_0(x)$, that included bright circular structures resembling the somata of molecular layer interneurons and elongated dark regions resembling blood vessels. The background image was normalized so that the brightest pixel was half as bright as the maximum brightness of the signal spatial filters. These time-dependent and static background signals defined a noiseless data set.

We modeled photon shot noise and other sources of variability using Poisson distributed intensity values. The probability of observing n photons in pixel, x , and time frame, t , was

$$\Pr[n] = e^{-\bar{n}_{x,t}} \frac{(\bar{n}_{x,t})^n}{n!}$$

where the mean intensity is given by $\bar{n}_{x,t} = A \sum_j s_j(x) a_j(t) + B F_0(x)$ and A, B are constant factors

defining the signal-to-noise ratio, S , and the contrast of the background image. We set $B = 5000$ to model typical brightness levels in our cerebellar data sets. We then chose A using a formula based on the desired values of S and the mean spike rate, r ,

$$A = \frac{S}{2} \left[1 + \sqrt{1 + 4 \frac{B}{Sr}} \right].$$

When analyzing artificial data sets we always chose the number of retained PC's after dimensional reduction, K , to be equal to or slightly greater than the total number of neuronal and glial signals. We set $K_{low} = 1, K_{high} = K$. We used $\mu = 0.5$ when performing spatio-temporal ICA (Figures 3 and S1) since this value was in the range for which cell sorting performed optimally for artificial data (Figure S3). The optimal value of μ for each data set depended on factors such as the average spike rate of cells, the imaging frame rate, the number of pixels and the number of time frames. Our experimental data had a larger ratio of pixels to time frames than our

artificial data sets, which led to our choice of $\mu = 0.1 - 0.2$ for analyzing the experimental data. For computational efficiency we limited the number of iterations of the gradient descent algorithm to 500.

Artificial Neocortical Movie Data Sets

To test the generality of our analysis procedure, we created simulations designed to mimic neocortical data sets. Simulated neocortical movies were created using a procedure similar to those of cerebellar simulations, but with modifications in a number of parameters reflecting differences in the morphology and dynamics of cellular signals. Neocortical simulations also included signals originating from the neuropil in the interstitial space between cell bodies.

Each movie covered a square field of view $100 \times 100 \mu\text{m}$ in size, with 50×50 pixels. The centroids of 13 somata were chosen randomly from a uniform distribution over the field of view, corresponding to a cellular density of 1300 somata per mm^2 . Each soma covered a circular area $6 \mu\text{m}$ in diameter. We calculated the relative weight of each pixel's contribution to a soma's spatial filter, $s_j(x)$, by finding the area of intersection of this circle with the pixel. Thus, pixels contained entirely within the interior of the circular soma contributed maximally to the spatial filter, while pixels on the edge of soma contributed in proportion to the overlap of the circle with the pixel. Neuropil signal originated from all parts of the field of view without a cell body. Thus, pixels completely outside of all somata contributed maximally to the neuropil spatial filter, whereas pixels on the edge of a soma contained a contribution from neuropil in addition to somatic signals. We blurred the cellular and neuropil spatial filters with a Gaussian kernel ($0.8 \mu\text{m}$ s.d.) to simulate the point spread function of a two-photon microscope in which the laser beam moderately under fills the objective back aperture (Gobel and Helmchen, 2007; Kerr and Denk, 2008). Finally, we normalized each spatial filter so that the maximum intensity of any pixel within the field of view was unity.

The time courses of somatic Ca^{2+} -dependent fluorescence signals, $a_j(t)$, were modeled by assuming that individual "events" correspond to bursts of 1-10 action potentials within one movie time frame (300 ms). Event onsets were randomly distributed with uniform probability density throughout the movie, with mean rate 0.1 Hz. We modeled each event with a time course proportional to $(t - t_0)e^{-(t-t_0)/\tau}$, with decay time constant $\tau = 200$ ms. The neuropil signal was modeled as a low frequency, sinusoidal oscillation with frequency 0.5 Hz (Kerr et al., 2005). Each artificial movie was 300 s in duration and the frame rate was 3.3 Hz. The fluorescence

background, $F_0(x)$, contained dark areas representing vessels and more brightly labeled regions, as in our experimental data. Poisson distributed shot noise was introduced as in our cerebellar simulations (see above).

Definitions of Signal Fidelity and Cross Talk

To quantify the performance of automated cell sorting, we compared the signals extracted from our artificial data sets to the underlying cellular signals used to generate the artificial movie. We computed the correlation coefficient between each true signal, $a_k(t)$, and each extracted signal trace, $\hat{a}_j(t)$:

$$cc_{j,k} = \frac{E[\hat{a}_j a_k] - E[\hat{a}_j]E[a_k]}{\left(E[\hat{a}_j^2] - E[\hat{a}_j]^2\right)^{1/2} \left(E[a_k^2] - E[a_k]^2\right)^{1/2}}$$

where E denotes averaging over all time frames. We then paired each extracted signal with one of the true signals in an iterative fashion. First, we identified the pair of signals with the largest correlation coefficient and assigned these two signals to each other. We then removed these two signals from the set of available pairs and repeated the procedure with the remaining true and extracted signals. We proceeded in this way until each extracted signal was assigned to one true signal. We defined the fidelity for each extracted signal as the cross correlation with the true signal with which it was paired in this process. The correlation coefficients between all the other true signals and the extracted signal yielded the cross talk. Our overall measure of cross talk is the median of the K largest such correlation coefficients, where K is the total number of extracted signals. In Figures S4C and S4D showing results from neocortical simulations, cross talk was measured as the correlation between each cell's signal and the neuropil signal.

Linear Regression Analysis

We compared ICA to a theoretical upper bound on its performance provided by a linear regression analysis. The latter computes an optimal unmixing matrix, \mathbf{W}^{opt} , that minimizes the mean square difference between each cell's true signal, $a_j(t)$, and the regression estimate,

$\hat{a}_j(t) = \sum_k \mathbf{W}_{jk}^{opt} \mathbf{V}_{tk}$. The optimum is given by $\mathbf{W}^{opt} = \mathbf{C}^{-1} \mathbf{V}^T \mathbf{A}$ where $\mathbf{C} = \mathbf{V} \mathbf{V}^T$ is the covariance matrix of the principal component time courses. Note that this procedure uses the matrix of true signal time courses, \mathbf{A} , and thus should outperform blind signal extraction methods like ICA. To estimate the performance of this regression estimator without over fitting, we divided the

simulated movie data into disjoint “training” and “testing” sets. 70% of movie time frames were randomly chosen for the training set, and the unmixing matrix \mathbf{W}^{opt} was computed using these frames. The remaining 30% of movie frames were used to provide an independent measure of the fidelity and crosstalk of the signals extracted by this method.

Region of Interest (ROI) Analysis

We also compared ICA to an analysis method designed to represent a best-case scenario for the results a human data analyst could achieve with a region of interest (ROI) approach. The underlying assumption is that for each cell the analyst possesses several examples of movie time frames that she knows to contain examples of activations of that cell, caused by spikes in neurons or Ca^{2+} transients in glial processes. In practice, uncertainty and error in the estimation of these exemplary spike times will degrade the performance of an ROI cell sorting procedure compared with the idealized version we studied. Thus, our results likely understate the superiority of ICA over ROI analysis.

We defined ROIs for each Purkinje dendrite by computing the average of the simulated fluorescence in each pixel over all movie time frames coinciding with the occurrences of a simulated spike in that cell. This spike-triggered average generally yielded a filter with high amplitudes in the pixels contained within the cell, and low amplitudes in pixels reflecting only background fluctuations in fluorescence or signals from other, uncorrelated cells. We used a similar approach to define an ROI for each glial Ca^{2+} transient. Each glial signal contained only one fluorescence transient during the simulated movie, but the transient persisted ~ 10 times longer than neuronal action potential transients. We therefore defined glial ROIs by averaging fluorescence during a period of 500 ms duration (5 movie time frames) around the peak of the transient.

For both neuronal and glial ROIs, we smoothed the triggered average image by convolving with a Gaussian blurring function of 2 pixels standard deviation. We then applied a threshold, defined as 0.8 times the filter’s maximum amplitude, to generate a binary mask. This mask defined the final ROI (Figure 3B). To compute the signal corresponding to each ROI, we averaged the fluorescence in all pixels within the corresponding binary mask for each time frame.

Animals and Animal Surgery

Mouse surgeries were performed under isoflurane anesthesia (1.5–2.5%, mixed with 1.0–2.0 L/min O_2). Body temperature was monitored with a rectal thermistor placed underneath the

animal's belly and maintained at 36-37°C using a heating blanket. We assessed the depth of anesthesia by monitoring pinch withdrawal, eyelid and corneal reflexes, breathing rate, and vibrissae movements. Our experiments involved two surgical procedures. Several days prior to recordings the animal's skull was exposed and cleaned above neocortex and cerebellum. A custom metal plate allowing cranial access and repeated head-fixation was attached to the skull with dental acrylic cement (Coltene/Whaledent, H00335). On the recording day, a craniotomy (1-4 mm in diameter) was opened (cerebellar vermis: bregma -6.5 mm; 0.5 mm lateral; neocortex: bregma \pm 0.0 mm; 2.0 mm lateral). Exposed tissue was irrigated with warm artificial cerebral spinal fluid (ACSF; 125 mM NaCl, 5 mM KCl, 10 mM D-Glucose, 10 mM HEPES, 2 mM CaCl_2 , 2 mM MgSO_4 ; pH to 7.4 with NaOH). To dampen heartbeat- and breathing-induced brain motion, the craniotomy was filled with agarose (1.5-2.0%; Type III-A, highEEO; Sigma) in ACSF and covered with a coverslip immobilized onto the head plate, creating an optical window for imaging. The dura was left intact for Ca^{2+} -indicator injections, but removed immediately prior to agarose application.

Fluorescence Labeling

Targeted bulk-loading of cerebellar or neocortical tissue with the fluorescent Ca^{2+} -indicator Oregon green 488 1,2-bis(2-aminophenoxy)ethane-N,N,N',N'-tetraacetate-1 acetoxymethyl (OGB-1-AM; Molecular Probes) was performed as described (Sullivan et al., 2005; Nimmerjahn et al., 2004; Nimmerjahn et al., 2009; Stosiek et al., 2003). The final concentration of OGB-1-AM and DMSO in our pipette solution was 500 μM and 2-5%, respectively. To obtain a homogenous, widespread stain of superficial cortical layers we performed several pressure ejections (3-3.5 psi pressure) of 1-2 min duration using a picospritzer (Picospritzer III, General Valve, Fairfield, NJ). Typically five injections were performed in a grid-like pattern with \sim 150-250 μm spacing. To counterstain astroglia, the red fluorescent astrocyte marker sulforhodamine 101 (SR101; 200-250 μM in ACSF; Invitrogen) was generally added to the injection solution or applied topically (1-5 min) (Nimmerjahn et al., 2004; Nimmerjahn et al., 2009).

Two-Photon Imaging in Behaving Mice

Two-photon imaging was done as previously described (Nimmerjahn et al., 2009). In brief, after surgical implantation of a head-plate into the cranium and a recovery period of 1-4 days, mice were habituated in 3-9 sessions (typically 1 session/day) to accept increasingly longer periods (15-90 min sessions; 1.5-7.5 h total habituation time) of head restraint while walking on a 11.9-cm-diameter exercise ball (Habitrail, model 62065) covered with copper mesh. The ball rotated

on one axis, allowing the animal to stride forward or backward at liberty. We used a custom upright two-photon microscope equipped with an ultrashort pulsed Ti:sapphire laser (Tsunami, Spectra-Physics), two fluorescence detection channels, and a 20x water-immersion objective (Olympus, 0.95 NA, XLUMPlanFI). To excite OGB-1-AM, the laser was tuned to 800 nm. For simultaneous excitation of OGB-1-AM and SR101 we used 800-820 nm. Prior to Ca^{2+} imaging mice generally recovered from isoflurane for >60 min to minimize aftereffects of anesthesia. To minimize the possibility of photo-induced alterations in tissue physiology, the duration and average power of continuous illumination were <12 min and <35 mW at a given brain location. We used MP-Scope data-acquisition software for laser-scanning microscopy (Nguyen et al., 2006), which uses a 5 MS/s data acquisition card (PCI-6110, National Instruments) for simultaneous acquisition and display of up to two imaging and two analog input channels. Images contained 256 x 256 pixels (10.2–11.1 Hz frame rate) or 128 x 128 pixels (20.4 Hz).

Behavioral Assessment

Rotations of the exercise ball were monitored using an optical encoder (US Digital, E7PD-720-118). In addition, mouse behavior was recorded in Mini-DV format using a digital Camcorder (Sony DCR-VX2000).

Electrophysiology

To record the electrophysiological complex spikes corresponding to optically detected Ca^{2+} -spikes in Purkinje cell dendrites (Flusberg et al., 2008; Ozden et al., 2008; Welsh et al., 1995), we performed targeted extracellular single-unit recordings in the cerebellar cortex of isoflurane anesthetized mice studied simultaneously by two-photon imaging as described above. Ca^{2+} -transients in distal Purkinje cell dendrites were visualized using a 40x 0.8 NA water-immersion objective (Zeiss, Achroplan). Extracellular Purkinje cell activity was recorded using beveled borosilicate glass electrodes (A-M Systems, Inc.; OD, 1.0 mm; ID, 0.58 mm; internal solution, 2 M NaCl and in some cases 200-500 μM sulforhodamine 101; pipette resistance, 5-10 M Ω). Electrodes were positioned under visual control near OGB-1-AM labeled Purkinje cell somata. Purkinje cells were identified by the brief pause in simple spike activity following each complex spike. The raw electrode signal was amplified, filtered (A-M Systems, Model 1800 microelectrode AC amplifier; low cut-off, 300 Hz; high cut-off, 5 kHz), digitized (20 kHz) and stored on disk for off-line analysis. Two-photon imaging data of dendritic Ca^{2+} -activity was recorded simultaneously (frame rate, 20.35 Hz) in superficial regions of the molecular layer (depth, 20-50 μm) located directly above the microelectrode tip. The correspondence between

dendritic Ca^{2+} -spikes and electrophysiological complex spike activity was verified through a spike-triggered average analysis written in MATLAB, in agreement with prior combined optical and electrophysiological studies of Purkinje cells (Flusberg et al., 2008; Ozden et al., 2008).

Electrical Stimulation of Limbs

To test whether evoked Ca^{2+} spikes could be detected by our cell sorting procedures, we electrically stimulated the limbs while the mouse was under anesthesia. Electrical stimuli were delivered via a pair of 30-gauge needles attached to a stimulus isolator (Isostim A320; World Precision Instruments) and connected to a 12-bit high-speed 1MS/s analog output board (PCI-6711, National Instruments). Stimulation was controlled by custom software written in LabVIEW (Version 7.1; National Instruments) and consisted of a train of 60 square pulses, each of 20 ms duration and 0.5-10 mA amplitude, delivered at 20 Hz. Electrical stimulation caused limb withdrawal that was highly noticeable. Signals delivered to the stimulation needles were recorded at 10 kHz simultaneously with optical recordings, using MP-Scope data-acquisition software (see above).

Receiver Operating Characteristic (ROC) Analysis

Receiver operating characteristic (ROC) analysis characterizes the tradeoff between two competing objectives in the detection of spikes. On the one hand, the threshold for a spiking event should be set low enough so that most of the true spikes are correctly detected (a high hit rate). On the other hand, the threshold should be kept high enough so that very few spikes are reported at times when no spike actually occurred (low rate of false positives). The ROC curve plots the false positive rate versus the hit rate for a range of values of the spike threshold. The area under this curve is a measure of the spike discriminability and can be interpreted as the probability of correctly distinguishing a spike from a non-spike (Fawcett, 2006).

We wrote custom MATLAB routines to generate ROC curves for paired electrophysiological and two-photon imaging recordings, and to compute the ROC area. After extracting an estimate of each cellular signal by ICA cell sorting, we applied temporal deconvolution and used the resulting deconvolved signal for ROC analysis.

Identification of Periods of Active Locomotion

We used an optical encoder to record the mouse's running speed on the exercise ball. The encoder was sufficiently sensitive to detect even 0.5 deg rotations of the exercise ball during periods when the mouse did not actively walk or run. We first smoothed the encoder signal by

convolving with a Gaussian kernel of standard deviation 0.5 s. We defined the onset times of active movement as times at which the smoothed velocity signal surpassed a threshold of 2 mm/s. The offset times of the periods of movement were defined as when the speed fell below 1 mm/s. As a conservative measure, to avoid including any data from actively moving periods in the analysis of resting data, we padded the period of active movement with 2 s before the onset and after the offset of each movement. This procedure allowed us to detect periods of continuous movement, which may include brief pauses lasting less than a few seconds.

Detection of Cerebellar Microzones

We partitioned all identified cells into disjoint microzones using k-means clustering (MacKay, 2003; Ozden et al., 2008), implemented using the MATLAB function *kmeans*. We represented each spike train as a vector, with value 1 for time frames at which a spike was detected and 0 otherwise. The cross-correlation between spike trains provided the distance function that we used in the clustering procedure (Welsh et al., 1995). We repeated k-means clustering for each data set 20 times with random initial cluster centroids, and chose the solution with the largest average cross-correlation between each spike train and the centroid of its cluster. To assess the stability of microzones over time, we divided each recording lasting 7-10 minutes into 1-minute duration data subsets and repeated the clustering procedure for each subset. We also performed clustering separately on data recorded during rest and active locomotion.

SUPPLEMENTAL REFERENCES

- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters* 27, 861-874.
- Flusberg, B.A., Nimmerjahn, A., Cocker, E.D., Mukamel, E.A., Barretto, R.P., Ko, T.H., Burns, L.D., Jung, J.C., and Schnitzer, M.J. (2008). High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nat Methods*.
- Gobel, W., and Helmchen, F. (2007). In vivo calcium imaging of neural network function. *Physiology (Bethesda)* 22, 358-365.
- Haralick, R.M., and Shapiro, L.G. (1992). *Computer and Robot Vision*, Vol. 1. (Addison-Wesley), pp. 28-48.
- Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10, 626-634.
- Hyvarinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw* 13, 411-430.
- Karthikesh, A.H.a.R. (2000). Sparse priors on the mixing matrix in independent component analysis. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pp. 452-477.

- Kerr, J.N., and Denk, W. (2008). Imaging in vivo: watching the brain in action. *Nat Rev Neurosci* 9, 195-205.
- Kerr, J.N., Greenberg, D., and Helmchen, F. (2005). Imaging input and output of neocortical networks in vivo. *Proc Natl Acad Sci U S A* 102, 14063-14068.
- MacKay, D.J.C. (2003). Information theory, inference, and learning algorithms. (Cambridge, UK; New York: Cambridge University Press), pp. 284-292.
- Mitra, P.P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophys J* 76, 691-708.
- Nguyen, Q.T., Tsai, P.S., and Kleinfeld, D. (2006). MPScope: a versatile software suite for multiphoton microscopy. *J Neurosci Methods* 156, 351-359.
- Nimmerjahn, A., Kirchhoff, F., Kerr, J.N., and Helmchen, F. (2004). Sulforhodamine 101 as a specific marker of astroglia in the neocortex in vivo. *Nat Methods* 1, 31-37.
- Nimmerjahn, A., Mukamel, E.A., and Schnitzer, M.J. (2009). Motor behavior activates Bergmann glial networks. *Neuron* 62, 400-412.
- Ozden, I., Lee, H.M., Sullivan, M.R., and Wang, S.S. (2008). Identification and clustering of event patterns from in vivo multiphoton optical recordings of neuronal ensembles. *J Neurophysiol* 100, 495-503.
- Schiffmann, S.N., Cheron, G., Lohof, A., d'Alcantara, P., Meyer, M., Parmentier, M., and Schurmans, S. (1999). Impaired motor coordination and Purkinje cell excitability in mice lacking calretinin. *Proc Natl Acad Sci U S A* 96, 5257-5262.
- Sengupta, A.M., and Mitra, P.P. (1999). Distributions of singular values for some random matrices. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 60, 3389-3392.
- Servais, L., Bearzatto, B., Hourez, R., Dan, B., Schiffmann, S.N., and Cheron, G. (2004). Effect of simple spike firing mode on complex spike firing rate and waveform in cerebellar Purkinje cells in non-anesthetized mice. *Neuroscience letters* 367, 171-176.
- Stone, J.V., Porrill, J., Porter, N.R., and Wilkinson, I.D. (2002). Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *Neuroimage* 15, 407-421.
- Thevenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process* 7, 27-41.
- Welsh, J.P., Lang, E.J., Sugihara, I., and Llinas, R. (1995). Dynamic organization of motor control within the olivocerebellar system. *Nature* 374, 453-457.
- Yaksi, E., and Friedrich, R.W. (2006). Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca²⁺ imaging. *Nat Methods* 3, 377-383.
- Zhang, K., and Chan, L. (2006). ICA with Sparse Connections. Paper presented at: IDEAL (Burgos, Spain, Springer).