Cell Reports, Volume *21*

Supplemental Information

Tracking the Same Neurons across Multiple Days in Ca^{2+} Imaging Data

Liron Sheintuch, Alon Rubin, Noa Brande-Eilat, Nitzan Geva, Noa Sadeh, Or Pinchasof, and Yaniv Ziv

Supplemental Figures and Legends

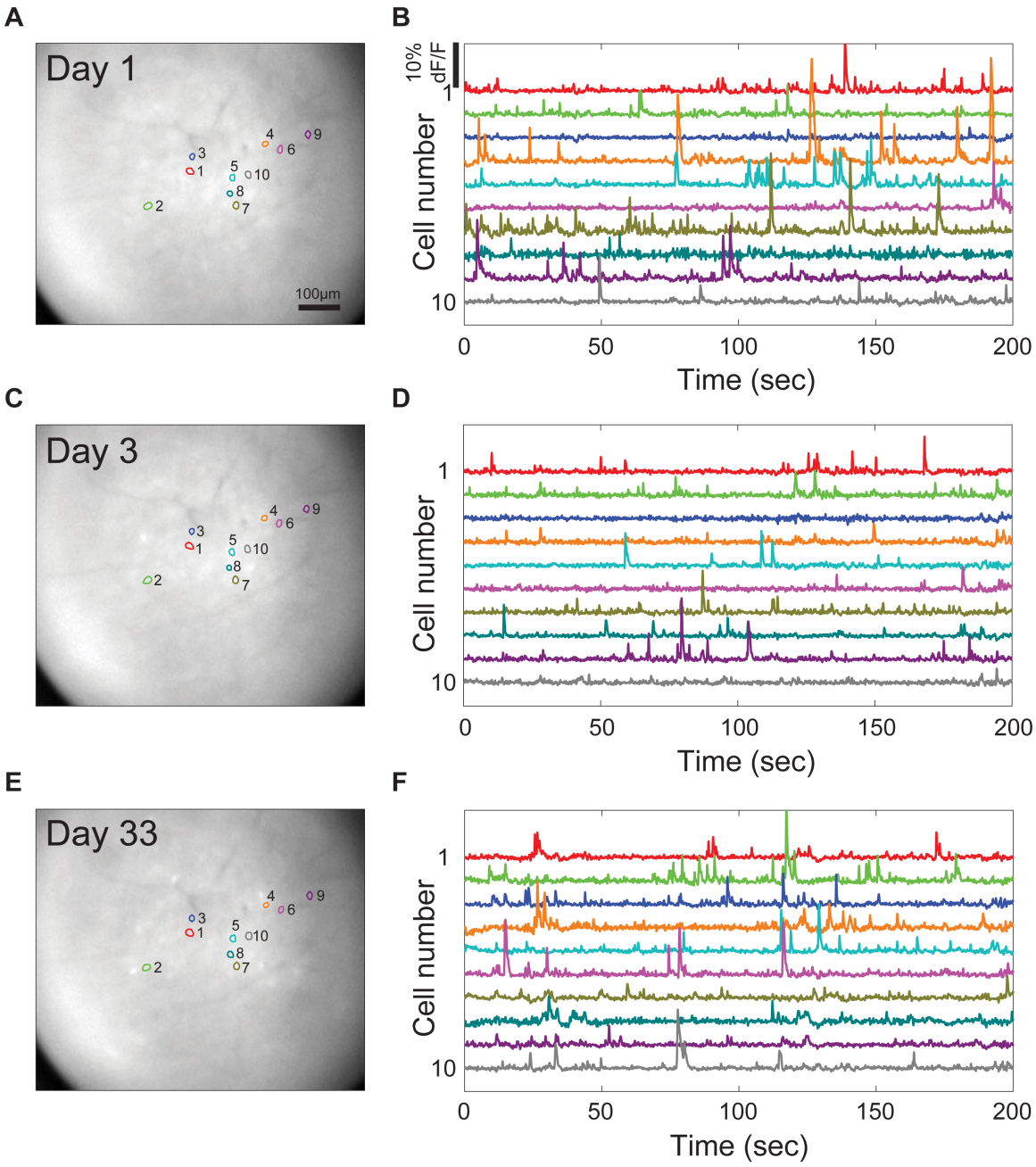

Figure S1. The same cells are tracked over three sessions spanning more than a month. Related to Figure 1.

Identification of individual cells and extraction of their Ca^{2+} dynamics using principal component analysis and independent component analysis. (A,C,E) Contours of 10 spatial footprints overlaid on a representative single frame from day 1 (A), day 3 (C) and day 33 (E). (B,D,F) Corresponding Ca^{2+} traces for the same cells shown in A. Typical Ca^{2+} dynamics are observed for individual neurons. Data taken from three sessions recorded in the hippocampal CA1 of a Thy1-GCaMP6f transgenic mouse while freely exploring the same environment.

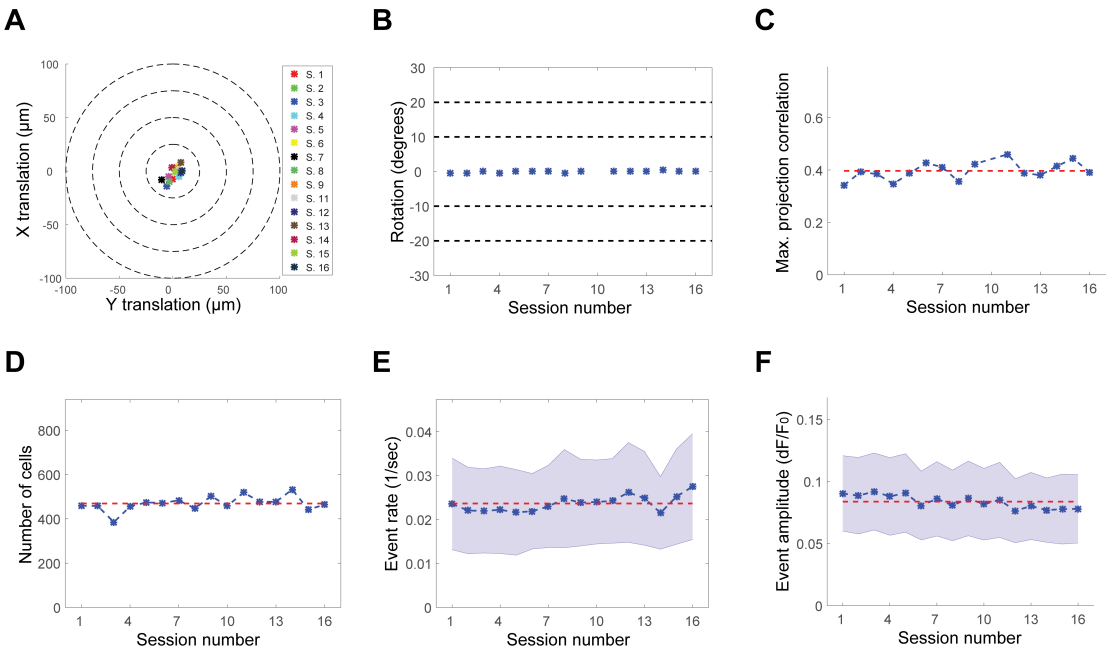

Figure S2. A stable preparation is maintained across all days of the experiment. Related to Figure 1.

(A,B) The translations (A) and rotations (B) that result in maximal correlation between the projection of the spatial footprint centroids in each session and that of the reference session. Session ten was chosen as a reference. (C) The maximal correlation between the projections of the spatial footprint centroids in each session and that of the reference session (blue asterisks), and the average maximal correlation across all sessions (red dashed curve). (D) The number of detected cells in each imaging session (blue asterisks) and the average number of cells across all sessions (red dashed curve). (E) The average cell event rate in each imaging session (SD is shown as shaded blue area), and the average cell event rate across all sessions (red dashed curve). (F) The average cell event amplitude in each imaging session (SD is shown as shaded blue area), and the average cell event amplitude across all sessions (red dashed curve). A-C can potentially point out deviations in the FOV of each session compared to the reference session, while significant deviations in the cellular activity detected in each session compared to the average cellular activity could potentially be uncovered in D-F. Data in all panels was taken from 16 sessions recorded on eight different days in the hippocampal CA1 of a mouse while freely exploring the same environments.

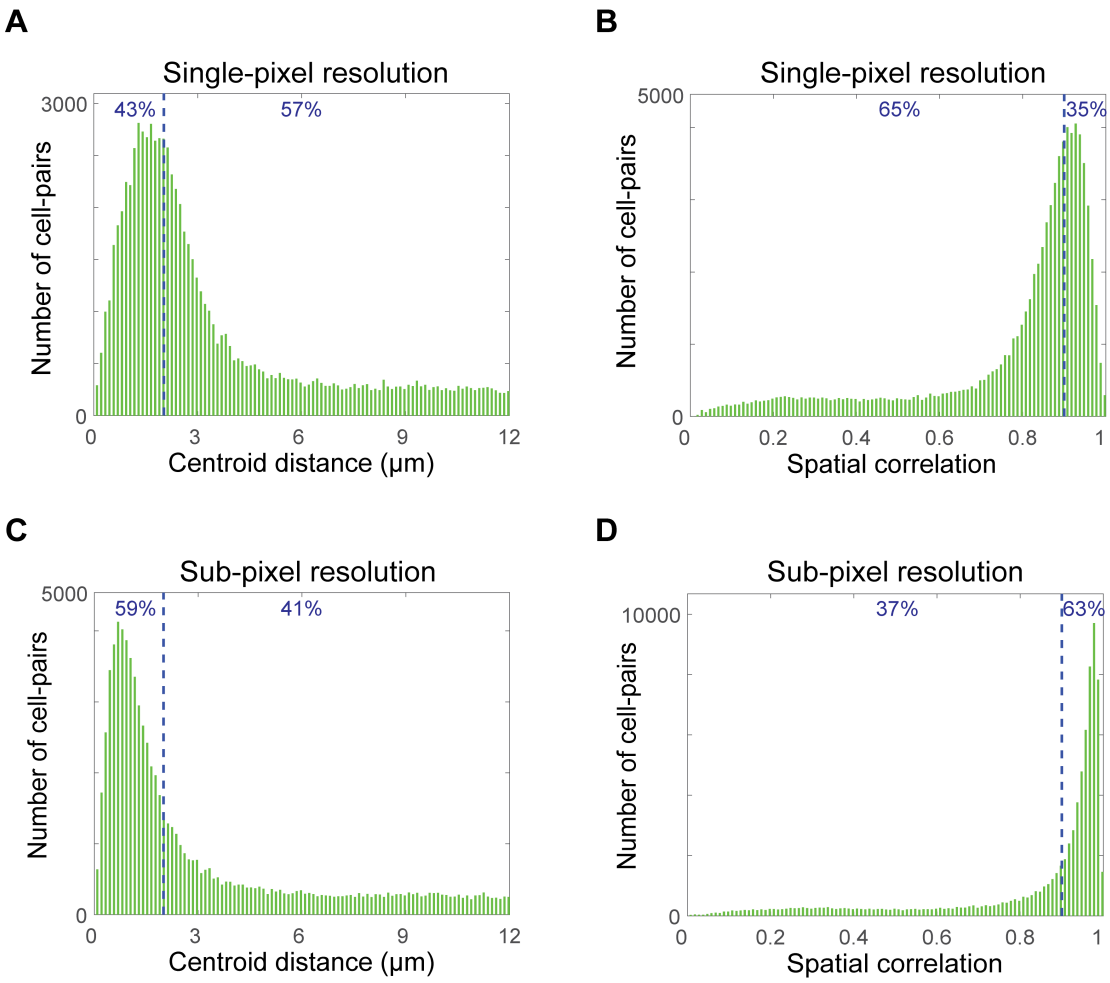

Figure S3. Image alignment with sub-pixel resolution increases the spatial footprint similarity between pairs of nearest neighbor cells. Related to Figure 2.

(A-D) Distributions of centroid distances and spatial correlations between pairs of nearest neighbor cells across days obtained after alignment (rigid-body registration) with single-pixel (A and B) and sub-pixel (C and D) resolution. We limited the distribution to cell-pairs with centroid distances <12μm. The blue dashed lines show centroid distances =2μm and spatial correlation =0.9. Note the higher probabilities for nearest neighbors to have centroid distances <2μm and spatial correlation >0.9 for alignment with sub-pixel resolution in comparison with their single-pixel counterparts. This fine spatial resolution of alignment was attainable because of the large number of centroid locations that were used to align each session. The data presented here were pre-processed with a pixel size of 2.3X2.3μm which is not negligible compared to the typical cell size. The effect of sub-pixel resolution on alignment should be less significant for imaging data with higher spatial resolution. Data in all panels was taken from 16 sessions recorded on eight different days in the hippocampal CA1 of a mouse while freely exploring the same environments.

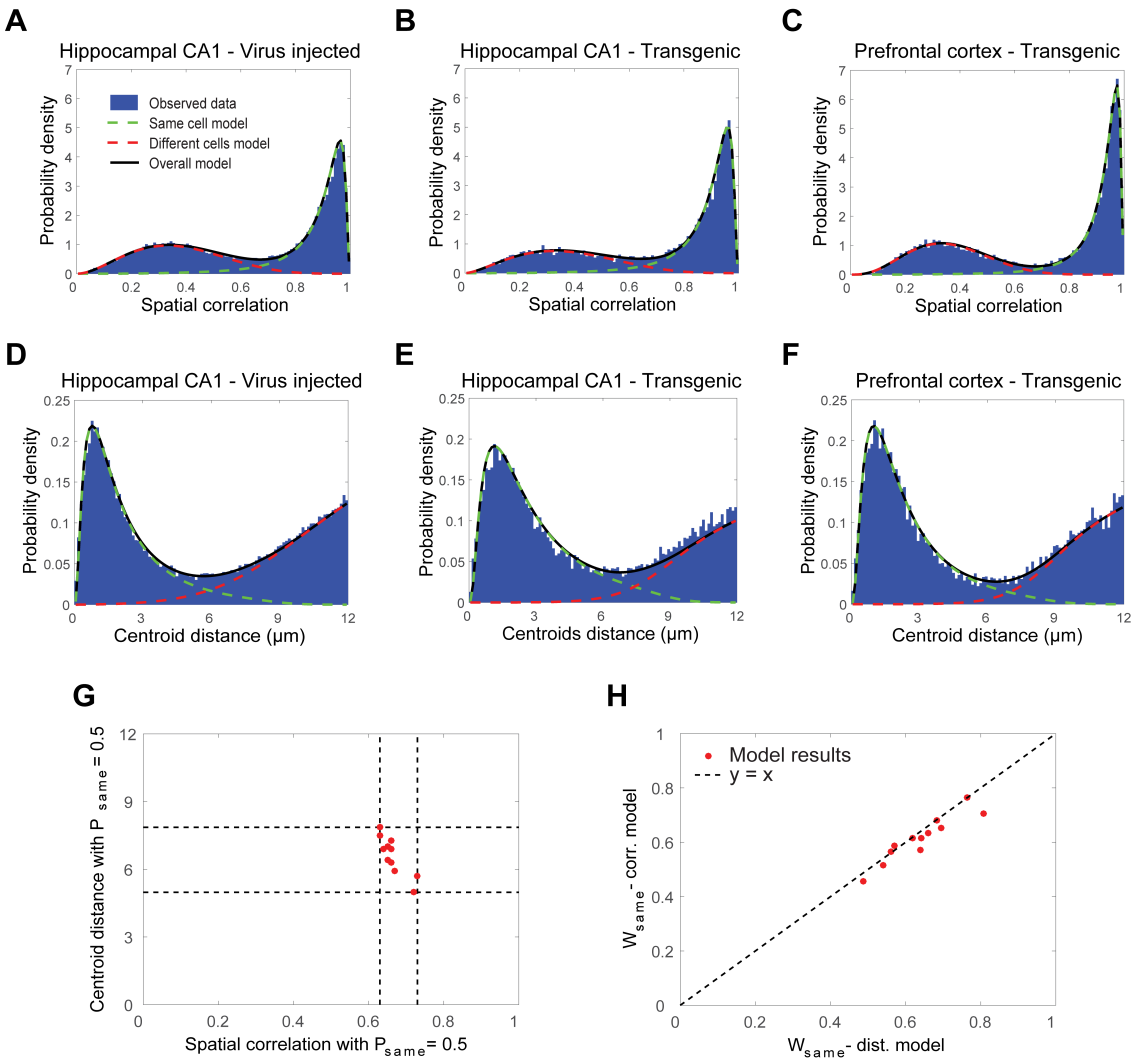

Figure S4. Examples of the distributions of spatial correlations and centroid distances modeled for each mouse individual. Related to Figure 2.

(A-F) Distributions of spatial correlations (A-C) and centroid distances (D-F) in three different mice for neighboring cell-pairs (blue bars), and the modeled distributions of same cells (dashed green curves), different cells (dashed red curves), and their weighted sum (solid black curves). Estimated distributions were obtained by assuming that the data consists of a weighted sum of two subpopulations, corresponding to same cells and different cells, and finding the parameters that best fit the data. Data and models in A and D are for 16 sessions recorded on eight different days in the hippocampal CA1 of a GCaMP6f virus injected mouse while freely exploring the same environments. Data and models in B and E are for seven sessions recorded on seven different days in the hippocampal CA1of a Thy1-GCaMP6f transgenic mouse while freely exploring the same environment. Data and models in C and F are for six sessions recorded on six different days in the prefrontal cortex of a CaMKII-GCaMP6s transgenic mouse while freely exploring the same environments. (G) The intersection between the estimated distribution of same cells and different cells (corresponding to a registration threshold of P_{same}=0.5) for the centroid distances model versus the spatial correlations model. Each dot represents a different mouse (N=12). Black dashed lines show the minimum and maximum of these values across all mice. Note that data from different mice yielded different registration thresholds. (H) The weight of the subpopulation of same cells (W_{same}) obtained by the centroid distances (dist.) model versus W_{same} obtained by the spatial correlations (corr.) model. Each dot represents a different mouse (N=12). Black dashed line represents the y=x curve. Note that the spatial correlations and centroid distances models independently reach similar W_{same} values.

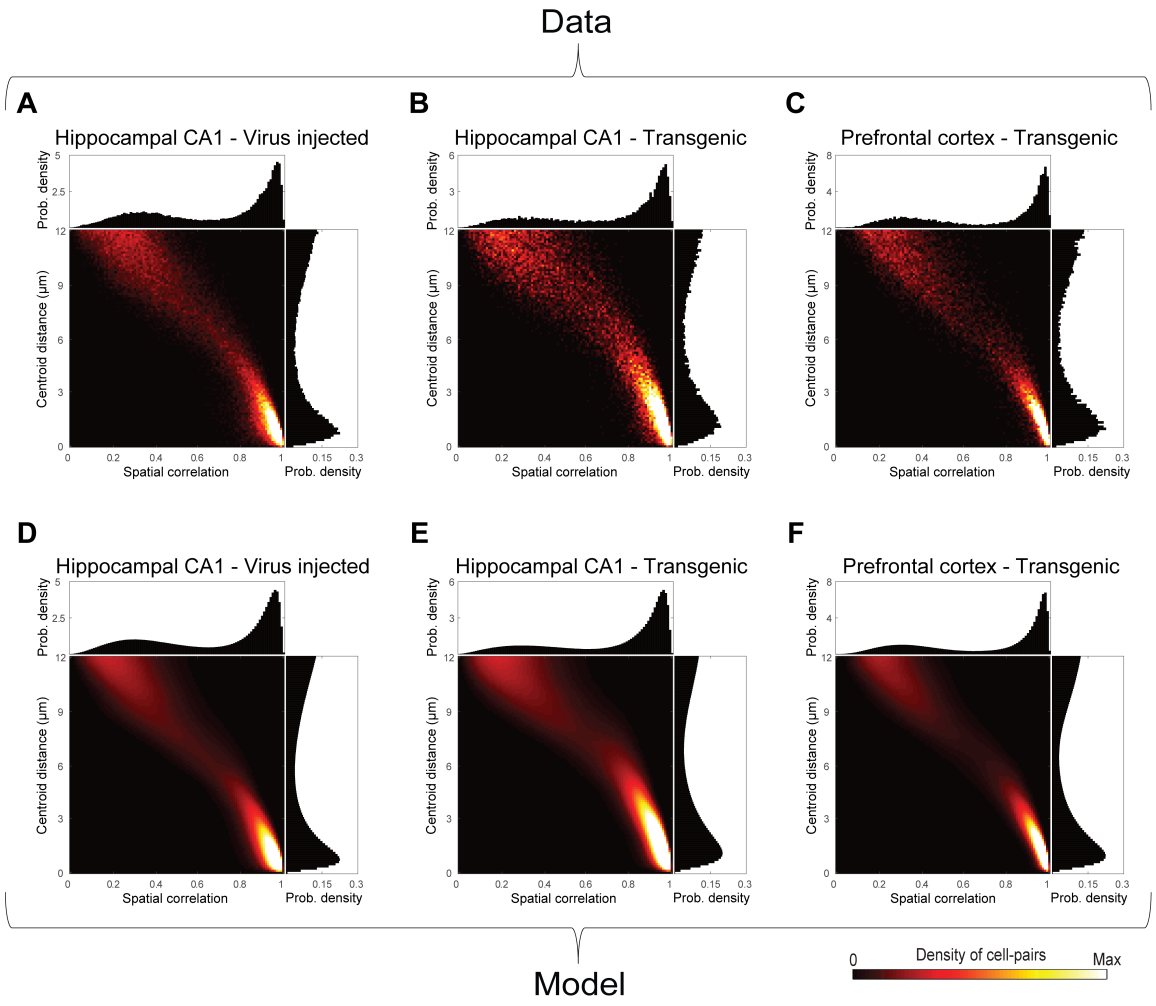

Figure S5. Examples of the joint distribution of spatial correlations and centroid distances modeled for each mouse individually. Related to Figure 2.

Joint distribution of spatial correlations and centroid distances and their corresponding marginal distributions for neighboring cell-pairs as observed in the data of three different mice (A-C), and the weighted sum of the models for same cells and different cells (D-F). The color scale was set to reach 0.25 of the maximal value to enable visualization of both subpopulations. Data and models in A and D are for 16 sessions recorded on 8 different days in the hippocampal CA1 of a GCaMP6f virus injected mouse while freely exploring the same environments. Data and models in B and E are for seven sessions recorded on seven different days in the hippocampal CA1 of a Thy1-GCaMP6f transgenic mouse while freely exploring the same environment. Data and models in C and F are for six sessions recorded on six different days in the prefrontal cortex of a CaMKII-GCaMP6s transgenic mouse while freely exploring the same environments.

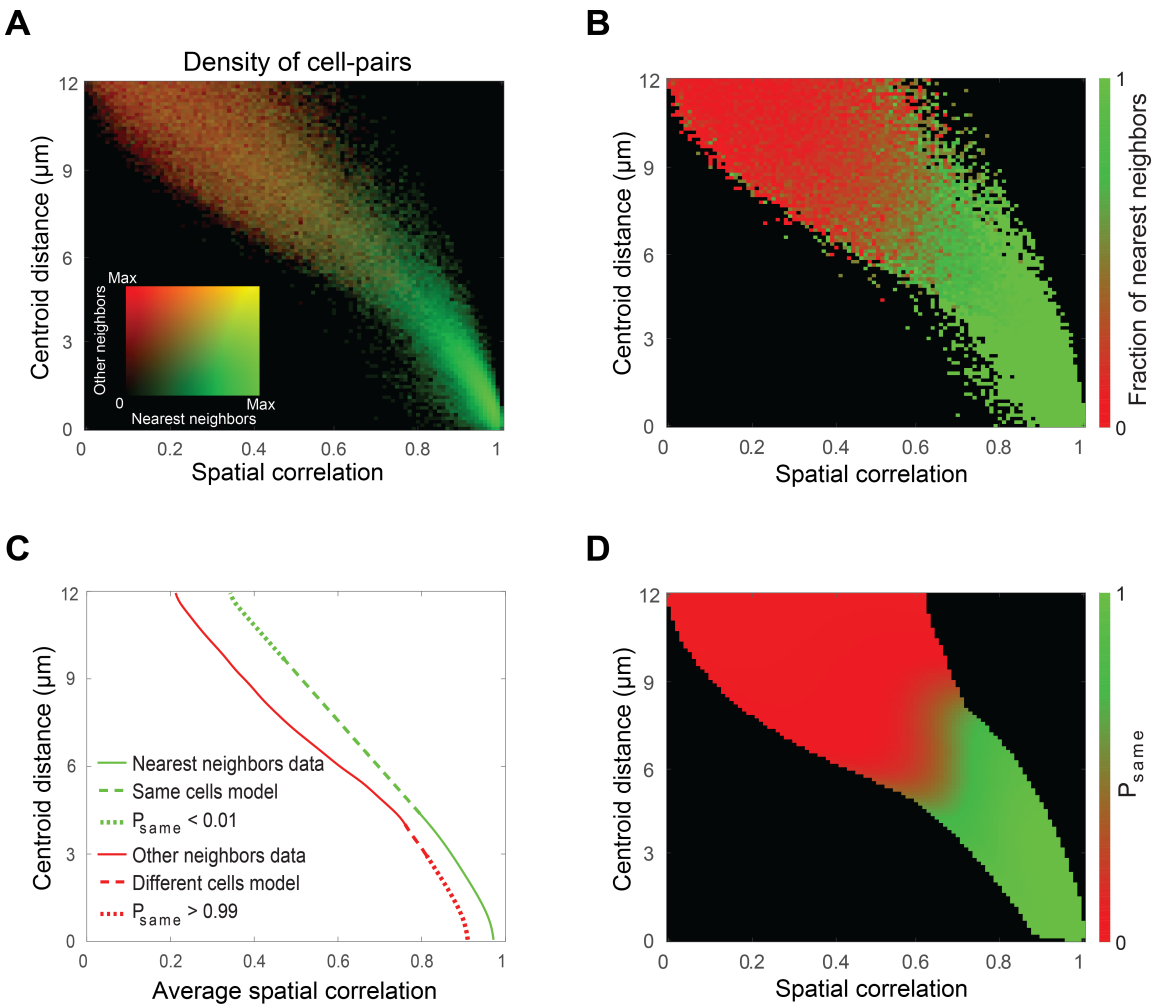

Figure S6. Obtaining the joint model for the joint distribution of spatial correlations and centroid distances based on nearest neighbors and other neighbors. Related to Figure 2.
(A) An RG overlay of the density of nearest neighbor (green) and other neighbor (red) cell-pairs given the centroid distance and spatial correlation. (B) Ratio of the density of nearest neighbor to all neighbor cell-pairs given the centroid distance and spatial correlation. Note the high fraction of nearest neighbor cell-pairs for high spatial correlations and low centroid distances. (C) The average spatial correlation as a function of the centroid distance for nearest neighbors (solid green curve), and other neighbors (solid red curve), and extrapolations of the data (dashed and dotted green and red curves), used to obtain the models for same cells and different cells, respectively. Dotted green curve represents centroid distances for which $P_{same}<0.01$, a region for which the extrapolation for the same cells model is insignificant, while the dotted red curve represents centroid distances for which $P_{same}>0.99$, a region for which the extrapolation for the different cells model is insignificant. (D) The probability for two cells from two different sessions to be the same cell (P_{same}) given the centroid distance and spatial correlation, as estimated by the model. P_{same} is presented exclusively for centroid distances and spatial correlations where the model's probability for cell-pairs of either type (same cell or different cells) is $>10^{-5}$. Data and model in all panels are for 16 sessions recorded on eight different days from GCaMP6f-expressing CA1 neurons of a mouse while freely exploring the same environments.

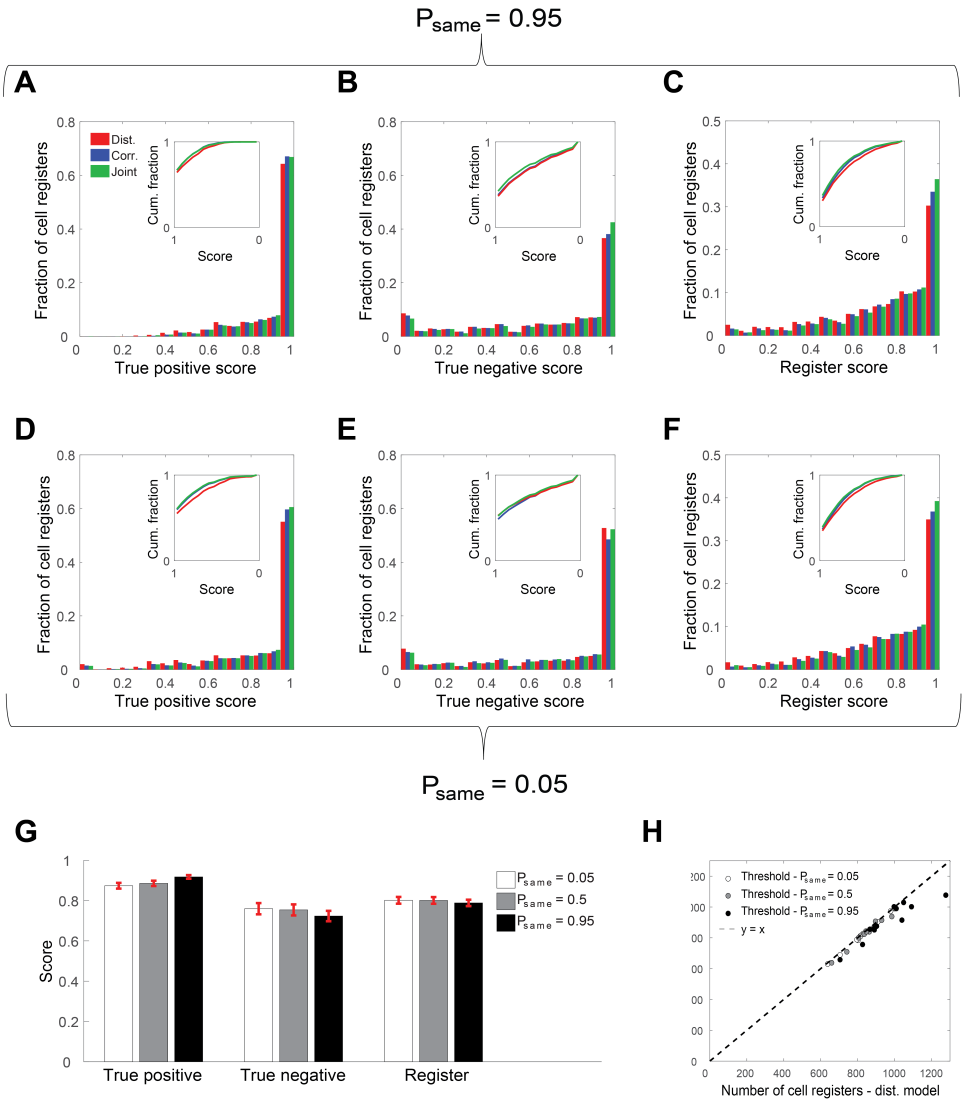

Figure S7. False positive and false negative rates can be controlled by the registration threshold. Related to Figure 5.

True positive (A and D), true negative (B and E), and register (C and F) scores for cell registers with thresholds of P_{same}=0.95 (A-C) and P_{same}=0.05 (D-F), for the centroids distances (red), spatial correlations (blue), and joint (green) models. Insets, cumulative fraction of cell registers as a function of the score reversed from 1 to 0. (G) Average true positive, true negative and register scores for the joint model with thresholds of P_{same}=0.05 (white bars), P_{same}=0.5 (gray bars), and P_{same}=0.95 (black bars). Error bars show SEM. The average true positive score increases while the average true negative score decreases with the registration threshold. Note, that using a more relaxed registration threshold (P_{same}=0.05) results in a slightly higher average register score than a stringent threshold (P_{same}=0.95). This result is partially explained by the fact that if more than one candidate exceeds the registration threshold, only the one with highest P_{same} will be registered as the same cell, thus eliminating a large fraction of false positives. (H) Final number of cell registers obtained by centroid distances based registration versus spatial correlations based registration. Dots represent the thresholds P_{same}=0.05 (white circles), P_{same}=0.5 (gray circles), and P_{same}=0.95 (black circles) for each mouse. Black dashed line represents the y=x curve. Note that the spatial correlations and centroid distances models obtain similar numbers of cell registers and that increasing the registration threshold increases the final number of cell registers. Data in all panels are pooled from N=12 mice.

Supplemental Experimental Procedures

Animals and surgical procedures

All procedures were approved by the Weizmann's Institute Institutional Animal Care and Use Committee. We used data from a total of 12 male mice. Mice were housed in cages with running wheels in a reverse light cycle facility, and were 8-12 weeks old at the beginning of the study. For Ca^{2+} imaging in the hippocampal CA1, we used five Thy1-GCaMP6f mice (Jackson Laboratory, #025393) (Dana *et al.*, 2014), and five AAV injected C57BL/6 wild type mice (data from four out of five wild type mice were taken from our previous work (Rubin et al., 2015)). For Ca^{2+} imaging in the prefrontal cortex, we used two CaMKII-tTA and rtTA-GCaMP6s double transgenic mice (CamKII-GCaMP6s, for short) (Tg(tetO-GCaMP6s)2Niell; Jackson Laboratory, #024742; B6.Cg-Tg(CaMK2a-tTA)1Mmay/DboJ; Jackson Laboratory, #007004; all mice were bred on C57BL/6 background). C57BL/6 wild type and Thy1-GCaMP6f mice were housed with 1-4 cage-mates. CamKII-GCaMP6s mice were single housed. All surgical procedures were conducted under isoflurane anesthesia (1.5-2% volume). We used a viral vector to express GCaMP6s or GCaMP6f in the CA1 of wild type mice. These mice underwent two surgical procedures. In the first, we injected into the CA1 400 nL of the viral vector AAV2/5-CaMKIIa-GCaMP6s or AAV2/5-CaMKIIa-GCaMP6f (Chen et al., 2013) (~2 X 10^{13} particles per ml, packed by University of North Carolina Vector Core). Stereotactic coordinates were: -1.9 mm anterio-posterior, -1.4 mm mediolateral, -1.6 mm dorsoventral from bregma. Mice were allowed to recover in their home-cages for at least one week before the next surgical procedure, in which we implanted a glass guide tube directly above the CA1, as previously described (Ziv et al., 2013; Rubin et al., 2015). This procedure was similar for both virus injected and Thy1-GCaMP6f transgenic mice. CamKII-GCaMP6s mice were implanted with a micro-prism lens (800μm diameter) in the prefrontal cortex. Stereotactic coordinates of the implantation were: 1 mm anterior-posterior, 0 mm mediolateral, -1.8 mm dorsoventral from bregma.

Ca^{2+} imaging and behavioral setup

Preparatory process

For time-lapse imaging in freely behaving mice using an integrated miniature fluorescence microscope (nVistaHD, Inscopix), we followed a previously established protocol (Ziv et al., 2013; Rubin et al., 2015). Briefly, at least three weeks after the guide tube or micro-prism lens implantation, we examined Ca^{2+} indicator expression and tissue health by imaging mice under isoflurane anesthesia using a two-photon microscope (Ultima IV, Bruker, Germany), equipped with a tunable Ti: Sapphire laser (Insight, Spectra Physics, Santa Clara, CA). At

this stage, we inserted into the guide tube of CA1 implanted mice a 'microendoscope' consisting of a single gradient refractive index lens (0.44 pitch length, 0.47 NA, GRINtech GmbH, Germany). We selected for further imaging only those mice that exhibited homogenous GCaMP6 expression and healthy appearance of the tissue. For the selected CA1 implanted mice, we affixed the microendoscope within the guide tube using ultraviolet-curing adhesive (Norland, NOA81, Edmund Optics, Barrington, NJ). Next, we attached the microscope's base plate to the dental acrylic cap using light cured acrylic (Flow-It ALC, Pentron, Orange, CA). We used a similar procedure to attach the base plate to the skull of the mice that were implanted with a micro-prism lens. All mice were returned to their home cages for a few days following the aforementioned procedure.

Ca^{2+} imaging in freely behaving mice

We used data from three experiments. In the first, we trained the five wild type mice to run back and forth on two elevated 96cm straight or 'L' shaped linear tracks (Rubin et al., 2015). Before beginning with Ca^{2+} imaging, we trained the mice for 8–11 days, until the mice ran at least 60 times the entire length of each track in two consecutive days. We imaged the mice every other day for 15 days, making for eight recording days. Each day of the experiment consisted of two 15 minutes long sessions, separated by 4–5 hours. In the second experiment, we imaged the five Thy1-GCaMP6f mice while they were freely exploring a spatial environment (a total of seven 20-55 minutes long sessions, separated by 1-28 days). In the third experiment, we imaged the prefrontal cortex of the two CamKII-GCaMP6s mice on six consecutive days, while the mice were freely exploring two different environments (3.5 minutes per day in each environment). Ca^{2+} imaging of the virus injected and transgenic mice was performed at 20Hz and 10Hz, respectively. To record mouse behavior, we used an overhead camera (DFK 33G445, The Imaging Source, Germany), which we synchronized with the integrated microscope.

Two-photon imaging data

We obtained published two-photon imaging data from the Allen Brain Observatory, 2016 (http://observatory.brain-map.org/visualcoding). We used data from ten different experiments in head fixed mice (experiment numbers: 511510945; 511510779; 511510699; 511510670; 511510911; 511510718; 511510870; 511510836; 511856567; 511498742). The experiments measured visual responses from GCaMP6f expressing neurons in different cortical areas and layers. Each experiment consisted of three recording sessions from the same mouse, separated by at least one day.

Processing of Ca^{2+} imaging data

We processed imaging data using commercial software (Mosaic, version 1.1.1b, Inscopix) and custom MATLAB routines as previously described (Ziv et al., 2013; Rubin et al., 2015). To increase computation speed, we spatially down-sampled the data by a factor of two in each dimension (final pixel size of 2.3 X 2.3μm). To correct for non-uniform illumination both in space and time, we normalized the images by dividing each pixel by the corresponding value of that pixel in a smoothed version. The smoothed version was obtained by applying a Gaussian filter with a radius of 100μm on the movies. Normalization also enhanced the appearance of the blood vessels, which were later used as stationary fiducial markers for image registration. We used a rigid-body registration to correct for lateral displacements of the brain. This procedure was performed on a high contrast subregion of the normalized movies for which the blood vessels were most prominent. The movies were then transformed into relative changes in fluorescence, $\Delta F(t)/F_0 = (F(t) - F_0)/F_0$, where F_0 is the value for each pixel averaged over time. For the purpose of cell detection, the movies were down-sampled in time by a factor of five or two for data recorded at 20Hz and 10Hz, respectively. We detected *spatial footprints* (i.e., weighted regions of interest consisting of each pixel's contribution to the cell's fluorescence) corresponding to individual cells using an established cell detection algorithm that applies principal and independent component analyses (PCA and ICA; (Mukamel, Nimmerjahn and Schnitzer, 2009). For each spatial footprint, we used a threshold of 50% of the footprint's maximum intensity, and each pixel that did not cross the threshold was set to zero. Once the cells were detected, further cell sorting was performed to find the spatial footprints that follow a typical cellular structure. This was done by measuring the footprints' area and circularity and discarding those whose radius was smaller than 5μm or larger than 15μm, or which had a circularity smaller than 0.8. In some cases, the output of the PCA-ICA algorithm included more than one component that corresponded to a single cell. To eliminate such occurrences, we examined all pairs of cells with centroid distances <18μm, and whenever their traces had a correlation >0.9, the cell with the lower average event peak amplitude was discarded. As an independent technique for cell detection, we also used CNMF-E (Zhou et al., 2016), an extension of the constrained nonnegative matrix factorization method (Pnevmatikakis et al., 2016) for one-photon microendoscopic data.

Detection of Ca^{2+} events

Ca^{2+} activity was extracted by applying the thresholded spatial footprints to the full temporal resolution (20Hz or 10Hz) $\Delta F(t)/F_0$ movies. Baseline fluctuations were removed by subtracting the median trace (20sec sliding window). The Ca^{2+} traces were smoothed with a

low-pass filter with a cutoff frequency of 2Hz. Ca^{2+} candidate events were detected whenever the amplitude crossed a threshold of 4 or 5 median absolute deviations (MAD), for GCaMP6s or GCaMP6f, respectively. We considered for further analysis only candidate Ca^{2+} events with decay time equal or longer than 600msec or 200msec for GCaMP6s or GCaMP6f, respectively, consistent with typically observed indicator decay times (Chen et al., 2013). To avoid the detection of several peaks for a single Ca^{2+} event, only peaks that were 4 or 5 MAD higher than the previous peak (within the same candidate event) and 2 or 2.5 MAD higher than the next peak for GCaMP6s or GCaMP6f, respectively, were regarded as true events. We set the Ca^{2+} event occurrence to the time of the peak fluorescence. To mitigate the effects of crosstalk (i.e., spillover of Ca^{2+} fluorescence from neighboring cells), we adopted a conservative approach, allowing only one cell from a group of neighbors (<18μm apart) to register a Ca^{2+} event in a 200msec time window (the event with highest peak $\Delta F(t)/F_0$ value). If two neighboring cells had a correlation >0.9 in their events, the cell with the lower average peak amplitude was discarded. Once the events were identified, further event sorting was performed to find the cells with sufficient signal-to-noise ratios. This was accomplished by measuring the event rate and the average event peak amplitude for each cell and discarding those whose event rate was smaller than 0.01Hz or which had an average event amplitude smaller than 1% $\Delta F(t)/F_0$.

Registration of cells across sessions

We developed a probabilistic method for registering cells across multiple sessions based on the similarity in the spatial footprint of their cellular activity. We used two similarity measures, the distance between centers of mass of two spatial footprints (*centroid distance*) and the Pearson correlation between spatial footprints (*spatial correlation*), to decide whether or not pairs of neighboring cells from two different sessions are the same entity. Note, that the spatial correlation is not merely a measure of shape similarity (as in spatial cross-correlation), but also accounts for differences in the cells' locations within the FOV. We registered cells using three main stages: (1) image alignment: rigid-body transformation was applied to all sessions according to a single reference session to correct for translation and rotation differences between sessions; (2) the distribution of spatial footprint similarities between neighboring cells from different sessions was calculated and then modeled as a weighted sum of the distributions of two subpopulations representing the same cells and different cells. This allowed us to estimate the probability for any pair of neighboring cells from two different sessions to be the same cell (P_{same}) given their spatial correlation and

centroid distance; (3) an iterative clustering procedure (Bansal, Blum and Chawla, 2004) that registers cells based on the P_{same} obtained from the probabilistic model.

(1) Image alignment

For each session, we projected the centroid locations of all the cells' spatial footprints onto a single image. We computed the cross-correlations between the projections of each session with the projections of a reference session, examining different rotations. We corrected for translation and rotation differences between the sessions by applying the translations and rotations that resulted in the maximal cross-correlation, yielding each cell's location in the reference coordinate system. We also corrected for sub-pixel translational differences between the sessions by applying a Gaussian fit to the cross-correlation in each dimension. Sub-pixel translational correction decreased the centroid distances and increased the spatial correlations between the nearest neighbors (Figure S3).

(2) Probabilistic modeling of the data

We calculated the distributions of centroid distances and spatial correlations for neighboring cell-pairs across sessions. We defined neighboring cell-pairs as pairs with centroid distances <12μm across sessions, and assumed that cell-pairs with larger centroid distances are different cells. To obtain reliable registration we sought to model the distributions of centroid distances (centroid distances model) and spatial correlations (spatial correlations model) for neighboring cell-pairs, as a weighted sum of the distributions of two subpopulations, representing same cells and different cells. The specific choice of the functional forms we used to model the distributions of same cells and different cells were guided by the distributions obtained from the data for nearest neighbors and other neighbors.

Because there is a certain minimal distance between different cells in the FOV, we approximated the density of cells given their distance from any given cell as a sigmoid function. Because the area of a ring surrounding each cell linearly increases with the distance, we modeled the distribution of centroid distances for different cells as a multiplication of a sigmoid function by a linear function. To model the distribution of centroid distances for the same cells, we used a lognormal distribution, to match the requirement of a non-negative distribution that resembles the distribution for nearest neighbors in the data. Similarly, the distribution of spatial correlations for same cells was modeled as a lognormal distribution. Since the correlation cannot exceed 1, and the distribution of spatial correlations for nearest neighbors peaks close to 1 and decays towards 0, we defined the origin as 1 by taking 1 minus the spatial correlation value. Because lognormal distribution is >0 for any positive value, whereas correlation is bounded, we multiplied the obtained model for same

cells by a sigmoid function. The distribution of spatial correlations for other neighbors is single peaked and ranges between 0 and 1. We therefore modeled the distribution of spatial correlations for different cells using a beta function. For each model, we found the parameters of the distributions and weights that best fit the data in terms of the mean squared error (performed for each mouse, individually).

Applying similar principles, we modeled the joint distribution of centroid distances and spatial correlations (joint model) for neighboring cell-pairs as a weighted sum of the distributions of two subpopulations. We used the previously estimated distributions of centroid distances and their weights as the marginal distribution of the joint model. We then estimated the distribution of spatial correlations for any given centroid distance (Figure S6). We calculated the average spatial correlation for nearest neighbors and other neighbors given the centroid distance to approximate the spatial correlation for same cells and different cells, respectively. We used the average spatial correlation measured for nearest neighbors at small centroid distances (mostly the same cells) to linearly extrapolate the average spatial correlation for large centroid distances where we could not be certain that the nearest neighbors were indeed the same cells. Assuming that other neighbors correspond to different cells, we used the average spatial correlation measured for other neighbors at high centroid distances, and extrapolated the data to small centroid distances for which data was sparse. For a given centroid distance, the nearest neighbors showed higher spatial correlations than other neighbors. We therefore measured the difference between the average spatial correlation for nearest neighbors and other neighbors per centroid distance and extrapolated the mean spatial correlation for different cells assuming a constant difference. In practice, this extrapolation had little impact on the joint model since it was calculated mostly for a centroid distances range in which the proportion of different cells was < 0.01 (Figure S6C – note the narrow range of the dashed red curve versus the wide range of dotted red curve). For simplicity, the variance of the spatial correlations at a given centroid distance was assumed to be the same for both subpopulations and was estimated based on the standard deviation of all neighboring cell-pairs and the estimated average spatial correlations of each subpopulation. To obtain the joint model, for each mouse individually, we modeled for any given centroid distance, the distribution of spatial correlations for the same cells as a lognormal distribution and for different cells as a beta distribution.

To obtain the probability for any pair of neighboring cells from different sessions, given their spatial correlation and centroid distance, we calculated $P_{same}(dist,corr)$, according to Bayes' rule:

$$P_{same}\left(dist,corr\right)=\frac{P\left(dist,corr\mid same\right)\ W_{same}}{p\left(dist,corr\mid same\right)\ W_{same}+p\left(dist,corr\mid diff\right)\ W_{diff}}$$

where W_{same} and W_{diff} are the estimated weights (prior probabilities) for the two subpopulations of same cells and different cells, respectively. P(dist,corr|same) and P(dist,corr|diff) are the estimated conditional joint distributions of centroid distances and spatial correlations for same cells and different cells, respectively. Similar calculations were performed for each of the one-dimensional models (spatial correlation and centroid distance). This allowed us to obtain P_{same} for any pair of cells across sessions in the data.

(3) Cell registration with a clustering procedure

Following the observation that cells are spatially clustered into clusters representing the same cells across imaging sessions, we applied an iterative clustering procedure aimed at maximizing within-cluster P_{same} and minimizing between-cluster P_{same}. To obtain a clustering of cells to serve as initial conditions for the iterative clustering procedure, we registered cells according to the spatial correlations model with a registration threshold of P_{same}=0.5. We created an initial list of cells comprising of the cells detected in the first session. Each session was then inspected and each cell that had a pairing candidate on the list, with which it had a P_{same}>0.5, was registered to that cell. Cells that did not have such a pairing candidate were added to the list as new cells. In cases where more than one candidate crossed the threshold, the cell-pair with the highest P_{same} was registered to be the same cell. The iterative clustering procedure was then applied. In each of the iterations, all cells were inspected and each cell either: (1) remained in the same cluster; (2) was transferred to a more similar cluster; or (3) initiated a new cluster. The decision was made by finding for each cell the candidate from all the different sessions with which it had the highest P_{same}. If the P_{same} was lower than the registration threshold, the cell formed a new cluster. If P_{same} was higher than the registration threshold, the cell joined the cluster of that candidate (if they were already in the same cluster, then nothing changed). If there was already a cell from that session with a lower P_{same} in the cluster, the new cell replaced the old one in the cluster, but if the existing cell had a higher P_{same}, then nothing changed. The iterative process continued until the clustering procedure converged and no more clustering changes occurred.

Quantification of registration accuracy

Cell-pairs

By applying P_{same} to our data, we obtained the certainty of registration associated with any pair of neighboring cells across sessions. We defined uncertain registrations of cell-pairs as

those with a probability ≤0.95 to be the same cell and ≤0.95 to be different cells ($0.05 \leq P_{same} \leq 0.95$), and calculated the fraction of uncertain cell-pairs out of all neighboring cell-pairs. Additionally, we used the estimated probability distributions of same cells and different cells to calculate a receiver operating characteristic (ROC) curve obtained from 1,000 evenly spaced registration thresholds of P_{same} between 0 and 1. For the centroid distances model, the false negative rate was estimated as the fraction of the model of same cells that is higher than the registration threshold:

$$False\,negative\,rate = \frac{\#\,false\,negatives}{\#\,false\,negatives + \#\,true\,positives}$$

The false positive rate was estimated as the fraction of the model of different cells that is lower than the registration threshold:

$$False\,positive\,rate = \frac{\#\,false\,positives}{\#\,false\,positives + \#\,true\,negatives}$$

Similar calculations were performed for the spatial correlations and joint models. We defined the overall estimated registration accuracy:

$$accuracy = \frac{\#\,true\,negatives + \#\,true\,positives}{\#\,false\,negatives + \#\,true\,positives + \#\,false\,positives + \#\,true\,negatives}$$

Additionally, we calculated the Gini coefficient G_1 to quantify the separability between the distributions of two subpopulations as $G_1 = 2AUC \quad 1$, where AUC is the area under the ROC curve.

Cell registers

To evaluate the performance of our registration method for multiple sessions, we defined a stringent register score index for each registered cell, which takes into account false positives, false negatives, and non-exclusive cell registrations from all the pairwise combinations of registered sessions across all sessions. For each session, a cell was either active or inactive according to the cell registration, resulting in pairs of sessions where the cell was active in both (active-active), active in only one (active-inactive), or active in none (inactive-inactive — no evaluation is required in these cases). For each cell register in the data, a true positive score was calculated as the total number of reliable active-active cell-pairs ($P_{same}>0.95$) out of the total number of active-active pairwise combinations of all sessions. A true negative score was calculated as the total number of reliable active-inactive pairs ($P_{same}<0.05$ for all candidates) out of the total number of active-inactive pairwise combinations of all sessions. An exclusivity score was calculated as the total number of exclusive active-active cell-pairs (for all additional candidates $P_{same}<0.05$) out of the total

number of active-active pairwise combinations of all sessions. The register score was calculated as the total number of reliable pairs out of the total number of active-active and active-inactive pairwise combinations of all sessions, where each pair was considered reliable only if it met all three aforementioned criteria. Accordingly, we defined the register score as:

$$register\ score = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{\substack{m=1 \\ m \neq k}}^{N} \delta(k,m)$$

where N is the total number of sessions, N is the number of active sessions for a given cell register, and $\delta(k,m)$ is 1 if a cell-pair in the cluster is reliable, and 0 otherwise.

Scalability for a large number of sessions

To examine if registration accuracy changes with the number of registered sessions, we applied our registration method to data recorded in 16 sessions on eight different days spanning two weeks (N=5 mice). Registration was performed simultaneously for multiple sessions with the clustering procedure, choosing random subsets of the 16 sessions comprising of 4, 8, or 12 sessions, and for the entire set of 16 sessions. The average register score was measured as a function of the number of registered sessions.

Validation of cell registration

Exclusivity and transitivity measures

To validate our probabilistic approach, we defined *exclusivity* and *transitivity* measures. Exclusivity requires that only one cell from session 1 be paired with a given cell from session 2. Transitivity requires that if a cell from session 1 and a cell from session 2 are paired and the cell from session 2 is paired with a cell from session 3, then the cells from sessions 1 and 3 be also paired. To measure exclusivity, we calculated the distribution of P_{same} with all additional candidates only for cells that have a pairing candidate with a $P_{same}>0.5$ in another session. Cells with additional pairing candidates with a $P_{same}>0.5$ were considered non-exclusive. To measure transitivity, we calculated the distribution of P_{same} for pairs of cells from two different sessions where each had a $P_{same}>0.5$ with the same candidate from a third session. Of such cell-pairs, those with a $P_{same}<0.5$ were considered non-transitive. To compare exclusivity and transitivity of the data with those expected by chance, we obtained shuffled data by measuring the centroid distances between cell-pairs from different sessions, where each session was taken from a different mouse. We then performed the same analyses on the shuffled data.

Place field stability

We also validated our probabilistic approach by applying our registration method to data recorded from the CA1 of the hippocampus of N=5 mice while freely exploring linear tracks (Rubin et al., 2015). We analyzed mouse behavior videos using a custom MATLAB (Mathworks) routine that detected the mouse's center of mass in each frame, calculated its velocity and applied a rectangular smoothing window of 250msec. For place field analysis, we considered periods wherein the mouse ran >3cm/sec. We divided each track into 24 bins (4cm each) and computed the time spent in each bin and the number of Ca^{2+} events per bin, and smoothed these two maps ('occupancy' and 'Ca^{2+} event number') using a truncated Gaussian kernel (σ=1.5 bins, size=5 bins) (Ziv et al., 2013; Rubin et al., 2015). We then computed the activity map (event rate per bin) for each neuron by dividing the smoothed map of Ca^{2+} event numbers by the smoothed map of occupancy. We separately considered place fields for the two running directions on the linear track. We defined each of the place field's position to be the peak value of the activity map. For each place field with ≥5 events for a given session, we computed the spatial information (in bits per event) using the unsmoothed events-rate map of each cell, as previously described (Markus et al., 1994):

$$Spatial\ information = \sum_i p_i \left(r_i / \bar{r}\right) log_2 \left(r_i / \bar{r}\right)$$

where p_i is the probability of the mouse to be in the i^{th} bin (time spent in i^{th} bin/total session time); r_i is the Ca^{2+} event rate in the i^{th} bin; $\bar{r}$ is the overall mean Ca^{2+} event rate; and i is running over all the bins. We then performed 1,000 distinct shuffles of animal locations during Ca^{2+} events, accounting for the spatial coverage statistics for the relevant session and direction, and calculated the spatial information for each shuffle. This yielded the p value of the measured information relative to the shuffles. Cells with spatial information higher than that of 95% of their shuffles were considered significant place cells. We then calculated the place field positional shift and place field correlation for cell-pairs from different sessions for which at least one cell in the pair was considered significant. The place field positional shift was calculated as the difference in place field peak positions, and the place field correlation was calculated as the Pearson correlation between the place field activity maps. The place field maps used for these two measures included both running directions on the running track. To compare the place field stability of the data with the stability expected by chance, the identity of cells were randomized (shuffled data) across sessions. We then performed the same analyses on the shuffled data.

Simulated data

Simulated data was obtained by using representative spatial footprints of cells that were detected in our data. We then placed each representative spatial footprint in a random location in the FOV to serve as the original centroid location of the cell. We added a constraint that did not allow for two different cells to have very close centroid locations. The minimal centroid distance between the original locations of neighboring cells (different cells) was set to 7±1μm (normally distributed) to match the distribution of the observed data. In each session, a fraction of the cells were simulated as active and their locations were set as the original centroid locations with a jitter. The radius of the jitter for each cell in each session was drawn from a lognormal distribution and the angle was uniformly distributed. The parameters for the active fraction of cells and for the jitter were chosen to match those of the observed data. To obtain simulated data with different noise levels, we systematically increased the radius of the jitter by increasing the mean of the lognormal distribution. We then applied our registration procedure to the simulated data. This allowed us to estimate the false positive and false negative rates from the simulated data model and compare them with the actual false positive and false negative rates obtained when applying either our cell registration method or a centroid distance threshold to the simulated data. Finally, we simulated the effects of registration errors on measurements of the stability of the functional properties of cells. We defined different cells (shuffled data) as having a certain probability (10±1%) to share the same coding properties across sessions. To test whether registration errors could obscure conclusions drawn from data with different population effect sizes, we systematically changed the percentages of same cells that maintain their coding properties across sessions (10-15±1%). Effect size was defined as the difference between the mean stability of the population and shuffled data, divided by the stability standard deviation.

Supplemental Information References

Bansal, N., Blum, A. and Chawla, S. (2004) Correlation Clustering, *Mach. Learn.*, 56, pp. 89–113.

Chen, T. W. *et al.* (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity, *Nature*, 499, pp. 295–300.

Dana, H. *et al.* (2014) Thy1-GCaMP6 Transgenic Mice for Neuronal Population Imaging In Vivo, *PLoS One*, 9, p. e108697.

Markus, E. J. *et al.* (1994) Spatial information content and reliability of hippocampal CA1 neurons: Effects of visual input, *Hippocampus*, 4, pp. 410–421.

Mukamel, E. A., Nimmerjahn, A. and Schnitzer, M. J. (2009) Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data, *Neuron*, 63, pp. 747–760.

Pnevmatikakis, E. A. *et al.* (2016) Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data, *Neuron*, 89, p. 299.

Rubin, A. *et al.* (2015) Hippocampal ensemble dynamics timestamp events in long-term memory, *Elife*, 4, p. e12247.

Zhou, P. *et al.* (2016) Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data, *arXiv*, p. e1605.07266.

Ziv, Y. *et al.* (2013) Long-term dynamics of CA1 hippocampal place codes., *Nat. Neurosci.*, 16, pp. 264–6.