Supplementary Materials for

Persistence of neuronal representations through time and damage in the hippocampus

Walter G. Gonzalez, Hanwen Zhang, Anna Harutyunyan, Carlos Lois*

*Corresponding author. Email: clois@caltech.edu

Published 23 August 2019, *Science* **365**, 821 (2019)
DOI: 10.1126/science.aav9199

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/content/365/6455/821/suppl/DC1)

Movies S1 to S9 (.mp4)

Materials and Methods

Materials and Methods

Animals.

Male and female C57BL6J-Tg-Thy1-GCaMP6s 6 to 20-week old (Jackson Labs stock: 025776) were housed in a reverse 12 h light/dark photocycle and provided food ad libitum. Mice were single housed post-surgery until the end of the experiment. Experimental animals were selected randomly and included both sexes. All animal procedures were approved and performed following institutional guidelines (Caltech IACUC).

Unilateral/Bilateral endoscope implantation

Mice were anesthetized with a single dose of 100/10 mg/kg ketamine/xylazine before the surgery and placed into a stereotactic frame. The body temperature was maintained with a passive heating pad at 37 ºC. Ketoprofen 5 mg/kg and buprenorphine SR 1 mg/kg was subcutaneously injected prior to surgery. Bupivacaine 1 mg/kg solution was added dropwise along the surgical incision prior to wound closure and animals were maintained on ibuprofen 30 mg/mL (in the water) ad libitum for at least 3 days post-surgery. Animals were in a recovery period for at least 4 weeks before attachment of the microendoscope.

Mice underwent unilateral or bilateral surgeries to implant GRIN lenses (1.8 mm 0.25 pitch 0.55 NA) directly dorsal to CA1. Before implantation, we performed a 1.8 mm diameter craniotomy centered around the coordinates (relative to bregma: 1.8 mm and -1.8 mm lateral; -2.0 mm posterior) using a FG1/4 carbide bur. Freshly prepared artificial cerebrospinal fluid (aCSF) was applied to the exposed tissue throughout the surgery to prevent dehydration. Using a blunt 26-gauge needle, the dura, cortex, and portion of the corpus callosum were quickly aspirated under continuous perfusion with aCSF. Aspiration was stopped once a thin layer of horizontal fibers was left on the surface of the hippocampus. The cortical cavity was perfused with more aCSF and small pieces of moist gelfoam were placed on the surface of the craniotomy to prevent excessive bleeding while avoiding contact with the surface of hippocampus. Once the surface of the hippocampus was clear of blood, the GRIN lens was slowly lowered in into the brain using a stereotaxic arm to a depth of 1.30 mm below the surface of the skull. Removal of the cortex and insertion of the GRIN lens was performed in 10 minutes or less (in each hemisphere) to prevent bulging of the hippocampus due to the decreased dorsal pressure. Two skull screws were placed anterior to bregma (1.8/-1.8 mm lateral; 1.0 mm anterior) and both the screws and lens were secured with cyanoacrylate glue and dental cement. The exposed end of the GRIN lens was protected with transparent Kwik-seal glue and animals were returned to a clean cage. Two weeks after the surgery, mice were anesthetized with 1.0 to 2.0 % isoflurane, the glue covering the GRIN lens was removed and a microendoscope was aligned with the GRIN lens. The miniature microscopes were connected to a portable computer for live view of the fluorescence image which was used to guide the final alignment and focal plane of the microscope and lens. The microscope was permanently attached to the implant with dental acrylic and the focal sliding mechanism on the microendoscope was sealed with superglue.

Local CA1 damage.

Damage was induced unilaterally by increasing the LED power of the microendoscope to the maximum allowable level. The power of the 470 nm blue LED on the miniscope was measured to produce 500-700 μW at this setting. The power measurements is done at the end of the GRIN lens facing CA1 using a power meter (Thorlabs PM100D, S155C probe). The brain was illuminated for at least 30 minutes during the foraging and linear track tasks. In one animal we performed local brain heating on two sessions to increase the damage area (top row of Fig. 3A). The extent of the lesion was assessed by the presence of abnormal burst of activity and by the presence of continuously green cells on the following day. Sessions with abnormal activity were selected based on the low place/time field correlation observed in Fig. S11 and by the increase number of co-active neurons, as shown in Fig. 3B. Three animals were exposed to high LED power and all three developed abnormal activity in the form of large number of co-active neurons. However, the mouse with the most extensive damage required over a month for the field of view to become clear enough to be registered. When this animal was exposed to the linear track it formed place/end cells but the similarity to the original representation was lower than the other two mice.

Mouse behavior.

Mice were maintained on a reverse photo cycle and three days prior to the start of behavior recording they were restricted to 2.0 mL of water per day. After three days of water restriction, mice were brought to the recording room and connected to a computer system through a commutator with a 2.0-meter-long custom cable (Mouser, 538-50MCX-37). Animals were habituated to handling and being tethered to the cable for at least 3 days. During this time, mice were connected to the computer and allowed to explore their home cage but not the linear track. After habituation, mice were placed in the middle of the track and imaging was started within ~20 seconds of the mice being introduced in the linear track. Mice were exposed to the maze every day during the learning period and first 5 days following re-exposure. However, during the training period or 5 days after re-exposure, recording sessions were at different day intervals ranging from 2-7 days. The initial facing orientation of the mouse in the linear track was not controlled.

Each behavior session consisted of a 10-minute recording of the mouse freely moving in its home cage, without the cage lid or food dispenser, immediately followed by a 20-minute recording of the mouse running on a linear track. The mouse home cage is rectangular 20 cm x 35 cm x 15 cm with transparent walls while the linear track is a 1.5 m x 12 cm x 15 cm made of white plastic. The linear track was cleaned with 70 % ethanol before introducing another mouse. Three group of cues were place on both walls of the linear track and consisted of black stripes (2 cm wide) at different angles (Fig. S1). The linear track was equipped with two automatic liquid dispensing ports at either end delivering between 10-50 μL of sugar water (15% sucrose in DI water). The system would require the animal to run to opposite ends of the track to receive water, a green LED (right side) and a red LED (left side) indicated which port the mouse needed to run towards. A beeping sound was played once the mouse activated the IR sensor. Delayed reward experiments were performed by decoupling the delivery of sugar reward from IR sensor activation, thus the animal was required to wait for 5 seconds. The beeping sound was not delayed. The position of the mouse was tracked by an ultra-wide-angle webcam

at 25 Hz and 640x360 pixels positioned about 1.8 meters above the maze. Mouse position was extracted using OptiMouse (*31*).

Custom miniaturized fluorescence microscopes.

Miniaturized microendoscope were custom made following previous designs (*19*). The main differences between the microendoscope used here and previous designs is the reduction in size achieved by reverse engineering the CMOS sensor used in (*32*) and by minimizing the wall thickness and footprint of the microendoscope body used in (*17*). The CMOS sensor used here is adapted from a miniature CCTV marketed online as model MC900 but can also be purchased under other names. As purchased, the video camera circuit consists of two 1 mm thick PCB boards: a primary sensor PCB containing an Omnivision 7960 CMOS sensor and some peripheral resistors and capacitors connected through a 6 pin connector to a secondary PCB with an 24.545 MHz clock and a voltage regulator circuit. These two circuits were carefully separated by desoldering the 6 pin connector. The clock and voltage regulator circuits can be connected to the sensor PCB through cables as long as 20 cm but electrostatic noise can affect the clock output on longer distances. Here we soldered a female 6 pin connector to the CMOS PCB and a male connector to the clock and voltage regulator connectors. This approach allowed us to decrease the weight of each microendoscope by as much as 200 mg during the period the animal is not being recorded.

The clock and voltage regulator PCB were connected to an USB analog video to digital converter. LED intensity was controlled through the ADC output of a signal generator. To decrease the footprint and weight we designed the body of the microendoscope with wall thickness less than 1 mm. The body of the microendoscope was machined from 1.25 x 2.5 x 2.5 cm^3 black acetal (Delrin) blocks using a 5-axis CNC machine (PocketNC). All CAD designs and CNC trajectories were generated by Fusion 360 (AutoCAD) and are provided in the supplementary information. Light from the excitation LED was filtered through a 470/40 nm bandpass emission filter, fluorescence light was separated from the excitation light by a 495 nm long pass dichroic mirror and filtered through a 520/50 nm bandpass filter (Chroma technologies). Light was collimated onto the CMOS sensor through a 12.5 mm achromatic lens (Edmund Optics). The FOV covered 600 µm x 479 µm at a resolution of 720 x 576 pixels, 0.83 µm/pixel (NBS1952 calibration target, Thorlabs).

Calcium imaging.

Video Acquisition: **signal** from the microendoscope CMOS sensor was acquired using a UVC compliant USB analog video to digital converter (EasyCAP DC60). The video feed was captured and saved by videoLAN media player (www.videolan.org) using custom MATLAB scripts. Data acquisition was set at 25 Hz and display resolution at 720 x 576 pixels using a YUV4:2:2 codec and AVI file encapsulation. All three cameras were synchronized to start simultaneously and were verified to have latencies smaller than 10 ms. Raw videos were offline transcoded to lossless H.264 -MPEG-4 AVC codec and MP4 encapsulation and the first 2 seconds of each video were deleted. Transcoded videos were filtered using a high quality 3-dimensional low pass filter (*hqdn3d*) with spatial and temporal smoothing of 4x4 pixels and 2 frames, respectively. Denoised videos were then 4x down-sampled by a moving window averaging of 4 frames. All transcoding,

smoothing, and down sampling was performed by the open source program *ffmpeg* (www.ffmpeg.org) controlled through custom MATLAB scripts. Down sampled videos were motion corrected using the recursive fast Fourier transform approach provided in the MATLAB script *sbxalign.mat* from Scanbox. For batch analysis all 4x down-sampled videos were concatenated into a larger video and then motion corrected.

Signal extraction: motion corrected videos were analyzed using CNMFe (MATLAB variable names shown in parenthesis). We further down sampled our 4x accelerated data with a 2-fold spatial (ssub) and temporal down-sampling (tsub) (*33*, *34*). The data was smoothed with a gaussian kernel of width between 6-8 pixels (gSig), and neurons were constrained to a diameter of 30 pixels (gSiz). A ring model (bg_model) was used for the background with radius of 30 (ring_radius). Neurons with spatial overlap greater than 0.65 (merge_thr) and centroid distance less than 5 pixels (dmin) were merged. Only ROIs with minimum peak-to-noise ratio of 5 (min_pnr) and minimum spike size of 3 (smin) were extracted. The foopsi deconvolution method by CNMFe was employed in all datasets. In some cases we selected more stringent parameters to account for different imaging quality across animals. To compare neuronal activity during home cage exploration and running in the linear track in each session, 7450 frames of calcium imaging during the 20-minute linear track task and 3750 frames during home cage exploration were analyzed simultaneously. Individual session analysis was used to compare neurons active during home cage exploration and linear track as well as for determining whether neurons were active every day using pixel intensity correlation. In addition, field stability across days was determined from datasets where up to 200,000 frames of linear track or home cage data was analyzed simultaneously with CNMFe. The resulting neuronal activity was also utilized to confirm registration procedures described below. Data analysis was performed using the neural activity output (*neuron.S*) from CNMFe.

Data analysis

Neuronal activity extraction: using CNMFe we extracted the background subtracted raw calcium activity of each neuron (neuron.C_raw), the deconvoluted calcium activity (neuron.C), and the neural activity (neuron.S). The footprint of a neuron was recovered by selecting the contour from the neuron.Coor matrix generated by CNMFe using a threshold of 0.6. This threshold selects the most intense pixels (60^{th} percentile) as the contour of the neuron from the complete contour extracted by CNMFe. The peak to noise ratio (PNR) was calculated by a built-in function from CNMFe or by a custom script, both approaches calculate the PNR by obtaining the maximum signal of the deconvoluted calcium transients (neuron.C) and dividing it by the standard deviation of the noise of each neuron. The noise level for each neuron was calculated by subtracting the deconvoluted calcium activity from the raw calcium activity (noise = neuron.C_raw – neuron.C). We noticed that not all regions of interest (ROIs) generated by CNMFe can be considered neurons with certainty, some may be segmented dendrites or background fluctuations (**Fig. S2b-c**). We removed these ROIs from the pool of registered ROIs by selecting only those with areas between 30 to 250 $pixel^2$ and inverse circularity less than 4.0. To determine these cut-off parameters, we plotted the overall distribution of PNR, Area, and circularity for each dataset and selected the values so that only the most linear part of the distribution would be included in the data. The same

parameters were used for all animals. The stability of a neuron could also be affected by the detection limit of our microendoscope, to avoid bias towards instability we only analyzed neurons whose PNR was larger than 8. Thus, in concatenated datasets only neurons satisfying these shape and signal constrain on at least one session were included in our analysis. In individual datasets, when comparing home cage exploration and track data, the neuron must pass such criteria at least in one of the environments in one session. Throughout the manuscript we refer to "firing rate" of a neuron. These values are not the number of action potentials per second. Instead, "firing rate" here refers to the binarized neural activity (neuron.S) so that any non-zero value was set to 1, summed over the complete recording and divided by the recording time (20 minute track or 10 minute home cage).

Identifying active cells: a cell was defined to be active if the maximum amplitude of the deconvoluted calcium activity (neuron.C) was larger than 3 standard deviation of the noise of that neuron (neuron.C_raw minus neuron.C) in that session. The standard deviation of the noise and maximum amplitude of the spike was calculated for each session individually. Neuronal activity below 3 standard deviation (approximately 20 ± 11 % of all spikes) were not used to determine if a neuron was active or not but were included in all other analyses. Alternatively, we have also generated motion corrected images and videos of the correlated pixel fluctuations to support the results that most ROIs have some level of activity on most days (**Fig. 1G, S4, and Movie 3, 4, 6**). Correlation images provide an alternate visual inspection of activity per session and are independent of the deconvolution ability of CNMFe or registration accuracy across days.

Identifying place cells: response fields of place cells were extracted by identifying periods when mice ran continuously faster than 3 cm/second for more than 0.4 seconds. Together, these thresholds eliminate periods during which mice were grooming, rearing, or turning. The length of the linear track was divided into bins spanning 3 centimeters (50 bins). The average firing rate of a neuron in each bin was calculated by the sum of all calcium activity in a bin divided by the amount of time the mouse spent in that bin. The average firing rate of a neuron was then normalized by the total number of spikes in order to generate normalized tuning profile of each neuron. Neurons were classified as place cells if: (1) the place field is at least 15 cm wide; (2) calcium transients were present > 30 % of individual lap traversals through the place field; and (3) the cell contains significantly greater spatial information than chance. Spatial information is calculated using (*35*) :

$$SI = \frac{1}{\lambda}\sum_i \lambda_i \log_2\left(\frac{\lambda_i}{\lambda}\right) P_i$$

where λ is the overall average calcium transient rate of the cell, λ_i is the average calcium transient rate in spatial bin *i* and P_i is the probability the mouse is in spatial bin *i*. Chance level spatial information for each neuron is calculated by shuffling the time stamps of the calcium transients and calculating the spatial information of the shuffled neuronal activity, this is done for 1000 iterations. The spatial information of the cell is considered significant if it is higher than 95% of the shuffled traces.

Identifying end cells: the linear track was equipped with a LED light that would turn off once the animal activated the IR sensors at the water reward port. The ON/OFF transition of the LED in the behavior video was extracted and used as a timestamp. The LED timestamp was set as time = 0 and a window of 32 frames or the time the animal was immobile (velocity less 1 cm/sec, whichever was smaller) was used for analysis. Neurons were classified as end cells if: (1) they fired at least 20 % of the times the animal activated the water port; (2) the neuron fired 30% more within its field than outside (within a field if defined as ± 10 % from the maximum of amplitude of the tuning curve); and (3) the cells contain significantly greater information than that encoded in a dataset where spikes where randomized. Information content was calculated using the same equation above but using λ as the average activity during immobility, λ_i is the average activity in frame *i* after activation of the water port. The variable P*i* in this case represent the probability that the animal was not moving during frame *i*.

Quantifying stability: we investigated the stability CA1 representation by analyzing neurons which were classified as place/end cells through several metrics including: (1) cell overlap, (2) fraction consecutively active, (3) centroid shift, (4) directional stability, (5) similarity, and (6) field correlation.

Cell overlap is the fraction of place/end cells on one day that remain classified as place/end cells N days apart. In this approach, for each animal we consider all possible session pairs that are separated by N days. For instance, session 1 and 2 (N=1), 3 and 7 (N=4), 10 and 16 (N=6), etc. We then calculate the fraction of place/end cells that remain classified as such for each session pair at each N day interval and pool the data across animals. It is expected that this approach would yield more samples on sessions 1 day apart than on 50 days apart, which leads to larger error towards long timescales in Fig. 2A. Periods of learning, trained, and re-exposure were analyzed simultaneously. Damage and recovery periods were not included.

Fraction consecutively active refers to the number of neurons that were classified as place/end cells on consecutively recorded sessions (Fig. 2A, right panel). For each neuron we calculated all consecutive sessions it was classified as a place/end cell and counted how many neurons were place/end cells on one session only, two sessions, three session, etc. For instance, a given place/end cell could have a response field on one session, lose it the next session, comeback for 3 sessions consecutively, lose it for one day, and then become responsive again for 4 more sessions consecutively. Such a neuron would be classified as being three and four sessions consecutively active. The probability distribution for 2 to 40 consecutive sessions was then calculated. Neither of these two metrics are sensitive to changes in response field position, thus a place cell changing directional preference or peak of activity in the linear track on two sessions would be considered to be overlapping and consecutively active.

Centroid shift was defined as the difference between the centroid position of place/end cells in two sessions at N days intervals. The centroid position was obtained by averaging the normalized response field of each place/end cell above a threshold of 0.5. When neurons had multiple peaks in their response fields, we only considered the centroid closer to the reward port the animal was running towards. We then plotted the distribution of centroid shifts at 1, 10, and 20-day intervals as well as the fraction of place/end cells shifting their centroids by less than 10% as a function of day intervals. For

place cells, 10 % is equal to 5 bins (15 cm) and for end cells is equal to 3 frames (480 ms). The cell overlap, fraction consecutively active, and centroid shift were compared to a distribution in which the same number of place/end cells randomly gained/lost a response field at random positions in the maze or time after arrival at the reward port.

Directional stability is the fraction of neurons (shown in %) retaining a response field with the same directional preference between two sessions. Directional stability was calculated by finding all place/end cells that had retained the same directional preference between two sessions and dividing by the session with fewer number of place/end cells.

Similarity: for any two sessions separated by N days we found the place/end cells that retained their classification as place/end cells with the same directional preference. For every place/end cell we calculate the correlation of their response field between the two sessions N days apart and counted what fraction of these neurons had a correlation above 0.4. The threshold of 0.4 was selected based on the average correlation observed in naïve animals exposed to the linear track. The same calculation was performed for place/end cells randomly selected between two sessions at the same interval. The field similarity was obtained by subtracting the fraction of neurons with correlation above 0.4 in the random sample from the actual data.

Field correlation was calculated similarly to the field similarity but instead of calculating the fraction above 0.4 correlation we calculated the median correlation across all neurons retaining their classification as place/end cells with the same directional preference between two session. In this case, the random correlation level was not subtracted but it is shown in **Fig. 2f and 3f**.

Identifying rotated sessions: we noticed that in some sessions the direction selectivity of a large number of place/end cells would change abruptly. That is, place/end cells with response fields when the animal runs in the right direction and arrives at the right side of the maze in one session would change this rightward preference to a leftward preference in the next session, and vice versa. To quantify the magnitude of this effect we defined a session to have undergone a rotation using two methods. Method 1: a representation in a session was classified to be rotated if (1) place/end cells in one session with preference in one direction would be classified as place/end cells with preference in the opposite direction on the next session; (2) the number of place/end cells changing their directional preference was larger than the number maintaining their preference; and (3) reassigning the directional preference of place/end cells in a rotated session improved the directional stability by 2-fold or more. This definition was used in **Fig. S9a**. Alternatively, we also calculated changes in directional preference by considering only place/end cells that retain a field response across two consecutive days. Method 2: (1) we determined if a place/end cell had a left, right, or bidirectional firing preference; (2) changes in classification from place to end cell or vice versa were not considered as a change in directional preference; and (3) only cells with fields on two consecutive sessions or more were analyzed. Cells that underwent changes in classification from place to end cells were not excluded from the analysis. Next, in each two consecutive sessions we determined the number of place/end cells changing the direction of their field and divided it by the total number of cells passing our criteria. We refer to this parameter as the flipping ratio, which fluctuates from 0 to 1 with a clear bimodal distribution (**Fig. S9b**). Sessions with flipping ratio higher than 0.5 were classified as being rotated **Fig**

S9c. The two methods identified 98 % of the same sessions (90 and 92 sessions out of 330 sessions).

Network graphs: an adjacency matrix was generated by calculating the pairwise Pearson correlation between neuronal activity (neuron.S) of all neurons during a 20-minute linear track or 10-minute home cage recording. Only statistically significant ($p<0.05$) correlation values above 0.10 were used to build the adjacency matrix. The correlation threshold was also tested at 0.15 or 0.05 and similar results were observed **(Fig. S21)**. The weight of the edge between two neurons was set to be equal to the correlation coefficient. The networks are plotted with the network graphical tool (Gephi 0.9.2, www.gephi.org) using the built-in ForceAtlas2 layout. Modularity was calculated with a Markov diffusion built-in function from Gephi using a diffusion time of 0.7. Cell groups are defined as cells making up each module extracted by the modularity calculation. All the modules or the 11 largest modules (whichever was smaller) were analyzed and a neuron was classified as belonging to one of the cell groups or none. Unless stated otherwise, cell group participation was not changed across days and neurons were assigned to belong to the same cell group across days. The connectivity of a node in a graph was calculated by counting how many edges a node made with all 11 modules. The cell group connectivity was defined as the fraction of edges from node A to other nodes residing in the same cell group divided by all edges from node A. Metrics including degree, average clustering coefficient, eigenvector centrality were calculated using built in and custom MATLAB functions.

Synthetic dataset: it is possible that the observed correlations of neuronal activity could be explained by random coincidence of activity arising from field overlap between place/end cells. To test if random firing of neurons within their field could lead to high levels of correlation, we generated a synthetic dataset sharing similar properties as the original. In this synthetic dataset the spikes of all neurons from the original dataset were randomly shuffled in time. Spikes in the synthetic dataset were generated with a uniform random distribution so that: 1) in each 20 minute session, the real and synthetic neuron maintained the same number and amplitude of firing events as the original; and 2) the position in the maze were a spike occurred was preserved. This was achieved by discretizing the mouse position in the maze into 450 bins, assigning a value between 1 and 449 when running right and a value of -1 to -449 when running left. At the ends of the maze, when the animal is not moving, the position was assigned to be -450 at the left side of the maze, +450 at the right. Then, for each spike of each neuron we did the following: 1) find in which frame the neuron fired, 2) get the position in the maze when that firing occurred (using values defined above), 2) find all the times the animal was at that position in the maze during that session, 4) randomly reassign the spike to one of the times the animal was in that position but the neuron was not active. This approach would not significantly affect neurons that fired exclusively at one position every time the animal was there. However, these cells are rare, since even robust place cells would show a variability in the location along the maze where they fire.

Linear decoding of behavior with place/end cells: The position of a mouse on frame i on day A was decoded using the neuronal activity (neuron.S) of place/end cells in a

moving window between frames *i*-5 and *i*. This resulted in an n x 5 training matrix (n being the number of place/end cells and their signal in 5 frames). The training matrix was used as an input into a generalized linear model (*glmfit,* normal distribution) with the mouse position at frame *i* as the target output. The mouse position was discretized in 50 bin intervals, each bin is 3 cm. The linear model was trained with randomly selected 10 minutes of place/end cell activity and mouse position. Decoding mouse position on the same day was performed with the remaining 10 minutes of data. For time-lapse decoding the linear decoder was trained with the full 20 minutes of place/end cell activity on day A. The linear model was then used to decode mouse position on day B using the neuronal activity on day B of the same place/end cells used to train the decoder on day A. The decoder was also tested on a random dataset in which the neural activity of each neuron was randomly shuffled in time.

Linear decoding of behavior with cell groups: Decoding mouse position using cell groups was performed in a similar fashion, but instead of using place/end cell activity we used the summed neuronal activity (*neuron.S*) of neurons making up each cell group. Cell group activity was determined at session N by: (1) calculating the maximum projection of the adjacency matrix across 5 trained sessions (sessions 5-10), (2) calculating the cell groups using Markov diffusion on the maximum projection of the adjacency matrix, and (3) assigning the same cell group to all sessions. The IDs of neurons making each cell group were used to calculate their integrated group activity in all other sessions. Thus, we used the *neuron.S* output from CNMFe, identified cells in a group, and integrated the neuronal activity for each group. Resulting in a 11 x 5 training matrix composed of signal from 11 cell groups in a window of 5 frames and one target (discretized mouse position) as a training set. Distribution of training data in time-lapse decoding using cell groups was performed similarly to place cell decoding. In both decoding approaches, the output of the linear decoder was smoothed by moving window average of 5 frames. The mean absolute error between the decoded and real mouse position was used to quantify the accuracy of the decoder. The accuracy of time-lapse decoding was calculated by taking the median across all mice in both hemispheres.

Nonlinear decoding of behavior with cell groups: To account for the nonlinear and asymmetric nature of fields encoded by cell groups we also performed position decoding using a nonlinear input-output time series neural network (*timedelaynet*). This network was set to have 10 hidden layers and a delay of 800 ms (5 frames). The network was trained with 70% of randomized cell group activity (11 inputs) and mouse position, validated with 15 % and tested with 15% of the remaining data. The network trained in one session was used to decode N sessions apart.

Decoding behavior with graph topology: we also decoded mouse position by approximating the mouse position based not on which neurons are active but rather on which functionally connected neurons in a graph are active. For day A we extracted where in the linear track each neuron was active and calculated the preferred firing position of each neuron by taking the median position of all its spikes. Thus, each neuron was assigned a positional preference in the maze ranging from 0 to 50 (3 cm bins). Then, the neuronal activity (neuron.S) was used to calculate the pair-wise neuronal correlation and build a network graph where functional connections between two neurons was proportional to the Pearson's correlation of their activity. The mouse position was decoded at frame i by integrating the activity of each neuron between frames *i* and *i*-5.

Neurons with nonzero neuronal activity in the 5 frame window were considered active. We then found the nearest neighbor of each active neuron in the graph using Dijkstra's Shortest Path First algorithm in which the weight of an edge between two nodes is set as the distance. Only near neighbors that are also active in frame *i* to *i*-5 and within a radius of 0.5 were considered. We modified each neuron's preferred firing position in the time window by averaging it with the preferred firing position of the nearest neighbors. This approach ensures that the decoded position from a neuron that is highly interconnected would be weighed more than the position decoded from a neuron with sparse connectivity in the graph. Finally, the mouse's decoded position at frame *i* was determined to be the median of the preferred firing position of all neurons active and their neighbors. We compared the performance of all decoders with a randomized dataset in which spikes were randomly shuffled in time. The error reported represent the mean absolute error ± sem between the mouse position and the decoded position. All periods including learning, trained, and re-exposure were decoded simultaneously. Sessions with rotated representations significantly decreased the decoders efficiency but were not removed from the analysis. Diagrams showing different decoding approaches are shown in Fig. S14.

Classifying cells using graphs: here we seek to identify whether classification of neurons as place/end cells or neither can be achieved on a day based solely on synchrony and graph topology without any behavioral input. To achieve this, we first obtained the cell group classification of each neuron from the network graph using the Markovian diffusion approach (see above). We also calculated the cell group connectivity of each neuron, a metric which determines how many edges a node makes with its own cell group divided by all the edges it makes with all other neurons. Cell group connectivity is determined from the graph connectivity in each session. The cell group classification of a neuron was updated if in one session its cell group connectivity towards another cell group increased above 0.5, otherwise it remained in the same cell group. We observe approximately 80 % of place cells do not drift into other cell groups. We then assigned end cells an index of 1, place cells an index of 2, and cells with no statistically significant response field an index of 0. Cells classified as place and end cells were not analyzed. On one day, we ranked each cell group based on how many place/end cells each cell group contained. We assigned the two cell groups with the most place cells as a place cell group, and the two with the most end cells as the end cell group. On another session we identified the cells belonging to the original place/end cell groups as place/end cells. Cells that had a cell group connectivity of 0 or were not part of the place/end cell group were set as having no response field. Note that behavioral information only from one trained day is necessary in order to assign a label to each cell group. We compared the accuracy of classification by comparing how many correct classifications arise by using firing rate. In this approach, we discretized firing rate into 11 bins and performed the same analysis on these binned firing rates as we did with cell groups.

Decoding centroid shifts using graphs: we observe that response fields of place/end cells drift from day to day. Here we investigated whether these drifts can be decoded from changes in graph topology without the need of behavioral information. First, we calculated the centroid of all place cells on each session, end cells and cells without a

response field were not included. We then calculated the network graphs from neuronal activity of all neurons in two sessions separated by N days, session A and B. From neuronal activity on session A we identified the nearest neighbors in the graph within a radius of 0.5 of all cells having a nonzero cell group connectivity. For each neuron, regardless of its classification as place/end cell, we then calculated the mean centroid position of all its place cell neighbors. On session B, we performed the same analysis using neuronal activity from that session but the centroid of the place cell neighbors was the same as that determine in session A. Thus, every cell in session B was assigned a neighbor centroid as long as it had neighbors who were place cells on the session used as training. The error of this decoding approach was determined by taking the mean absolute deviation of the decoded centroid and the experimentally determined response field centroid on session B. Without knowledge of behavior on session B, the best estimate is to assume that response fields do not drift. Thus, we compared our decoder with the error observed if fields were maintained constant from session A to B.

Predicting future response field drifts: here we investigate whether future changes in response fields are to some extent encoded during performance of the task. In this case, we used neuronal activity and behavioral data from one session to make a prediction about another session N days in the future. We hypothesized that if a place cell has near neighbors in a graph whose response field centroid is different than its own then it is more likely to drift its field in the direction of the average centroid of its neighbors. Thus, using data from session A we subtracted the near neighbor centroids of a place cell from its own response field centroid. We then plotted all the actual field drift between two session as a function of the neighbor-node differences and fitted the data to a linear equation. Linear fitting was performed on less than 10% of randomly selected data from all the mice. The linear relationship allowed us to make predictions about future field drifts using data from one session. We compared the linear predictions to that obtained by the linear relationship between field drift and firing rate, random guess of the field drift, or assuming no field drift. In all cases the simple neighbor-node linear relationship outperformed the other prediction models.

Histology

Mice were perfused transcardially with 4 % paraformaldehyde in phosphate buffer saline solution at pH 7.4 (PBS). Brains were extracted and fixed overnight at 4 ºC in the same PFA solution. Brains were embedded in 3 % agarose and sectioned in a vibratome into 80 µm thick sections. Slices were washed in PBS for 30 minutes at room temperature, permeabilized for 15 minutes in PBS+0.3 % Triton X100, and washed again with PBS. Slices were blocked for at least 1 hour at room temperature in PBS+10 % fetal bovine serum solution (FBS) before incubation with primary antibodies in PBS+1% FBS overnight at 4 ºC. Chicken polyclonal anti-GFP (1:1000 dilution, Aves Lab, GFP-10-10) and mouse monoclonal anti-RGS14 (1:250 dilution, UC Davis, 75-170) antibodies were used to label GCaMP6s positive or CA2 neurons, respectively. After incubation overnight with primary antibodies, slices were washed three times with PBS and incubated for 90 minutes with secondary antibodies (Alexa Fluor 488 goat anti-chicken, and Alexa Fluor 555 anti-mouse, Invitrogen) diluted to 1:1000 in PBS+1% FBS. Lastly,

sections were mounted on a glass slide with Fluoromount (F4680, Fluoromount Aqueous Mounting Medium).

Quantification and statistical significance

All statistics were done with nonparametric and parametric approaches either using a Wilcoxon rank-sum test, two-way ANOVA, or a 1000 iterations of a bootstrap shuffling procedure. Correlation p-values were calculated using a two-sided t-test. The p-values for the linear regression fits (R-squared values) were obtained by testing whether the model fits better than a constant (F-statistic vs constant model). Measurement of effect size were performed with the parametric Cohen's d metric or the nonparametric Cohen's U3 **(see supplementary Table S1 and S2)**. Values are shown as median ± standard deviation (SD) unless stated otherwise. We did not record from 3 CA1 regions after the no-task period and data from these mice were only included when analyzing learning and trained periods.

Data availability

Sample raw data and processed data is available at (https://doi.org/10.22002/d1.1229). Custom MATLAB scripts, fusion360 CAD files for the design of the custom microendoscope, and CNMFe parameters used are available at (https://doi.org/10.22002/d1.1229).

Supplementary Text

Cell Registration across days

Accurate alignment of neurons across sessions spanning days to months has noticeably become a challenging aspect of calcium imaging. Taking from studies using multiphoton imaging, registration approaches using single photon epifluorescence involve alignment of neurons between sessions based on their spatial footprint. However, the spatial footprint of extracted neurons using microendoscopes can suffer from artifacts arising from changes in firing rates, extraction algorithms, GRIN lens aberration, and motion correction artifacts. These artifacts can be compounded when dense cellular labeling, as observed in transgenic animals, is analyzed. To overcome these limitations, we employed two novel approaches: 1) interleaved analysis and 2) batch analysis.

CellReg registration assigns scores to each registration, thus providing a useful metric to asses registration confidence (*36*). However, validating registration across days using such approaches is not possible since no ground truth is available. Alternatively, the statistical nature of the CNMFe and CellReg algorithms makes it possible to analyze how minor changes in neuronal footprint can affect signal extraction and alignment. We explored the effect of analysis noise by using a 20-minute recording of CA1 activity in a mouse running in the linear track. We generated two videos, one is the original 20-minute video, and another is a temporally concatenated version of the original video (40 minute long). We motion corrected and analyzed both videos separately using CNMFe as described in the methods section. Temporal concatenation of the video increases the number of frames and changes the statistics used by CNMFe to extract neuron footprints, thus leading to minor changes in the footprints. CNMFe effectively extracted similar fraction of ROIs from both videos (649 in original vs 646 in concatenated). The footprint

extracted by CNMFe was thresholded so that only the pixels above the 60th percentile formed the ROI footprint. Next, we tested whether CellReg was able to register ROIs from these two videos using the spatial correlation of their footprint. Using a 0.5 confidence level in CellReg we found that CellReg registered 85 % of ROIs (549 out of 646) and at a confidence of 0.95 the fraction dropped to 76 % (492 out of 646). The misaligned ROIs also led to 15 % (0.5 confidence) and 24 % (0.95 confidence) increase in the number of ROIs. Thus, we observe that minor fluctuations introduced by CNMFe can lead to inaccurate registration of neurons across days when only using footprints. To investigate if cell registration across days could be improved we tested two additional method of analysis.

Interleaved analysis: in this approach a session of data was common between two days to be aligned. This common session served the purpose of decreasing noise in the extracted neuron footprints and provided firing data common to both datasets. Minor motion artifacts between two sessions were corrected by a fast Fourier transform method using the PNR image output from CNMFe. The spatial correlation among all neurons with centroid distances below 15 pixels between the two alignment datasets was calculated. The correlation of the CNMFe deconvoluted neural activity was also calculated for all neuron pairs with centroid distance less than 15 pixels. Only spatial and temporal correlation above 0.5 and 0.3, respectively, were considered. An alignment coefficient was calculated by multiplying the spatial correlation and temporal correlation of every potential neuron pair. The alignment coefficient was binned in 100 intervals and the probability distribution calculated. A plot of the probability distribution showed a clear bimodal distribution (Fig. S2f), a threshold of approximately 0.5 was selected manually. All neuron pairs below this threshold were deleted, and the remaining were aligned based on an iterative selection of pairs with the highest alignment coefficient. This approach was validated by aligning a video to itself, where we observed that including temporal information increased aligning accuracy by about 10%.

Batch analysis: Here, we take advantage of small motion artifacts across days in order to motion correct and analyze several days together in one animal. This ensures, that even if the neuron decreases its firing rate significantly, CNMFe will still draw an ROI around the neuron and extract its firing activity or show that it is not active. Because in some cases the FOV drifts in a non-rigid manner, we restricted our analysis to less than 30 sessions. In some figures we show the correlation image of neurons per session **(Fig. 1g and S4a)**. These images were obtained by analyzing videos of CA1 calcium activity during home cage exploration and linear track (30-minute videos). The resulting correlation image from CNMFe was then motion corrected using a Fourier transform method as mentioned above.

Using interleaved analysis, we investigated whether using footprint and the activity profile of a neuron can enhance the registration across days. CellReg uses two metrics (spatial correlation and centroid distance) in order to build a distribution of potentially same or different cell pairs. However, these metrics are not fully independent, potentially leading to significant overlap and high uncertainty in the registration. Registration can be improved by temporally concatenating two days and setting the activity of a neuron as a constraint for alignment, a metric which is independent and orthogonal to both centroid and shape. However, this requires the ability to analyze multiple days simultaneously, which is something we can easily achieve using chronic implants thanks to the minimal

drift across consecutive days. To test the effectivity of this approach, we performed the same analysis as above but including in the analysis that neurons should not only look similar (0.5 spatial correlation) and be located in the same region (<15 pixels centroid distance) but should also have similar activity (0.3 temporal profile). Using this method, we can align 93 % of all the ROIs (599 out of 646 ROIs), leading to 7% increase in the number of total ROIs. There were 80 ROIs not satisfying our shape and PNR constrain (see above) in the original video and 66 in the concatenated version. After removal of these ROIs, the fraction aligned using our method (90 %) or CellReg (82 %) did not change dramatically. Using our method, we calculated that analysis of two versions of the same data by CNMFe resulted in footprints with areas differing by as much as 19 ± 18 $pixel^2$.

We tested whether the improvement persisted when analyzing multiple sessions. In this case, we analyzed three 20-minute recordings of CA1 activity in the linear track separated by 1 day **(Fig. S2a)**. Two sets of recordings were generated. Set A contains temporally concatenated recordings of day 1 and day 2 (in that order) and set C contains recordings from day 2 and 3 **(Fig. S2d).** The recording of day 2 is common to both sets and the calcium activity of such neurons can be assigned as a constrain on the alignment.

In batch analysis we concatenated all three videos into a 60-minute video (Set D) and analyzed it with CNMFe. Set D can be used as validation, in this set we observe that most (97 %) ROIs are active in all 3 days and there is a total of 731 ROIs **(Fig S2d)**. CNMFe extracted 702 ROIs in Set A and 700 ROIs in Set C. CellReg is able to align 74 % of these ROIs at 0.5 confidence (521 out of 700) and 63 % at 0.95 confidence (442 out of 700). Due to the misalignment, a total of 881 ROIs are detected in both sets, an error of 21 % when compared to the 731 ROIs detected by CNMFe. Using the interleaved method, we can align 87 % of the ROIs (607 out of 700) and identify a total of 794 ROIs (13 % error compared to 731 ROIs by CNMFe). CNMFe found that 97 % of ROIs are active all three days, the interleaved approach identifies 88 %, and CellReg found 72 %. All three approaches lead to the same answer; most neurons are active all three days and similar results were obtained when extending the analysis to longer timescales. Due to the more reliable registration obtained by batch analysis using CNMFe as well as the higher detection levels **(Fig. S2b-c)**, this method was used in this study.

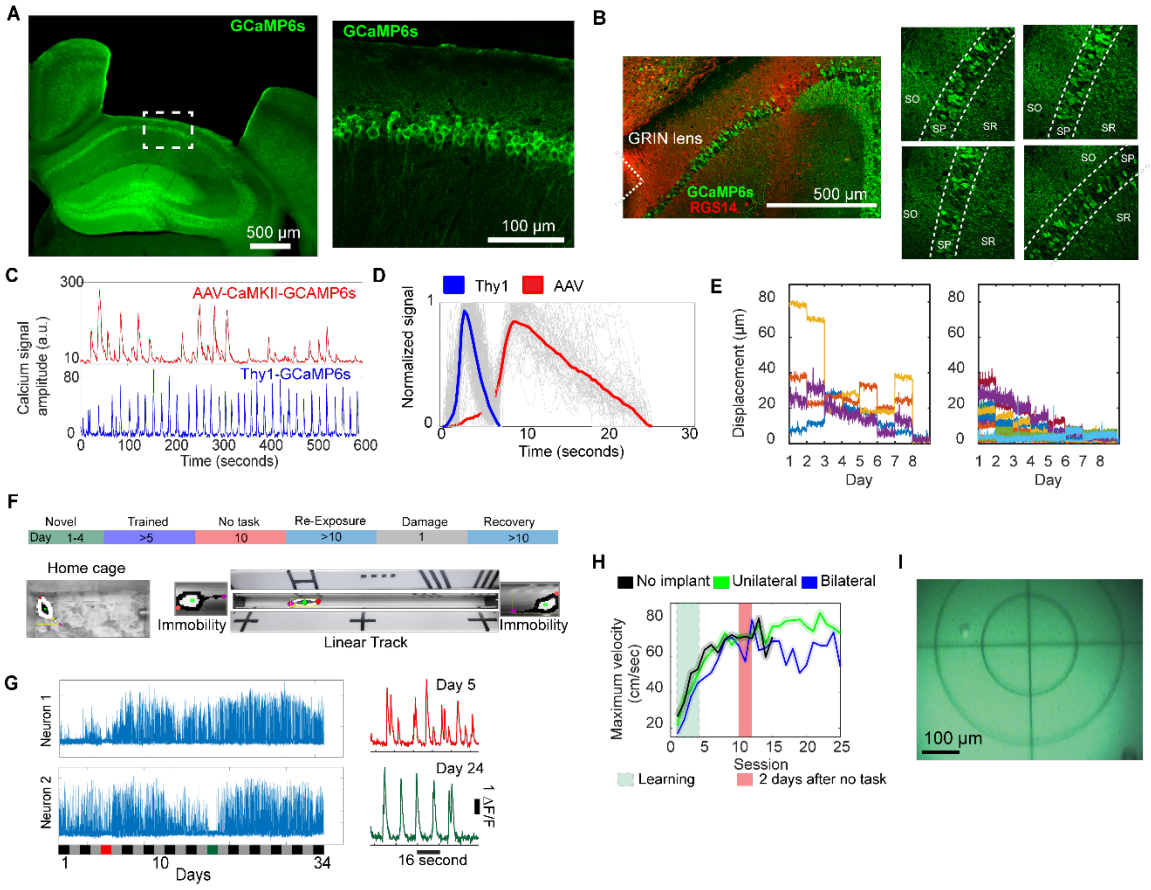

Fig. S1.

Simultaneous bilateral imaging of CA1 activity in freely moving Thy1GCaMP6s mice. (A) (Left) Histological labeling of GCaMP6s positive neurons in a coronal section of a mouse that carried the 1.8 mm GRIN lens implant for over 12 months after the surgery. GCaMP6s positive neurons labeled. (Right) GCaMP6s positive neurons under the GRIN lens implant near the field of view (rectangle left panel), 8 months of data from the field of view of this animal is shown in Fig. S4a. (B) (left) The edge of the GRIN lens implant is over 500 µm away from CA. Antibody staining of the regulator of G-protein signaling RGS14 (a marker for the CA2 region) is shown in red and GCaMP6s positive neurons are shown in green. The edge of the tissue damaged by the GRIN lens implant is shown for reference (white dotted line). (Right) Four coronal sections showing homogeneous labeling of deep and superficial CA1 pyramidal neurons in a Thy1GCaMP6s mouse (SO, stratum oriens, SP: stratum pyramidale, SR: stratum radiatum). **(C)** Calcium transients in CA1 neurons of a mouse running in a linear track using adeno associated viruses to induce expression of GCaMP6s (red trace) and using transgenesis to induce expression of GCaMP6s (blue traces). AAV data was obtained from www.miniscope.org. **(D)** The fastest decaying calcium transients we observed in Thy1GCaMP6s mice had a half-life of 0.56 ± 0.14 s (mean ± SD, n=554), about 5-fold faster than that observed using AAV GCaMP6s, (2.9 ± 0.5 s, n = 63, $p<10^{-32}$, rank-sum

test). This is similar to the faster dynamics observed in L2/L3 neurons of the anterior lateral motor cortex of transgenic animals compared to viral vector mediated GCaMP6s expression (*18*). **(E)** Motion artifacts across sessions after re-attachment of a commercial microscope (left panel) or chronic implant of the microscope used here (right panel). Four mice shown on the left panel and data from 8 mice (13 CA1 recordings) shown on the right panel. Analysis of session to session motions shown in Fig 1b. Reattachment data adapted from (*30*). **(F)** Experimental schedule and tracking of mice in the home cage (left) and linear track (right) during periods of running and immobility (nose, tail, and center of mass shown in red, pink, and green). Cues shown on top and bottom of the linear track. **(G)** (left) background corrected calcium signal from two neurons showing continuous activity throughout several days. (right) Inset showing several seconds of activity on two days of neuron 2. **(H)** Bilateral and unilateral implants do not affect the maximum velocity in the linear track. Rectangles highlight period of novelty (first 4 sessions) and initial 2 days following re-exposure to the task (n = 3 no implant, n = 4 bilateral, n = 5 unilateral). Task performance does not degrade if the animals are not exposed to the task (red rectangle). **(I)** Field of view of the microscope used here (circles have diameters of 250 and 500 μm).

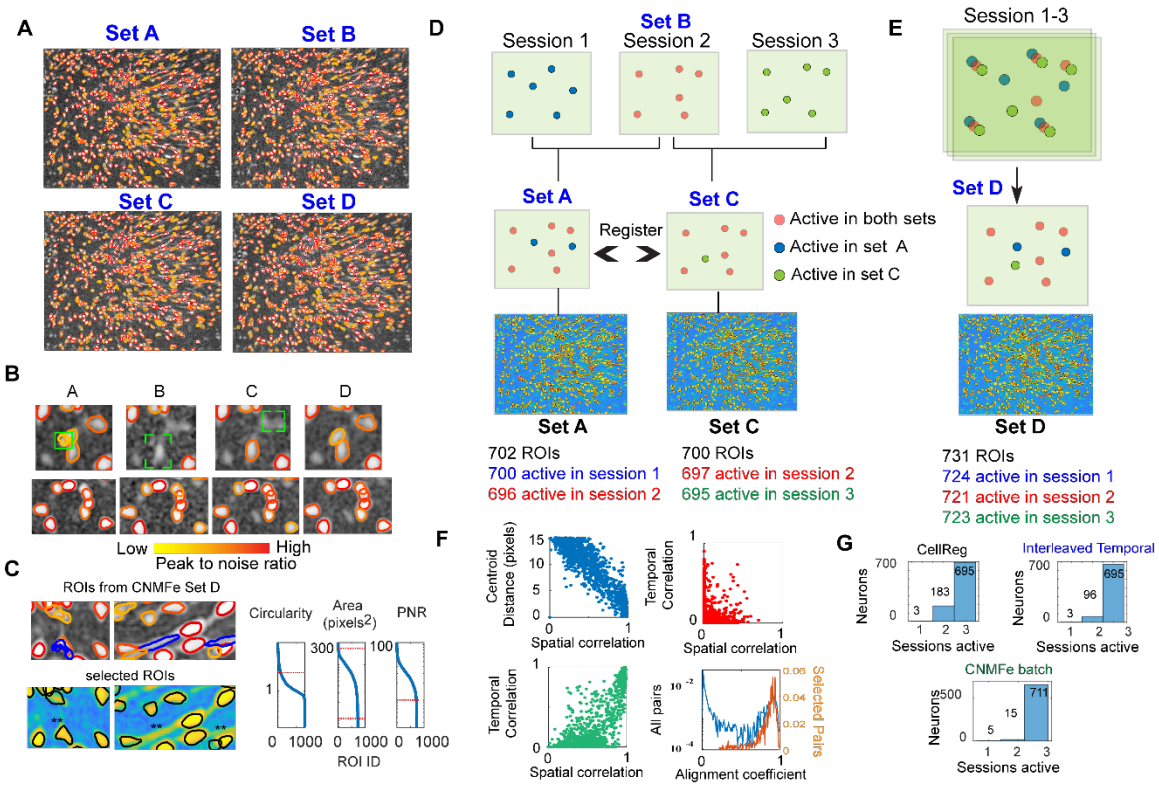

Fig. S2.

Using temporally concatenated datasets improves cell registration across days. **(A)** Three 20-minute recording of CA1 activity in a mouse running in the linear track on three different days were combined into 4 sets. Set A includes day 1 + 2 (14900 frames), Set B is only day 2 (7450 frames), Set C includes days 2 + 3 (14900 frames), and Set D contains days 1 + 2 + 3 (22350 frames). **(B)** Concatenation leads to better signal extraction using CNMFe. Set D performs better than all other sets. Note the regions highlighted by the green rectangle where there is some signal but not enough to be identified as a neuron. The lower panels show another region where four very close neuron are better extracted by combining all videos. **(C)** Two small regions from set D identifying residual noise (left, blue trace) and dendrites (right, blue trace) as regions of interest (ROI). These ROIs were removed from the analysis (asterisks in lower panel) by setting shape and peak-to-noise ratio (PNR) thresholds (right panel). **(D-E)** Diagram showing concatenation approach and CNMFe output. Active neurons were determined from the neural activity (neuron.S) output from CNMFe. Neurons with nonzero neural activity were considered active.. **(F)** Steps used in the interleaved approach tested here. (top left) Relationship between centroid distance and spatial correlation of all ROIS with centroids less than 15 pixels apart. (Top right) Temporal and spatial correlation of ROIs with centroids further than 15 pixels. (Bottom left) Temporal and spatial correlation of ROIs with centroids closer than 15 pixels. (Bottom right) Alignment coefficient of all ROI pairs (blue) and those selected to be the same ROIs across Set A and Set C (orange trace). Note the better separation of distributions in the lower left panel compared to the top left panel. **(G)** Results from all three approaches indicate that most neurons are active most days.

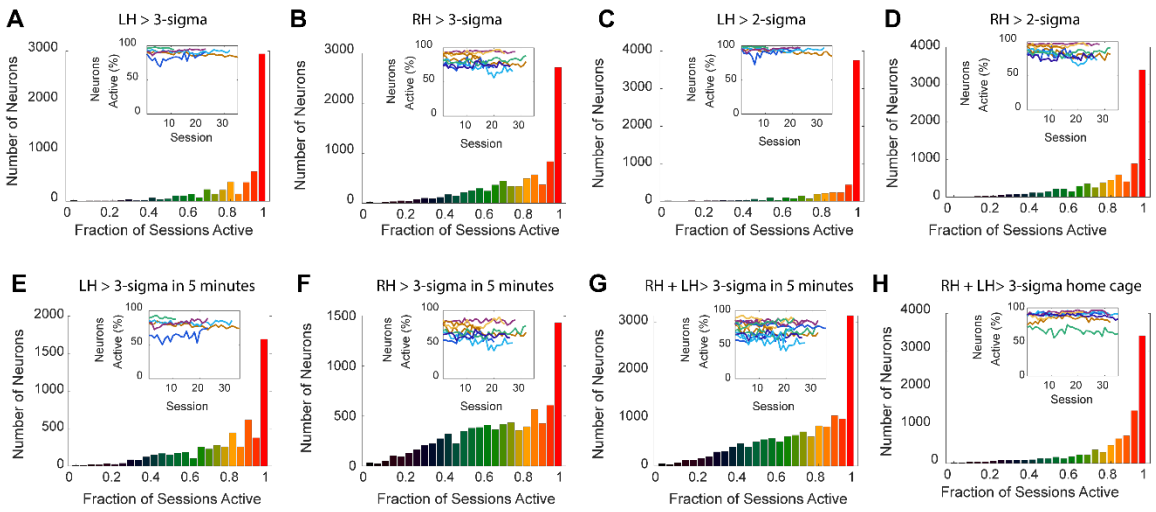

Fig. S3.

Neurons are observed to be active most days. **(A-B)** Distribution of neurons active for specific fraction sessions recorded. Left hemisphere (LH) and right hemisphere (RH) analyzed separately (n = 5 and n=8 mice respectively). The inset shows the fraction of neurons from the total that were active in each session for each mouse. See methods for classification of a neuron as being active. A higher fraction of neurons was active in each session in the left hemisphere than the right hemisphere (90 ± 4 %, n = 131 sessions vs 81 ± 9 %, n = 212 sessions, rank-sum test, $p<10^{-13}$). However, the fraction of neurons active on all sessions was similar across the left (50 ± 20 %) and right hemispheres (38 ± 21 %, rank-sum test, p>0.3). (**C-D**) The fraction of neurons active each session or on all sessions increases if the threshold for defining activity is lowered to 2-sigma. The fraction active in each session in the left (94 ± 4 %) is still higher than the right (85 ± 8 %, rank-sum test, $p<10^{-15}$) but the fraction active in all sessions remains similar (67 ± 20 % in the left vs 60 ± 24 % in the right, rank-sum test, p>0.1). (**E-G**) Considering only the first 5 minutes the mice were in the linear track shows a similar trend. In the left hemisphere 81 ± 9 % of neurons were active in each session, compared to 68 ± 11 % in the right hemisphere (rank-sum test, $p <10^{-10}$). In both hemispheres the same fraction of neurons were active on all sessions (29 ± 17 % for the left and 18 ± 11 % for the right hemisphere, rank-sum test, p>0.2). Counting both hemispheres in the first 5 minutes, the fraction active each session is 75 ± 11 % and 21 ± 14 % on all sessions. (**H**) A similar neuron participation was observed during a 10 minute recording of mice in their home cage (88 ± 4 % of neurons were active in each session and 43 ± 2 % were active every session, due to computational constrains we only included data from 5 right hemispheres and 2 left hemispheres in the foraging analysis).

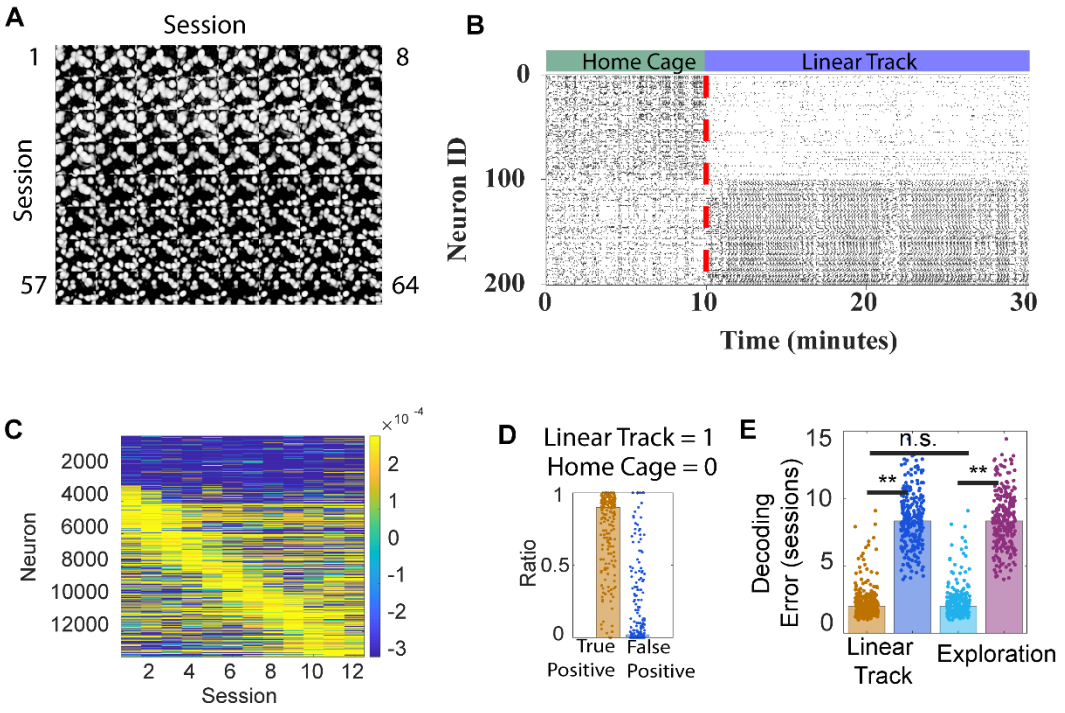

Fig. S4.

Neurons change their firing rate across environments and days. (A) Correlated pixel intensity image of the same region shown in Fig. 1G but across 64 sessions spanning 8 months. There is a 2-month interval between session 37 and 38. Note the gradual change in ROIs due to small drifts in the focal plane of illumination. (B) Despite their stable participation, neurons change their firing rate across environments and days. Raster plot of deconvoluted activity from 100 neurons with increased activity during exploration (ID: 1-100) and while running in the linear track (ID: 101-200). Vertical line shows the transition from one environment to another. Data from the right hemisphere of one animal shown but similar results were observed in both hemispheres of all animals (not shown). (C) Firing rate changes between days. Normalized burst deviation from the population median for all mice across days. For each row we calculated how many spikes (neuron.S) were above or below the median population spike amplitude (excluding periods of neuronal inactivity) for that session and then normalized so that the sum of all spike deviations equals one. Neurons were sorted by their maximum deviation from the median (data from both hemispheres of all mice). (D) Accuracy of a fitting neural network trained with the neuronal activity of 200 neurons with preference for the linear track or home cage (panel B) to identify whether the animal was in the linear track (logical 1) or in the home cage (logical 0). Average true-positive-rate 0.86 and false-positive 0.11 ($p<10\text{-}10$, each dot represents a session, data from both hemispheres of all mice). (E) Bayesian classifier trained with neuronal rate changes (panel C) in the right hemisphere to identify the environment and session, respectively. Average decoding error for the track (orange) and home cage (cyan) were similar (1.8 ± 0.5 and 1.7 ± 0.5). The

blue and violet bars show the error if the inputs (firing rates, rows in panel C) are randomized (data from both hemispheres of all mice,** = $p < 10\text{-}10$).

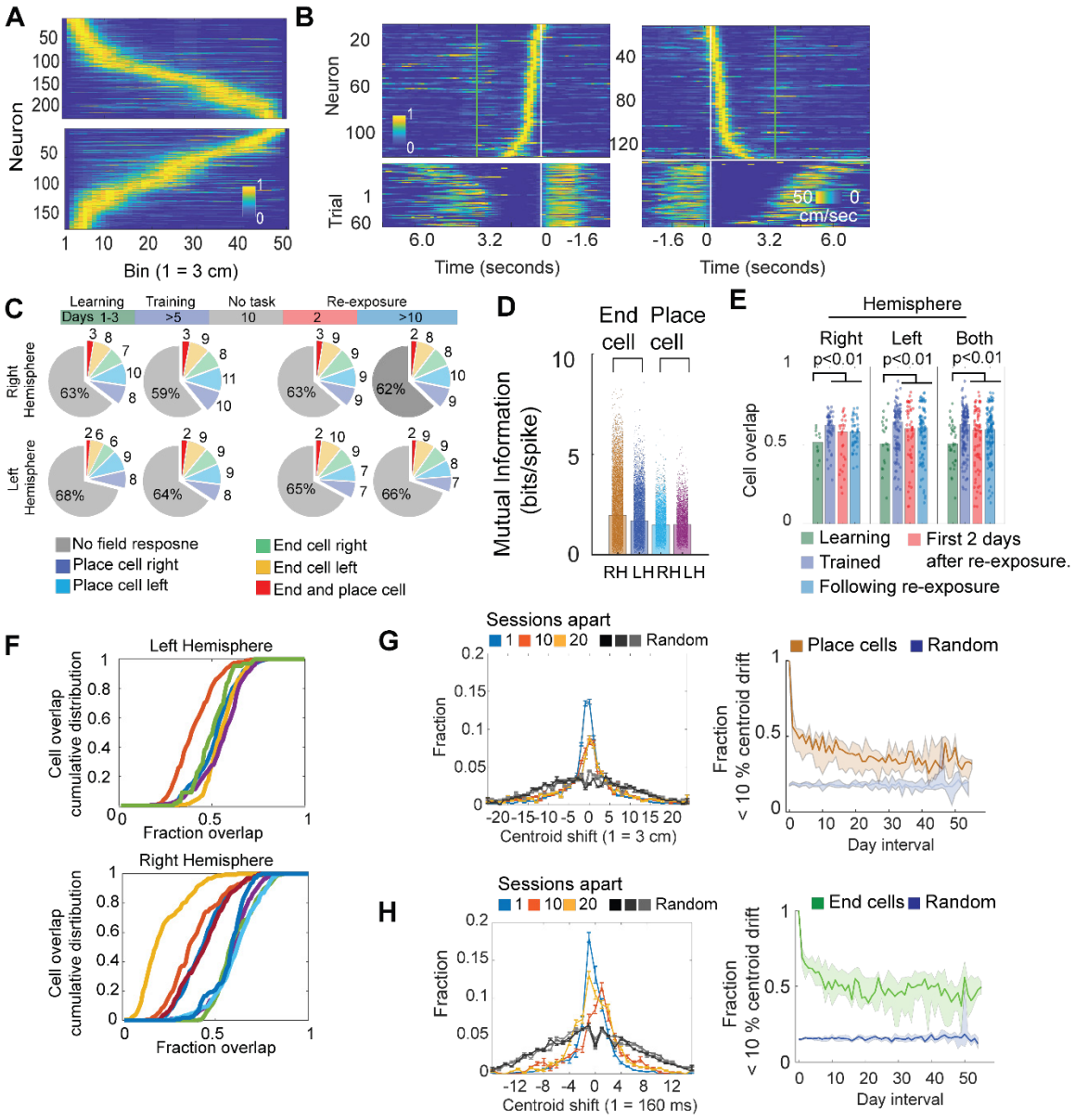

Fig. S5.

Custom microendoscopes in transgenic mice can detect neurons with spatial and temporal fields at the ends of the maze. **(A)** Normalized tuning profiles of neurons with statistically significant place field when running right (top) and when running left (bottom), only data from the right hemisphere in one mouse during one session is shown. **(B)** Normalized tuning curves of end cells firing at the left side (left) and right side (right) of the track after activation of the sugar reward port shown in white lines (the white-green lines indicate the boundaries of the window used to calculate end cell response fields). (Bottom) Mouse velocity on each running trial aligned to the time after reward delivery (white line). **(C)** During periods of mobility and immobility we observe similar fractions of neurons with statistically significant fields (in right hemisphere place cells: 13 ± 7 %, end cells: 11 ± 3 %, p>0.8 rank-sum test, in the left hemisphere place cells: 10

± 8 %, end cells 14 ± 2 %, p>0.8 rank-sum test). There was no significant difference between the fraction of place or end cells across hemispheres (p>0.35 and p>0.6 rank-sum test, respectively). A small fraction of neurons were classified as end and place cells (~2 %). End cells were slightly more direction selective than place cells (% bidirectional, 3 ± 8 vs 6 ± 5, $p<10^{-13}$, n =342 sessions) but failed to fire within their field more often than place cells (% burst within field, 65 ± 13 % vs 79 ± 6 %, $p<10^{-5}$, al mice both hemispheres). **(D)** Mutual information of place and end cells in all mice across all sessions (each dot is one cell). The mutual information of end cells was slightly higher than place cells (in bits/spike, 1.8 ± 1.1 for end cells and 1.5 ± 0.7 for place cells, p=0.004, Kolmogorov-Smirnov test on cumulative distributions). **(E)** Cell overlap of place and end cells during learning (first 4 sessions, 54 ± 12 %), trained (5 sessions to before the no-task period, 66 ± 15 %), the first 2 days following re-exposure (61 ± 15 %) and the subsequent days (64 ± 12 %). Each do represent as session pair during each period and cell overlap was calculated independently for place or end cells, rank-sum test. **(F)** Cumulative distribution of cell overlap between all session intervals in individual mice in each hemisphere. **(G)** (Left) Probability distribution of centroid drifts of place cells at intervals of 1, 10, and 20 sessions (median and SD shown, n = 13). The random distribution was obtained by randomly aligning neurons between the same session intervals and calculating the centroid differences. (Right) The fraction of place cells whose centroid drifts by less than 15 cm as a function of session intervals (10 % of the linear track length). Solid line represent the median and shadow the 95 % bootstrap confidence interval. **(H)** Same as panel G but using end cells whose response fields are measured in time after arrival at the reward port.

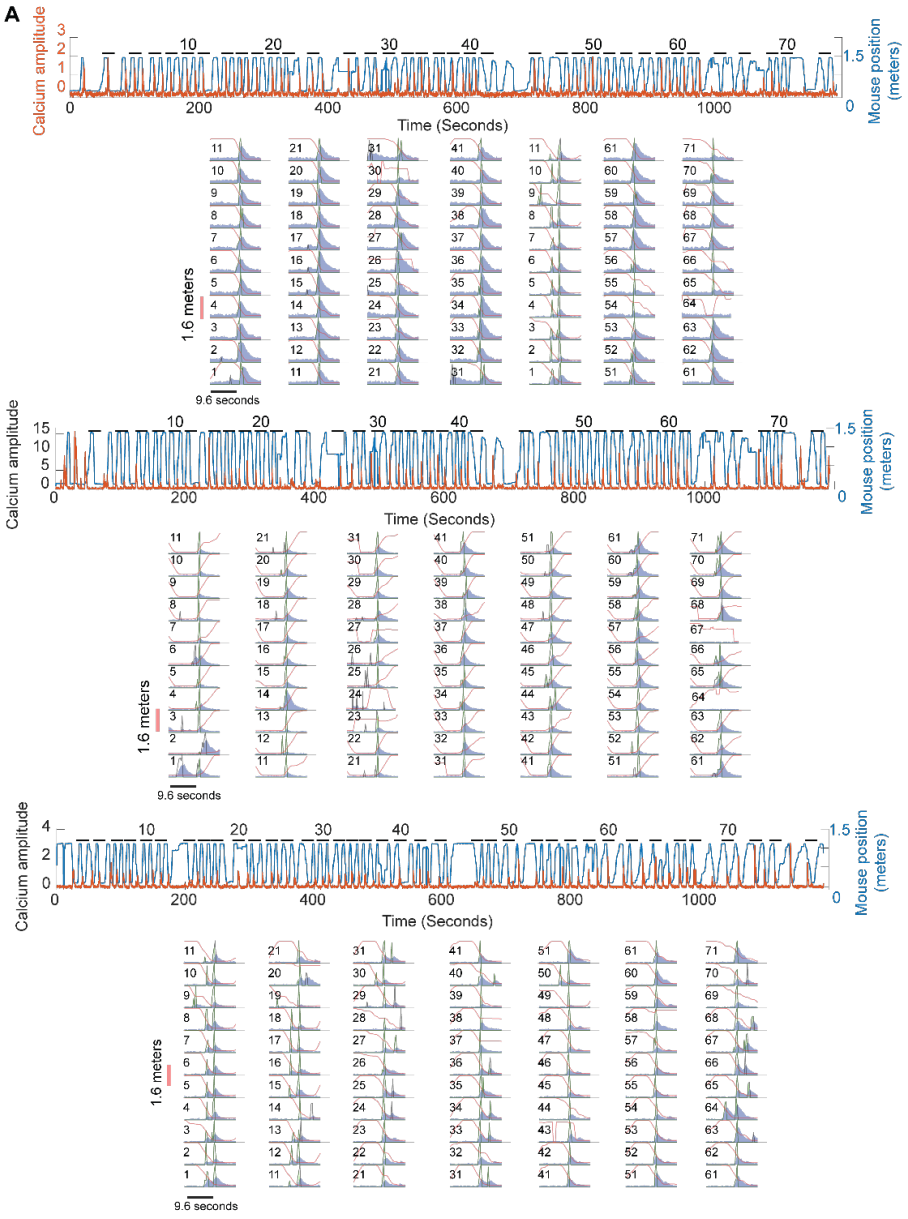

Fig. S6.

Three place cells in the linear track showing robust response to their place fields. The top panels show the complete 20-minute session in the linear track and the bottom panels a ~ 20 second window near the maximum activity of the cell. The blue trace represents the background subtracted calcium transient and it is normalized so that the maximum during the session is equal to one. The black and green lines represent the spikes included and not included in the analysis, depending on whether they occur while the animal was moving or not, respectively. The red line shows the position of the mouse in the track. Numbers indicate the trial number and every even trial is shown on the top panel.

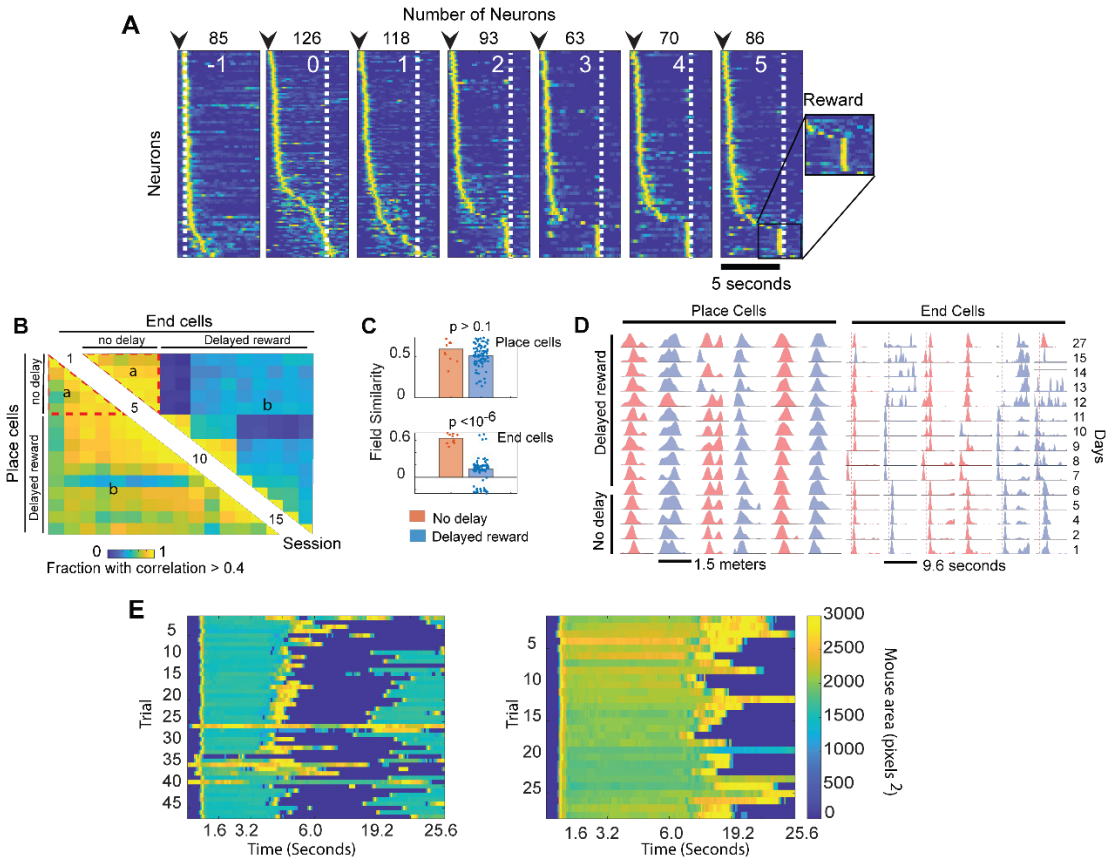

Fig. S7.

Changes in reward time affects the field response of end cells but not place cells. **(A)** Tuning curves of end cells at the right side of the maze during a delayed reward period. White numbers indicate sessions after delay and the black numbers the number cells identified in the session. Black arrows show activation of the IR sensor and the dashed vertical line the time of water delivery. Neurons are sorted by the delay between their maximum activity and activation of the water port. Neuron IDs updated between days, each row may correspond to different neurons (data from one mouse in one hemisphere shown). **(B)** The field similarity of end cells changes after the delayed reward period, but place fields remain stable. Pairwise fraction of neurons with field correlation >0.4 between pairs of sessions. Colors represent the median value of a session pair. Numbers indicate the session. Sessions were grouped as indicated by the dashed red boundaries, letters are labels for each region. **(C)** Quantification of field similarity of place cells and end cells before and after introducing a 5 second delay to the delivery of sugar water (rank-sum test, each dot represents median correlation of a session pair). **(D)** Normalized response fields of six place cells (left) and end cells (right) during the delayed reward period, colors legends as in Fig. 3D. Vertical red line marks activation of the water port. **(E)** To quantify whether mice moved or not during the delayed reward period we used the mouse area extracted with OptiMouse. Normally we observed that mice did not move for about 4 seconds after activating the water reward port (left); however, mice did not move for about 15 seconds when the reward was delayed by 5 second (right) (data from one mouse in one session shown).

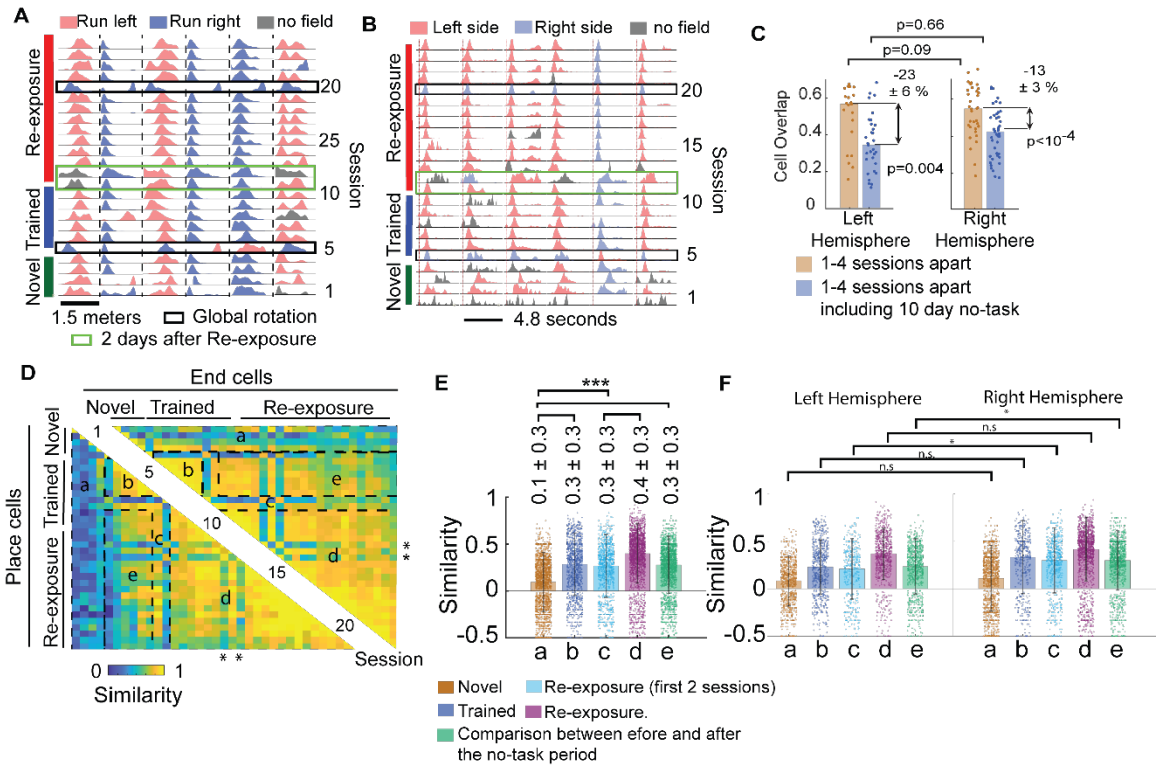

Fig. S8.

Response fields of place and end cells persist through periods of no exposure to the linear track. **(A)** Response fields of six place cells across 35 days from one mouse. Black rectangles show sessions in which response underwent an spontaneous and global change of direction (see Fig. S9). Bidirectional cells not marked. Green rectangle are the two sessions after re-exposure. **(B)** Response fields of six end cells with response fields during periods of immobility in the right hemisphere across 35 days. Black rectangles show sessions in which 88 ± 8 % of fields changed direction. Green rectangle are the two sessions after re-exposure. Red vertical line indicates activation of the water reward port. **(C)** Re-exposure to the linear track after a period >10 days of no task leads to an additional 13-23 % of place/end cells becoming unresponsive to a field compared to the decay observed across 1-4 sessions. This figure is analyzed differently than Fig. 2C in order to get a wider range of comparison between session intervals. This approach was taken to minimize the potential effect of re-exposure on cell overlap and to be able to analyze each hemisphere individually. In this case we calculate the cell overlap across all possible intervals between sessions 1-4 before the gap (shown in blue numbers in Fig. 2D). Then we compare what the cell overlap is between all possible intervals between the session before the gap (blue "1") and sessions 1, 2, 3 after the gap shown in red (median ± sem, rank-sum test). This calculation was performed among all sessions before and after the gap. **(D)** Each pixel represents the fraction of neurons with a response field on two sessions whose response field has a correlation larger than 0.4. Thus, a value of 1 indicates that all the response fields of place/end cells active on these two sessions were correlated above 0.4. Numbers along the diagonal indicate the session. Sessions corresponding to periods of learning, trained, re-exposure, and recovered were grouped as indicated by the dashed rectangle and letters (a, b, c, d, e, respectively). This is a different

animal than Fig. 2E and is shown to highlight periods of rotations happening during re-exposure (region d, marked with two asterisk). **(E)** Field similarity of place/end cells during learning, trained, and re-exposure periods. Field similarity is defined as the fraction of cells with correlation above 0.4 minus the fraction with correlation above 0.4 if neurons were randomly paired between days. *** is $p<10^{-45}$, both hemispheres of all mice combined. **(F)** Field similarity of place/end cells in the each hemisphere across different periods of the task (color legends as in panel E, n = 5 left hemisphere, n = 8 right hemisphere, * indicates a small effect $p<10^{-5}$). See table S1 and S2 for measure of effect size for field similarity shown in panel E and F.

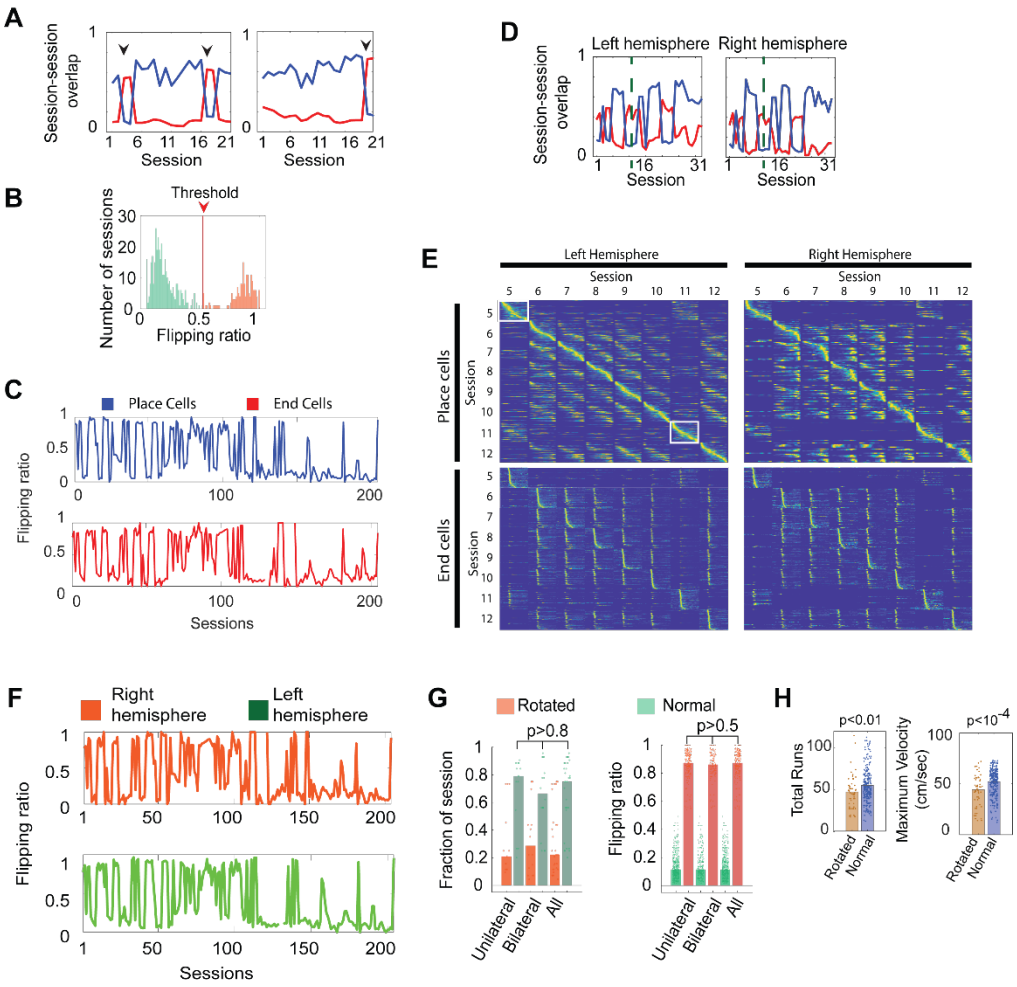

Fig. S9.

Place and end cell response fields undergo spontaneous global rotations. **(A)** Direction specific overlap of place and end cells between two consecutive sessions in the right hemisphere of one mouse. The blue trace represents the cell overlap between neurons in session N and N+1 with the same directional preference and the red trace the cell overlap of neuron with preference for one direction in session N with neurons of the opposite direction on session N+1. An increase in red and decrease in blue traces indicates change in directionality of place/end cells, shown with black arrows. The left and right panels are calculated from two animals being exposed to the linear track on the same days. Rotation events are not correlated across animals, indicating that these changes are not due to major changes in the environment. **(B)** Distribution of session with normal or rotated place/end cell response fields, red line indicates threshold used to identify rotated sessions (all mice both hemispheres, place and end cells counted separately). This distribution was obtained by calculating the number of place/end cells that changed the directional preference of their field between two consecutive sessions divided by the total number of place/end cells that retained their response field between two consecutive sessions (see methods). Approximately ~ 22 % of all sessions were rotated (146 rotated, 256 normal, n=408 sessions in all mice, both hemispheres). Between normal sessions we observe 13 ± 9 % of place/end cell change the directional preference

of their response field; however, in rotated sessions 91 ± 10 % change their directional preference. **(C)** Place/end cells change their direction preference simultaneously. Plot shows the flipping ratio of place cells and end cells in both hemispheres (n = 204, 4 bilateral mice, correlation 0.79, $p<10^{-10}$). **(D)** Flipping ratio of place and end cells in four bilateral mice highlight simultaneous rotations across hemispheres (correlation 0.81, $p<10^{-10}$). **(E-F)** Place/end cell overlap and normalized tuning profiles in a mouse with bilateral implants showing that rotation of response fields involves place/end cells across both hemispheres simultaneously. Dashed vertical line highlight a 10-day period of no task. **(G)** The fraction of sessions with rotated representations was similar for unilateral and bilateral animals (0.17 ± 0.24 unilateral, n = 5, 0.31 ± 0.27, bilateral, n=4, and 0.22 ± 0.25 % for both, rank-sum test, place/end cells of each hemisphere analyzed independently). The flipping ratio (fraction of neurons undergoing rotations in each rotated session) was also similar across mice with unilateral or bilateral implants (0.88 ± 0.09 unilateral, n = 5, 0.88 ± 0.10, bilateral, n=4, and 0.88 ± 0.10 % for both, rank-sum test). **(H)** During sessions with rotated representations mice ran less distance (total number of laps, 0 ± 18 vs 52 ± 20, p = 0.02, rank-sum test) and they reached a lower maximum velocity (48 ± 15 vs 55 ± 13, rank-sum-test, $p<10^{-4}$, each dot represents a session).

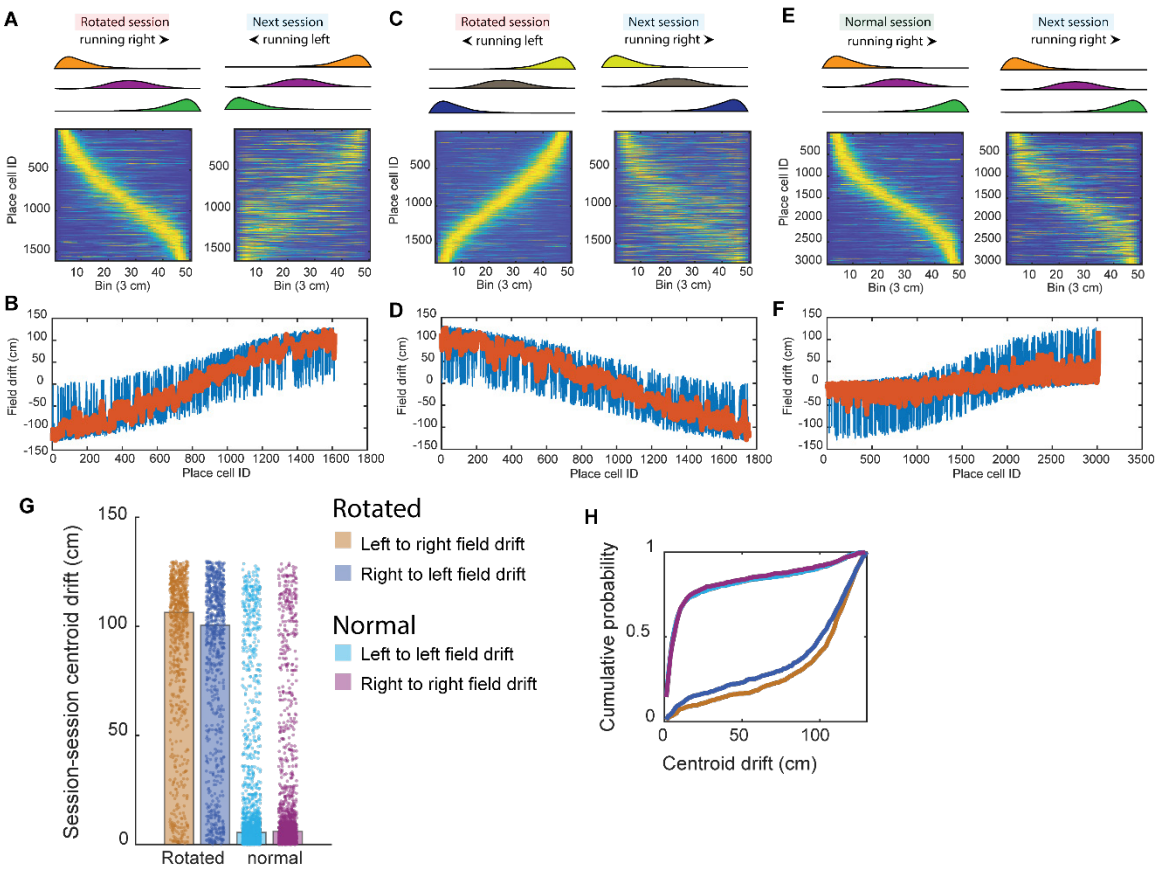

Fig. S10.

Place cells flip their directional preference while maintaining the position of the response field with respect to the reward. (A) (left panel) Response fields of place cells firing in the rightward direction in a rotated sessions (flipping ratio > 0.5). (Right Panel) In the following session, the cells from the left panel were classified as place cells with leftward preference and the majority of field centroids relocate to the opposite end of the maze (cell ID ordering preserved on both panels, 1613 place cells in both hemispheres of all mice). (B) Subtracting the leftward field centroids from the rightward field centroids reveals the overall drift in response field centroids between a rotated and normal. The larger shift in centroids near the rewards positions demonstrates that when place cells change their directional preference, the field centroids maintain their distance from the reward regardless of other cues in the linear track. The blue line is the centroid drift of each place cell and orange line is a moving window average of 10 cells. (C-D) The same analysis as in panel A-B but using place cells firing in the leftward direction in sessions with flipping ratio > 0.5 and rightward directional preference on the following session. (E) (left panel) As a control, we analyzed the response fields of rightward place cells in normal sessions with flipping ratio < 0.5 and compared the field response of these cells on the next session (right panel). These place cells retained their rightward firing preference and response fields. (F) As expected, even though a small fraction of cells change their field centroids but the majority does not change by more than 8 cm (median, 3014 place cells). (G) Place cells in rotated session change their field centroid by 104 ±

39 cm with respect to the next session. In normal sessions, place cells retain directional preference and shift their field centroids by 5.8 ± 35 cm (rotated vs normal $p<10^{-50}$, rank-sum test). (H) Cumulative distribution of field centroids shift of rotated and normal session (colors as in panel G).

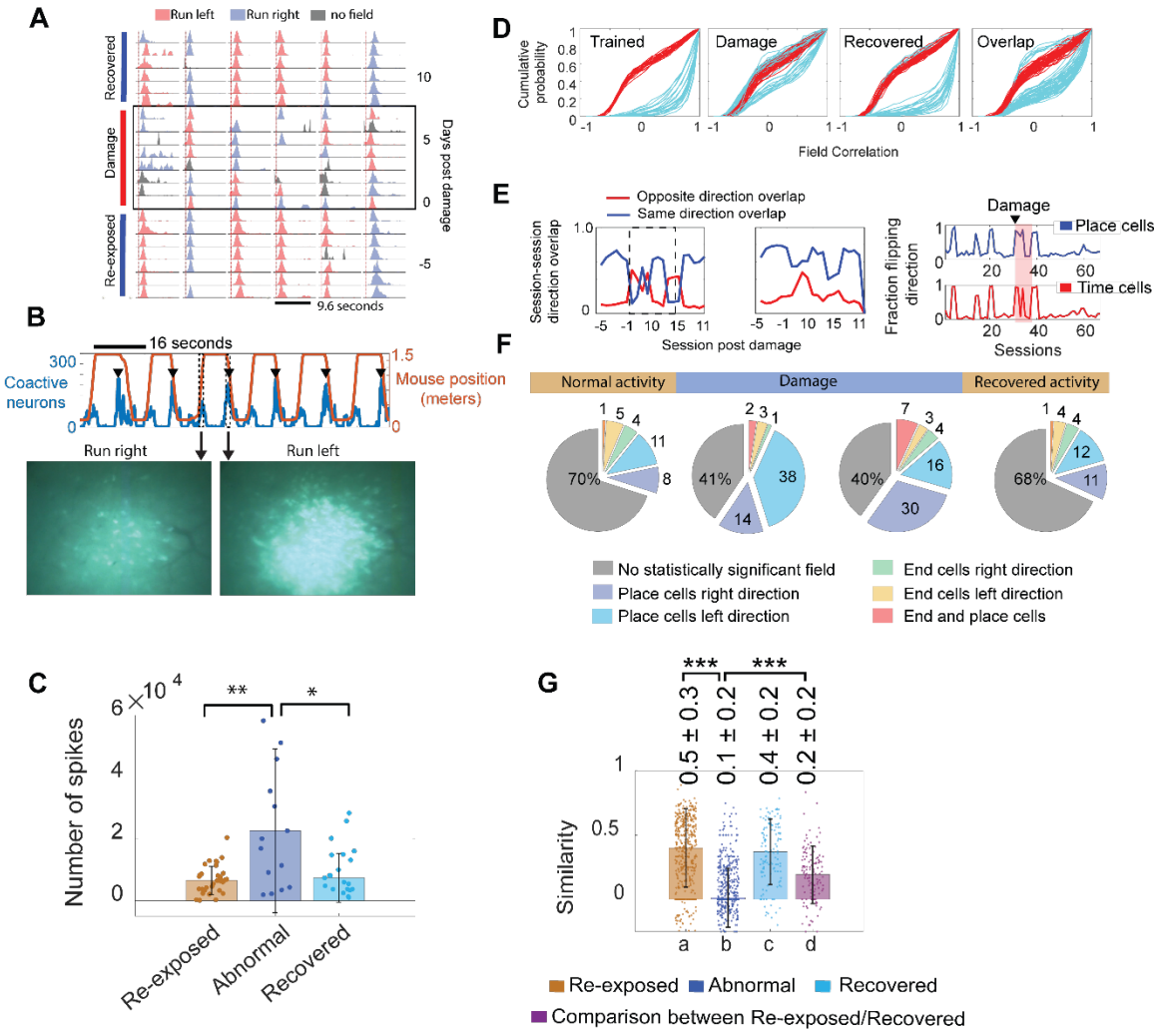

Fig. S11.

High illumination damage induces place specific abnormal activity, increase in place cells, and increased rotation events. **(A)** Normalized response fields of six place time (right) cells before, during, and after the lesion in CA1. **(B)** (top) Blue trace represents the number of coactive neurons during a frame (160 milliseconds), reaching approximately ~250 neurons (black arrows) in one direction but only ~ 100 in the other. Orange trace shows the position of the mouse in the linear track. (Bottom) Single frame capture during a normal running right period and a burst while running left. **(C)** Absolute number of spikes while the mouse runs in the right direction and drinks at the right port minus the number of spikes while the animal runs in the left direction and drinks at the left port. Each dot represents a session, n = 3 mice, ranksum-test, * is p<0.01, ** is p<0.001. **(D)** Cumulative distribution of pairwise field correlation between place cells during learning, trained, and recovered periods. Red traces represent the field correlation of randomly assigned place cell pairs between two sessions. Cyan traces represent the correlation of place cells that retained their responses across two sessions. The overlap traces represent the distribution of correlation between trained and after the damage. **(E)** Following damage to CA1 the incidence of place and time field rotation increased in this animal (4

out of 7 sessions, left panel) compared to before and after the abnormal activity (8 out of 60 sessions, right panel). The middle panel is the fraction of place/end cells that would have similar field if their fields were rotated by 180 degrees. The right panel represents sessions recorded across 7 months and analyzed in 3 datasets of 21-25 sessions each. **(F)** The fraction of neurons with statistically significant place field also increased during abnormal activity (normal 18 ± 10 % compared to 43 ± 13 %, $p<10^{-4}$, rank-sum test, 3 mice). **(G)** Field correlation of place/end cells during learning, trained, and re-exposure periods. *** is $p<10^{-5}$, 3 mice right hemisphere, see Table S3-S4 for quantification.

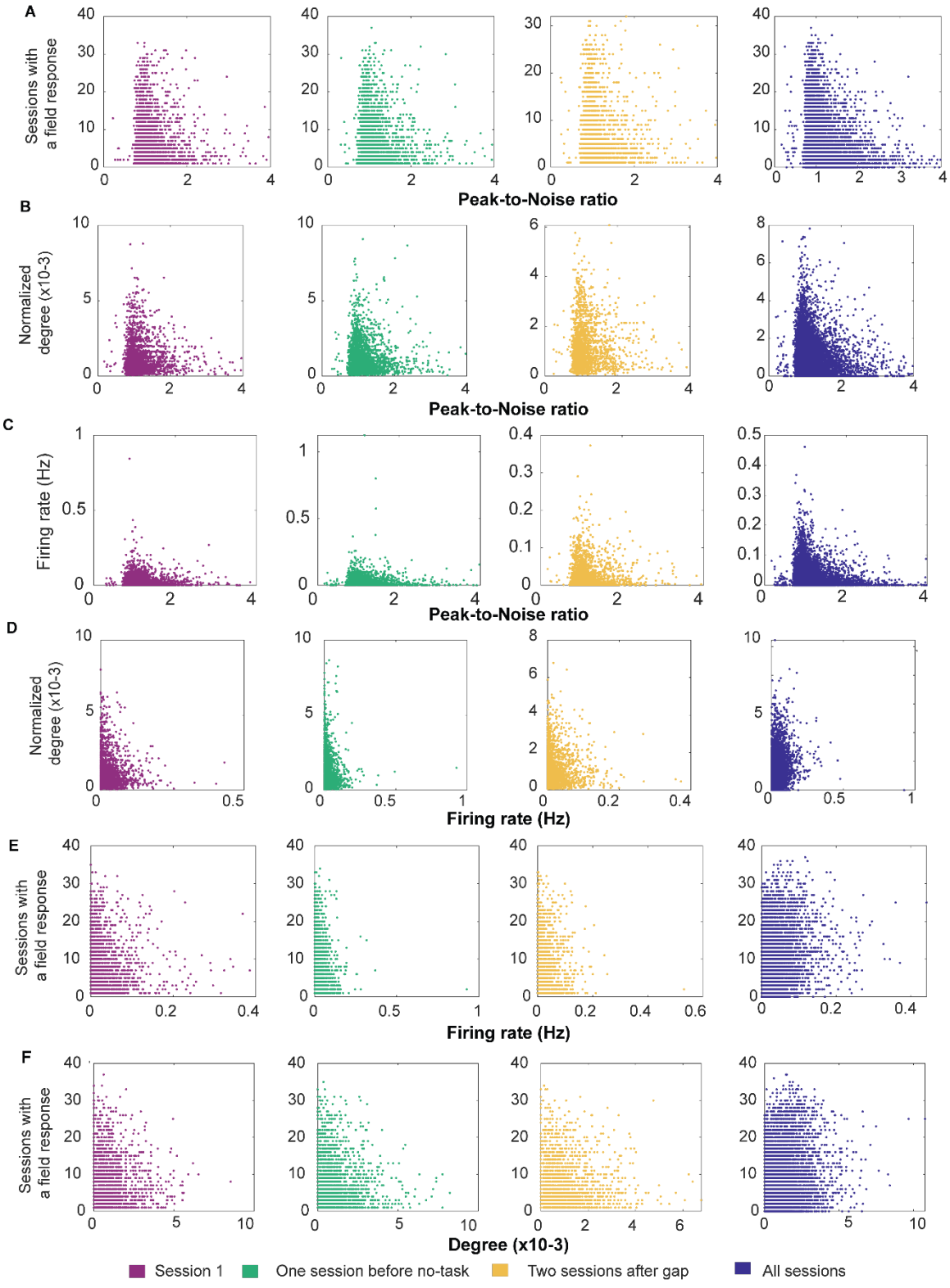
A
Sessions with a field response
Peak-to-Noise ratio
B
Normalized degree (x10-3)
Peak-to-Noise ratio
C
Firing rate (Hz)
Peak-to-Noise ratio
D
Normalized degree (x10-3)
Firing rate (Hz)
E
Sessions with a field response
Firing rate (Hz)
F
Sessions with a field response
Degree (x10-3)
Session 1
One session before no-task
Two sessions after gap
All sessions

Fig. S12.

Place/end cell response field instability is not due to detection problems. We observe a near exponential distribution of peak-to-noise ratios (PNR) (**Fig. S2c)**. However, there is no linear relationship between **(A)** the number of days a neuron was classified as having a field response and its PNR (R-squared <0.04, $p>10^{-10}$). **(B-C)** The degree of a neuron in a graph (R-squared <0.01, $p>10^{-10}$) or the firing rate (R-squared <0.02, $p>10^{-10}$) had also no dependence on PNR. **(D)** There was a small relationship between the degree of a neuron and the firing rate (R-squared <0.15, $p<10^{-10}$). **(E-F)** As a population, The number of days a neuron was classified as having a field response showed a slight relationship to the firing rate (R-squared <0.16, $p>10^{-10}$) and degree (R-squared <0.09, $p<10^{-10}$). Note that figure S12e and S12f are calculated using all neurons, whereas plots in figure 4 and S13 are calculated using only place and end cells. Data from both hemispheres in all mice was pooled together and 10000 random samples were selected for each plot. The scatter plot of data from all sessions was fitted with a linear model to quantify the linear relationship.

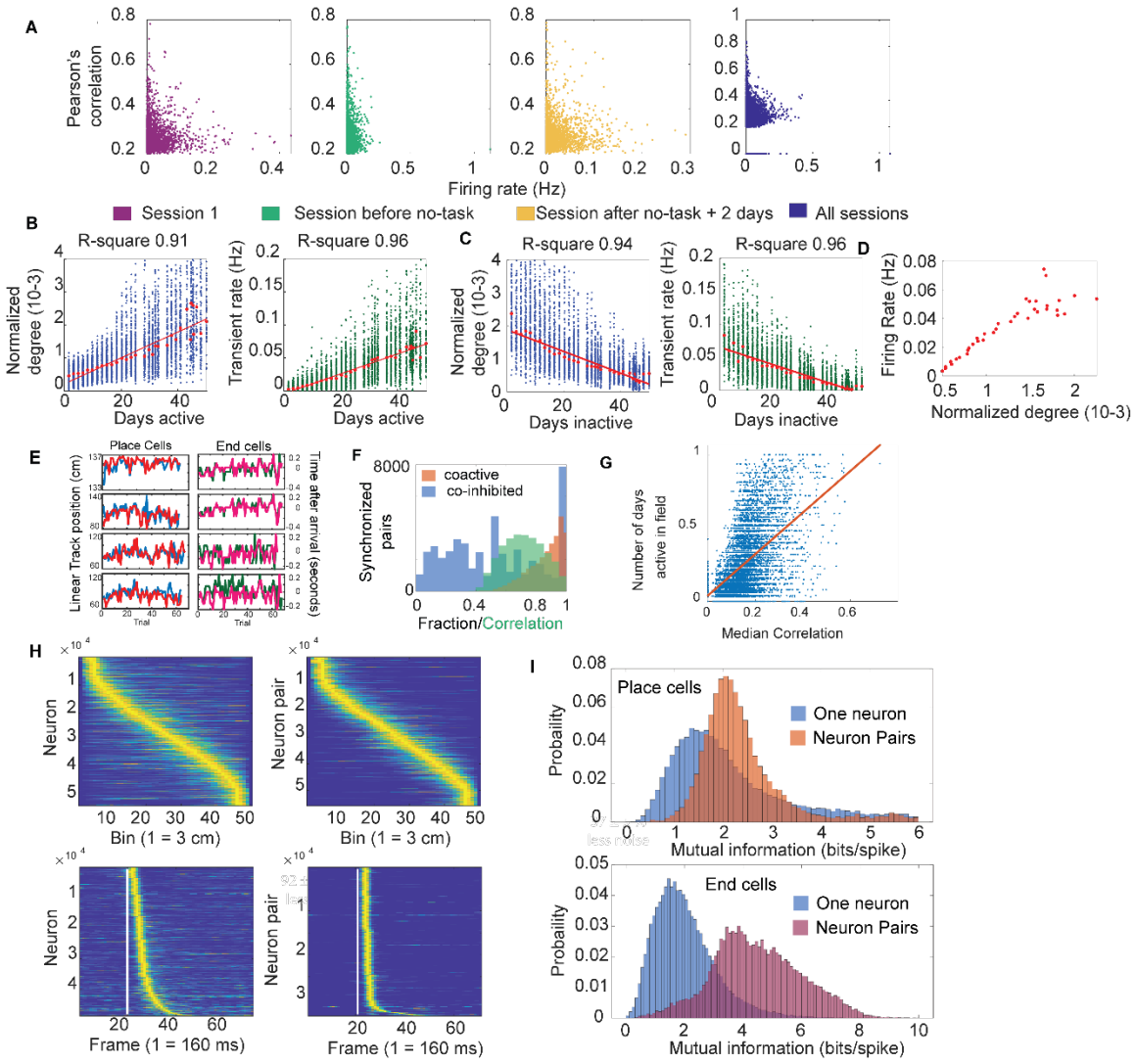

Fig. S13.

Degree is proportional to the firing rate but neuron pairs have higher information content. (A) We observe that the Pearson's correlation between two neurons is independent of the firing rate of each neuron in the pair (how many deconvoluted calcium events per seconds). Indicating that more active neurons do not have higher correlation (R-squared <0.05, $p<10^{-10}$). **(B)** Even though we observe no relationship between how many neuron pairs a neuron has (degree) and its firing rate (Figure S12d). The relationship between degree or firing rate and the number of days a neuron has a field response becomes significant when only considering place/end cells. Indicating that more active place/end cells have more neuron pairs and are more stable ($p<10^{-10}$). **(C)** There is also an inverse relationship between degree or firing rate and the number of days a neuron had a response field. Neurons which were never classified as having a field response had the lowest degree/firing rate. In panels B and C each dot represents an individual place/end cell active for N days, red dots represent the median for place/end cells active N days, and the solid line represent the best fit with a linear model (both

hemispheres, all mice). **(D)** There is a linear relationship between degree and firing rate of place/end cells (both hemispheres, all mice, median, R-squared 0.82, $p<10^{-10}$). **(E)** The correlation between neurons is not by chance since neuron pairs deviate from their field centroid in a correlated manner. Four place-place and end-end neuron pairs shown. The field 'position' of end cells is measured in seconds after or before activation of the reward port. **(F)** Neuron pairs are co-active 90 ± 3 % (orange distribution) of the time and when one does not fire in its field, its neuron pair does not fire 50 ± 20 % of the time (blue distribution). Synchronized pairs also show correlated fluctuations of their fields (green distribution, 0.77 ± 0.17, $p<10^{-8}$). **(G)** The number of sessions either neuron in a pair remains responsive to a field is proportional to their level of synchrony (each dot is a neuron pair, all mice both hemispheres, R-square 0.35, $p< 10^{-4}$). **(H)** Neuron pairs encode information with less noise than individual neurons. The left panels show all the place (top) and end cells (bottom) recorded on all mice and the right panels are the response fields of all neurons when out of sync spikes are removed. The vertical white line in the end cell tuning curves shows activation of the reward port. **(I)** The mutual information of place and end cells increases when removing out of sync spikes. End cell neuron pairs 4.4 ±. 1.5 bits/spike and 1.8 ±. 1.0 bits/spike for individual neurons, $p<10^{-50}$ rank-sum test. Place cell neuron pairs 2.2 ±. 0.9 bits/spike vs individual neurons 1.8 ± 1.3 bits/spike, $p<10^{-50}$ rank-sum test. For calculation of neuron pair information content we deleted every spike that happened out of sync in neuron pairs. Only neuron pairs with significant correlation above 0.2 considered. Then we used the noise free (out of sync signal removed) neuronal activity to calculate place and time fields as described in the methods. Data from both hemispheres on all sessions in the linear track pooled together.

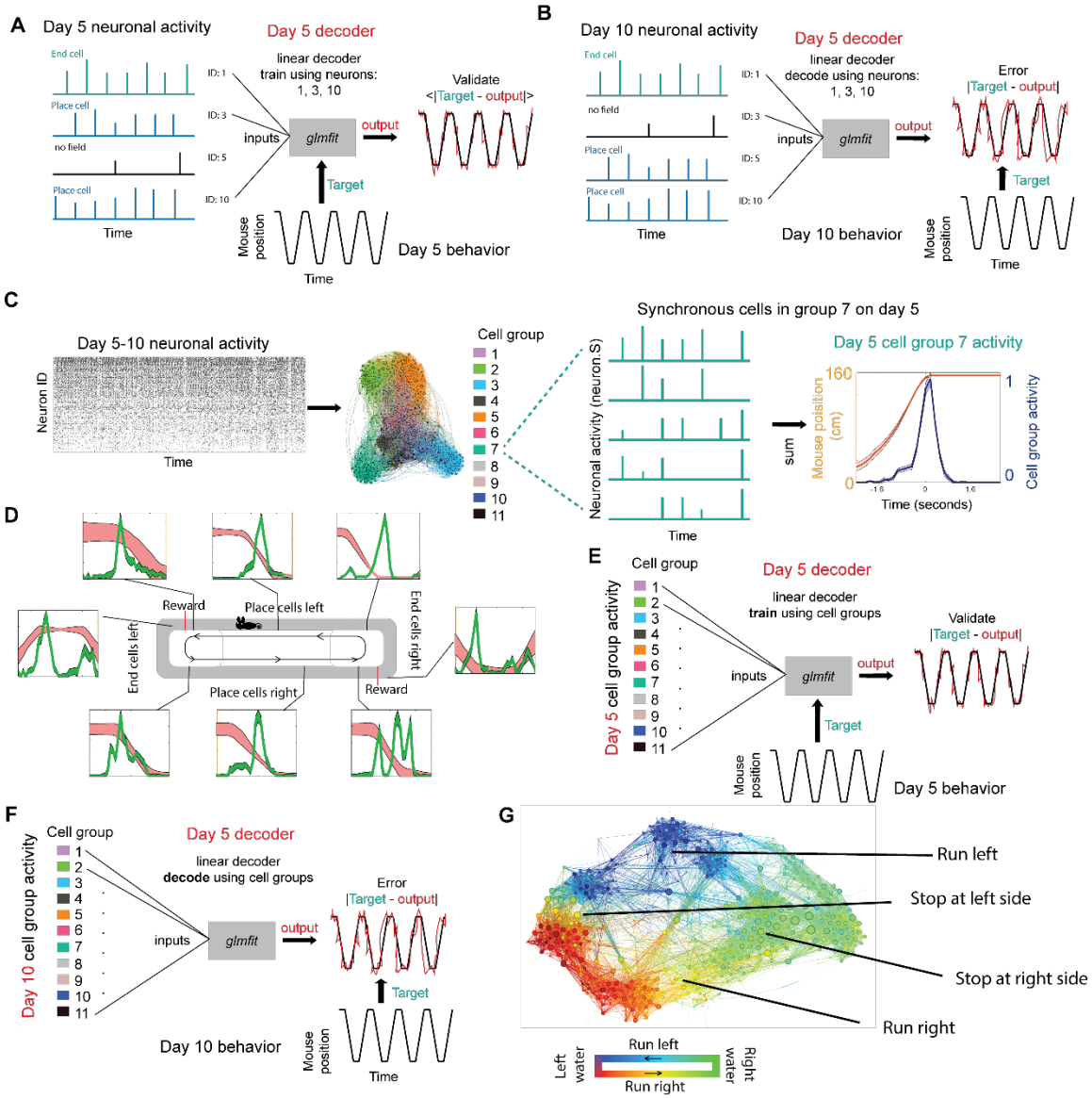

Fig. S14.

Different decoding approaches to estimate the mouse position from brain activity. **(A)** A generalized linear decoder was trained with half the neuronal activity (neuron.S) of all place and end cells identified in one session and the mouse position. The position of the mouse was decoded on the remaining half of the data. **(B)** The decoder trained with one day of data was then tested with the activity of the same neurons on another day. **(C)** Correlated neuronal activity from all neurons in sessions 5-10 were used to build a graph and identify the neurons in each cell group. The activity of neurons in a group in one day was then integrated (blue trace in rightmost panel) and was used as an input into the GLM decoder or the delayed neural network. **(D)** Diagram showing the median activity of each cell group (green lines) along the linear track (position shown in red shadow) in each direction of motion. **(E)** Half the activity of each cell group was used as an input into a generalized linear model with the position of the mouse as the target. Same day decoding was performed with the remaining half of the data. **(F)** Across day decoding

was done with all the cell group data but using a previously trained decoder. **(G)** Decoding of the mouse position using graph topology was performed by first assigning a positional preference to each node in the a graph. This preference was calculated from the median location where each neuron (node) was active during a 20-minute session in the linear track. At each time window the location of the mouse was decoded based on the median positional preference the active neuron and its nearest neighbors in the graph (see supplementary methods).

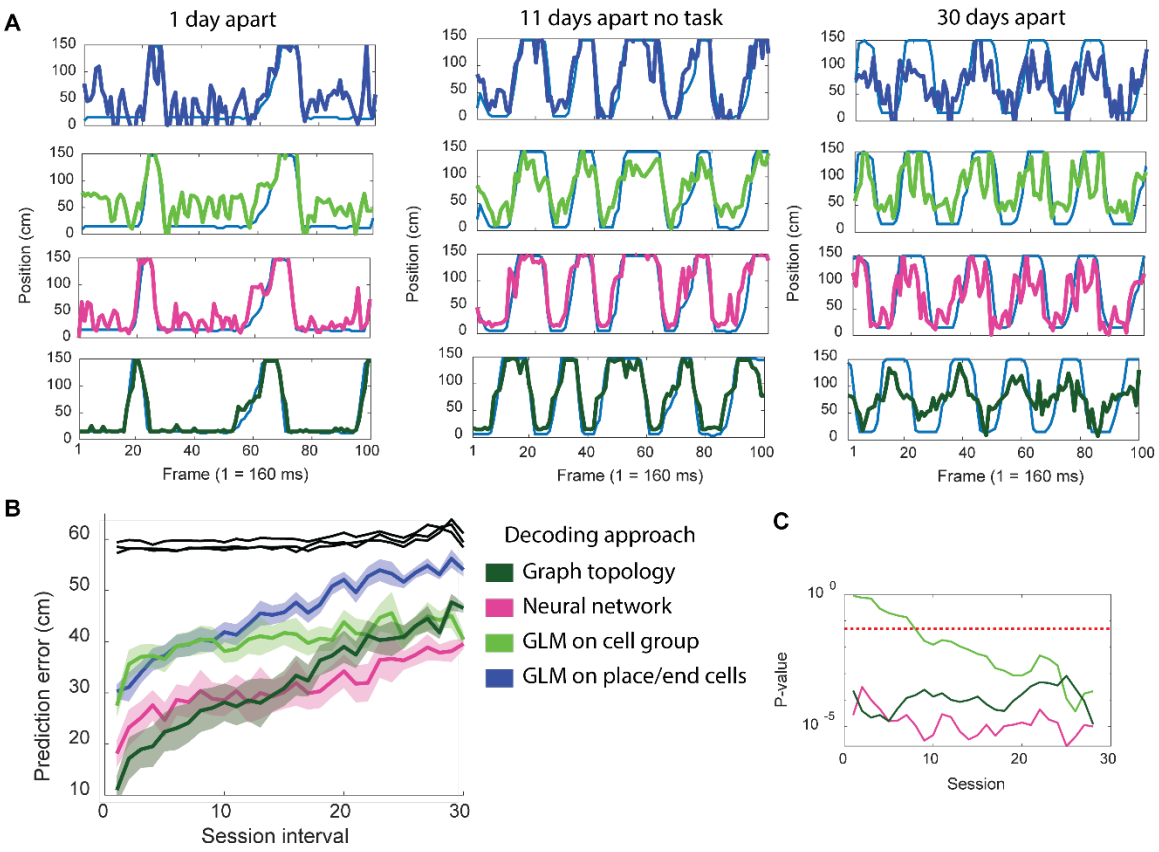

Fig. S15.

Linear and nonlinear decoders using graph information outperform decoders trained with place/end cell activity. (A) (Left) Example decoding of mouse position in the linear track using (from top to bottom) a generalized linear decoder (GLM) trained with place/end cell activity, GLM trained with cell group activity, a time delayed neural network trained with cell group activity, and decoding using graph topology and firing preference of each neuron. All decoders were trained with data from the day before, no behavior information was used other than to validate the results. (Middle) The same as the left panel but using a decoder that was trained with data from the day before the no-task period and tested on the first day after the animal was exposed to the task. (right) Position decoding using a decoder trained with data from the fourth session and tested on the 24th session after a total of 30 days including 10 days that the animal was not exposed to the task. All the data is from the same mouse and right hemisphere. **(B)** Quantification of performance for all decoders as a function of session interval (both hemispheres from all mice analyzed). Lines and shadow represent the median ± SEM, solid black lines are the predictions in which the neuronal activity was randomly shuffled in time. **(C)** Statistical significance of difference between decoding efficiency from place/end cells (blue trace) compared to the other decoders shown in panel B. The p-value was calculated between a session (x-axis) and 2 sessions ahead. Thus, the p-value at session 10 includes data from session 10 to session 12 of the cell group decoder compared to the same range of dates from place/end cells. We observe that cell group decoder outperform

place/end cells after session 10 using the GLM decoder and on all sessions using the other two methods.

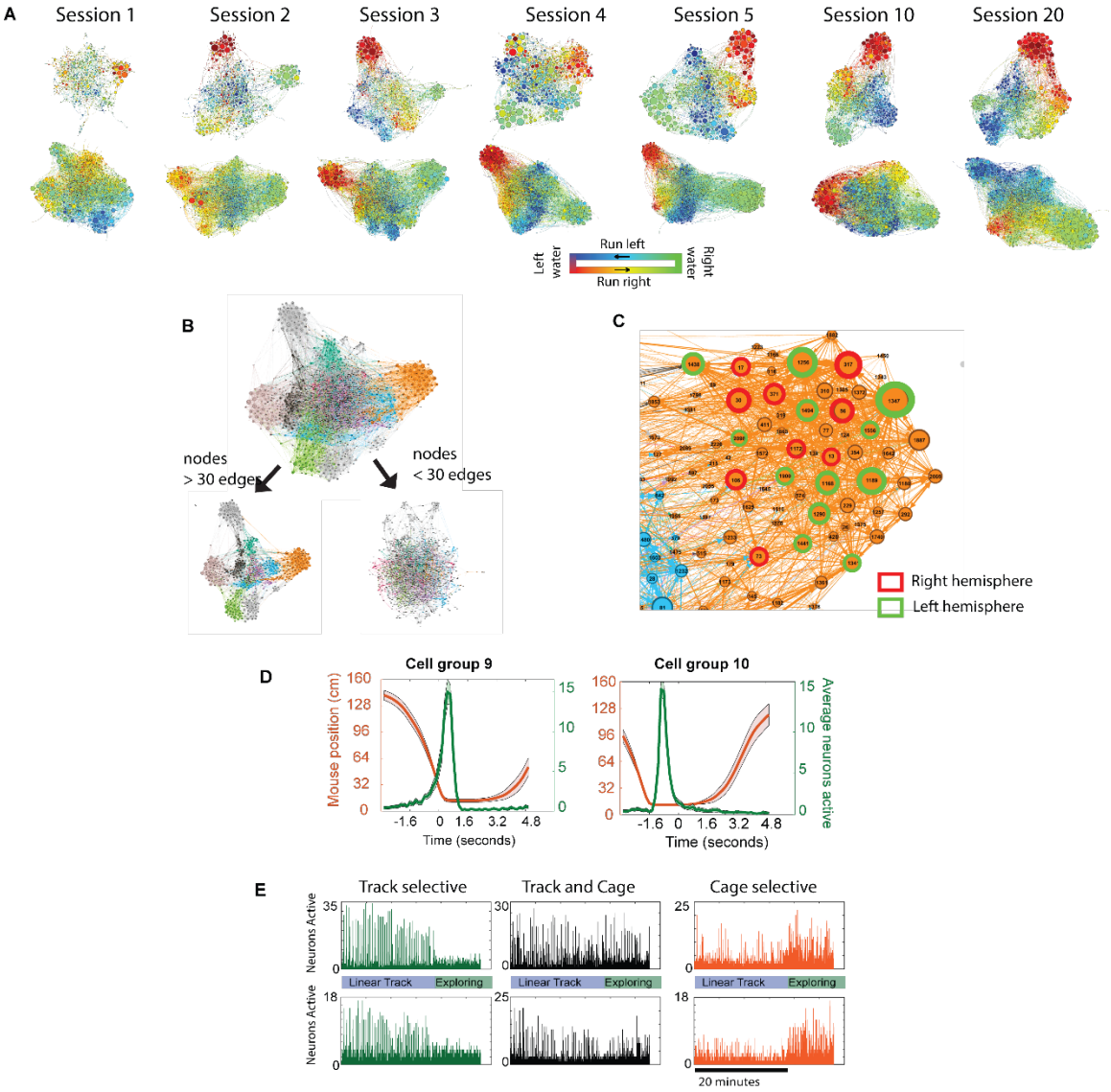

Fig. S16.

Intra- and interhemispheric synchrony in CA1 develop during learning and encodes information about the task. **(A)** Network graph of two mice during the learning phase (first 5 sessions) and after the mice have become habituated to the task (sessions 10 and 20). Note that opposite ends of the linear track (green vs red/blue) are separated by a largest distance, indicating that neurons on either of these ends have little correlated activity, as expected. **(B)** Network graph of neuronal activity in both hemispheres of a mouse running in the linear track. Colors indicate cell groups identified using the Markov diffusion approach. The graph can be decomposed into a graph with small world topology and a random graph by selecting nodes with more than or less than 30 edges, respectively. Thus, highly synchronized neurons form part of a functional network with small world topology spanning both hemispheres. **(C)** Close in view of the cell group in panel B with a subset of neurons from the left hemisphere (green) and right hemisphere (red) shown. **(D)** Two cell groups with place (left) and end of the maze (right) preference. Note the asymmetric shape of the distributions, indicating that as a group, neurons could

use non-linear activity in order to encode position. **(E)** Two cell groups with robust synchronous activity in the linear track but not the home cage (green), two cell groups with synchronous activity both in the home cage and linear track (black), and two cell groups with synchronous activity in the home cage. The number of neurons active represents the sum of all the neurons active (nonzero neuron.S) at each frame (160 ms).

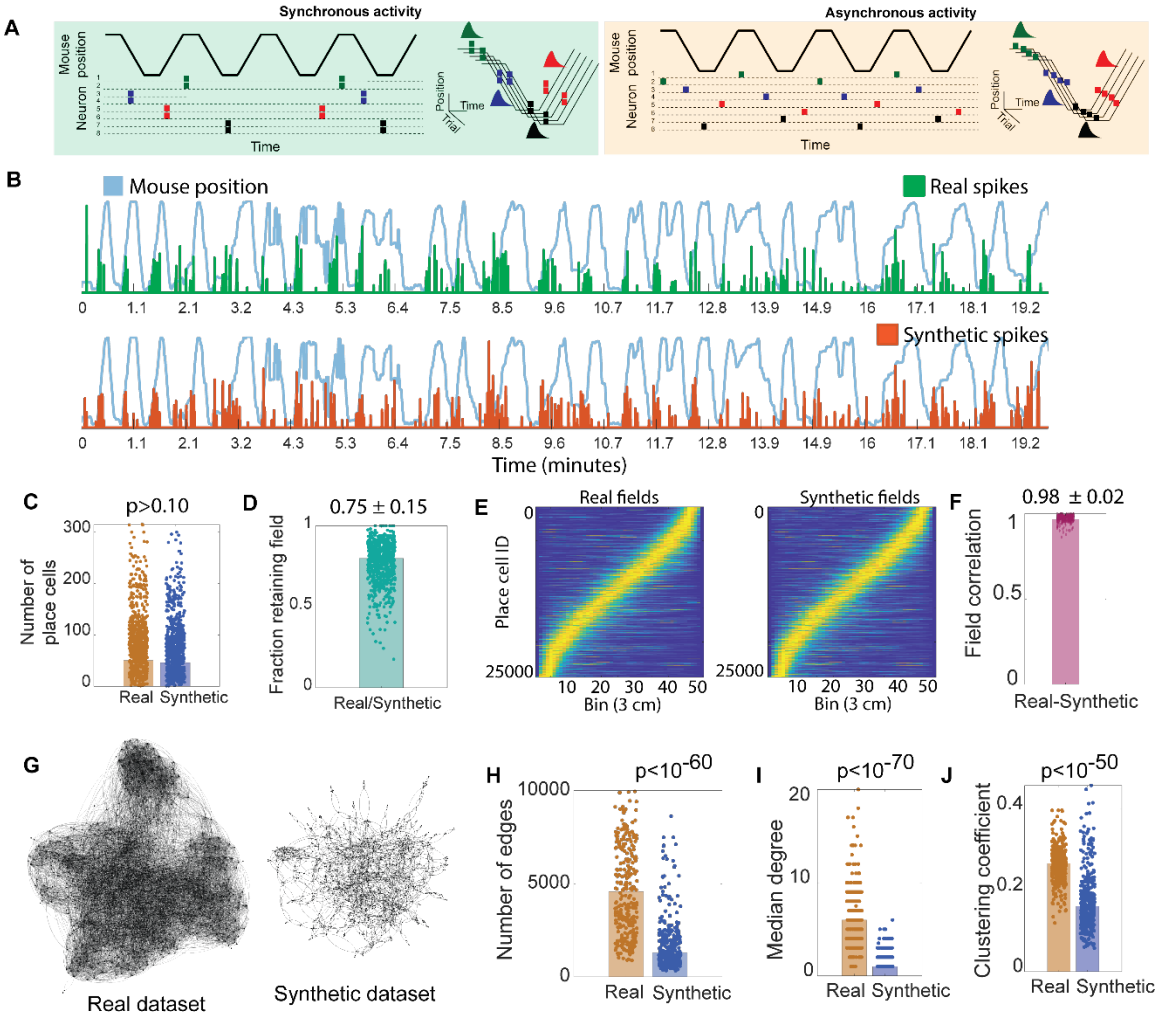

Fig. S17.

Synchrony in CA1 is not due to the overlap of response fields. Neuronal dynamics in CA1 are regulated in such a manner than neurons encoding similar information are more synchronous with other neurons than with neurons encoding different or no information. However, it would be expected that neurons with overlapping response fields could fire simultaneously just by chance. **(A)** To investigate the relationship between place field overlap and synchrony we generated a synthetic dataset by randomly shuffling the temporal information encoded in each spike while preserving the spatial information. This approach is expected to reduce synchronous activity but maintain the response field overlap and firing rates. **(B)** The synthetic dataset shuffled the spikes in a session but preserved the activity with respect to the behavior. (Top) A real neuron firing specifically when the animal runs in one direction and (bottom) the same neuron with its activity shuffled while maintaining the direction and spatial preference but with a random spike temporal profile (mouse position in the maze shown in light blue). **(C)** The number of place cells was not affected by the shuffling approach (median ± sem, 51 ± 66 in the real dataset vs 46 ± 60 in the synthetic dataset, p>0.1, rank-sum test). A total of 348 sessions with place cells for each direction of motion represented by two different points. **(D)** The majority of place cells retained a response field in the same running direction after

shuffling of the spikes (0.75 ± 0.15, n = 696). The small difference likely arises from the smoothing applied to the mouse position data in order to extract direction of motion. **(E)** Response field of place cells in the real and synthetic dataset sorted by IDs of the real dataset. Place fields of all mice in all sessions while the mouse runs in the left direction shown. **(F)** Neurons classified as place cells in the real and synthetic dataset had identical response fields (field correlation 0.99 ± 0.02, n = 695, right and left place cell analyzed separately). **(G)** However, shuffling the spikes resulted in dramatic changes in the neuronal synchrony and graph topology (only the first session in one mouse shown). **(H)** There is an overall decrease in the number synchronous pairs (edges) in the synthetic dataset compared to the real dataset (4580 ± 4314 vs 1294 ± 1637, n=248 sessions, rank-sum test). **(I)** The number of synchronous pairs a neuron has (degree) also significantly decreased (6.0 ± 4.4 vs 1 ± 0.9, n = 348 sessions, rank-sum test). **(J)** The clustering coefficient of the graph also decreased in the synthetic dataset (0.27 ± 0.04 vs. 0.16 ± 0.07, n = 348, rank-sum test). Overall, we observe that synchronous activity of neurons in CA1 cannot be simply explained by response field overlap.

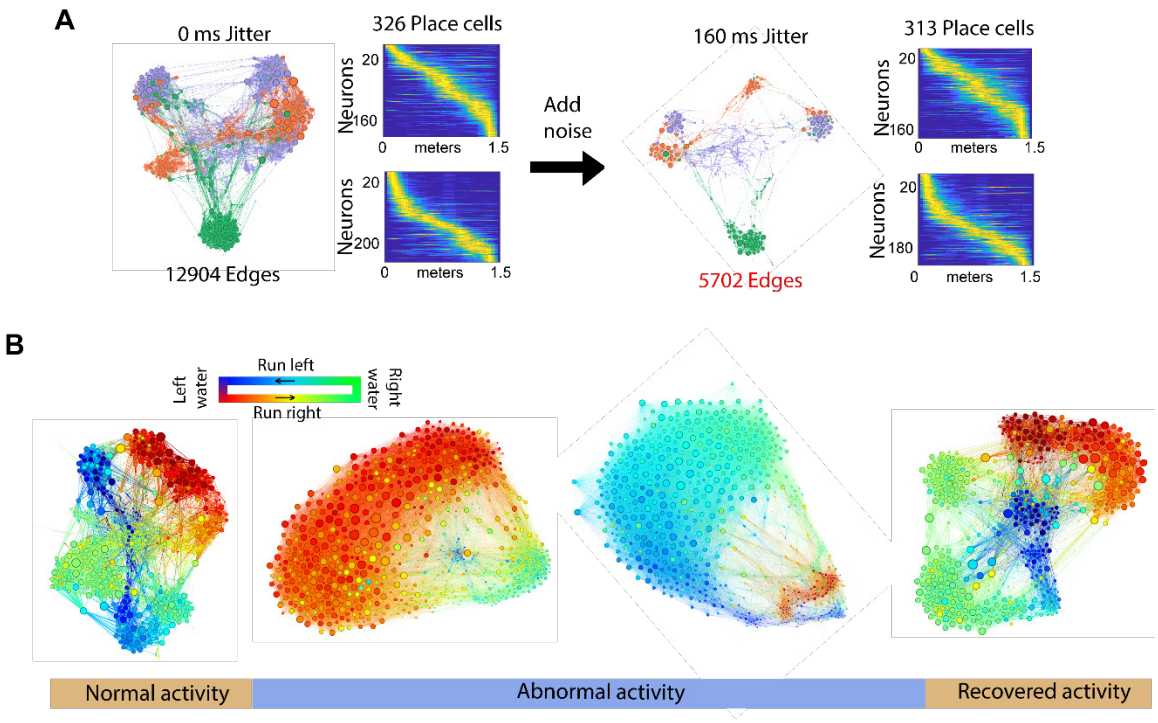

Fig. S18.

Graph topology is sensitive to noise and can capture large cell groups corresponding to direction specific abnormal activity after a lesion in CA1. (A) Introducing a small jitter noise to every deconvoluted neural spike leads to a drastic decrease (55 %) in the number of edges in a graph. However, the same amount of noise does not decrease the number of place cells detected. (B) Network graph during the linear track before lesions to CA1 (left), after lesion and during abnormal activity (middle), and after recovery (right). Note the large change in directional preference of the large cluster during abnormal activity. Data from one mouse shown.

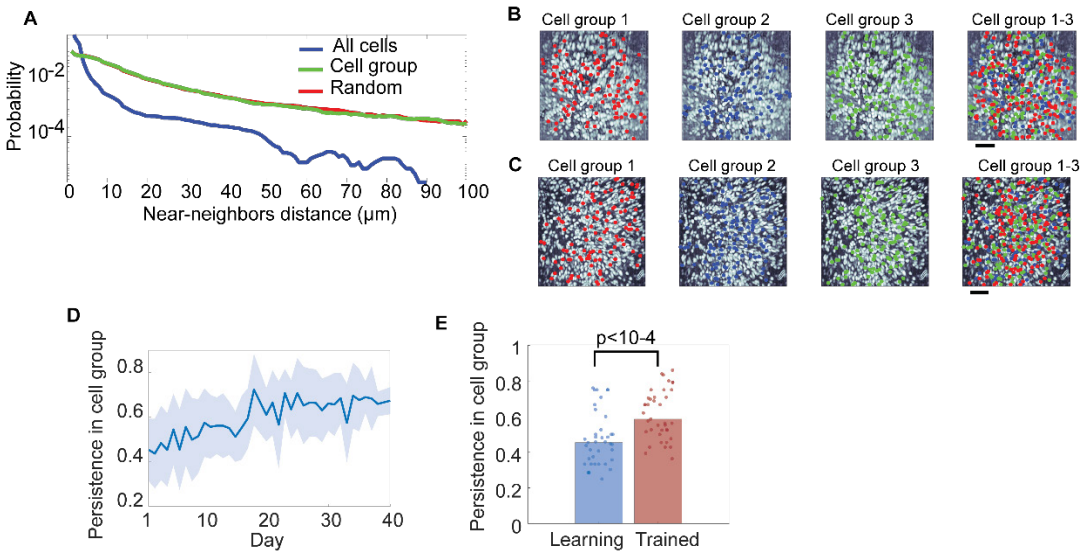

Fig. S19.

Neurons in cell groups are randomly distributed across the imaging field of view and become more stable in their cell group after the learning period. (A) The distance between neurons in a cell group is similar to that between neurons randomly sampled in the field of view ($p>0.7$, Kolmogorov-Smirnov, n = 12709 neurons) but different than the distances observed if the closest neurons are selected ($p<10^{-28}$, Kolmogorov-Smirnov). The blue trace represents the distance between one neuron and its 6 closest neighbors when all neurons in the field of view were sampled. The green trace represents the distance between a neuron in a cell group and its 6 closest neighbors who are also part of the same cell group. The red trace is the distance between one neuron and 6 other neurons in a pool of neurons randomly selected from the field of view. The number of neurons randomly selected is the same as the number of neurons in the cell group. In all cases, the center of mass of the extracted neuron contour was used. Data from both hemispheres from all mice analyzed simultaneously. (B and C) Neurons in cell groups are interspersed across the field of view. Spatial location of neurons (colored dots) composing the three largest cell assemblies in two mice are shown. Scale bar is 100 μm. (D) Neurons in a cell group populate different cell groups during learning but converge to their most stable cell group once the animal has been trained (both hemispheres from all mice combined, line is median and shadow is SD). Initially, we calculated to which cell group a neuron belonged as mentioned in the supplementary methods. Then, on each session we reassigned the cell group participation if a node made more than 50% of edge connections with another cell group. For each node we calculate the fraction of sessions it spent in each cell group across all sessions ("residence"). Lastly, on each session the cell group of a neuron was determined and the residence in that group was assigned to that neuron. Thus, if a neuron populated the same cell group on 9 out of 10 sessions (residence 0.9) it would be assigned a value of 0.9 on 9 sessions and 0.1 on one session. A neuron that spend 5 sessions in one cell group and 5 in another (residence 0.5 and 0.5) would be assigned a value of 0.5 on all 10 sessions. A neuron that spent 1 session in each cell group during 10 sessions (residence 0.1) would be given a value of 0.1 on all 10 sessions. (E) During learning the residence of a neuron in a cell group was lower than

once the mice were trained (3 sessions before the no-task period, 0.46 ± 0.15 vs 0.59 ± 0.14, rank-sum test, n = 39 sessions).

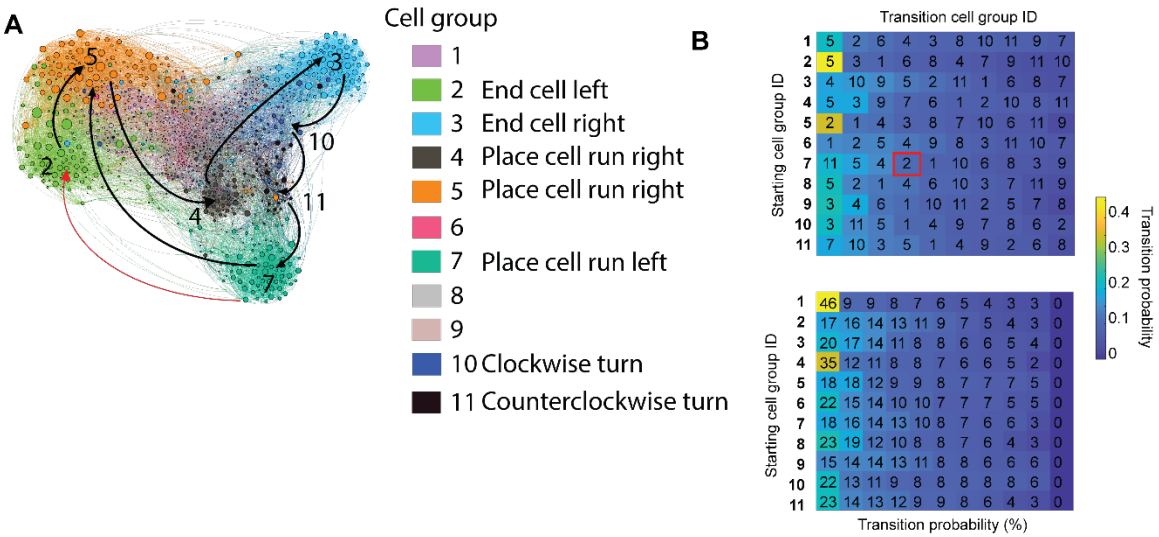

Fig. S20.

Node connectivity in a graph provides insight into alternating sequential behaviors. **(A)** Network graph of neuronal activity of one mouse in one session. Nodes are colored by the cell group they belong to, the size of the node is proportional to its degree. Arrows indicate a sequence of behaviors identified using the transition probability shown on the right panel. **(B)** The color in each box represents the transition probability of one cell group into another (cell group shown with numbers), the probabilities are sorted so the highest is on the left. The transition probability was calculated by counting how many edges a node in a cell group makes with nodes in another cell group. Edges within the cell group are not counted. The numbers of edges were then divided by all the edges made by that cell group, excluding self-connections. The bottom panel shows the actual probabilities shown as a percent. The arrows in panel A were obtained in the following manner. Start at cell group 2, second row on right panel. From group 2 the next most likely transition is to cell group 5. From cell group 5, the most likely transition (ignoring loop propagation, and non-place/time cell group 1) is to cell group 4. From cell group four the next transition is 3, then cell group 10 to cell group 11 and to cell group 7. From cell group 7 the next most likely transition is cell group 5, which is classified as place cells in the right direction. This is an unlikely transition since before the animal runs right it must stop and drink. So, the only transition missing to complete the cycle is from cell group 7 to 2, which is the fourth transition (shown in red arrow and rectangle). There are many explanations for the lack of a complete cycle, including the fact that end and place cells in the same direction show little correlation due to their asymmetric population fields **(Figure S16d)**. A different pathway could be possible if more than 11 cell groups would be identified using a faster Markov time (Figure S19F). Interestingly, the transition from end cells on the right (cell group 3, light blue) to place cells in the left direction (cell group 7, green) has two intermediate cell groups (11 and 10). These two cell groups correspond to when the animal turns clockwise or counterclockwise (**Movie 9**).

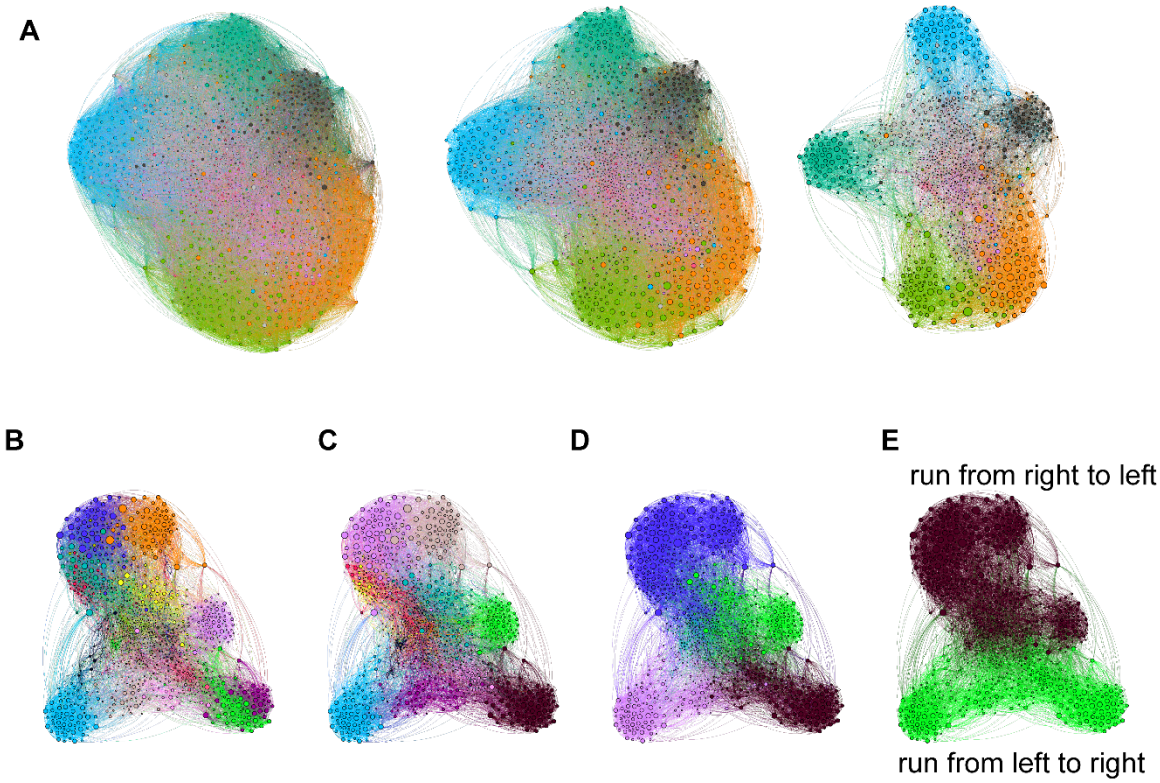

Fig. S21.

Selecting different correlation threshold or Markov diffusion time does not change the overall results presented and may provide finer detail into mouse behavior. (A) In our analysis we only look at neurons whose correlation is above 0.1 with a p value below 0.05 (two-sided t-test). Selecting different thresholds results in higher or lower number of edges (left, threshold 0.0 results in ~ 75000 edges; middle, threshold of 0.05 results in 35000 edges; right, a threshold of 0.1 results in 13700 edges). The number of nodes was mostly unchanged by the threshold. We selected a threshold of 0.1 because it provided a manageable amount of data and required a reasonable amount of time to process. However, the number of cell groups was only slightly changed (10 at correlations above 0, 11 at correlation above 0.05, and 14 at correlations above 0.1). We limited our analysis to 11 mainly because we cannot assign any behavioral properties to modules beyond 6 or 7. A second more important parameter is the Markov diffusion time. This parameter is proportional to how many random walks would be required for a walker to diffuse out of a cell group. (C) Short timescales split the graph into several groups which may correspond to finer variations of the mouse behavior in the track (time set to 0.3). Notice the time cell group being split into two (bottom right, pink and green). At timescales of (D) 0.7 and (E) 1.0 we start to observe a number of cell groups corresponding to direction specific place and time cells. (F) We noticed that longer Markov times split the graph in two groups comprising each direction in the track (time set at 5). Color randomly assigned between panels C to F.

Table S1.

Comparing stability across hemispheres of place and end cells response fields during different phases of the task.

	Both Hemispheres			RH			LH				MES**	
	Directional stability	SD	n	Directional stability	SD	n	Directional stability	SD	n	p-value*	Cohen's d	Cohen's U3
Learning	0.19	0.14	2335	0.17	0.15	1442	0.20	0.13	893	0.008	0.06	0.53
Trained	0.36	0.22	1360	0.31	0.23	589	0.38	0.21	771	0.1231	0.12	0.57
Re-exposed (2 days)	0.30	0.17	1052	0.33	0.17	641	0.21	0.18	411	0.0004	**0.26**	**0.61**
Re-exposed (3+ days)	0.42	0.21	2053	0.44	0.21	1341	0.37	0.21	712	0.0006	0.17	0.59
Trained/Re-exposed	0.28	0.15	2316	0.30	0.14	1365	0.26	0.15	951	0.037	0.10	0.58
	Correlation	SD	n	Correlation	SD	n	Correlation	SD	n	p-value*	Cohen D	Cohen U3
Learning	0.37	0.31	2348	0.37	0.32	1455	0.38	0.29	893	0.31	0.05	0.52
Trained	0.62	0.35	1360	0.72	0.37	589	0.57	0.34	771	1.7E-5	0.14	0.64
Re-exposed (2 days)	0.64	0.36	1052	0.69	0.36	641	0.57	0.35	411	5.0E-5	**0.21**	**0.63**
Re-exposed (3+ days)	0.79	0.34	2053	0.80	0.37	1341	0.79	0.25	712	0.38	0.22	0.53
Trained/Re-exposed	0.68	0.33	2316	0.69	0.33	1365	0.66	0.33	951	1.4E-5	0.12	0.55
	Similarity	SD	n	Similarity	SD	n	Similarity	SD	n	p-value*	Cohen D	Cohen U3
Learning	0.10	0.32	1919	0.11	0.33	1176	0.09	0.31	743	0.30	0.01	0.55
Trained	0.31	0.33	1223	0.36	0.35	518	0.27	0.31	705	2.6E-6	0.19	0.64
Re-exposed (2 days)	0.27	0.31	956	0.30	0.29	592	0.24	0.33	364	3.4E-4	**0.26**	**0.58**
Re-exposed (3+ days)	0.40	0.32	1930	0.41	0.34	1266	0.38	0.27	664	0.25	0.09	0.57
Trained/Re-exposed	0.30	0.31	2070	0.31	0.29	1229	0.27	0.31	841	2.9E-4	0.14	0.55

Statistical tests shown between hemispheres. *Wilcoxson rank-sum test. ** Measure of effect size (MES) is shown using the Cohen's d parameter defined as |mean(X)-mean(X2)|./mean(SD(X)+SD(X2)), where X and X2 represent data from each hemisphere. Cohen's d <0.2 is considered non-significant, >0.2 is a small effect, and >0.5 is a strong effect. Cohen's U3 is a nonparametric approach that measures the proportion of data points in the lower group which are smaller than the median of the higher group. Cohen U3 <0.55 is considered non-significant, 0.55<U3<0.65 is a small effect, and >0.65 is a strong effect. Division by zero lead to unequal n values.

Table S2.

Hypothesis testing and measure of effect size of place/end cells stability between different periods of the task.

Wilcoxson rank-sum test

	learning	Trained	Re-exposed (2 days)	Re-exposed (3+ days)	Trained/re-exposed
learning	1	1.56E-70	7.4E-51	2.8E-199	1.4E-99
Trained	1.6E-70	1	0.04	5.2E-21	0.03
Re-exposed (2 days)	7.4E-51	0.004	1	9.9E-38	0.16
Re-exposed (3+ days)	2.8E-199	5.2E-21	9.9E-38	1	1.4E-52
Recovered	1.4E-99	0.03	0.16	1.4E-52	1

Cohen's d measure of effect size

	learning	Trained	Re-exposed (2 days)	Re-exposed (3+ days)	Trained/re-exposed
learning	0.00	0.62	0.54	0.91	0.60
Trained	0.62	0.00	0.09	0.29	0.04
Re-exposed (2 days)	0.54	0.09	0.00	0.39	0.05
Re-exposed (3+ days)	0.91	0.29	0.39	0.00	0.34
Recovered	0.60	0.04	0.05	0.34	0.00

Cohen's U3 measure of effect size

	learning	Trained	Re-exposed (2 days)	Re-exposed (3+ days)	Trained/re-exposed
learning	0.50	0.81	0.77	0.91	0.80
Trained	0.81	0.50	0.55	0.65	0.53
Re-exposed (2 days)	0.77	0.55	0.50	0.73	0.53
Re-exposed (3+ days)	0.91	0.65	0.73	0.50	0.74
Recovered	0.80	0.53	0.53	0.74	0.50

Recovered is the overlap between the trained and re-exposed periods

Table S3.

Values shown in Fig. 3E and Fig. S11F.

	Directional stability	SD	n
Re-exposed	0.42	0.20	342
Damaged	0.13	0.13	574
Recovered	0.47	0.19	128
Recall	0.20	0.10	128
	Correlation	SD	n
Re-exposed	0.83	0.36	340
Damaged	0.44	0.37	558
Recovered	0.83	0.29	128
Recall	0.61	0.34	126
	Similarity	SD	n
Re-exposed	0.50	0.28	340
Damaged	0.06	0.25	558
Recovered	0.41	0.25	128
Recall	0.17	0.22	126

Division by zero leads to small differences in n. Right hemisphere of 3 mice. Recall is the similarity between the re-exposed and recovered periods.

Table S4.

Hypothesis testing and measure of effect size for similarity during recovery and damage.

Wilcoxson's rank-sum test p-value

	Trained	Damaged	Recovered	Recall
Re-exposed	1	4.38E-52	0.21133	5.98E-18
Damaged	4.38E-52	1	9.22E-30	3.09E-05
Recovered	0.21133	9.22E-30	1	3.31E-14
Recall	5.98E-18	3.09E-05	3.31E-14	1

Cohen's d measure of effect size

	Trained	Damaged	Recovered	Recall
Re-exposed	0.00	1.21	0.05	0.94
Damaged	1.21	0.00	1.24	0.36
Recovered	0.05	1.24	0.00	0.96
Recall	0.94	0.36	0.96	0.00

Cohen's U3 measure of effect size

	Trained	Damaged	Recovered	Recall
Re-exposed	0.50	0.94	0.59	0.94
Damaged	0.94	0.50	0.91	0.64
Recovered	0.59	0.91	0.50	0.90
Recall	0.94	0.64	0.90	0.50

See table S1 and methods for definitions. Recall is the similarity between the re-exposed and recovered periods.

Movie S1

(Top row) Unprocessed 15 second recording of the left hemisphere in one animal and right hemisphere in another while running in the linear track. (Bottom row) Temporally down-sampled (4-fold) and spatially smoothed 15 second video of CA1 activity while two different mice run in the linear track. Left hemisphere is shown on the left panel and right hemisphere on the right.

Movie S2

Simultaneous bilateral recording of a mouse while running in the linear track. Image is background subtracted. Left hemisphere shown on the left and right hemisphere on the right. Neuronal activity extracted from this animal is shown on figure 1e.

Movie S3

(Left) Correlation image of CA1 activity across two months not motion corrected, each frame represents a 30-minute session (home cage exploration and linear track), 45 sessions in total. (right) The same data but motion corrected. The sudden shift between session 14 and 15 is due to the 10-day period of no task.

Movie S4

Motion corrected correlation image of CA1 activity in one mouse recorded for 8 months (76 sessions, data from home cage recordings and linear track). Sessions between 33 and 42 not included due to abnormal activity due to damage prevent accurate registration using individual sessions. The data presented in the manuscript is motion corrected and analyzed in a different manner (see methods).

Movie S5

Direction specific burst activity in CA1 in three mice after damage. Note the continuously fluorescent cells, these cells will eventually become inactive. All three mice have a leftward direction burst in the video, but the direction could change between days (see figure S11 and S13).

Movie S6

The same data used in movie 5 but at a closer view showing how some neurons persist in the field of view while other stop being active after the CA1 lesion.

Movie S7

(Top left) Changes in the topology of a network graph of CA1 activity in a mouse in its home cage across days. Note the small cluster of nodes leaving and joining the larger cluster. (Top right) Changes in the topology of a network graph of CA1 activity while the mouse becomes familiar with the linear track. (Bottom) The same data but nodes are colored by cell groups determined on the last day once the mouse is familiar with the environment. The video proceeds in reverse order, trained periods are shown in the first frames and learning are the last. Note that modules retain a large portion of neurons across days but eventually fall apart during learning. Also, the diffusion of colors indicates that some neurons can change their role in the network.

Movie S8

The number of edges between clusters can be used to extract the sequence of events in the behavior (see figure S21). In this video, we show that a sequential behavior has a clear sequential activity in the graph. Cell groups are shown in color once at least 5 neurons in the group are active, only on cell group per frame shown.

Movie S9

In some cases, we observe that sequences in a graph can bifurcate, activating different groups. In graph shown in this video we have at the right side of the maze four cell groups: cell group 3 shown in light blue (time cells), cell group 10 in dark blue, 11 in black, and 7 in green (place cells). Using the graph connectivity, we observe that the pathway 3-10-7 and 3-11-7 have very similar transition probabilities. Here we show 20 frame segments of behavior centered at when either cell group 10 (left video) or cell group 11 (right) are active. We observe that each of these sequences correspond to turning counterclockwise or clockwise.

Additional Data (*29*)

Sample raw data and processed data is available at (https://doi.org/10.22002/d1.1229). Custom MATLAB scripts, fusion360 CAD files for the design of the custom microendoscope, and CNMFe parameters used are available at (https://doi.org/10.22002/d1.1229).

References and Notes

1. W. B. Scoville, B. Milner, Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957). doi:10.1136/jnnp.20.1.11 Medline

2. N. J. Cohen, L. R. Squire, Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science* **210**, 207–210 (1980). doi:10.1126/science.7414331 Medline

3. J. O'Keefe, J. Dostrovsky, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971). doi:10.1016/0006-8993(71)90358-1 Medline

4. A. Czurkó, H. Hirase, J. Csicsvari, G. Buzsáki, Sustained activation of hippocampal pyramidal cells by 'space clamping' in a running wheel. *Eur. J. Neurosci.* **11**, 344–352 (1999). doi:10.1046/j.1460-9568.1999.00446.x Medline

5. D. Aronov, R. Nevers, D. W. Tank, Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* **543**, 719–722 (2017). doi:10.1038/nature21692 Medline

6. M. R. Mehta, C. A. Barnes, B. L. McNaughton, Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8918–8921 (1997). doi:10.1073/pnas.94.16.8918 Medline

7. M. A. Wilson, B. L. McNaughton, Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993). doi:10.1126/science.8351520 Medline

8. J. Debiec, J. E. LeDoux, K. Nader, Cellular and systems reconsolidation in the hippocampus. *Neuron* **36**, 527–538 (2002). doi:10.1016/S0896-6273(02)01001-2 Medline

9. S. M. Zola-Morgan, L. R. Squire, The primate hippocampal formation: Evidence for a time-limited role in memory storage. *Science* **250**, 288–290 (1990). doi:10.1126/science.2218534 Medline

10. Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. El Gamal, M. J. Schnitzer, Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266 (2013). doi:10.1038/nn.3329 Medline

11. J. D. Zaremba, A. Diamantopoulou, N. B. Danielson, A. D. Grosmark, P. W. Kaifosh, J. C. Bowler, Z. Liao, F. T. Sparks, J. A. Gogos, A. Losonczy, Impaired hippocampal place cell dynamics in a mouse model of the 22q11.2 deletion. *Nat. Neurosci.* **20**, 1612–1623 (2017). doi:10.1038/nn.4634 Medline

12. T. Hainmueller, M. Bartos, Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* **558**, 292–296 (2018). doi:10.1038/s41586-018-0191-2 Medline

13. C. G. Kentros, N. T. Agnihotri, S. Streater, R. D. Hawkins, E. R. Kandel, Increased attention to spatial context increases both place field stability and spatial memory. *Neuron* **42**, 283–295 (2004). doi:10.1016/S0896-6273(04)00192-8 Medline

14. E. A. Mankin, G. W. Diehl, F. T. Sparks, S. Leutgeb, J. K. Leutgeb, Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron* **85**, 190–201 (2015). doi:10.1016/j.neuron.2014.12.001 Medline

15. K. Z. Tanaka, A. Pevzner, A. B. Hamidi, Y. Nakazawa, J. Graham, B. J. Wiltgen, Cortical representations are reinstated by the hippocampus during memory retrieval. *Neuron* **84**, 347–354 (2014). doi:10.1016/j.neuron.2014.09.037 Medline

16. H. A. Zariwala, B. G. Borghuis, T. M. Hoogland, L. Madisen, L. Tian, C. I. De Zeeuw, H. Zeng, L. L. Looger, K. Svoboda, T.-W. Chen, A Cre-dependent GCaMP3 reporter mouse for neuronal imaging in vivo. *J. Neurosci.* **32**, 3131–3141 (2012). doi:10.1523/JNEUROSCI.4469-11.2012 Medline

17. D. J. Cai, D. Aharoni, T. Shuman, J. Shobe, J. Biane, W. Song, B. Wei, M. Veshkini, M. La-Vu, J. Lou, S. E. Flores, I. Kim, Y. Sano, M. Zhou, K. Baumgaertel, A. Lavi, M. Kamata, M. Tuszynski, M. Mayford, P. Golshani, A. J. Silva, A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534**, 115–118 (2016). doi:10.1038/nature17955 Medline

18. H. Dana, T.-W. Chen, A. Hu, B. C. Shields, C. Guo, L. L. Looger, D. S. Kim, K. Svoboda, Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo. *PLOS ONE* **9**, e108697 (2014). doi:10.1371/journal.pone.0108697 Medline

19. K. K. Ghosh, L. D. Burns, E. D. Cocker, A. Nimmerjahn, Y. Ziv, A. E. Gamal, M. J. Schnitzer, Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011). doi:10.1038/nmeth.1694 Medline

20. C. J. MacDonald, K. Q. Lepage, U. T. Eden, H. Eichenbaum, Hippocampal “time cells” bridge the gap in memory for discontiguous events. *Neuron* **71**, 737–749 (2011). doi:10.1016/j.neuron.2011.07.012 Medline

21. K. Kay, M. Sosa, J. E. Chung, M. P. Karlsson, M. C. Larkin, L. M. Frank, A hippocampal network for spatial coding during immobility and sleep. *Nature* **531**, 185–190 (2016). doi:10.1038/nature17144 Medline

22. R. Boehringer, D. Polygalov, A. J. Y. Huang, S. J. Middleton, V. Robert, M. E. Wintzer, R. A. Piskorowski, V. Chevaleyre, T. J. McHugh, Chronic Loss of CA2 Transmission Leads to Hippocampal Hyperexcitability. *Neuron* **94**, 642–655.e9 (2017). doi:10.1016/j.neuron.2017.04.014 Medline

23. D. O. Hebb, *The Organization of Behavior* (Wiley, 1949).

24. K. D. Harris, J. Csicsvari, H. Hirase, G. Dragoi, G. Buzsáki, Organization of cell assemblies in the hippocampus. *Nature* **424**, 552–556 (2003). doi:10.1038/nature01834 Medline

25. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsáki, Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008). doi:10.1126/science.1159775 Medline

26. Y. N. Billeh, M. T. Schaub, C. A. Anastassiou, M. Barahona, C. Koch, Revealing cell assemblies at multiple levels of granularity. *J. Neurosci. Methods* **236**, 92–106 (2014). doi:10.1016/j.jneumeth.2014.08.011 Medline

27. K. Z. Tanaka, H. He, A. Tomar, K. Niisato, A. J. Y. Huang, T. J. McHugh, The hippocampal engram maps experience but not place. *Science* **361**, 392–397 (2018). doi:10.1126/science.aat5397 Medline

28. T. Kitamura, S. K. Ogawa, D. S. Roy, T. Okuyama, M. D. Morrissey, L. M. Smith, R. L. Redondo, S. Tonegawa, Engrams and circuits crucial for systems consolidation of a memory. *Science* **356**, 73–78 (2017). doi:10.1126/science.aam6808 Medline

29. W. Gonzalez, Persistence of neuronal representations through time and damage in the hippocampus, Version 1.0, CaltechDATA (2019); https://doi.org/10.22002/d1.1229.

30. N. R. Kinsky, D. W. Sullivan, W. Mau, M. E. Hasselmo, H. B. Eichenbaum, Hippocampal Place Fields Maintain a Coherent and Flexible Map across Long Timescales. *Curr. Biol.* **28**, 3578–3588.e6 (2018). doi:10.1016/j.cub.2018.09.037 Medline

31. Y. Ben-Shaul, OptiMouse: A comprehensive open source program for reliable detection and analysis of mouse body and nose positions. *BMC Biol.* **15**, 41 (2017). doi:10.1186/s12915-017-0377-3 Medline

32. W. A. Liberti III, L. N. Perkins, D. P. Leman, T. J. Gardner, An open source, wireless capable miniature microscope system. *J. Neural Eng.* **14**, 045001 (2017). doi:10.1088/1741-2552/aa6806 Medline

33. E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T. M. Jessell, D. S. Peterka, R. Yuste, L. Paninski, Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* **89**, 285–299 (2016). doi:10.1016/j.neuron.2015.11.037 Medline

34. P. Zhou, S. L. Resendez, J. Rodriguez-Romaguera, J. C. Jimenez, S. Q. Neufeld, A. Giovannucci, J. Friedrich, E. A. Pnevmatikakis, G. D. Stuber, R. Hen, M. A. Kheirbek, B. L. Sabatini, R. E. Kass, L. Paninski, Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *eLife* **7**, e28728 (2018). doi:10.7554/eLife.28728 Medline

35. W. E. Skaggs, B. L. McNaughton, K. M. Gothard, in *Advances in Neural Information Processing Systems 5*, S. J. Hanson, J. D. Cowan, C. L. Giles, Eds. (Morgan-Kaufmann, 1993), pp. 1030–1037.

36. L. Sheintuch, A. Rubin, N. Brande-Eilat, N. Geva, N. Sadeh, O. Pinchasof, Y. Ziv, Tracking the Same Neurons across Multiple Days in Ca^{2+} Imaging Data. *Cell Rep.* **21**, 1102–1115 (2017). doi:10.1016/j.celrep.2017.10.013 Medline