ARTICLES
https://doi.org/10.1038/s41593-022-01086-6

nature neuroscience

Learning binds new inputs into functional synaptic clusters via spinogenesis

Nathan G. Hedrick[1,2,3,4] ✉, Zhongmin Lu[1,2,3,4], Eric Bushong[3,5,6], Surbhi Singhi[1,2,3,4], Peter Nguyen[1,2,3,4], Yessenia Magaña[1,2,3,4], Sayyed Jilani[1,2,3,4], Byung Kook Lim[1,2], Mark Ellisman[3,5,6] and Takaki Komiyama[1,2,3,4] ✉

Learning induces the formation of new excitatory synapses in the form of dendritic spines, but their functional properties remain unknown. Here, using longitudinal in vivo two-photon imaging and correlated electron microscopy of dendritic spines in the motor cortex of mice during motor learning, we describe a framework for the formation, survival and resulting function of new, learning-related spines. Specifically, our data indicate that the formation of new spines during learning is guided by the potentiation of functionally clustered preexisting spines exhibiting task-related activity during earlier sessions of learning. We present evidence that this clustered potentiation induces the local outgrowth of multiple filopodia from the nearby dendrite, locally sampling the adjacent neuropil for potential axonal partners, likely via targeting preexisting presynaptic boutons. Successful connections are then selected for survival based on co-activity with nearby task-related spines, ensuring that the new spine preserves functional clustering. The resulting locally coherent activity of new spines signals the learned movement. Furthermore, we found that a majority of new spines synapse with axons previously unrepresented in these dendritic domains. Thus, learning involves the binding of new information streams into functional synaptic clusters to subserve learned behaviors.

Learning is known to involve the restructuring of neural circuits to facilitate changes in brain functions[1]. Among the putative substrates of such circuit-level changes is the formation of new dendritic spines: small protrusions from dendrites that serve as the primary sites of excitatory connections in the brain. While new spine formation has been directly observed in the cortex of animals during learning[2–4], the functional properties of new spines have only been inferred, precluding a detailed description of how they actually contribute to learning-related changes. For example, many lines of evidence now support that spines are functionally clustered on single dendritic branches[5–8], and new spines have been suggested to follow this pattern[4], implying that new spines cooperate with nearby spines to subserve the generation of learned behaviors. However, the activity patterns of spines that give rise to clusters containing new spines, and how the resulting clusters are activated during learned behaviors, remain unknown. Furthermore, the processes that ensure that new spines connect with axon partners that are appropriate for learning are unknown. In this study, we sought to better understand the function of new spines that form during motor learning by investigating the detailed structural and functional properties of the dendritic sub-domains where they form. To this end, we performed longitudinal functional imaging of dendritic spines to examine whether and how learning-related information is encoded in new spines. We then performed correlated electron microscopy (EM) to investigate the presynaptic axons providing inputs onto the new spines and their neighboring spines.

Results

As a robust platform of learning that induces new spine formation, we used a motor learning task amenable to simultaneous imaging in which head-fixed mice press a lever with their left forelimb in response to an auditory cue to receive a reward[9] (Fig. 1a,b and Methods). With daily training in this task, mice develop a reproducible movement that is accompanied by reproducible population activity in L2/3 of the primary motor cortex (M1)[9]. Furthermore, this learning induces new spines on the apical dendrites of these M1 L2/3 neurons[9,10]. Consistent with these reports, mice trained in this task in the current study showed a significant improvement in the success rate, reaction time and movement reproducibility over 2 weeks (Fig. 1c–e and Extended Data Fig. 1a,b). To examine the spine activity and structural plasticity of M1 L2/3 neurons, we sparsely expressed either the calcium sensor GCaMP6f or the glutamate reporter iGluSnFR[11–13], and performed two-photon imaging of single optical planes of dendrites in the apical arbor in L1 during behavioral sessions. Up to three fields of view (FOVs) were imaged in each animal on different imaging sessions, each being at least 500 µm apart. To minimize photodamage, we imaged each FOV three times, with a 5-day interval between imaging days for a given field (Extended Data Fig. 1c). Dendrites tolerated these conditions well, and showed no significant deterioration in structure nor global reductions in the frequency of GluSnFR events (Methods and Extended Data Fig. 1d,e). In a subset of animals ($n=4$), we performed post hoc correlated EM to gain ultrastructural readouts of the imaged dendrites and surrounding areas (Methods and Extended Data Fig. 2). For these animals, an additional imag-

[1]Neurobiology Section, Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA. [2]Center for Neural Circuits and Behavior, University of California, San Diego, La Jolla, CA, USA. [3]Department of Neurosciences, University of California, San Diego, School of Medicine, La Jolla, CA, USA. [4]Halıcıoğlu Data Science Institute, University of California, San Diego, La Jolla, CA, USA. [5]Center for Research in Biological Systems, University of California, San Diego, School of Medicine, La Jolla, CA, USA. [6]National Center for Microscopy and Imaging Research, University of California, San Diego, La Jolla, CA, USA. ✉e-mail: nghedric@gmail.com; tkomiyama@ucsd.edu

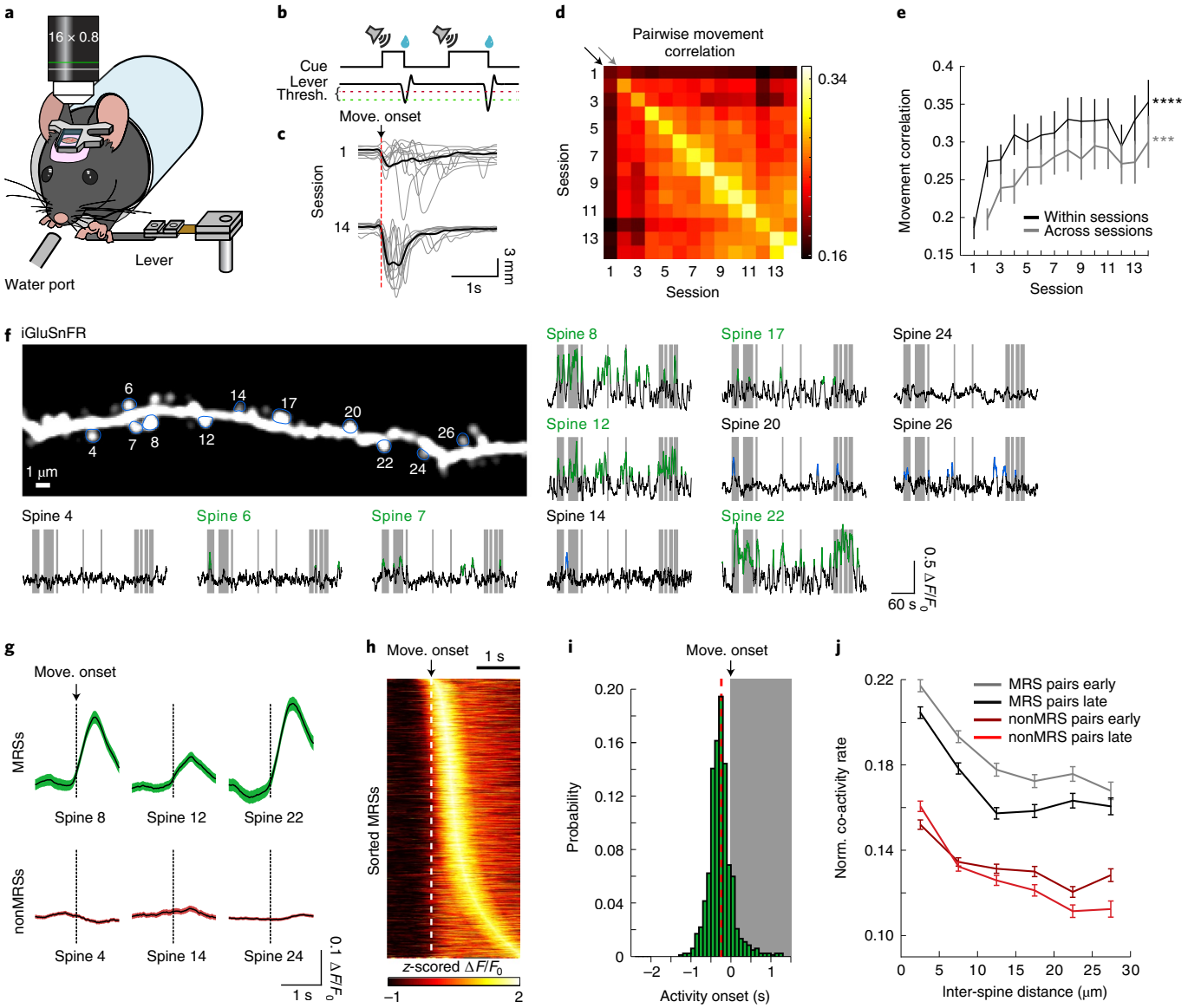


Fig. 1 | Characterization of movement-related signals in dendritic spines in M1 during motor learning. a, Schematic of experimental setup. **b**, Task structure. **c**, Lever movement traces during rewarded trials in sessions 1 and 14 for one mouse. Gray, ten individual trials; black, average of all trials. **d**, Correlogram of lever trajectory correlation within and across sessions ($n = 53$ mice). Each box corresponds to the median pairwise correlation coefficients of rewarded movement trajectories over a 3-s window (as shown in **c**), averaged across animals. **e**, The mean lever trajectory correlation increases both within (black line; center diagonal from **d**; $P = 0.00005$, Pearson's correlation) and across (gray line; +1 diagonal from **d**; $P = 0.0002$, two-sided Pearson's correlation coefficient) learning sessions. $n = 53$ mice. Data points correspond to means ± s.e.m. **f**, Example average-intensity projection of a single-plane in vivo imaging time series of an iGluSnFR-expressing apical dendrite of an L2/3 excitatory neuron in M1, along with associated fluorescence traces of a subset of spines (demarcated with blue outline). Numbers correspond to associated spine traces. Spines classified as MRSs are labeled in green. Portions of each trace classified as 'active' are demarcated with overlaid green (for MRSs) or blue (for other spines) lines. **g**, Comparison of movement-onset-aligned average fluorescence traces of MRSs (green) and nonMRSs (red). Spine numbers correspond to the spines in **f**. **h**, Movement-onset-aligned, *z*-scored activity from all imaged MRSs pooled across all sessions, sorted by peak timing. $n = 2{,}554$ MRSs. **i**, Onset timing histogram of MRS activity with respect to movement onset. Gray shaded region corresponds to the period after movement onset ($t = 0$). Median onset value (−280 ms) is plotted as a vertical dashed red line. $n = 2{,}554$ MRSs. **j**, Pairs of MRSs show higher co-activity rates than nonMRS pairs in both early (1–3) and late (11–13) sessions (two-way analysis of variance (ANOVA; $F = 418.3$, d.f. = 3) with post hoc test using the least-significant difference (LSD): early MRS versus early nonMRS: $P = 5 \times 10^{-114}$; late MRS versus late nonMRS: $P = 5 \times 10^{-157}$). Closer pairs show higher co-activity rates in all groups (two-way ANOVA, main effect of distance: $F = 78.1$, d.f. = 6, $P = 1 \times 10^{-96}$; group × distance interaction: $F = 2.8$, d.f. = 18, $P = 6 \times 10^{-5}$; first distance bin versus all other bins: $P < 0.0001$ for all groups), and co-activity rates significantly negatively correlated with distance (early MRS pairs: $r = -0.15$, $P = 1 \times 10^{-25}$; early nonMRS pairs: $r = -0.09$, $P = 1 \times 10^{-8}$; late MRS pairs: $r = -0.12$, $P = 4 \times 10^{-12}$; late nonMRS pairs: $r = -0.18$, $P = 6 \times 10^{-30}$). Co-activity rates were normalized in each spine pair by the geometric mean of the activity rates of the two spines in the pair. $n = 894$ early MRSs/6,990 early MRS–MRS pairs; 873 early nonMRSs/9,434 early nonMRS–nonMRS pairs; 818 late MRSs/6,556 late MRS–MRS pairs; 1,097 late nonMRS–nonMRS pairs. Cases of only one MRS being present on an imaged dendrite (6 total MRSs per dendrite) were necessarily excluded from co-activity analysis of MRSs. Means ± s.e.m.

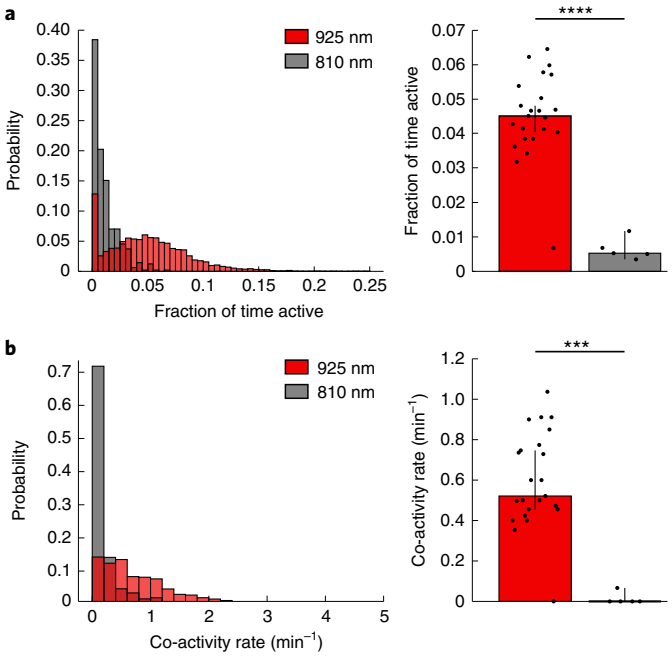


Fig. 2 | iGluSnFR fluorescence signals are dependent on glutamate sensitivity. Comparison of iGluSnFR signals using excitation at 925 nm, which was used for the experiments described in this study, with signals using excitation at 810 nm, which is near the isosbestic point of iGluSnFR such that iGluSnFR signals are independent of glutamate concentration. Separate cohorts of mice were used for 810-nm and 925-nm imaging. **a**, Comparison of the fraction of time spines were considered 'active' while imaging iGluSNFR with excitation at 925 nm (red) versus 810 nm (gray). Left, histograms of the fraction of time active (that is, total time above threshold divided by the total imaging time) for all spines imaged across all animals in both groups. Right, bar graph summary of the fraction of time active for each animal. $P < 0.001$, two-sided rank-sum test. Median ± bootstrap 95% confidence intervals (CIs). $n = 5$ mice/17 fields/33 dendrites/484 spines imaged at 810 nm; $n = 23$ mice/45 fields/76 dendrites/1,915 spines for 925 nm. Individual data points show average for each mouse. **b**, Comparisons of co-activity rates based on excitation wavelength. Left, histograms of all co-activity rates across all spine pairs within the same imaging field imaged at the corresponding wavelength. $n = 5$ mice/17 fields/33 dendrites/484 spines/4,909 condendritic spine pairs imaged at 810 nm; $n = 23$ mice/45 fields/76 dendrites/1,915 spines/94,060 condendritic spine pairs imaged at 925 nm. Right, bar graph summary of the co-activity rates of each animal imaged at the corresponding wavelength. $P = 4 \times 10^{-4}$, two-sided rank-sum test. Median ± bootstrap 95% CIs. Individual data points show the median for each mouse. $n = 5$ mice imaged at 810 nm; $n = 23$ mice imaged at 925 nm.

ing bout was performed on the final behavioral session (session 14) to ensure acquisition of the most up-to-date structural state of the dendrite.

To help contextualize patterns of new spine formation, we first sought to characterize the synaptic activity of preexisting spines on the apical dendrites of L2/3 neurons in M1. The prevalence of action potential-related global events that invade spines in GCaMP6f-expressing dendrites made it difficult for us to achieve a reliable readout of synaptic activity at individual dendritic spines. Even though we acknowledge that post hoc subtraction methods have been successfully used to calculate spine-specific activity[5,14,15], in our data these global events appeared to be particularly strong, and we sought an approach to selectively visualize inputs to individual spines. iGluSnFR is sensitive primarily to glutamate, and therefore does not suffer from contamination from global signals. Thus, we limited analysis of synaptic activity to iGluSnFR experiments ($n = 23$ mice/45 imaging fields/76 dendrites/1,915 unique spines). One of the GCaMP animals was used for structural and EM analyses. iGluSnFR signals of individual preexisting spines were heterogeneous (Fig. 1f), presumably reflecting diverse presynaptic inputs to individual spines. Based on the iGluSnFR signals from each spine, we were able to subdivide spines into those spines showing significant lever movement-related activity (movement-related spines, or MRSs, defined similarly to movement-related neurons in previous publications[9,10] as those spines that show significantly more activity during lever movements compared to no-movement periods; Methods, Fig. 1g–i and Extended Data Fig. 3a) and those spines whose activity did not correlate with lever movements (nonMRSs). Overall, 51% (2,554/5,005) of imaged spines pooled across all sessions were MRSs. MRSs were more consistently active during movement periods than nonMRSs (Fig. 1g and Extended Data Fig. 3b), with 78% (1,991/2,554) exhibiting peak activity within 1 s of movement onset, and 88% (2,258/2,554) showing activity onset timing that preceded the movement (Fig. 1h,i). For subsequent analyses, we used a previously reported[9,16] event-detection method to define active periods for each spine (Extended Data Fig. 3c). We found that, despite the majority of MRS activity occurring during movements, individual MRSs were active during a minority (~20%) of movements (Extended Data Fig. 3d). Given such sparse and heterogeneous movement-related inputs to spines, we next explored potential patterns of spatiotemporal coordination of MRSs and nonMRSs on single dendrites. Because many lines of evidence now support the idea that functionally related spines are 'clustered'—that is, that they are closer together on a single dendrite[4–7,15,17–22], we next asked whether such functional clustering is present for preexisting spines on the apical dendrites of L2/3 excitatory neurons in M1. To characterize functional clustering, we examined the frequency of spine pairs on a dendrite exhibiting activity events simultaneously. Consistent with previous reports, we indeed found that preexisting spine pairs show a pronounced distance-dependent relationship in their co-activity rates, such that closer spine pairs are more frequently co-active than more distant pairs on the same dendrite (Fig. 1j). This pattern was consistent in both early (sessions 1–3) and late (sessions 11–13) learning sessions. Critically, imaging iGluSnFR near its isosbestic point[11] (810 nm of excitation), where iGluSnFR fluorescence is insensitive to glutamate concentration, yielded a drastic reduction in detected events, and median co-activity rates of zero (Fig. 2), indicating that our observations are not due to glutamate-independent artifacts.

Given the observed functional clustering of preexisting dendritic spines (particularly MRSs), we considered the possibility that such clustering informs the sites of new spine formation. A similar idea has been put forward previously based on the observation that new spines in the motor cortex seem to cluster around apparently potentiated, previously formed new spines[4]. However, a description of the synaptic activity of preexisting spines and its relationship to new spine formation is still lacking. We therefore explored the spatial relationship between MRSs and new spines (Fig. 3a), under the hypothesis that the presence of MRSs on a dendrite promotes nearby spinogenesis. To identify new spines, we compared the motion-corrected, time-averaged, single-plane images from each of the three imaging sessions (early, middle and late) for a given field, allowing longitudinal structural analysis of dendrites and spines without the need for coexpression of a cell fill. Subsequent EM analysis of a subset of animals showed that 100% of spines identified in this way contain structural synapses (Extended Data Fig. 4). Using this approach, we observed at least one new spine formation event on 50/76 (66%) of imaged dendrites, with a median overall new spine density of 4.0/100 µm across all dendrites, similar to the densities described in a previous report[2].

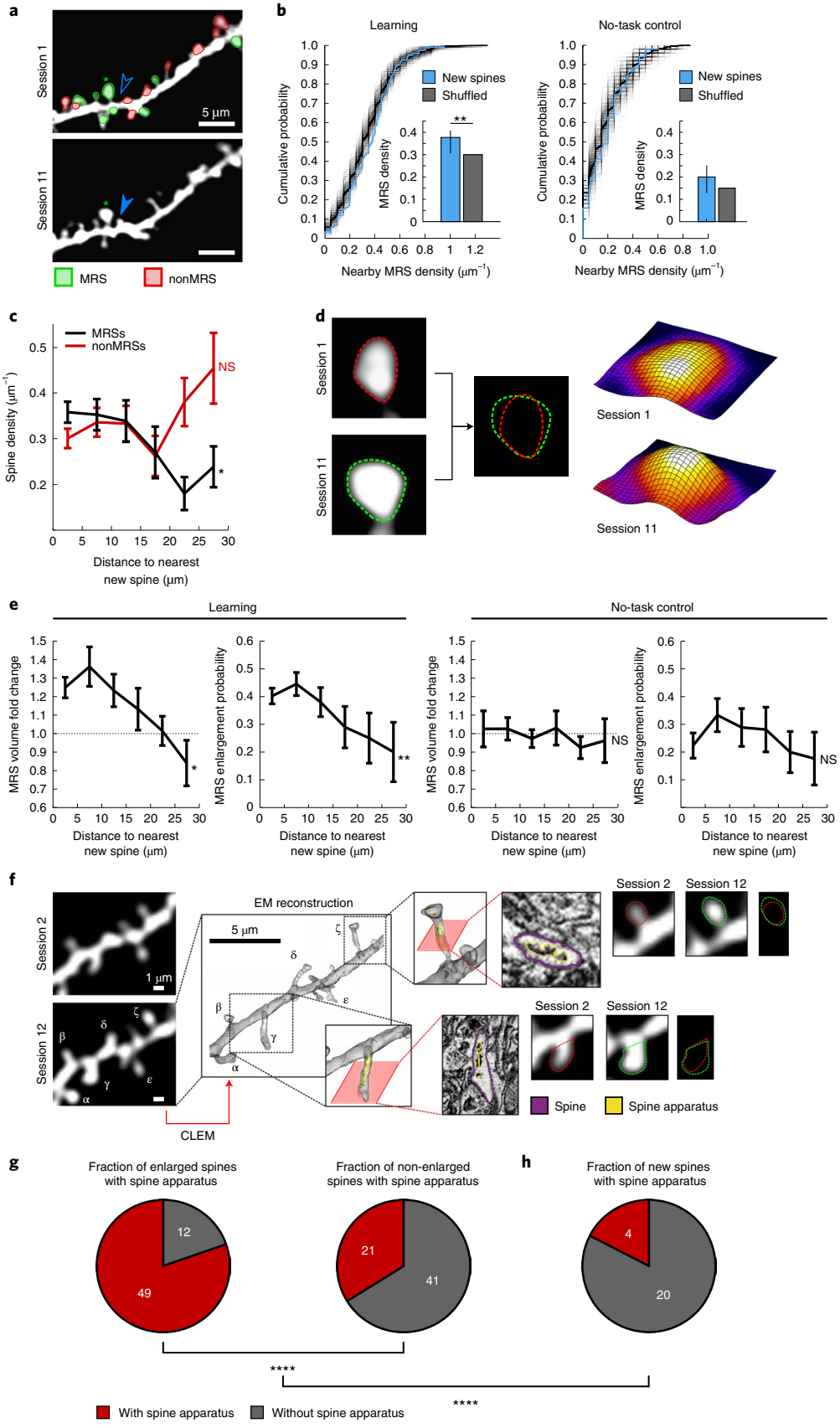
To understand the extent to which our observations are specific to task learning, we exposed an independent cohort of mice to a contingency-broken version of the lever-press task (hereafter 'no-task control'). In the no-task control, water was administered at the end of every trial, irrespective of the motion of the lever (Methods). Despite the absence of the requirement, mice nonetheless often moved the lever. However, movements did not become more reproducible over days, unlike in the learning condition (Extended Data Fig. 5). In the no-task control, we found the activity event frequency of spines to be comparable to the learning condition (Extended Data Fig. 5d compared to Extended Data Fig. 1e). MRSs in the no-task condition were less dense (Extended Data Fig. 5e) and less co-active (Extended Data Fig. 5h compared to Fig. 1j) than the learning condition, but were nonetheless functionally clustered. Consistent with a learning-related role of new spine formation, we found the density of new spines in the no-task condition to be significantly lower than in the learning condition by twofold (Extended Data Fig. 5f). Additionally, the learning group displayed a higher rate of new spine survival than the no-task group (Extended Data Fig. 5g), suggesting the engagement of new spines during learning. Thus, directed motor learning increases the engagement of movement-related signals and increases the rate of new spine formation and stabilization.

To address whether the formation sites of new spines during learning were influenced by nearby task-related input activity (that is, MRSs), we tracked the dendritic positions of MRSs from early imaging sessions, before new spine formation. We found that there is a higher density of MRSs in the vicinity (10 µm) of future new spine formation sites than expected by chance (that is, randomizing new spines' locations; Fig. 3b). This was specific to MRSs, as the overall spine density did not predict new spine locations (Extended Data Fig. 6). This analysis in the no-task control exhibited a similar trend but failed to yield significance (Fig. 3b). Consistently, the density of MRSs significantly decreased as a function of distance from new spine formation sites in the learning condition (Fig. 3c), but not in the no-task control (Extended Data Fig. 5i). This effect during learning was specific to MRSs, as nonMRSs showed no such distance-dependent change in density (Fig. 3c). These data suggest that dendritic micro-domains with higher MRS density early in learning locally promote the formation of new spines.

What is the mechanism that favors new spine formation near MRSs? It has been reported that synaptic potentiation at a spine can lead to priming of the surrounding dendrite[23,24], creating a localized dendritic environment wherein plasticity, possibly including new spine formation, is more likely. We therefore asked whether MRSs nearby sites of new spine formation showed signs of having undergone synaptic potentiation during learning, using enlargement of spines as an indicator of synaptic potentiation[25–27] (Fig. 3d). Enlargement was assessed by comparing the estimated spine volume (Methods) from early sessions (before new spine formation) to the sessions in which the new spine was first visible. We found that both the mean spine volume fold change and the probability of spine enlargement (fold change $\geq 1.1\times$) was higher for MRSs near new spines compared to MRSs more distant from new spines (Fig. 3e), suggesting that MRS potentiation locally enhances new spine formation. This effect was not observed in the no-task control, wherein overall spine volume changes were less pronounced and showed no distance-dependent relationship with new spines (Fig. 3e). These results were consistent across a range of arbitrary thresholds for spine enlargement (Extended Data Fig. 7a,b). While it is conceivable that spine volume estimates could be confounded by the frequency of activity events in spines, we found no positive correlation between spine volume estimates and event frequency (Extended Data Fig. 7c,d). Further, recalculating spine volume after spine-wise

Fig. 3 | New spines form near enlarged, movement-related spines in the motor cortex during motor learning. a, In vivo images (average-intensity projections of time series) of an iGluSnFR-expressing dendrite from early (top) and late (bottom) learning sessions, with MRSs (green) and nonMRSs (red) labeled. Arrowhead indicates a new spine. Asterisks indicate spines used for spine volume examples in **e**. **b**, New spines form in regions of higher MRS density. Left, cumulative probability plot of the density of MRSs within 10 µm of new spines (light blue) in the learning condition compared to shuffled values found by randomizing new spine locations 10,000 times (black) from the same data. Right, analogous plot for the no-task control condition. Individual shuffles are plotted in gray. Insets, the median density of nearby MRSs was significantly higher for new spines than for shuffled data ($P=0.003$), an effect that was similar but did not reach statistical significance in no-task controls ($P=0.06$; P value calculated as the fraction of shuffles following the null hypothesis via a one-sided query of whether the median of each shuffle was greater than the median of the real data). Median ± bootstrapped 95% CIs. $n=118$ new spines (learning group); $n=51$ new spines (no-task control group). **c**, MRS density, but not nonMRS density, decreased as a function of distance from new spines (two-way ANOVA (MRS label × distance), significant interaction of MRS label with distance ($F=2.27$, d.f. = 5, $P=0.046$)), showing that the effect of MRS status on the density measurement depends on distance from the new spine. Correspondingly, MRS density was significantly negatively correlated with distance (Pearson's correlation coefficient, $r=-0.15$, $P=0.009$), while nonMRS showed no such correlation ($r=0.10$, $P=0.09$)). Mean ± s.e.m. $n=118$ new spines; 697 MRSs present on 50 new spine-containing dendrites. All statistical tests are two-sided. **d**, Left, in vivo images of an example MRS showing 58% volume enlargement over learning (displayed spine is indicated with an asterisk in **a**). Spine outline from early (green) and late (red) sessions (defined automatically by intensity values above background; Methods) are shown in each image, and overlaid in the middle image for comparison. Right, three-dimensional surface plots of the spine showing a clear increase in intensity across the spine head. **e**, New spines formed nearby enlarged MRSs. Left, MRSs near sites of new spine formation showed a larger mean spine volume change (left; $r=-0.09$, $P=0.04$, Pearson's correlation coefficient) and were more likely to show enlargement (volume > 1.1 × early session volume; right) than MRSs in sites further from new spines ($P=5\times10^{-5}$; Pearson's correlation coefficient of enlargement probability versus distance). $n=118$ new spines; 697 MRSs present on 50 new spine-containing dendrites. Right, in the no-task control condition, neither the fold change in MRS volume (Pearson's correlation coefficient, $r=-0.015$, $P=0.79$) nor the probability of MRS enlargement (Pearson's correlation coefficient, $r=0.02$, $P=0.68$) changed as a function of distance to new spines. $n=51$ new spines; 339 MRSs on new spine-containing dendrites. Mean ± s.e.m. All statistical tests are two-sided. **f**, Preexisting spines that show enlargement over learning tend to house spine apparatus. Left-most images show early and late in vivo images showing dendrites used in CLEM. Second-to-left images show EM reconstruction of the dendrite shown on the left image, with Greek letters indicating fiducial spines for alignment. The two outlined regions indicate spines under examination. Zoomed-in reconstruction images for both of these spines are provided on the right, and include reconstructed spine apparatus (yellow) contained within, as well as a schematized sectioning plane from which EM micrographs are drawn. Beside each zoomed-in window is a corresponding EM micrograph showing the spines of interest (purple outline) and highlighted spine apparatus (yellow outline). Right-most images show early and late in vivo images showing spine enlargement, with early (red) and late (green) spine outlines provided for comparison. Overlaid outlines for the spines are shown to the right for clarity. **g**, Pie charts showing the fraction of spines housing a spine apparatus. The spines that enlarged during learning in vivo (left) are more likely to house a spine apparatus than non-enlarged spines (right; Pearson's chi-squared test of independence; $P=2\times10^{-7}$). **h**, Pie chart showing the fraction of new spines that house a spine apparatus. This fraction (0.17) is significantly smaller ($P=5\times10^{-8}$; Pearson's chi-squared test of independence) than the overall fraction of preexisting spines from **g** (70/123 = 0.57). NS, not significant.

removal of activity periods produced values that tightly correlated with the original estimates (Extended Data Fig. 7e), collectively suggesting that our volume estimates are not strongly affected by spine activity. In addition, when inspected in post hoc electron micrographs, enlarged spines showed a very high probability (0.8) of housing a spine apparatus, an elaborated form of endoplasmic



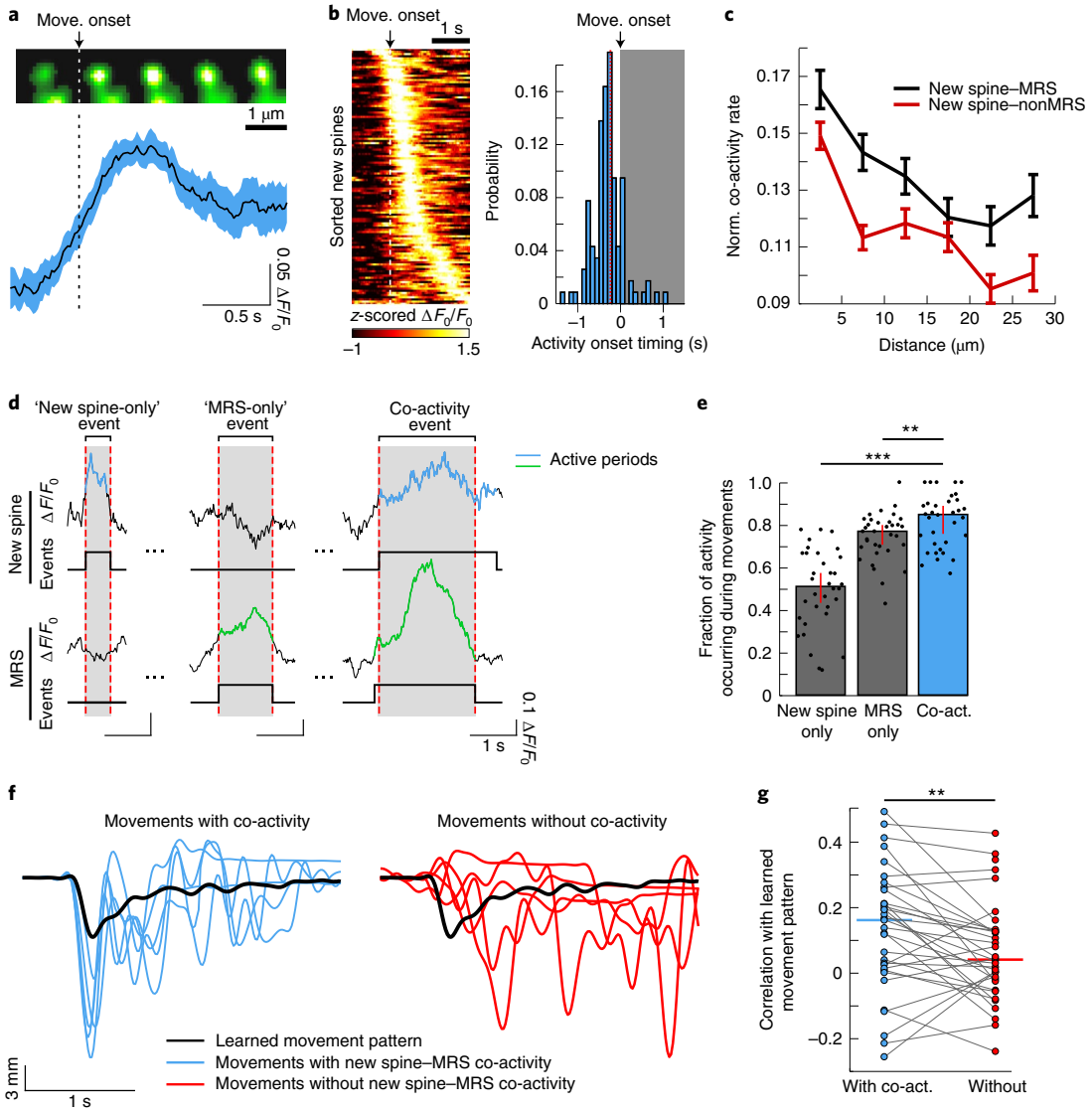


Fig. 4 | Functional clusters of new spines and movement-related spines preferentially encode learned movements. a, Top, images of a new spine from peri-movement periods, averaged across movements. Vertical dashed line indicates movement onset. Bottom, average movement onset-aligned activity of the new spine shown above. Mean ± s.e.m. **b**, Left, movement onset-aligned, *z*-scored activity of all imaged new spines, sorted by timing of peak activity. Right, histogram of onset timing of new spine activity with respect to movement onset. Gray shaded region indicates time periods after movement onset ($t = 0$). Median onset timing (−232 ms) is indicated with a vertical red dotted line. $n = 118$ new spines. **c**, New spines show higher co-activity rates with MRSs than with nonMRSs (two-way ANOVA (MRS label × distance), main MRS label effect: $F = 31.12$, d.f. = 1, $P = 3 \times 10^{-8}$). All pairs showed co-activity that depended on interspinal distance (main distance effect: $F = 20.4$, d.f. = 5, $P = 5 \times 10^{-20}$), and each group showed a significant decrease over distance, wherein the 0–5-μm bin was significantly higher than other distance bins (post hoc test using LSD, new spine–MRS: $P < 0.001$ versus all other bins; new spine–nonMRS: $P < 0.005$ versus all other bins). Further, co-activity rates were significantly negatively correlated with distance (Spearman's rank coefficient, new spine–MRS versus distance, $r = -0.16$, $P = 1 \times 10^{-7}$; new spine–nonMRS: $r = -0.20$, $P = 5 \times 10^{-14}$). $n = 1{,}658$ new spine–MRS pairs; 2,207 new spine–nonMRS pairs. Mean ± s.e.m. **d**, Example traces illustrating a new spine-only activity event, an MRS-only activity event and a co-activity event. Shown are both the $\Delta F/F_0$ trace (upper) as well as binarized 'event' traces (lower) indicating active and inactive periods for each spine. Colored regions on the $\Delta F/F_0$ trace also indicate active periods for each spine. Gray blocks demarcate the periods defined by each of the three activity types. **e**, New spine–MRS co-activity occurred preferentially during movements. The fraction of new spine–MRS co-activity events occurring during movements was higher than the fraction of new spine-only ($P = 4 \times 10^{-7}$) and MRS-only events ($P = 3 \times 10^{-4}$; two-sided sign-rank test correcting for multiple comparisons using the false discovery rate (FDR) method). $n = 34$ fields. Median ± bootstrapped 95% CIs. **f**, Example lever traces of individual movements that coincide with new spine–MRS co-activity (left; blue lines) compared to movements that lack new spine–MRS co-activity (right; red lines). The 'learned movement pattern' (the average of all rewarded movements over the late (11–14) learning sessions) is overlaid on each group (black line). **g**, Movements with new spine–MRS co-activity (light blue) were more similar to the learned movement pattern than movements lacking such activity (red; $P = 0.003$; two-sided signed-rank test). Each data point corresponds to the median correlation value of all new spine–MRS pairs within a given field ($n = 34$ fields). The medians of these values are indicated as horizontal lines of the corresponding color.

reticulum within spines that is associated with spine maturity[28] and long-term potentiation[29]. This value was significantly higher than other spines (0.34) from the same dendrites (Fig. 3f,g and Extended Data Fig. 7f). Thus, our in vivo spine volume measurements agree with an intracellular correlate of spine maturity acquired from an orthogonal measurement, indicating that spine enlargement in our

data is a reliable measure of spine potentiation. We also note that an especially small fraction (0.17) of new spines house a spine apparatus, supporting the idea that these are truly newly formed spines (Fig. 3h). Taken together, these data suggest that learning-related new spine formation occurs preferentially in sites of clustered synaptic strengthening of spines showing task-related activity.

We next sought to characterize the activity of new spines after they form. Like preexisting spines, new spines showed heterogeneous activity patterns. Nevertheless, 75% (89/118) of new spines showed activity whose peak was within 1 s of movement onset, and 83% (98/118) displayed activity onset that preceded movements, similar to preexisting MRSs (Fig. 4a,b and Extended Data Fig. 3). Does this activity of new spines exhibit a local co-activity structure similar to preexisting spines? We found that, much like preexisting MRSs, new spines show higher average co-activity rates with MRSs than with nonMRSs (Fig. 4c). Furthermore, new spines showed a distance-dependent decay of co-activity rates with preexisting spines (both MRSs and nonMRSs), suggesting that new spines participate in functional clusters. In contrast, new spines observed in no-task controls exhibited a substantially reduced version of such functional clustering (Extended Data Fig. 5j), suggesting that the local functional coordination of new spines depends on learning. Given that new spine activity is coordinated with nearby MRSs, we then examined how the co-activity of new spines and MRSs is related to lever movements. We found that new spine–MRS co-activity was significantly more specific to movement periods than was either of the constituent spines' (that is, the new spine or MRS) activity in isolation (Fig. 4d,e). Taken together, these data reveal that the activity of spines that formed during task learning show strong co-activity with nearby, task-related synapses, and that co-activity of new spines and MRSs is particularly prominent during movements. These observations were less clear in the no-task controls, suggesting that the local coordination of activity involving new spines is specific for the acquisition of the novel motor skill (Extended Data Fig. 5).

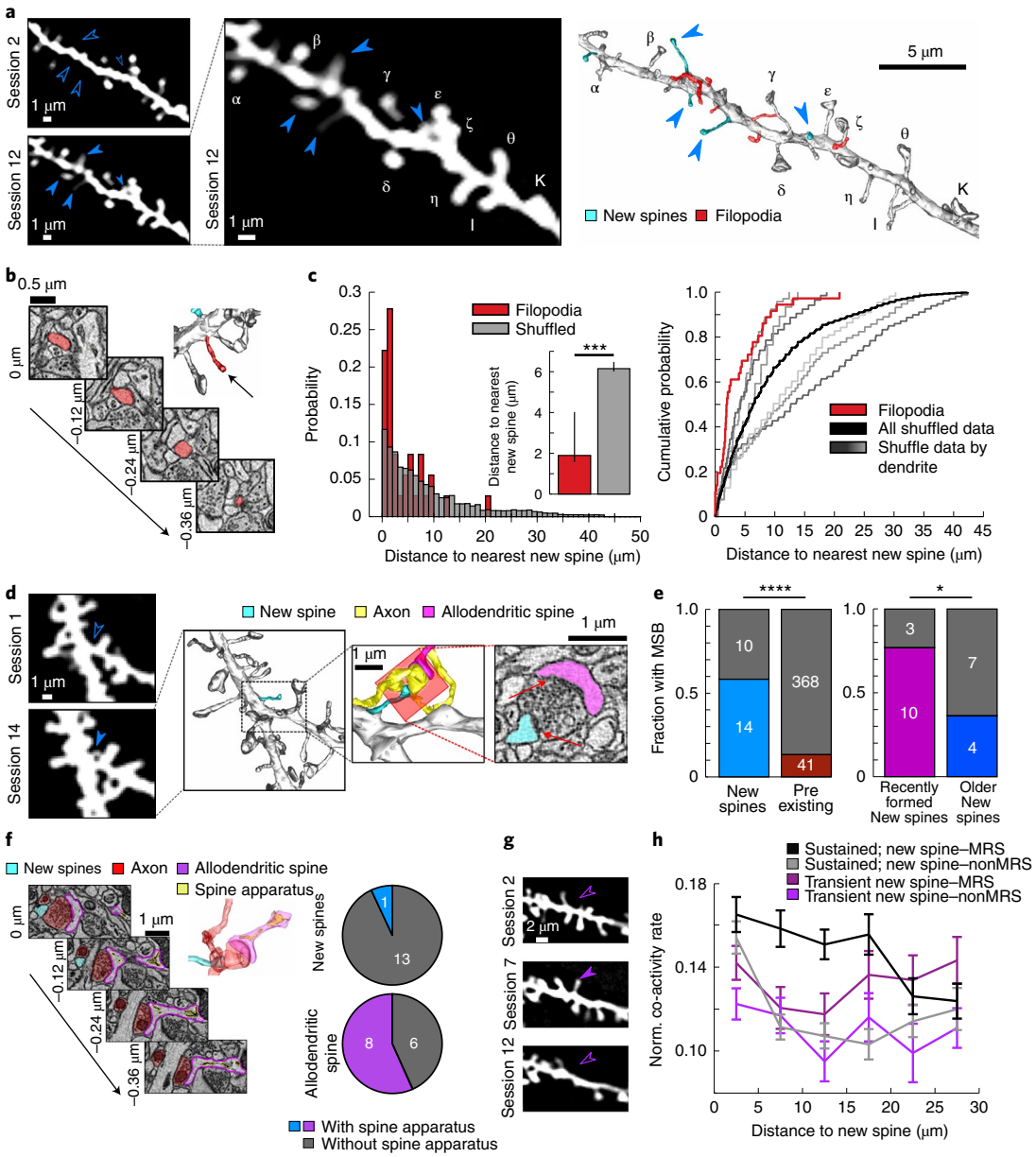
Following the observation that new spine–MRS co-activity is strongly associated with movements, we further investigated the quality of the movements associated with such co-activity. Given the reported role of M1 in the learning of reproducible movements[9,10,30–32], we hypothesized that co-activity of new spine–MRS pairs would contribute to the generation of the learned movement. We thus compared individual movements to the learned movement pattern, defined as the average kinematics of rewarded lever presses from late (11–14) learning sessions[9]. We found that movements that coincide with the co-activity of new spine–MRS pairs were more similar to the learned movement pattern than movements lacking such co-activity (Fig. 4f,g). These results are consistent with the idea that co-activity of new spine–MRS pairs contributes to the generation of the learned movement.

Interestingly, these movement features of new spine–MRS co-activity (the movement-period enrichment, and the preferential encoding of the learned movement pattern) were observed

Fig. 5 | Correlated light and electron microscopy reveals patterns of microstructures surrounding functional clusters. a, Example CLEM images showing filopodial clustering around new spines. Left, early (session 2) and late (session 12) in vivo images (average projections of time series) showing new spine formation (blue arrowheads) over learning. Middle, magnified late-session in vivo image with labeled fiducial structures (Greek letters) for alignment with EM image. Right, EM reconstruction of the same dendrite, with new spines (cyan structures, blue arrowheads) and fiduciary structures (Greek letters) labeled. Filopodia (red structures) not visible in vivo are revealed by EM to cluster around new spines. **b**, Example filopodium (red) identified in EM showing lack of synaptic contact (absence of postsynaptic density (PSD) and aggregated presynaptic vesicles) over the full length of the structure. Inset in the upper right shows the fully reconstructed filopodium (arrow). **c**, Left, histogram of the distances between filopodia and their closest new spine demonstrating that filopodia (red) cluster around new spines. True distances were compared to chance, calculated by shuffling the filopodia locations 10,000 times (gray). Inset, bar graph summary of data. Median ± bootstrapped 95% CIs. ($n = 36$ filopodia/6 dendrites/5 fields/4 animals; $P = 0.0003$). P value calculated as the fraction of shuffles following the null hypothesis by a one-sided comparison of whether the median of each shuffle was greater than or equal to the median of the real data. ***$P < 0.001$. Right, cumulative probability distributions of the distance between filopodia and their nearest new spine compared to shuffled data. Distributions of shuffles separated by individual dendrites are shown in gray, and the overall shuffled distribution in black. **d**, CLEM identification of new spine synapsing with MSBs. Left, example in vivo images of a dendrite showing new spine formation by the final session. Background was manually removed for clarity. Middle, EM reconstruction of the region surrounding the new spine (highlighted in cyan), along with a zoomed-in image of the reconstructed region (dotted-line box shows region of interest (ROI); other spines were removed in the zoomed-in image for clarity), with the axon associated with the new spine and the allodendritic spine added. Right, EM image showing the new spine synapsing with an MSB; both synapses on the bouton are indicated with red arrows. **e**, Summary of the prevalence of MSBs on different spine types. Left, new spines (light blue) show a higher fraction of MSBs than preexisting spines (red; chi-square test, $P = 1 \times 10^{-7}$). Right, MSBs were more common on recently formed new spines (those that formed between the middle and late sessions; magenta) than older new spines (those that formed between the early and middle sessions; dark blue; Pearson's chi-squared test of independence, $P = 0.04$). **f**, Left, example images (four sections) of the group of a new spine, an MSB and an allodendritic spine from **d**, showing the presence of a spine apparatus (yellow outlined region) in the allodendritic spine. Inset above shows the volume reconstruction, with a partially transparent allodendritic spine showcasing the internal spine apparatus. Right, new spines connected to an MSB were less likely to house a spine apparatus than the paired allodendritic spines, suggesting that the allodendritic spines are more mature than the new spines ($P = 0.005$; Pearson's chi-squared test of independence). **g**, Example in vivo images over early, middle and late sessions of learning showing 'transient' new spine formation; that is, the new spine appeared in the middle session and was eliminated by the late session (open arrowheads indicate preformation and post-elimination spines; filled arrowhead indicates new spine). **h**, New spine survival is related to the local co-activity structure with task-related spines. Both sustained and transient new spines showed significantly higher co-activity rates with MRSs than nonMRSs on middle learning sessions (three-way ANOVA (MRS label × distance × survival), main MRS label main effect: $F = 30.86$, d.f. = 1, $P = 3 \times 10^{-8}$; survival of new spine main effect: $F = 7.75$, d.f. = 1, $P = 0.005$; sustained new spine–MRS versus sustained new spine–nonMRS co-activity: $P = 2 \times 10^{-8}$; transient new spine–MRS versus transient new spine–nonMRS co-activity: $P = 0.004$, post hoc test using the LSD). Sustained new spine–MRS co-activity rates were significantly higher than both transient new spine groups (versus transient new spine–MRS pairs: $P = 0.009$; versus transient new spine–nonMRS: $P = 8 \times 10^{-9}$). A main effect of distance was observed (main distance effect: $F = 5.46$, d.f. = 5, $P = 1 \times 10^{-4}$), and the MRS label × distance × survival interaction was significant ($F = 2.38$, d.f. = 5, $P = 0.04$), suggesting that the distance-dependent relationship of MRS status with co-activity depends on the survival of the new spine, and justifying the statistical separation of the data at these three levels. Individual inspection of the data grouped in this way (that is, one group corresponding to each line) reveals that only sustained new spines showed a significant negative correlation of co-activity rates with distance (sustained new spine–MRS: $r = -0.16$, $P = 8 \times 10^{-5}$; sustained new spine–nonMRS: $r = -0.14$, $P = 5 \times 10^{-4}$; transient new spine–MRS: $r = 0.02$, $P = 0.74$; transient new spine–nonMRS: $r = -0.08$, $P = 0.17$). $n = 28$ transient new spines, 418 transient new spine–MRS pairs, 481 transient new spine–nonMRS pairs; 57 sustained middle-session new spines, 765 sustained new spine–MRS pairs and 1,024 sustained new spine–nonMRS pairs. Mean ± s.e.m.

regardless of the distance between the spines (Extended Data Fig. 6b,c). This result is reminiscent of a previous observation that distributed inputs along a dendritic branch, when synchronized, are equally capable of eliciting a nonlinear dendritic response[33]. However, the number of movements coinciding with the co-activity of new spine–MRS pairs had a significant relationship with the distance, with more movements being represented by closer pairs (Extended Data Fig. 6d). Thus, while new spine–MRS pairs across the length of dendrites are equally equipped to encode learned movements, closer pairs do so more often. These results collectively support the notion that the clustered coincident inputs onto new spines and MRSs contribute to the reproducible activity patterns of M1 L2/3 neurons reliably associated with the learned movement pattern[9].

How could new spines form so as to achieve the coordinated activation of new spines and MRSs? In other words, how could new spines find the appropriate presynaptic axons whose activity is correlated with inputs to nearby MRSs? A number of studies have favored a model in which filopodial outreach from dendrites is a mechanistic predecessor of new spine formation[34–36]. Such a process would allow efficient sampling of nearby axons, potentially giving the postsynaptic neuron access to axons showing correlated activity with those already synapsing on the parent dendrite. To examine this 'filopodial hypothesis' in the context of motor learning, we turned to our post hoc EM volumes and identified putative filopodia (that is, thin dendritic protrusions lacking a detectable synapse). These filopodia were not reliably visualized with in vivo imaging, perhaps due to their small caliber and high motility (as described in brain slices[37]; Fig. 5a,b). While several studies have imaged filopodia-like structures in vivo[38–41], this approach is limited by the inability to differentiate such structures from thin, but nonetheless synaptic, spines. In contrast, EM permits a higher-confidence assessment of synapses and thus a more accurate identification of true filopodia. We found that filopodia identified in EM were particularly abundant nearby sites of new spine formation (Fig. 5a,c), supporting the presence of 'hot spots' wherein multiple filopodia are (perhaps repeatedly) extended to explore nearby axons, and successful connections emerge as new spines[42].



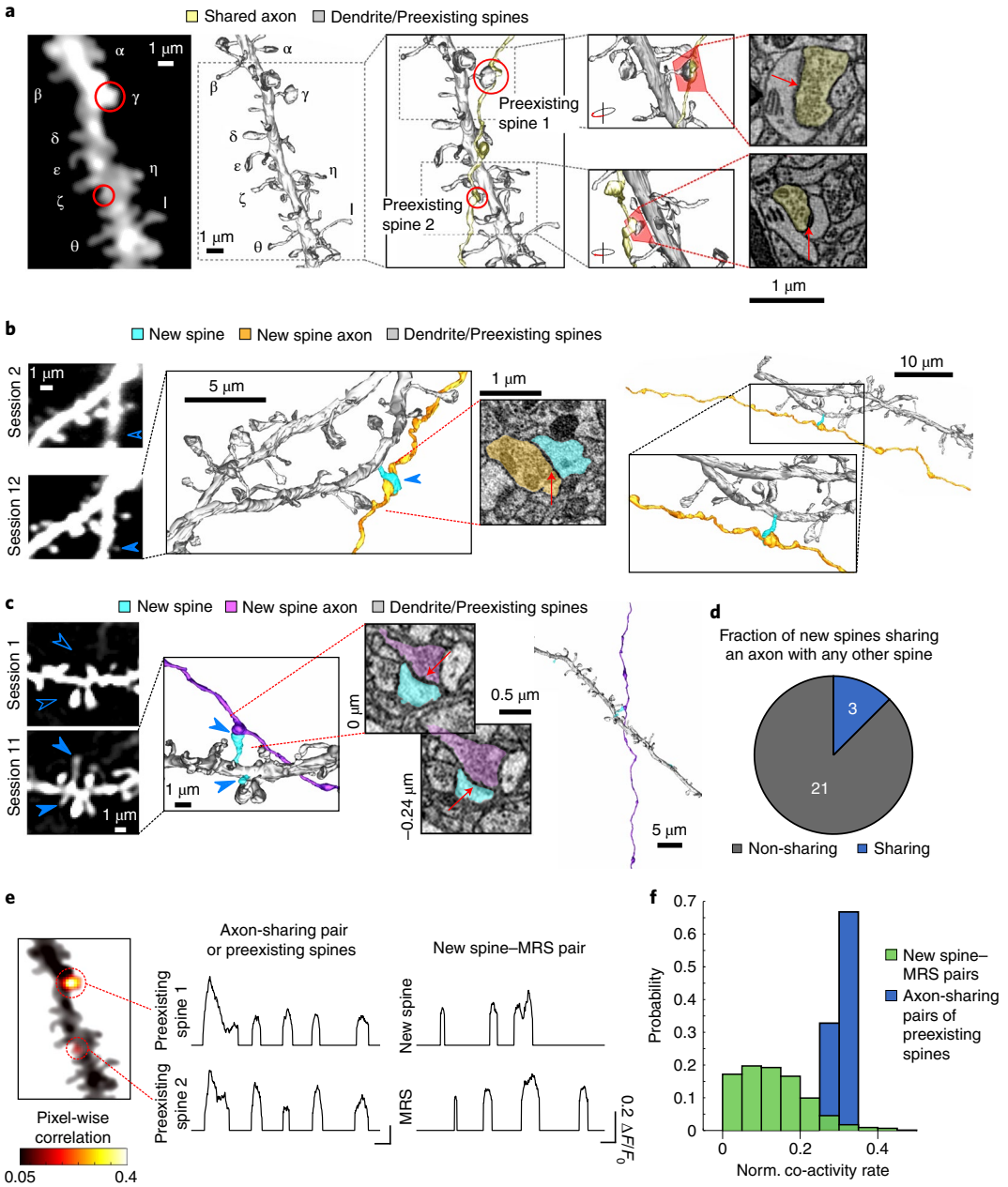


Fig. 6 | Most new spines represent novel connections on the dendritic segment. a, Example of axon sharing between preexisting spines. Left-most image, in vivo image (average projection of time series) of a dendritic segment found to have a pair of preexisting spines that share an axon (red circles). Fiducial structures for EM alignment are denoted by Greek letters. Second-to-left images, corresponding EM reconstructions of the same dendritic segment. Fiducial structures marked with Greek letters. Reconstruction on the right shows a zoom-in portion of the dendrite (demarcated by dashed box on the left), with reconstruction of an axon (yellow) shared by the two preexisting spines (red circles) added. A zoomed-in image (demarcated by gray dashed-line boxes) are provided on the right. Zoomed-in versions are rotated for clarity. Within each zoomed-in image, the approximate EM slicing plane (red planes) are indicated. Right, EM micrographs of each spine showing the presence of a synapse (red arrows) with the axon of interest. **b**,**c**, Example reconstructions of new spines and their associated axons demonstrating the lack of axon sharing by new spines. Left, in vivo images showing new spine formation between early and late sessions (empty arrowheads for preformation new spine location, filled arrowheads for new spines). Middle, as in **a**, EM reconstructions of the corresponding dendritic regions and EM images illustrating synaptic connections with the axon of interest. Right, zoomed-out reconstructions of the dendritic and axonal segments. The approximate EM sectioning plane (red) is indicated. The axon in **b** runs parallel to the dendrite, and does not contact nearby spines on the targeted dendrite (zoomed-in window provided for clarity). The axon in **c** makes contact with another spine, but does not form a synapse, and the remainder of the axon runs nearly orthogonal to the imaged dendrite and does not ramify within the captured volume. **d**, Most new spines do not share axons with nearby spines within the EM volume (Pearson's chi-squared test of independence; $P = 4 \times 10^{-6}$, sharing versus non-sharing). **e**,**f**, Most new spine–MRS pairs do not show co-activity rates consistent with axon sharing. **e**, Left, pixel-wise correlation map of the dendritic segment shown in **a**, relative to the central pixel in 'preexisting spine 1' (top circle). High correlation values outside the basis spine (preexisting spine 1) are only present in 'preexisting spine 2', the spine found to share an axon with preexisting spine 1, as in **a**. Middle, example activity event traces illustrating the highly coherent activity of the confirmed axon-sharing pair of spines shown on the left. Right, example activity event traces of a new spine–MRS pair from the same dendrite. **f**, Histograms of normalized co-activity rates for all new spine–MRS pairs (green; $n = 1{,}658$ pairs) and all preexisting spine pairs that are confirmed in EM to share axons (blue; $n = 3$ pairs). New spine–MRS pairs show a lower median co-activity rate (0.12) versus confirmed axon-sharing pairs (0.32; $P = 0.009$, two-sided rank-sum test).

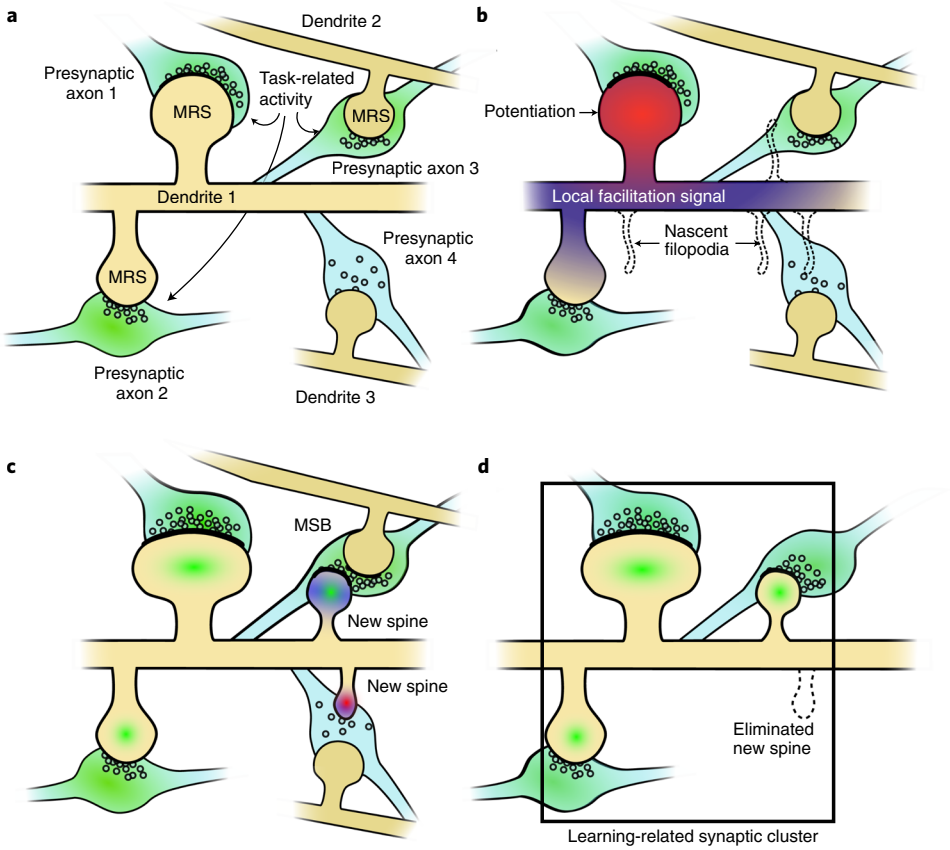


Fig. 7 | Proposed model of locally facilitated spinogenesis during motor learning. a, During early stages of learning, task-related inputs (presynaptic axons 1–3; task-related activity indicated by green haze) to preexisting spines (MRSs) on M1 L2/3 apical dendrites support task performance. **b**, A subset of these task-related synapses undergo synaptic potentiation (red fill in indicated spine), likely leading to local facilitation in the dendrite. The locally facilitated region becomes a hot spot for nascent filopodia (dashed outlines), which sample the surrounding neuropil for preexisting axonal boutons. **c**, Some of the filopodia make contact with a preexisting bouton, forming new dendritic spines (blue), and resulting in MSBs shared with allodendritic spines (spines from dendrites 2 and 3). Thereafter, new spines that show synchronized activity (green haze within spines) with surrounding task-related spines are selectively maintained, while those with less synchronous activity (red haze) are eliminated. This process selects new spines that contribute to functional clustering of task-related information. **d**, The surviving new spines participate in a new or modified synaptic cluster wherein correlated, task-related inputs to new spines and nearby spines (green haze) robustly drive the postsynaptic activity and the execution of the learned movement. The elimination of new spines receiving locally asynchronous activity (**c**) prevents the inclusion of functionally different inputs into the dendritic repertoire. Dendrite 2 was excluded from the drawing to reflect the unknown fate of MSBs (that is, whether they split into distinct boutons or the other spine retracts).

Do these filopodia establish connections with de novo presynaptic boutons of an axon? Previous reports have suggested that most new spines synapse with multisynaptic boutons (MSBs)[36,43,44], a finding that was used to support the 'filopodial outreach' hypothesis of new spine formation[43]. To test whether a similar phenomenon is present in our dataset, we used correlated light and electron microscopy (CLEM) to reconstruct the axonal boutons synapsing with new spines. We found that a majority (58%) of new spines indeed synapsed with axonal boutons hosting multiple synapses, a fraction that is considerably higher than that seen on preexisting spines (10%; Fig. 5d,e), consistent with previous reports[36,43,44]. These data further increase the confidence that new spines identified in this study are truly newly formed spines.

Synapsing on MSBs appears to be a transient feature of new spines, as a significantly lower fraction of older new spines (36%) are connected to MSBs than more recently formed new spines (77%; Fig. 5e). Thus, MSBs may split, or the other synapse on the MSB may be eliminated, as the new spines mature, potentially explaining why such structures are rarely observed in preexisting spines. The other synapses on these MSBs were always observed to be allodendritic; that is, they connect with spines from different parent dendrites. This observation is in contrast to previous reports from the hippocampus and cerebellum[45,46], where MSBs were often found to synapse with multiple spines from the same dendrite. Thus, cortical plasticity during motor learning appears to subserve different strategies than these brain regions. The allodendritic spines of the MSBs in our data are significantly more likely to house a spine apparatus than the new spines themselves (Fig. 5f), suggesting that the allodendritic spines are more mature, and thus more likely to predate the associated new spine. Taken together, these data support the model that new spines form by filopodia leveraging a nearby preexisting synapse on another dendrite to establish a new synaptic connection.

How could such an exploratory basis for new spine formation consistently achieve structured, local co-activity with MRSs? One possibility is that new spines are selected for survival based on their co-activity with other spines on the same dendrite. Such a possibility is supported by a recent study in hippocampal slices showing that spine plasticity is based on local activity synchrony with other nearby spines[47]. We therefore tested this possibility, taking

advantage of the fact that some of the new spines that formed in the middle imaging sessions were 'transient'; that is, they disappeared by the late sessions (Fig. 5g). We found that the survival of new spines could be predicted by their local co-activity pattern: new spines surviving through late learning sessions showed stronger co-activity preferentially with nearby MRSs when compared to transient new spines (Fig. 5h). This functional separation was absent in the no-task condition; surviving new spines did not show higher co-activity with nearby MRSs than with other groups (Extended Data Fig. 8m). These data thus suggest that, after formation, new spines showing strong co-activity with nearby task-related spines are selected for survival during learning.

Finally, we considered two potential mechanisms that could underlie the high co-activity rates of new spine–MRS pairs. First, the new spine and nearby MRSs could share the same presynaptic axon, thus representing a duplication of existing connections. Alternately, the new spine and nearby MRSs could receive distinct—but correlated—inputs. To examine these possibilities, we performed thorough EM reconstructions of axons synapsing onto the dendrites imaged in vivo for the entire EM volumes ($n = 330$ axons; Extended Data Fig. 2c). There were indeed clear cases of axon sharing between preexisting spines (Fig. 6a and Extended Data Fig. 8), consistent with previous reports in the hippocampus[48,49] and somatosensory cortex[50]. When we analyzed the presynaptic axons of new spines, however, we found that the vast majority of new spines did not share a local axonal segment with any other spines within the EM volume (Fig. 6b–d). Thus, while we cannot formally reject the possibility that a subset of new spines shares a presynaptic neuron with another MRS through separate axonal branches whose branch point falls outside of our EM volume, the majority of new spines that form over motor learning likely represent new connections for the dendritic segment that are nevertheless functionally related to preexisting connections.

To corroborate these findings, we next characterized the activity patterns of spine pairs that are confirmed to share presynaptic axons through the EM analysis. We reasoned that these cases would provide a range of co-activity values expected for true cases of axon sharing, thus allowing us to infer the probability that new spine–MRS pairs share axons based on their activity. As expected, confirmed axon-sharing spines showed highly correlated activity (Fig. 6e), and events were overwhelmingly synchronized (Fig. 6e), supporting our categorization of such pairs using CLEM. Conversely, despite their relatively high co-activity rates, new spine–MRS pairs often showed activity in one spine that was absent in the other (Fig. 6e). Indeed, we found that ~88% (1,455/1,658 pairs) showed co-activity rates that were lower than the minimum rate observed in confirmed axon-sharing pairs (Fig. 6f), indicating that most new spines do not show co-activity rates consistent with sharing an axon with other spines on the same dendrite.

Discussion

Here, we present for the first time, to the best of our knowledge, a description of the functional properties of inputs to new synapses and their microenvironment during learning in vivo. Our results provide a mechanistic model of how spinogenesis during learning is regulated to contribute to learning-related reorganization of dendrites. In this model (Fig. 7), dendritic regions with a high density of task-related preexisting spines, especially those undergoing synaptic potentiation, become primed for spinogenesis, and extend filopodia to sample nearby preexisting axonal boutons, forming MSBs and temporarily sharing the axonal bouton with an allodendritic spine. Among the nascent spines forming new connections, those exhibiting high levels of co-activity with nearby task-related spines are selected for survival. As a result, the selected new connections contribute to the spatiotemporal clustering of learning-related inputs, thus affording these inputs disproportionate weights to robustly drive the postsynaptic neurons[19,51,52]. The importance of learning in this process is highlighted by the various differences observed in the no-task control, chief among them that new spines form less often, are less stable and show markedly reduced spatial and functional clustering with MRSs. Nonetheless, we also note that some of the observations in the no-task control show a similar trend to the learning condition. We interpret this result as representative of ongoing low-level plasticity—and possibly learning—that would be expected to occur in any animal for which behaviors are not constrained to a particular task.

While the majority of our analyses focused on all movements, limiting our analysis to rewarded movements yields similar results, suggesting that our results reflect the acquisition of the learned movement (Extended Data Fig. 9). Further, our results are robust to modifying the threshold for MRSs (Extended Data Fig. 10). We also provide data suggesting that most connections formed by new spines are not duplications of connections already represented on the parent dendrite, but the binding of new information streams into the dendritic repertoire. Even though it seems difficult for new spines to specifically target a new axon that is nonetheless functionally related to existing connections, filopodial exploration followed by subsequent activity-based selection provides a plausible mechanism to achieve this feat. It is tempting to speculate that extra-synaptic glutamate, for example, from spillover, could serve as a signal that guides filopodia to new candidate axons. Indeed, one study in brain slices has shown that the uncaging of glutamate near a dendritic shaft can be sufficient in certain contexts to drive new spine formation, and that the probability of successful spinogenesis can be increased with stimulation of nearby spines[24]. Of great future interest is a description of the brain regions of origin of the inputs synapsing on the new spines, as well as those synapsing on correlated MRSs, to understand how integration of information at the level of dendrites facilitates learning. Finally, the coordinated activity of new spines and nearby task-related spines preferentially encodes learned behaviors (Fig. 4), suggesting a privileged role of new spines in generating or shaping learned behaviors. Our data thus represent a synaptic substrate for the learning of motor behaviors.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-022-01086-6.

Received: 26 January 2021; Accepted: 26 April 2022;
Published online: 2 June 2022

References

1. Makino, H., Hwang, E. J., Hedrick, N. G. & Komiyama, T. Circuit mechanisms of sensorimotor learning. *Neuron* https://doi.org/10.1016/j.neuron.2016.10.029 (2016).
2. Hofer, S. B., Mrsic-Flogel, T. D., Bonhoeffer, T. & Hübener, M. Experience leaves a lasting structural trace in cortical circuits. *Nature* **457**, 313–317 (2009).
3. Xu, T. et al. Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* **462**, 915–919 (2009).
4. Fu, M., Yu, X., Lu, J. & Zuo, Y. Repetitive motor learning induces coordinated formation of clustered dendritic spines in vivo. *Nature* **483**, 92–95 (2012).
5. Iacaruso, M. F., Gasler, I. T. & Hofer, S. B. Synaptic organization of visual space in primary visual cortex. *Nature* https://doi.org/10.1038/nature23019 (2017).
6. Wilson, D. E., Whitney, D. E., Scholl, B. & Fitzpatrick, D. Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. *Nat. Neurosci.* **19**, 1003–1009 (2016).
7. Scholl, B., Wilson, D. E. & Fitzpatrick, D. Local order within global disorder: synaptic architecture of visual space. *Neuron* https://doi.org/10.1016/j.neuron.2017.10.017 (2017).

8. Kerlin, A. et al. Functional clustering of dendritic activity during decision-making. *Elife* https://doi.org/10.7554/eLife.46966 (2019).
9. Peters, A. J., Chen, S. X. & Komiyama, T. Emergence of reproducible spatiotemporal activity during motor learning. *Nature* **510**, 263–267 (2014).
10. Chen, S. X., Kim, A. N., Peters, A. J. & Komiyama, T. Subtype-specific plasticity of inhibitory circuits in motor cortex during motor learning. *Nat. Neurosci.* **18**, 1109–1115 (2015).
11. Marvin, J. S. et al. An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat. Methods* https://doi.org/10.1038/nmeth.2333 (2013).
12. Marvin, J. S. et al. Stability, affinity, and chromatic variants of the glutamate sensor iGluSnFR. *Nat. Methods* https://doi.org/10.1038/s41592-018-0171-3 (2018).
13. Hires, S. A., Zhu, Y. & Tsien, R. Y. Optical measurement of synaptic glutamate spillover and reuptake by linker optimized glutamate-sensitive fluorescent reporters. *Proc. Natl. Acad. Sci. USA* https://doi.org/10.1073/pnas.0712008105 (2008).
14. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
15. Kerlin, A. et al. Functional clustering of dendritic activity during decision-making. *eLife* https://doi.org/10.7554/eLife.46966 (2018).
16. Peters, A. J., Lee, J., Hedrick, N. G., O'Neil, K. & Komiyama, T. Reorganization of corticospinal output during motor learning. *Nat. Neurosci.* **20**, 1133–1141 (2017).
17. Takahashi, N. et al. Locally synchronized synaptic inputs. *Science* **335**, 353–356 (2012).
18. De Roo, M., Klauser, P. & Muller, D. LTP promotes a selective long-term stabilization and clustering of dendritic spines. *PLoS Biol*. https://doi.org/10.1371/journal.pbio.0060219 (2008).
19. Kastellakis, G., Cai, D. J., Mednick, S. C., Silva, A. J. & Poirazi, P. Synaptic clustering within dendrites: an emerging theory of memory formation. *Prog. Neurobiol.* **126**, 19–35 (2015).
20. Kleindienst, T., Winnubst, J., Roth-Alpermann, C., Bonhoeffer, T. & Lohmann, C. Activity-dependent clustering of functional synaptic inputs on developing hippocampal dendrites. *Neuron* https://doi.org/10.1016/j.neuron.2011.10.015 (2011).
21. Makino, H. & Malinow, R. AMPA receptor incorporation into synapses during LTP: the role of lateral movement and exocytosis. *Neuron* **64**, 381–390 (2009).
22. Makino, H. & Malinow, R. Compartmentalized versus global synaptic plasticity on dendrites controlled by experience. *Neuron* https://doi.org/10.1016/j.neuron.2011.09.036 (2011).
23. Hedrick, N. et al. Rho GTPase complementation underlies BDNF-dependent homo- and heterosynaptic plasticity. *Nature* **538**, 104–108 (2016).
24. Kwon, H.-B. & Sabatini, B. L. Glutamate induces de novo growth of functional spines in developing cortex. *Nature* **474**, 100–104 (2011).
25. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. R. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**, 761–766 (2004).
26. Cichon, J. & Gan, W.-B. Branch-specific dendritic Ca^{2+} spikes cause persistent synaptic plasticity. *Nature* **520**, 180–185 (2015).
27. Murakoshi, H., Wang, H. & Yasuda, R. Local, persistent activation of Rho GTPases during plasticity of single dendritic spines. *Nature* **472**, 100–104 (2011).
28. Spacek, J. & Harris, K. M. Three-dimensional organization of smooth endoplasmic reticulum in hippocampal CA1 dendrites and dendritic spines of the immature and mature rat. *J. Neurosci.* https://doi.org/10.1523/jneurosci.17-01-00190.1997 (1997).
29. Deller, T. et al. Synaptopodin-deficient mice lack a spine apparatus and show deficits in synaptic plasticity. *Proc. Natl. Acad. Sci. USA* https://doi.org/10.1073/pnas.1832384100 (2003).
30. Hwang, E. J. et al. Disengagement of motor cortex from movement control during long-term learning. *Sci. Adv*. https://doi.org/10.1126/sciadv.aay0001 (2019).
31. Conner, J. M., Culberson, A., Packowski, C., Chiba, A. A. & Tuszynski, M. H. Lesions of the basal forebrain cholinergic system impair task acquisition and abolish cortical plasticity associated with motor skill learning. *Neuron* **38**, 819–829 (2003).
32. Kawai, R. et al. Motor cortex is required for learning but not for executing a motor skill. *Neuron* **86**, 800–812 (2015).
33. Losonczy, A. & Magee, J. C. Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron* **50**, 291–307 (2006).
34. Niell, C. M., Meyer, M. P. & Smith, S. J. In vivo imaging of synapse formation on a growing dendritic arbor. *Nat. Neurosci*. https://doi.org/10.1038/nn1191 (2004).
35. Ziv, N. E. & Smith, S. J. Evidence for a role of dendritic filopodia in synaptogenesis and spine formation. *Neuron* https://doi.org/10.1016/S0896-6273(00)80283-4 (1996).
36. Toni, N. et al. Synapse formation on neurons born in the adult hippocampus. *Nat. Neurosci*. https://doi.org/10.1038/nn1908 (2007).
37. Dailey, M. E. & Smith, S. J. The dynamics of dendritic structure in developing hippocampal slices. *J. Neurosci*. https://doi.org/10.1523/jneurosci.16-09-02983.1996 (1996).
38. Lendvai, B., Stern, E. A., Chen, B. & Svoboda, K. Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex in vivo. *Nature* **404**, 876–881 (2000).
39. Zuo, Y., Lin, A., Chang, P. & Gan, W. -B. Development of long-term dendritic spine stability in diverse regions of cerebral cortex. *Neuron* **46**, 181–189 (2005).
40. Cruz-Martín, A., Crespo, M. & Portera-Cailliau, C. Delayed stabilization of dendritic spines in fragile X mice. *J. Neurosci.* **30**, 7793–7803 (2010).
41. Holtmaat, A. J. G. D. et al. Transient and persistent dendritic spines in the neocortex in vivo. *Neuron* **45**, 279–291 (2005).
42. Frank, A. C. et al. Hotspots of dendritic spine turnover facilitate clustered spine addition and learning and memory. *Nat. Commun*. https://doi.org/10.1038/s41467-017-02751-2 (2018).
43. Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E. & Svoboda, K. Spine growth precedes synapse formation in the adult neocortex in vivo. *Nat. Neurosci.* **9**, 1117–1124 (2006).
44. Yang, Y. et al. Selective synaptic remodeling of amygdalocortical connections associated with fear memory. *Nat. Neurosci*. https://doi.org/10.1038/nn.4370 (2016).
45. Lee, K. J. et al. Motor skill training induces coordinated strengthening and weakening between neighboring synapses. *J. Neurosci*. https://doi.org/10.1523/JNEUROSCI.0848-12.2013 (2013).
46. Toni, N., Buchs, P. A., Nikonenko, I., Bron, C. R. & Muller, D. LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature* https://doi.org/10.1038/46574 (1999).
47. Niculescu, D. et al. A BDNF-mediated push–pull plasticity mechanism for synaptic clustering. *Cell Rep*. https://doi.org/10.1016/j.celrep.2018.07.073 (2018).
48. Bloss, E. B. et al. Single excitatory axons form clustered synapses onto CA1 pyramidal cell dendrites. *Nat. Neurosci*. https://doi.org/10.1038/s41593-018-0084-6 (2018).
49. Chicurel, M. E. & Harris, K. M. Three-dimensional analysis of the structure and composition of CA3 branched dendritic spines and their synaptic relationships with mossy fiber boutons in the rat hippocampus. *J. Comp. Neurol.* https://doi.org/10.1002/cne.903250204 (1992).
50. Kasthuri, N. et al. Saturated reconstruction of a volume of nocortex. *Cell* https://doi.org/10.1016/j.cell.2015.06.054 (2015).
51. Govindarajan, A., Kelleher, R. J. & Tonegawa, S. A clustered plasticity model of long-term memory engrams. *Nat. Rev. Neurosci.* **7**, 575–583 (2006).
52. Poirazi, P. & Mel, B. W. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* **29**, 779–796 (2001).

Methods

Animals. All animal procedures were performed in accordance with guidelines set forth and protocols approved by the UCSD Institutional Animal Care and Use Committee and the National Institutes of Health (NIH). Mice (C57BL/6 for all experiments described in this study) were grouped housed in disposable cages with standard bedding in a temperature-controlled and humidity-controlled room (~21 °C and 42% humidity) with a reversed light cycle (10.00–22.00: dark). All experiments were performed during the dark cycle. After surgeries, animals were singly housed. Males and females were randomly used for surgeries, with no selection criteria other than surgery outcome.

Surgery. Adult mice (6 weeks or older) were anesthetized with isoflurane in an enclosed, ventilated chamber (5% isoflurane with a constant flow rate of 1 l min^{-1} at 0.1 bar and 21°C) until a deep plane of anesthesia was reached, as indicated by low muscle tone and slowed breathing rate. Enrofloxacin (Baytril; 10 mg per kilogram body weight) and dexamethasone (2 mg per kilogram body weight) were injected subcutaneously to prevent infection and brain swelling, respectively. Skin and connective tissue over the dorsal surface of the skull was removed, the skull was slightly scored with a scalpel and a custom stainless-steel headplate was glued to the skull surface. A craniotomy (~3 mm diameter) was performed, as previously described[9,10], over the right caudal forelimb area around the central coordinate of ~300 µm anterior and ~1,500 µm lateral from bregma. Viruses (AAV1-CMV-PI-CRE and AAV1-Syn-FLEX-GCaMP6f from Addgene/UPenn Vector Core; AAV1-Syn-FLEX-SF-iGluSnFR-A184S, construct generously received from L. Looger) were diluted to achieve sparse expression of iGluSnFR or GCaMP6f (1:1 mixture of iGluSnFR or GCaMP6f and 1:5,000–10,000 dilution of Cre in saline + 0.5% FastGreen for visualization of injections) and injected into the region of the caudal forelimb area of the exposed cortex using beveled glass pipettes (~12–25 µm inner diameter). Each injection consisted of a ~20 nl volume at a depth of ~250 µm from the pial surface to target layer 2/3. Injection volumes were dispensed over the course of ~2 min. Multiple (3–5) injections were performed in each craniotomy, separated by at least 500 µm. Pipettes were left in the brain for 4 min after injection to avoid backflow. Chronic imaging windows consisting of a 3-mm diameter plug glued to a larger (~5 mm) glass base were then implanted in the craniotomy. The window was held in place with gentle pressure while the edges were affixed to the skull with small amounts of surgical glue (VetBond). Buprenorphine was injected subcutaneously at the end of surgery for pain management.

Water restriction. Animals were allowed to recover from surgery for ~10–14 d, after which they were progressively water restricted (2 ml per day for 3 d, 1.5 ml for 3 d, then 1 ml per day for the remainder) for ~14 d. Weight was constantly monitored to ensure loss of no more than 30% of starting body weight.

Behavior. After water restriction, mice were trained in the lever-press task for 14 d. Simultaneous two-photon imaging was performed on sessions 1–3 ('early'), 6–8 ('middle') and 11–13 ('late'). The lever comprised a piezoelectric flexible force transducer (LCL-113G, Omega Engineering) attached to a 1/14-mm-thick brass rod. The lever position was continuously recorded using a data acquisition device (LabJack) and software (Ephus, MATLAB, MathWorks) working with custom software running on LabVIEW (National Instruments), which monitored threshold crossing. The behavioral setup was controlled by MATLAB software (Dispatcher, Z. Mainen and C. Brody) communicating with a real-time system (RTLinux). A 6-kHz tone was presented to indicate a period during which a lever press was rewarded with water (~8 µl per trial) paired with a 500 ms,12-kHz tone, followed by an intertrial interval of 8–12 s. Successful lever presses were defined as those crossing two thresholds ~1.5 mm and ~3 mm below the resting position) within 200 ms. The 3-mm threshold defined the target lever displacement, while the 1.5-mm threshold ensured that the mouse did not hold the lever near the target threshold. Failure to perform a successful press during cue presentation triggered a white noise punishment signal and the start of the next intertrial interval. 'Non-cued' presses during the intertrial interval were neither rewarded nor punished. Mice were exposed to 100 trials each day, or until the mouse became disengaged (no movements for 20+ trials) or satiated (no licking in response to water delivery).

In vivo two-photon imaging. Imaging was performed using a commercial two-photon microscope (B-Scope, ThorLabs) equipped with a ×16/0.8-NA objective (Nikon) and a Ti-Sa laser (MaiTai, Newport) tuned to 925 nm (or 810 nm for control experiments). The laser power coming through the objective was controlled with a Pockel's cell, and ranged from 10 to 40 mW for these experiments. Image acquisition was controlled through ScanImage software. Imaging was always performed in awake animals. Images (256 × 512 pixels at either ×8.5 or ×12.1 zoom, corresponding to ~60 × 120 µm and 42 × 85 µm, respectively) were recorded at approximately 58.3 Hz in 5-min-on, 5-min-off intervals for the duration of the behavioral session. Such interleaved imaging was performed so as to limit phototoxicity to the dendrites. The median number of trials imaged per day using this method was 60 (95% CI = (57, 62)). Further, imaging for a single field was performed in 5-d intervals, with three fields being selected for each animal, such that field 1 was imaged on sessions 1, 6 and 11, field 2 on sessions 2, 7 and 12, and field 3 on sessions 3, 8 and 13. In our hands, such an imaging schedule preserved the health of most dendrites, preventing obvious dendritic blebbing and/or photo-bleaching. Separate fields were always at least 500 µm apart, taking advantage of the multiple injection sites. In 4 of the 45 iGluSnFR fields, imaging was performed only on early and late sessions.

A small subset of imaging fields were processed for subsequent EM. In these cases, a high-resolution, low-zoom (1,024 × 1,024 pixels at ×1 and ×5 zoom, corresponding to ~1 mm × 1 mm and 250 × 250 µm fields, respectively) *z*-stacks of the target region were acquired from the pial surface to the target soma at 1-µm intervals.

Identification of target dendrites. Fluorescent dendrites visible near the pial surface (presumed layer 1) were typically observed within the first month after viral injections. Dendrites in the superficial layers were targeted, with an average imaging depth of 35.1 ± 0.9 µm (mean ± s.e.m.). As the injections were targeted to L2/3 of M1, expression of a reporter construct should be enriched in L2/3 excitatory neurons. Because expression in L5 is also possible, however, we took care to trace target dendrites in L1 back to their parent apical dendrites and follow the apical branch back to its parent soma to assess the laminar location of the cell whenever possible. In addition to their markedly different laminar depths, the apical dendrites of L2/3 excitatory neurons are also morphologically distinct from those of L5 excitatory neurons, with L2/3 typically presenting a more 'shrub-like' appearance (shorter primary apical dendrite with long, tortuous higher-order branches) versus the 'tree-like' appearance of L5 neurons (long primary apical dendrite with more planar arborization/'tuft' in L1). Furthermore, a previous study[53] has shown that the spine density of L2/3 excitatory neurons (~0.73 per µm) is higher than that of L5 excitatory neurons (~0.47 per µm). Consistent with accurate targeting of L2/3 dendrites, the median spine density from the data shown in this paper is ~0.67 per µm.

Field identification across longitudinal imaging sessions. On the first imaging session for each field (that is, sessions 1–3), bright-field images of the surrounding vasculature were acquired, and the depth of FOV from the pial surface was recorded, allowing landmark-driven identification of the same field in future sessions. On subsequent sessions, the single-plane average projection image of the time series acquired on the first session was used as reference to frame the imaging field as similarly as possible. Brief (~10–20 s) time series were first acquired, subjected to motion correction (see description below) and projected so as to provide a comparison image against the previous session image. Care was taken to adjust the *z*-plane so as to maximally reproduce the appearance of the reference image before commencing the experiment.

Evaluation of dendritic health. Due to the risk of phototoxic effects from imaging, dendritic health was carefully monitored. The punctuated imaging schedule used in this study (Extended Data Fig. 1c) was designed through pilot experiments in which imaging power and duration were decreased until dendritic health was maintained through the end of the experiment. Dendritic health was evaluated based on: (1) spine density (analysis detailed below), with large decreases signaling poor cellular health; (2) dendritic morphology, with any 'bleb'-like structure reflecting cellular death; and (3) fluorescent event frequency (detailed below), with global decreases potentially indicating damage. In our data, median spine densities are not different between early (0.67 spines per µm) and late (0.61 spines per µm) sessions (Extended Data Fig. 1e). Event frequencies were largely stable, showing similar values for MRSs across early and late sessions, with a small decrease apparent in nonMRSs. Taken together, these data suggest that dendritic health was maintained in the data presented in this study.

Movement analysis. Movement analyses were performed as previously described[9,10]. Briefly, lever displacement traces (voltage recordings from the force transducer) were downsampled from 10 kHz to 1 kHz, then filtered using a four-pole 10 Hz low-pass Butterworth filter, after which the velocity of the lever was determined by smoothing the difference of consecutive points with a moving average window of 5 ms. The envelope of the lever velocity was then extracted using a Hilbert transform, and movement bouts were defined by the envelope crossing a threshold of 4.9 mm per second. Each movement bout was extended by 75 ms on either side. Bouts separated by less than 500 ms were considered continuous. Movement start and end times were defined as the points at which the lever exceeded or fell below the thresholds defined by rest periods before and after the movement bouts. Thresholds were defined as the resting position plus the 99th percentile of the noise distribution, in turn defined as the difference between the Butterworth smoothed trace and the original trace. For reaction time analysis, the trials in which mice were moving the lever within 100 ms before the cue start time were excluded from analysis. The median number of trials excluded for this purpose was 28.4 (95% CI = (23.8, 32.7)), corresponding to 35% (95% CI = (32.2, 37.9)) of trials.

Movement correlations within sessions were calculated using the median of all pairwise correlations of rewarded movements that started after cue onset within a single session. Movement correlations across sessions were found using the median correlations of all possible pairs of movements between sessions (one movement

taken from one session and the other movement taken from the other session). The movement correlation heat map shown in Fig. 1d was computed by taking the mean within-session and across-session correlation computed above for each animal.

For correlations with the learned movement pattern, the learned movement pattern was defined as the average of rewarded movements that started after cue onset from the late (11–14) learning sessions. Movements coincident with new spine–MRS pair co-activity were defined as any movements executed during the session (that is, not limited to rewarded movements) that overlapped with co-activity. Prolonged movements (lasting >3 s) typically corresponded to repeated movements in succession, and were therefore excluded from this analysis.

Image analysis. Lateral motion of imaging time series was corrected using custom full-frame cross-correlation image alignment[54]. Motion within each frame was negligible due to the fast frame rate. To register fields across sessions, the average projection images of each of the time series corresponding to a particular imaging field were subjected to image alignment. To assist with spine categorization and new spine identification, duplicates of the session-aligned images were iteratively deconvolved (Diffraction PSF 3D; FIJI/ImageJ). ROIs were manually drawn using custom MATLAB software. For dendritic spines, elliptical ROIs were drawn around the center of the spine head beyond the edge of detectable fluorescence above background (Extended Data Fig. 4). Apparent spines along the z axis of the dendritic shaft were included in analysis. Series of regularly spaced elliptical ROIs were also drawn along the length of the dendrite, with the center of each ellipse serving as the point along a poly-line being used to calculate dendritic distance between spines. A single, large ROI was drawn in an empty region of the field to estimate background.

For display purposes in figures, images were manually cropped around dendrites of interest for visual clarity. Care was taken (using fluorescence traces as reference) to ensure that no structures belonging to the dendrite of interest were removed in the process.

Data exclusion. We attempted to capture 3 imaging fields for each of the 23 animals used in this experiment. In a subset of animals, fewer than 3 fields were fully captured, resulting in a total field count of 61.

Dendrites that presented any signatures of poor health (see 'Evaluation of dendritic health'), such as blebbing, significant bleaching and/or globally reduced/absent activity, at any point during the experiment were excluded from all analyses. Of the 61 imaging fields acquired for this dataset, 4 fields (6.6%) were excluded due to poor dendritic health.

Fields that showed deteriorated optical quality (typically characterized by higher background and lower visibility of the dendrite, as well as cellular debris in the immediate environment of the target dendrite) or had newly visible structures blocking the target dendrite, were excluded if dendritic spines could not be confidently characterized. This resulted in the exclusion of an additional 12 fields, yielding 45 total fields used in this study.

Fluorescence analysis. Fluorescence time series were produced by averaging the pixels within each ROI for all imaging frames. The time-varying baseline (F_0) of a fluorescence trace was estimated by smoothing inactive portions of the trace, using a previously described iterative procedure[9]. Briefly, this process identified 'active' and 'inactive' portions of the trace, removing active portions and using the LOESS-smoothed inactive portions (interpolated across active periods) to estimate the time-varying baseline. The normalized $\Delta F/F_0$ trace was then calculated, where ΔF was found by subtracting the baseline trace from the raw trace, and F_0 is the calculated time-varying baseline.

Activity events were detected based on previously described methods[9]. Briefly, noise was estimated for each $\Delta F/F_0$ trace as the standard deviation of negative fluorescence values mirrored about the origin. This noise estimate was then used to set two thresholds, one being 2× the noise to find active portions of the trace, and another being 1× the noise to define the baseline. Active portions of the trace were defined as when the 1-s LOESS-smoothed $\Delta F/F_0$ trace crossed the active threshold and extended backwards to begin when the baseline threshold was crossed by the unsmoothed trace. Binarized traces with the value of 1 for active frames and zero otherwise were then produced for each ROI. Such binarized traces were used for all co-activity analysis and event frequency calculations. Using this approach, we found that the median event rates for each spine type were as follows: MRSs, early: 5.3 events per min (95% CI = (5.1, 5.6)), late: 5.1 events per min (95% CI = (4.8, 5.4)); late versus early MRSs: $P = 0.46$); nonMRSs, early: 5.0 events per min (95% CI = (4.7, 5.3)), late: 4.4 events per min (95% CI = (4.2, 4.6); early versus late nonMRS: $P = 7 \times 10^{-7}$) and new spines: 4.9 events per min (95% CI = (4.4, 5.8)).

For structural analysis, average projection images of the entire motion-corrected time series were produced. To estimate spine volume in each session, the integrated fluorescence intensity of pixels with intensity values above background (the average pixel intensity across the designated background ROI) over a given spine ROI was divided by the average fluorescence intensity of the nearby region of dendrite for normalization. Such normalization should account for global changes in the expression level of the sensor. Local dendritic fluorescence intensity was estimated by using the dendritic ROIs (described above) within 5 µm of the base of the spine.

Activity onset analysis. To estimate the timing of the activity onset of individual spines during movements, the fluorescence traces of each spine were aligned to the onsets of movements overlapping with the activity of the spine. We only considered movements that did not have another movement within 1 s before movement onset and 2 s after movement onset to avoid contamination of activity related to other movements. Activity averaged across movements was then used for peak detection, as follows: first, peak activity was defined using the 'findpeaks' function (MATLAB) in a 2-s window starting 1 s before and ending 1 s after movement onset, with a minimum distance between peaks set at 0.5 s, and a minimum peak height corresponding to the median + standard deviation of the full 3-s peri-movement period being inspected. If multiple peaks were found, each was given a score accounting for both the amplitude of the peak as well as the temporal proximity to movement onset ((1/abs(peak timing − movement onset timing) × peak amplitude), and the peak with the highest score was considered as the target peak activity. After peaks were identified, the velocity of the activity trace, defined as the first derivative of the robust-loess-smoothed activity trace (10 frames per ~170-ms window), was used to identify changes in the slope of the activity from negative to positive, allowing the identification of the beginning of rising phases of activity. Onset timing was then defined by searching backwards in time from the end of the rising phase of the target peak (defined as 75% of the target peak amplitude) to one of two criteria: either (1) when the activity velocity trace fell below zero, corresponding to the start of the rising phase, or (2) when the smoothed activity fell below the median of the full 3-s window of of peri-movement activity.

Spine structural classification. Average projections of time series from each session were registered with respect to the first imaging session for that field (described in 'Image analysis') to allow for comparison across days. A duplicate set of these images were iteratively deconvolved with 'Iterative Deconvolve' plugin for ImageJ) as a guide for spine detection. We excluded spines that were too close to each other to be accurately separated for fluorescence trace extraction in the original two-photon image series projection. Any visible dendritic protrusions emanating from the dendrite were considered putative spines. Bright, punctate regions of at least 0.5 µm in diameter overlapping with the dendrite in the imaging plane were also considered spines. Both assumptions were corroborated with subsequent EM reconstructions of a subset of imaged dendrites (for example, Extended Data Figs. 4b and 7g). Spines that appeared in the same location across sessions, or whose neck originated from the same dendritic region were considered to be the same spine. The spines that were represented across all sessions in this way were considered stable, 'preexisting' spines. Spines that were no longer visible in later imaging sessions (eliminated spines) were only considered for early session analyses, such as the analysis of early MRS density. Dendritic protrusions in later sessions that were not present in previous sessions were considered new spines. New spines that were present only for the 'middle' learning sessions were considered 'transient' new spines, and were used only for the analyses in Fig. 3, where specifically indicated. Spines that transiently disappeared in the middle session and then reappeared in the late session were rare, and were excluded from analysis.

Using the above approach, we found that the overall spine density does not significantly change between early (0.67 spines per µm) and late (0.61 spines per µm) sessions ($P = 0.19$, rank-sum test; Extended Data Fig. 1d).

Movement-related classification. Spines were classified as movement-related on each individual session, as previously described[9]. Briefly, the dot product of binarized lever traces (movements versus non-movements, as detailed above) and continuous $\Delta F/F_0$ traces was calculated for each spine. This value was then compared to the dot products when shuffling the movement periods 10,000 times. The dot products of each of the shuffled traces with $\Delta F/F_0$ traces were then compared to the values of the actual data. Actual values that were above the 97.5th percentile of the shuffled distribution were considered 'movement-related'.

Distance analysis. In all analyses regarding the 'distance from new spine', a given spine's distance value corresponded to the dendritic distance (as determined from single-plane two-photon images) from the base of the spine to base of the nearest new spine. Note that this dendritic distance differs from Euclidian distance (direct distance between two points in the plane) in that the curvature of dendrites was considered. For spine density and volume change probability analyses, if multiple new spines were present on a dendrite, only the nearest new spine was considered. The territory of each new spine was bounded at the halfway point between the new spines.

For analysis of the density of MRSs surrounding new spines, we focused our analysis on dendrites that showed formation of at least one new spine. For each spine, any MRSs that fell within 10 µm from the spine in either direction along the dendrite were considered. The number of MRSs was then divided by the dendritic distance considered (20 µm in most cases, but occasionally less when a new spine is within 10 µm of the edge of a dendrite). Chance estimates of the number of nearby MRSs were performed by randomizing each new spine location across all dendrites used for analysis in Fig. 3b (that is, those that showed at least one new spine) 10,000 times. Randomized locations were assigned to dendrites with a spatial resolution of 0.5 µm and using the full length of the imaged dendrite. For

example, a 45-µm-long dendrite could have a randomized new spine location at 0 µm, 0.5 µm, 1 µm, 1.5 µm and so on, up to 45 µm. As before, if a simulated new spine was within 10 µm of the edge of the simulated dendrite, then only the true total dendritic distance considered was used to calculate MRS density.

For functional spine density measurements, the total numbers of either MRSs or nonMRSs on a single dendrite were counted with respect to the closest new spine, and divided by the total distance measured along the imaged dendrite. Boundaries for each new spine were defined as the extent of the current distance bin being measured (successive 5-µm steps in either direction; see below), the edge of a dendrite, or—in cases where multiple new spines were present on a single dendrite—the halfway point to the nearest, other new spine. Any bins that corresponded to less than 5 µm in total length (for example, when multiple new spines were close to the edge of a dendrite) were excluded from analysis.

All distance values were binned in 5-µm increments to simplify visualization. Bins correspond to $dBin_n < d \leq dBin_{n+1}$, such that the 2.5-µm bin represents interspinal distance values from 0 to 5 µm, the 7.5-µm bin corresponds to $5 < d \leq 10$ µm, the 12.5-µm bin to $10 < d \leq 15$ µm, and so on.

Co-activity analysis. Co-activity rates between all possible spine pairs in a single field were calculated using binarized event traces (binarization process defined in 'Fluorescence analysis') for the entire imaging session. All periods where activity events were present in both spines (that is, frames for which both binarized activity traces are logical true) were considered co-active periods. A single co-activity event was defined as the entire duration that both spines were continuously co-active. The co-activity rates were calculated as the number of such co-active events per unit time. Co-activity rates were then normalized to the geometric mean of the activity frequencies of both spines. Because the geometric mean is highly correlated with the calculated co-activity rates (Extended Data Fig. 3e), normalization by this value allows better comparison between the relative co-activity between spine pairs showing different overall frequencies. All co-activity rates were calculated across the entire trace (that is, in both movement and non-movement periods, as well as across all trial epochs, including intertrial intervals and cue periods). Analyses for Fig. 4d,e were performed by removing co-active events from the traces of both constituent spines in each new spine–MRS pair, producing 'new spine-only' and 'MRS-only' activity traces. Such traces reflect periods when one spine is active while the other spine in the pair is silent, allowing the differentiation of coherent and desynchronized activity for a given spine pair. The fraction of events of each event 'type' (that is, new spine-only activity, MRS-only activity and co-activity) occurring during movement periods (defined in 'Movement analysis') thus represents the specificity of each signal to movement.

Simulation of spine activity. Spine activity was simulated to provide proof-of-principle data illustrating that the geometric mean of activity event frequencies scales better with co-activity rates than does the arithmetic mean (Extended Data Fig. 3). Binarized event traces were simulated by assigning random activity blocks a value of '1'. The durations of these simulated events were assigned by randomly sampling from the durations of the real spine imaging data. Simulated event frequencies ranged from 0 to 25 events per min. The timing of simulated events was random, with the constraint that no activity block overlaps another. If such overlap occurs, the timing of events is randomized until the criterion was met. Co-activity of all possible simulated spine pairs was defined as elsewhere: whenever two spines' binarized activity event traces were equal to 1, this was considered a co-activity period.

New spine density analysis. New spine density was calculated as the total number of new spines formed on a given dendrite normalized by the dendritic length. Normalizing the new spine number to the total number of preexisting spines on the analyzed dendrites (which we found to be tightly correlated to the dendritic length; $r = 0.83$, $P = 1 \times 10^{-35}$; Pearson's correlation coefficient) yielded nearly identical results (data not shown).

Correlated light and electron microscopy. *Sample preparation.* After completion of the final imaging session, data were analyzed so as to identify dendrites that showed new spine formation. After selection of dendrites of interest, animals were administered a retro-orbital injection of fluorescent dextran (FITC-Dextran; Sigma-Aldrich 52194) and lectin (Tomato Lectin, DyLight 594; Vector Laboratories DL-1177) dyes to visualize vasculature, with lectin permanently marking the vessels following transcardial perfusion.

Mice were then anesthetized with an intraperitoneal injection of ketamine/xylazine and transcardially perfused with a brief flush of Ringer's solution containing heparin and xylocaine, followed by approximately 80 ml of 0.5% glutaraldehyde/4% prilled paraformaldehyde in 0.15 M sodium cacodylate buffer containing 2 mM calcium chloride ('caco'). The brain was removed from the cranium and post-fixed for approximately 1 h on ice in the same fixative. The brain was then manually dissected with a razor blade to allow for horizontal sectioning of the cortex on a vibratome (Leica VT1000S). The vibratome blade was brought as close to the cortex as possible by eye and 150-µm-thick sections were then cut. The area under the imaging window was usually captured in one or two sections.

Transmitted light images of the sections were collected with a dissecting scope at low magnification to reveal the vasculature and the images were aligned relative to each other in Photoshop (Adobe). By comparing this map with the images of the vasculature taken before perfusion, it was possible to accurately locate the area of interest within a particular vibratome slice. The slice was then stained for 1 h on ice with DRAQ5 (Biostatus) diluted 1:1,000 in caco. The slice was washed three times in caco. The area of interest was located and confocal volumes collected using ×20 and ×60 water objectives on an inverted microscope (FluoView, Olympus).

The slice was post-fixed overnight at 4 °C in caco containing 2.5% glutaraldehyde. The slice was then washed with solutions of caco and caco containing 100 mM glycine. The slice was stained with the following series of solutions, thoroughly washing with distilled water after each step: 2% osmium tetroxide/1.5% potassium ferrocyanide in caco for 1 ho at room temperature (r.t.), 0.5% aqueous thiocarbohydrazide for 30 min at r.t., 2% aqueous osmium tetroxide for 1 h at r.t., 2% aqueous uranyl acetate overnight at 4 °C and Walton's lead solution for 30 min at 60 °C. The slice was then dehydrated with the following series of solutions, at 10 min for each step: 70% ethanol, 90% ethanol, 100% ethanol, 100% ethanol, dry acetone, dry acetone. The slice was placed into a 50:50 acetone:Durcupan ACM solution (Sigma-Aldrich) overnight on a rotator. The Durcupan was made with 11.4 g component A, 10 g component B, 0.3 g component C and 0.1 ml component D. The slice was placed into fresh 100% Durcupan in a vacuum chamber for two consecutive nights and then flat-embedded in Durcupan between two glass slides coated with liquid release agent (Electron Microscopy Sciences), using pieces of Aclar 33C as spacers to prevent crushing tissue. The Durcupan was cured at 60 °C for 48–72 h.

The slice was mounted on the end of a small aluminum rod and a low-resolution microCT volume (~2.6 mm pixel size) was collected at 80 kV (Zeiss Versa 510 XRM). The vasculature pattern was used to locate the area of interest in the embedded section. The block was trimmed down to less than 1 mm × 1 mm in size, mounted to a serial block-face EM (SBEM) specimen rivet using conductive silver epoxy (Ted Pella), and left at 60 °C overnight. Following trimming of the block using an ultramicrotome, a higher-resolution microCT scan was collected (~1 mm pixel size) to allow for precise targeting of the SBEM stage to the ROI. An SBEM volume was collected on either a Zeiss Merlin SEM or a Zeiss Gemini 300 SEM equipped with a Gatan 3View and OnPoint backscatter detector system. The SBEM volumes were collected at 2.5 kV EHT with 5 nm *XY* pixels, 50 nm *Z* steps and 1 ms dwell time. Because volumes were often very near the surface of the brain, necessitating imaging of areas of empty resin, focal charge compensation with nitrogen gas was used to eliminate charging artifacts.

The 'tiltxcorr' program of IMOD[55] was used to generate the final SBEM volume by applying cross-correlation to eliminate any minor jitter between slices. The SBEM and confocal volumes were co-registered with the 'Landmark Image Warp' module in Amira (versions 2019/2020), using nuclei and other distinctive features visible in both modalities as landmarks.

EM segmentation. Structures of interest in the EM volumes were manually segmented using IMOD software (version 4.9)[55] running on Cygwin Terminal (version 3.3). Segmentation was performed on binned (5–10 pixels in *xy*) images to reduce computer memory load. Target dendrites were first identified using the overlaid confocal-EM images, and the accuracy of the overlaid image was then confirmed by tracing 'landmark' spines and dendritic features that were particularly obvious in in vivo images (for example, large, solitary spines, dendritic curves or bifurcations, and so on). Individual contours were drawn for each spine, dendrite, axon and subcellular structure (for example, spine apparatus) using the sculpt tool. All visible protrusions from the target dendrite were segmented, irrespective of their visibility in in vivo images. The segmentation of each structure was evaluated by three individuals.

EM synapse classification. For a given spine, the presynaptic axon was identified based on the presence of an apparent synapse between the two structures. Synapses were defined based on (1) the presence of apparent PSD, that is, a darkened band at the membranous edge of (primarily the head of) dendritic spines visible across at least 2–3 sections, (2) the presence of a vesicle-housing axonal bouton immediately opposed to the PSD, (3) the collection of vesicles within the bouton around the putative synaptic site, that is, directly opposed to the PSD across at least 2–3 slices, and (4) a synaptic cleft, appearing as a small but distinct space between the presynaptic and postsynaptic membranes. In some instances, multiple axons formed apparent synapses on a single spine. In most of these cases, one of the two axons, when traced, connected primarily to the shaft of other dendrites, suggesting that these instances likely represent mostly inhibitory axons. These instances were also concomitant with less defined PSDs, consistent with inhibitory synapses. We did not have the spatial resolution in EM to categorize the vesicular shape of such inputs, and therefore cannot conclusively identify them as inhibitory. However, all instances of such axons were fully traced, and were considered as viable candidates in our 'axon-sharing' analysis. However, we found no instances of such axons connecting with multiple spines on the same dendrite.

EM filopodia classification. Putative filopodia were categorized within EM images as dendritic protrusions that were at least 1 µm in length and lacked an apparent synapse (as described above). Filopodia identified in this way were then compared

to in vivo two-photon image series (deconvolved, average projection images from early, middle and late learning sessions) to see if they corresponded to spines identified in vivo that might have shrunken or been classified as 'eliminated' by shrinking below a detectable threshold. In general, spines classified as eliminated were also absent in EM, and we only encountered one instance of a putative filopodia being in the location of a spine classified as eliminated. We excluded this case from analysis.

EM distance calculations. Dendritic distances were calculated in EM based on open objects drawn in the center of visible dendritic portions across and within slices. The resulting line segments were then visually inspected to ensure that they roughly corresponded to the center of mass of the dendritic branch along its length. Spine and filopodia locations were logged as the point at which the center of mass of the 'neck' of the structure merged with the dendrite.

EM spine volume calculation. Spine volume calculations in EM (Extended Data Fig. 4e) were made in IMOD software by taking the area of contours drawn around individual dendritic spine heads multiplied by the *z*-step size (typically 4 nm) to obtain volume.

Spine apparatus classification. Spine apparatus were defined as densely stained (dark), typically laminar structures that invaded the spine head and/or the spine neck and spanned a majority of the slices covered by the spine. Such spine apparatus were frequently observed to connect with apparent endoplasmic reticulum-like structures in the parent dendrite, which aided in classification. All spine apparatus classifications were made while blinded to the spine volume change conditions calculated in in vivo images.

Estimation of chance distances between new spines and filopodia. The estimation of the chance level of filopodia appearing nearby new spines was performed as follows: for a given number of filopodia on a particular dendrite, the location of each was randomized, finding the 'shuffled' location's distance to the nearest new spine. This was repeated for all the data 10,000 times, and the median new spine–filopodium distance was calculated for each shuffle. The *P* value was defined as the number of such values that were less than or equal to the median value of the real data, as a fraction of the total shuffles. This *P* value thus represents the fraction of simulated occasions that the null hypothesis was supported (that is, that filopodia are not closer to new spines than chance).

Statistics and reproducibility. For all sets of experiments described here, pilot experiments were run to assess the intrinsic statistics of our readouts, as well as associated failure rates. Otherwise, no formal statistical method was used to predetermine sample sizes. However, our sample sizes are similar to those reported in previous publications. Data were excluded only in cases of poor optical quality in imaging experiments or when cellular health was deemed poor (see 'Evaluation of dendritic health'). Animals in this study were not selected for allocation into experimental groups based on any other prerequisite features other than general wellbeing. Control experiments were performed on randomly selected mice and interleaved with learning-group mice. Investigators were blinded for all EM reconstructions (that is, the properties of interest of spines were not known during their reconstruction) and ROI drawing for in vivo images (that is, the properties of interest that might be affected by ROI drawing are not yet known at the time of drawing), but were otherwise not blinded to allocation during experiments and outcome assessment. Non-parametric statistics were used when possible to avoid assumptions of data normality. Multiple comparisons were corrected for using the FDR method. All tests performed were two-sided. Sample sizes (*n*) are as follows: total mice: 23 iGluSnFR; 30 GCaMP6f; 5 iGluSnFR (imaged at 810 nm); 14 iGluSnFR (no-task controls); number of fields: 45 iGluSnFR, 66 GCaMP6f; 17 iGluSnFR (imaged at 810 nm); 25 iGluSnFR (no-task controls) number of dendrites: 76 iGluSnFR, 137 GCaMP6f; 33 iGluSnFR (imaged at 810 nm); 40 iGluSnFR (no-task controls); total number of unique imaged spines (iGluSnFR): 1,915 (median 23 spines per dendrite); 484 iGluSnFR (imaged at 810 nm); 1,787 iGluSnFR (no-task controls); total number of early session spines (iGluSnFR): 1,767; total number of middle-session spines (iGluSnFR): 1,582; total number of late-session spines (iGluSnFR): 1,656; number of early MRSs: 898; number of middle-session MRSs: 836; number of late MRSs: 820. Number of mice showing new spines: 21/23 iGluSnFR, 25/30 GCaMP6f; number of dendrites showing new spines: 50/76 iGluSnFR, 62/137 GCaMP6f; total number of new spines: 118 iGluSnFR; 140 GCaMP6f; 51 iGluSnFR (no-task control). *N* value in figures regarding new spines only represents animals/dendrites that show new spine formation. Summaries of spine numbers, including MRS and new spine counts, can be found in Supplementary Tables 1 (learning group) and 2 (no-task control group).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data are available upon request from the corresponding authors.

Code availability

Code used to analyze data and generate figures for this paper are available upon request from the corresponding authors.

References

53. Tjia, M., Yu, X., Jammu, L. S., Lu, J. & Zuo, Y. Pyramidal neurons in different cortical layers exhibit distinct dynamics and plasticity of apical dendritic spines. *Front. Neural Circuits* **11**, 43 (2017).
54. Mitani, A. & Komiyama, T. Real-time processing of two-photon calcium imaging data including lateral motion artifact correction. *Front. Neuroinform.* https://doi.org/10.3389/fninf.2018.00098 (2018).
55. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

Acknowledgements

We thank K. O'Neil, Q. Chen, O. Arroyo, L. Hall, P. Yao and E. Hall for technical assistance; M. Banghart, B. Bloodgood and members of the laboratory of T.K., especially A. Ramot, for discussion; and E. Gjoni for assistance with EM analysis. This research was supported by grants from NIH (R01 NS091010, R21 NS112750, R21 NS109722, R01 DC014690, R01 EY025349 and P30 EY022589), Pew Charitable Trusts and David & Lucile Packard Foundation to T.K.; NIH (F32 NS103267 and K99 NS114175) to N.G.H.; U24 NS120055, R01 DA038896, R24 GM137200 and R01 GM082949 to M.E.; and U01 MH114829, R01 DA049787 and R01 NS121231 to B.K.L.

Author contributions

This work was conceived by N.G.H. and T.K. All in vivo experiments were performed by N.G.H. and Z.L. with assistance from Y.M. and S.J. and analyzed by N.G.H. and T.K. CLEM was performed by N.G.H, Z.L., E.B. and M.E., and analyzed by N.G.H., S.S. and P.N. B.K.L. provided viral vectors. The paper was written by N.G.H. and T.K. with inputs from all authors.

Competing interests

The authors declare no competing interests.

Additional information

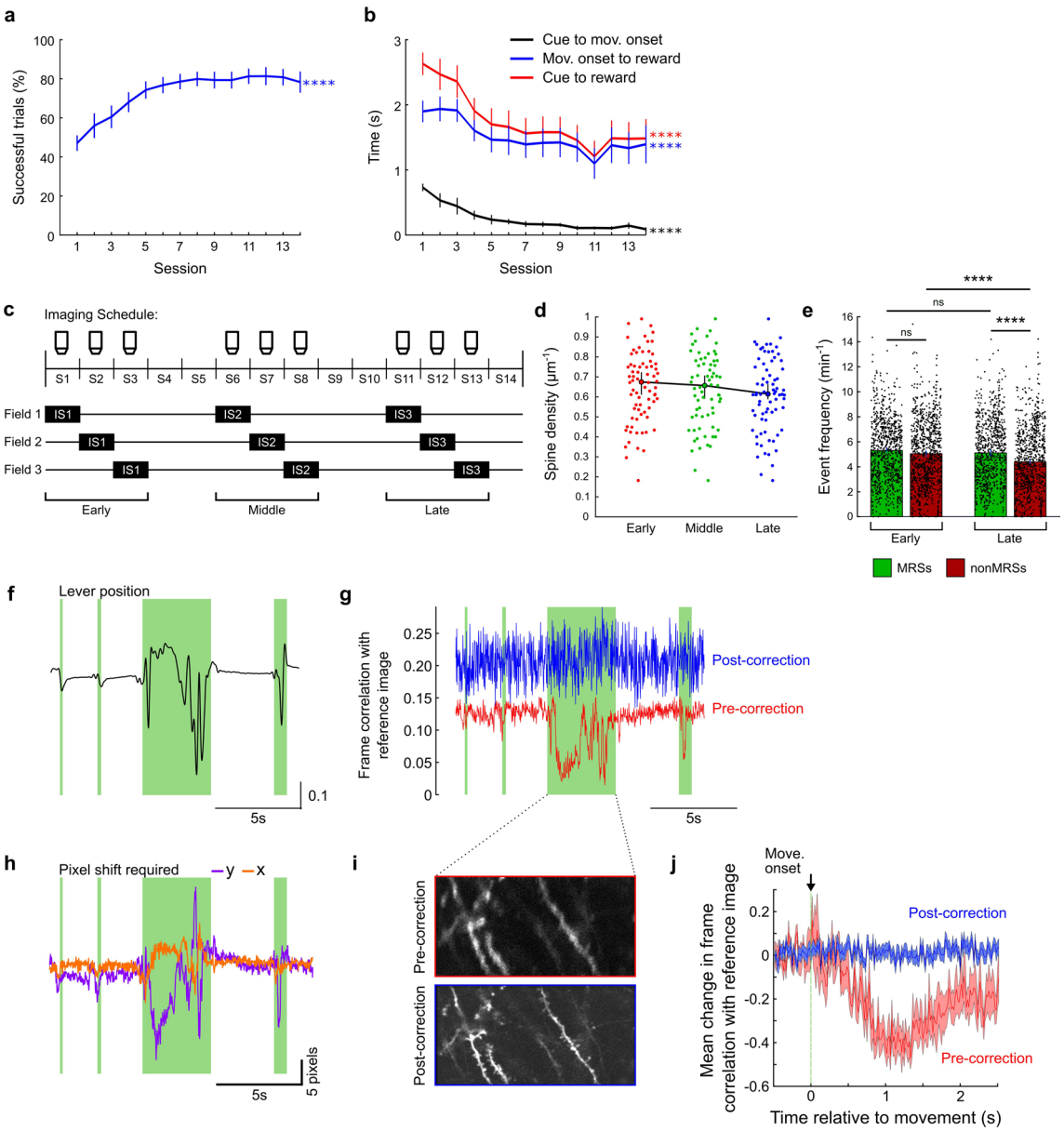
Extended data is available for this paper at https://doi.org/10.1038/s41593-022-01086-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-022-01086-6.

Correspondence and requests for materials should be addressed to Nathan G. Hedrick or Takaki Komiyama.

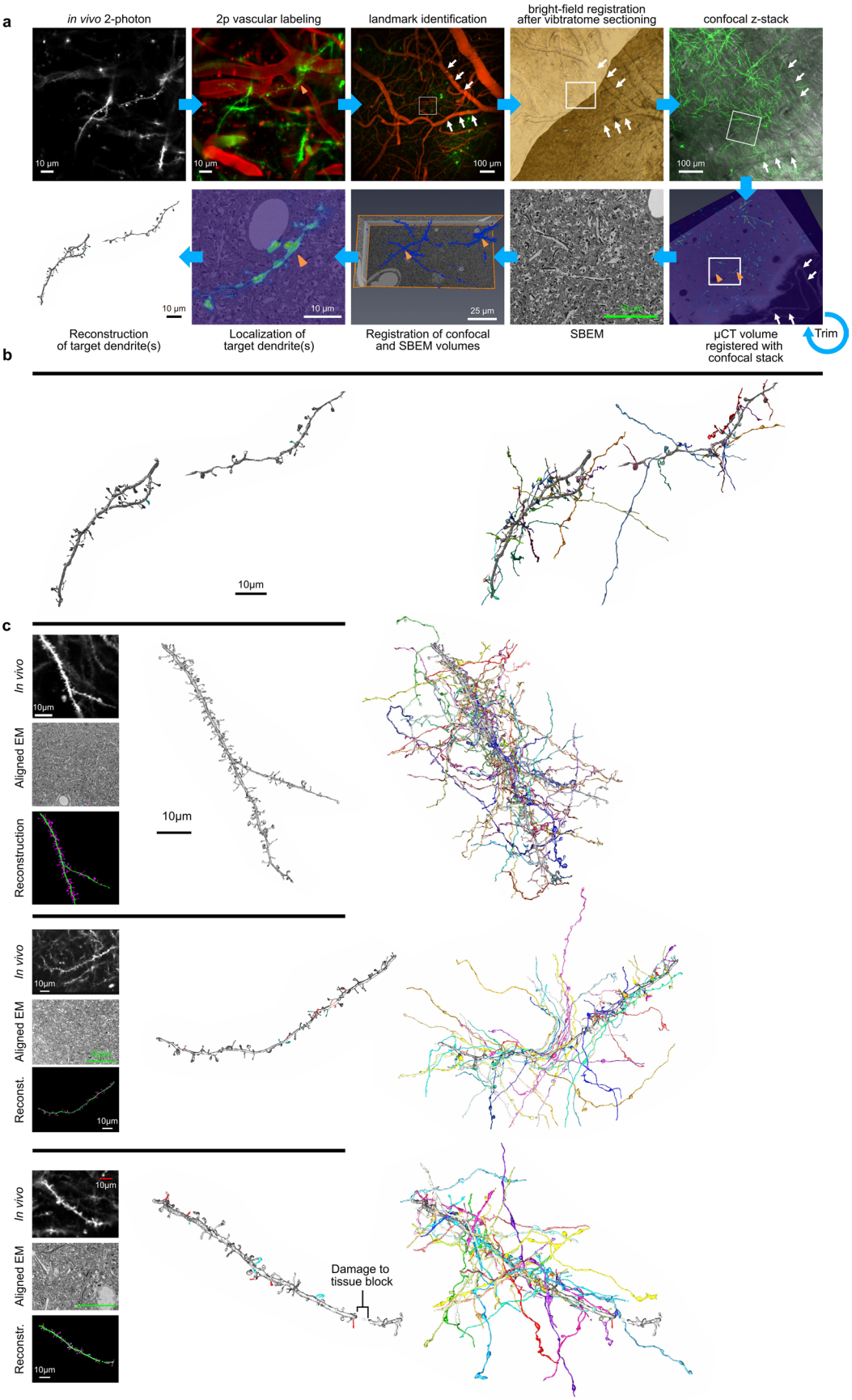
Peer review information *Nature Neuroscience* thanks Carlos Portera-Cailliau and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



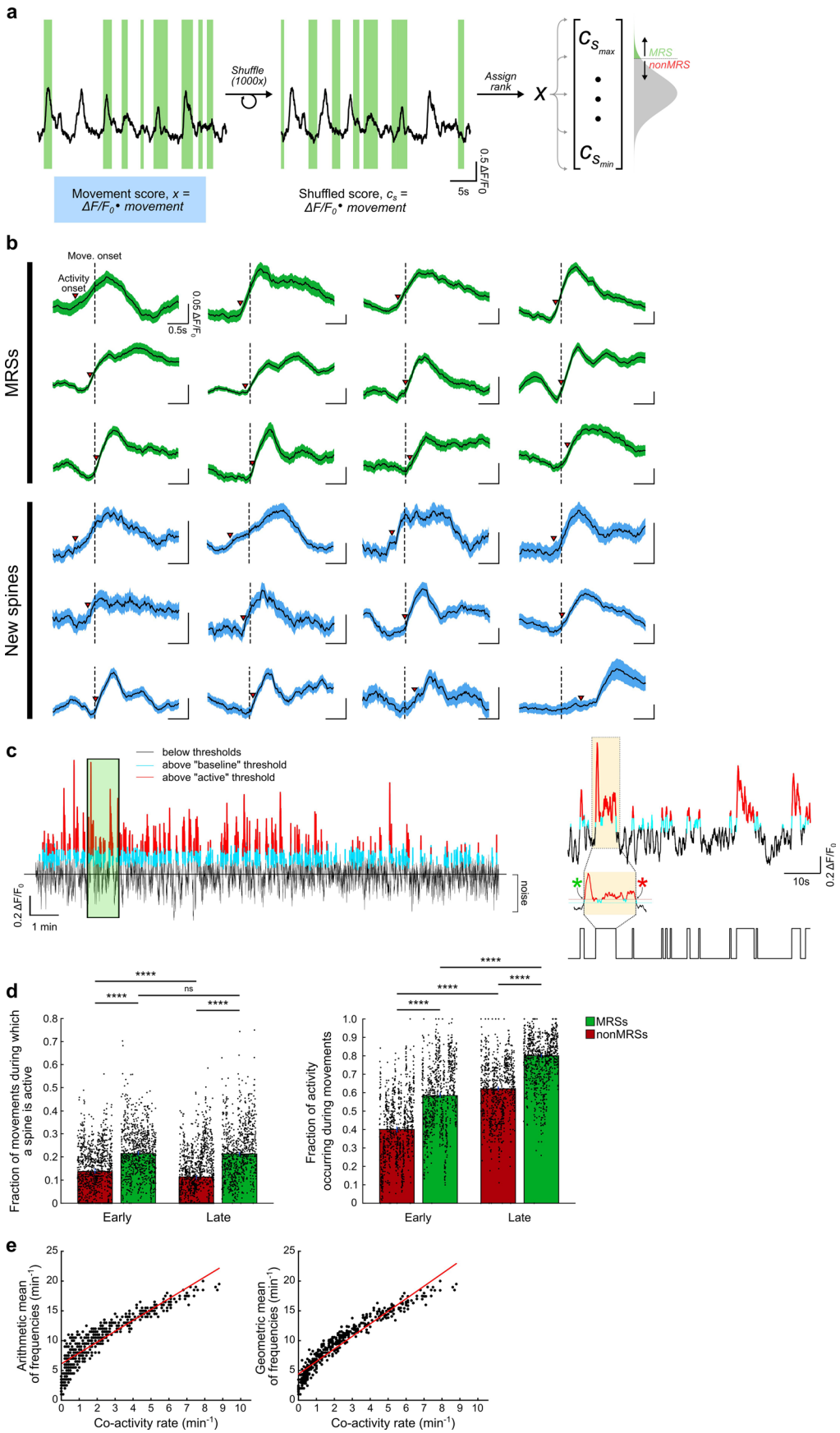
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Experimental setup, learning metrics, and image quality assurance. (a,b) Task performance improves over days of training. (**a**) The percentage of trials resulting in reward significantly increases over learning (p = 2e-31; Pearson's correlation coefficient). Data points correspond to the mean fraction of successful trials ± SEM. **** p < 0.001. n = 53 mice. **(b)** The reaction time (time from cue onset to movement onset; black) as well as the time from cue to reward (red) and movement onset to reward significantly decrease over learning (p = 8e-35 for cue-to-reward; p = 7e-50 for cue-to-movement; p = 6e-4 for movement onset to reward, Pearson's correlation coefficient). Data points correspond to mean time ± SEM. **** p < 0.001. n = 53 mice. **(c)** Schematic of imaging schedule. In each animal, three dendritic fields were imaged on each of the first three training days (Imaging Session 1, IS1), after which each field was imaged in 5-day intervals for two additional imaging sessions (IS2 and IS3). This yielded three imaging sessions for each field: early, middle, and late sessions. **(d)** Spine density does not significantly change over learning (1-way ANOVA, main effect of session: $F = 0.94$, $d.f. = 2$, p = 0.39). Individual dendrites are plotted as colored dots; black line corresponds to the median spine density ± bootstrapped 95% CI.n = 76 dendrites / 23 mice.**(e)** Spine event frequency is stable over learning for MRS (p = 0.46, rank-sum test), and decreases for nonMRSs (p = 1e-5, rank-sum test). Spine event frequencies are higher for MRSs in late sessions (early: p = 0.17, late: p = 7e-7). Bars represent median ± bootstrapped 95% CI. n = 898 early MRSs, 820 late MRSs, 869 early nonMRSs, 788 late nonMRS. Multiple comparisons were corrected using the FDR method. **(f)** Example lever position trace, showing movement periods in green shadings. **(g)** Example of frame-wise 2d correlation with the corresponding calculated reference image (that is, an iteratively aligned and average-projected version of the imaging field) associated with the behavioral window shown in (**d**). Prior to motion correction (red line), frame correlations with the reference image show with a decrease during movements. After motion correction, mean frame correlations are higher regardless of movements (blue line). **(h)** Example x-y pixel shifts used for motion correction of each frame in the behavioral window shown in (**f,g**), showing that extra correction during movements successfully compensated for movement artifacts. **(i)** Average projection images from the ~4 s behavioral window indicated in (**g**). Prior tomotion correction, images are blurry and individual structures are difficult to resolve because of misalignment across frames (top). After motion correction, images are sharp, and individual spines are visible (bottom). **(j)** Motion correction generates stable frames over movement. When aligned to movement onset, frame correlations with the reference image (as in (**e**)) decrease sharply during movements in the raw, pre-correction images (red line), but are stable during movements for the post-correction frames (blue line). Curves represent mean change in reference imaged correlation. Shaded portions correspond to SEM.



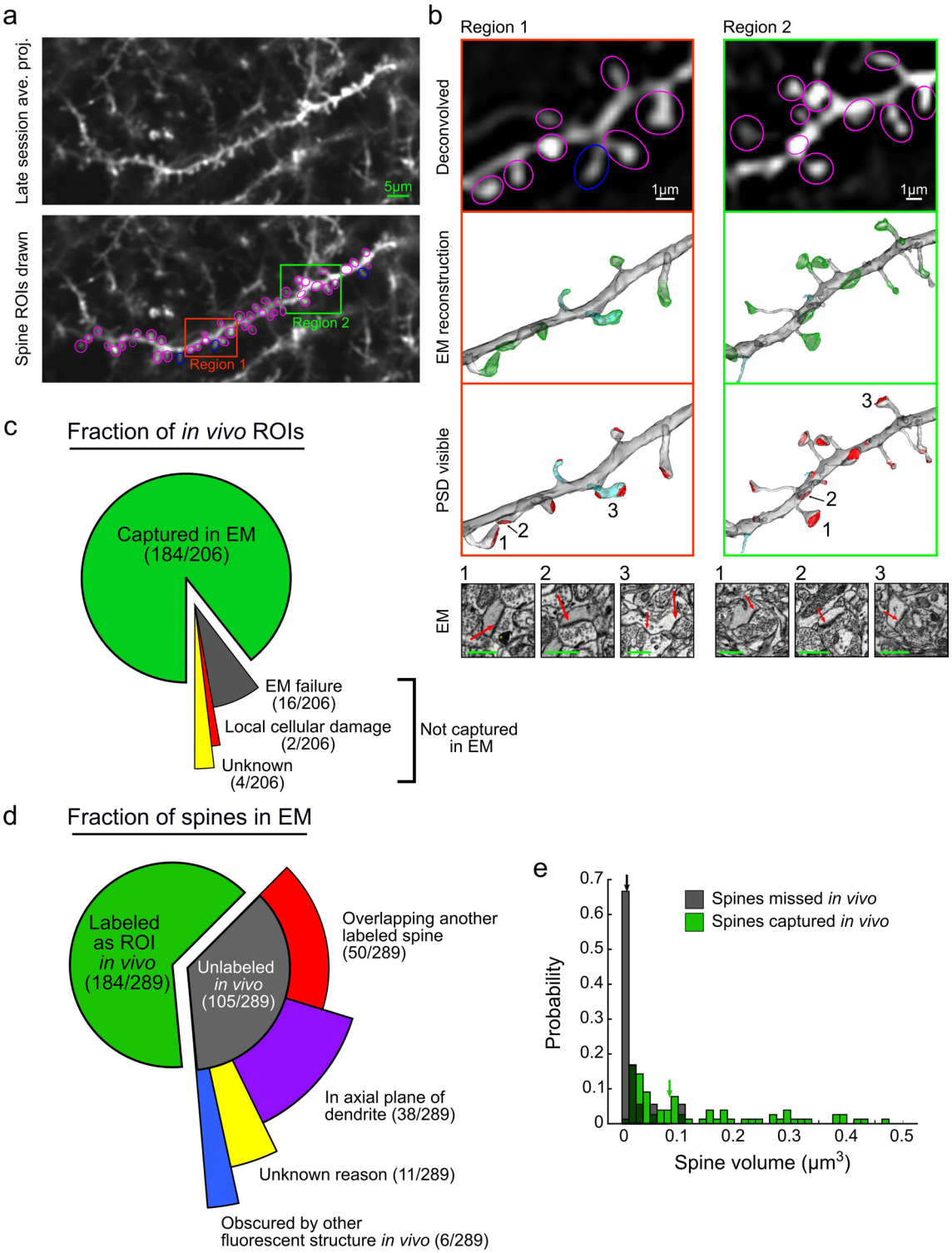
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Workflow of *in vivo* imaging followed by correlated *ex vivo* electron microscopy. (a) Workflow of correlated light and electron microscopy (CLEM) to identify the *in vivo*- imaged dendrites for EM. From top left moving clockwise: *in vivo* images of dendrites imaged during an experiment are acquired a final time for reference of the most updated structural information. Vascular maps are then generated by 2p *in vivo* imaging of TRITC-Dextran (injected retro-orbitally on final day of *in vivo* imaging), capturing an ~200μm z-stack in the area surrounding the dendrite of interest for detailed structural landmarks (panel 2 and white square in panel 3) and a low-zoom (usually 1 zoom/ ~1000μm^2, panel 3) to provide the larger structural context with respect to the vasculature. Arrows indicate the vasculature used for aligning the subsequent images. After the animal is perfused with fixative, thin sections are cut, and bright-field imaging is used to find the vasculature and approximate the location of the dendrite (white square in panel 4). Confocal images are then acquired using the slice containing the target vasculature so as to ensure the location of the target dendrite (white square in panel 5). The slice of interest is then processed for EM, thereby permanently rendering the tissue photo-inaccessible. The tissue is then exposed to iterative steps of X-Ray tomography and trimming (guided by software-based non-rigid alignment) so as to isolate the dendrite within a small volume amenable to EM. Shown is an alignment of the confocal stack with the x-ray tomogram. Colored arrowheads point to the dendrites of interest identified in subsequent images. Serial-section scanning-blockface EM is then performed on the final tissue block, and the confocal images are aligned to the output using the accumulated fiducial structures over previous imaging modalities. The final result (panel 8) is an aligned stack of confocal fluorescence and EM data, allowing a fluorescence overlay onto the EM stack. Using this alignment, the target dendrite and constituent spines are then reconstructed using the *in vivo* 2p image as reference for the domains of interest. These experiments were repeated independently in n = 4 mice with comparable results (shown below). **(b)** After reconstruction of the target dendrite(s) and spines (left), the synaptically connected axonal partner of each spine was traced to the full extent of the imaging volume or to the furthest point of confident identification (right). In this volume, 59 axons were successfully reconstructed to various extents. **(c)** Results of all remaining CLEM experiments from the current study. Left, *in vivo* 2-photon average projections, aligned EM data, and the corresponding reconstruction of the target dendrites are shown. Reconstruction images are angled so as to accentuate the correspondence between 2p and EM. Middle, zoomed-in visualizations of the reconstructions of target dendrites prior to reconstruction of synaptically connected axons. The number of axons reconstructed in each volume were: 137 (top), 74 (middle), and 60 (bottom), for a total of 330 axons (including the 59 axons from (**b**)). Note that the axonal partners of spines near the ends of dendrites were often not reconstructed due to their proximity to the edge of the sample volume.



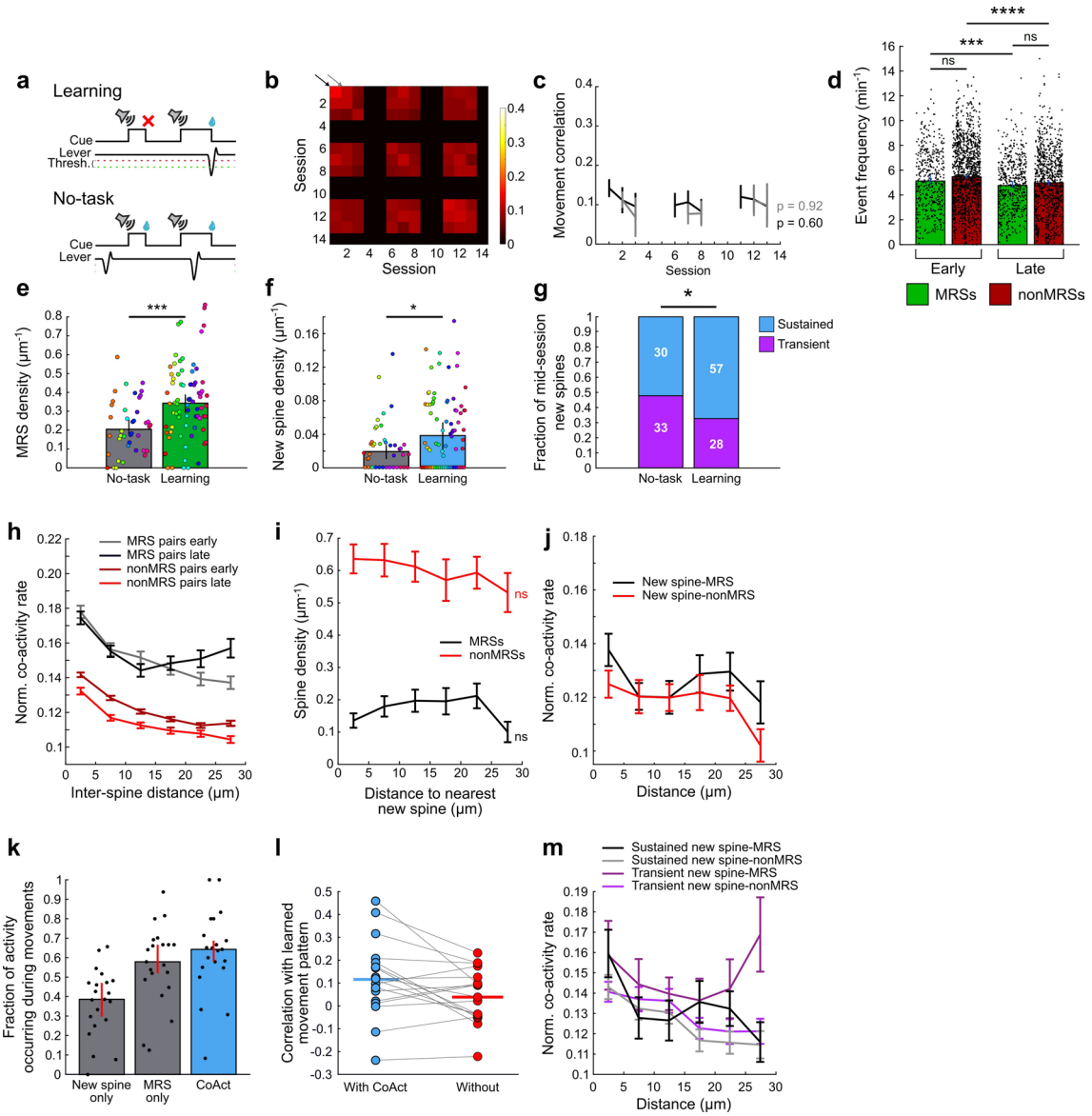
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Characterization of movement-related signals at single dendritic spines. Schematic of MRS definition. The dot product of the $\Delta F/F_0$ trace with binarized movement traces defines the movement score, *x*, for a given spine. Movements are then shuffled in time 1000x (without breaking individual movements), and the corresponding shuffled scores calculated. The real and shuffled scores are then ranked together. Those spines whose movement score is above the 97.5th percentile are considered MRS. Example traces of movement onset-aligned activity for 12 individual MRSs (green, top) and 12 new spines (blue, bottom). Movement onset is indicated by black vertical dashed line. Activity onset timing is indicated with red triangles. Mean ± SEM. Explanation of event detection based on previously published[9,16] methods. Left, full $\Delta F/F_0$ trace for an example spine. For each trace, noise is estimated based on the negative portion of the trace, reflected about the x-axis. Values above 1x the standard deviation of this trace are considered above the 'baseline' threshold (light blue), and values above 2x the standard deviation are considered above the 'active' threshold (red). Green rectangle indicates portion of the trace inspected in right. Upper right, zoomed-in section of the trace for easier visualization. Tan region indicates example region used to demonstrate end and start times of active periods. End times (red asterisk) are defined as when a smoothed version of the trace falls below the active threshold, and start times (green asterisk) are defined by moving backwards in time from each end point to the first point where the raw trace falls below the baseline threshold. All frames meeting these criteria are assigned a value of 1, and all other frames a value of 0. In this way, events are considered continuous unless they fall below the baseline threshold. The resulting binarized trace (bottom right) indicates when spines are active and inactive. Relationship of spine active periods to movements. Using the binarized activity described in (**c**), the relationship of spines' activity can be shown as either the fraction of movements during which a given spine is active (left) or the fraction of activity for a given spine that occurs during movements (right). Spines are subdivided into MRSs and nonMRS, as described in (**a**). Bars correspond to the medians ± 95% CI. MRSs are active during a higher fraction of movements than nonMRSs in both early ($p = 8e\text{-}52$; rank-sum test) and late ($p = 2e\text{-}70$; rank-sum test) sessions, and while MRSs are active during a similar fraction of movements across sessions ($p = 0.95$; rank-sum test), nonMRSs show a significant decrease ($p = 5e\text{-}7$; rank-sum test). Similarly, a higher fraction of MRS activity occurs during movements than nonMRSs in both early ($p = 1e\text{-}72$; rank-sum test) and late ($p = 4e\text{-}93$; rank-sum test) sessions, though both MRSs ($p = 9e\text{-}101$; rank-sum test) and nonMRS ($p = 4e\text{-}105$; rank-sum test) show an increased fraction of activity occurring during movements by late sessions. All comparisons were two-sided. Multiple comparisons were corrected using the FDR method. **** $p < 0.001$. The co-activity rates of simulated spine pairs (Methods) correlate better with the geometric mean of the constituent spine pairs' event frequencies than the arithmetic mean, demonstrating that the geometric mean is a more suitable normalization factor for co-activity rates. Overall statistics for arithmetic mean: $r = 0.91$, $p < 0.0001$; geometric mean: $r = 0.95$, $p < 0.0001$. The r-value was higher for the fit between geometric mean and co-activity rates in 1000/1000 shuffles.



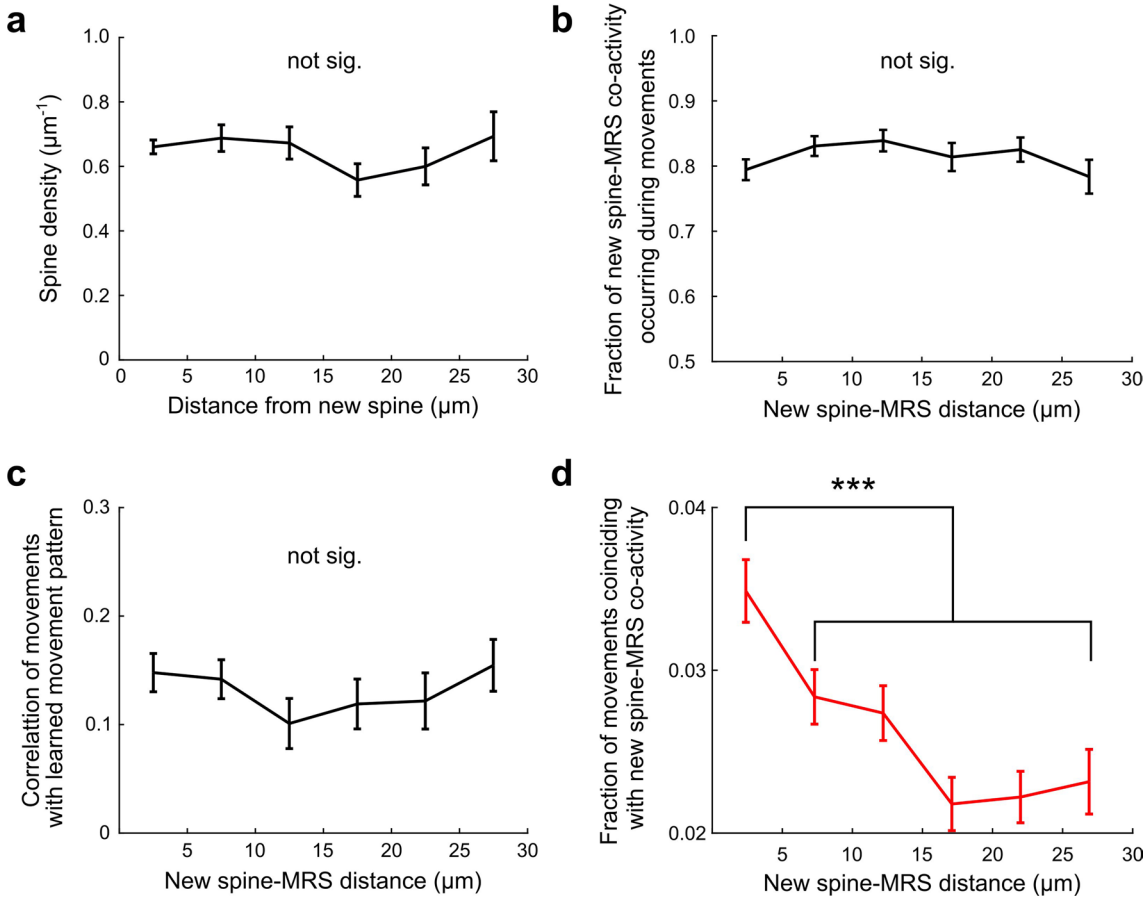
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Spine identification and CLEM-based corroboration. Example of ROIs. Top, *in vivo* image under consideration. Bottom, elliptical ROIs manually drawn in the initial analysis of this dendrite. Magenta ellipses correspond to ROIs that were successfully located in EM, while blue ellipses show those that were not confidently identified. Two example regions are indicated for closer inspection in (**b**). **(b)** Zoomed-in regions shown in (**a**) for closer inspection of ROI drawing. Top row: deconvolved *in vivo* images of the two regions shown in (**a**). Deconvolved images were used throughout this study to aid in spine identification. Second row: corresponding EM reconstructions showing the same spines imaged *in vivo*. Third row: reconstructions of the associated regions of dendrite, but with post-synaptic density (PSD) reconstructions shown in red. Fourth row: example EM micrographs demonstrating PSDs on a selection of spines (numbers correspond to spines in above image). Scale bars in the fourth row (green) correspond to 0.5 μm. **(c)** Summary of the fraction of ROIs identified *in vivo* that were successfully located in EM. The vast majority of structures drawn as spines *in vivo* were also found in EM. Of those structures that were not located, most were due to technical failures in EM acquisition (1 of the 16 of such failures was due to debris from a previous slice obfuscating the region of the target dendrite, and 15 were due to slices that were skipped, likely due to uneven cutting). We also identified one highly localized dendritic region that appeared damaged/blebbed despite a healthy appearance *in vivo*, preventing the assessment of two spines observed *in vivo*. The four remaining ROIs were not located due to unknown reasons, suggesting either mislabeling or elimination of these spines between the final *in vivo* imaging session and EM processing (see a potential example of this in the blue-encircled spine in (**b**). **(d)** Summary of the fraction of all spines identified in EM (within the dendrites captured *in vivo*) that were also labeled as ROIs *in vivo*. The majority of spines that were not identified *in vivo* were co-axial with either another (usually larger, see (**e**)) spine (red slice), or the dendritic branch (purple slice). A small subset of spines located in EM were obscured by another fluorescent structure (for example, a labeled axon) *in vivo* (blue slice). The remaining spines were not identified *in vivo* for unknown reasons, which might include the rapid formation of spines between the final imaging session and EM processing, or most of these spines being below the detectable size for *in vivo* imaging (see (**e**)). **(e)** Summary histograms of the volumes (as calculated in EM) of spines that were also captured *in vivo* (green) and spines that were not identified *in vivo* (gray). The vast majority of spines that were identified in EM but not labeled *in vivo* were small in size; 67% of 'missed' spines were within the 3rd percentile of the 'captured' population. The medians of the 'missed' and 'captured' groups are indicated with black and green arrows, respectively.

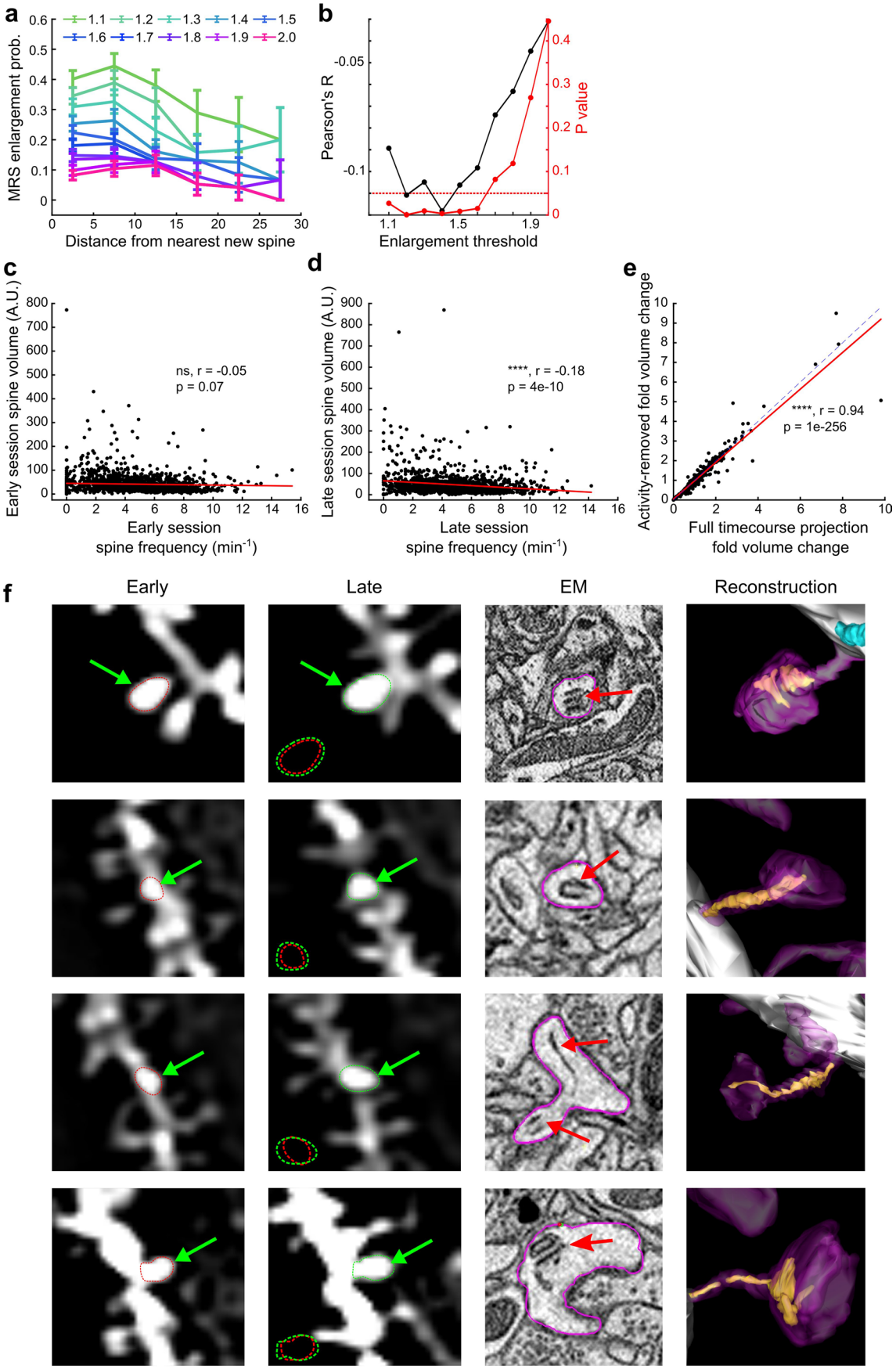


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Analysis of no-task controls. (a) Schematic of 'no-task' condition compared to the typical 'learning' condition. Unlike in the learning condition, the no-task condition administers water rewards at the end of each cue period, regardless of the movement of the lever. Mice were head-fixed in the experimental apparatus only during the 9 imaging sessions. **(b)** Heat map of lever press correlations across sessions of the 'no-task' experiment. Black portions correspond to days during which mice were not placed in the experimental apparatus. n = 14 mice. **(c)** Lever movement correlations, both within and across adjacent sessions, did not increase over time in the no-task controls. Significance of the relationship was evaluated with Pearson's correlation coefficient, with p-values shown in figure text. Data points correspond to mean ± SEM correlation values for each session. n = 14 mice. **(d)** Activity event frequency of different spine types for the no-task condition. In both early and late sessions, MRSs and nonMRSs show comparable event frequencies (early: p = 0.07, late: p = 0.4, rank-sum test). MRSs and nonMRS slightly decrease (MRSs: p = 0.0035, nonMRSs: p = 9e-8) their activity event frequencies over the experiment. Multiple comparisons corrected using the FDR method. When pooled with the learning-group data, only late MRSs are significantly higher in the learning group than in the no-task control after correcting for multiple comparisons (p = 0.006, rank-sum test). Conversely, nonMRS showed higher event frequencies in both early and late sessions for the no-task group (early: p = 2e-4, late: p 0.002, rank-sum test). n = 433 early MRSs, 1270 early nonMRSs, 506 late MRSs, 918 late nonMRSs. Bars represent median ± 95% CI. Individual data points (spines) shown as black dots. **(e)** MRS density was lower in the no-task group (0.21/μm, 95% CI = [0.17, 0.25]) than the learning group ((0.34/μm, 95% CI = [0.29, 0.38]; p = 4e-4, rank-sum test). Bars represent median values ± 95% CI. Circles correspond to individual dendrites, color-coded by animal. Note that all imaged dendrites are shown, including those that show no new spine formation. n = 40 no-task group dendrites, 76 learning-group dendrites. **(f)** The new spine density on imaged dendrites was lower in the no-task group (0.02/μm, 95% CI = [0.010 0.027]) than the learning group (0.039/μm, 95% CI = [0.022, 0.054]) (p = 0.04, rank-sum test). Bars represent median values. Circles correspond to individual dendrites, color-coded by animal. Note that all imaged dendrites are shown, including those that show no new spine formation. n = 40 no-task control dendrites, 76 learning-group dendrites. **(g)** A higher fraction of new spines was transient in the control group vs. learning group (p = 0.02, Chi-squared test). **(h)** Functional clustering was intact in the no-task condition, but co-activity was generally less pronounced. Much like the learning condition, MRSs showed higher overall co-activity levels than nonMRSs (3-way ANOVA [MRS status × distance × session], main effect of MRS status: $F = 976.58$, $d.f. = 1$, p = 2e-211; post-hoc using least-significant difference: p < 0.0001 for MRS vs. nonMRS across all distance bins when collapsing across sessions or when preserving session identity). We observed a main effect of distance ($F = 109.38$, $d.f. = 5$, p = 5e-211), and the first distance bin was higher than all other bins (p = 1e-30, post-hoc using LSD), indicating that functional clustering is present. By comparing to the learning-group data (Fig. 1j), we find that overall co-activity values were higher in the learning group (4-way ANOVA [MRS status × distance × session × condition], main effect of condition, p = 6e-54; post-hoc using LSD). We also observe a significant MRS status × condition interaction ($F = 106.27$, $d.f. = 1$, p = 7e-25), with post-hoc analysis revealing that both MRSs (p = 2e-57) and nonMRS (p = 1e = 5) show higher co-activity rates in the learning condition than their no-task counterparts. n = 6376 early MRS pairs, 49894 early nonMRS pairs, 8088 late MRS pairs, 47274 late nonMRS pairs. Y-axis scale was set to match that of the learning group (Fig. 1j). Mean ± SEM. **(i)** Neither MRS density (r = 0.03, p = 0.64) nor nonMRS density (r = −0.10, p = 0.16, Pearson's correlation coefficient) changed as a function of distance to the nearest new spine in the no-task controls. Mean ± SEM. **(j)** New spines formed during the no-task control are weakly functionally clustered with MRSs. Like in the learning condition, new spines formed in the no-task condition show slightly higher overall co-activity rates with MRSs vs. nonMRSs (2-way ANOVA [MRS status × distance], main effect of MRS status: $F = 4.2$, $d.f. = 1$, $p = 0.04$) and closer spine pairs generally had higher co-activity (main effect of distance: F = 2.3, d.f. = 5, p = 0.04). No significant interaction term was observed. When comparing to the learning condition data (Fig. 4c), we observed a significant main effect of MRS status and distance (3-way ANOVA [MRS status × distance × condition], MRS status: $F = 22.85$, $d.f. = 1$, p = 2e-6; distance: $F = 14.87$, $d.f. = 5$, p = 2e-14), but not condition (that is no-task vs. learning). However, we found a significant interaction between distance and condition (F = 5.66, d.f. = 5, p = 3e-5), suggesting that the spatial relationship of co-activity rates differs across the two conditions. Indeed, post-hoc comparisons show that only the learning condition shows significant functional clustering, such that the first distance bin is greater than all other distance bins (p < 0.0001 for learning condition at distance = 0–5um vs. all other bins; p > 0.05 for no-task condition at distance = 0–5um vs. all other bins). n = 789 new spine-MRS pairs, 2029 new spine-nonMRS pairs. Y-axis scale was set to match that of the learning group (Fig. 3c). Mean ± SEM. **(k)** Like in the learning condition, new spine-MRS pairs show activity that is enriched during movements, more so than new spine activity alone (p = 7e-5; sign-rank test) or MRS activity alone (p = 0.037; sign-rank test). Multiple comparisons were corrected for using FDR. n = 21 fields showing new spine-MRS co-activity / 25 total fields. Bars represent median ± 95% CI. **(l)** Movements containing new spine-MRS co-activity are not more correlated with the learned movement pattern (LMP) than movements lacking such activity (p = 0.43, sign-rank test). Data points correspond to the median correlation of movements occurring with (blue) or without (red) co-activity of new spine-MRS pairs in a given field. n = 18 fields showing new spine-MRS co-activity during at least 3 movements / 21 fields showing new spine-MRS co-activity / 25 total fields. **(m)** The functional separation of sustained new spine-MRS pairs from other pair types is conspicuously absent. New spine-MRS pairs, irrespective of survival, showed slightly higher overall co-activity rates (3-way ANOVA [MRS status × distance × survival], main effect of MRS: p = 0.001), and spine pairs were generally functionally clustered (main effect of distance, p = 0.02), but no interaction terms were significant. Notably, we observed a main effect of new spine survival (p = 0.01), but post-hoc inspection revealed that transient new spine-MRS co-activity was actually slightly higher than the other groups (p < 0.02 for all comparisons). Mean ± SEM.

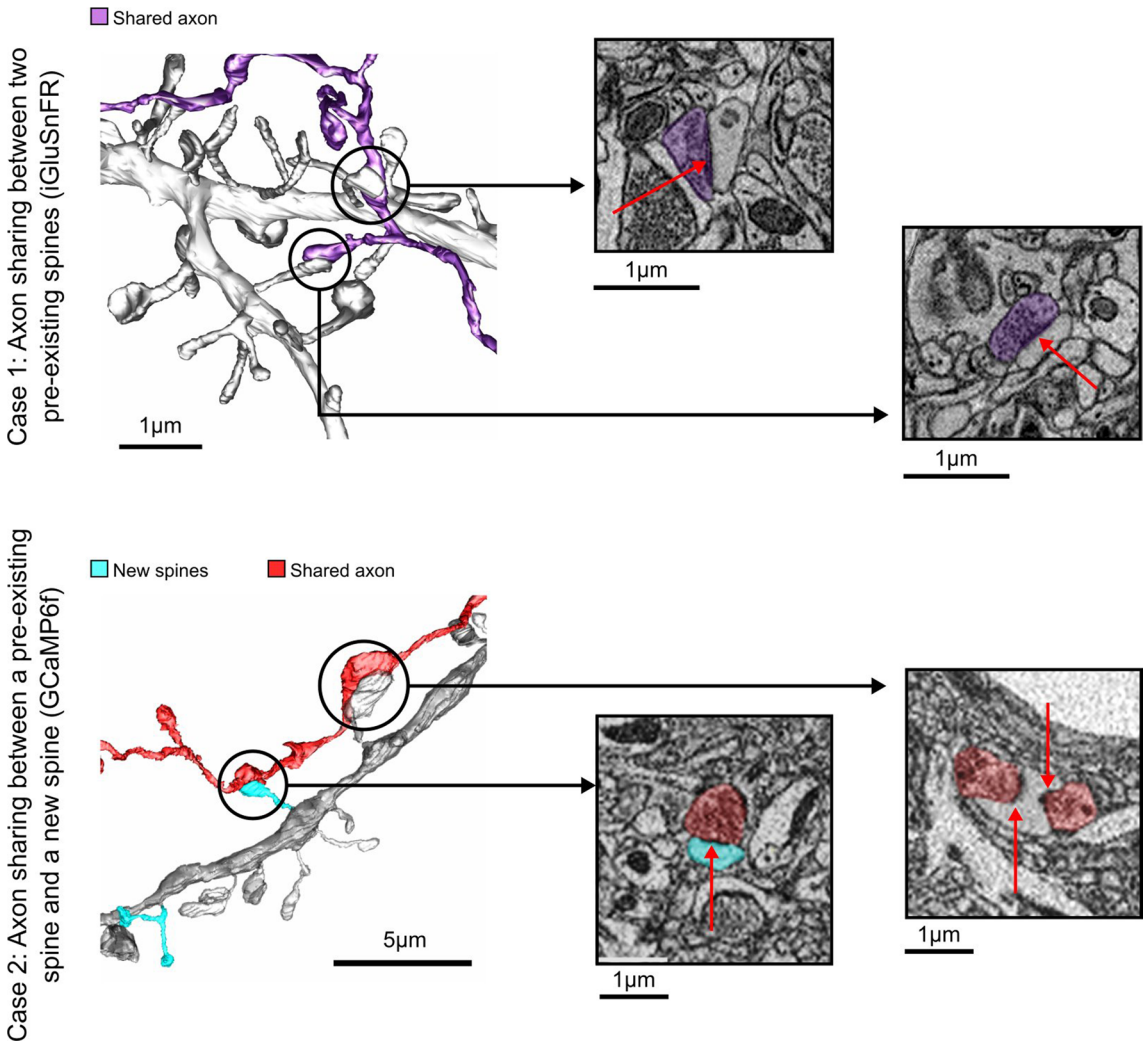


Extended Data Fig. 6 | Additional spatial analyses of dendritic structural and functional features with respect to new spines. (a) Overall spine density as a function of distance from new spines. While there is a trend towards higher spine densities closer to new spines, the effect is not significant (Pearson's correlation coefficient; $r = -0.10$, $p = 0.13$). Data points correspond to the mean spine density ± SEM of all dendrites displaying new spine formation ($n = 50$ dendrites across 21 mice). **(b)** A large fraction of new spine-MRS co-activity occurs during movements regardless of the distance between the spines (ANOVA with post-hoc test using the least-significant difference; main effect of distance $p = 0.11$). Data points correspond to the mean ± SEM of all new spine-MRS pairs ($n = 1658$ pairs across 50 dendrites from 21 mice). **(c)** Movements coincident with new spine-MRS coactivity have similarly high correlations with the learned movement pattern regardless of the distance between the spines (ANOVA with post-hoc test using the least-significant difference; main effect of distance: $p = 0.85$). Data points correspond to the mean (± SEM) correlation of movements coincident with co-activity of a given new spine-MRS pair ($n = 1658$ pairs / 50 dendrites / 21 mice). **(d)** The fraction of movements coincident with co-activity of a particular new spine-MRS pair shows a significant decrease with increasing distance between the spines, illustrating that a higher number of movements is encoded by closer spine pairs (ANOVA with post-hoc test using the least-significant different; main effect of distance: $p = 0.9e\text{-}05$; $p < 0.02$ for post-hoc comparison of first distance bin vs. all other distance bins). Mean ± SEM. ($n = 1658$ pairs / 50 dendrites / 21 mice).

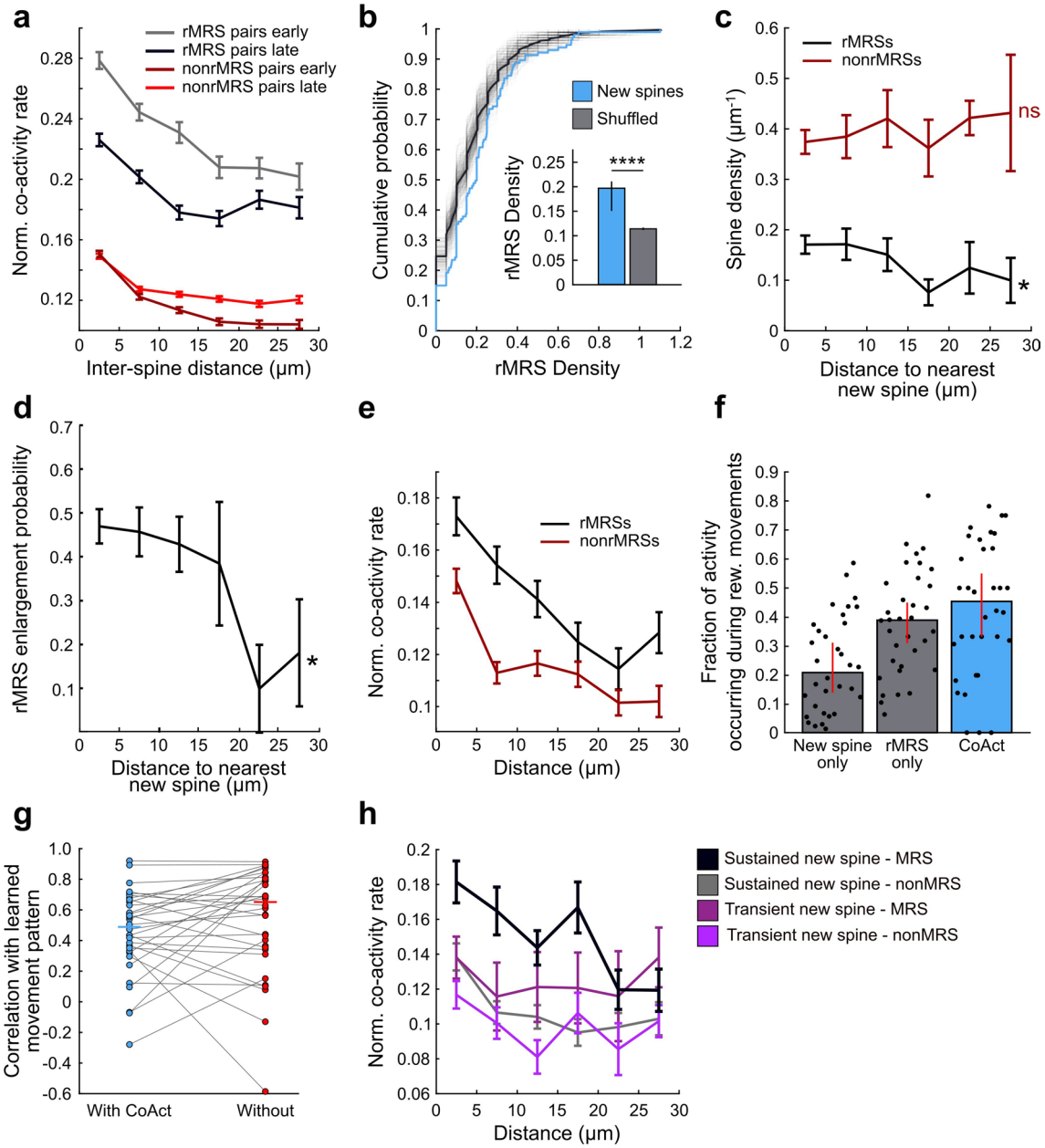


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Supporting evidence of *in vivo* spine volume estimates using iGluSnFR. Effect of different enlargement threshold cutoff values (from 1.1, light green, to 2, magenta) on the relationship between the probability of MRS enlargement and distance to the nearest new spine. Data points correspond to the mean probability of spine enlargement (± SEM) for MRSs in a given distance bin for each new spine imaged. n = 118 new spines, 697 MRSs present on 50 new-spine-containing dendrites across 21 mice. Summary of Pearson's statistics for data shown in (**a**). The relationship is significant for threshold values up to 1.5-fold change in spine volume. Data points correspond to either the calculated Pearson's correlation coefficient (black, left axis), or the corresponding p-value (red, right axis) calculated from statistical tests on the data groups shown in (**a**). **(c,d)** Relationship between estimated spine volume and spine event frequency in early (**c**) and late (**d**) sessions, revealing a lack of positive correlation, suggesting that our methods using iGluSnFR fluorescence do not overestimate the volume of highly active spines. Linear fit of data shown in red. Significance of the relationship was determined from Pearson's correlation coefficient, with r and p-values shown in figure text. **** $p < 0.001$, ns: not significant. **(e)** Evaluation of the effect of removing active periods from image projections prior to calculation of spine volume. In 11 of the 23 'learning group' animals (769 spines), spine volume was re-calculated after removal of all frames in which a given spine was considered 'active'. 'Activity-removed' spine volume estimates correlate strongly and significantly with spine volume estimates from full-length time series projections. Unity line shown in dashed blue. Linear fit of data shown in red. Significance of the relationship was determined from Pearson's correlation coefficient, with r and p-values shown in figure text. **** $p < 0.001$ **(f)** Example *in vivo*, EM, and reconstruction images illustrating the presence of spine apparati, a signature of mature and potentiated spines, in spines that showed enlargement *in vivo* during learning. Green arrows indicate spines of interest. Each spine of interest in the *in vivo* images is outlined for spine size comparisons (red = early; green = late), and the early and late outlines are overlaid in the bottom left corner of each late session. Red arrows in the EM images indicate spine apparati. Reconstructed spine apparati are shown as yellow structures within spines (magenta).

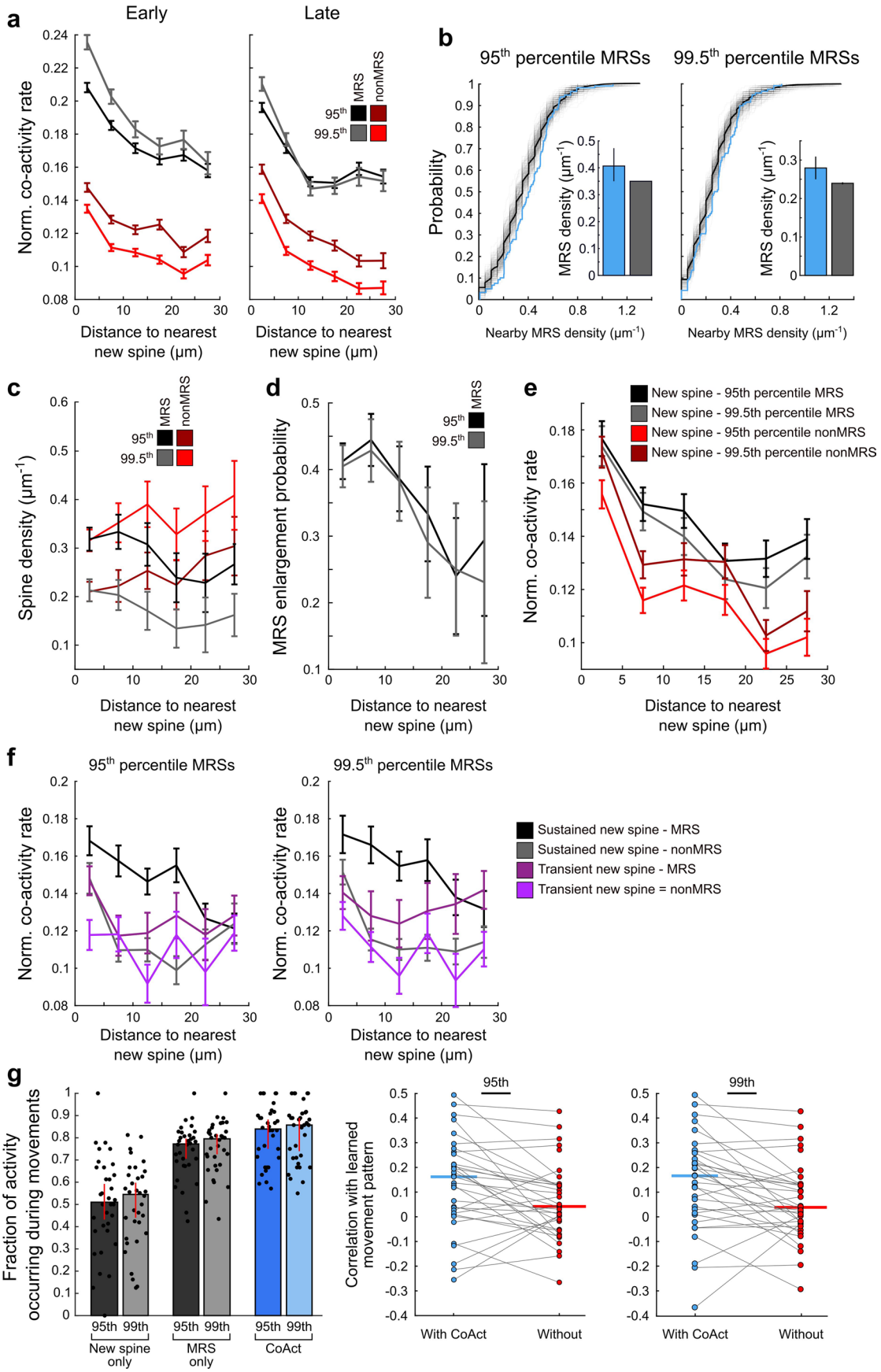


Extended Data Fig. 8 | Cases of axon-sharing between spines. Two additional example cases of axon sharing between spines on imaged dendrites. The left-most column shows reconstructions of the dendrites and its spines along with the axon being shared between two spines. Each synapse occurring with the dendrite of interest is circled, and corresponding EM images illustrating synaptic contact are shown on the right (further indicated with red arrows). The top example shows a double connection between two pre-existing spines immediately after a dendritic branch point. The bottom example shows a double connection formed by a new spine and a nearby pre-existing spine.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Limiting analysis to rewarded movement-related spines (rMRSs) yields reproduces main findings. Pre-existing rMRSs show strong functional clustering. Data points correspond to mean ± SEM. n = 443 early rMRSs / 2052 condendritic rMRS-rMRS pairs; 1472 early nonrMRS-nonrMRS pairs; 571 late rMRSs / 3278 late nonrMRS-nonrMRS pairs. **(b)** rMRS density nearby (≤10μm) new spines is higher than expected by chance, as estimated by randomizing the location of new spines. Statistical significance was determined by a one-sided comparison of the median of each shuffle across all new spines to the real data median; the resulting fraction of shuffles that are consistent with the null hypothesis (that is that there is *not* a higher nearby MRS density than expected by chance) corresponds to the p value; $p < 0.001$. Inset, median nearby rMRS density of all new spines compared to the median chance estimate. Error bars correspond to bootstrapped 95% CI. n = 118 new spines across 50 dendrites. **(c)** rMRS density decreases as a function of distance from the nearest new spines, whereas nonrMRS density does not. MRS density: Spearman's $r = -0.16$, $p = 0.01$; nonMRS density: Spearman's $r = 0.03$, $p = 0.67$). Data points correspond to mean ± SEM. n = 118 new spines across 50 dendrites. **(d)** rMRS enlargement probability decreases as a function of distance from the nearest new spine. Spearman's $r = -0.11$, $p = 0.048$). Data points correspond to mean ± SEM. n = 118 new spines; 443 rMRSs present on 50 new spine-containing dendrites. **(e)** New spines are functionally clustered with rMRSs in a manner similar to MRSs. Data points correspond to mean ± SEM. n = 1091 new spine-rMRS pairs; 2774 new spine-nonrMRS pairs / 118 new spines across 50 dendrites / 21 mice. **(f)** The fraction of activity of new spine-rMRS pairs occurring during movements trends higher than either new spine activity alone ($p = 1e\text{-}6$) or rMRS activity alone ($p = 0.1$; sign-rank rest correcting for multiple comparisons using the FDR method). n = 34 fields. Bars correspond to the median fraction of cases for each group present in each of the 34 fields imaged. Error bars represent bootstrapped 95% CI. **(g)** Rewarded movements associated with new spine-rMRS co-activity are not more correlated with the learned movement pattern (LMP) than rewarded movements lacking such activity ($p = 0.14$). Thus, it appears that preferential occurrence of co-activity during rewarded movements lead to the similarity of movements to LMP shown in Fig. 4g. **(h)** The survival of new spines can be predicted by co-activity with nearby rMRSs. New spine-rMRS pairs showed higher overall co-activity rates than other spine pairs (3-way ANOVA [rMRS status × distance × survival], main effect of rMRS status: $p = 3e\text{-}11$), and co-activity values were significantly impacted by distance (main effect of distance, $p = 0.0006$). We observed a main effect of new spine survival ($p = 0.007$), and trends for interaction terms distance × survival ($p = 0.055$) and rMRS × distance × survival ($p = 0.058$). Given the trend in the three-way interaction term, we performed post-hoc comparisons of the four groups across each distance bin. Sustained new spine-rMRS co-activity at close distances (within 5μm) was significantly higher in 17 of the 18 possible comparisons against the other three groups (p values range from 7e-12 to 0.02 for the 17 significant comparisons, $p = 0.47$ for the final comparison), and was higher than values within the same group at longer distances (>15, $p < 0.008$). Conversely, transient new spines show co-activity that is not higher at closer distances (transient new spine-MRS pairs' co-activity within 5μm is not different than any other distance bin (p values range from 0.27 to 0.73); transient new spine-nonMRS pairs' show co-activity that is significantly higher ($p = 0.02$) than one other bin and significantly lower ($p = 0.046$) than another bin), suggesting that no significant functional clustering is present between transient new spines and other spines. Further, neither transient spine group showed a significant correlation with distance (transient new spines-MRS co-activity vs. distance, $r = 0.06$, $p = 0.39$; transient new spines-nonMRS co-activity vs. distance: $r = -0.07$, $p = 0.14$), while sustained new spine co-activity is significantly negatively correlated with distance (sustained new spine-MRS co-activity vs. distance: $r = -0.18$, $p = 0.0001$; sustained new spine-nonMRS co-activity vs. distance: $r = -0.14$, $p = 6e\text{-}5$). Thus, surviving new spines' co-activity with rMRSs is higher and more spatially selective than transient new spines. n = 28 transient new spines, 230 transient new spine-rMRS pairs, 669 transient new spine-nonrMRS pairs; 57 sustained middle-session new spines, 530 sustained new spine-rMRS pairs, 1259 sustained new spine-nonrMRS pairs. Mean ± SEM.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Main findings are robust against changes in the threshold for MRS definition. Shifting the threshold for defining MRSs to being greater than either the 95th percentile of shuffles or the 99.5th percentile of shuffles has negligible effects on the main findings. **(a)** Functional clustering of pre-existing spines is similar when using both 95th and 99.5th percentiles to define MRSs. 95th percentile cutoff: MRSs are still more co-active than nonMRS (3-way ANOVA, main effect of MRS status: $F = 1092.04$, d.f. = 1, $p = 2e\text{-}233$), and spines show distance-dependent co-activity rates (main effect of distance: $F = 132.07$, d.f. = 6, $p = 3e\text{-}138$), such that closer pairs are more co-active (Pearson's correlation coefficient, r values range from 0.14–0.22 and $p < 0.0001$ for all 8 separate groups shown). More conservative cutoff results in higher co-activity rates of MRS pairs. n = 687 early MRSs / 4318 early MRS-MRS pairs / 1080 early nonMRSs / 13621 early nonMRS-nonMRS pairs / 685 late MRSs / 4687 late MRS-MRS; 1230 late nonMRS / 12318 late nonMRS-nonMRS pairs at the 99.5th percentile; n = 954 early MRSs / 7719 early MRS-MRS pairs / 813 early nonMRSs / 8534 early nonMRS-nonMRS pairs / 868 late MRSs / 7313 late MRS-MRS pairs / 788 late nonMRSs / 8967 late nonMRS-nonMRS pairs. **(b)** The higher density of nearby MRSs is robust to the different MRS thresholds. Main plots show cumulative probability distributions of the nearby MRS density for real new spines (light blue) and chance, estimated by shuffling new spine locations 1000 times for each new spine (250 individual shuffles shown in gray, median of all shuffles shown in black). Statistical significance was determined by comparing the median of each shuffle across all new spines to the real data median; the resulting fraction of shuffles that are consistent with the null hypothesis (that is that there is not a higher nearby MRS density than expected by chance) corresponds to the p value. 95th percentile MRS density: $p = 0.003$. 99.5th percentile MRS density: $p = 0.012$. n = 118 new spines across 50 dendrites. **(c)** The distance-dependent decrease in MRS density is similar across different thresholds. 95th percentile: MRSs: Spearman's rank coefficient, $r = -0.12$, $p = 0.045$, nonMRS: $r = 0.07$, $p = 0.212$. MRSs show a significantly different slope than nonMRS ($p < 0.001$). 99.5th percentile: MRSs: $r = -0.11$, $p = 0.052$. n = 1390 new spine-MRS pairs / 2475 new spine-nonMRS pairs at 99.5th percentile; 1731 new spine-MRS pairs / 2131 new spine-nonMRS pairs at 95th percentile. **(d)** Volume increase probability with respect to new spine location is similar across different MRS thresholds. 95th percentile cutoff: $r = -0.07$, $p = 0.062$ (Pearson's correlation coefficient. 99.5th percentile cutoff: r = −0.−8, $p = 0.06$ (Pearson's correlation coefficient). n = 1390 new spine-MRS pairs / 2475 new spine-nonMRS pairs at 99.5th percentile; 1734 new spine-MRS pairs / 2131 new spine-nonMRS pairs at 95th percentile. **(e)** New spine functional clustering is nearly identical across different thresholds. 95th percentile MRSs: New spine-MRS pairs are more co-active than new spine-nonMRS pairs (2-way ANOVA, main effect of MRS status: $F = 18.71$, d.f. = 1, $p = 2e\text{-}5$). New spine co-activity generally shows a distance-dependent change in co-activity rates (main effect of distance: $F = 20.97$, d.f. = 6, $p = 1e\text{-}20$), an effect that did not depend on MRS status, (MRS status by distance interaction: $F = 1.5$, d.f. = 5, $p = 0.19$). Both groups showed higher co-activity at the 0–5μm bin than all other bins (post-hoc using LSD: $p < 0.05$ for all). n = 1390 new spine-MRS pairs / 2475 new spine-nonMRS pairs at 99.5th percentile; 1734 new spine-MRS pairs / 2131 new spine-nonMRS pairs at 95th percentile. **(f)** The functional separation of transient and sustained new spines with respect to their co-activity with nearby spines is comparable for both thresholds. 95th percentile cutoff: as in other conditions, new spine-MRS pairs are more co-active than new spine-nonMRS pairs (3-way ANOVA, main effect of MRS status: $F = 24.91$, d.f. = 1, $p = 7e\text{-}7$), and pairs are differentially co-active at closer distances (main effect of distance: $F = 5.77$, d.f. = 5, $p = 3e\text{-}5$). Only surviving new spines show a significant negative correlation with distance (surviving new spine-MRS vs. distance, $r = -0.16$, $p = 2e\text{-}5$ [Spearman's: $r = -0.13$, $p < 0.001$]; surviving new spine-nonMRS vs distance, $r = -0.13$, $p = 0.003$ [Spearman's r = 0- 0.13, $p = 0.003$]; transient new spine-MRS co-activity vs distance, $r = -0.07$, $p = 0.2$ [Spearman's $r = -0.09$, $p = 0.13$], transient new spine-nonMRS co-activity vs. distance: $r = -0.04$, $p = 0.54$ [Spearman's $r = -0.06$, $p = 0.38$], Pearson's correlation coefficient). n = 339 transient new spine-MRS pairs / 560 transient new spine-nonMRS pairs / 618 sustained new spine-MRS pairs / 1171 sustained new spine-nonMRS pairs at 99.5th percentile; 451 transient new spine-MRS pairs / 448 transient new spine-nonMRS pairs / 833 sustained new spine-MRS pairs / 956 sustained new spine-nonMRS pairs at 95th percentile. **(g)** The fraction of activity occurring during movements is highly similar for different MRS thresholds. The fraction of new spine-MRS co-activity events occurring during movements is higher than the fraction of new spine-only (95th: $p = 1e\text{-}6$; 99.5th: $p = 7e\text{-}5$) and MRS-only (95th: $p = 0.003$; 99.5th: $p = 2e\text{-}4$, rank-sum test) events. n = 34 fields. Median ± 95% CI. **(h)** The correlation of movements coinciding with new spine-MRS co-activity with the learned movement pattern (LMP) is robust to MRS threshold. Both thresholds yield MRSs that, when co-active with new spines, signal movements more similar to the LMP than movements lacking such activity (95th: $p = 0.002$, 99.5th: $p = 0.0009$, sign-rank test). n = 34 fields. Individual data points correspond to the median correlation of all movements concurrent with co-activity (blue) or lacking such activity (red) of each imaged field. Overall median of each group shown in color-coded line.

nature portfolio

Corresponding author(s): Takaki Komiyama

Last updated by author(s): Oct 22, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Two-photon imaging was performed using ScanImage (v5) software running on MATLAB (2018a). Confocal images were acquired using FluoView software (version 4.02). Electron microscopy images were acquired using Gatan 3View software (version 3.22). Behavioral data was acquired via the continuous monitoring of lever position via a LabJack acquisition device and Ephus (legacy version from original publication) software running on MATLAB (version 2011b, Mathworks) working with custom code running on LabVIEW (version 9.0, National Instruments) to monitor threshold crossing. The behavioral setup was controlled by MATLAB software (Dispatcher).
Data analysis	Analysis of two-photon data was performed using a combination of previously published code from the lab and additional custom code to fit the needs of the current experiments (all in MATLAB), as well as FIJI/ImageJ software (version 2.3, NIH) for reference in dendritic structural analysis. Electron microscopy data was processed and analyzed using a combination of IMOD (version 4.9) running through Cygwin terminal (version 3.3) and Amira (version 2019/2020) software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data and code available upon request from corresponding author.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes required for this study were estimated based on pilot studies, but no formal statistical tests were run to pre-determine sample size. However, sample sizes were similar to those reported in similar publications. The intrinsic failure rate of the experiments, the success rate of identifying new dendritic spines over learning, and signal-to-noise ratio of the imaging sensor were critical factors in initial estimates. Pilot experiments were conducted with approximately 5 animals, and sample sizes were estimated accordingly.
Data exclusions	Data were only excluded from the data used in this study if cellular health showed commonly cited signs of deterioration, or if the fluorescent sensor used showed obvious signs of photo-bleaching (i.e visibility was considerably lower on during or across imaging sessions so as to prevent analysis). Parameters for the experiments (e.g. laser power) were selected based on these viability criteria during pilot experiments not presented in this manuscript, whereafter individual occurrences of data exclusion were all recorded and reported in the Methods section under "Data Exclusion"
Replication	Trends in data were evaluated based on pilot experiments (as described in 'Sample Size' above), and reproduction of the data was considered successful if the same trends were observed in the rest of the samples. All of the data presented in this manuscript adhered to this process, and were thus considered successful replications.
Randomization	Animals used in this study were not selected based on any other prerequisite features other than general animal wellbeing (e.g. normal grooming and social behavior, no obvious infections, etc.) for allocation into a particular experimental group. Control experiments presented in this study were performed on randomly selected mice and interleaved with learning-group mice. When possible, equal numbers of mice from each cage were used for each group so as to minimize batch effects of each cohort.
Blinding	When possible, independent analyses were performed blinded. Electron microscopy analyses - after localization of the target portion of the tissue - were performed blinded to the labels from 2-photon data by individuals removed from the 2-photon analysis. Un-blinding was only performed after data analysis was complete. Experimenters were not blinded to data collection conditions, as different hardware or software settings were central to the experimental design. However, decision parameters accessible during the experiments (e.g. dendrite selection) were: 1) optimized for data quality only, 2) conceptually many steps removed from the central parameters used in this study (e.g. movement-relatedness), making it extremely unlikely that the data could be biased in one predictable direction, and 3) impossible to use as a priori prediction criteria due to learning-related changes. Thus, knowledge of conditions is extremely unlikely to produce bias in our measurements. Analysis of 2-photon data was not blinded to the condition of the experiment. This is due to the fact that the primary experimenters also performed all the analysis; the repeated exposure to and familiarity with imaging fields makes blinding to the experiment ID impossible, and since the experimenter could not be blinded to the experiments themselves (for the reasons described above), they cannot be reasonably blinded during analysis. However, ROI drawing and subsequent analyses were compared across experimenters with nearly identical results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	All animal procedures were performed in accordance with guidelines set forth and protocols approved by the UCSD Institutional Animal Care and Use Committee and the National Institutes of Health. Mice (mus musculus, c57bl/6, males and females), 6 weeks and older were group-housed in disposable cages with standard bedding in a temperature- and humidity-controlled room (~21°C and 42% humidity) with a reversed ligh cycle (10am-10pm: dark). All experiments were performed during the dark cycle. After surgeries, animals were singly housed. Male and female mice were selected randomly for surgical preparation, with no further selection criteria other than surgery outcome.
Wild animals	This study did not involve wild animals
Field-collected samples	This study did not involve animals collected from the field
Ethics oversight	UCSD Institutional Animal Care and Use Committee

Note that full information on the approval of the study protocol must also be provided in the manuscript.