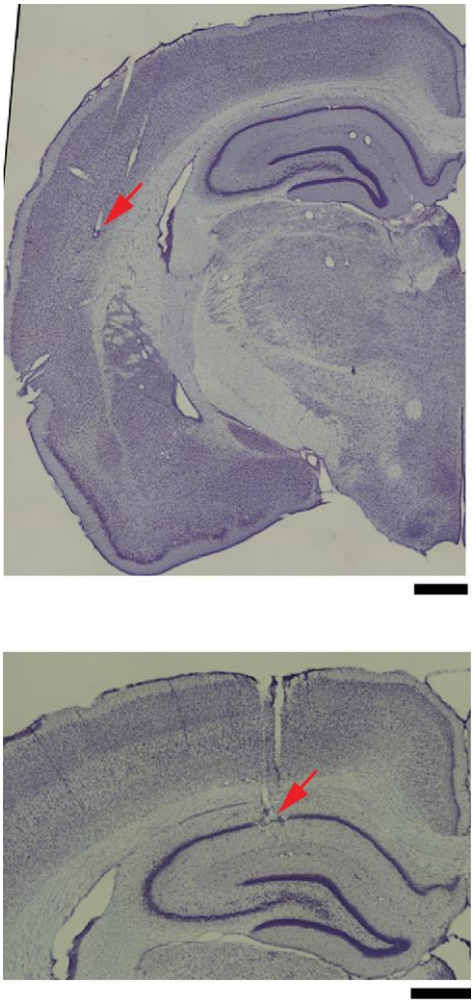
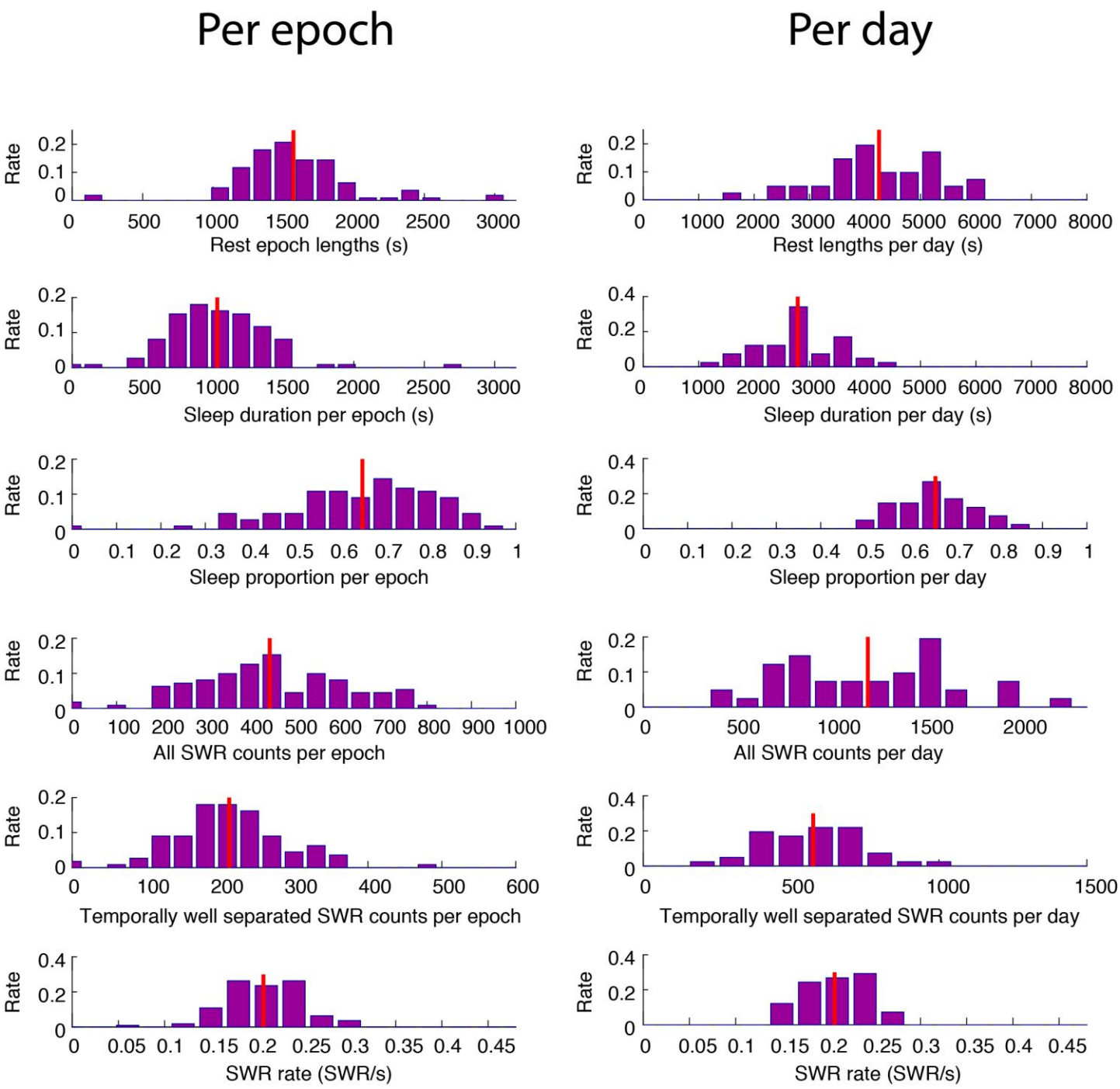


Supplementary Figure 1

Histology

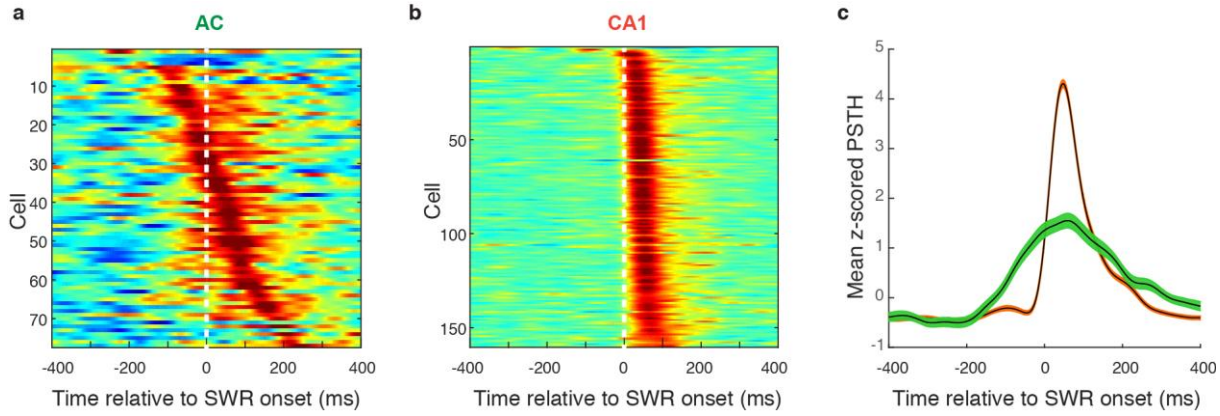
Nissl stained histological coronal sections showing example electrode tracks and lesion locations. Top: auditory cortex, Bottom: dorsal CA1. Scale bars are 1 mm.



Supplementary Figure 2

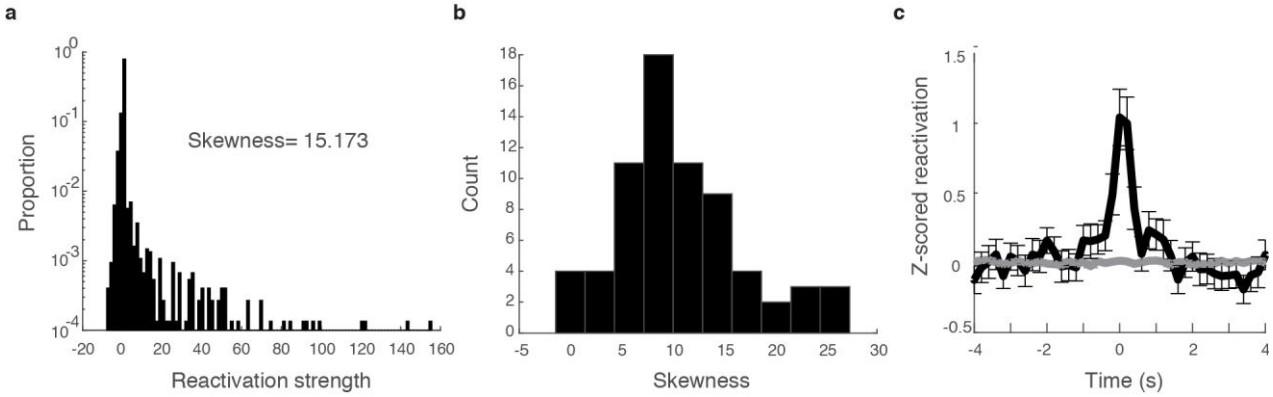
Rest, sleep and SWR counts

Rest, sleep and SWR data per silent sleep epoch (left) and per day across silent sleep epochs (right). From top to bottom: Silent rest epoch durations, NREM sleep during silent rest epochs, NREM sleep proportion out of rest time, all SWR counts, temporally well-isolated SWR counts (after removal of SWRs at temporal offsets of <500ms from other SWRs, as being used for SWR modulation and GLM predictions), SWR rates of temporally well-isolated SWR counts.



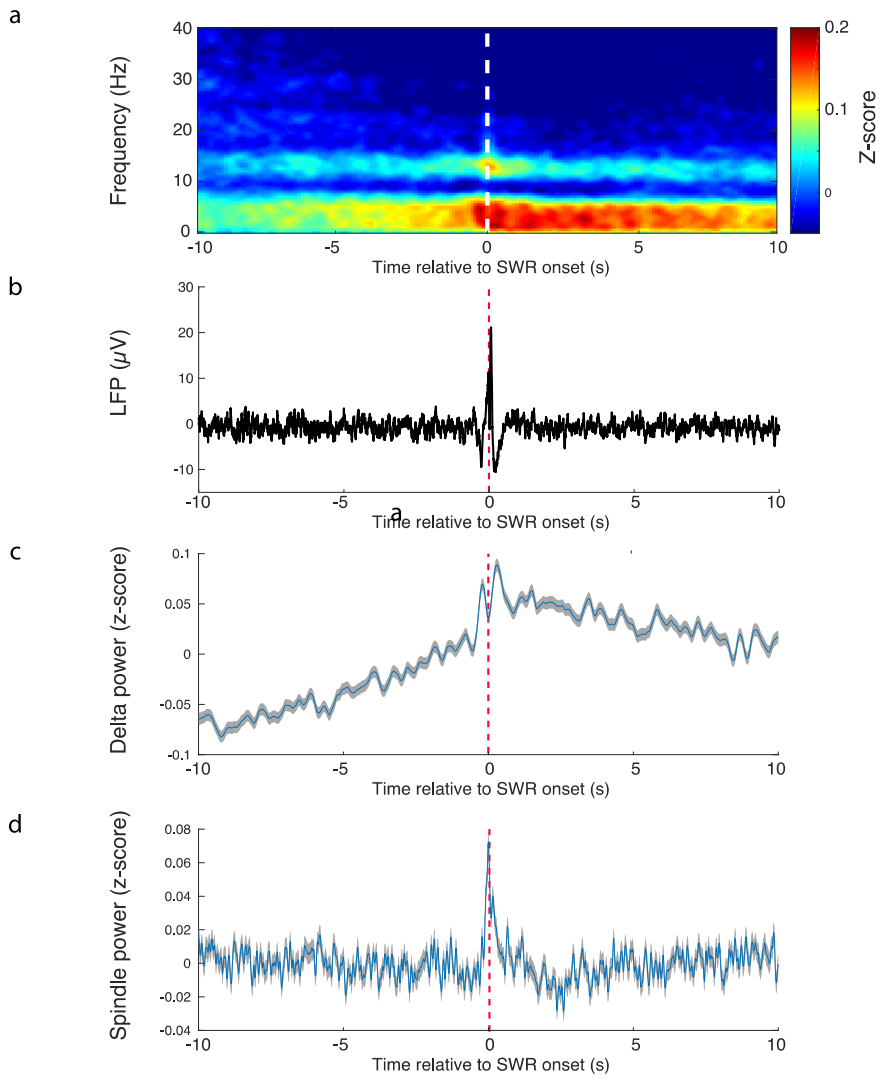
Supplementary Figure 3

Same temporal pattern of SWR modulation using alternative SWR-modulation criterion
Parallel figures to Fig. 2a-c but when applying an SWR-modulation criterion of $P<0.05$ across entire daily datasets. The same temporal patterns are apparent.
(a) SWR-triggered mean spiking histograms of all SWR-modulated AC neurons. Each row shows the z-scored mean SWR-triggered spiking histogram of one neuron. Neurons are ordered by timing of peak firing in -250 ms to 250 ms window relative to SWR onset
(b) Same as (a) for CA1 neurons.
(c) Mean z-scored SWR-triggered spiking histogram across all SWR-modulated neurons for AC (green) and CA1 (orange). Shaded area indicates sem. Traces correspond to the mean±sem of data in (a) and (b).



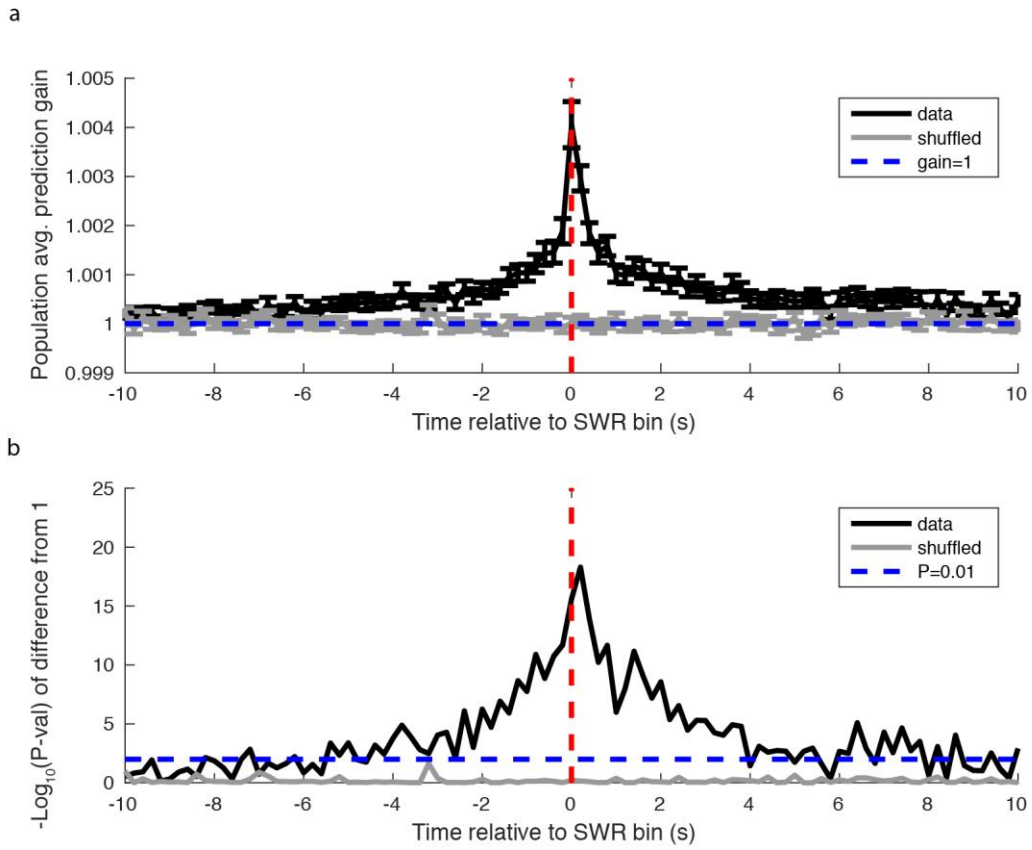
Supplementary Figure 4

Reactivation of awake patterns during sleep in auditory cortex
(a) Histogram of reactivation amplitudes for the example in Fig. 1g.
(b) Distribution of reactivation skewness values across all epochs
(c) SWR-triggered reactivation as in Fig. 1h, but using full correlation matrix from awake state as template instead of first principle component. This alternative analysis demonstrates that the reactivation shown in Figure 1 is not dependent on the choice of the first principle component as the reactivation measure.



Supplementary Figure 5

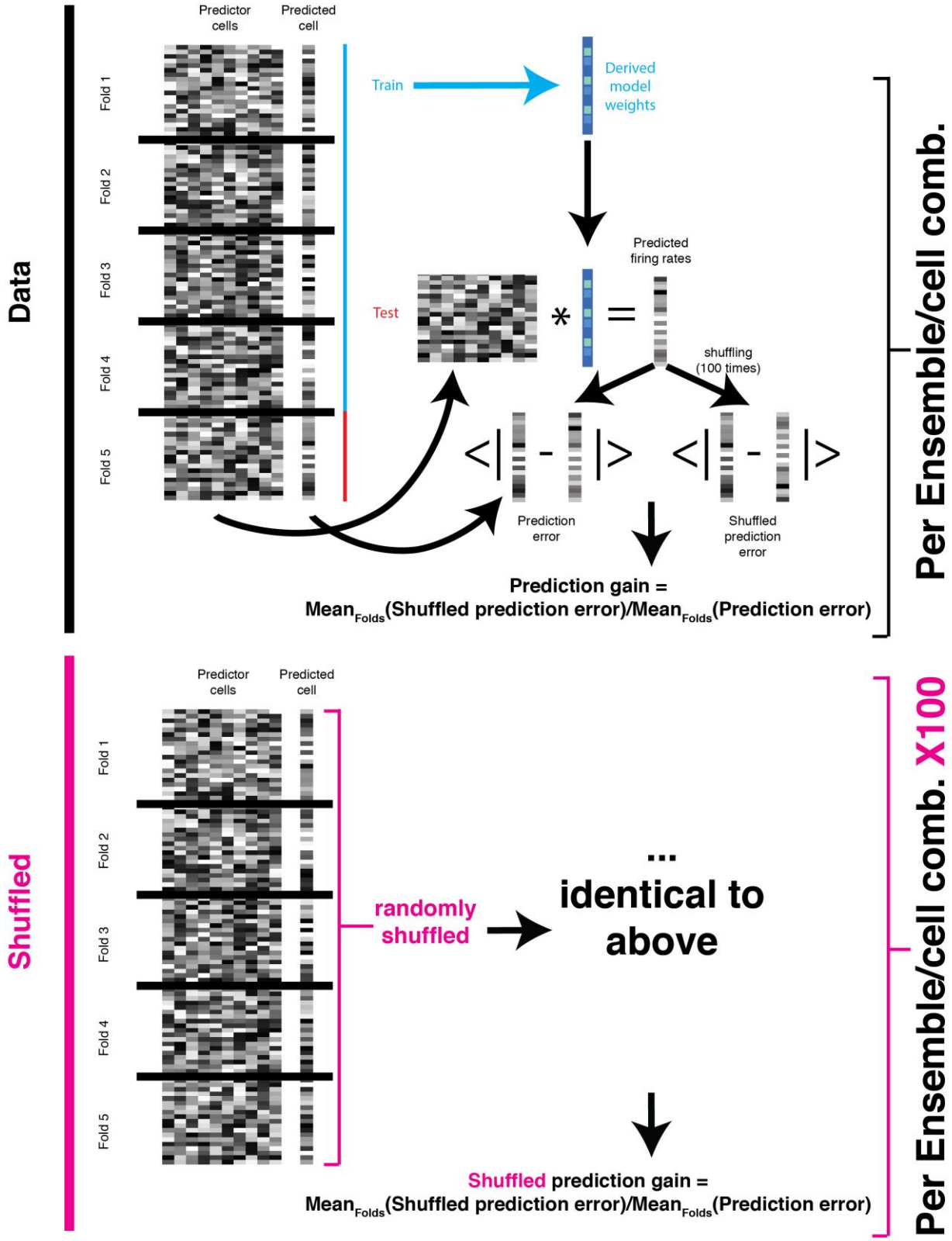
SWR-triggered LFP in auditory cortex
(a) Mean z-scored SWR-triggered spectrogram of AC local field potential. We extracted the LFP around all SWRs across the dataset and calculated the spectrogram around each SWR. Spectrograms were averaged across tetrodes in the same recordings, and z-scored by the average power across frequencies. The spectrogram shows an increase in power in the delta and spindle bands during SWRs.
(b) Mean SWR-triggered raw local field potential
(c) SWR-triggered power in the delta band. We filtered the LFP in the delta band (1-4 Hz) and identified its envelope using the Hilbert transform. We extracted the delta envelope around all SWRs in the dataset and z-scored each of them. The shaded area is mean±SEM. As seen in 'a', the delta power increases during SWRs, with an increase before SWR onset.
(d) SWR-triggered power in the spindle band. Similar to 'c' but in the spindle band (8-20 Hz). As seen in 'a', power in the spindle band shows an increase during SWRs.



Supplementary Figure 6

Predicting SWR occurrence from AC ensemble spiking patterns

To predict SWR occurrences we used a cross-validated GLM approach. For each silent sleep epoch we binned all AC spike counts into 200 ms bins and created a vector of 0/1's of the same length indicating bins in which SWRs occurred. We used 5-fold cross validation: the data was divided into 5 equally sized datasets, balanced in the number of SWRs. For each fold, 4 folds were used to train a binomial GLM model to predict SWR occurrence from the AC firing patterns, and the remaining fold was used as a test set. To quantify performance of each fold, the GLM model was applied to predict the likelihood of SWR occurrences in all bins from the AC ensemble spiking in the test fold. The test error was defined as the mean absolute difference between the prediction and real rates. Test errors were averaged across folds. As control, we calculated the error across 100 shuffles of the test sets. Prediction gain for each dataset was defined as the error of the shuffled data divided by the error of the original data.
(a) Prediction gain of SWR occurrence. Each column at time t depicts the prediction gain of SWR occurrence at time 0 from predictor AC patterns at time t. Prediction was measured as described above. Gray is shuffled data, in which the AC patterns were shuffled relative to SWR times. Error bars are SEM.
(b) Significance of median larger than 1, corresponding to the data in 'a'. P-values were measured using a one-sided Wilcoxon signed-rank test and presented as $-\log_{10}(P)$

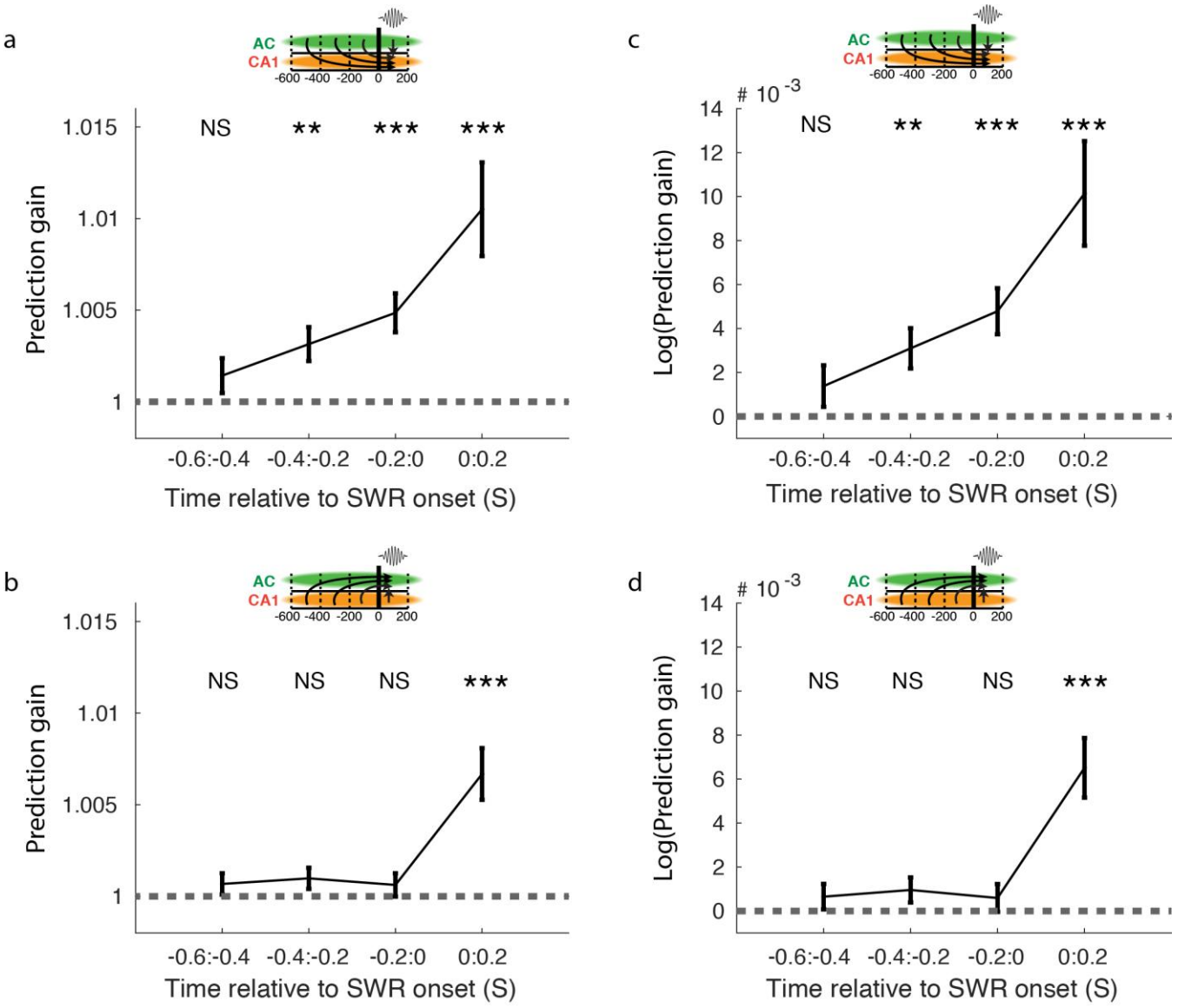


Nature Neuroscience: doi:10.1038/nn.4457

Supplementary Figure 7

Visual illustration of GLM cross-validation method

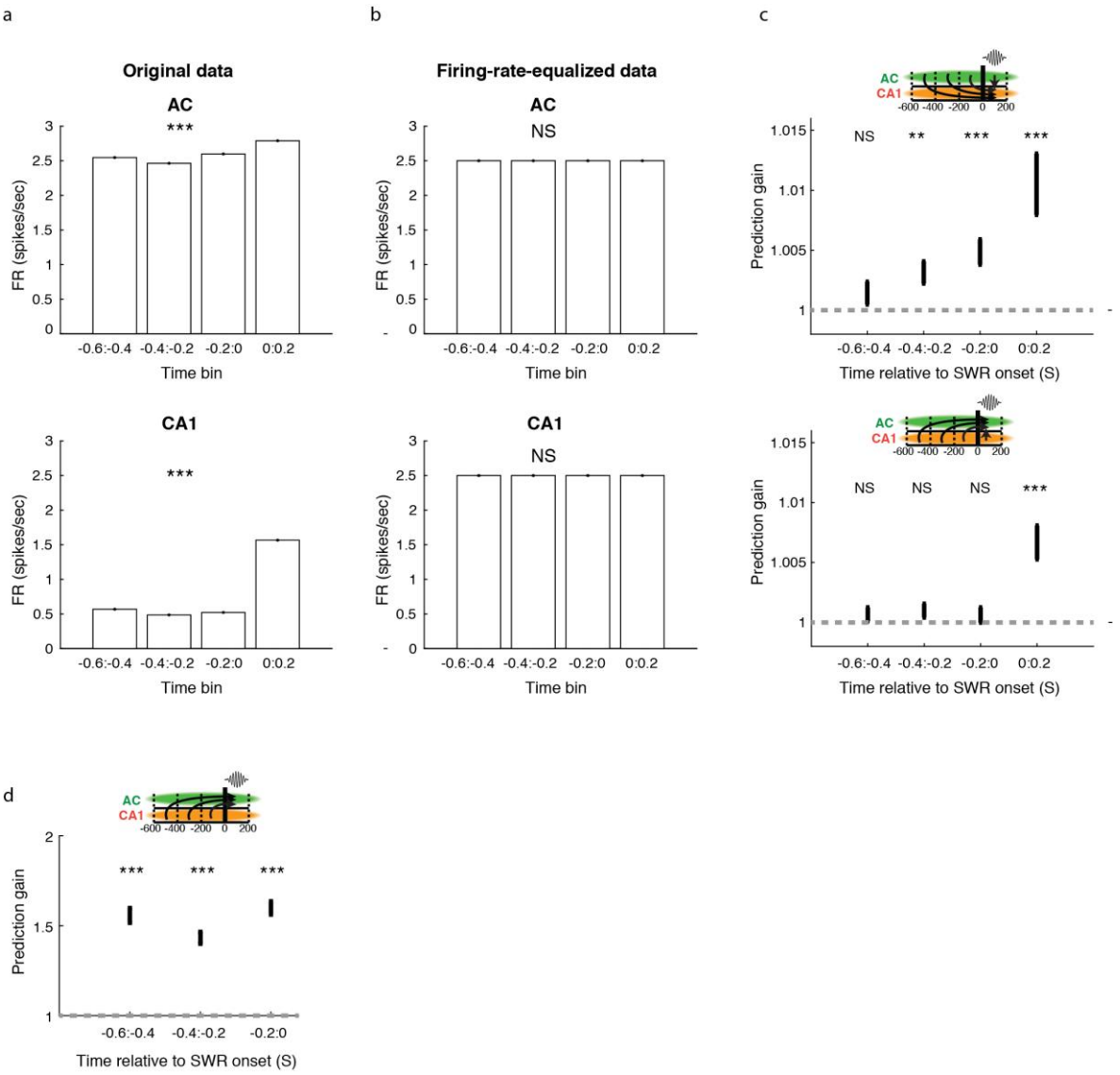
The overall procedure was as follows: For each combination of predictor ensemble and predicted cell in a given time window, the predictor data consisted of the ensemble patterns (binned spiked counts) across SWRs, and the predicted data consisted of the single-cell SWR responses (binned spiked counts) across SWRs. We then performed 5-fold cross validation by randomly partitioning the daily SWRs into 5 equal-sized sets. For each of the 5 folds, we used 4 of 5 folds to train the GLM model, and the remaining fold to test. For the test phase, the model derived from the training phase was applied to the predictor ensemble data in the test set, yielding predictions for the predicted cell firing across SWRs. The prediction error was the mean absolute difference between the predicted spike counts and the real spike counts. A baseline prediction error was calculated as the mean prediction errors across 100 random shuffles of the predicted firing rates across SWRs. The real and shuffled prediction errors were then averaged across the 5 folds. Prediction gain for one predictor ensemble/predicted cell combination in one time window was defined as the shuffled prediction error divided by the real prediction error. For comparison, we repeated the exact same procedure described above on 100 random shuffles of the entire original dataset, where shuffling entailed random matching of activity patterns in the predictor and predicted data.



Supplementary Figure 8

Alternative statistical tests for GLM prediction significance

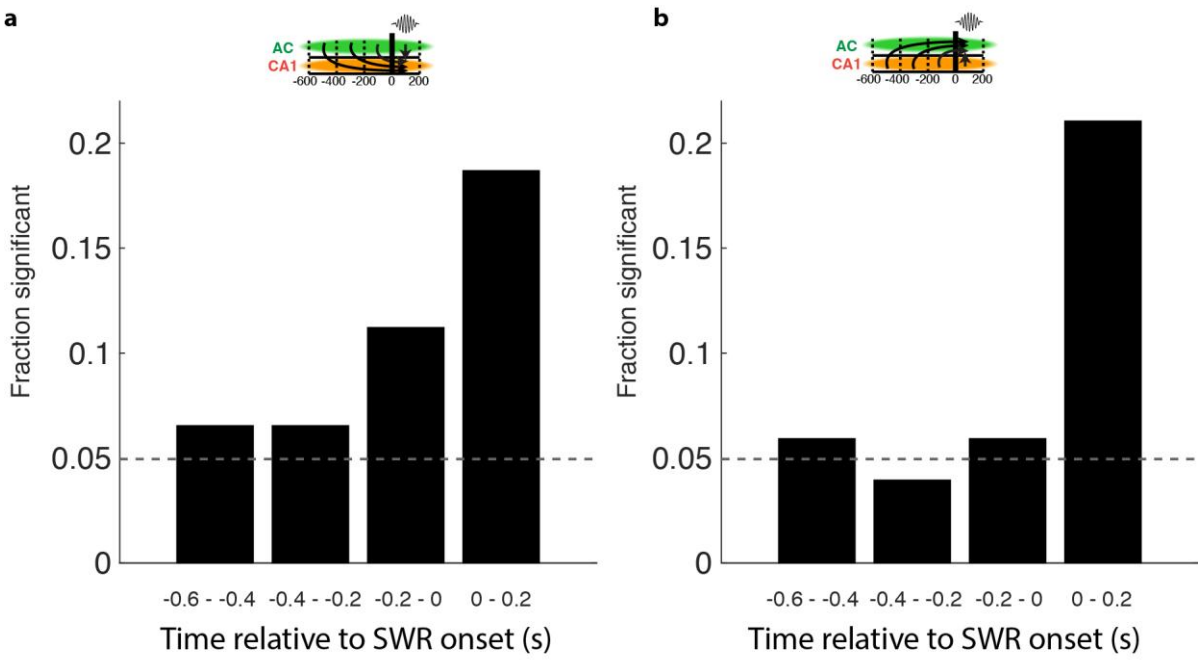
(a) Prediction of CA1 single-cell spiking during SWRs from ensemble spiking patterns in AC across varying time windows (n = 107 predicted CA1 cells). Black error bars indicate mean ± s.e.m. prediction gain for real data. Grey dashed line indicates 1 (no gain). Columns represent varying time windows used as predictor data; predicted data is always the SWR time window (0 – 200 ms). This panel parallels Fig. 2g, but significance of prediction gain is tested against 1 using a two-tailed Wilcoxon signed rank test. CA1 spiking during SWRs could be significantly predicted better than shuffled data from AC ensemble spiking patterns in the -400 to -200 ms window (p= P=0.0061), -200 to 0 ms window (P=9.35e-5) and 0 to 200 ms window (P=1.22e-8), but not from the -600 to -400 ms window (P=0.7) (two-tailed Wilcoxon signed rank test of difference from 1).
(b) Same as (a), but in the opposite direction: prediction of AC single-cell spiking during SWRs from ensemble spiking in CA1 (n=152 predicted AC cells) using statistical test of difference from 1. AC spiking during SWRs could not be significantly predicted from pre-SWR CA1 ensemble spiking patterns, but could be predicted from CA1 spiking during the SWR (-600 to -400: P=0.845; -400 to -200: P=0.371; -200 to 0: P=0.769; 0 to 200: P=1.05e-6, two-tailed Wilcoxon signed rank test of difference from 1)
(c) Same as (a), but using an additional test of whether log(prediction gain) is different from 0. Results show the same pattern: -600 to -400: P=0.705; -400 to -200: P=0.0063; -200 to 0: P=9.6e-5; 0 to 200: P=1.29e-8, two-tailed Wilcoxon signed rank test of difference from 0.
(d) Same as (b), but using an additional test of whether log(prediction gain) is different from 0. Results show the same pattern: -600 to -400: P=0.851; -400 to -200: P=0.377; -200 to 0: P=0.76; 0 to 200: P=1.1e-6, two-tailed Wilcoxon signed rank test of difference from 0.



Supplementary Figure 9

Firing rate differences do not account for prediction patterns

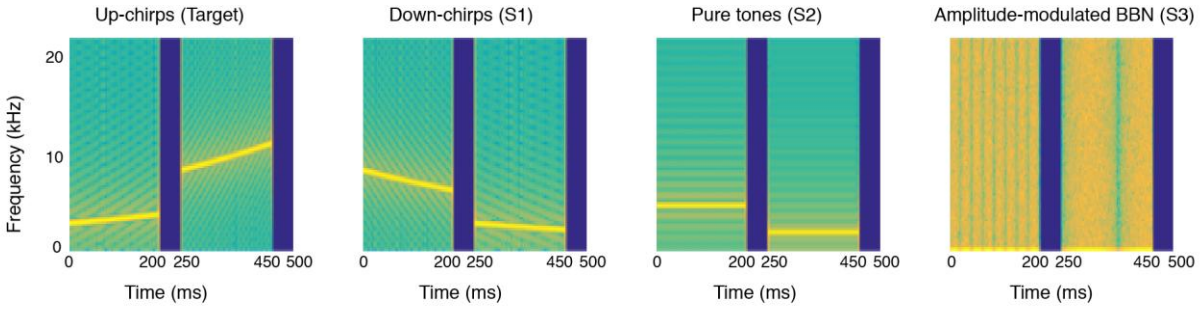
(a) Firing rates of AC (top) and CA1 (bottom) cells across time bins relative to SWR times. Data are pooled across cells, SWRs and experiments. Firing rates in both AC and CA1 show a significant difference across time bins (P<10-10, one-way ANOVA).
(b) Same as in (a) but after firing rate equalization. For firing rates of a given predictor ensemble in a given time window, equalization was done by multiplying the firing rates of all neurons in the ensemble across all SWRs by a factor of 2.5/(observed mean) to yield a resulting mean of 2.5 across cells and SWRs. The resulting equalized firing rates did not show a significant difference across time bins (p>0.99, one-way ANOVA).
(c) Prediction gains paralleling Fig. 2g-h and Suppl. Fig. 8, when applied on firing-rate equalized data (b). The same prediction patterns across time bins and brain regions were observed as with original data. In the AC to CA1 direction (top): -600 to -400: P=0.697; -400 to -200: P=0.0061; -200 to 0: P=9.35e-5; 0 to 200: P=1.22e-8, two-tailed Wilcoxon signed rank test of difference from 1. In the CA1 to AC direction (bottom): -600 to -400: P=0.845; -400 to -200: P=0.371; -200 to 0: P=0.769; 0 to 200: P=1.05e-6, two-tailed Wilcoxon signed rank test of difference from 1.
(d) Upper bound of prediction gains of firing patterns in pre-SWR CA1 windows predicting AC SWR spiking. Upper bound was estimated by artificially sorting the original firing rates of each predictor cell across SWRs to match the firing rate order of the predicted cell. The upper bound of ~50% prediction gain indicates that lack of prediction in this direction in the original data is not due to lack of sufficient spikes. $P<10^{-10}$ in all three windows, two-tailed Wilcoxon signed rank test of difference from 1



Supplementary Figure 10

Rate of significantly predictive ensemble–cell combinations

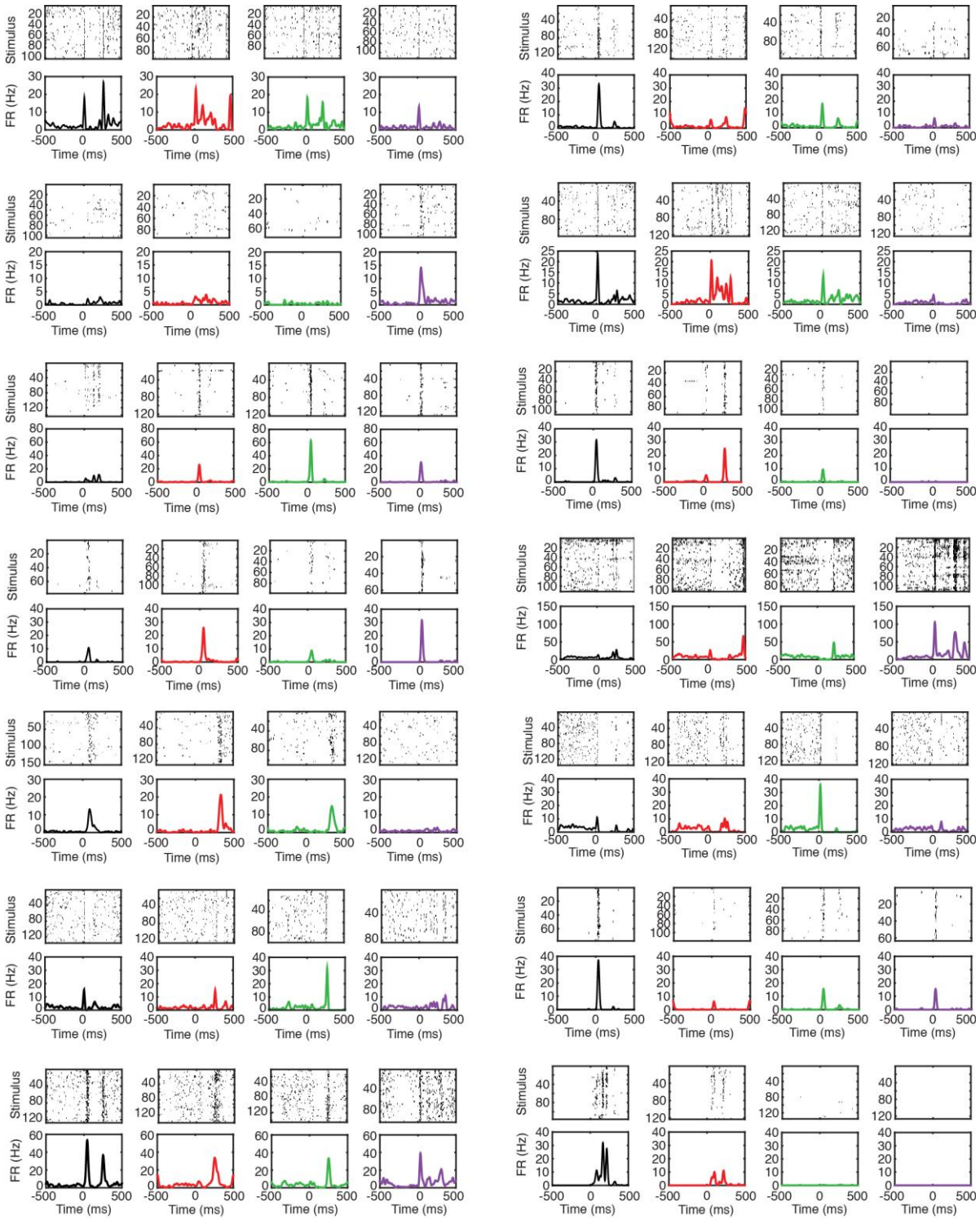
a. Fraction of individually significant (P<0.05) ensemble-cell predictions in the AC->CA1 direction. Significance was tested by comparing the prediction gain of each cross-validated model to the prediction gains obtained from 1000 shuffles. The P-value was defined as the rate of shuffles with gain higher than the data.
b. The same as in ‘a’ for the CA1->AC direction.



Supplementary Figure 11

Sound spectrograms

Spectrograms of the 4 auditory stimuli used.

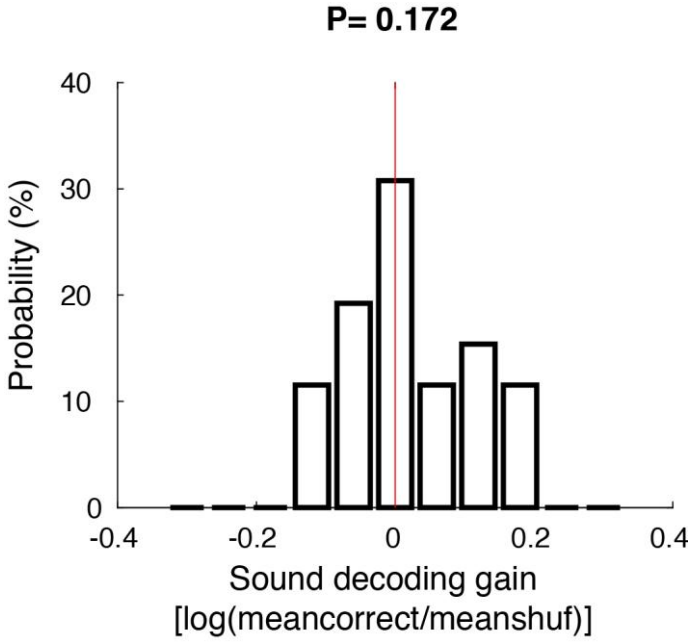


Supplementary Figure 12

Additional examples of AC sound responses during sleep

Responses of 14 example AC cells to the auditory stimuli presented during sound sleep sessions. Each group of 8 panels shows the responses of one AC cell to the 4 different stimuli in one sleep sound epoch as in Fig. 3b. These data illustrate the strength, temporal profiles and diversity of the observed responses.

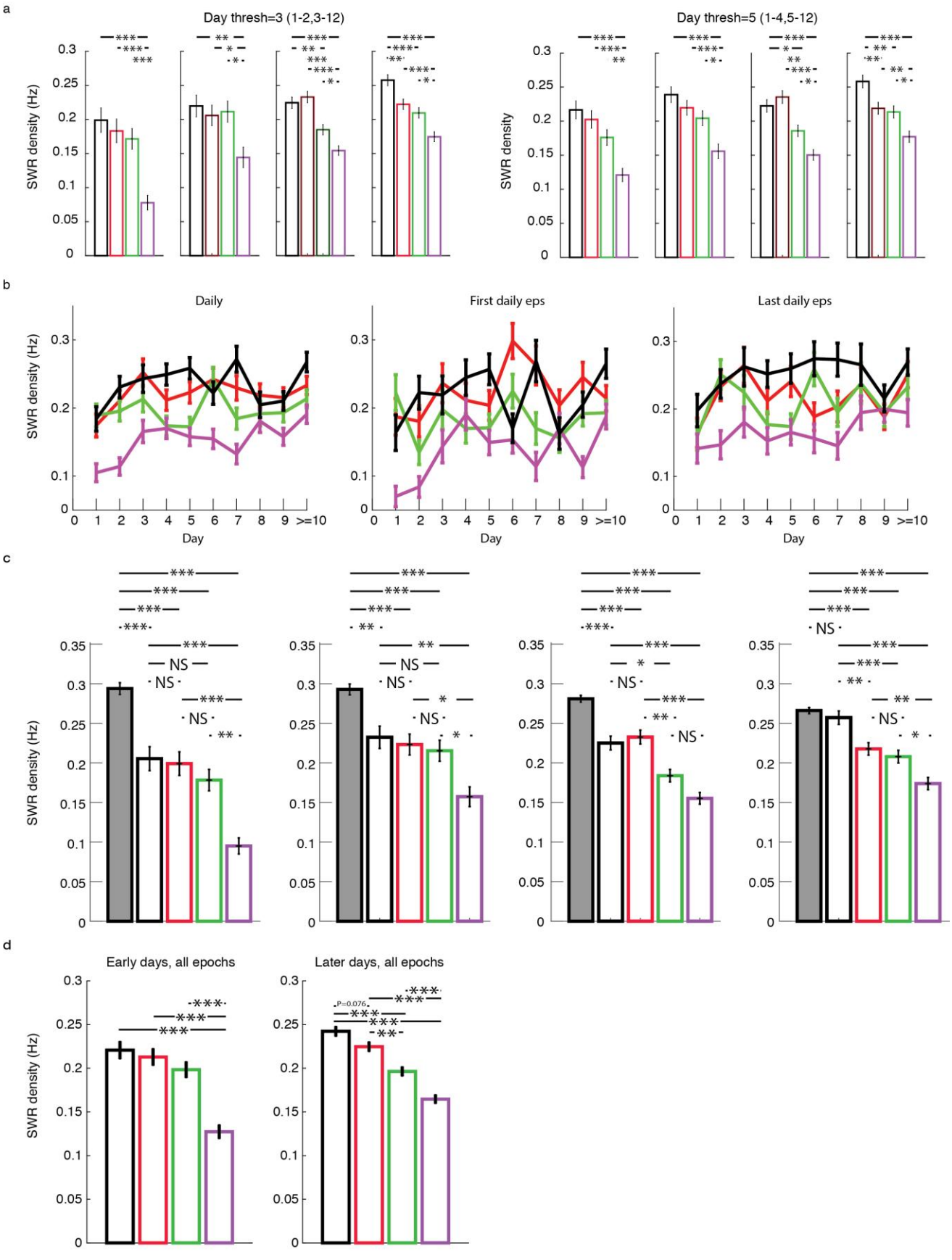
Nature Neuroscience: doi:10.1038/nn.4457



Supplementary Figure 13

Sound decoding from CA1 ensemble patterns during SWRs

Decoding was performed as in Figure 3e, but for CA1 ensemble firing patterns during the SWR. Decoding gain was not significantly different from 0 (P=0.172, cross-validated linear discriminant analysis, population test for mean larger than 0 using one-tailed t-test)



Nature Neuroscience: doi:10.1038/nn.4457

Supplementary Figure 14

Stimulus-specific SWR density across learning

(a) Stimulus-specific SWR density across learning, as in Fig. 5b, using alternative cutoffs for First/Later days, showing similar results. Left: First days are 1-2, Later days are >=3. Right: First days are 1-4, Later days are >=5.
(b) SWR density across days, for daily (first and last epochs combined, left), first daily epochs (middle) and last daily epochs (right). SWR density of non-target stimuli did not show a decreasing trend across days.
(c) Stimulus-specific SWR density across learning, as in Fig. 5b, but including SWR density in no-sound times (gray filled bars). Note that at later stages of learning, SWR density during target sound presentation did not differ from during no-sound.
(d) Stimulus-specific SWR density in First versus Later days. The first and last daily epochs were combined.

Nature Neuroscience: doi:10.1038/nn.4457