Optical and computational dissection of emergent prefrontal rewiring to encode fear memory

Masakazu Agetsuma[1, 2, 3, 4] *, Issei Sato[5], Yasuhiro R Tanaka[6], Luis Carrillo-Reid[7], Atsushi Kasai[8], Yoshiyuki Arai[3], Miki Yoshitomo[1], Takashi Inagaki[1], Hitoshi Hashimoto[8, 9, 10, 11, 12], Junichi Nabekura[1], and Takeharu Nagai[3]

1, Division of Homeostatic Development, National Institute for Physiological Sciences, 38 Nishigohnaka Myodaiji-cho, Okazaki, Aichi, 444-8585, Japan
2, Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan
3, SANKEN (The Institute of Scientific and Industrial Research), Osaka University, Mihogaoka 8-1, Ibaraki, Osaka 567-0047, Japan
4, Division of Molecular Design, Research Center for Systems Immunology, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
5, Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo. 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
6, Brain Science Institute, Tamagawa University, 6-1-1 Tamagawagakuen, Machida, Tokyo, 194-8610, Japan
7, Instituto de Neurobiologia, National Autonomous University of Mexico, Boulevard Juriquilla 3001, Juriquilla, Queretaro, CP 76230, Mexico
8, Graduate School of Pharmaceutical Sciences, Osaka University, Yamadaoka 1-6, Suita, Osaka 565-0871, Japan
9, United Graduate School of Child Development, Osaka University, Kanazawa University, Hamamatsu University School of Medicine, Chiba University, and University of Fukui, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
10, Division of Bioscience, Institute for Datability Science, Osaka University, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan
11, Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
12, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

*, Corresponding author: Masakazu Agetsuma
(e-mail) age@nips.ac.jp (phone) +81-564-55-7854

・Using chronic two-photon imaging and sparse modeling, we successfully discriminated neural ensembles encoding conditioned responses (CR ensembles).

・We confirmed that the CR ensembles were distinct from neurons encoding regular locomotion and emerged as a result of fear conditioning.

・Enhanced coactivity and functional connectivity were specifically observed in CR ensembles as a result of fear conditioning.

・Further graphical modeling revealed the signature of the construction of the conditioned stimulus-unconditioned stimulus (US) association circuit by rewiring around the US responsive pattern completion cells in an experience-dependent manner.

Abstract

The prefrontal cortex regulates various emotional behaviors and memories, and prefrontal dysfunction can trigger psychiatric disorders. While untangling the internal network may provide clues to the neural architecture underlying such disorders, it is technically difficult due to the complexity and heterogeneity of the network. Here we propose an optical and computational dissection of the internal prefrontal network utilizing chronic two-photon imaging and a sparse modeling algorithm, which enabled the discrimination of newly emerged neuronal ensembles specifically encoding conditioned fear responses. Further graphical modeling revealed that neurons responding to the unconditioned stimulus during fear conditioning became a core of the ensembles with an enhanced capability for pattern completion, demonstrating activity-dependent rewiring upon the associative learning.

Introduction

The prefrontal cortex (PFC) is an important brain region that regulates various types of behaviors and memories: aversive and appetitive memories, decision-making, and higher-order cognitive functions[1-5]. The importance of the PFC is evolutionarily conserved in mammals, from humans to primates to rodents[4, 6-8], although the functional and anatomical analogy across species is still debated[7, 9-12]. Dysfunction of the PFC impairs the ability to organize positive and negative valences, and may lead to various psychiatric diseases, including depression, schizophrenia, and fear-related disorders (e.g. post-traumatic stress disorder)[13-17]. How the PFC can distinctively encode and regulate such diverse information, however, remains unclear. Information processing based on the neural population, including neurons with mixed selectivity, is suggested to be key for prefrontal computations[18], but further studies are required to uncover the mechanism underlying the acquisition of novel emotional memories and related functions. Resolving the mechanisms underlying the implementation of newly acquired aversive memories in the internal PFC network in parallel with existing information encoders could contribute to elucidating the neural architecture involved in such psychiatric disorders.

In rodents, the dorsal part of medial prefrontal cortex (dmPFC, also called prelimbic cortex) is important for the retrieval of fear memory[19-24]. Previous studies revealed activated individual neurons[25] or enhanced synchrony of neural populations[22] in the dmPFC during the conditioned response (CR, i.e. fear memory-evoked freezing behavior). Pharmacological and optogenetic silencing impair the CR[19, 20], suggesting that fear memory is normally stored using the dmPFC network.

Therefore, in the present study, to untangle the computational architecture in the internal prefrontal networks dealing with multiple information in parallel, we investigated how the dmPFC network newly and distinctively encodes fear memory in an experience-dependent manner. Chronic two-photon neural activity imaging performed in vivo to simultaneously record neural activities from hundreds of neurons was combined with a regularization and variable selection algorithm to discriminate the neural ensembles specifically encoding the CR in a dmPFC internal network. Using this optical and computational dissection method, we successfully identified neural ensembles encoding the CR (CR ensembles). We confirmed that the CR ensembles were distinct from neurons encoding regular locomotion and emerged as a result of fear conditioning. The CR ensembles were predictive of animal behavior during fear memory retrieval, while the predictiveness collapses during the inter-trial-interval. Both enhanced coactivity and functional connectivity also emerged specifically in the CR ensembles, suggesting the possible rewiring behind the memory consolidation. Interestingly, neurons responsive to an unconditioned stimulus (US) became predominantly involved in the CR ensemble, and further graphical modeling revealed that those neurons also became more strongly connected to the internal CR network and more predictive of the conditioned stimulus (CS). Altogether, our findings revealed that neurons responding to the US during fear conditioning became a core of the CS-US association circuit with an enhanced capability for pattern completion, encoding fear-memory-driven behaviors in an experience-dependent manner in parallel with the pre-existing regular locomotion network.

Results

To uncover the mechanisms underlying how neural populations, or ensembles, acquire and regulate specific fear memory distinctively to the other information that the mPFC processes in parallel, we developed a system to perform cued-fear conditioning and memory retrieval under a two-photon microscope to directly record the neural activities of hundreds of neurons with single-cell resolution in awake mice, and compared the neural activities of the same sets of neurons before, during, and after fear conditioning (Figs.1, 2). The mice were head-fixed under the objective, and placed on a running disk through which the mouse locomotion (whether the mouse was locomoting, stationary, or expressing a freezing response) was recorded (Fig. 1A). Tones and foot shocks were delivered as the CS and US, respectively. Two different tones were used, one was associated with the US (CS+) and the other was not (CS–), as described in previous studies[22, 26]. Behavioral analyses revealed that the mice learned to decrease their locomotion specifically during the CS+, only after the fear conditioning (day [D]4) as a conditioned response (CR), but not before the fear conditioning (D3) (Fig.1B, C). In most of the analyses, the neural representation during the first three trials on D3 (D3-early [D3E]) were compared with those during the first three trials on D4 (D4E) to investigate the change before and after the fear conditioning and memory consolidation, while the data obtained during the last three trials on D3 (D3-late [D3L]) were used to assess the late conditioning phase. The data obtained during the last three trials on D4 (D4-late [D4L]) were used to assess the extinction phase[22, 26], when the CS+-evoked suppression of the mouse locomotion observed during D4E was extinguished after repeated exposure to the CS+ (Fig. 1D, E). Overall, these behavioral data established that our behavioral system and the fear conditioning protocol were useful for observing a change in the neural representation after the associative memory consolidation.

To monitor the neural activities in the dmPFC by two-photon microscopy, we implanted a 2-mm microprism along the midline to optically access the dmPFC region. Although the size of the prism was larger than that of prisms used in previous work[27], there was sufficient space between the bilateral dmPFC to enable the smooth insertion of the prism without injuring the prefrontal area and callosal fibers (Fig. 2A). Using a genetically encoded Ca^{2+} indicator, GCaMP6f, expressed by an adeno-associated virus (AAV), the activities from a wide region of the prefrontal area were visualized chronically (Supplementary Movie 1). In the present study, we focused on activities in the dmPFC area (Fig. 2B).

Prior to investigating population coding in the dmPFC before and after acquisition of the fear memory, we first summarized the single-neuron responses to the CS+ and CS– (Figs. 2C-E and S1). We found that approximately 60% of neurons exhibited a significant change in neural activity during the CS+ and/or CS–, and approximately 20% of neurons showed significant responses to both the CS+ and CS–. The distributions of these types of neurons were consistent throughout the learning process (Figs. 2E and S1). This type of “mixed selectivity” (responsive to variable task-relevant aspects) has been reported in the primate PFC[18] as well as in the mouse caudal mPFC during a decision-making task[3]. The potential advantage of the mixed selectivity was proposed to enhance the number of tasks that each neural circuit, with a limited number of neurons, can handle, through high-dimensional neural representations implemented by a population of neurons[18, 28]. This encouraged us to further analyze the population coding for fear memory.

Our goal in this study was to dissect the computational architecture composed by a neural population in the dmPFC enabling the distinctive acquisition of a novel CS-US association. For this purpose, we first extracted a group of neurons encoding the CR (CR ensemble). Unlike previous studies utilizing unsupervised learning algorithms such as Principal Component Analysis or Non-Negative Matrix Factorization, which first seek embedded structures in neural data and further test which structure is most likely to correlate with or explain target behaviors[22, 29, 30], we intended to directly extract ensembles encoding the CR by a supervised and model-based machine learning algorithm, elastic net[31] (Figs. 2F-J, 3, and S2). The elastic net is a regularization and variable selection algorithm based on the regression model (Fig. 2G; see the Methods for details) and designed to automatically select variables[31], which enabled us to systematically identify neurons encoding the target behaviors. This method allowed us to directly extract not only CR ensembles but also the neural ensembles encoding regular locomotion (RL ensemble; Fig. 2K), independently in the same mice, and to compare them and verify whether neurons in CR ensembles were unique or mostly overlapped with RL ensembles (Figs. 2F, 3).

We extracted CR ensembles using data obtained during the CS+ presentation of D4E (retrieval session). Because these CR ensembles were discriminated by the data and the behavioral labels during the CS+, and not by comparison between those during CS+ and those during the presentation of other stimuli, our method did not produce any bias to the CS+ in selecting CR ensemble neurons. We evaluated the fitting and decoding performance of the obtained model, and confirmed that the obtained CR ensemble was highly predictive for the CR during the retrieval session (Fig. 2H) (mean ± SE of the prediction accuracy, 0.9450 ± 0.0265, n=7 mice; also shown later in Fig. 3F). As for the spatial distribution, the identified CR ensemble neurons were spatially intermingled over the field of view, as shown in Figs. 2H and 3A.

Another advantage of the elastic net, compared with the conventional sparse modeling Lasso, is that the hyper parameter “alpha” enables us to adjust the sparseness of the selected population, which is very important to avoid missing neurons encoding the target information, especially when an analyzed neural network includes strongly correlated neural pairs, which is likely the case for our data considering the results shown below, as well as previous electrophysiological observations[22]. When searching for the optimal alpha for each individual circuit, we intended to minimize CR-related information remaining after removing identified CR ensemble neurons (Figs. 2I, J, and S2; see also Methods). Wide range of the alpha values for each individual circuit was tested, and the decoding performance of neurons remaining after the removal of the CR ensemble neurons selected at each alpha was evaluated (Fig S2). This systematic optimization procedure revealed the general trend that a larger alpha tended to select a smaller number of CR ensemble neurons (Fig. S2B, top), and though the decoding performance of the smaller number of selected CR ensembles was very high, equivalent to that of the others (Fig. S2B, middle), the removal of such a smaller portion from the whole set of neurons was not enough to substantially diminish the information encoded by the remaining neurons (Figs. S2B, bottom, and S2D, F), suggesting that the CR was redundantly encoded in the dmPFC, while the RL was not (Fig. S3). After determining the optimal alphas for individual circuits, we observed a substantial reduction of the decodability by the neurons that remained after removing all the selected CR ensemble neurons (Figs. 2I, J, and S2).

Following the optimization of the hyper parameter alpha, we evaluated the specificity and uniqueness of the extracted CR ensembles. We confirmed that most of the neurons involved in the CR ensemble were unique and did not overlap with the RL ensemble (Fig. 3A, B).

We then conceived the hypothesis that the unique CR ensemble might dominantly and exclusively explain the behaviors of the mice during CS +-evoked memory retrieval as an encoder of the acquired fear memory. If this is true, RL ensembles, distinct from CR ensembles (Fig. 3B), should have diminished decodability for the behavior during CS+ during fear memory retrieval. To test this possibility, we checked the decoding performance of the RL ensembles for the behaviors observed during the CS+ at each of the learning steps (Figs. 3 and S4). The decoding performance by the RL ensembles to the RL was similar between pre- and post-memory consolidation (Fig. 3D). The decoding performance of the RL ensembles to the behaviors during CS+ presentation at D3E (before fear memory consolidation) was similar to that for the RL (Fig. 3D, E). In contrast, the decoding performance of the RL ensembles to the behaviors during the CS+ on D4E (during fear memory retrieval) was significantly reduced compared with that of D3E (Figs. 3C, E). There was a small, but not significant, change during the fear conditioning (D3E vs D3L; Fig. S4), and importantly, the reduced decodability of the behavior during CS+ at D4E (memory retrieval) was substantially recovered after the extinction training (no significant difference between D3E and D4L, and a significant difference between D4E and D4L; Figs. 3E and S4).

In contrast, the decodability of CR ensembles was specific to the CR and not applicable to the RL on D4 (Fig. 3F). These results established that the CR, or the behavior during the memory retrieval, was dominantly explained by the CR ensembles, and support the idea that the CR ensembles that we specifically extracted from all recorded neurons might be a dominant and specific group of neurons encoding the CR during memory retrieval, emerge after consolidation of the fear memory and are suppressed after extinction.

In these CR ensembles, we observed a slight but significant increase in CS+ activatable neurons, but no change in CS+ inactivated neurons after fear conditioning (Fig. S5). In contrast, other cells (neurons that were not included in the CR ensembles: Non-CR ensemble [Non-CRE] neurons) exhibited no significant changes in the CS+ activated neurons, with a significant increase in CS+ inactivated neurons. Neurons in the RL ensembles did not exhibit any change in CS+ responsiveness. We detected no significant change in CS– responsiveness in any of the categories. These results indicated that there might be some mechanism that makes neurons involved in the CR ensembles dominantly activated by the CS+ after memory consolidation.

To further evaluate and characterize the identified CR ensemble, we compared the change in the coactivity of the neural network by calculating the pairwise correlation coefficients (R)[32] between pre- and post-memory consolidation. We found that, after the fear conditioning, only the positively correlated fraction was enhanced specifically within the CR ensembles, and not in the outside network (Non-CRE) (Fig. S6A). Statistical analyses demonstrated that this enhancement in positive correlation after the fear conditioning, as well as the enhanced ratio of significantly and positively correlated pairs, specifically occurred in the CR ensembles (Figs. 4A and S6A-C). Analyses based on the shuffled data, where the activity of each neuron was preserved but the temporal order was randomly shuffled neuron by neuron, revealed no significant difference between the CR ensembles and Non-CRE (Figs. 4B, and S6A, C), suggesting that the enhancement of the coactivity in the real data did not derive from the enhanced neural activation. Similar results were observed in the CR ensembles excluding the RL-ensemble overlapped neurons (Fig. S6A-C). In addition, changes in the coactivity across the categories (coactivity between CR ensembles and Non-CRE) were significantly smaller than those within the CR ensembles (Fig. S6C). These results led us to hypothesize that the functional connectivity within the CR ensembles was specifically enhanced as a result of the fear conditioning.

To test this hypothesis, we introduced a probabilistic graphical model method, the conditional random field (CRF) model[33, 34], that evaluates the conditional probability that a group of neurons fire together given that one neuron is active (Fig. 4C). Among the various mathematical algorithms used to evaluate possible functional connectivity of neural networks and ensembles, the CRF model is substantially more reliable because the results of the calculation (functional connectivity) have already been carefully evaluated by two-photon holographic optogenetics and consequential behavioral modulation[33, 34]. Using this method, we found that, after the fear conditioning (D4E), the functional connectivity was significantly higher in the CR ensembles (Fig. 4D). This method also allowed us to evaluate the information coding of any arbitrary labels, e.g. CS+, and we found that the CS+ information encoded by the CR ensembles was also significantly higher than that of Non-CRE (Fig. 4E). Importantly, the neurons in the CR ensembles were discriminated by the data and the behavioral labels during the CS+, not by comparison between those during CS+ and those during the presentation of other stimuli, suggesting that our method did not produce any bias to the CS+ in selecting CR ensemble neurons. Therefore, this result indicates that the CR ensembles dominantly conveyed not only the CR information but also the CS+ information. In addition, we found that the enhancement in both the functional connectivity and information coding for CS+ derived in an experience-dependent manner after the fear conditioning, predominantly in the CR ensemble cells (Fig. 4F, G). In contrast, the changes in information coding for the CS– were not significantly different between the CR ensembles and the Non-CRE (Fig. 4G). These results suggest that newly emerged CR ensembles derived as a result of the rewiring of the functional connectivity, perhaps via activity-dependent modulation during the fear conditioning. This led us to search the possible existence of a signature for this plasticity in the neural activity data.

Interestingly, during the fear conditioning, we observed that some of the dmPFC neurons strongly responded to the US (Fig. 5A). Statistical analyses demonstrated that neurons responsive to the US during the fear conditioning were predominantly and significantly more involved in the CR ensemble after the fear conditioning (Fig. 5B, C), suggesting that these US-responsive neurons (USR), or US-evoked inputs to the dmPFC, might modulate network connectivity within the dmPFC network and strengthen the specific connections stemming from the USR during or after the fear conditioning, perhaps leading to the formation of the CS-US association network encoding the CR as a result of the memory consolidation.

Further analyses based on the CRF modeling revealed that the USR actually became functionally more connected within the CR ensemble than non-US responsive neurons, while these differences were not observed in Non-CRE (Fig. 5D). This higher connectivity was a result of the fear conditioning (Fig. 5E). The information coding for the CS+ was also significantly higher in the USR, specifically in the CR ensembles (Fig. 5F). According to a previous study, higher functional connectivity and higher decoding performance of sensory stimuli are typical features of pattern completion cells whose activation could efficiently enhance the entire ensemble activity for a specific sensory stimulus and promote the stimulus-associated behaviors of mice[33]. Therefore, considering our results altogether, it is suggested that the USR were predominantly integrated into the CR ensembles as a result of the fear conditioning, maybe by some activity-dependent modulation like Hebbian plasticity (i.e. fire together, wire together), and the eventual functional connectivity stemming from these USR may have a key role in regulating memory retrieval by enabling the specific association between the US-information network and the CS+ network.

Discussion

How the PFC encodes and regulates variable memories and cognitive functions is a long-standing question. In the present study, we tackled this question using an optical and computational dissection method. A model-based machine-learning algorithm enabled us to untangle the internal prefrontal network and to identify neural ensembles encoding the CR ensemble distinctively from the RL ensemble. Further graphical modeling revealed how those specific circuits were newly constructed, and suggested a possible activity-dependent circuit modulation mechanism associating the US-network with the CS+-network in an experience-dependent manner to encode the CR, a fear memory-guided behavior. This emergence of the novel memory circuit was successfully detected by chronic cellular recording from the same set of hundreds of neurons in each awake mouse during fear conditioning and retrieval/extinction tasks.

More than 60 years ago, Hebb proposed that repeated coactivation of a group of neurons might create a memory trace through the enhancement of synaptic connections[35]. Our findings indicated that neurons strongly responding to the US during the fear conditioning became more dominantly involved in the CR ensembles after the memory consolidation. Also, within the CR ensembles, the USR became more densely connected to the other neurons, as the network hub, and more linked to the CS+. These results suggest that Hebbian plasticity might underlie the rewiring of the prefrontal memory structure, enabling the emergence of a strong link between the US signaling pathway and the CS+ signaling pathway.

CR information was redundantly encoded in the dmPFC. The advantage of the redundancy is not clear, but because fear memory is critical for animal survival, it is possible that the redundant coding for the fear memory is not inefficient at all, but rather evolutionarily crucial. On the other hand, the redundancy can also be considered inefficient in terms of the short-term cost. Because the dmPFC is known to be involved in long-term memory[20, 36], it would be interesting to investigate whether the redundantly encoded information for the CR is maintained or diminishes. Also, the memory is not stored solely in the dmPFC, but brain-wide networks process memory[20, 36, 37]. This redundancy might be related to the brain-wide regulation of memory, which could be studied by labeling the downstream or upstream structures for additional anatomical dissection using virus-based anterograde or retrograde fluorescent labeling techniques simultaneously with GCaMP6f imaging in dmPFC.

As we have successfully dissected the specific neural ensembles encoding the CR as well as more detailed structure of the CR ensemble, testing the causality of the identified structure to behavior by holographic optogenetics[33] could be intriguing. On the other hand, we found that the dmPFC also usually responds to auditory signals (Fig. 2C-E, S1) and encodes RL (Fig. 3). Because enhancing the sensory coding can boost performance in a decision-making task as shown by activation of the primary visual cortex[33], further mathematical dissection and additional anatomical dissection as discussed in the preceding paragraph would be the next step to more precisely find the “memory”-corresponding structure in experiments using holographic optogenetics.

Funding

Supported by the Japan Science and Technology Agency, PRESTO (to M.A.), JSPS

KAKENHI Grant (grant number JP18K06536, JP18H05144, JP20H05076, JP21H02801 to M.A.; JP20H03357, JP20H05073, JP21K18563 to Y.R.T.; JP20H05065 to A.K.), JSPS Bilateral Program (JPJSBP1-20199901 to M.A.), AMED (grant number JP21dm0207086 to M.A.; JP21dm0207017 to H.H.), the grant of Joint Research by the National Institutes of Natural Sciences (NINS program No 01112008 and 01112106 to M.A.), and grants from Brain Science Foundation and Shimadzu Foundation to M.A.,

Acknowledgements

We thank M. Sato, Y. Hayashi, M. Matsuzaki, Y. Masamizu, K. Takao, K. Miyamichi, A. Yamanaka, K. Seiriki, J. Chikazoe, H. Tsukada, D. Peterka, A. Packer, A. Noritake, and T. Imai for technical advice and discussion. We thank the Bionanophotonics Consortium under the MEXT Scientific Research on Innovative Area "Spying Minority in Biological Phenomena" (Grant Number, 23115001) for assistance in microscopy, and the ISIR Machine shop for development of the running disk system. We also thank the members of the laboratory for their help, especially T. Matsuda and T. Wazawa for technical advice, and T. Kobayashi for help with the mice.

Author Contributions

M.A. conceived and coordinated the whole project. M.A. designed and performed behavioral experiments with the support of A.K., H.H., and T.N.; M.A. constructed in vivo imaging system and performed imaging experiments with the support of Y.A. and T.N.; M.A. performed data analyses with the support of I.S, Y.R.T., L.C.-R., M.Y., T.I., and J.N.; M.A. wrote the paper, with contributions from all authors.

Methods

Animals. All animal experiments were carried out in accordance with the Institutional Guidance on Animal Experimentation and with permission from the Animal Experiment Committee of Osaka University (authorization number: 3348), or in accordance with National Institutes of Health guidelines and approved by the National Institute for Physiological Sciences Animal Care and Use Committee (approval number 18A102). Male C57BL/6 or PV-Cre mice (Jax: 008069) mice housed under a 12-h light/dark cycle with free access to food and water were used for all experiments. Behavioral experiments were performed during the dark cycle (i.e. when mice were normally awake) using single-housed mice. Mice at 4–6 months of age were used for the behavioral and imaging experiments.

Virus injection. To express GCaMP6f, a genetically encoded calcium indicator to monitor the neural activity, we used a gene expression system based on the AAV vector. Viruses were injected into mice at postnatal day (P) 50-120 for in vivo experiments, at least 1 month before the microprism implantation, which was followed by the in vivo experiments 1–3 months after the implantation. Injection procedures were performed as described previously[32], with some modifications. During surgery, the mice were anesthetized with isoflurane (initially 2% [partial pressure in air] and then reduced to 1%). A small circle (~1 mm in diameter) of the skull was thinned over the left mPFC using a dental drill to mark the site for a small craniotomy. AAV1/CamKII.GCaMP6f was obtained from the University of Pennsylvania Vector Core, and injected into the left mPFC (slightly away from the imaging target area to avoid damaging the field of view) at three sites (depth 1.0, 1.5, and 2.0 mm from the pial surface, volume 375 nl/site) to cover the dorsal mPFC, over a 5-min period at each depth using a UMP3 microsyringe pump (World Precision Instruments). The X-Y coordinates for the injection site was usually 0.5 mm lateral to the midline and 2.0 mm rostral to bregma, but if large blood vessels obstructed the position, we shifted the insertion site slightly to avoid the vessels. The beveled side of the injection needle was faced to the midline so that the needle could be smoothly inserted and the virus would cover the surface layers of the mPFC. We designed our injection protocol (especially the volume and depth) carefully to widely cover the mPFC areas, while the anatomical coordinates of the field of view for the two-photon imaging were precisely targeted using the position of the pial surface and the sinus, which were usually visible through the imaging window prepared as shown below, as a guide (the field of view ranged from a depth of ~0.9-1.9 mm and centered at a depth of ~1.1-1.5 mm from the pial surface and the sinus).

In vivo two-photon imaging. In vivo two-photon imaging was performed as described previously[27, 32], with modifications to pair with our new experimental system. At 1–3 months after the virus injection, the mice were anesthetized with isoflurane (initially 2% [partial pressure in air] and reduced to 1%). A titanium head plate described in a previous paper by Goldy et al.[38] was selected for the present study to minimize the area laying over the ear and to minimize the blockage of auditory input through the ear. The head plate was attached to the skull with dental cement. For the subsequent microprism implantation, a square cranial window (~2.3 x 2.3 mm) was carefully made with minimal bleeding above the right mPFC, the hemisphere opposite to the virus injection site. An implantable microprism assembly[27], comprising a 2-mm right angle glass microprism (TS N-BK7, 2mm AL+MgF2, Edmund) bonded to a 2x2 mm square cover glass (No.1; Matsunami) for the middle position and a 4x4 or 3x4 mm glass window at the surface position of the imaging window, was prepared and inserted into the subdural space within the fissure along the midline as described previously[27] to avoid harming any nerves surrounding the mPFC network in both hemispheres, allowing for visualization of the left mPFC, which was previously injected with the GCaMP6f virus, through the imaging window. The area directly beneath the microprism was compressed but remained intact. This insertion procedure sometimes caused a small amount of bleeding that covered the imaging site, but even in that case, the imaging window became clear after waiting at least a month before performing the experiments. As reported before[27], the mice recovered quickly and displayed no gross impairments or behavioral differences compared with non-implanted mice, enabling chronic imaging of the dmPFC in behaving mice.

The activity of dorsal mPFC neurons was recorded by imaging fluorescence changes with a FVMPE-RS two-photon microscope (Olympus) and a Mai Tai DeepSee Ti:sapphire laser (Spectra-Physics) at 920 nm, through a 4x dry objective, 0.28 N.A. (Olympus) or a 16x water immersion objective, 0.80 N.A. (Nikon). Mean (±SE) frame rate was 8.96 ± 0.87 (frames/s). GCaMP6f signals were detected via the band-pass emission filter (495-540nm).

As the GCaMP6f was expressed under the regulation of the CaMKII promoter[39, 40], all of the recording targets were assumed to be excitatory neurons[41]. Scanning and image acquisition were controlled by FV30S-SW image acquisition and processing software (Olympus). To smoothly set the mice below the objective lens for the imaging, light and minimal-duration isoflurane (2.0% for less than 2-3 min) anesthesia was used, and behavioral and imaging experiments were started 5 min after the mice awoke and began locomoting on the running disk, which was visually confirmed via the video camera (VLG-02, Baumer) under infrared light-emitting diode illumination (850nm: LDL-130X15IR2-850, CCS Inc.). To detect neural activity from the same set of neurons in each mouse over multiple days, the depth from the surface of the brain (dmPFC area) and configuration of blood vessels and basal GCaMP6f signals in each field of view were recorded and referenced as described previously[42].

Fear conditioning, memory retrieval, and extinction under the microscope.

The experiments were designed according to previous studies, with some modification to optimize conditions for the two-photon microscope system[21, 22, 26]. The heads of the mice were fixed under the objective lens for two-photon imaging, allowing them to run freely on the running disk placed below them, and locomotion and the freezing response were measured by the rotation of the running disk, as previously described[43]. Experiments were performed in a completely dark environment to protect the detector (photo multiplier tube) for the two-photon imaging from the room light. We prepared two different types of running disks to establish two different contexts, as used in conventional fear conditioning experiments for head-unfixed mice[21, 22, 26]. Disk A was made of light-colored plastic with ridges from the center to the rim that the mice could grip to allow them to easily rotate (and walk on) the disk[43]. Disk A was used for habituation (D1 and D2) and for retrieval and extinction (D4). Disk B was built for the fear conditioning (D3), and comprised a grid made of stainless steel bars (Fig. 1A), which was attached to a foot shock generator (SGA-2010, O'HARA & CO., LTD) via an electrical slip ring so that electrical current to this running disk for the foot shock (US) could be stably delivered to the mouse irrespective of whether the running disk was rotating. The behavioral sessions on each day began only after the mouse was constantly locomoting for more than 5 min. The running disks and the surrounding area (inside the cage for the microscope) were cleaned with 70% ethanol before and after each experiment. To score freezing behavior, the speed of the mouse locomotion was measured by the rotation speed of the running disk[43], and mice were considered to be stationary (during no CS presentation) or freezing (during CS+/retrieval) if no movement was detected for at least 1 s. On D1 and D2, the mice underwent an adaptation session with disk A for an hour each day, to familiarize them with the novel environment. On D3, the mice underwent a habituation session in context B, in which they received four presentations of the CS− and CS+ alternately (total CS duration, 30 s for each trial; consisting of 50-ms pips at 1 Hz repeated 30 times; pip frequency, 7.5 kHz or white-noise, respectively, 80-dB sound pressure level (60-dB basal room noise produced by the air conditioning system, and 20-dB for the CS)). The habituation session was immediately followed by discriminative fear conditioning[21, 22, 26] on the same day by pairing the CS+ with a US (1-s foot shock, 7 CS+–US pairings).The intensity of the foot shock was usually 0.05~0.1 mA, but when mice showed no responses at all, which was probably caused by that a part of the running disk became dirty or wet by mice and the foot shock might be suppressed by this during the experiment, an intensity of 0.25~0.45 mA was used. The onset of the US coincided with the onset of the last sound pip of each 30-s CS trial. The CS−and the CS+ trials were performed alternately (inter-trial intervals, 50–150 s). On D4, conditioned mice underwent a retrieval session followed by an extinction session on disk A during which they received 4 presentations of the CS– and 12 presentations of the CS+. During the experiment (D1-4), the mouse was continuously encouraged to locomote by administering a 4-ul drop of saccharin water per 100 cm of locomoting, provided through a spout placed near their mouth[42] so that the freezing response could be discriminably detected as decreased locomotion (Fig. 1). The mice were not water-deprived. The locomotion speed and timings of the tones and the foot shock were synchronously recorded with image acquisition (GCaMP6f imaging in dmPFC) using NI software (Labview; National Instruments) and NI-DAQ (National Instruments). The results shown in Fig.1 show that this protocol led to the mice successfully learning the CS+-US association, and show a reduction in locomotion in response to the CS+, but not the CS–, and not before but only after the fear conditioning session, enabling us to observe changes in neural representations in the dmPFC as a result of the fear conditioning.

Imaging data analyses and statistics. The raw images of the GCaMP6f signals in the dmPFC were processed to correct for brain motion artifacts using the enhanced correlation coefficient image alignment algorithm[44]. To apply the same regions of interest (ROIs) for analyzing the images obtained across multiple days, the movies from the same mouse were precisely aligned with each other using the same enhanced correlation coefficient algorithm as above, while, for a local shift (shift of a few pixels in a small number of neurons among all recorded cells), the corresponding ROIs were manually adjusted.

The ROIs for the detection of neural activity were automatically selected using a constrained nonnegative matrix factorization algorithm in MATLAB as described previously[45], with some manual adjustment. Further steps to process the GCaMP6f signals for measurements of the signal change (ΔF/F) of each neuron were performed as described previously[32, 46]; although the same constrained nonnegative matrix factorization package for ROI detection also provides an option for signal processing that was not sufficiently optimized to analyze our data, which were obtained over several days with more than 30,000 frames each day. Fluctuations in the background fluorescence, which contains synchronous fractions across nearby neurons[45, 46], was subtracted before calculating the ΔF/F of GCaMP6f signals as described previously[32]. Briefly, a ring-shaped “background ROI” was created for each ROI 2–5 pixels away from the border of each neuronal ROI to a width of 30–35 pixels, and the size was adjusted to contain at least 20 pixels in each background ROI after completing the following steps. From the background ROI, we removed the pixels that belonged to any neuronal ROIs, and the ROIs that contained artificially added pixels (black pixels added at the edge of the image due to the motion correction procedure) at any time-point. We then removed the pixels that, at some time-point(s), showed signals exceeding that of the neuronal ROI by two standard deviations of the difference between each background ROI pixel time series and the neuronal ROI time series. The resulting background ROI signals were averaged at each time-point, and a moving average of the time series was calculated. Using the moving average instead of the raw background ROI signal was helpful to minimize the production of an artificially large increase or decrease at each time-point due to the subtraction, which could have altered the analyses of the timing of neural activations. Pixels within each neuronal ROI were also averaged to give a single time course, and then the background ROI signal was subtracted. Then, the ΔF/F of GCaMP6f signals of all neurons in each circuit was calculated. For most of the analyses and comparisons of the results from multiple mice, the ΔF/F data were further z-normalized within each experiment (same mouse, same day) as described previously[21, 26]. On the other hand, particularly for the CRF modeling used to evaluate the functional network connectivity, the spike probabilities were inferred from the ΔF/F as an alternative estimate of neuronal activation using a constrained sparse nonnegative calcium deconvolution method[45]. We used the code "constrained_foopsi.m"[45], and the parameters used in the calculation were not manually selected but estimated from the data by the code. After inference of the spike probability and further thresholding by two standard deviations, the obtained binominal data were further binned (bin size: 1 s). Importantly, the results obtained by CRF modeling were consistent with the results of the coactivity analyses based on the ΔF/F (and z-normalized ΔF/F) (Fig.4), providing substantial support that the analyses based on both estimates complemented each other for the data analyzed in the present study. While neurons for the analyses were initially automatically detected, neurons responding to noisy signals with no apparent calcium transient at any time during the experimental days were identified by visual inspection and excluded from further analysis.

For the statistical analysis, we used MATLAB (MathWorks, Natick, MA). The Wilcoxon signed rank tests for paired comparisons or the Wilcoxon rank sum test (equivalent to Mann-Whitney U test) for unpaired comparisons was used to determine statistical significance ($P < 0.05$) unless otherwise indicated. Two-tailed tests were selected for all statistical analyses. All p-values less than 0.0001 are described as "P<0.0001" (or ****). Graphs were produced by MATLAB (MathWorks) or Excel (Microsoft). When comparing two groups (e.g. D3 vs D4) consisting of the results of multiple mice, in addition to the analyses using original data (e.g. N=7 vs N=7 [D3 vs D4]), we performed bootstrap resampling to more systematically estimate representative values (e.g. mean or median) of each mouse or each group where the number of recorded neurons in each field view varied. When statistically comparing original data (e.g. comparing D3 vs D4), we used a paired permutation test that does not require any assumptions regarding the data distribution, though the p-values obtained by this method and the evaluated statistical significance were very similar to those obtained by the paired t-test in almost all cases. For the analyses based on bootstrap resampling followed by statistical comparison, random resampling (with accepting overlapped sampling) from each mouse was performed in total with the same number as that of the original data of each mouse for each resampling round, and the means (e.g. of 7 mice each day) and the means of the difference or ratio (e.g. difference between D3 vs D4 averaged over mice) were calculated. This was repeated 2000 times to derive the distribution (of 2000 bootstrap replications) for each estimate, and the statistical significance was evaluated based on the 95% confidence interval.

In the present study, to compare changes in neural responses and ensemble representations before and after the fear memory consolidation without any bias, we did not exclude neurons that showed no response to the CS on D4 from the analyses, which was done in some previous experiments (e.g. Herry et al., 2008[26]). Neurons for the analyses were automatically selected based on the neural responses, as described above, and all neurons that exhibited clear activity during at least one of the experimental days were included for the analyses irrespective of whether it was during the CS presentation or only during no CS presentation, considering the previous work suggesting that not only the neurons that typically respond to the CS, but also other types of neurons (including those of mixed selectivity) are important for population coding in the prefrontal network[18].

The significance of CS-induced neural responses was determined according to previous studies[21, 26]. Signals during CS presentation were normalized to baseline activity using a z-score transformation, as described previously[21, 26]. The CS-induced neural activity for each stimulus was then calculated as the mean of the activity during ~1 s from each stimulus onset (depending on the imaging frame rates, we set the number of frames to be used for this calculation so that sampling duration was closer to 1 s but the frames that overlapped with the next stimulus onset was excluded). The last sound pip of each 30-s CS trial was also excluded from this analysis because, during fear conditioning, the last sound pip of the CS+ overlapped with the US (we excluded the last pip data not only for analysis of CS+-evoked responses during fear conditioning but for all data analyses on both D3 and D4, for both CS+ and CS–). They were averaged over blocks of 3 CS trials consisting of 87 individual sound pips in total, for D3E (first three trials during the fear conditioning session), D3L (last three trials during fear conditioning on D3), D4E (first three trials on D4, as responses during fear memory retrieval), and D4L (last three trials only for CS+ on D4 as responses during extinction), respectively, or used to statistically test whether the responses of each neuron were significantly different from zero (baseline) and to define CS-activated / -inactivated neurons.

To define US responsive neurons, because the number of US were limited (7 stimuli in total for each mouse), the mean z-score of each neuron for 1.5 s from the US onset was calculated, and US responsive neurons were defined as neurons with responses of one standard deviation or larger. The number of USR was very limited (zero or only a few for some of the mice), and therefore all the analyses shown in Fig. 5 were performed with pooled data from all mice (N=7 mice).

To evaluate the coactivation of neural activity in the dmPFC network, we calculated cell-to-cell pair-wise correlations within each ensemble using Pearson's correlation coefficient, from the GCaMP6f signals (z-normalized ΔF/F) of two cells over the duration of the CS+ presentation, as described before[32]. The calculated correlation coefficients (R) were statistically analyzed. As a complementary analysis, we also used the inferred spike probability to analyze the functional connectivity, as explained in the section describing the CRF model, which revealed consistent results as shown in the results section. We further performed analyses based on surrogate datasets, as described in previous studies[32, 47]. For this, the total activity of each neuron was preserved, but only the timing was shuffled randomly within each neuron, followed by calculation of the correlation coefficients of shuffled data.

Extraction of neuronal ensembles. To directly differentiate neural populations (ensembles) encoding the CR (i.e. suppressed locomotion triggered by CS+ during the memory retrieval) and those encoding RL (i.e. stationary or locomotive state during no CS presentation), we used the elastic net[31], a regularization and variable selection algorithm that enabled us to systematically extract neurons encoding respective target behaviors. For this, we used the "lassoglm" function of MATLAB R2019b. Because this method allowed us to identify different ensembles for different behaviors independently from the same mice, we used this to verify whether neurons in CR ensembles were unique or mostly overlapped with RL ensembles (Figs. 2F and 3). Compared with the conventional sparse modeling method called Lasso (least absolute shrinkage and selection operator), the advantage of the elastic net is that the hyper parameter "alpha" additively enables the adjustment of the size of selected neurons depending on the data; when the analyzed data include strongly correlated pairs, which appeared to be the case for our data as shown in Figs. 4 and S6, conventional Lasso removes redundant predictors and selects only one or a part of such a synchronous population, but in the elastic net, lowering the alpha value increases their inclusion, which is helpful toward preventing missing encoder neurons.

When extracting the CR ensemble, we used data only during the CS+ presentation of D4E (retrieval session) and identified neurons informative for distinguishing whether animals exhibited freezing behavior or were locomoting during the CS+ so that the auditory information of the CS was not considered for identifying the ensemble neurons. While mice exhibited the CR as suppressed locomotion during the fear memory retrieval session (Fig. 1), they also showed more or less locomotion intermittently, and both labels (freezing and locomotive) are required to perform the regression based on the elastic net (Fig. 2G); only the data containing at least 10% of each label (freezing and locomotive) were used to discriminate ensembles in the present study. On the other hand, for extracting the RL ensemble, we used data only during the no-CS presentation (for D3 and D4).

Learning the elastic net is formulated as follows.

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(-y_i\log\tilde{y}_i-(1-y_i)\log(1-\tilde{y}_i)\right)+\lambda P_\alpha(\beta)\right),$$

where

$$\tilde{y}_i = \frac{1}{1+\exp(-(\beta_0 + x_i^\top \beta))} \; (i = 1, \ldots, N)$$

$$P_\alpha(\beta) = \frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1 = \sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2 + \alpha\,|\beta_j|\right)$$

and N is the number of observations; y_i is the behavior (freezing/stationary y_i=1 or locomotive y_i =0) at observation i; x_i is data (neuronal activity), a vector of p values at observation i; λ is a positive regularization parameter; parameters β_0 and β are a scalar variable and a p-dimensional vector, respectively. As λ increases, the number of nonzero components of β decreases. The elastic net is a hybrid of ridge regression and lasso regularization: when alpha (α) = 1, elastic net is the same as lasso, while, as α shrinks toward 0, elastic net approaches ridge regression. For other values of alpha (α), the penalty term Pα(β) interpolates between the L^1 norm of β and the squared L^2 norm of β. Lasso is sensitive to correlations between variables and can choose one if there are two highly correlated and useful variables, whereas elastic net is more likely to select both useful variables, which leads to more stable variable selection. The tuning parameter λ controls the overall strength of the penalty. β_j is the coefficient for the corresponding neuron j estimated by this model. Because this method is designed to sparsely leave the coefficients β_j for the respective neurons, we could identify neurons with a non-zero coefficient as ones of substantial decodability (i.e. ensemble neurons). The lambda value with minimum expected deviance, as calculated during cross-validation, was selectively used to define these beta coefficients for each dataset. To avoid an imbalance of the number of original labels for respective states (e.g. freezing or locomotive for CR ensembles) for the training, the same number of data points from respective states were randomly selected to prepare the training data despite an overlap, a total of 900 samples for each, and used to produce the model. We found that the eventual model and non-zero-coefficient neurons slightly varied trial by trial. To accurately define each ensemble, we repeatedly performed this procedure (random sampling and modeling) 100 times to obtain the distribution of each beta value. Gaussian fitting was performed to define the centroid and the 95% confidence interval of each distribution of each beta, and then the 95% confidence interval was used to determine whether or not they were significantly different from zero (enabling us to maintain sparsity), with the centroid being used to define the final beta values of non-zero coefficient neurons to build the model. To evaluate the fitting and decoding performance of the obtained model, the prediction accuracy and the area under the curve (AUC) of receiver operating characteristic curve (ROC) were calculated, respectively, revealing that those scores were very similar and highly correlated with each other (Fig. S4).

Based on the above-described procedure, we next optimized the alpha values. Ideally, if all the informative neurons can be extracted into the selected CR ensembles, the remaining neurons should have poor decoding performance. According to this idea, to optimize the alpha value, after building a model at each alpha for each mouse ("AUC original" in Fig. S2A), we compared the difference in decoding performance between "AUC CRE-rem" and "AUC nonCRE-rem" (Figs. 2I and S2A). AUC CRE-rem is the AUC value calculated by an elastic net model built with the neurons, excluding the original CR ensemble neurons. On the other hand, AUC nonCRE-rem is the AUC value calculated by the neurons, excluding neurons other than original CR ensemble neurons, randomly selected, and the number of excluded neurons was the same as the number of original CR ensemble neurons (so that the number of neurons used to calculate AUC nonCRE-rem were set to be the same as that used for AUC CRE-rem calculation). The "AUC difference" (Fig. S2A) between those two values was calculated to estimate the degree of remaining information, and in principle, we defined the best alpha based on the maximum AUC difference for each mouse independently. In addition, for further statistical evaluation to define the optimal alpha as explained below, we repeated these procedures 10 times for both "AUC CRE-rem" and "AUC nonCRE-rem".

As shown in Fig. S2B, although the decoding performance of the original CR ensembles (i.e. AUC original in Fig. S2A) was not affected by the alpha (Fig. S2B, middle), the size of the CR ensemble was affected, and a smaller alpha generally resulted in a larger number of selected neurons for each CR ensemble (Fig. S2B, top), suggesting that the CR information might be redundantly encoded in the dmPFC as discussed in detail later. On the other hand, the influence of the alpha on the AUC difference was more complicated. As explained above, we defined the best alpha based on the maximum AUC difference for each mouse independently, but in some exceptional cases as shown in Fig. S2D (mouse #3), when the other alpha(s) showed a AUC difference(s) not significantly far from the maximum AUC difference, the alpha of the smallest of the ensembles among those alphas, i.e. largest alpha among them, was selected to avoid unnecessarily including additional neurons that did not improve the AUC difference (e.g. in mouse #3, alpha = 0.1, 0.05, 0.01 showed similar AUC differences and there was no statistically significant difference between them [Wilcoxon rank sum test, alpha of maximum AUC difference vs the other alpha, n=10 estimates for each calculated as explained above], so in this case, the largest alpha 0.1 among those three was selected to define the CR ensemble for this mouse).

These results revealed two important points. First, searching around the alpha value may be important in some cases. Considering this, we also searched alphas in the case of RL ensembles (Fig. S3), and found that there was no difference among the various alphas, for the RL ensembles, even if we tested an additional number of reference frames (means of the neural activities over the past or future several frames were used as neural activity data to predict a single label at each single time-point, which showed no significant difference from each other, evaluated by the Friedman test, a non-parametric statistical test similar to the parametric one-way repeated measures ANOVA). Therefore, in the present study, we fixed the alpha to define RL ensembles at 0.75 for most of the analyses, except for the data in Figs. S3 and S4, where we evaluated the influence of the alpha for RL ensembles.

Second, fear memory triggering the CR might be redundantly encoded in the dmPFC. As discussed above, although decoding performance of the original CR ensembles was not affected by the alpha (Fig. S2B, middle), the size of the CR ensemble was affected, and a smaller alpha generally resulted in a larger number of selected neurons for each CR ensemble (Fig. S2B, top). In addition, when the alpha was fixed at alpha (A) =0.9 (a larger alpha (than 0.9) did not work for some circuits in our data), while the uniqueness of the CR ensembles was maintained and the ratio of the CR ensemble neurons overlapping with RL ensembles was 26.84% (Fig. S2E), which was very similar to the case of alpha-optimized CR ensembles (Fig. 3), the size of this CR ensemble (A=0.9) was two times smaller than that of the alpha-optimized CR ensembles (Fig. S2F). Importantly, 97.82% of the neurons selected at A=0.9 were also selected in the alpha-optimized CR ensembles (Fig. S2F), suggesting that the neurons selected at the largest alpha might be more reliable and robust for the decoding among all the informative neurons. In addition, even after the removal of such "core" neurons, the remaining neurons also possessed information for the CR (Fig. S2B, D), indicating that the CR information was redundantly encoded in the dmPFC. Because this redundancy was specific to the CR ensemble and not observed in the RL ensemble, it would be interesting to investigate possible changes in this redundancy when the memory is recalled as a long-term memory (e.g. 30 days after the memory consolidation).

To evaluate the dominance of the CR ensembles vs the RL ensembles, we applied the CR decoder to predict the RL, and vice versa (Figs. 3 and S4).

CRF models to evaluate functional connectivity. To evaluate the functional connectivity between neurons in the recorded network and the pattern completion capability of each neuron, we used conditional random fields (CRFs) as described previously[33], which models the conditional probability distribution of a given neuronal ensemble firing together. We used CRFs to capture the contribution of specific neurons to the overall network activity defined by population vectors belonging to a given neuronal ensemble. We generated a graphical model in which each node represents a neuron in a given ensemble and edges represent the dependencies between neurons. For training, 80% of the recorded data randomly selected from all time frames was used, and for cross-validation, the remaining 20% was used. For this analysis, binned neural activity data (1 s) were used. The model parameters were determined by the local maximum of the likelihood function in the parameter space. We constructed a CRF model in two steps: (1) structure learning, and (2) parameter learning. For the structure learning, we generated a graph structure using $\ell 1$-regularized neighborhood-based logistic regression[34]. Here λs is a regularization parameter that controls the sparsity (or conversely, the density) of the constructed graph structure, leaving only relevant functional connectivity, including both coactive and suppressive relationships. A previous study showed that this number of connections was enhanced as a result of optogenetic rewiring of the local network[34], demonstrating the reliability of the functional connectivity estimated by CRFs models. Therefore, we also calculated the ratio of these remaining connections per all the possible connections for each neuron as a "functional connectivity" score for each node, after carefully screening the optimal λs value by maximizing the log-likelihood of the observations at the following parameter learning step. When comparing the connectivity between different ensembles (e.g. within-CR-ensemble vs within-Non-CRE) or different cell types (e.g. USR vs non-US responsive neurons), we first calculated a whole network connectivity without separating the ensembles, and further separated them into different categories. To measure which neurons were the most informative for a given stimulus (CS+ or CS–), we computed the standard ROC, taking as ground truth the timing of a particular CS. The AUC from the ROC curve that represents the performance of each neuron was calculated to compare the encoded information in different ensembles, different neuron types, and different days (e.g. before vs after the fear memory consolidation). As was recently demonstrated [33], high ranks for this value indicate high potential to recall the neural and cognitive representation of a given stimulus.

1. Jezzini, A., Bromberg-Martin, E.S., Trambaiolli, L.R., Haber, S.N. & Monosov, I.E. A prefrontal network integrates preferences for advance information about uncertain rewards and punishments. *Neuron* **109**, 2339-2352 e2335 (2021).
2. Spellman, T., Svei, M., Kaminsky, J., Manzano-Nieves, G. & Liston, C. Prefrontal deep projection neurons enable cognitive flexibility via persistent feedback monitoring. *Cell* **184**, 2750-2766 e2717 (2021).
3. Reinert, S., Hubener, M., Bonhoeffer, T. & Goltstein, P.M. Mouse prefrontal cortex represents learned rules for categorization. *Nature* **593**, 411-417 (2021).
4. Miller, E.K. & Cohen, J.D. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* **24**, 167-202 (2001).
5. Burgos-Robles, A.*, et al.* Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. *Nature neuroscience* **20**, 824-835 (2017).
6. Calhoon, G.G. & Tye, K.M. Resolving the neural circuits of anxiety. *Nature neuroscience* **18**, 1394-1404 (2015).
7. Le Merre, P., Ahrlund-Richter, S. & Carlen, M. The mouse prefrontal cortex: Unity in diversity. *Neuron* **109**, 1925-1944 (2021).
8. Shin, L.M. & Liberzon, I. The Neurocircuitry of Fear , Stress , and Anxiety Disorders. *Neuropsychopharmacology* **35**, 169-191 (2009).
9. Sakurai, T.*, et al.* Converging models of schizophrenia--Network alterations of prefrontal cortex underlying cognitive impairments. *Prog Neurobiol* **134**, 178-201 (2015).
10. Uylings, H.B. & van Eden, C.G. Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Prog Brain Res* **85**, 31-62 (1990).

11. Laubach, M., Amarante, L.M., Swanson, K. & White, S.R. What, If Anything, Is Rodent Prefrontal Cortex? *eNeuro* **5** (2018).

12. Kesner, R.P. Subregional analysis of mnemonic functions of the prefrontal cortex in the rat. *Psychobiology* **28**, 219-228 (2000).

13. Fenster, R.J., Lebois, L.A.M., Ressler, K.J. & Suh, J. Brain circuit dysfunction in post-traumatic stress disorder: from mouse to man. *Nature reviews. Neuroscience* **19**, 535-551 (2018).

14. MacLean, J.N., Fenstermaker, V., Watson, B.O. & Yuste, R. A visual thalamocortical slice. *Nature methods* **3**, 129-134 (2006).

15. Curley, A.A. & Lewis, D.A. Cortical basket cell dysfunction in schizophrenia. *J Physiol* **590**, 715-724 (2012).

16. Sigurdsson, T., Stark, K.L., Karayiorgou, M., Gogos, J.a. & Gordon, J.a. Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature* **464**, 763-767 (2010).

17. Hashimoto, T.*, et al.* Alterations in GABA-related transcriptome in the dorsolateral prefrontal cortex of subjects with schizophrenia. *Molecular psychiatry* **13**, 147-161 (2008).

18. Rigotti, M.*, et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585-590 (2013).

19. Corcoran, K.A. & Quirk, G.J. Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 840-844 (2007).

20. Do-Monte, F.H., Quinones-Laracuente, K. & Quirk, G.J. A temporal shift in the circuits mediating retrieval of fear memory. *Nature* **519**, 460-463 (2015).

21. Courtin, J.*, et al.* Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature* **505**, 92-96 (2014).

22. Dejean, C.*, et al.* Prefrontal neuronal assemblies temporally control fear behaviour. *Nature* **535**, 420-424 (2016).

23. Klavir, O., Prigge, M., Sarel, A., Paz, R. & Yizhar, O. Manipulating fear associations via optogenetic modulation of amygdala inputs to prefrontal cortex. *Nature neuroscience* **20**, 836-844 (2017).

24. Jercog, D.*, et al.* Dynamical prefrontal population coding during defensive behaviours. *Nature* **595**, 690-694 (2021).

25. Burgos-Robles, A., Vidal-Gonzalez, I. & Quirk, G.J. Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **29**, 8474-8482 (2009).

26. Herry, C.*, et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454**, 600-606 (2008).

27. Low, R.J., Gu, Y. & Tank, D.W. Cellular resolution optical access to brain regions in fissures: imaging medial prefrontal cortex and grid cells in entorhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 18739-18744 (2014).

28. Fusi, S., Miller, E.K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Current opinion in neurobiology* **37**, 66-74 (2016).

29. Rozeske, R.R.*, et al.* Prefrontal-Periaqueductal Gray-Projecting Neurons Mediate Context Fear Discrimination. *Neuron* **97**, 898-910 e896 (2018).

30. Ghandour, K.*, et al.* Orchestrated ensemble activities constitute a hippocampal memory engram. *Nat Commun* **10**, 2637 (2019).

31. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320 (2005).

32. Agetsuma, M., Hamm, J.P., Tao, K., Fujisawa, S. & Yuste, R. Parvalbumin-Positive Interneurons Regulate Neuronal Ensembles in Visual Cortex. *Cerebral cortex* **28**, 1831-1845 (2018).

33. Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A. & Yuste, R. Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. *Cell* **178**, 447-457 e445 (2019).

34. Carrillo-Reid, L.*, et al.* Identification of Pattern Completion Neurons in Neuronal Ensembles using Probabilistic Graphical Models. *The Journal of Neuroscience* (2021).

35. Hebb, D.O. *The Organization of Behavior: A Neuropsychological Theory* (Wiley, 1949).

36. Kitamura, T.*, et al.* Engrams and circuits crucial for systems consolidation of a memory. *Science* **356**, 73-78 (2017).

37. Tovote, P., Fadok, J.P. & Luthi, A. Neuronal circuits for fear and anxiety. *Nature reviews. Neuroscience* **16**, 317-331 (2015).

38. Goldey, G.J.*, et al.* Removable cranial windows for long-term imaging in awake mice. *Nat Protoc* **9**, 2515-2538 (2014).

39. Resendez, S.L.*, et al.* Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. *Nat Protoc* **11**, 566-597 (2016).

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

41. Lee, S.-H.*, et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379-383 (2012).

42. Masamizu, Y.*, et al.* Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task. *Nature neuroscience* **17**, 987-994 (2014).

43. Inagaki, S.*, et al.* Imaging local brain activity of multiple freely moving mice sharing the same environment. *Sci Rep* **9**, 7460 (2019).

44. Evangelidis, G.D. & Psarakis, E.Z. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE transactions on pattern analysis and machine intelligence* **30**, 1858-1865 (2008).

45. Pnevmatikakis, E.A.*, et al.* Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* **89**, 285-299 (2016).

46. Peters, A.J., Chen, S.X. & Komiyama, T. Emergence of reproducible spatiotemporal activity during motor learning. *Nature* **510**, 263-267 (2014).

47. Miller, J.-e.K., Ayzenshtat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4053-4061 (2014).

Figure Legends

Fig. 1. Cued fear conditioning during two-photon microscopy. (A) Schematic diagram showing the system used to perform the cued-fear conditioning and memory retrieval under a two-photon microscope. (B) (top) Experimental protocol. CS, conditioned stimulus; US, unconditioned stimulus; FC, fear conditioning. (bottom) An example of the changes in locomotion over time of a mouse on day (D) 4 (first four trials). (C-E) Fear conditioning under the microscope produced CS+-specific memory consolidation. Comparisons of the locomotor speed between before the tone onset and during the tone presentation are shown in (C-D). Before the fear conditioning (on D3), the mice (N = 23) exhibited no significant change in locomotion during the CS+ and CS- presentations (C, left, and D). After the fear conditioning (i.e. during fear retrieval; the first four trials on D4), however, the CS+ suppressed locomotion as a CR, while the CS- induced no significant change (C, right, and D). After repeated presentations of the CS+ (fear extinction; 5th-12th trials on D4), the CS+-evoked CR became smaller until no significant change in locomotion was observed upon CS+ presentation (D). (E) Statistical comparison among responses to the CS- and those to the CS+ at each testing phase on D4 during the tone presentation revealed that locomotion during CS+ was significantly lower only during trials 1–4 on D4, and not after repeated presentations to the CS+ (5th-12th trials). Note that locomotion during pre-tone-onset (before) was not significantly different between the CS- and CS+ conditions. *$p<0.05$; **$p<0.01$; n.s., not significant by Wilcoxon signed-rank test (the Friedman test followed by post-hoc multiple comparisons revealed similar results for panel E). Error bars, s.e.m.

Fig. 2. Extraction of neural ensembles encoding conditioned responses. (A) Microprism implantation along the midline for optical access to the mPFC without cutting nerves. (B) In vivo two-photon microscopy to detect single-cell neural activity visualized by GCaMP6f, chronically (day [D] 3 and D4) from the same set of neurons observed through the prism. Scale bar, 250 μm. (C) Summary of neural responses during the retrieval session (D4-early [D4E], mean of three trials) to the CS+ or CS-. Mean of neural responses in each category (significantly activated [bright red or blue], inactivated [dark red or blue], and others [dark gray]), as well as the mean of all cells (light gray) are plotted. (D) Scatter plot showing responses of individual neurons to the CS+ and CS- in an example mouse during D4E. Each dot represents the mean response of each neuron. Blue, red, and green colors indicate that cells had a significant response as described in the panel. These features for all the mice are summarized in panel E. (E) Summary of response profiles at each phase (D3E, D3-late [D3L], and D4E, respectively; N=7 chronically recorded mice). (F) Mice exhibited low locomotion rates during the CS+ as a CR, or during the inter-trial interval as a regular stationary state. To investigate the specificity of the neural ensembles encoding the CR, we also independently determined the ensembles encoding regular locomotion (RL) for the comparison. (G) Schematic diagram showing how we extracted the CR ensembles. See the Methods for details. (H) An example of the CR ensemble and encoded neural representation

of the behavior. (top) Extracted neurons are drawn with a bold margin, and the mean activity during CR (freezing) is shown in color. (bottom) Time course changes of neural representation encoded by the CR ensemble (same ensemble shown in the top panel). Black dots on the top of the graph and pink color in the graph indicate the timing of the actual CR, while the blue line shows information (i.e. “y” in panel G) decoded by this CR ensemble. The plots show a part of the whole length of the data, and overall decoding accuracy was 97.36% in this example. TP, time points (i.e. image frames). (I) We optimized the parameter “alpha” by calculating the decoding performance of the remaining neurons after removing the CR ensemble extracted by each alpha (CRE-removed), and comparing its decoding performance with the control (Non-CRE removed: decoding performance after non-CR-ensemble [Non-CRE] neurons were removed). An example of this comparison is shown in J, and more details are shown in Fig. S2. (J) Comparison of the decoding performance between CRE-removed and Non-CRE removed, revealing the poor neural information in the Non-CRE removed. This is the result from the same mouse shown in panel H. (K) Schematic diagram showing how we extracted the RL ensembles.

Fig. 3. Emergence of unique CR ensembles after fear conditioning. (**A**) An example Venn diagram and an example spatial map showing the overlap between CR ensemble neurons and RL ensemble neurons in an example mouse. (**B**) Summary of the overlap between the CR ensemble neurons and RL ensemble neurons of all mice (N=7, n=1165 neurons). (**C-E**) Decoding locomotion during RL (inter-trial interval) or CS+ by RL ensembles. (**C**) In an example mouse, an RL ensemble (RLE) that showed high accuracy for decoding performance to predict RL (top) also showed high decoding performance in predicting locomotion during CS+ at day 3-early (D3E), but the performance dropped when it was applied to the prediction of locomotion during CS+ at D4E. (**D**) Original decoding performance of the RL ensembles (i.e. predictability for RL) were not significantly different between D3 and D4. (**E**) (left) Decoding performance of RL ensembles to locomotion during CS+ at D4E (i.e. during fear retrieval) was significantly lower than that for D3E (i.e. early fear conditioning phase). (right) The change in decoding performance was statistically evaluated. Decoding performance was not significantly different between D3E and D3-late (D3L), or between D3E and D4L. (F) Decoding locomotion during CS+ by CR ensembles. Decoding performance was significantly decreased when the CR ensembles were applied to predict RL. Within D3, N=10; D3 vs D4 and within D4, N=7 pairs. A non-paired comparison (Wilcoxon rank sum test) was performed for panel D, while for the other comparisons in E and F, a paired permutation test was performed. For the decoding performance, we plotted the accuracy scores, while the AUC was very similar as shown in Fig. S4. **p<0.01; n.s., not significant. Red bars, median; box in panel E (left) indicates 25th and 75th percentiles.

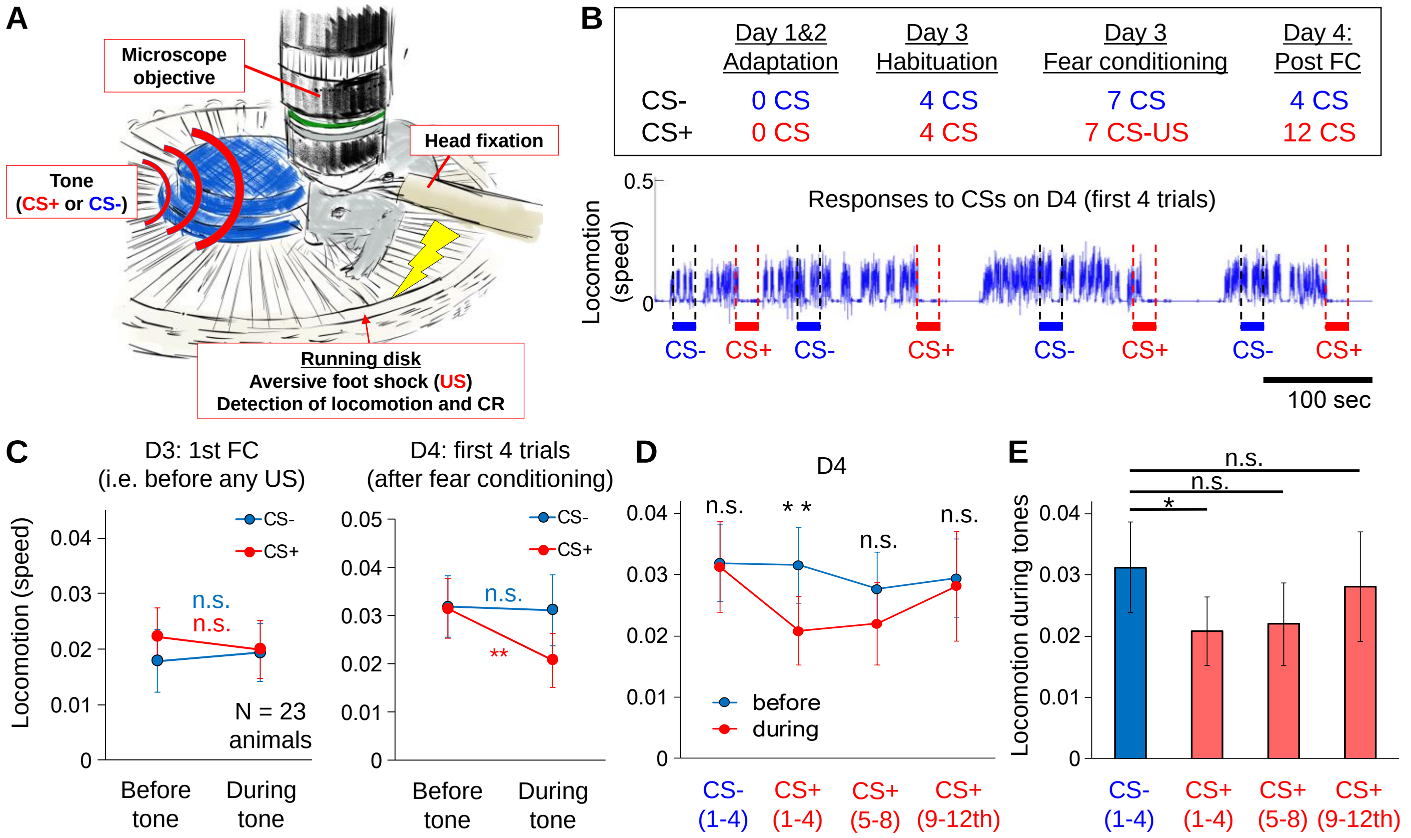
Fig. 4. Enhanced coactivity, functional connectivity, and CS+ encoding in CR ensemble neurons after fear conditioning. (A) Summary of changes in the correlation coefficients (R) of all chronically observed networks (N=7 mice). R differences (day 4-early [D4E] minus

D3E) are plotted as a result of bootstrap resampling (2000 times) performed to systematically compare all the raw R data of whole mice, for CR ensemble neurons (CRE) or for those other than CR ensemble neurons (Non-CRE). Because the change in R between D3E and D4E was observed only for positive correlations specifically in CRE as shown in Fig. S6, we statistically tested changes in the mean, and the 85th and 50th percentiles (as well as the 90th and 80th percentiles in Fig. S6), and the ratio of pairs with a significantly high correlation (RSHC). The asterisk or n.s. beside each column indicates the result of the statistical comparison between D3E and D4E, whereas an asterisk or n.s. at the top of each panel shows the result of the comparison between CRE and Non-CRE. (B) Statistical comparisons similar to panel A, but using shuffled data. (C) Functional connectivity between neurons in an example circuit. Among all the possible connections for all pairs of neurons, the CRF model enables the estimation of functional connections, as well as the dependencies of connected pairs. In this panel, the top 50% edge potentials were visualized. (D) During D4E, the functional connectivity within CRE was significantly higher than that of Non-CRE. (E) During D4E, the mean of cellular predictability for CS+ in CRE was also significantly higher than that in Non-CRE. (F) Change in functional connectivity within CRE of an example circuit. This is the same as the circuit shown in C, but only the connectivity of the CRE neurons marked by the red ellipses were analyzed. Left panel shows the change in the connectivity between D3E and D4E, while the right panel shows the change in the ratio of functional connectivity per all possible connections for individual nodes (i.e. individual neurons). (G) Summary of changes in functional connectivity and cellular decoding performance for CS+ and CS- of all observed networks (N=7 mice). As in panels A and B, differences (D4E minus D3E) of these scores are plotted as a result of bootstrap resampling (2000 times) to compare CRE and Non-CRE, or CRE-noRLE (CRE neurons excluding those overlapping with RL ensemble neurons) and Non-CRE. A paired permutation test was used for the statistics in D and E. The Wilcoxon signed-rank test was used for the statistics in F. The data obtained by bootstrap resampling were statistically analyzed as described in the Methods. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; n.s., not significant. Red bars, median; gray boxes in panels A, B, F, G indicate 25th and 75th percentiles.

Fig. 5. US responsive neurons were, by definition, pattern completion cells in the CR ensemble networks. (A) A part of the recorded neurons in the dmPFC showed increased activity upon US presentation on day 3 (D3) during fear conditioning. Mean activity over 7 trials of all (top) or US-responsive (middle) neurons, and the mean ± s.e.m. of respective categories (bottom) are plotted. Green dotted line indicates the onset of the US, and yellow bar indicates the 1-s duration of the US presentation. (B) Summary of US responses of CR ensemble neurons (CRE) and others (Non-CRE). All individual neurons for the respective categories are plotted. (C) Neurons responding to the US on D3 were dominantly involved in the CRE on D4 after the fear conditioning. The difference between CRE vs Non-CRE, as well as CRE-noRLE vs Non-CRE, was statistically evaluated. (D) Comparison of functional connectivity between US responsive neurons (USR) and others (nonUSR). In the CRE network, USR became more connected within the network than nonUSR, while there was no

significant difference between USR and nonUSR outside of the CRE (Non-CRE). (E) The higher connectivity of USR on D4 was experience-dependent. Functional connectivity of USR on D4 was significantly higher in CRE, while there was no significant difference between them in Non-CRE. (F) USR in the CRE exhibited significantly higher decoding performance of CS+ than nonUSR, which was not the case in Non-CRE. Because the number of USR was limited (only 5.63% under the present definition), the analyses shown in this figure were performed with data pooled together from all mice (N=7 mice). Fisher's exact test was used for the statistics in C, a non-paired comparison (Wilcoxon rank sum test) was used in D and F, and the Wilcoxon signed-rank test was used in E. *p<0.05; **p<0.01; n.s., not significant. Red bars, median; gray boxes in panels D-F indicate 25 and 75 percentiles.

Fig. 1



	Day 1&2 Adaptation	Day 3 Habituation	Day 3 Fear conditioning	Day 4: Post FC
CS-	0 CS	4 CS	7 CS	4 CS
CS+	0 CS	4 CS	7 CS-US	12 CS

Fig. 2

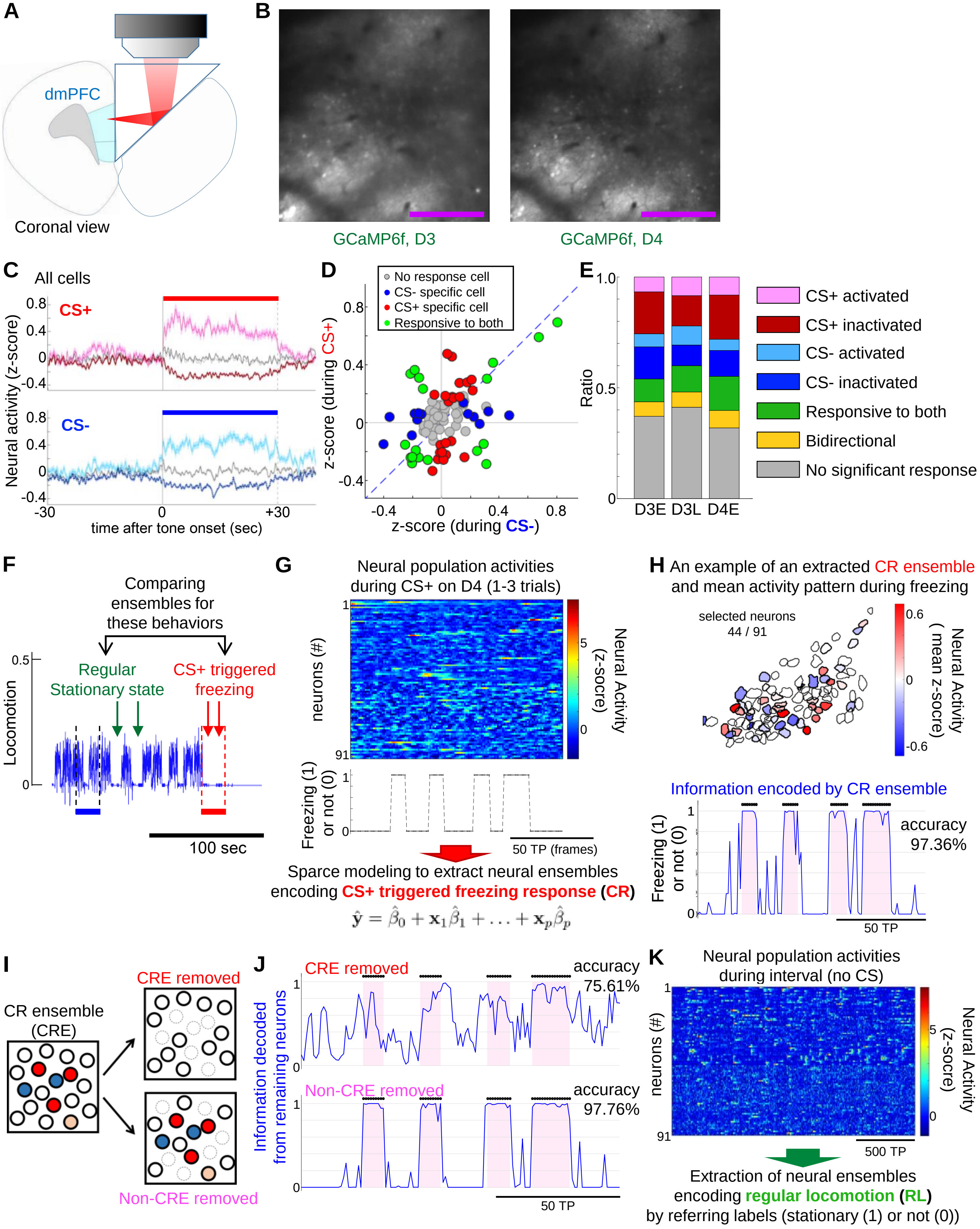

A
dmPFC
Coronal view
B
GCaMP6f, D3
GCaMP6f, D4
C
All cells
CS+
CS-
Neural activity (z-score)
0.8
0.4
0
-0.4
-30
0
+30
time after tone onset (sec)
D
No response cell
CS- specific cell
CS+ specific cell
Responsive to both
z-score (during CS+)
z-score (during CS-)
-0.4
0
0.4
0.8
E
Ratio
1.0
0.5
0
D3E D3L D4E
CS+ activated
CS+ inactivated
CS- activated
CS- inactivated
Responsive to both
Bidirectional
No significant response
F
Comparing ensembles for these behaviors
Regular Stationary state
CS+ triggered freezing
Locomotion
0.5
0
100 sec
G
Neural population activities during CS+ on D4 (1-3 trials)
neurons (#)
1
91
Neural Activity (z-socre)
5
0
Freezing (1) or not (0)
1
0
50 TP (frames)
Sparce modeling to extract neural ensembles encoding CS+ triggered freezing response (CR)
$\hat{\mathbf{y}} = \hat{\beta}_0 + \mathbf{x}_1\hat{\beta}_1 + \ldots + \mathbf{x}_p\hat{\beta}_p$
H
An example of an extracted CR ensemble and mean activity pattern during freezing
selected neurons 44 / 91
0.6
0
-0.6
Neural Activity (mean z-socre)
Information encoded by CR ensemble
Freezing (1) or not (0)
1
0
accuracy 97.36%
50 TP
I
CR ensemble (CRE)
CRE removed
Non-CRE removed
J
Information decoded from remaining neurons
CRE removed
accuracy 75.61%
Non-CRE removed
accuracy 97.76%
1
0
50 TP
K
Neural population activities during interval (no CS)
neurons (#)
1
91
Neural Activity (z-socre)
5
0
500 TP
Extraction of neural ensembles encoding regular locomotion (RL) by referring labels (stationary (1) or not (0))

Fig. 3

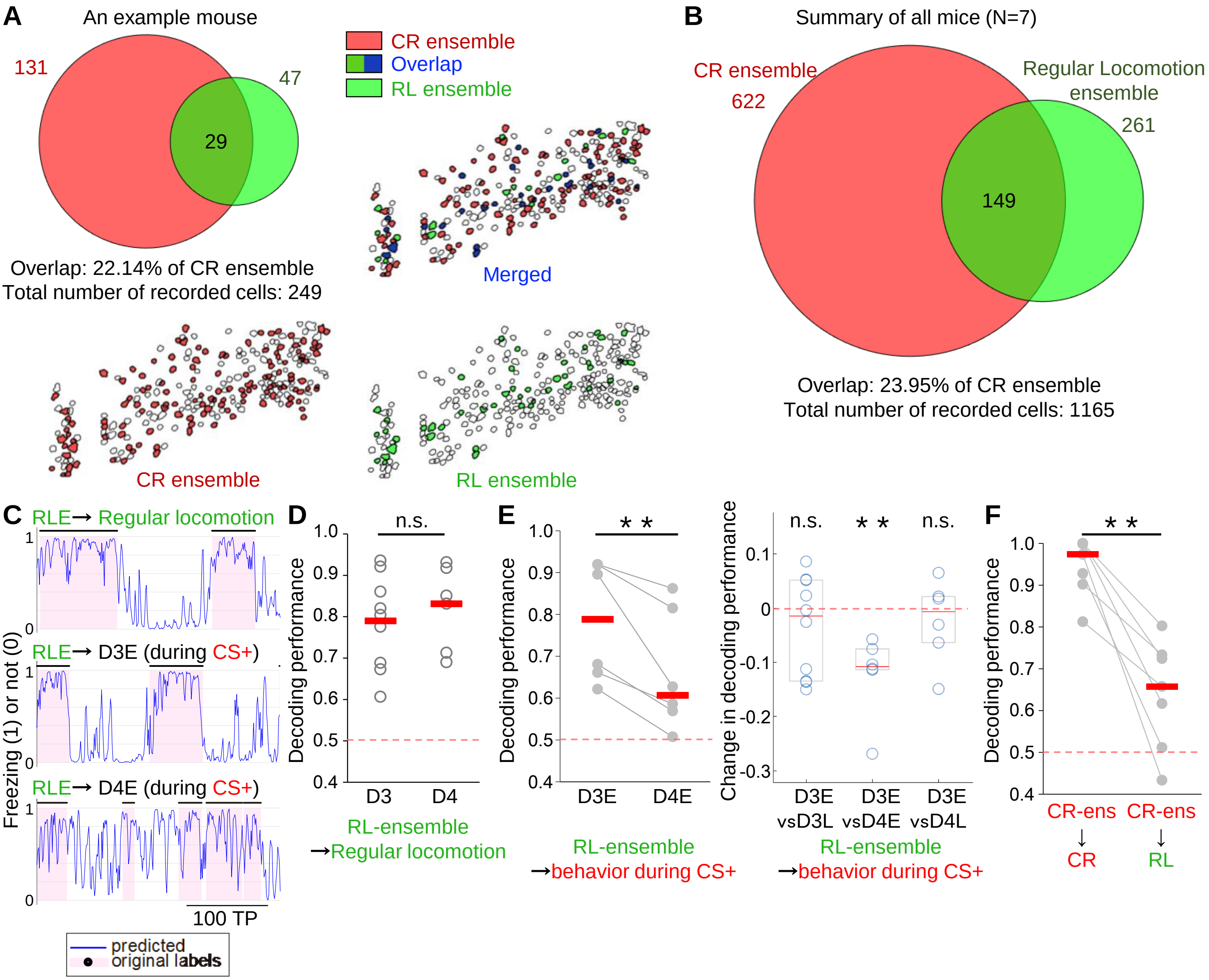

A
An example mouse
131
47
29
CR ensemble
Overlap
RL ensemble
Overlap: 22.14% of CR ensemble
Total number of recorded cells: 249
Merged
CR ensemble
RL ensemble
B
Summary of all mice (N=7)
CR ensemble
622
Regular Locomotion ensemble
261
149
Overlap: 23.95% of CR ensemble
Total number of recorded cells: 1165
C
RLE→ Regular locomotion
RLE→ D3E (during CS+)
RLE→ D4E (during CS+)
Freezing (1) or not (0)
1
0
100 TP
predicted
original labels
D
n.s.
Decoding performance
1.0
0.9
0.8
0.7
0.6
0.5
0.4
D3
D4
RL-ensemble
→Regular locomotion
E
* *
Decoding performance
D3E
D4E
RL-ensemble
→behavior during CS+
Change in decoding performance
0.1
0
-0.1
-0.2
-0.3
n.s.
* *
n.s.
D3E vsD3L
D3E vsD4E
D3E vsD4L
RL-ensemble
→behavior during CS+
F
* *
Decoding performance
CR-ens ↓ CR
CR-ens ↓ RL

Fig. 4

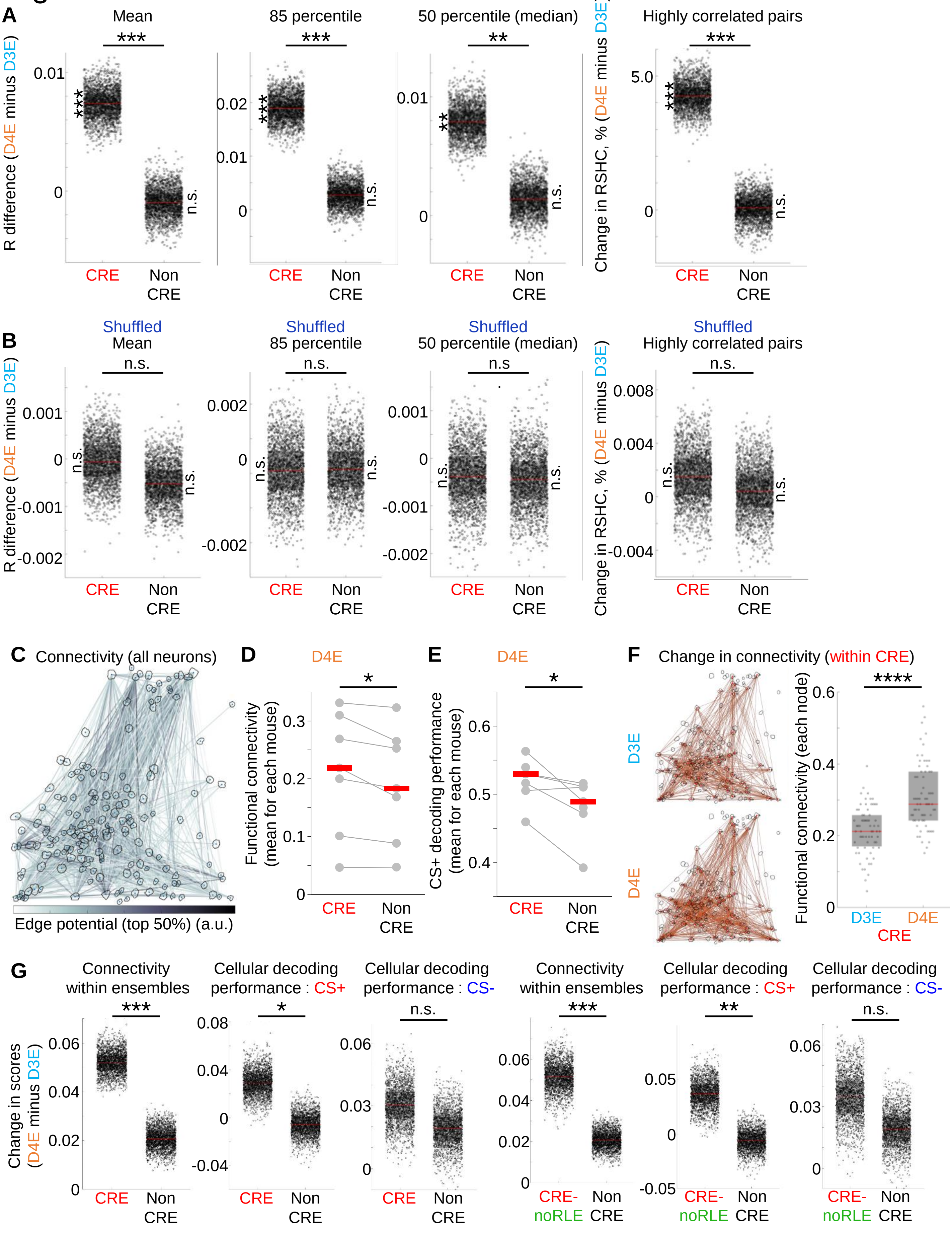

A
Mean
***
R difference (D4E minus D3E)
0.01
0
***
n.s.
CRE
Non CRE
85 percentile
***
0.02
0.01
0
***
n.s.
CRE
Non CRE
50 percentile (median)
**
0.01
0
**
n.s.
CRE
Non CRE
Highly correlated pairs
***
Change in RSHC, % (D4E minus D3E)
5.0
0
***
n.s.
CRE
Non CRE
B
Shuffled
Mean
n.s.
R difference (D4E minus D3E)
0.001
0
-0.001
-0.002
n.s.
n.s.
CRE
Non CRE
Shuffled
85 percentile
n.s.
0.002
0
-0.002
n.s.
n.s.
CRE
Non CRE
Shuffled
50 percentile (median)
n.s
0.001
0
-0.001
-0.002
n.s.
n.s.
CRE
Non CRE
Shuffled
Highly correlated pairs
n.s.
Change in RSHC, % (D4E minus D3E)
0.008
0.004
0
-0.004
n.s.
n.s.
CRE
Non CRE
C
Connectivity (all neurons)
Edge potential (top 50%) (a.u.)
D
D4E
*
Functional connectivity (mean for each mouse)
0.3
0.2
0.1
0
CRE
Non CRE
E
D4E
*
CS+ decoding performance (mean for each mouse)
0.6
0.5
0.4
CRE
Non CRE
F
Change in connectivity (within CRE)
D3E
D4E
****
Functional connectivity (each node)
0.6
0.4
0.2
0
D3E
D4E
CRE
G
Connectivity within ensembles
***
Change in scores (D4E minus D3E)
0.06
0.04
0.02
0
CRE
Non CRE
Cellular decoding performance : CS+
*
0.08
0.04
0
-0.04
CRE
Non CRE
Cellular decoding performance : CS-
n.s.
0.06
0.03
0
CRE
Non CRE
Connectivity within ensembles
***
0.06
0.04
0.02
0
CRE-noRLE
Non CRE
Cellular decoding performance : CS+
**
0.05
0
-0.05
CRE-noRLE
Non CRE
Cellular decoding performance : CS-
n.s.
0.06
0.03
0
CRE-noRLE
Non CRE

Fig. 5

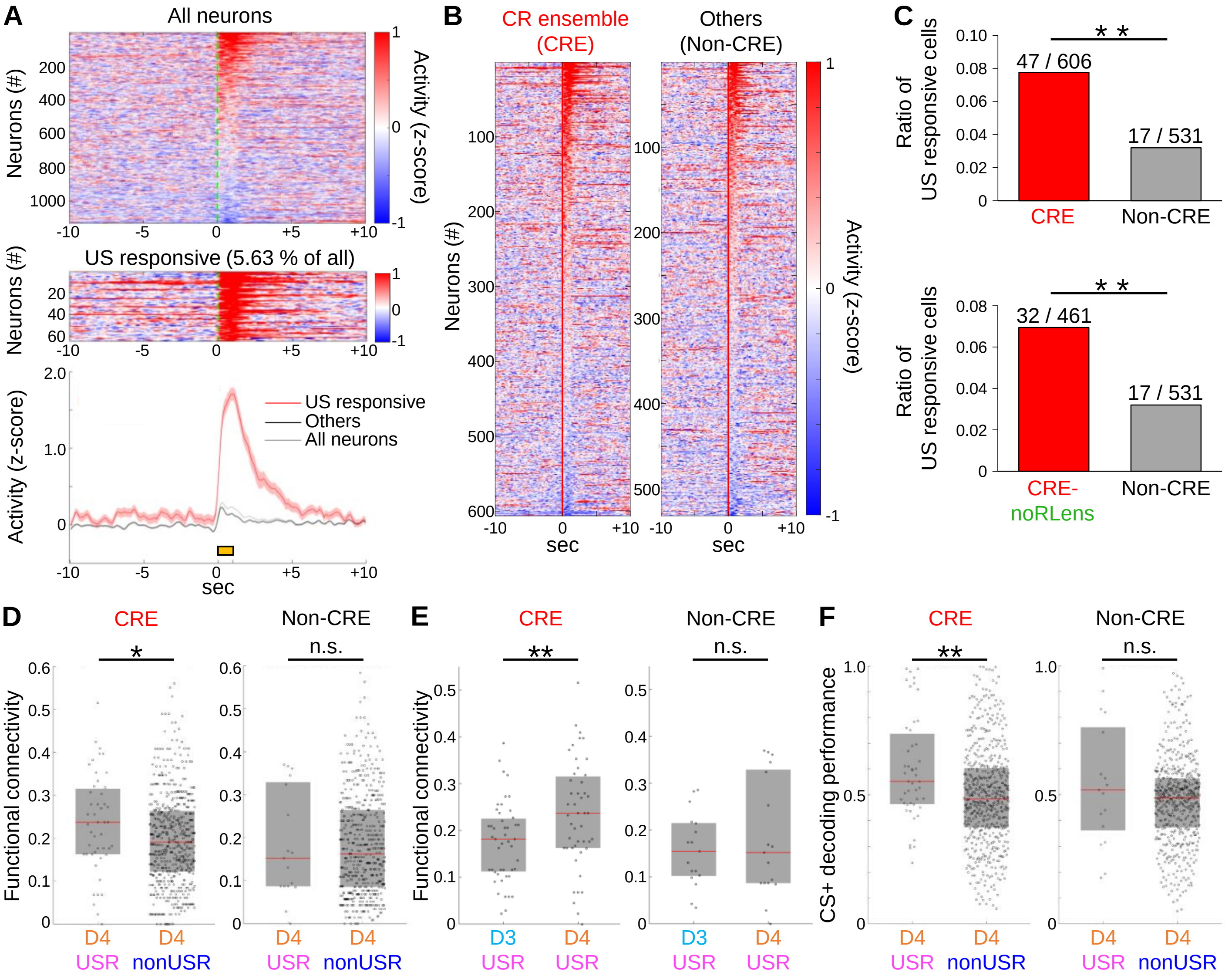

A
All neurons
Neurons (#)
200
400
600
800
1000
-10
-5
0
+5
+10
1
0
-1
Activity (z-score)
US responsive (5.63 % of all)
20
40
60
2.0
1.0
Activity (z-score)
US responsive
Others
All neurons
sec
B
CR ensemble (CRE)
Others (Non-CRE)
Neurons (#)
100
200
300
400
500
600
sec
C
Ratio of US responsive cells
0.10
0.08
0.06
0.04
0.02
* *
47 / 606
17 / 531
CRE
Non-CRE
32 / 461
CRE-noRLens
D
CRE
Non-CRE
*
n.s.
Functional connectivity
0.6
0.5
0.4
0.3
0.2
0.1
D4 USR
D4 nonUSR
E
**
D3 USR
D4 USR
F
CS+ decoding performance
1.0
0.5

Fig. S1

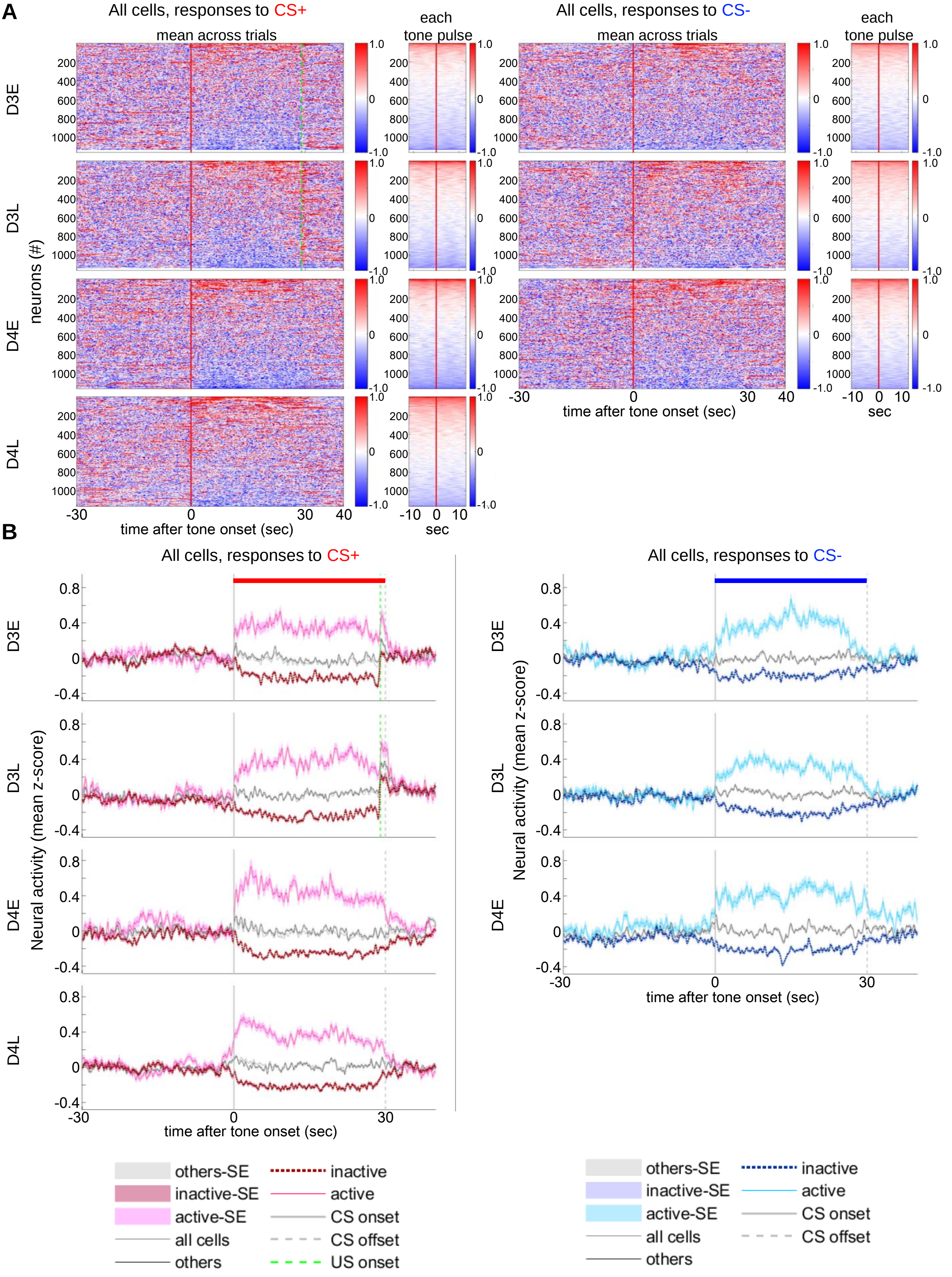

A
All cells, responses to CS+
mean across trials
each tone pulse
All cells, responses to CS-
mean across trials
each tone pulse
D3E
D3L
D4E
D4L
neurons (#)
200
400
600
800
1000
1.0
0
-1.0
-30
0
30
40
time after tone onset (sec)
-10
0
10
sec
B
All cells, responses to CS+
All cells, responses to CS-
Neural activity (mean z-score)
0.8
0.4
0
-0.4
-30
0
30
time after tone onset (sec)
others-SE
inactive-SE
active-SE
all cells
others
inactive
active
CS onset
CS offset
US onset

Fig. S1. Summary of responses to CS+ and CS- of individual neurons on day (D) 3 and D4.

To consider possible temporal changes in responses to the CSs before and after the fear conditioning, only results from mice in which neural activities were successfully recorded on both D3 and D4, from the same sets of neurons, were analyzed. (A) Mean activity over 3 CS trials, or over 87 onsets of 50-ms tone pulses during the 3 trials (D3-early[D3E]/D3-late [D3L] or D4E/D4L, respectively) for all individual neurons is plotted separately (n=1165). (B) Mean (± s.e.m.) CS responses of each category, at each temporal phase (D3E/D3L, D4E/D4L), are plotted separately.

Fig. S2

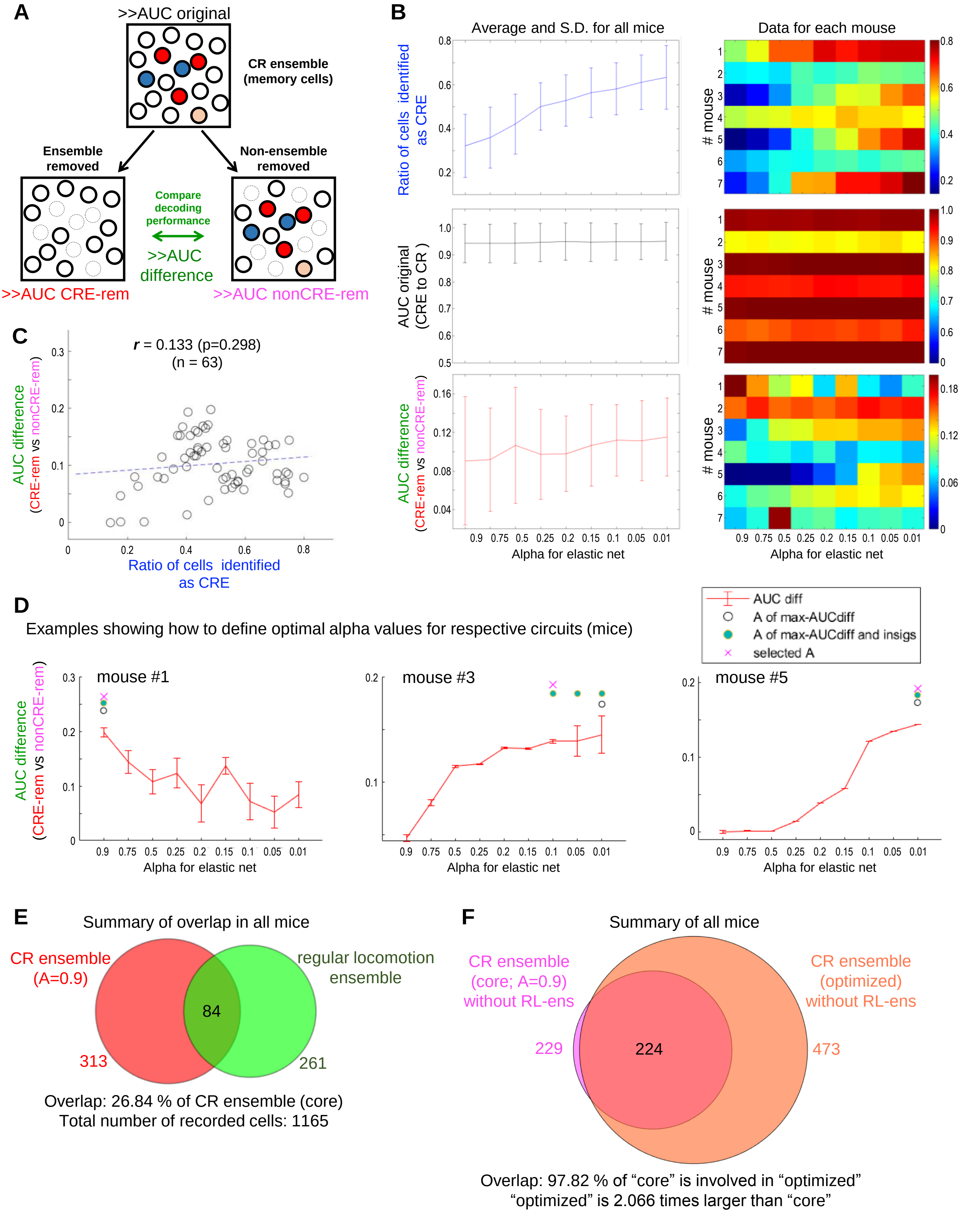

A
>>AUC original
CR ensemble
(memory cells)
Ensemble
removed
Non-ensemble
removed
Compare
decoding
performance
>>AUC
difference
>>AUC CRE-rem
>>AUC nonCRE-rem
B
Average and S.D. for all mice
Data for each mouse
Ratio of cells identified
as CRE
0.8
0.6
0.4
0.2
# mouse
1
2
3
4
5
6
7
AUC original
(CRE to CR)
1.0
0.9
0.8
0.7
0.6
0.5
AUC difference
(CRE-rem vs nonCRE-rem)
0.16
0.12
0.08
0.04
0.18
0.06
0
0.9 0.75 0.5 0.25 0.2 0.15 0.1 0.05 0.01
Alpha for elastic net
C
r = 0.133 (p=0.298)
(n = 63)
AUC difference
(CRE-rem vs nonCRE-rem)
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
Ratio of cells identified
as CRE
D
Examples showing how to define optimal alpha values for respective circuits (mice)
AUC diff
A of max-AUCdiff
A of max-AUCdiff and insigs
selected A
mouse #1
mouse #3
mouse #5
AUC difference
(CRE-rem vs nonCRE-rem)
0.3
0.2
0.1
0
0.9 0.75 0.5 0.25 0.2 0.15 0.1 0.05 0.01
Alpha for elastic net
E
Summary of overlap in all mice
CR ensemble
(A=0.9)
regular locomotion
ensemble
84
313
261
Overlap: 26.84 % of CR ensemble (core)
Total number of recorded cells: 1165
F
Summary of all mice
CR ensemble
(core; A=0.9)
without RL-ens
CR ensemble
(optimized)
without RL-ens
229
224
473
Overlap: 97.82 % of “core” is involved in “optimized”
“optimized” is 2.066 times larger than “core”

Fig. S2. Optimization process of the size of CR ensembles by adjusting alpha values for the elastic net, which also revealed the redundant coding feature.

(A) The AUC of the ROC was calculated to evaluate the decoding performance, and to optimize the alpha value, after building a model at each alpha for each mouse (AUC original), we compared the difference in decoding performances between “AUC CRE-rem” and “AUC nonCRE-rem”. AUC CRE-rem is the AUC value calculated by an elastic net model built with the neurons excluding original CR ensemble neurons. AUC nonCRE-rem is the AUC value calculated by the neurons excluding neurons other than original CR ensemble neurons. The “AUC difference” between those two values was further calculated, and in principle, we defined the best alpha based on the maximum AUC difference for each mouse independently (see more details in the Methods and panel D). (B) Summary and raw data for ratio of neurons identified as CR ensemble (per whole neurons of each circuit) (top), AUC original (middle), and AUC difference (bottom), at each alpha. (C) Pearson’s correlation was used to calculate the r and p values, revealing that the ratio of neurons identified as CR ensembles (per whole neurons) and the AUC difference were not significantly correlated (n=63 samples [7 mice x 9 alphas] were analyzed to determine the possible relationship). (D) Examples showing how to determine the optimal alpha values for respective circuits (mice). In principle, we defined the best alpha based on the maximum AUC difference for each mouse independently, but in some examples as in mouse #3, several alphas revealed statistically insignificant results among the AUC differences. In this case, the largest alpha among those with the same AUC difference was selected. See more details in the Methods. (E) When alpha was fixed at alpha(A)=0.9, the ratio of the CR ensemble neurons that overlapped with RL ensembles was 26.84%, similar to the case of an optimized alpha as shown in Fig. 3. (F) The size of this CR ensemble (A=0.9) was two times smaller than that of the alpha-optimized CR ensembles. 97.82% of the neurons identified at A=0.9 were also selected in the alpha-optimized CR ensembles, suggesting that the neurons selected at the largest alpha 0.9 might be more reliable and robust for the decoding among all the informative neurons in the dmPFC. In addition, even after the removal of such “core” neurons, the remaining neurons also possessed information for the CR (as shown in B and D), indicating that the CR information was redundantly encoded in the dmPFC. Error bars, s.e.m.

Fig. S3

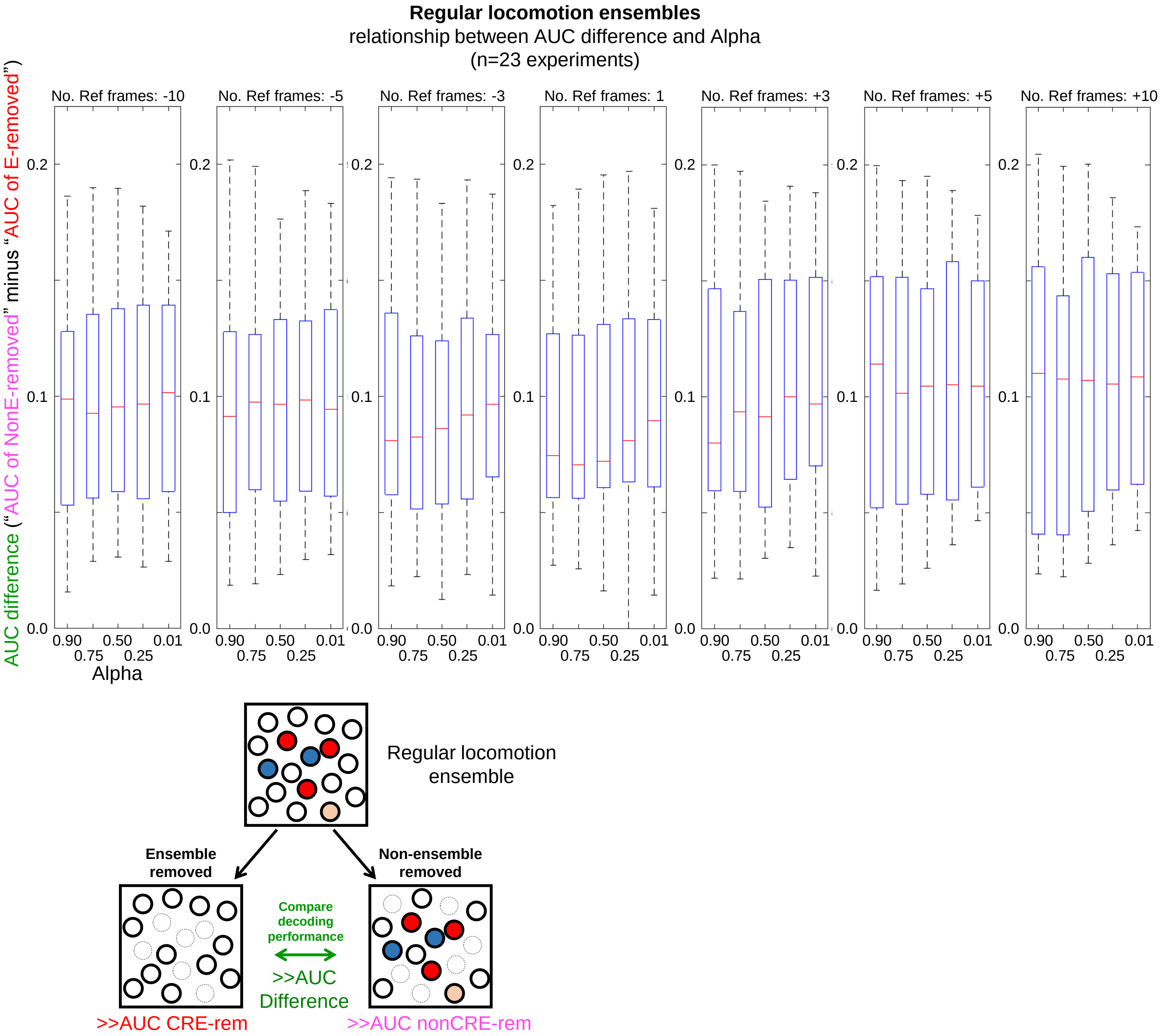
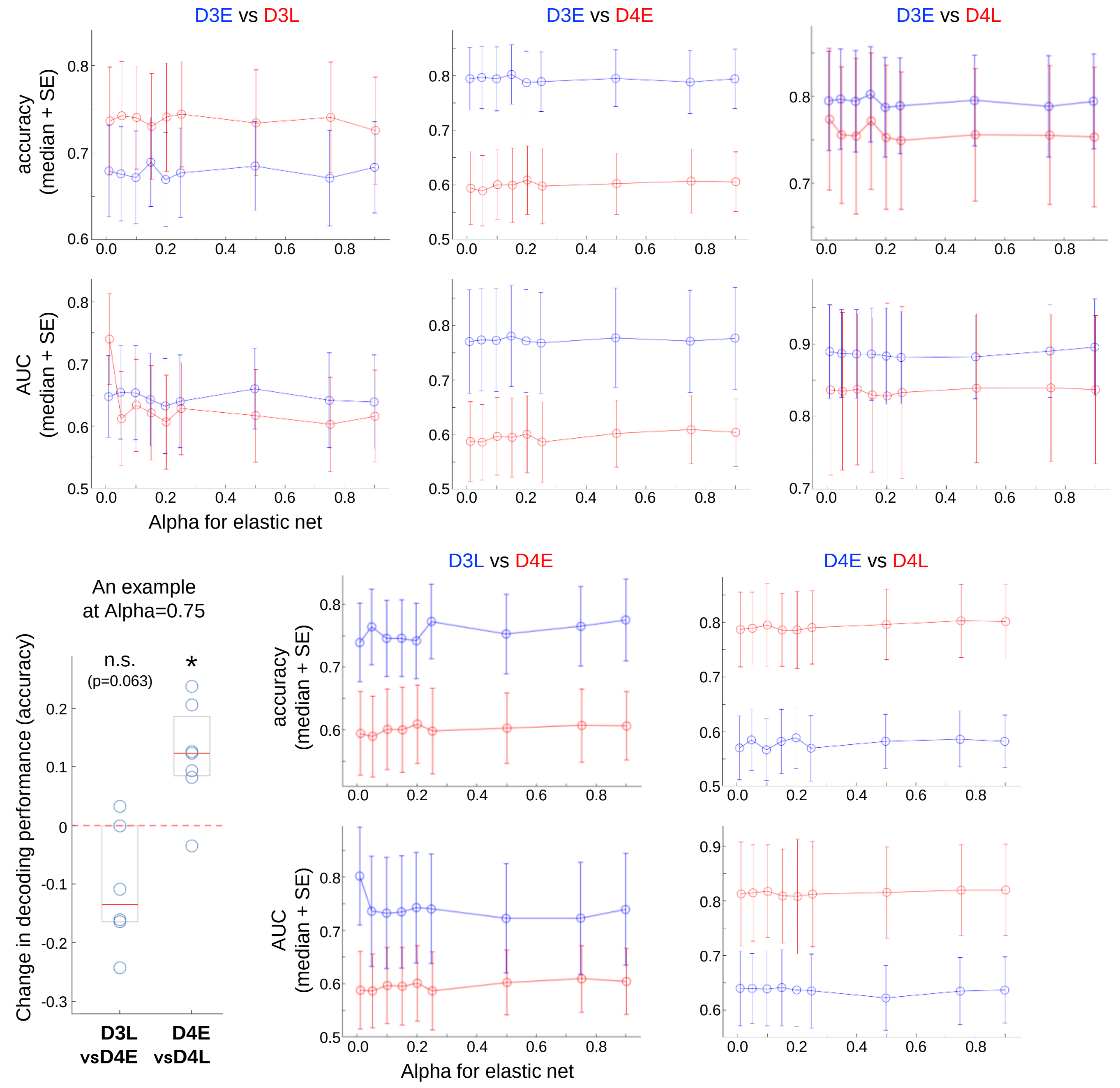


Fig. S3. RL ensembles were not affected by the alpha of the elastic net.

Because optimization of the alpha (hyper parameter for elastic net) was necessary to discriminate CR ensembles (Fig. S2), we also investigated the relationship between the alpha and AUC difference for RL ensembles. In addition to various alpha values, we tested various numbers of reference frames (means of the neural activities over several past or future frames were used as neural activity data to predict a single label at each time-point) to determine the potential difference in the decoding performance. We found no significant differences, however, among the various alphas, or among the different numbers of reference frames, which were evaluated by the Friedman test. Analyses shown in Fig. S4 also showed a similar independency of alpha values in the decoding performance of the RL ensembles. According to these results, we decided to fix the alpha at 0.75 to model RL ensembles, and to fix number of reference frames to one. Red bars, median; the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively; whiskers extend to the most extreme data points not considered outliers (outliers were calculated by the "boxplot" function of MATLAB R2014a).

Fig. S4



p value
at each alpha
(paired permutation test)

Alpha	0.01	0.05	0.10	0.15	0.20	0.25	0.50	0.75	0.90
accuracy, D3EvsD3L	0.107	0.219	0.260	0.113	0.250	0.309	0.215	0.244	0.195
accuracy, D3EvsD4E	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
accuracy, D3EvsD4L	0.531	0.500	0.500	0.531	0.531	0.594	0.469	0.531	0.250
accuracy, D3LvsD4E	0.156	0.094	0.156	0.125	0.219	0.156	0.125	0.063	0.031
accuracy, D4EvsD4L	0.031	0.016	0.078	0.016	0.078	0.031	0.016	0.016	0.031
AUC, D3EvsD3L	0.779	0.273	0.350	0.211	0.354	0.365	0.102	0.322	0.365
AUC, D3EvsD4E	0.219	0.188	0.219	0.219	0.250	0.188	0.188	0.219	0.156
AUC, D3EvsD4L	0.313	0.375	0.688	0.625	0.313	0.375	0.688	0.688	0.563
AUC, D3LvsD4E	0.219	0.563	0.750	0.656	0.750	0.563	0.625	0.563	0.438
AUC, D4EvsD4L	0.094	0.031	0.031	0.031	0.281	0.094	0.031	0.031	0.031

Fig. S4. Summary of state-dependent change in decoding performance of RL ensembles to predict behaviors during the CS+.

To evaluate the decoding performance, we calculated the accuracy and AUC (of the ROC) as described in the Methods. We fixed the alpha for the RL ensembles at 0.75 because there was no difference among the various alphas in any estimates, as shown here and in Fig. S3. For the selected results of alpha=0.75, the data of individual circuits are also shown in the left middle panel, as in Fig. 3E. P values at each alpha (calculated by paired permutation test) are also summarized in the table.

Fig. S5

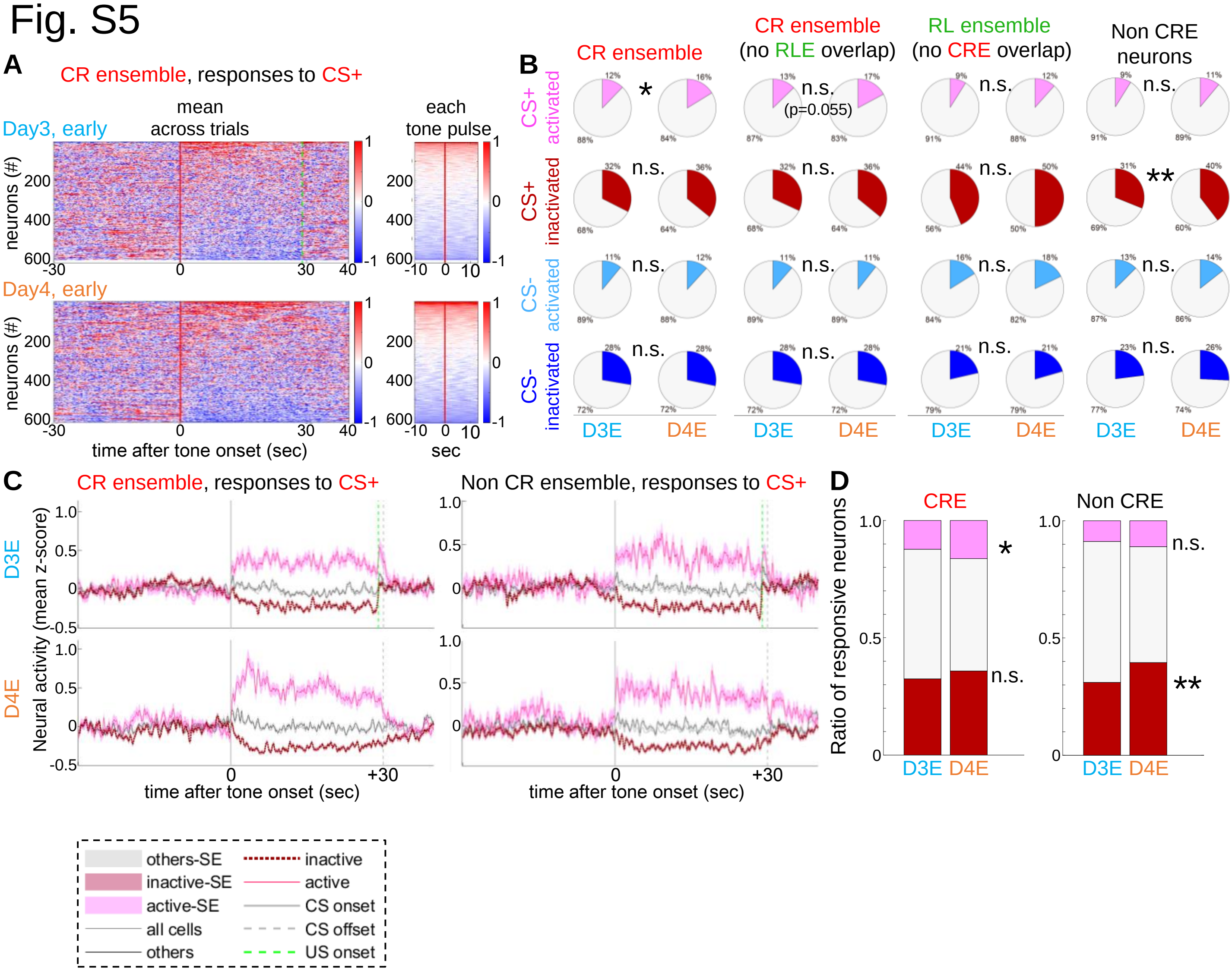


Fig. S5. Summary of response profiles of each ensemble to CS+ and CS- on day (D) 3 and D4.

(A) Different from Fig. S1, the mean activity over 3 CS trials, or over 87 onsets of 50-ms tone pulses during the 3 trials (D3-early [D3E] or D4E, respectively) in all individual neurons identified as CR ensemble neurons are plotted, indicating the enhanced responses at D4E compared with D3E. (B) Pie charts summarizing changes in CS responses. In CR ensembles neurons, CS+ activated neurons were slightly but significantly increased, while no change was observed for other features. On the other hand, in non-CRE neurons, only CS-inactivated neurons were significantly increased. (C) Mean CS responses (± s.e.m.) of each category in either CRE or non-CRE, at each temporal phase (D3E, D4E), are plotted separately. (D) Selected features in C were re-plotted as stacked bar graphs. A chi-square test was performed for the statistics in B and D. *p<0.05; **p<0.01; n.s., not significant.

Fig. S6

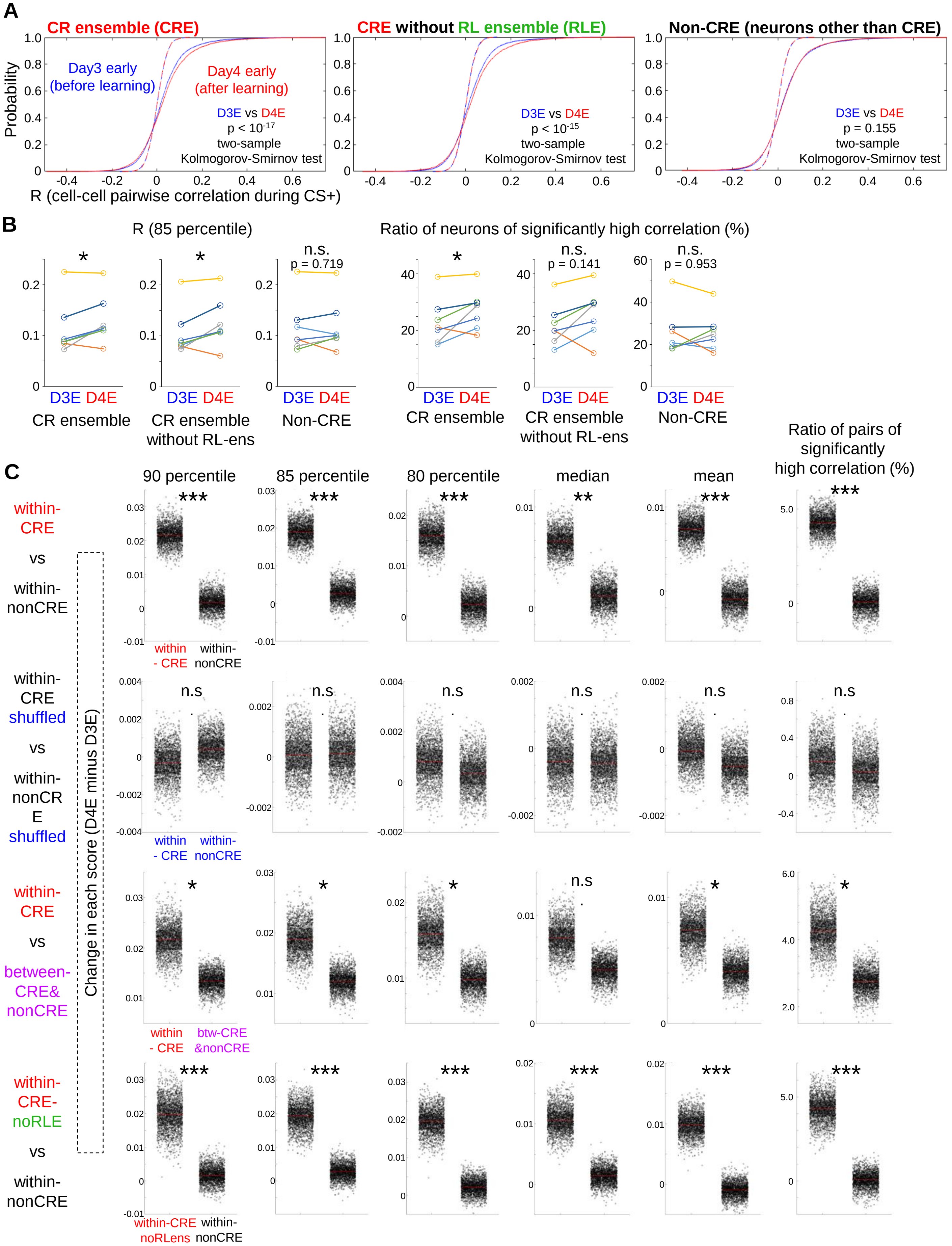

A
CR ensemble (CRE)
CRE without RL ensemble (RLE)
Non-CRE (neurons other than CRE)
Probability
1.0
0.8
0.6
0.4
0.2
0
-0.4
-0.2
0.2
0.4
0.6
Day3 early (before learning)
Day4 early (after learning)
D3E vs D4E
p < 10-17
p < 10-15
p = 0.155
two-sample Kolmogorov-Smirnov test
R (cell-cell pairwise correlation during CS+)
B
R (85 percentile)
Ratio of neurons of significantly high correlation (%)
*
n.s.
p = 0.719
p = 0.141
p = 0.953
D3E D4E
CR ensemble
CR ensemble without RL-ens
Non-CRE
Ratio of pairs of significantly high correlation (%)
C
90 percentile
85 percentile
80 percentile
median
mean
***
**
n.s
within-CRE vs within-nonCRE
within-CRE shuffled vs within-nonCRE shuffled
within-CRE vs between-CRE&nonCRE
within-CRE-noRLE vs within-nonCRE
Change in each score (D4E minus D3E)
within - CRE
within-nonCRE
btw-CRE &nonCRE
within-CRE noRLens
0.03
0.02
0.01
-0.01
0.004
0.002
-0.002
-0.004
5.0
0.8
0.4
-0.4
6.0
4.0
2.0

Fig. S6. Change in coactivity specifically observed in CR ensembles after fear conditioning.

(A) Cumulative curves were drawn for the data pooled by random resampling from all mice (2000 datapoints from each mouse, a total of 14,000 datapoints from 7 mice) to visualize the change in coactivity within CR ensemble neurons (CRE), CRE neurons without overlap with RL ensembles (CRE-noRLE), and neurons other than CRE (Non-CRE). The results of the statistical comparison between day 3-early (D3E) and D4E using a two-sample Kolmogorov-Smirnov test are shown in the respective panels. Dotted lines show the results of shuffled data (no statistically significant difference in all cases). (B) Comparison between D3E and D4E for coactivity-related values was performed with the original data. In addition to the systematic analyses based on the boot strap resampling as shown in panel C and in Fig. 4A-B, tests based on the raw data (i.e. representative values from the individual circuits shown here) also revealed a significant enhancement of the coactivity specifically in CRE, even though the number of the samples is limited (N=7). This also demonstrated that the results shown by the bootstrap resampling did not derive from artificially enhanced marginal differences. (C) Detailed analyses of the change in coactivity in dmPFC circuits after the fear conditioning (i.e. D3E vs D4E). To evaluate the enhancement of positive correlation, we calculated the 90th, 85th, 80th, and 50th percentiles, mean, and ratio of pairs of the significantly high correlation (RSHC) for each category. Some of the top two rows overlapped with the results shown in Fig. 4A-B, but here we additionally revealed results for within-CRE vs between-CRE&nonCRE (coactivity between CRE and Non-CRE), suggesting that enhanced coactivity within the CRE after the fear conditioning was specific. Results of CRE-noRLE were also consistent with those of CRE. A paired permutation test was used for the statistics in B. The data obtained by bootstrap resampling in C were statistically analyzed as described in the Methods. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; n.s., not significant. Red bars, median; gray boxes in panel C indicate the 25th and 75th percentiles.

Supplementary Movie 1.

An example of GCaMP6f signals in a field of view. ΔF/F is shown in red, over a background of the averaged-image shown in gray.