Lack of pattern decorrelation in the prelimbic circuit disables social discrimination in MeCP2-deficient mice

Pan Xu1,2#, Yuanlei Yue1,2#, Juntao Su1, Xiaoqian Sun1, Hongfei Du3, Zhichao Liu4, Rahul Simha5, Chen Zeng3, Hui Lu1,2*

1GW Institute for Neuroscience, The George Washington University, Washington, DC 20037, USA
2Department of Pharmacology and Physiology, School of Medicine and Health Sciences, The George Washington University, Washington, DC 20037, USA
3Department of Statistics, Columbia College of Art and Sciences, The George Washington University, Washington, DC 20037, USA
4Department of Physics, Columbia College of Art and Sciences, The George Washington University, Washington, DC 20037, USA
5Department of Computer Science, School of Engineering and Applied Science, The George Washington University, Washington, DC 20037, USA

*Corresponding author: Hui Lu, Ph.D.
GW Institute for Neuroscience, The George Washington University
Washington, DC 20037, USA
Email: huilu@email.gwu.edu
Phone: (202)9944166
Fax: (202)9942870
#These authors contributed equally to this work.
Abstract
The prefrontal cortical (PFC) circuit plays a central role in processing social information. PFC dysfunction has been demonstrated in many autism-associated disorders, including Rett syndrome (RTT), which is caused by loss-of-function mutations in the X-linked gene MECP2. The prelimbic PFC is hypoactive in RTT, but the precise mechanisms underlying the social avoidance remain obscure. Here we studied MeCP2-deficient mice, using in vivo calcium imaging to record neuronal activities in the prelimbic medial PFC (mPFC) while mice chose whether to interact with an object or a fellow mouse. We found that prelimbic mPFC hypoactivity restricts the responsiveness of the circuit and limits its ability to decorrelate patterns encoding social and nonsocial stimuli. Optogenetic stimulation of the prelimbic circuit throughout the social interaction restored pattern discrimination and social interactivity. This work shows that what appears to be social avoidance is actually an inability to discriminate social from nonsocial cues.

Keywords
Prefrontal cortex, prelimbic circuit, social avoidance, pattern decorrelation, stimulus classification, social discrimination, MeCP2, Rett syndrome, in vivo calcium imaging, optogenetic stimulation
Introduction

Social interactions involve a range of complex considerations, from awareness of one's internal state to assessment of myriad social cues and social memory\(^1,2\). Much of the processing of socially relevant information takes place in the prelimbic PFC circuit, which supports cognition, memory, and decision-making in humans\(^3,4,5\) and other animals\(^6,7,8,9\). Pyramidal neurons within the PFC network integrate social and affective information such as direct eye contact and friendly gestures\(^10\). Animal studies showed that neural activity in the prelimbic cortex, part of the mouse medial PFC (mPFC), carries both spatial and social information, the combination of which underlies the murine tendency to return to locations where they have previously interacted with a conspecific\(^7\). When a mouse approaches or interacts with a fellow mouse\(^6,11,12\), or senses the odor of social cues\(^13\), mPFC neurons become more active, but it is unclear how this activity encodes different stimuli to guide social discrimination.

Given the mPFC's centrality in social information processing, it is not surprising that mPFC impairment has been found to underlie social interaction abnormalities\(^12,14,15\), a salient feature of autism spectrum disorder (ASD)\(^16\). Among the more well-studied autism-associated disorders is Rett syndrome (RTT), which is caused by mutations in the X-linked gene that encodes the epigenetic regulator methyl-CpG binding protein 2 (MECP2)\(^17\). Girls with RTT appear to develop normally for the first year or two of life, then undergo a period of regression during which they lose acquired motor, cognitive, and linguistic skills and develop the autistic features of gaze avoidance, hand stereotypies, and loss of interest in social interaction and communication. MeCP2-deficient mice replicate most of the features of the disease, including social avoidance and anxiety\(^18,19\). Ex vivo brain slice studies show that reduced excitatory postsynaptic currents in the mPFC of MeCP2-deficient mice alter excitatory/inhibitory balance\(^20\), which is known to produce autism-like features\(^14\). The hypoactivity of the mPFC neurons that project to the ventral hippocampus underlies social memory deficits in a mouse model of RTT\(^21\), while increasing the activity of mPFC pyramidal neurons restores long-term retrieval of auditory conditioned fear memory and improves the respiratory dysrhythmias that afflict both RTT children and MeCP2 mutant mice\(^22\).

What is still unknown is how MeCP2 deficiency affects the mPFC's functional coding of social information. Could stimulation of the mPFC rescue the impaired coding and restore social behavior? We used in vivo Ca\(^{2+}\) imaging via a head-mounted miniature microscope to monitor prelimbic mPFC pyramidal neurons in MeCP2-deficient female mice while they performed a social approach assay in a modified linear chamber. We found that the hypoactive prelimbic circuit is less responsive to both social and nonsocial stimuli, is impaired in discriminating between the two, and loses the ability to refine its coding with experience.
Results

RTT mice lack a preference for social novelty

To explore the social behavior of our female Mecp2+/− mice (hereafter referred to as RTT mice), we used a modified three-chamber social approach task to test behaviors of five-month-old RTT mice and their wild-type (WT) littermates. The subject mouse was placed in the middle of a 60-cm linear chamber, with each flanking end-chamber containing either another mouse or an object (Fig. 1a, Supplementary Fig. 1a). Before the test, the subject mouse was habituated for 10 minutes in the middle 40-cm chamber while the 10-cm end compartments were empty. With the subject remaining in the same middle chamber, there were then three ten-minute testing sessions: in sessions 1 and 2 (S1 and S2), a mouse (M1) and an object (O) were placed in opposite end chambers. In session 2 (S2), the occupants of the end chambers were switched. In session 3 (S3), a new mouse (M2) replaced the object, so the subject mouse had a choice of two mice to interact with (or not). We measured the amount of time the subject mouse spent interacting with the occupant in either end chamber, as evidenced by sniffing.

WT mice were much more interested in their fellow mouse than in the object, but RTT mice spent equal time with both (Fig. 1b). This relative lack of social interest was also manifested by slower approach toward and less sniffing of M1 compared to the object (Fig. 1c, Supplementary Fig. 1b). These differences between genotypes narrowed in session 2 when M1 and O switched positions: this suggests that subject mice were curious about the new position of the object in the chamber that had just contained M1 (“where’s the mouse gone?”). In session 3, however, when the choice was between two mice, WT mice spent significantly more time exploring the new mouse (M2), while RTT mice spent about the same amount of time with M1 and M2 (Fig. 1b, Supplementary Fig. 1c). In fact, the RTT mice always spent 12-15% of their time interacting with whatever was in an end chamber, whereas WT mice spent about twice as much time interacting with M1 or M2 as they did with objects. This behavior was not due to differences in motor function, as locomotion in RTT mice was not impaired in the open field test at this age (Supplementary Fig. 2a). Thus, RTT mice are less sociable than WT mice, though the reason for this is not clear.

The selectivity of the prefrontal mPFC neural ensembles to different stimuli differs in RTT mice

During the social interaction test, we monitored the activity of excitatory pyramidal neurons in the prefrontal region of the mPFC using head-mounted miniature microscopes (Fig. 2a). We identified 60-120 neurons expressing the Ca2+ indicator, GCaMP6m, from each mouse (Fig. 2b), without bleaching effects in recording (Supplementary Fig. 3a, b). To verify that the neurons in the prefrontal region are involved in the representation of (social and nonsocial) stimuli, we identified neurons that respond preferentially to interaction with a specific stimulus by calculating the neuron’s activity at the start of each
behavior and then comparing this activity level with the neuron's own chance level of activity. Neurons whose activity strongly correlated with the start of a particular behavior were classified as part of an ensemble active for a specific interaction (ON neurons), while those whose activity level anti-correlated with the behavior were classified as OFF neurons (Fig. 2c). We thus identified six ensembles that were selectively tuned (either ON or OFF) to social, object, or new social stimulus, respectively. As a whole, the ensembles reliably and effectively encoded specific stimuli by compensating for temporal variation in the activity of individual neurons (Supplementary Fig. 4a). ON and OFF neurons were scattered in the field of view (Fig. 2d), and increased or decreased in activity around the onset of the interaction with a specific stimulus in both genotypes (Fig. 2e-f). RTT ensembles, however, especially the social-ON and new social-ON ensembles, consistently showed a lower amplitude of response (Fig. 2d).

We then compared the proportion of neurons engaged in different ensembles over the three sessions (Fig. 2g, Supplementary Fig. 4). Regardless of the stimulus, the RTT neural ensembles recruited fewer ON neurons than WT in S1 (Supplementary Fig. 4). In S3, RTT mice had more social-ON neurons tuned to M1 than WT (~35% vs. ~29% in WT), and fewer tuned to M2 (~31%) but more OFF neurons (45%) than WT (Supplementary Fig. 4). The more important comparison, however, is within-group across sessions. Whereas WT mice had significant changes in from one session to the next in the percentage of Social-ON neurons (Fig. 2g), there was no significant change between sessions for the RTT mice. It is interesting that by S3, RTT mice showed a greater percentage of Social-ON neurons relative to WT (~35% vs. 29%), perhaps in compensation for the overall lower firing rate (Fig. 2f). For the object stimulus, RTT mice again had fewer ON neurons (~29% vs. ~37% in WT) in S1, but the proportion of ON neurons in S2 actually became greater than WT (~38% vs. ~34%). Across sessions, WT mice recruited fewer ON neurons and more OFF neurons tuned by M1 or O interactions.

The response selectivity of the ensemble could be represented by transient rate changes in response to the different stimuli. The ON/OFF ensembles always showed their highest/lowest value in the vicinity of the stimuli they represented, for both WT and RTT mice (Supplementary Fig. 5). The transient rates of ON and OFF ensembles changed as mice moved into other areas. Overall, however, this response selectivity was sharper in WT than in RTT mice. Actually, when we considered only the immediate vicinity of one stimulus (where interactive sniffing happened), the corresponding stimulus-tuned ON ensemble displayed a much lower transient rate in RTT mice than WT, whereas the OFF ensembles did not differ between the two groups (Supplementary Fig. 5, right). We next examined the amplitude of these ensembles. The more-attractive stimuli in each session elicited greater amplitudes in the response of the corresponding interaction-tuned ON ensemble, whereas the corresponding OFF ensembles had similar responses to both sides (Supplementary Fig. 6a). Conversely, for the interaction with the less-attractive stimuli (object or old social), the tuned OFF ensembles had a decreased amplitude compared to
interaction with the other stimulus (Supplementary Fig. 6b). Therefore, the ON and OFF ensembles cooperate with each other in coding information about social salience and novelty. It is notable that the social-ON ensemble in RTT mice also showed strong selectivity to the familiar mouse (M1) in S3 (Supplementary Fig. 6, right), which accords with the lack of preference for social novelty in RTT mice.

Neural ensembles recruit neurons to encode complex information, so it is not surprising that the ensemble is not homogeneous. In both WT and RTT mice, some neurons were recruited to multiple ensembles to participate in encoding more than one type of information (Supplementary Fig. 7a). We calculated the percentage of these types of multi-functional neurons and found that RTT mice had fewer active neurons, which were tuned to encode both social and nonsocial information, in S1, indicating impairment of information coding in the RTT mPFC circuit. However, in S3, RTT mice recruited more neurons that responded specifically to the familiar (“old”) social stimuli and more inactive neurons that were tuned OFF to both old and new social stimuli (Supplementary Fig. 7b). This was consistent with the increased percentage and amplitude of social-ON neurons in the RTT ensemble, contributing to their longer interaction with M1 in S3. In different sessions, we also noticed that this recruitment process was very dynamic (Supplementary Fig. 7c). The percentage of ON or OFF neurons that remained tuned to M1 across all three sessions was very low (below 8%), but only the WT ensemble showed a difference between the proportions of stable ON and OFF neurons (Supplementary Fig. 7d). Consistent with Fig. 2g, RTT animals recruit fewer neurons to the ensemble for social interaction, and fewer of these neurons remain stable in responding specifically to social stimuli.

Pyramidal neurons in the mPFC of RTT mice are hypoactive during interactions

To examine how these differences in RTT mice affect population coding to different stimuli, we first analyzed firing rates in the prelimbic region across 3 sessions during stimuli interactions. We found that while the mPFC ensembles in WT mice became more active upon the mouse's approach to either M1 or O, this increase was very mild in RTT mice, whose transient rates were significantly lower during their approaches to both fellow mice and objects, across all three sessions (Fig. 3a-c). There was no difference in transient rates or amplitudes while the subject mouse visited the regions of the central chamber that were closer to the end chambers during habituation, meaning that interaction-induced activity changes were not induced by spatial information (Supplementary Fig. 8a). Notably, WT and RTT neurons showed similar transient rates during interactions with the social and object stimuli, but both showed higher amplitudes while interacting with M1 than O—and also with M2 over M1 (Supplementary Fig. 8b). This suggests that amplitude is related to the representation of stimulus salience and novelty, whereas the underactive Ca²⁺ transient rate underlies the relative lack of social interest of RTT mice. The average Ca²⁺ transient rate of the mPFC ensemble for both WT and RTT mice showed a significant
correlation with the amount of time they spent near another mouse (Fig. 3d, left), which was not the case for the time spent near the object (Fig. 3d, right).

As the behavioral and neural responses to interaction with the same stimulus changed across different sessions, we wondered whether the above neural activity-behavior correlation remained. For M1 interactions, the cross-session correlation between neural activity and interaction duration with M1 moved in opposite directions for the WT and RTT groups (Fig. 3e-g). In contrast, there was no clear trend for the cross-session differences between neural activity and interaction duration with O for any individual mouse (Fig. 3h-j). This result further supports the notion that hypoactivity of the mPFC excitatory ensemble contributes to the social deficit of RTT mice. However, the information coded by the firing rate seems unable to distinguish the social and nonsocial stimuli because the difference in firing rates did not correlate with the time the animal spent with the stimulus (Supplementary Fig. 8c).

The prelimbic circuit of RTT mice fails to differentiate the correlation of its activity pattern during social discrimination

The temporal relationship between spikes also encodes information. To explore the neural correlations in the prelimbic circuit during different interactions and how they are altered in the RTT mice, we first observed the temporal correlation of neural activities. In both WT and RTT mice, the pattern of mPFC neuron temporal correlation was specific to the social or object stimulus (Fig. 4a). The pairwise correlation coefficients for different stimuli across the whole mPFC neural population from all mice were distributed around zero, but with various widths (Fig. 4b). In WT mice, the full-width at half maximum (FWHM) of the correlation distribution was lower when mice interacted with the more-attractive stimulus than with the less-attractive one (Fig. 4c), which might be due to the more significant pattern decorrelation in the process of social (or new social) interactions. The decorrelation of neural activity patterns is reported to play an important role in odor classification, and the lower correlation variability could increase the signal-to-noise ratio, thereby improving information coding. When we focused on the neuron pairs with high correlation (threshold = ±0.2), the WT neurons tended to be less correlated in their encoding of social (or new social) stimulus, since the coefficient was lower for the more-attractive stimulus in each session (Fig. 4d). This difference between the neural correlations was significantly correlated with the index of discriminating M1 vs. object and the M2 vs. M1 (Fig. 4e) but was not apparent in RTT mice (Fig. 4c-e), reflecting a failure of their prelimbic circuit to perform the activity pattern decorrelation required for discriminating the social from nonsocial stimuli. This change in temporal correlation and the difference within ON/OFF ensembles were consistent with the above results (Supplementary Fig. 10a). Pattern decorrelation in the mPFC prelimbic circuit is therefore associated with social discrimination.
Experience-dependent refinement of social coding was impaired in RTT mPFC circuits

In response to one specific stimulus (M1 or O) across sessions, the WT prelimbic ensemble became more synchronized in activity, indicating the circuit is storing information. This change was specific for interacting with M1, as it did not occur while the animal was near M1 or grooming (Supplementary Fig. 10b). Loss of MeCP2 prevented this increasing synchronization (Fig. 4f). Meanwhile, the neuronal activity of this circuit in response to M1 kept decreasing across sessions (Fig. 3e). To understand the changes of activity patterns in the mPFC prelimbic circuit, we used variational auto-encoding (VAE, see Methods) (Fig. 4g). This dimension-deduction method searches for two-dimensional features that best represent the spatiotemporal pattern of each imaging frame. We examined the frames for social interaction from all three sessions and found that, consistent with the increased synchrony we observed above (Fig. 4f), the relative radius of the feature distribution kept decreasing across sessions in WT mice; the decrease between S2 and S3 was significant. We propose that this diminution reflects the fact that as the neural circuit becomes familiar with a stimulus, the pattern variation narrows (i.e., becomes more consistent) in response to that stimulus. The feature distribution for the RTT group, by contrast, remained similar in terms of the relative radius, indicating their failure to refine the neural circuit response (Fig. 4h). Next, we tested whether the changes in feature distribution were associated with social behavior. The decrease of the feature distribution radius correlated strongly with the reduced duration of interaction with M1 in individual WT mice (Fig. 4i), while no apparent correlation was found in the RTT group. To exclude the possibility that the longer time spent by WT mice interacting with M1 in the first two sessions contributed to the greater difference in Ca\textsuperscript{2+} dynamics, we used the first 150 images collected in each interacting episode to perform the same analysis and found a similar difference between WT and RTT groups (Supplementary Fig. 9). These results suggest that the experience-dependent refinement of the mPFC prelimbic circuit in encoding social behavior was impaired in RTT mice.

Enhancing mPFC pyramidal neuron activity improves social interactivity of RTT mice

Hypoactivity caused by loss of MeCP2 led to a narrower range of responsiveness to stimuli and a failure of pattern decorrelation. To determine which of these changes causes the social deficit, we decided to increase the firing rate of the ensemble and see whether this action can improve the social behavior of the RTT mice. We chose to de-repress these hypoactive excitatory neurons by suppressing inhibitory parvalbumin (PV) interneurons, which are thought to exert more influence in maintaining excitatory/inhibitory balance in the cortex than other types of interneurons. We expressed inhibitory halorhodopsin NpHR (EYFP control protein) in the PV neurons and GCamP6m in the excitatory pyramidal neurons by injecting AAV1-flox-eNpHR-EYFP (AAV1-flox-EYFP as a control) and AAV1-
CaMKII-Gcamp6m bilaterally into the prelimbic region of PV-Cre RTT mice (Supplementary Fig. 11a). This strategy enabled us to optogenetically manipulate PV neuron activity and simultaneously monitor excitatory neuron activity (Fig. 5a). Optogenetic suppression of PV neurons significantly raised the transient rate of the mPFC in RTT mice expressing NpHR, compared to control mice expressing EYFP only (Fig.s 5b, Supplementary Fig. 11a). The transient amplitude was not influenced by the manipulation (Supplementary Fig. 11b).

Because RTT mPFC neurons showed less activity during the interactions with the social stimulus (M1 in S1 and S2, and M2 in S3), we first wondered whether it is sufficient to rescue the social behavior of RTT mice if the neurons were manipulated only when the animal was in the zone near the social stimulus in each session (social zone in S1 and S2, and new social zone in S3) (Fig. 5c). This location-specific optogenetic manipulation significantly increased the mPFC neural transient rate, but elicited little improvement in the social performance of NpHR mice or the neural correlation upon social interactions (Fig.s 5d-g, Supplementary Fig. 11c). However, when we delivered light stimulation for the whole 10-min session of the social interaction test (Fig. 5h), this intervention significantly improved the social performance of RTT mice, increasing the time they spent in the vicinity of the social (or novel social) stimulus to a WT level during S1 and S3 (Fig. 5i, Supplementary Fig. 11d). PV-neuron suppression increased the transient rate markedly in responding to the stimulus interactions, whether for the whole mPFC neural population (Fig. 5i) or just the stimulus-tuned ON ensembles (Supplementary Fig. 11e). The close correlation between neural activity and behavior was maintained, such that the elevated neural activity in NpHR-expressing RTT mice indicated the longer duration of their social interaction (Fig. 5j). At last, like the WT mice, the temporal neural correlations within the mPFC circuit of NpHR-expressing RTT mice exhibited pattern decorrelation during social interaction and an experience-dependent correlation increase in responding to M1 interactions across sessions (Fig. 5k, l). These results show that the social deficit of RTT mice can be improved only by enhancing excitatory neuronal activity (in this case by suppressing inhibitory input from PV neurons) in the whole arena, while pattern decorrelation was induced for discriminating the stimuli with different interest saliences. On the other hand, the location-specific manipulation of neural ensembles was not sufficient to restore normal social behavior, which requires precise information coding for the salience of both lateral stimuli. To exclude the possibility that spatial-specific manipulation induced spatial preference or aversion, we did the place preference test with the prelimbic optogenetic interference in one randomly chosen half of an empty chamber. The mice spent a similar amount of time exploring both sides of the chamber, verifying that prelimbic manipulation did not have a spatial effect (Supplementary Fig. 11f). Those findings support a causal relationship between the failure of the pattern decorrelation in the mPFC prelimbic excitatory circuit and the social deficit of RTT mice.
Discussion

Most of our knowledge regarding autism-associated changes in cortical neuron function is based on ex vivo studies, and little is known about pathogenic changes in circuit dynamics in freely moving, behaving animals. Specifically, there has been no evidence to demonstrate how information coding of hypoactive mPFC excitatory neurons leads to social dysfunction of RTT mice. Here, we investigated how mPFC excitatory neurons encode social behavior in freely behaving RTT and WT mice. The results showed the mPFC neurons encode the interaction with different stimuli but that the hypoactivity caused by loss of MeCP2 constrains the circuit's responsiveness to different stimuli. Thus, what appears to be social avoidance in RTT mice—and perhaps other autism-related disorders that have hypoactive mPFC—is actually an inability to discriminate between social and inanimate stimuli, caused by failure to decorrelate neuronal activity in the prelimbic circuit. Moreover, suppressing the PV inhibitory neurons optogenetically to increase the responsiveness of the excitatory circuit only to the social stimuli is unable to restore the social interactivity, unless the optogenetic manipulation is applied throughout the whole behavioral assay to rescue both the responsiveness and the pattern decorrelation required for stimulus discrimination. This finding demonstrates a causal relationship and not merely a correlation between the pattern decorrelation of prelimbic neuronal activity and normal social interaction. In sum, the hypoactivity caused by loss of MeCP2 not only weakens the responsiveness to all kinds of stimuli but also impairs the pattern decorrelation of neural activities during stimulus classification, and the latter is the key cause of the social deficit in MeCP2-deficient mice.

The mPFC neurons in RTT mice showed lower activity in response to social interactions, and the transient rates showed a close correlation with social performance. This is consistent with previous evidence that mPFC pyramidal neurons show decreased excitation in brain slices from MeCP2-null mice \(^{20,34}\). The weakened, homogenized responses to both social and nonsocial stimuli in RTT make it difficult for the mice to discriminate and choose (Fig. 1b, c). The fact that increasing the response strength to one stimulus is not capable of promoting the animal’s preference for that stimulus implies that preference for a stimulus does not depend only on the strength of the response to the stimulus, but on something else.

We found that in WT mice, the pattern decorrelation of mPFC activity was more significant in the process of interacting with the attractive social stimulus compared to the other stimulus, contributing to the different correlation patterns for two lateral stimuli. The central function of pattern decorrelation is to convert input patterns representing different stimuli into separate output patterns to the downstream circuit to guide the subsequent behavior. Decorrelation is important for odor classification: in zebrafish, decorrelation of odor-evoked activity patterns in the olfactory bulb reduces the overlap between activity patterns encoding different, but structurally similar, odors \(^{35}\). Pattern decorrelation is also observed in the dentate gyrus of the hippocampus \(^{36,37}\), which pre-processes information for storage and classification in
hippocampal area CA3. In our study, the WT prelimbic circuit uses pattern decorrelation as a simple classification strategy to distinguish social from nonsocial stimuli, leading to data processing 'inequality' which can facilitate decision-making. The obvious benefit of such a simple strategy is to minimize time and effort in classifying different stimuli. In the context of MeCP2 deficiency, the activity patterns in response to different stimuli fail to diverge and create outputs that are distinguishable to the downstream target regions. We propose that this is the cause of the social deficit of RTT mice because recovering the capacity for prelimbic pattern decorrelation rescues their social deficit, despite the fact that MeCP2 deficiency affects the whole brain of this mouse model. Our result echoes a recent study using a mouse model of another autism-associated neurodevelopmental disorder showing that both the categorization of sensory stimuli and the refinement of social representations (provided by odor) were impaired in the mPFC circuit of mice lacking the autism-associated gene Cntnap2.

In WT mice, there is less variation in neural correlations during the initial interaction with the social stimulus (M1 in S1 and M2 in S3) than during subsequent exposures (M1 in S2 and S3). This may be because more attention is devoted to a new stimulus, and a more consistent inter-neuronal correlation likely improves the signal-to-noise ratio. Over multiple interactions, the tail of the correlation distributions lengthens, as the circuit patterns for social representation become more stable in WT mice. This experience-dependent refinement of the mPFC circuit was absent in RTT mice. We also found that the RTT mPFC was not able to stabilize or become streamlined in its information encoding across experimental sessions, indicating that the circuit was impaired in storing learned information. This is likely due to the impaired plasticity in the mPFC circuit. This speculation is supported by a previous ex vivo study demonstrating that loss of MeCP2 reduces excitatory postsynaptic currents, the ratio of NMDA to AMPA currents, and evoked population activity in the mPFC.

Anxiety is pronounced both in Rett syndrome and in Mecp2-deficient mice and could underlie the "social avoidance" that is so frequently seen in autism. There is another possibility, however: social behavior such as sniffing and grooming is typically rewarding for mice and involves mPFC-associated dopaminergic signaling, particularly in the prelimbic cortex. It may be that dysfunction of TH-positive dopaminergic neurons caused by loss of MeCP2 function diminishes the perceived rewards of social interaction. Dysfunctions of the reward system have been observed in other neurological disorders as well.

In summary, our results demonstrated that prelimbic hypoactivity-induced failure of stimulus classification plays a causal role in the social deficit female Mecp2 heterozygous mice, a physiologically relevant model of Rett syndrome. Supportive evidence is also presented by previous findings that restoring prefrontal cortex excitation/inhibition balance rescues social behavior in an autism mouse model, CNTNAP2-deficient mice. Although multiple other brain regions besides the mPFC are involved
in social interactivity \(^1\), and loss of MeCP2 results in excitatory hypoconnectivity in many forebrain structures \(^42\), restoring mPFC function locally is sufficient to rescue the social abnormality of RTT. Together with a previous study reporting that activation of the mPFC of RTT mice could reverse cognitive and respiratory symptoms \(^22\), we propose that stimulating the mPFC could be a useful strategy to mitigate many core symptoms of RTT patients. Given that dysfunction of the PFC has been reported in patients with autism spectrum disorder \(^43, 44\), modulating the PFC neural could be a strategy for treating many autism-associated disorders.
Methods

Animals

Mouse maintenance and use were in accordance with NIH Guidelines and with the approval of the Institutional Animal Care and Use Committee of the George Washington University. Mecp2\(^{+/−}\) mice on the 129S6SvEvTac strain were obtained from Dr. Huda Zoghbi’s lab in Baylor College of Medicine, while Camk2-cre and PV-Cre mice of pure C57BL/6 background were purchased from Jackson (JAX#005359). Female Camk2-cre, Mecp2\(^{+/−}\) (RTT) and Camk2-cre, Mecp2\(^{+/+}\) (WT) mice were obtained by breeding male Camk2-cre mice and female Mecp2\(^{+/−}\) mice, used for excitatory neuron imaging. Similarly, male PV-Cre mice were mated with female Mecp2\(^{+/−}\) mice, and the Mecp2\(^{+/−}\) pups were used for optogenetic manipulation research. We conducted experiments on WT and RTT mice of 3.5 to 4 months, following the same experimental procedure. Animals were given ad libitum access to standard mouse chow and water, housed 4 to 5 per cage in a temperature (23 ± 1 °C) and humidity (50 ± 10%) controlled room with a 12 hr light-dark cycle.

Virus injection and GRIN lens implantation

For imaging excitatory neurons, AAV1.GCaMP6m virus (Baylor College of Medicine) were stereotaxically injected as previously described. Briefly, Camk2-cre mice (RTT and WT) were anaesthetized and placed in a stereotaxic frame (Neurostar, Tübingen, Germany), then a 1.1 mm-diameter craniotomy (AP: +1.95 mm, M/L: -0.5 mm) was made with a high-speed rotary stereotaxic drill (Model 1474, AgnTho’s AB, Lidingö, Sweden). The virus was injected unilaterally (Nanojector II, Drummond Scientific) into the left region of mouse prelimbic cortex, with the stereotaxic coordinates from bregma: +1.95 mm anterior-posterior (AP), -0.35 mm medial-lateral (ML), -2.3~2.5 dorsal-ventral (DV) following a high resolution atlas. A total of 600 nL virus (diluted with 600 nL PBS) were injected at the rate of 30 nL/min and the needle was left in place for an additional 5 min after injection.

For optogenetic manipulation of PV neurons with simultaneous calcium imaging of pyramidal neurons, two viruses, including 600 nL halorhodopsin AAV5-EF1a-DIO-eNpHR3.0-EYFP (AAV5-EF1a-DIO-EYFP as control virus, Addgene, Watertown, MA,) and 600 nL of CaMKII-dependent virus AAV5-CaMKII-GCaMP6m (obtained from Neuroconnectivity core at Baylor College of Medicine, Houston, TX) were injected successively into the prelimbic region (AP: +1.85 ~+2.00 mm; ML: -0.35 mm; DV: -2.3~2.5 mm) of the PV-cre, Mecp2\(^{+/−}\) mice, following the similar surgical protocol as above. Then, a 1-mm diameter GRIN lens (Inscopix, Palo Alto, CA) was lowered into left prelimbic region (AP: +1.95 mm; ML: ± 0.35 mm; DV: -2.1~2.3 mm), 0.2 mm above virus injection site, at the speed of 50 μm/min, and then cemented in place (Metabond S380, Parkell). Mice were allowed to recover on a heat pad and thereafter monitored for 7 days closely during which they received injection of analgesic.
Baseplate attachment

Three to four weeks after surgery, the virus expression in the anesthetized mouse was checked with a miniaturized microscope (Inscopix, Palo Alto, CA). If GCaMP+ neurons were visible and clear, the microscope attached with baseplate would be hung on mouse’s skull window and lowered to assess the focus plane. Then, the baseplate was dental cemented onto the skull and capped with a cover, with microscope unattached. Before the behavioral test, mice were habituated to the environment of the test room with the dummy microscope mounted and handled for 5-7 days, 40 min each day.

Behavioral tests

Each day only one behavioral test was conducted and in each test the chamber was cleaned with 70% ethanol between trials. Topscan behavior analysis system (Clever Sys, VA) was used to monitor the animal behaviors, which would send a TTL signal simultaneously to trigger the microscope recording neural activity at the beginning of each test.

Social approach test. This test was carried on according to the previous protocol 6 with some improvement. To facilitate in vivo imaging with microscope, the conventional three-chamber apparatus was modified to one open square box (45×10×20 cm) with two small removable lateral chambers (10×10×40 cm), and chambers was separated by 1-cm-spaced thin metal wires, allowing mice to interact with stimuli. Test consists of three 10-min sessions, which was conducted following 10-min habituation in center open box. In session 1 (S1), an age- and weight-matched strange same-sex conspecific (the first social stimulus, M1) and inanimate object (nonsocial stimulus, O) were separately placed two lateral chambers randomly. In session 2 (S2), the positions of those two stimuli were switched, which was designed to diminish the spatial influence in stimuli-induced neural activities. Mice’s preference to social stimuli than inanimate object is used to evaluate their sociability. In session 3 (S3), a second age- and weight- matched same-sex conspecific (new social stimulus, M2) was used to replace the object in lateral chamber. It was used to evaluate social novelty preference, reflected by mice’s propensity to spend more time with a new conspecific than with a familiar one. At the beginning of each session, the testing mouse was placed in the center of open box and allowed to explore freely. The time spent involved in social interaction, object interaction, social zone, object zone and middle zone were calculated.

Open field test (OFT). This test was conducted in a square box (dimensions: 50 × 50 × 50 cm). Mouse was gently placed in the central field and allowed to explore freely during 10-min testing session. Locomotor activity was recorded by camera. The center is defined as the central 25 cm x 25 cm square area, while corner is sector area with 12.5 cm radius in each corner. Total distance traveled and time spent in the center or corner of the box was calculated.
**Place preference paradigm (Social CPP).** In this assay, the open-field square box was separated even with a wall into stimulation zone and non-stimulation zone, and a door in the middle of the wall allowed mice transfer freely between two zones. The locations of two zones were counter-balanced. Optogenetic LED light was delivered to mice only when they entered into the stimulation zone, whereas the excitation LED was kept on during 10-min testing session. The movement and duration of mice in different zones were recorded and compared.

**Calcium imaging with Miniature microscope**
Imaging in freely moving mice were performed using a head-mounted miniaturized microscope (nVista HD 2.0.4, Inscopix, Palo Alto, CA), which triggered by a TTL pulse from Topscan system to achieve the simultaneous acquisitions of calcium signal and behavioral video. Microscope was mounted onto mouse’s head right before imaging and imaging data were acquired at a frame rate of 15 Hz and at 1024 × 1024 pixels. The LED power was set to 0.3-1 mW and the gain was 1 to 2 depending on fluorescence intensity. Each individual mouse used the same imaging parameters of itself across three sessions.

**Optogenetic manipulation during imaging**
The nVoke imaging system (2.0, Inscopix, Palo Alto, CA) was used for combined optogenetic and imaging experiments, with two LED lights transmitted through GRIN lens into prelimbic region. The excitation LED of blue light (448 nm) was set to 0.5-1.0 mW for GCaMP imaging, with an analog gain of 1.0-2.0, whereas the optogenetic LED of amber light (590 nm, 3.5 mW, continuous) was used for activity inhibition of neurons expressed with NpHR. The optogenetic manipulation was delivered over the whole session or only when mice stay in social zone. These two strategies of optogenetic manipulation were applied to the same group of mice with a one-month interval between the two tests. The sequence of the two strategies was randomly picked.

**Histology**
Recording sites were verified by histological examination of lesions induced by lens implantation. Mice were anesthetized (i.p.) with an overdose of ketamine (400 mg/kg) and xylazine (20 mg/kg), then perfused transcardially with phosphate buffer solution (PBS) first, followed by 4% paraformaldehyde (PFA). Mice brains (with skulls and baseplate) were post-fixed with 4% PFA for 3 days. Then, brains were removed and sliced using a vibrate slicer (Vibratome Series 1000, St. Louis, MO) into 50-100 µm sections and mounted on slides. Slides were incubated and stored in 1:1000 Hoechst in 1x PBS (Invitrogen, Carlsbad, CA) to label cell nuclei. Slides were observed and imaged on a Confocal...
Microscope (Zeiss LSM 710, Oberkochen, Germany) and fluorescence of viral expressions and location of GRIN lenses could be recorded.

**Data analysis**

Behavior. Behavioral data were automatically tracked by top-down movies using Topscan behavioral data acquisition software (CleverSys, Reston, VA). The 2D locations of mice were also tracked and defined as different zones, such as social zone, object zone, and middle zone in the test chamber. Based on this, the type and duration of detailed behavioral events involved in different tests were recognized and extracted by the software, including social interaction with M1 (O), stay in social zone, grooming in the social zone, and approaching to M1 (O).

Calcium image processing. Calcium images were processed off-line using Inscopix Data Processing software (version 1.2.1). Briefly, raw movies were processed with the preprocessing, spatial filter, and motion correction subsequently. For normalizing the calcium signal, the average projection of filtered movies was generated as the background fluorescence ($F_0$), and instantaneous normalized $Ca^{2+}$ fluorescent signals ($\Delta F/F$) was calculated according to the formula, $\frac{(\Delta F/F)_i = (F_i - F_0)}{F_0}$, where $i$ represents each frame. Then, Individuals cells were identified using the principle component and independent component (PCA-ICA) analyses with no spatial or temporal down-sampling, and the regions of interest (ROI) were selected as candidate cells based on signal and image. Time-stamped traces of neurons were exported to Python (v3.0), where custom-written scripts were used for analysis onwards.

For calculating the calcium activity, spikes from fluorescence traces were predicted using unsupervised learning method sparse non-negative deconvolution 47, 48 and near-online OASIS algorithm 49. To observe the neural activity involved in one specific behavior, we aligned the frames of image and frames of behavioral data with each other and marked the image frames with the corresponding behavioral event labels.

**Identification of interaction-tuned neural ensembles**

To identify the interaction-tuned neural ensembles in each session, we evaluated the response preference of each neuron to one specific stimulus interaction. In brief, we first calculated the actual similarity (Sa) between vectors of calcium trace (ck) and behavior interaction (b), using the formula: $2b \cdot c_k / (|b|^2 + |c_k|^2)$ 23. Then, the behavior vector was randomly shuffled used to calculate a new similarity (Ss) with a neural trace for a neuron, which was repeated 5000 times in order to generate Ss distribution histogram. The neuron was classified as ON if its Sa was greater than 99.95% of the Ss distribution, conversely, it was an OFF neuron if Sa was less than 0.05% of the Ss distribution. Otherwise, it belonged to Others ensemble with Sa falling between 0.05% and 99.95%.
After the neuron classification, the proportion of ON, OFF, and Other neurons of each session were calculated for each mouse. Meanwhile, the transient rate and amplitude of those neuron ensembles were also calculated for each mouse in each session.

Neuron-behavior correlation analysis for neural ensemble. The correlation between calcium neural activity and specific behaviors within one neural ensemble was evaluated in two ways. One is that the Pearson correlation coefficient between the averaged calcium activity of all neurons and behavior vector, which represent the coding ability of a neuron ensemble. The other is to calculate the average of individual Pearson correlation coefficients between each individual neural activity and behavior vector, which reflect the representing ability of all individual neurons within an ensemble.

Neural ensemble overlap percentage. For a pair of ensembles, such as X and Y, we first found the number of the overlapped neurons (N_{XY}) between two ensembles. Meanwhile, the neuron number of ensemble X and Y were marked as N_X and N_Y. Thus, the overlap percentage was calculated by the formula: \( \frac{2N_{XY}}{N_X+N_Y} \). The range of overlap was from 0 to 1.

Neural correlation. For evaluating the functional connectivity within neurons, the average of pairwise Pearson correlation coefficients of calcium traces in the neuron population was calculated. For a specific behavior, the corresponding neural functional connectivity was indicated by the averaged pairwise Pearson correlation coefficient of the calcium traces, which were involved in that behavior duration. To observe the correlation between those high correlated neurons, we only observed the high correlation coefficient values, which exceed the threshold we set.

Variational autoencoder encoder (VAE) analysis of neuron image frames. This analysis was based on the autoencoder model, which contains two basic parts of encoder and decoder. The encoder compresses the input image into shrunk latent space, while the decoder reconstructs the input image using the latent space information with the mean-square-error (MSE) between original and reconstructed images minimized \(^{29,30}\). VAE pushes further by adding a Gaussian distribution restriction of the latent space distribution and additional noise on the latent space. In order to represent the neural circuit pattern by dimensional reduction of information, we only used the encoder part, which is able to extract the most intrinsic, important, and behavior-correlated features in the latent 2D projection.

Before training, the frame images, as input, were reconstructed to reduce background noise, through filling neuron pixels with neural activity and the rest pixels with 0. We also shrank the frame image to reduce computation complexity. Then, VAE model was implemented using Keras (https://keras.io/examples/variational_autoencoder/). A typical VAE architecture was adopted following the previous report about image dataset of MNIST handwriting digits (http://yann.lecun.com/exdb/mnist/). The encoder was set with two intermediate convolutional layers (with 64 and 32 3×3 filters, respectively; stride 2) followed by a fully connected layer of 16 and the latent
embedding layer of 2. During model training, the loss for optimization consists of reconstruction loss which stands for the MSE by the decoder, and KL divergence between the target Gaussian distribution and the actual distribution of the latent layer coding. The optimizer used for training was RMSProp created by Hinton in 2012. A batch size of 128 and a training epoch of 200 was adopted. The training epoch was set long enough so that the loss for all samples has converged with little fluctuation. Other parameters follow the default setting of Keras. The frame projections in the 2D latent space were used for downstream analysis.

In the 2D projection of frames in each session, the center location of the frames in the 2D space was calculated, and the distribution range (radius) was defined as the average distance of all frames to the center in the 2D latent space.

Statistics

All statistical analyses were performed using SPSS Statistics (version 24, IBM, Armonk, NY), Excel (Microsoft, Redmond, WA), and Python custom scripts. Since all the data could pass the normal distribution test (D'Agostino and Pearson), a two-tailed paired sample or unpaired t-test was applied in comparison. For multiple-factors comparison, two-way RM ANOVA was used, followed by Bonferroni-corrected post hoc comparisons. Statistical significance was taken as */# \( P<0.05 \), **/## \( P<0.01 \), ###/### \( P<0.001 \). All data are represented as mean ± SEM unless otherwise specified. The linear curve was fitted to indicate transient rate-behavior correlation, and the correlation was tested by regression analysis.

Data and materials availability

All data and codes in the main text or the supplementary materials are available upon request.
References

1. Chen P, Hong W. Neural circuit mechanisms of social behavior. *Neuron* **98**, 16-30 (2018).

2. Anderson DJ. Circuit modules linking internal states and social behaviour in flies and mice. *Nature Reviews Neuroscience* **17**, 692-704 (2016).

3. Adolphs R. The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* **60**, 693-716 (2009).

4. Euston DR, Gruber AJ, McNaughton BL. The role of medial prefrontal cortex in memory and decision making. *Neuron* **76**, 1057-1070 (2012).

5. Bicks LK, Koike H, Akbarian S, Morishita H. Prefrontal Cortex and Social Cognition in Mouse and Man. *Front. Psychol.* **6**, 1805 (2015).

6. Lee E, *et al.* Enhanced neuronal activity in the medial prefrontal cortex during social approach behavior. *J. Neurosci.* **36**, 6926-6936 (2016).

7. Murugan M, *et al.* Combined social and spatial coding in a descending projection from the prefrontal cortex. *Cell* **171**, 1663-1677. e1616 (2017).

8. Watson KK, Platt ML. Social signals in primate orbitofrontal cortex. *Curr. Biol.* **22**, 2268-2273 (2012).

9. Zhou T, *et al.* History of winning remodels thalamo-PFC circuit to reinforce social dominance. *Science* **357**, 162-168 (2017).

10. Grossmann T. The role of medial prefrontal cortex in early social cognition. *Front. Hum. Neurosci.* **7**, 340 (2013).

11. Kim Y, *et al.* Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Report* **10**, 292-305 (2015).

12. Brumback AC, *et al.* Identifying specific prefrontal neurons that contribute to autism-associated abnormalities in physiology and social behavior. *Mol. Psychiatry* **23**, 2078-2089 (2018).

13. Levy DR, *et al.* Dynamics of social representation in the mouse prefrontal cortex. *Nat. Neurosci.* **22**, 2013-2022 (2019).

14. Yizhar O, *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171-178 (2011).

15. Selimbeyoglu A, *et al.* Modulation of prefrontal cortex excitation/inhibition balance rescues social behavior in CNTNAP2-deficient mice. *Sci. Transl. Med.* **9**, eaah6733 (2017).

16. Kanner L. Autistic disturbances of affective contact. *Nervous Child* **2**, 217–250 (1943).
17. Sandweiss AJ, Brandt VL, Zoghbi HY. Advances in understanding of Rett syndrome and MECP2 duplication syndrome: prospects for future therapies. The Lancet Neurology 19, 689-698 (2020).

18. Kaufmann W, et al. Social impairments in Rett syndrome: characteristics and relationship with clinical severity. J. Intellect. Disabil. Res. 56, 233-247 (2012).

19. Samaco RC, McGraw CM, Ward CS, Sun Y, Neul JL, Zoghbi HY. Female Mecp2(+/−) mice display robust behavioral deficits on two different genetic backgrounds providing a framework for pre-clinical studies. Hum. Mol. Genet. 22, 96-109 (2013).

20. Sceniak MP, Lang M, Enomoto AC, James Howell C, Hermes DJ, Katz DM. Mechanisms of functional hypoconnectivity in the medial prefrontal cortex of Mecp2 Null mice. Cereb. Cortex 26, 1938-1956 (2016).

21. Phillips ML, Robinson HA, Pozzo-Miller L. Ventral hippocampal projections to the medial prefrontal cortex regulate social memory. Elife 8, e44182 (2019).

22. Howell CJ, et al. Activation of the medial prefrontal cortex reverses cognitive and respiratory symptoms in a mouse model of rett syndrome. eNeuro 4, e0277 (2017).

23. Liang B, et al. Distinct and dynamic ON and OFF neural ensembles in the prefrontal cortex code social exploration. Neuron 100, 1-15 (2018).

24. Hamm JP, Peterka DS, Gogos JA, Yuste R. Altered cortical ensembles in mouse models of schizophrenia. Neuron 94, 153-167.e158 (2017).

25. Antic SD, Hines M, Lytton WW. Embedded ensemble encoding hypothesis: The role of the “Prepared” cell. J. Neurosci. Res. 96, 1543-1559 (2018).

26. Cohen MR, Maunsell JHR. Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. 12, 1594-1600 (2009).

27. Mitchell JF, Sundberg KA, Reynolds JH. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron 63, 879-888 (2009).

28. Xu W, Morishita W, Buckmaster Paul S, Pang Zhiping P, Malenka Robert C, Südhof Thomas C. Distinct neuronal coding schemes in memory revealed by selective erasure of fast synchronous synaptic transmission. Neuron 73, 990-1001 (2012).

29. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, (2013).

30. Kingma DP, Welling M. Stochastic gradient VB and the variational auto-encoder. In: 2nd International Conference on Learning Representations (ICLR)). 2014 edn (2014).

31. Atallah BV, Bruns W, Carandini M, Scanziani M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. Neuron 73, 159-170 (2012).
32. Goel A, et al. Impaired perceptual learning in a mouse model of Fragile X syndrome is mediated by parvalbumin neuron dysfunction and is reversible. Nat. Neurosci. 21, 1404-1411 (2018).

33. Xue M, Atallah BV, Scanziani M. Equalizing excitation-inhibition ratios across visual cortical neurons. Nature 511, 596-600 (2014).

34. Katz DM, et al. Rett syndrome: crossing the threshold to clinical translation. Trends Neurosci. 39, 100-113 (2016).

35. Niessing J, Friedrich RW. Olfactory pattern classification by discrete neuronal network states. Nature 465, 47-52 (2010).

36. Leutgeb JK, Leutgeb S, Moser M-B, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 315, 961-966 (2007).

37. McHugh TJ, et al. Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. Science 317, 94-99 (2007).

38. Treves A, Tashiro A, Witter MP, Moser EI. What is the mammalian dentate gyrus good for? Neuroscience 154, 1155-1172 (2008).

39. Samaco RC, et al. Loss of MeCP2 in aminergic neurons causes cell-autonomous defects in neurotransmitter synthesis and specific behavioral abnormalities. PNAS 106, 21966-21971 (2009).

40. Dichter GS, Richey JA, Rittenberg AM, Sabatino A, Bodfish JW. Reward Circuitry Function in Autism During Face Anticipation and Outcomes. J. Autism Dev. Disord. 42, 147-160 (2012).

41. Harris LT, McClure SM, van den Bos W, Cohen JD, Fiske ST. Regions of the mPFC differentially tuned to social and nonsocial affective evaluation. Cognitive, Affective, & Behavioral Neuroscience 7, 309-316 (2007).

42. Kron M, et al. Brain activity mapping in mecp2 mutant mice reveals functional deficits in forebrain circuits, including key nodes in the default mode network, that are reversed with ketamine Treatment. J. Neurosci. 32, 13860-13872 (2012).

43. Carper RA, Courchesne E. Localized enlargement of the frontal cortex in early autism. Biol. Psychiatry 57, 126-133 (2005).

44. Oblak A, Gibbs T, Blatt G. Decreased GABAA receptors and benzodiazepine binding sites in the anterior cingulate cortex in autism. Autism Res. 2, 205-219 (2009).

45. Belzung C, Turiault M, Griebel G. Optogenetics to study the circuits of fear- and depression-like behaviors: a critical analysis. Pharmacology Biochemistry and Behavior 122, 144-157 (2014).

46. Franklin TB, et al. Prefrontal cortical control of a brainstem social behavior circuit. Nat. Neurosci. 20, 260-270 (2017).
47. Vogelstein JT, et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology* **104**, 3691-3704 (2010).

48. Pnevmatikakis EA, Giovannucci A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291**, 83-94 (2017).

49. Friedrich J, Zhou P, Paninski L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).
Acknowledgments

We would like to thank Dr. Huda Zoghbi at Baylor College of Medicine for providing the Mecp2-deficient mice. We also thank V.L. Brandt for helpful discussions and critical comments on the manuscript and Ethan Jin for significant input to data analysis.

**Funding:** This work was supported by funding from the NIH grant 5R00NS089824 and Brain & Behavior Research Foundation 2017 NARSAD Young Investigator Grant, and The George Washington University 2018 Cross-Disciplinary Research Fund to H. L and C.Z.

**Author contributions**

P.X., Y.Y., and H.L. designed and performed the experiments. P.X., J.S., X.S., H.D. and Z.L. analyzed data. P.X., Y.Y., J.S., X.S., H.D., Z.L., R.S., C.Z., and H.L. reviewed and interpreted data. P.X., J.S., X.S., H.D. and Z.L., and H.L. wrote the manuscript with critical input from Y.Y., R.S. and C.Z.

**Additional information**

- **Supplemental Information** including 11 figures is in another file.
- **Competing interests:** Authors declare no competing interests.
**Fig. 1** Female *Mecp2*+/− (RTT) mice display lack of interest in social interactions

a. The social approach test. The subject mouse is in the middle 40-cm-long chamber; the 10-cm end compartments contain either an object or another mouse. In the first 10-minute test session (S1), a strange mouse (M1) and an object (O) were placed in the end chambers; session 2 (S2) used the same stimuli but swapped their positions; in session 3 (S3) a new mouse (M2) replaced O, so that the subject mouse must choose whether to interact with a familiar or strange mouse (M1 vs. M2).

b. Top: raster plots of the percentage of time mice spent in interacting M1 in S1 and S2, and M2 in S3 during 10-minute session. The time is divided into 20-second bins; the color ranges from deep red (0%) to yellow (100%). Bottom: the percentage of time that WT (n=12, gray) and RTT (n=11, light red) mice spent in interaction with M1, O, and M2 in each session. Data are represented as mean ± SEM. * P < 0.05, WT vs. RTT; # P < 0.05, ## P < 0.01, ### P < 0.001, M1 vs. O, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

c. Discrimination indices of sociability and social novelty preference of WT and RTT mice. Discrimination index indicates the difference in the amount of time spent with each stimulus in a given session (M1 vs. O in S1 and S2; M2 vs. M1 in S3), relative to the total time of stimuli interactions out of 10-minute testing. Data are represented as mean ± SEM. * P < 0.05, WT (n=12) vs. RTT (n=11), two-tailed Student’s t-test.
Figure 2

(a) Diagram of experimental setup with miniScope and Blue LED.

(b) Field of View with cells labeled 1 to 5.

(c) Graph showing ΔF/F (Z-score) for Social-ON and Social-OFF conditions.

(d) Graph showing ΔF/F changes around onset for Social-ON and Social-OFF conditions.

(e) Heatmap of ΔF/F changes for Social-ON and Social-OFF conditions.

(f) Graph showing percentage (%) of Social-ON and New social-ON across S1, S2, and S3.

(g) Graph showing percentage (%) of WT and RTT for Social-ON and New social-ON.
**Fig. 2 Neurons in the mPFC respond to different stimuli**
a Schematic of the experimental approach. *Left:* the calcium signal of the freely behaving mice in the chamber was imaged with a head-mounted miniature microscope, while their behaviors were recorded by a high-speed camera. *Middle:* imaging setup and histology of the PL region with GRIN lens implantation. Scale bar, 500 μm. *Right:* the confocal image of PL neurons showing GCaMP6m expression. Scale bar, 10 μm.
b *Left:* the field of view under a GRIN lens in one mouse with identified neurons numbered and colored. *Right:* fluorescence traces of example neurons marked in the left panel.
c Raster plots of individual neural activities and calcium traces of averaged group activity of social-ON and social-OFF neurons from one representative mouse in S1. Vertical pink bars indicate discrete episodes of social interaction.
d Spatial distributions of social-ON, social-OFF and other neurons from one WT and one RTT mouse at S1.
e Raster plots of calcium activities of individual social-ON and social-OFF neurons around the onset of social interaction (2 s before and 8 s after), sorted by the time points of maximal (ON neurons) or minimal activities (OFF neurons) appearance.
f *Left:* averaged calcium traces of different stimulus-tuned ON/OFF neural ensembles around the onset of the corresponding interactions (10 s before to 10 s after). Solid lines and shaded regions represent the averaged value and SEM, respectively. *Right:* amplitude changes of ON and OFF ensembles (as shown in left traces) around the onset of the interactions with the stimuli they were tuned to (10 s after minus 10 s before). ΔF/F=change in fluorescence over baseline fluorescence intensity. Values are represented as mean ± SEM. *P < 0.05, **P < 0.01, RTT (n=11) mice vs. WT (n=12), two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
g Percentage of mPFC ON/OFF neurons tuned to different stimuli in WT (left) and RTT (right) mice as recorded in each session. Each dot indicates one mouse. Lines connect data from one mouse; thicker solid lines connect the averaged value from all the mice in each group. *P < 0.05, between sessions; # P < 0.05, WT (n=12) vs. RTT (n=11). Two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
Figure 3
Fig. 3 Reduced transient rate of mPFC neural circuit is related to the lack of interest in social interactions of RTT mice

**a** Top: the response field of the neural response. The normalized transient rates of all WT and RTT neurons are presented as solid lines, with standard error (SEM) conveyed by shaded regions. **Bottom:** heatmap of the spatial field of the averaged response of WT neurons.

**b** Responses of one representative neuron from WT and RTT mouse during social interactions. Trials were organized by the length of interactions.

**c** The averaged calcium transient rate when WT (n=12) and RTT (n=11) mice were engaged with different stimuli over the three sessions. Data are represented as mean ± SEM. * P < 0.05; **P < 0.01, two-tailed Student’s t-test.

**d** Correlation between transient rate and proportion of time spent interacting with mouse (left) or object (right) in individual WT (n=9) and RTT (n=8) mice in S1. Each dot represents an individual mouse; Pearson’s correlation coefficients were calculated across genotypes. *P < 0.05, **P < 0.01, regression.

**e** Transient rate changes of the mPFC neurons in the WT (n=9) and RTT (n=8) mice over three sessions of interacting with M1.

**f** Relationship between the transient rate and amount of time that WT and RTT mice respond to M1 in S1 and S3. Circles and triangles represent individual mice in S1 v.s. S3; dashed lines connect values from the same mouse over two different sessions.

**g** Relationship between the changes of transient rate and the duration of interaction of WT and RTT mice with M1 over sessions 1 to 3. Each dot represents an individual mouse. The Pearson’s correlation coefficients were calculated across genotypes. ***P < 0.001, regression.

**h, i, j** are similar to **e, f, g**, respectively, except in response to O in the first two sessions.
Figure 4

(a) Graph showing cell number over time (s) for Social and Object conditions.

(b) Graph illustrating the probability distribution of RTT.

(c) Scatter plot comparing correlation coefficients of WT and RTT conditions.

(d) Graph showing the correlation coefficient of highly correlated neurons.

(e) Graph depicting the discrimination index and correlation coefficient differences.

(f) Graphs comparing interaction with M1 and O for WT and RTT conditions.

(g) Variational Auto-encoder network frames and feature scatter plots.

(h) Graphs showing the radius relative to S1 for WT and RTT conditions.

(i) Graphs illustrating the percentage time of interaction with M1 for WT and RTT conditions.
Fig. 4 The reduced neural correlation difference between two stimuli underlies the impaired social behaviors of RTT mice

a Raster plots of calcium activity of individual mPFC neurons during social interaction and object interaction from representative WT and RTT mice. Calcium transient traces are sorted by the peak activity time of social interaction (left top) and object interaction (right bottom), while the calcium traces of sorted neurons in responding to object (right top) and social (left bottom) interactions were also plotted in the same sequence of the sorted plot.

b Distribution of pair-wise Pearson correlation coefficients among the all mPFC neurons of WT (n=9) and RTT (n=8) mice in responding to stimuli in each session. The black and red plots represent M1 interactions, while the lighter colors (grey and pink) represent the interaction with the other stimulus (O or M2). The full width at half maximum (FWHM) of peaks was calculated and displayed at the right top box of each panel.

c The averaged FWHM of correlation distribution of individual WT (n=9) and RTT (n=8) mice during interactions with different stimuli. Data are represented as mean ± SEM. *P < 0.05, two-tailed Student’s t-test.

d The averaged Pearson’s correlation coefficient of highly correlated neuronal pairs in the WT and RTT mice during different stimuli interactions. High correlation: |correlation coefficient| > 0.2. Error bars indicate SEM. *P < 0.05, two-tailed Student’s t-test.

e Correlation between the neural correlation difference in two lateral stimuli and the discrimination index of M1 vs. O (left) and M2 vs. M1 (right) in WT (n=9) and RTT (n=8) mice. Each dot represents an individual mouse; Pearson’s correlation coefficients were calculated across genotypes. *P < 0.05, **P < 0.01, regression.

f The averaged Pearson’s correlation coefficient of highly correlated neuronal pairs of the WT and RTT mice during interaction with M1 across three sessions (left) and O interactions across the first two sessions (right). High correlation: correlation coefficient| > 0.2. The coefficient of highly-correlated neurons kept increasing in WT (significantly in S2 versus S1), but not in RTT mice, which led to a significant difference between genotypes in S3. Error bars indicate SEM. Two-way RM ANOVA with Bonferroni-corrected post hoc comparisons. *P < 0.05, compare between different genotypes; # P < 0.05, S2 compared with S1.

g The variational auto-encoder approach (VAE) transforms the coding information within each imaging frame into two features (1 and 2), represented by one dot in a two-dimensional plot. Frames recorded during social (M1) interaction periods over three sessions were transformed and their corresponding feature distributions were color-coded dark blue, light blue, and red, respectively. The circular shaded area represents the distribution radius (See Methods).

h Top: Feature distributions of M1 interactions in S1, S2, and S3 from the representative WT and RTT mice. Bottom: changes of relative distribution radius (normalized to S1) responding to M1 interactions across three sessions of each individual WT (n=11) and RTT (n=7) mice. Each dot indicates the normalized average value. Lines connect data from one mouse; thicker solid lines connect the averaged value from all the mice in each group. Error bars indicate SEM. *P < 0.05, ns, no significance. Two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

i Correlation between changes in the duration of social interaction and relative radius between sessions (S1-S2 and S2-S3). Pearson’s correlations were calculated separately for WT (gray) and RTT (red) mice. Dots indicate individual mice. *P < 0.05, **P < 0.01, regression.
Figure 5

a) Diagram showing the experimental setup with light and neuronal signals.

b) Graphs showing the time course of interaction with M1 and M2.

c) Chart illustrating PL manipulation in the (new) social zone.

d) Bar graphs depicting the transient rates of M1 in S1, S2, and S3.

e) Graph showing the transient rate to M1 and the % time of interaction with M1.

f) Graphs displaying the FWHM of coefficient distribution for S1, S2, and S3.

g) Graphs illustrating the correlation of highly correlated neurons.

h) Chart showing PL manipulation in the whole arena.

i) Bar graphs representing the transient rate to M1, O, M2, and M1 in S1, S2, and S3.

j) Graphs depicting the transient rate to M1 and the % time of interaction with M1.

k) Graphs showing the FWHM of coefficient distribution for RTT and WT.

l) Graphs illustrating the correlation of highly correlated neurons.
Fig. 5 Elevating the activity of mPFC pyramidal neurons throughout the whole arena improves social interactions of RTT mice

a Recording configuration (left) and schematic of light path (right) for the head-mounted microscope enabling optogenetic parvalbumin (PV) suppression and simultaneous imaging of pyramidal neurons (PNs) in the PL region of the PV-Cre RTT mice. Inhibitory halorhodopsin NpHR (control protein, EYFP) and GCaMP6m were expressed by PV interneurons and excitatory PNs, respectively.

b Sample calcium traces of neurons from representative NpHR or EYFP mice recorded in their home cage with and without amber light stimulation. The shaded area represents the period with the amber light on.

c The first manipulation involved delivering amber light only when the mouse was in the social zone (SZ) or new social zone (NSZ). SZ, object zone (OZ), and NSZ indicate the 10-cm regions of the central chamber that are nearest the end chambers containing M1, O, and M2, respectively.

d The (new) social-zone manipulation of mPFC did not affect the time spent in interacting with the (new) social stimuli (top), though the transient rate (bottom) of the stimulated zone was elevated by the manipulation. *P < 0.05, NpHR (n=9) versus EYFP (n=8), two-tailed Student’s t-test.

e Pearson’s correlations between transient rate and time of M1 interactions in S1 of NpHR (n=9) and EYFP mice (n=8). Each dot represents one mouse; Pearson’s correlations were calculated across genotypes. *P < 0.05, regression.

f The averaged FWHM of correlation distribution of individual EYFP (n=9) and NpHR (n=8) mice during interactions with different stimuli. Data are represented as mean ± SEM. Two-tailed Student’s t-test.

g The averaged pairwise Pearson’s correlation coefficients of highly correlated mPFC neuron pairs of NpHR and EYFP mice in responding to M1 interaction in each session. High correlation: |correlation coefficient| > 0.2. No significant difference was found through sessions, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

h The second experimental design involved delivering optogenetic stimulation throughout the 10-min testing session.

i The whole-arena stimulation of mPFC increased the time spent (top) in interaction with M1 (in S1, S2) and M2 (in S3) and the transient rate (bottom) in responding to stimuli interactions of NpHR mice (n=9), compared with EYFP mice (n=8). *P < 0.05, two-tailed Student’s t-test.

j-l similar to e-g, except with the whole-arena manipulation of mPFC. The transient rate and time of M1 interactions still had a close relationship (j); the FWHM of neural correlation distribution of NpHR mice showed significant difference between two lateral stimuli k and the coefficient of highly-correlated neurons in response to M1 interaction kept increasing and became significantly elevated in S3, compared with EYFP mice (l). *P < 0.05, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons. NpHR (n=9) vs. EYFP (n=8).
Supplementary figures and legends:

a

Wean mice

Inject virus

Attach baseplate

Three-chamber social test

b

Supplementary Fig. 1 Five-month old female WT, but not MeCP2+/- mice ("RTT mice"), show preference for social interaction.

a Workflow of the surgical and testing experiments.

b Left. The ratio of time spent sniffing M1 to time spent sniffing O in S1 and S2, and M2 to M1 in S3. Right. Ratio of speed of approach to M1 vs. O in S1 and S2, and M2 vs. M1 in S3. *P < 0.05, RTT (n=11) mice vs. WT (n=12), two-tailed Student’s t-test.

c Percentage of time (out of 10 minute session) WT and RTT mice spent in interaction with M1 (left) and O (right) across sessions. Data are represented as mean ± SEM. *P < 0.05, RTT (n=11) mice vs. WT (n=12), two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
Supplementary Fig. 2 Five-month-old female RTT mice are more anxious than WT in the open field test but are not impaired in motor activity.

a Left: graphic presentation of area definition in an open field test (OFT) chamber. Right: the averaged total distance of the movement in the chamber for 10 minutes, and the percentage of time spent in the center area of WT and RTT mice at 5 months of age. The motor activity of RTT mice is equivalent to WT at this stage, but RTT mice are much more anxious.

b Relationship between the transient rate of the mPFC neurons and the running speed (per 10-second bin) of one representative mouse. There was no significant correlation.
Supplementary Fig. 3 The whole-session transient rates did not differ between WT and RTT mice.

a Raster plots of calcium activity of individual mPFC neurons over the whole testing session from one representative WT and one representative RTT mouse.

b Averaged mPFC transient rate of the 10-min testing session from WT (n=12) and RTT (n=11) mice. No significant difference was found across sessions (paired t-test) or between genotypes (two-tailed Student’s t-test).
Supplementary Fig. 4 Percentage of mPFC neurons in WT and RTT mice tuned ON or OFF to different stimuli.

Blue, light blue, and gold sections of each bar reflect S1, S2, and S3, respectively; the numbers therein are percentages, and arrows indicate a statistically significant increase or decrease \((P < 0.05)\) in RTT mice \((n=11)\) compared with WT \((n=12)\) in the same session; * \(P < 0.05\), indicates significant changes across sessions. Two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
Supplementary Fig. 5 Transient rates of mPFC neural ensembles across different zones did not change as much in RTT mice as in WT mice. 

Top: Diagrams showing the gradient of responses of social-ON and -OFF ensembles by spatial field from one representative mouse; darker gray means higher response. Bottom: The transient rate of stimulus-tuned ON/OFF ensembles in WT mice (black, grey) and RTT mice (red, pink) across different locations of the chamber. "Sniff" means the mouse was at the end of the chamber and interacting with a stimulus; social zone (SZ) and object zone (OZ) indicate the 10-cm regions of the central chamber that are nearest the end chambers in which M1 and O, respectively, reside; transition zone (TZ) means the middle chamber. Data are represented as mean ± SEM. *$P < 0.05$, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
Supplementary Fig. 6  Transient amplitudes of mPFC neuronal responses to different stimuli were similar in WT and RTT mice.

Amplitudes of various stimulus-tuned ON and OFF ensembles of WT and RTT mice responding to different stimuli over three sessions. Ensembles tuned to the more-attractive (M1 in S1, S2 and M2 in S3) in a and less-attractive stimuli (O in S1, S2 and M1 in S3) in b in each session. Values are displayed as mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001, O vs. M1, or M1 vs. M2, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
Supplementary Fig. 7 Neural recruitment to ensembles is altered in RTT mice.

a Left: four groups of overlapped neurons, categorized by how they respond to two different stimuli within one session (here is S1). Right: spatial distribution of the active neurons of WT (blue) and RTT (purple) mice.

b Proportion of the neural ensembles (as defined in A) in each session of WT and RTT mice. S=social, O=object; NS=new social (novel mouse); OS=old social (familiar mouse). The neurons that are most stable in WT mice are always specific to (new) social interaction, whereas those in the RTT mice are specific to the object in S2, and to the familiar mouse in S3. Data are shown as mean ± SEM. *P < 0.05, **P < 0.01, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

c Map of ensemble coding properties across three testing sessions form a representative WT and RTT mouse.

d Percentage of neurons that were always tuned to M1 (either ON or OFF) over all three sessions. Data are shown as mean ± SEM. *P < 0.05, RTT (n=11) versus WT (n=12), two-tailed Student’s t-test.
Supplementary Fig. 8 mPFC transient rate and amplitude changes to different stimuli in WT and RTT mice.

a Averaged transient rate (left) and amplitude (right) of mPFC neurons when WT and RTT mice were in the 10-cm ends of the central chamber.

b Averaged amplitudes when WT and RTT mice were engaged with different stimuli in each session. Data are presented as mean ± SEM. *P < 0.05, **P < 0.01, O vs. M1, or M1 vs. M2, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

c Correlation between the changes in transient rates and discrimination indices of M1 vs. O (left) and M2 vs. M1 (right) of WT (n=9) and RTT (n=8) mice. Each dot represents an individual mouse. Regression analysis found no significant correlation.
Supplementary Fig. 9 VAE analysis is unaffected by changes in frame number.

a Pearson’s correlation between relative distribution radiiuses (normalized to S1) based on the first 150 “social interaction” frames and those based on all “social interaction” frames. Each dot indicates the relative radius values of one mouse based on two different frames. All mice were included (n=18). Pearson’s correlations were calculated across genotypes. ***P < 0.001, regression.

b Changes of relative distribution radius (compared to S1) across three sessions averaged from the first 150 “social interaction” frames of each individual WT (n=11) and RTT (n=7) mouse. Each dot indicates the normalized averaged values. Lines connect data points from the same mouse. The thicker solid line connects the averaged value from all the mice of each group. Error bars indicate SEM. *P < 0.05, ns, no significance, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.
Supplementary Fig. 10 The neural correlation within the mPFC ON/OFF ensembles of WT and RTT mice.

a The averaged Pearson’s correlation coefficient of highly correlated neuronal pairs in the ON (left) and OFF (right) ensembles of WT (n=9) and RTT (n=8) mice. High correlation: |correlation coefficient| > 0.2. Error bars indicate SEM. *P < 0.05, **P < 0.01, two-tailed Student’s t-test.

b The averaged Pearson’s correlation coefficient of highly correlated neuronal pairs of the WT and RTT mice when they were in social zone near M1 (left) or grooming through three sessions (right). The high correlation: |correlation coefficient| > 0.2. Error bars indicate SEM. Two-way RM ANOVA with Bonferroni-corrected post hoc comparisons. There was no significant difference between genotypes.
Supplementary Fig. 11 Whole-arena optogenetic stimulation of mPFC enhances social interactions of RTT mice.

a Histology of PL with GRIN lens implantation and viral expression in pyramidal neurons (left square) and PV neurons (right square).

b Averaged mPFC transient rate of NpHR and EYFP mice during the 10-min testing session under two different PL manipulation conditions (see Fig. 5C and H). Changes across sessions are slight, which diminishes the likelihood of a bleaching effect. Two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

c Manipulation of the mPFC solely within the (new) social zone did not significantly influence the discrimination indices (left) or the correlation coefficient of highly correlated neuronal pairs (middle) of NpHR (n=9) mice, compared with EYFP (n=8) mice, but these two indicators had a close relationship with each other in two groups (right). High correlation: |correlation coefficient| > 0.2. Error bars indicate SEM. *P < 0.05, two-tailed Student’s t-test.
d Manipulation of the mPFC throughout the duration of the experiment (not just during social interaction) significantly increased the discrimination indices (left) and the correlation coefficient of highly correlated neuronal pairs (middle) of NpHR mice, compared with EYFP mice; these two indicators displayed a close relationship with each other in two groups (right). High correlation: |correlation coefficient| > 0.2. *P < 0.05, two-tailed Student’s t-test.

e Averaged transient rate (left) and amplitude (right) of NpHR (n=9) and EYFP (n=8) mice during ordinary home cage behavior, with and without PL optogenetic manipulation. Values were plotted as mean ± SEM. *P < 0.05, amber light ON versus OFF, two-way RM ANOVA with Bonferroni-corrected post hoc comparisons.

f Left: Place preference testing, with stimulation given only when the mouse moves into the stimulation site (orange). Right: the percentage of time NpHR and EYFP mice spent in each side. There was no significant difference between groups or conditions (two-way RM ANOVA with Bonferroni-corrected post hoc comparisons).