UCSF
UC San Francisco Electronic Theses and Dissertations

Title

Encoding of social behavior by psychiatrically relevant cell types of the medial prefrontal cortex

Permalink

https://escholarship.org/uc/item/23v1b8jx

Author

Turner, Marc Lawrence

Publication Date

2024

Peer reviewed|Thesis/dissertation

Encoding of social behavior by psychiatrically relevant cell types in the medial prefrontal cortex
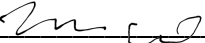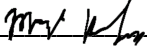
by
Marc Turner

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION
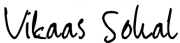of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Michael Brainard

Chair

Mazen Kheirbek

DEVANAND MANOLI

Vikaas Sohal

Committee Members

i

Encoding of social behavior by psychiatrically relevant cell types in the medial prefrontal cortex

Marc Lawrence Turner

Abstract

Our behaviors during social interactions are wide-ranging and highly contextual. The norms that dictate how we should act vary based on who we are interacting with, societal context, and past experiences. Social cognition refers to the ability of animals to utilize both experience and the knowledge of learned social norms to select appropriate behaviors for specific contexts. The medial prefrontal cortex (mPFC) has a long-established role in executive function and cognition, and has also been found to be critical for the production of normal social behavior. However, the precise role of the mPFC in social behavior and the neural dynamics that occur in this region during social interaction have not been explored in great detail. Elucidating the precise role of the mPFC during social interaction is critical to understanding how abnormal social phenotypes arise in the array of psychiatric illnesses that impinge upon the PFC.

One way to approach this problem is to utilize rodent models in combination with the wide array of molecular and genetic tools available in this system. These tools allow for studying the role of specific brain regions, neuronal subtypes, and genes in behavior. In particular, one tool that has expanded our understanding of the neural foundations of complex behavior is *in vivo* calcium imaging of neuronal activity with miniaturized, head-mounted microscopes. In combination with tools that allow for expression of the genetically encoded calcium indicator GCaMP in genetically-defined neuron subpopulations, this approach makes it possible to record

activity specifically from cell types that have relevance to psychiatric disorders while animals can move and behave freely.

Here, we describe two distinct studies that utilize this approach to record activity in the mPFC of mice from neuron populations that have been implicated in psychiatric disorders. In the first study, we record activity in layer 5 projection neurons of the mPFC in both wild-type mice and mice that have the *Tbr1* gene deleted specifically in cortical layer 5 neurons. *Tbr1* is a high-confidence autism risk-gene, and mPFC layer 5 projection neurons have been identified as a hub of autism risk-gene expression. We find that these *Tbr1* layer 5 conditional knockout animals (*Tbr1*-cKO) display abnormal social and anxiety-related avoidance behaviors. During social interactions, encoding of behavior by correlated activity of mPFC neurons is diminished in cKO animals, while correlated activity remains intact during anxiety-related behavior. We also identify signals in mPFC neural ensembles that are predictive of approach-avoidance decisions, but are lost in the *Tbr1* cKO mouse model.

In the second study, we record specifically from mPFC neurons that express the dopamine receptor D2R, which has been implicated in disorders including schizophrenia and depression. We also record from mPFC D2R+ neurons after knocking out the D2R, to assess the role of the receptor itself in socially recruited activity. These animals performed the 3-chamber social assay, and we find significant center chamber associated activity in D2R+ neurons that is lost when the D2R is deleted. Inhibition of mPFC D2R+ neurons specifically in the center chamber of this task leads to an overall increase in the number of social interactions.

Table of Contents

List of Figures

Chapter 1 – General Introduction

1.1   The role of the mPFC in behavior

The medial prefrontal cortex (mPFC) is a hub for information processing and cognition in the mammalian brain that is essential for the support of complex behaviors. Important cognitive functions such as attention, motivation, working memory, long-term memory, and emotional control, amongst others, have all been shown to rely on the mPFC (Euston et al., 2012; Miller, 1999; Sawaguchi & Goldman-Rakic, 1991). While this set of functions may seem vast and wide-ranging, they are all critical for integrating external signals with internal states and past experiences. Such integration is imperative for selecting appropriate behaviors in environments that are constantly changing, and in circumstances where there are competing internal drives and motivations. With dense afferent input from many sensory and limbic structures (Hoover & Vertes, 2007), and an array of efferent projections to both cortical and subcortical targets (Gabbott et al., 2005), the mPFC is well positioned for this integrative role. Thus, the activity within the mPFC can be viewed as promoting behaviors that lead to survival and success, while suppressing behaviors that are inappropriate or maladaptive. Disruptions to this brain region can lead to cognitive impairments, and have been linked to many psychiatric disorders, including schizophrenia, depression, and autism (Gregoriou et al., 2014; Mier & Kirsch, 2017; Rossi et al., 2009; Takumi et al., 2020). Given this important role in cognition, information processing, and behavioral control, it is unsurprising that the mPFC has been found to also be critical for normal social behaviors (Bicks et al., 2015; Kennedy & Adolphs, 2012).

1.2   Social circuitry of the brain

Social interactions amongst individuals are extremely important for the survival of many species as they allow for the sharing of resources and information, promote mating and procreation, and

can provide protection against potential threats. Broadly, social behaviors can be viewed as either affiliative and prosocial, or avoidant and antisocial. Appropriate behavioral selection depends on many factors, including sensory cues, internal states and emotions, and both personal and shared goals. In real time, an individual must react to a conspecific's actions and behaviors and update their own.

In support of this complex repertoire of social behaviors is a network of interconnected brain regions known as the 'social brain' (Bicks et al., 2015; Modi & Sahin, 2019; Yizhar & Levy, 2021). Despite the relative complexity of human social behavior compared to rodents, and even non-human primates, this network is largely conserved across these species. Activity in some of these brain areas has been shown to drive particular social behaviors. For example, stimulation of different neuron populations of the hypothalamus can directly elicit innate behaviors, such as mating and aggression (Lischinsky & Lin, 2020). Moreover, oxytocin releasing neurons of the hypothalamus which project throughout the brain have shown to gate social reward and promote parental behaviors (Carcea et al., 2021; Dölen et al., 2013). The amygdala is another region known to support social behaviors across species, with studies showing activity in this area reflects aspects of conspecific identity in both rodents and humans (Gothard, 2020; Li et al., 2017). Other regions within the social brain network include the hippocampus (HPC), nucleus accumbens (NAc), and ventral tegmental area (VTA), which have all been linked with certain aspects of social behavior. For example, the HPC is important for social memory, and the NAc and VTA both play critical roles in social reward (Chen & Hong, 2018).

In contrast with these other nodes of the social brain, the mPFC has not been linked strongly to any one aspect of social behavior. However, the mPFC possesses dense reciprocal connections with many of these other brain structures, including the hypothalamus, amygdala, NAc, and HPC (Ko, 2017). This interconnectivity with much of the social brain allows the mPFC

to integrate the various streams of social information along with internal states to exert top-down control and promote appropriate behaviors.

1.3   Encoding of social information in the mPFC

Given its position within the social brain network and much evidence in humans that the mPFC is critical for normal social behaviors, many recent studies have begun to probe social behavior-related activity within this region. By recording the activity of individual neurons in the mPFC during social interactions, we can begin to parse exactly what variables are being encoded and which cell types are playing a critical role.

Within the mPFC, there is a rich heterogeneity of neuronal subtypes. Primarily, there are excitatory pyramidal neurons and inhibitory interneurons, which make up ~80% and ~20% of neurons of the mPFC, respectfully. These neurons are organized into the classic multi-layered cortical architecture, with different layers receiving distinct afferent inputs from other brain areas. Moreover, these neurons can also be classified by their downstream projection targets, with excitatory neurons in more superficial layers (layers 2/3) predominantly targeting other cortical regions and the amygdala, and deep layers (layers 5/6) favoring subcortical areas such as the thalamus and striatum. Beyond this cellular diversity, the mPFC can also be divided into three main subregions: the anterior cingulate cortex (ACC), prelimbic cortex (PrL), and the infralimbic cortex (IL). These subregions have been shown to play differing roles in many of the functions attributed to the mPFC.

Studies recording single-unit activity neurons in the mPFC of rodents have reported neurons that either increase or decrease their activity during social interaction, implying that a many cell-types may be involved (Frost et al., 2021; Jodo et al., 2010; Liang et al., 2018). Moreover, one of these studies found that throughout the course of a 3-chamber social interaction test, these ensembles of 'ON' and 'OFF' neurons were dynamic (Liang et al., 2018).

This finding highlights that socially-recruited activity in the mPFC may not reflect just the animals behavior, but may also reflect the context and internal states of the animals as well. Indeed, one study identified a population of NAc-projecting neurons in the mPFC that exhibited a mixture of spatial and social encoding, becoming strongly active only when the mice were interacting in certain locations of the task arena (Murugan et al., 2017). Another study found neurons that encode conspecific sex (Kingsbury et al., 2020).

Beyond encoding of an individual's own behavior and context, studies in humans and primates have shown that mPFC activity can reflect the actions and states of conspecifics (Demolliens et al., 2017; Denny et al., 2012). More recent studies have found similar encoding in the mPFC of rodents. One study identified somatostatin (SST) interneurons as being particularly important for encoding the affective states of conspecifics (Scheggia et al., 2020). In another study, mPFC activity was recorded in pairs of interacting mice, and found separate groups of neurons that encoded self behavior and the behavior of the other mouse (Kingsbury et al., 2019). Beyond this, activity of these "other behavior" encoding neurons was correlated with activity in the conspecific's mPFC, and the strength of this correlated activity depended on the social ranks of the mice. Social rank is an extremely important variable in social decision making and action selection, and many studies have established the mPFC as playing a central role in encoding the relative status of animals from mice to humans (Wang et al., 2014; Zhou et al., 2017). Therefore, relative social rank may be an extremely important factor in shaping mPFC encoding of social behavior.

Ultimately, many of these studies have recorded activity from heterogenous populations of mPFC neurons, but studies have shown that mPFC neurons with different afferent inputs or downstream efferent targets can have varying impacts on social behavior. For example, mPFC to NAc projection neurons have been found to be especially important for social reward and recognition, and their activation can promote interactions (Murugan et al., 2017; Xing et al.,

2021). mPFC to basolateral amygdala (BLA) projections have also been shown to facilitate social recognition, while projections to the hypothalamus and periaqueductal gray (PAG) shape aggressive and defensive behaviors (Biro et al., 2018; Franklin et al., 2017; Tan et al., 2019). How the activity of these various projection neurons is coordinated during an ongoing social interaction remains unknown, and future recording studies should be directed specifically at these populations. Of particular interest is how the expression of known risk genes for psychiatric disorders overlap with these populations, and how perturbations of these genes impact these socially relevant populations compared to other mPFC neurons.

1.4 References

Bicks, L. K., Koike, H., Akbarian, S., & Morishita, H. (2015). Prefrontal Cortex and Social
Cognition in Mouse and Man. *Frontiers in Psychology*, *6*, 1805.
https://doi.org/10.3389/fpsyg.2015.01805

Biro, L., Sipos, E., Bruzsik, B., Farkas, I., Zelena, D., Balazsfi, D., Toth, M., & Haller, J. (2018).
Task Division within the Prefrontal Cortex: Distinct Neuron Populations Selectively
Control Different Aspects of Aggressive Behavior via the Hypothalamus. *Journal of
Neuroscience*, *38*(17), 4065–4075. https://doi.org/10.1523/JNEUROSCI.3234-17.2018

Carcea, I., Caraballo, N. L., Marlin, B. J., Ooyama, R., Riceberg, J. S., Mendoza Navarro, J. M.,
Opendak, M., Diaz, V. E., Schuster, L., Alvarado Torres, M. I., Lethin, H., Ramos, D.,
Minder, J., Mendoza, S. L., Bair-Marshall, C. J., Samadjopoulos, G. H., Hidema, S.,
Falkner, A., Lin, D., … Froemke, R. C. (2021). Oxytocin neurons enable social
transmission of maternal behaviour. *Nature*, *596*(7873), 553–557.
https://doi.org/10.1038/s41586-021-03814-7

Chen, P., & Hong, W. (2018). Neural Circuit Mechanisms of Social Behavior. *Neuron*, *98*(1), 16–
30. https://doi.org/10.1016/j.neuron.2018.02.026

Demolliens, M., Isbaine, F., Takerkart, S., Huguet, P., & Boussaoud, D. (2017). Social and
asocial prefrontal cortex neurons: A new look at social facilitation and the social brain.
Social Cognitive and Affective Neuroscience, *12*(8), 1241–1248.
https://doi.org/10.1093/scan/nsx053

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A Meta-analysis of Functional
Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for
Mentalizing in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742–
1752. https://doi.org/10.1162/jocn_a_00233

Dölen, G., Darvishzadeh, A., Huang, K. W., & Malenka, R. C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature*, *501*. https://doi.org/10.1038/nature12518

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The Role of Medial Prefrontal Cortex in Memory and Decision Making. *Neuron*, *76*(6), 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Franklin, T. B., Silva, B. A., Perova, Z., Marrone, L., Masferrer, M. E., Zhan, Y., Kaplan, A., Greetham, L., Verrechia, V., Halman, A., Pagella, S., Vyssotski, A. L., Illarionova, A., Grinevich, V., Branco, T., & Gross, C. T. (2017). Prefrontal cortical control of a brainstem social behavior circuit. *Nature Neuroscience*, *20*(2), 260–270. https://doi.org/10.1038/nn.4470

Frost, N. A., Haggart, A., & Sohal, V. S. (2021). Dynamic patterns of correlated activity in the prefrontal cortex encode information about social behavior. *PLOS Biology*, *19*(5), e3001235. https://doi.org/10.1371/journal.pbio.3001235

Gabbott, P. L. A., Warner, T. A., Jays, P. R. L., Salway, P., & Busby, S. J. (2005). Prefrontal cortex in the rat: Projections to subcortical autonomic, motor, and limbic centers. *Journal of Comparative Neurology*, *492*(2), 145–177. https://doi.org/10.1002/cne.20738

Gothard, K. M. (2020). Multidimensional processing in the amygdala. *Nature Reviews Neuroscience*, *21*(10), 565–575. https://doi.org/10.1038/s41583-020-0350-y

Gregoriou, G. G., Rossi, A. F., Ungerleider, L. G., & Desimone, R. (2014). Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nature Neuroscience*, *17*(7), 1003–1011. https://doi.org/10.1038/nn.3742

Hoover, W. B., & Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure and Function*, *212*(2), 149–179. https://doi.org/10.1007/s00429-007-0150-4

Jodo, E., Katayama, T., Okamoto, M., Suzuki, Y., Hoshino, K., & Kayama, Y. (2010). Differences in responsiveness of mediodorsal thalamic and medial prefrontal cortical neurons to social interaction and systemically administered phencyclidine in rats. *Neuroscience*, *170*(4), 1153–1164. https://doi.org/10.1016/J.NEUROSCIENCE.2010.08.017

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, *16*(11), 559–572. https://doi.org/10.1016/j.tics.2012.09.006

Kingsbury, L., Huang, S., Raam, T., Ye, L. S., Wei, D., Hu, R. K., Ye, L., & Hong, W. (2020). Cortical Representations of Conspecific Sex Shape Social Behavior. *Neuron*, *107*(5), 941-953.e7. https://doi.org/10.1016/j.neuron.2020.06.020

Kingsbury, L., Huang, S., Wang, J., Gu, K., Golshani, P., Wu, Y. E., Behavior, B., & Brain, B. A. (2019). Correlated Neural Activity and Encoding of Behavior across Brains of Socially Interacting Animals Interbrain correlation of neural activity X B Other X A Self X B Self X A Other ΣX A Self + ΣX A Other ΣX B Self + ΣX B Other ≈ Kingsbury et al. *Cell*, *178*, 429–446. https://doi.org/10.1016/j.cell.2019.05.022

Ko, J. (2017). Neuroanatomical substrates of rodent social behavior: The medial prefrontal cortex and its projection patterns. *Frontiers in Neural Circuits*, *11*. https://doi.org/10.3389/fncir.2017.00041

Li, Y., Mathis, A., Grewe, B. F., Osterhout, J. A., Ahanonu, B., Schnitzer, M. J., Murthy, V. N., & Dulac, C. (2017). Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. *Cell*, *171*(5). https://doi.org/10.1016/j.cell.2017.10.015

Liang, B., Zhang, L., Barbera, G., Fang, W., Zhang, J., Chen, X., Chen, R., Li, Y., & Lin, D. T. (2018). Distinct and Dynamic ON and OFF Neural Ensembles in the Prefrontal Cortex Code Social Exploration. *Neuron*, *100*(3), 700-714.e9. https://doi.org/10.1016/j.neuron.2018.08.043

Lischinsky, J. E., & Lin, D. (2020). Neural mechanisms of aggression across species. *Nature Neuroscience*, *23*(11), 1317–1328. https://doi.org/10.1038/s41593-020-00715-2

Mier, D., & Kirsch, P. (2017). Social-Cognitive Deficits in Schizophrenia. In M. Wöhr & S. Krach (Eds.), *Social Behavior from Rodents to Humans: Neural Foundations and Clinical Implications* (pp. 397–409). Springer International Publishing. https://doi.org/10.1007/7854_2015_427

Miller, E. K. (1999). The Prefrontal Cortex: Complex Neural Properties for Complex Behavior. *Neuron*, *22*(1), 15–17. https://doi.org/10.1016/S0896-6273(00)80673-X

Modi, M. E., & Sahin, M. (2019). A unified circuit for social behavior. *Neurobiology of Learning and Memory*, *165*, 106920. https://doi.org/10.1016/j.nlm.2018.08.010

Murugan, M., Jang, H. J., Park, M., Miller, E. M., Cox, J., Taliaferro, J. P., Parker, N. F., Bhave, V., Hur, H., Liang, Y., Nectow, A. R., Pillow, J. W., & Witten, I. B. (2017). Combined Social and Spatial Coding in a Descending Projection from the Prefrontal Cortex. *Cell*, *171*(7), 1663-1677.e16. https://doi.org/10.1016/j.cell.2017.11.002

Rossi, A. F., Pessoa, L., Desimone, R., & Ungerleider, L. G. (2009). The prefrontal cortex and the executive control of attention. *Experimental Brain Research*, *192*(3), 489–497. https://doi.org/10.1007/s00221-008-1642-z

Sawaguchi, T., & Goldman-Rakic, P. S. (1991). D1 Dopamine Receptors in Prefrontal Cortex: Involvement in Working Memory. *Science*, *251*(4996), 947–950. https://doi.org/10.1126/science.1825731

Scheggia, D., Managò, F., Maltese, F., Bruni, S., Nigro, M., Dautan, D., Latuske, P., Contarini, G., Gomez-Gonzalo, M., Requie, L. M., Ferretti, V., Castellani, G., Mauro, D., Bonavia, A., Carmignoto, G., Yizhar, O., & Papaleo, F. (2020). Somatostatin interneurons in the prefrontal cortex control affective state discrimination in mice. *Nature Neuroscience*, *23*(1), 47–60. https://doi.org/10.1038/s41593-019-0551-8

Takumi, T., Tamada, K., Hatanaka, F., Nakai, N., & Bolton, P. F. (2020). Behavioral neuroscience

of autism. *Neuroscience and Biobehavioral Reviews*, *110*.

https://doi.org/10.1016/j.neubiorev.2019.04.012

Tan, Y., Singhal, S. M., Harden, S. W., Cahill, K. M., Nguyen, D. T. M., Colon-Perez, L. M.,

Sahagian, T. J., Thinschmidt, J. S., De Kloet, A. D., Febo, M., Frazier, C. J., & Krause, E.

G. (2019). Oxytocin receptors are expressed by glutamatergic prefrontal cortical neurons

that selectively modulate social recognition. *Journal of Neuroscience*, *39*(17), 3249–

3263. https://doi.org/10.1523/JNEUROSCI.2944-18.2019

Wang, F., Kessels, H. W., & Hu, H. (2014). The mouse that roared: Neural mechanisms of social

hierarchy. *Trends in Neurosciences*, *37*(11), 674–682.

https://doi.org/10.1016/j.tins.2014.07.005

Xing, B., Mack, N. R., Guo, K. M., Zhang, Y. X., Ramirez, B., Yang, S. S., Lin, L., Wang, D. V.,

Li, Y. C., & Gao, W. J. (2021). A Subpopulation of Prefrontal Cortical Neurons Is

Required for Social Memory. *Biological Psychiatry*, *89*(5), 521–531.

https://doi.org/10.1016/j.biopsych.2020.08.023

Yizhar, O., & Levy, D. R. (2021). The social dilemma: Prefrontal control of mammalian

sociability. *Current Opinion in Neurobiology*, *68*, 67–75.

https://doi.org/10.1016/j.conb.2021.01.007

Zhou, T., Zhu, H., Fan, Z., Wang, F., Chen, Y., Liang, H., Yang, Z., Zhang, L., Lin, L., Zhan, Y.,

Wang, Z., & Hu, H. (2017). History of winning remodels thalamo-PFC circuit to reinforce

social dominance. *Science*, *357*(6347), 162 LP – 168.

https://doi.org/10.1126/science.aak9726

Chapter 2 - The role of *Tbr1* in encoding of socioemotional behavior in the mPFC

2.1 Abstract

Disruptions in many genes linked to autism spectrum disorder (ASD) affect synaptic function and socioemotional behaviors in mice. However, exactly how synaptic dysfunction alters neural activity patterns underlying behavior remains unknown. We addressed this using mice lacking the high confidence ASD gene *Tbr1* in cortical layer 5 (L5) projection neurons (*Tbr1* cKO mice). These mice have known deficits in synaptic input to L5 neurons and social behavior, and we additionally find abnormal anxiety-related avoidance. Calcium imaging of prefrontal L5 neurons revealed that despite markedly reduced overall activity in cKO mice, normal numbers of neurons were recruited into ensembles encoding social and anxiety-related behaviors. However, during social behavior, cKO ensemble stability and coactivity were diminished. Within anxiety assays, cKO ensembles no longer predicted approach-avoidance decisions. These results show how prefrontal ensembles encode socioemotional behaviors, and reveal that *Tbr1* loss disrupts multineuron coordination within these ensembles even when ensemble recruitment remains intact.

2.2 Introduction

Autism spectrum disorder (ASD) encompasses a range of neurodevelopmental conditions that are defined by the presence of specific core features, namely deficits in the social domain as well as repetitive behaviors. Despite these shared core features, the severity and precise nature of behavioral alterations vary greatly across patients. Given this heterogeneity, it is perhaps unsurprising that >100 genes have been identified which impart a large individual risk for ASD (Satterstrom et al., 2020; H. R. Willsey et al., 2022). Many of these genes have been linked to

specific cell-types, molecular pathways and cellular processes, and in many cases, mice which model disruptions in these genes have been shown to exhibit altered social behavior. Despite this progress in understanding the genetics of ASD and generating animal models with face validity, there remains a critical gap in our understanding between the cellular/molecular and behavioral levels. In particular, how exactly these diverse genetic disruptions alter the patterns of neural activity that normally support and promote complex behaviors is still largely unknown.

Within the field of ASD genetics, much effort has been made to identify common molecular and cellular pathways impacted by the risk-gene network. Additionally, the temporal and spatial expression of these risk genes have been leveraged to pinpoint convergent brain areas and developmental timepoints that may be particularly critical for ASD pathology (Satterstrom et al., 2020; A. J. Willsey et al., 2013). Notably, this work has highlighted deep layer projection neurons in prefrontal cortex (PFC) during midfetal development as a locus and timepoint of risk-gene convergence. From the ASD risk-gene network, T-box, brain, 1 (*TBR1*) was found to be the most connected to other risk-genes in the PFC at this critical developmental stage (A. J. Willsey et al., 2013). *TBR1* is a transcription factor that is expressed in all newborn cortical glutamatergic neurons during development, but its expression is only maintained in deep-layer cortical pyramidal neurons into adulthood. In these neurons, *TBR1* plays a critical role in maintaining deep-layer cell identity (Bedogni et al., 2010; Fazel Darbandi et al., 2018), and has also been shown to bind directly to the promoters of many other ASD risk-genes (Fazel Darbandi et al., 2020; Notwell et al., 2016).

We recently published a study of mice with the conditional, postnatal deletion of *Tbr1* from cortical layer 5 (L5) neurons showing that these conditional knockout (cKO) mice spend less time than their wild-type littermates interacting with novel conspecifics (Fazel Darbandi et al., 2020). This study and others have characterized cellular and molecular abnormalities in these mice (Fazel Darbandi et al., 2018, 2020, 2022). One finding from this work is that both

anatomical and physiological measures of excitatory and inhibitory synaptic inputs to L5 neurons are reduced in *Tbr1* L5 cKO mice. Furthermore, spines on L5 neurons in adult cKO mice had a persistently immature, filamentous morphology. Importantly, these L5 cKO mice are generated using a Cre-driver line (Rbp4-Cre) which leads to deletion at murine post-natal day 0 (P0), which roughly corresponds to human midfetal development (Workman et al., 2013).

Again, despite this detailed information on cellular, molecular and behavioral phenotypes in these cKO mice, it remains unknown how this genetic disruption affects the patterns of neural circuit activity which mediate behavior. In fact, there is limited information about how the activity of neurons within the prefrontal cortex is altered during behavior in mouse models of autism. Previous work from our lab has shown that social behavior recruits increased activity across an excessive number of medial PFC (mPFC) neurons in *Shank3* knockout (KO) mice (Frost et al., 2021). This results in a reduction in the ability of coordinated multineuron activity to specifically encode social interaction. Other work in *Cntnap2* KO mice has found increased baseline activity but decreased recruitment of mPFC neurons by social odors (Levy et al., 2019). Together with more variable neuronal activity, this impairs the encoding of whether odor stimuli belong to social vs. nonsocial categories. Whether these types of abnormalities are present in other mice with genetic disruptions related to autism, e.g., *Tbr1* cKO mice, is unknown. It also remains to be examined whether these sorts of abnormalities in multineuron encoding are specific to social behavior, or would generalize to other prefrontal-dependent behaviors as well.

In this study, we sought to address these questions using microendoscopic calcium (Ca2+) imaging during behavior in *Tbr1* cKO mice. Notably, we also find that these mice have abnormalities in anxiety-related avoidance, which, like social interaction, depends on the mPFC (Adhikari et al., 2011). This allows us to directly test whether similar abnormalities in neural circuit activity occur during both behaviors. Furthermore, previous work has shown that treatment with a single dose of lithium rescues synaptic and social interaction deficits in *Tbr1*

cKO mice (Fazel Darbandi et al., 2020). Thus, we could examine whether this pharmacologic intervention also produces any long-lasting alterations of *in vivo* activity in these mice.

In contrast to previous work in *Shank3* KO and *Cntnap2* KO mice, we find that *Tbr1* cKO mice have decreased L5 mPFC activity across a range of behaviors. Despite this, social behavior is able to recruit normal numbers of mPFC neurons in *Tbr1* cKO mice. However, similar to *Shank3* KO mice, the coordination of these neurons into coactive ensembles which encode social behavior is disrupted. Interestingly, this defect in multineuron coordination appears to be specific to social behavior and was not observed in the context of abnormal anxiety-related avoidance. We do find other abnormalities in the encoding of anxiety-related avoidance behaviors. Finally, we find that a single treatment with lithium ameliorates the deficit in activity of L5 cKO neurons.

2.3   Materials and Methods

Animal Subjects

All animal care procedures and experiments were conducted in accordance with the National Institutes of Health guidelines and approved by the Administrative Panels on Laboratory Animal Care at the University of California, San Francisco. Mice were housed in a temperature-controlled environment (22–24 °C) with ad libitum access to food and water. Mice were reared in normal lighting conditions (12-h light/dark cycle). *Tbr1* layer 5 knockout mice were generated by crossing *Tbr1*$^{f/+}$ with *Tbr1*$^{f/+}$*::Rpb4-cre*$^{+/-}$*::Ai14*$^{+/+}$ mice (Fazel Darbandi et al., 2020). Experimental mice were either *Tbr1*$^{+/+}$*::Rpb4-cre*$^{+/-}$*::Ai14*$^{+/-}$ *(Tbr1* WT) or *Tbr1*$^{f/f}$*::Rpb4-cre*$^{+/-}$*::Ai14*$^{+/-}$ *(Tbr1* cKO).

Surgeries

All animals used for microendoscopic recordings underwent two surgeries: the first for viral injection and the second for GRIN lens implantation. For both surgeries, mice were

anaesthetized with isoflurane (3.5% induction, 1-2% maintenance in 95% oxygen, flow rate 0.9L/min) and secured with ear bars in a stereotaxic frame (David Kopf instruments). Analgesics were administered via subcutaneous injection at surgery onset (buprenorphine, 0.1 mg/kg; meloxicam, 2 mg/kg). Body temperature was maintained using a heating pad. The scalp was incised along the rostro-caudal axis to expose the dorsal surface of the skull, which was then aligned using bregma and lambda as references. After surgery, animals were allowed to recover on a heated pad until ambulatory. All animals received an additional dose of meloxicam on the day following surgery.

Viral injections were performed on mice 10-12 weeks old. A dental drill was used to create an opening in the skull above the right mPFC (-1.7mm AP, +0.4mm ML relative to bregma). AAV9-hSyn-FLEX-jGCaMP7f-WPRE (AddGene) was diluted 3:1 in sterile saline immediately before injection. Virus was injected at four different depths, starting with the most ventral, (-2.75mm, -2.5mm, -2.25mm, and -2.0mm DV; 150nl at each depth) at a rate of 100nl/min with a 35-gauge microinjection syringe (World Precision Instruments) connected to a pump (UMP3 UltraMicroPump, World Precision Instruments). The needle was kept at each injection site for 5mins after infusion. Scalp incisions were sutured.

GRIN lens implant surgeries were performed two weeks after viral injection. A dental drill was used to create an 1mm diameter opening in the skull above the right mPFC (-1.7mm AP, +0.55mm ML relative to bregma). The cortex was aspirated to an approximate depth of 2.0mm, and a 1 mm diameter x 4 mm long integrated GRIN lens (Inscopix) was slowly lowered to a final depth of 2.3mm. Implants were secured to the skull with Metabond (Parkell). After surgery, mice were singly housed to prevent damage to their implants. Mice were allowed 3 weeks to recover from surgery before behavioral testing and microendoscopic imaging.

Behavioral Assays

For all cohorts, mice were habituated to experimenter handling for 3-5 days (5-10min per day). Mice were transported to the testing room 45-60min prior to test onset. For calcium imaging cohorts, mice were also habituated to the head-mounted microscope before testing (3-5 days, 45min per day). Behavioral videos were recorded with a video camera, and real-time animal tracking was performed using AnyMaze software.

Social Behavior Assays

For the initial behavioral cohort (**Fig. 2.1A-B**), group-housed animals were transferred to a clean holding cage 1 hour prior to testing. Animals were then, one-by-one, transferred to their home cage for testing. Mice were presented with a novel juvenile conspecific (sex-matched, 4-6 weeks old) for 5min, and then with a novel object (a clean 15mL Falcon tube) for 5min. Interaction time with both targets was scored manually.

For calcium imaging cohorts, mice (singly housed) underwent a serial interaction assay in their home cage. Mice were attached to the head-mounted microscope and an initial 10-minute baseline recording was performed. Following the baseline, four 5-minute interaction epochs were interleaved with 3-minute rest periods (with no interaction target). The four interaction epochs were (in order): a novel same-sex juvenile mouse (Mouse 1), a second novel same-sex juvenile mouse (Mouse 2), a 15mL plastic conical tube (Novel Object), and the first novel juvenile (Mouse 1, now labeled 'Familiar'). All interactions were manually scored.

Elevated Plus Maze

For the initial behavioral cohort (**Fig. 2.1A-B**), animals were placed into the center of the maze and allowed to freely explore for 10min. For calcium imaging cohorts, two days after social testing, animals were attached to the head-mounted microscope and an initial 10-minute baseline recording was performed in the home cage. Mice were then placed into one of the

closed arms of the maze and allowed to freely explore for 15min. For all cohorts, AnyMaze tracking was used for subsequent behavioral analyses.

Lithium Chloride Treatment

For the calcium imaging cohorts, a subset of animals were retested on both the serial interaction assay and the EPM after receiving a single IP injection of lithium-chloride (5mg/kg). Animals underwent initial social testing on day 1, initial EPM testing on day 3, and received LiCl on day four. Retesting on these assays was conducted 30 days after LiCl administration.

Data acquisition and processing

Calcium signals were imaged at 20Hz with 4x spatial downsampling using a head-mounted, one-photon microendoscope (nVoke2; Inscopix Inc.). Behavioral videos were recorded using AnyMaze software, and a TTL cable connecting the AnyMaze computer to the Inscopix DAQ box was used to align calcium activity and behavior videos. Calcium activity videos were preprocessed in Inscopix Data Processing Software (IDPS). Videos were spatially band-pass filtered with cut-offs set to 0.015 pixel-1 (low) and 0.300 pixel-1 (high), and then motion corrected. Preprocessed videos were exported from IDPS as tiff files and imported to MatLab for cell segmentation. Putative cellular ROIs were identified using a combined PCA/ICA approach (Mukamel et al., 2009). For each movie, putative neurons were manually sorted to remove ROIs with low SNR or overlapping boundaries. Cell sorting was done with the assistance of the signalSorter GUI from an open-source calcium imaging software package (CIAPKG) (Corder et al., 2019).

Cellular dF/F traces for sorted neurons were extracted and then used to generate a binary calcium event raster for each neuron using a previously described event detection algorithm (Frost et al., 2021). All subsequent analyses were done on binary event rasters.

Data analysis

All analyses of calcium activity data were performed in MatLab using custom scripts.

Determining Mean Cell Activity

To calculate a cell's mean activity during a particular behavior, first, all calcium imaging frames corresponding to that behavior were identified. Mean activity was then computed as the fraction of these behaviorally associated frames with a detected calcium event.

Correlation Analysis

To assess the pairwise correlations between neurons during a particular behavior, frames corresponding to that behavior were identified and extracted from the full recording. Pairwise correlations were then calculated on this behavior-specific activity matrix using the MatLab corr function. Any correlation with an associated p-value below 0.05 was taken to be significant, and results were reported as the portion of pairwise correlations that were significant.

Neural Network Classifiers and Identification of Behavioral Ensembles

To identify behaviorally recruited neural ensembles, a neural network classifier (NNC) was utilized (previously detailed in Frost et al., 2021). NNCs were trained to distinguish two different behavioral periods (B1 and B2) based on binary calcium event rasters. Each NNC contained a hidden layer comprised of 1000 artificial units. Each hidden layer unit received input from a random subset of real neurons (connection probability was set to 0.3). The binary matrix C (dimensions: N real neurons x n hidden units) described connectivity from real neurons to the hidden layer. Hidden layer units were all connected to a single output unit with weights given by the vector w. The output weight vector w was updated during training.

To train a binary NNC, all calcium imaging frames corresponding to B1 and B2 were identified. Frames associated with B1 and B2 were assigned output values of 1 and 0,

respectively. For identification of ensembles, frames were not split into separate training/testing sets because the purpose was to train the weight vector w, not to assess performance. Training consisted of 500 passes through the training frames, using a different, random order on each pass. If B1 and B2 had different numbers of associated frames, the larger set was randomly subsampled to match the number of frames in the smaller set; each training pass used a new random subset. Each pass through the data was further capped at 1000 frames to make the number of training frames comparable across animals. For each animal we trained 20 NNCs for each pair of behaviors. After training, the effective weight associated with each real neuron was computed as **C*w**. The resulting effective weight vectors were then averaged across all 20 NNCs, resulting in one effective weight vector for each behavioral comparison for each animal.

Given the labels of 1 and 0 for B1 and B2 frames, respectively, positive effective weights were associated with B1 encoding neurons and negative weights with B2 encoding neurons. To identify neural ensembles for specific behaviors from the effective weights, a threshold of 1 standard deviation (SD) was used. Thus, the 'B1 ensemble' was defined as neurons with an effective weight at least 1 SD above the average value, and the 'B2 ensemble' was defined as neurons with an effective weight at least 1 SD below the mean.

For the serial interaction assay data, 3 types of NNC were trained per animal. In all cases, B1 was designated as the baseline home cage recording period, while B2 was either Mouse 1 interaction, Mouse 2 interaction, or Familiar interaction. Only frames corresponding to active social interaction were used.

For the EPM assay data, 2 types of NNC were trained per animal. In both cases, B1 was designated as the baseline home cage recording period, while B2 was either Closed Arm exploration or Open Arm exploration.

SHARC and Swap Shuffling

SHARC is a method for generating surrogate datasets from an original matrix of neural activity that nonrandomly shuffles blocks of activity such that the shuffled data conforms to a target correlation matrix (Luongo et al., 2016). Thus, the resultant surrogate datasets preserve the pairwise correlation structure of the target correlation matrix that is provided. A modification of this method that also preserves the overall activity of each neuron was used (Frost et al., 2021).

SHARC surrogate datasets were generated for a particular behavior by first creating a subraster consisting only of frames corresponding to that behavior. A pairwise correlation matrix was calculated from this behavioral subraster and used as the target correlation matrix for SHARC in order to preserve the native correlation structure.

Swap-shuffled surrogate datasets were also generated from behavioral subrasters. This method involves making random, reciprocal swaps of detected calcium events between neurons. Resulting swap-shuffled surrogate datasets preserve the overall activity of each neuron and the total network activity of each frame in the native dataset, but the native correlation structure is broken.

To assess the relative contribution of neuronal coactivity in encoding of behavior, 20 SHARC and 20 Swap-shuffled surrogate datasets were generated for each of the behaviors that NNCs were trained on (Social baseline, Mouse 1, Mouse 2, Familiar, EPM baseline, Closed Arm exploration, and Open Arm exploration). These NNCs trained on real data (20 for each behavioral comparison) were then tested on corresponding SHARC and swap-shuffled datasets. For example, the 20 NNCs trained on Social baseline and Mouse 1 data were each tested on one SHARC dataset composed of Baseline and Mouse 1 data, and one swap-shuffled dataset composed of Baseline and Mouse 1 data. Because there were always more Baseline frames than frames for the other behavior, the Baseline surrogate data was randomly

subsampled to provide an equal number of frames. For each NNC, a performance improvement score was calculated as $(P_{SHARC} - P_{Swap}) / (P_{Swap} - 0.5)$, where P_{SHARC} is performance on SHARC data, P_{Swap} is performance on swap-shuffled data, and 0.5 is the chance performance level. The 20 different improvement scores were averaged to yield a single value for each behavioral comparison in each animal.

Quantification and statistical analysis

All statistical calculations were performed in MatLab or Graphpad Prism. All details of these statistical analyses are provided in the main text and figure legends.

2.4   Results

Selective deletion of *Tbr1* from layer 5 cortical neurons causes abnormal social and anxiety-related behaviors

To determine whether *Tbr1* expression in cortical L5 pyramidal neurons is required for normal socioemotional behavior, we crossed the *Rbp4-Cre* mouse line with a line harboring a floxed *Tbr1* allele, as described previously (Fazel Darbandi et al., 2020). This results in *Tbr1* deletion within L5 excitatory neurons of the neocortex at postnatal day zero (P0), ~8 days after *Tbr1* is first expressed. We refer to these *Tbr1*$^{fl/fl}$::*Rbp4-Cre*$^{+/-}$::*Ai14*$^{+/-}$ mice as '*Tbr1* cKO' mice, or just 'cKO.'

Previously, we showed that *Tbr1* cKO mice exhibit deficient social interaction as measured by a home-cage resident-intruder assay with a novel, same-sex juvenile conspecific (Fazel Darbandi et al., 2020). To determine if these mice display additional behavioral deficits, cKO mice and WT littermates (*Tbr1*$^{+/+}$::*Rbp4-Cre*$^{+/-}$::*Ai14*$^{+/-}$) underwent a battery of behavioral tests. As seen previously, cKO mice displayed abnormal social behavior as measured by interaction time with a novel same-sex juvenile mouse in the home cage (WT: 131 ± 28 sec,

cKO: 70 ± 24s, p = 0.000020, Welch's Unpaired T-test) (**Fig. 2.1A**). By contrast, there was no difference in the time WT vs. cKO mice spent exploring a novel object (data not shown).

Intriguingly, *Tbr1* cKO mice also had a deficit in anxiety-related avoidance as measured using the elevated plus maze (EPM). cKO mice spent significantly more time exploring the open arms (181 ± 90 sec in cKO vs. 106 ± 47 sec for WT, p = 0.027, 2-way ANOVA with Sidaks correction) and significantly less time in the closed arms (303 ± 79 sec in cKO vs. 381 ± 56 sec for WT, p = 0.020, 2-way ANOVA with Sidak's correction) (**Fig. 2.1B**).

In the open field test (OFT), there were no significant differences in the amount of time spent in the center or the perimeter. However, cKO mice had significantly more center entries and significantly higher distance traveled within the center (data not shown). While the EPM and OFT are not completely interchangeable assays, this increase in entries into and distance traveled within the OFT center aligns with the decreased open arm avoidance in the EPM, supporting the interpretation that *Tbr1* cKO mice avoid potentially anxiogenic regions less than their WT counterparts.

Activity of *Tbr1* cKO neurons is significantly reduced during behavior

To better understand how the loss of *Tbr1* affects neuronal circuits underlying social and anxiety-related behaviors, we used miniaturized, head-mounted microscopes to record Ca2+ signals from L5 mPFC neurons in both cKO mice and their WT littermates during freely moving behavior. Mice were injected at 11-13 weeks of age with a Cre-dependent virus encoding the fluorescent calcium indicator GCaMP7f (AAV9-hSyn-FLEX-GCaMP7f-WPRE). Thus, we recorded signals specifically from L5 neurons that were labeled by the Rbp4-Cre line, (i.e., the neurons from which *Tbr1* had been deleted) (**Fig. 2.1C**). Mice were implanted with a 1mm gradient refractive index (GRIN) lens 2-3 weeks post-injection, then were allowed to recover for 3 weeks before habituation to the microscope and behavioral testing.

Mice were first tested on a home cage social interaction assay, which consisted of an initial 10-minute baseline recording period followed by four 5-minute interaction epochs interleaved with 3-minute rest periods. The four interaction epochs were (in order): a novel same-sex juvenile mouse (Mouse 1), a second novel same-sex juvenile mouse (Mouse 2), a 15mL plastic conical tube (Novel Object), and the first novel juvenile (Mouse 1, now labeled 'Familiar') (**Fig. 2.1D**). Two days later, mice had a 10-minute baseline recording in their home cage, followed by 15 minutes of testing in the EPM (**Fig. 2.1E**).

Calcium signals were imaged at 20Hz using the Inscopix nVoke system. Following video preprocessing, putative cellular ROIs were identified using a combined PCA/ICA approach and manually sorted (Mukamel et al., 2009). Cellular dF/F traces were extracted and a custom event detection algorithm was used to obtain a binary calcium-event raster for each neuron (Frost et al., 2021) (Methods). All subsequent analyses were performed using these binary event rasters.

First, to determine how the loss of *Tbr1* affected the activity of L5 neurons, we calculated the mean activity of each neuron for relevant timepoints within both the serial interaction assay and the EPM. Mean activity was calculated for each neuron as the fraction of frames with a detected calcium event out of all frames corresponding to a given behavior. For the interaction assay, only frames corresponding to "active interaction" were used. "Active interaction" included all instances in which the test mouse was within ~1cm of, and oriented toward, the target (either a juvenile mouse or novel object). Despite similar activity of WT and cKO neurons in the initial baseline period (WT: 0.0399 +/- 0.0015, cKO: 0.0422 +/- 0.0020; data not shown), during the Mouse 1 (WT: 0.0627 +/- 0.0022, cKO: 0.0557 +/- 0.0034), Mouse 2 (WT: 0.0461 +/- 0.0022, cKO: 0.0372 +/- 0.0027), Novel Object (WT: 0.0570 +/- 0.0045, cKO: 0.02752 +/- 0.004069), and Familiar (WT: 0.04426 +/- 0.002409, cKO: 0.03879 +/- 0.003111) interaction epochs, cKO neurons were significantly less active than their WT counterparts (**Fig. 2.1D**). Similarly, in the EPM, cKO neurons were significantly less active in every zone (closed arms, WT: 0.0503 +/-

0.0025, cKO: 0.0237 +/- 0.0022; open arms, WT: 0.0442 +/- 0.0025, cKO: 0.0185 +/- 0.0024; center, WT: 0.0451 +/- 0.0024, cKO: 0.0210 +/- 0.0021) (**Fig. 2.1E**). cKO neurons were also significantly less active in the 10 minute baseline recording immediately prior to the EPM (WT: 0.0520 +/- 0.0018, cKO: 0.0236 +/- 0.0018), despite no baseline deficit two days prior.

A neural network classifier approach for identifying behaviorally recruited neuronal ensembles

To determine whether this reduction in L5 neuronal activity affected the encoding of behavior, we used a neural network classifier (NNC) to identify behaviorally relevant neuronal ensembles (**Fig. 2.2A**). Previous work from our lab showed that this approach can identify both individual neurons encoding behavior via changes in their activity levels and groups of neurons which increase or decrease their correlated activity (coactivity) during specific behaviors (Frost et al., 2021).

For each animal, NNCs were trained to distinguish two different behavioral periods (B1 and B2) based on calcium event data. For example, B1 could correspond to frames during the baseline home cage period (when the mouse is alone), whereas B2 could be frames corresponding to periods of active social interaction. Each NNC contained a hidden layer comprised of 1000 artificial units, each of which received input from a random subset of real neurons based on a connection probability of 0.3. The binary matrix **C** (dimensions: N real neurons x n hidden units) described connectivity from real neurons to the hidden layer. Hidden layer units were all connected to a single output unit. The vector **w**, containing the output connection weights, was updated during training. After training, the effective weight associated with each real neuron was computed as **C*w**.

During training, frames associated with B1 and B2 were assigned output values of 1 and 0, respectively. This means neurons encoding B1 and B2 correspond to those neurons with the

largest (most positive) and smallest (most negative) effective weights, respectively. To identify neural ensembles encoding specific behaviors, we used a threshold of 1 standard deviation (SD). I.e., the 'B1 ensemble' was defined as neurons with an effective weight at least 1 SD above the average value, and the 'B2 ensemble' was neurons with an effective weight at least 1 SD below the average value.

The major advantage of defining neuronal ensembles in this manner, versus simply identifying which neurons are most active during a particular behavior, is that NNCs can also identify neurons that encode behavior via changes in their correlations or coactivity with other neurons. By contrast, defining ensembles based simply on the activity of each neuron would overlook the potential for this type of inter-neuronal synergy or combinatorial coding. Our previous work showed that optimal linear decoders based on logistic regression or support vector machines are also insensitive to coactivity-based encoding (Frost et al., 2021).

Ensembles encoding different conspecifics become abnormally differentiated in *Tbr1* cKO mice

To test if *Tbr1* cKO mice have altered social encoding, three types of NNC were trained for each animal, yielding social ensembles corresponding to each type of social interaction: 'M1' (interaction with Mouse 1 when it is novel), 'Fam' (interaction with Mouse 1 when it is familiar), and 'M2' (interaction with Mouse 2). In each case, the NNC was trained using data from the baseline recording period (HC1) (Behavior 1) and the relevant social interaction period (Behavior 2) (**Fig. 2.2B**).

Overall, the portion of neurons identified as being part of a social ensemble was similar across mice and genotypes. 'M1', 'M2', and 'Fam' neurons were pooled by genotype (n = 84 social cells for WT; n = 85 social cells for cKO) and the overlap between the different ensembles was assessed (**Fig. 2.2C**). In both WT and cKO mice, the Fam and M2 ensembles were

significantly non-overlapping, i.e., the number of neurons in both ensembles was less than would be expected by chance (WT: 24/84 in M2 only, 29/84 in Fam only, 19/84 in both, 12/84 in neither, p = 0.0166; cKO: 30/85 in M2 only, 24/85 in Fam only, 12/85 in both, 19/85 in neither, p = 0.0098, one-tailed Fisher's exact test). Interestingly, in cKO mice, the M1 and M2 ensembles were significantly non-overlapping, whereas in WT mice these ensembles overlapped as expected by chance (cKO: 32/85 in M2 only, 30/85 in M1 only, 10/85 in both, 13/85 in neither, p = <0.0001; WT: 28/84 in M2 only, 20/84 in M1 only, 15/84 in both, 21/84 in neither, p = 0.2686; one-tailed Fisher's exact test) (**Fig. 2.2C**). This suggests that even though Mouse 1 and Mouse 2 interactions have something in common – both involve a novel conspecific – their encoding becomes abnormally distinct in cKO mice.

We can further confirm this interpretation by examining NNC effective weight vectors. Each element of an effective weight vector represents the degree to which the corresponding neuron encodes one behavior vs. the other, and there is a different effective weight vector for each classifier. Thus, we can take an ensemble identified from one classifier (i.e., the M2 ensemble from the Mouse 2 classifier) and look at the effective weights of those neurons in other classifiers to assess any similarities in encoding across behaviors. Doing this for M2 ensemble neurons, we see that their effective weights in the Mouse 1 classifier are shifted toward negative values in WT mice, indicating that the M2 ensemble also tends to encode Mouse 1 interaction (**Fig. 2.2D**). However, in cKO mice, effective weights of M2 neurons in the Mouse 1 classifier are significantly less negative and instead center around zero, implying that M2 neurons in cKO animals do not strongly encode Mouse 1 interaction. Thus, social encoding becomes more unstable across novel conspecifics in *Tbr1* cKO mice compared to WT animals.

To more generally examine overall activity levels within each neuronal ensemble and across behaviors, we plotted the average activity of neurons in the M1, M2 or Fam ensembles during Mouse 1, Mouse 2, and Familiar interactions (**Fig. 2.2E**). Notably, despite the marked

reduction in overall activity in cKO mice, activity levels of social ensemble neurons during social behavior were very similar in WT and cKO mice. The modulation of this activity during different types of social behavior was also similar across genotypes. For example, in both WT and cKO mice, the M1 ensemble was more active during Mouse 1 interaction than during Mouse 2 or Familiar interactions, and the Fam ensemble was more active during Familiar interaction than during Mouse 2.

Thus, focusing on neuronal ensembles reveals that even though overall population activity during behavior is reduced in cKO mice, activity levels within socially-encoding ensembles are grossly intact. Furthermore, using NNCs to identify these ensembles reveals that even though their activity levels are intact, their composition become more unstable, or more variable, in *Tbr1* cKO mice.

The encoding of social behavior by correlated activity is diminished in *Tbr1* cKO mice

We previously showed that pairwise neuronal correlations, or neuronal coactivity, in the mPFC encodes information about social behavior, over and above the information transmitted by neuronal activity levels (Frost et al., 2021). Moreover, this contribution of prefrontal correlations was diminished an autism model: *Shank3* KO mice. To assess whether the contribution of correlated activity to social encoding is similarly disrupted in *Tbr1* cKO mice, we measured the ability of NNCs trained on our data to classify social behavior in surrogate datasets that had been shuffled to either preserve or disrupt pairwise correlations between neurons (**Fig. 2.3A**). Specifically, we generated surrogate datasets using either the SHuffling Activity to Rearrange Correlations (SHARC) (Luongo et al., 2016) method so that the shuffled data conformed to a target correlation matrix (in this case preserving the original correlations), or randomly swapped activity between neurons to preserve activity levels across neurons and timepoints while disrupting correlations (Methods).

For each mouse, we identified activity rasters associated with the baseline period or each type of social interaction (Mouse 1, Mouse 2, and Familiar). Then we generated 20 SHARC and 20 Swap-shuffled surrogate datasets for each of these rasters. The NNCs described previously (**Fig. 2.2B**), which were trained on the original unshuffled data, were then tested using the corresponding SHARC and Swap surrogate datasets (i.e., the Mouse 1 NNC was tested on shuffled baseline and Mouse 1 datasets). To quantify the degree to which correlations encode information beyond the information encoded by activity levels alone, we calculated: $(P_{SHARC} - P_{Swap}) / (P_{Swap} - 0.5)$, where P_{SHARC} = NNC performance on SHARC shuffled data, P_{Swap} = NNC performance on swap-shuffled data, and 0.5 is chance performance level.

Across all three types of social interaction, the improvement in classifier performance achieved when using SHARC shuffling to preserve correlations was higher for WT than *Tbr1* cKO mice (average normalized improvement = 1.05 +/- 0.13 in WT vs. 0.28 +/- 0.09 in cKO; significant effect of genotype, p = 0.018 by 2-way ANOVA) (**Fig. 2.3B**). To more directly relate this finding, derived from measuring classifier performance using surrogate datasets, to the social ensembles we had previously identified in real data, we examined intra-ensemble correlations. For each pair of neurons within an ensemble, we assessed whether activity of that pair was significantly correlated during periods of social interaction. We found that the proportion of significant intra-ensemble pairwise correlations (PWCs) was consistently higher in WT compared to *Tbr1* cKO mice (M1 ensembles: 115/140 significant in WT vs. 92/143 significant in cKO, p = 0.0008 by χ2 test; M2 ensembles: 178/230 significant in WT vs. 114/186 significant in cKO, p = 0.0005 by χ2 test; Fam ensembles: 196/249 significant in WT vs. 94/149 significant in cKO, p = 0.0011 by χ2 test) (**Fig. 2.3C**). Thus, the normal role of prefrontal coactivity in encoding social interactions is diminished in cKO mice.

Prefrontal ensembles predict approach-avoidance decisions in WT but not cKO mice

Having characterized social encoding abnormalities in *Tbr1* cKO mice, we next turned our attention to encoding of anxiety-related behaviors within the EPM. Again, we used NNCs to derive behavioral ensembles. For each mouse, two NNCs were trained to distinguish periods of either closed or open arm exploration, from the baseline home cage period; the resulting two NNCs were then used to derive closed (CL) and open (OP) arm ensembles, respectively (**Fig. 2.4A**). Similar numbers of neurons were identified as belonging to a CL or OP ensemble across genotypes. Despite this, the magnitude of the effective weights of OP ensemble neurons in the OP Classifier was significantly lower in cKO mice than in WT (-19.89 +/- 2.57 for WT, -5.66 +/- 1.05 for cKO, Dunn's post-hoc test, p = 0.0003), indicating that encoding of the open arms in cKO mice is weaker (**Fig. 2.4B**).

Many previous studies have used multineuron recordings to show that different sets of neurons, most notably in the mPFC and ventral hippocampus, increase their activity in the open or closed arms (Adhikari et al., 2011; Lee et al., 2019). However, it has been much more difficult to transform these multineuron patterns of activity into signals that predict future behavior, e.g., decisions to enter vs. avoid the open arms. Our previous study on prefrontal vasoactive intestinal polypeptide (VIP)-expressing interneurons was able to do this, but by using bulk activity from a specific interneuron subtype rather than multineuron patterns of activity (Lee et al., 2019). One reason it has been challenging to identify predictive signals from multineuron recordings is that the number of open arm entries in a given mouse is relatively low, limiting the available data and making it impractical to, for example, train a classifier on approach vs. avoidance trajectories. We hypothesized that open and closed arm ensembles, derived from NNCs, might represent a useful tool for overcoming this challenge. Specifically, instead of training a classifier to identify ensembles specifically associated with approach vs. avoidance trajectories, we hypothesized that open and closed arm ensembles might themselves anticipate

upcoming decisions. To test this hypothesis, we examined the activity of CL and OP ensembles time-locked to the moment of EPM center entry. For each animal, all periods of center exploration lasting one second or longer were identified and then grouped based on the EPM arm (closed or open) the animal came from and subsequently entered. This yielded four run types: closed-closed, closed-open, open-open, and open-closed.

Normalized ensemble activity traces were calculated for each CL and OP ensemble as follows: first, the binary calcium event rasters for all ensemble neurons were summed, resulting in a time series representing how many ensemble neurons were active in each frame. Next, each time series was normalized to its maximum value over the entire EPM recording. Finally, 4-second windows around each entry to the center zone were used to average together all time-locked segments of these time series corresponding to each run type (closed-closed, closed-open, open-closed, open-open), first within each mouse and then across all mice of the same genotype. This yielded 16 averaged ensemble activity traces, each time-locked to the time of center entry (2 ensemble types x 4 run types x 2 genotypes).

Inspecting these ensemble activity traces revealed a relationship between ensemble activity and approach-avoidance decisions in WT mice. During the 2s preceding center entry, CL ensemble activity was higher than OP ensemble activity, specifically for closed-closed runs but not for closed-open runs. In other words, stronger activity in the CL ensemble, relative to the OP ensemble, tended to predict open arm avoidance. However, this relationship was no longer evident in cKO mice. To more clearly illustrate this, we plotted the differential ensemble activity for both of these run types (**Fig. 2.4C**). For this, the normalized OP activity vector was subtracted from the normalized CL activity vector for the entire EPM recording to get the CL-OP differential activity vector for each animal. Then, the mean differential activity was calculated for each type of run for a 4-second window around EPM center entry. To assess the statistical validity of these observations, we performed a multi-factor ANOVA on mean CL-OP differential

activity during the 2s preceding center entry for each genotype, using the run-type and animal

ID as factors. Indeed, this revealed a significant effect of run-type for WT, but not *Tbr1* cKO mice

(WT: Run Type main effect, p = 0.0069; Mouse ID main effect, p = 0.064; Run Type x Mouse ID

interaction, p = 0.23; cKO: Run Type main effect, p = 0.91; Mouse ID main effect, p = 0.17; Run

Type x Mouse ID interaction, p = 0.92).

The Encoding of EPM behavior by correlated L5 activity is not altered in Tbr1 cKO mice

Given our findings during social behavior, we wondered if correlated activity of mPFC L5

neurons is similarly perturbed in cKO mice during EPM exploration. As described previously for

social behavior, NNCs trained on real data recorded during EPM exploration were tested on

both SHARC and Swap shuffled datasets. In this case, there was no significant difference

between WT and cKO mice for the degree to which classifier performance was improved using

SHARC-shuffled instead of swap-shuffled data (**Fig. 2.4D**). There was also no difference in the

proportion of significant intra-ensemble PWCs across genotypes for either the CL or OP

ensemble (**Fig. 2.4E**).

Activity levels are no longer lower in Tbr1 cKO mPFC neurons after lithium treatment

A previous study of ours in *Tbr1* cKO mice found that a single dose of lithium chloride (LiCl)

could reverse both synaptic and social behavior abnormalities in these animals (Fazel Darbandi

et al., 2020, 2022). This raises the question of whether LiCl treatment would also normalize the

encoding of socioemotional behaviors. To address this, we injected a subset of our mice with a

single dose of LiCl (5 mg/kg) one day after initial EPM testing, then retested these animals 30

days later on both the serial interaction assay and the EPM (**Fig. 2.5A**). Of note, because the

number of these animals was very limited, we were not able to include an additional cohort of

mice which receive vehicle only. Thus, this experiment could only examine whether the

abnormalities in neuronal activity patterns we initially observed are still present in cKO mice

even after LiCl treatment, which is known to normalize abnormalities in social behavior and synapse number.

Whereas we had found a deficit in activity of *Tbr1* cKO neurons during all epochs of the serial interaction assay and in all zones of the EPM before LiCl treatment (**Fig. 2.1D-E**), after treatment, deficits were only seen during the M2 and Familiar interactions (**Fig. 2.5B**). Moreover, *Tbr1* cKO neurons were actually more active than their WT counterparts in all zones of the EPM, and during the Baseline period immediately prior to the EPM (**Fig. 2.5C**).

2.5   Discussion

While many genes have been identified in relation to the risk of autism spectrum disorders (ASD) and mouse models based on disruption of these genes have been shown to display abnormal socioemotional behaviors, there has remained a gap in understanding of how these genetic disruptions alter the normal neuronal activity that supports these behaviors. In this study, we sought to address this by recording single-neuron activity in the mPFC of an ASD model, Tbr1-cKO mice.

We find that these mice display reduced social interaction with a novel conspecific, as well as decreased anxiety-related avoidance in the elevated plus maze (EPM). During these behaviors, mPFC layer 5 Tbr1-cKO neurons display a marked reduction in activity. In the context of social interaction, behavioral encoding by neuronal coactivity is diminished, and correlations between social ensemble neurons are disrupted in cKO animals.

These results confirm a previous finding from out lab that correlated activity of neurons in the mPFC is important for social encoding, and that this information provided by correlations is lost in the Shank3 mouse model of ASD (Frost et al., 2021). Interestingly, mPFC neurons in the Shank3 mice of this previous study were more active than their WT counterparts, while we find a decrease in activity in our Tbr1-cKO mouse model. Still, these divergent effects on activity

lead to the same circuit-level deficit in correlations, indicating that this might be a convergent mechanism of disrupted behavior across ASD genetic models. One unifying framework for understanding the various genetic disruptions that can lead to ASD is the excitation-inhibition (or E/I) balance within the mPFC local circuitry (Sohal & Rubenstein, 2019). The balance of excitation and inhibition within a local microcircuit is crucial for information encoding and processing, as inhibition will tend to reduce noise, potentially by coordinating the activity of local excitatory neurons. Previous work has demonstrated that mPFC E/I balance is crucial for normal social behavior (Yizhar et al., 2011), and that restoring normal E/I balance in the *Cntnap2* mouse model of ASD also restores normal social behavior (Selimbeyoglu et al., 2017). Further, another study established that this CNTNAP2 mouse model had reduced pairwise correlations between mPFC neurons, although this finding was not specifically in the context of social interaction (Levy et al., 2019). Altogether, these findings point to ASD risk-genes effecting normal mPFC E/I balance, which in turn disrupts local correlated activity that normally supports ongoing social behavior.

We also identified distinct, but overlapping, ensembles of neurons in the mPFC that encode different types of social interactions. Specifically, the social interaction paradigm we utilized involved interactions with two different novel conspecifics, along with a repeat presentation of one of these conspecifics as a "familiar" interaction. This design allowed us to evaluate the consistency of mPFC representations across different social interactions. We found that ensembles encoding the two different novel interactions were significantly non-overlapping in the cKO animals, but not in WT animals, implying that representations of different conspecifics becomes abnormally distinct in the cKO. Further, we find that neurons encoding the familiar mouse (which was a second exposure to Mouse 1), are similarly activated during Mouse 1 and Familiar interactions, but significantly less active during Mouse 2 interactions for

both genotypes. This strongly implies that neurons of the mPFC have responses tuned to conspecific identity, and that this encoding is not necessarily disrupted by the loss of *Tbr1*.

Interestingly, in the context of the EPM, we find that behavioral encoding by coactivity is unaffected in cKO animals, potentially pointing to an inherent difference in the way that social and anxiety-related behaviors are encoded in the mPFC. However, we do find that overall encoding of the anxiogenic, open arms of the EPM is weakened, which may explain why these mice tend to explore the open arms more than WT mice. Further, the activity of open and closed arm ensembles was predictive of approach-avoidance decisions in the center of the EPM in WT animals, but not in cKO mice. Interestingly, a recent study identified two populations of ventral hippocampus (vHPC) to mPFC projection neurons that influence approach-avoidance decisions, and which may thus underlie the predictive activity that we report (Sánchez-Bellot et al., 2022). Given that previous studies of these Tbr1-cKO animals demonstrated a loss of synaptic input to L5 cKO neurons (Fazel Darbandi et al., 2020), the loss of predictive encoding in the cKO animals may reflect diminished vHPC input to mPFC. Interestingly, previous work from our laboratory found disruptions in vHPC-mPFC communication associated with decreases in avoidance of the EPM open arms in another mouse model of ASD (*Pogz*$^{+/-}$ mice) (Cunniff et al., 2020). Further work could test this idea that diminished vHPC input to mPFC contributes to the loss of this predictive signal within mPFC.

Finally, we tested the ability of lithium treatment to reverse these deficits in cKO activity and encoding. Previously, a single dose of lithium was shown to reverse synaptic and behavioral abnormalities in these Tbr1 cKO animals (Fazel Darbandi et al., 2020). While our lithium treatment cohort was underpowered to probe effects on ensembles and correlations, this treatment does appear to reverse the deficient activity of mPFC cKO neurons during behavior. While a larger cohort of mice including saline-treated control animals would be needed to confirm these findings and delve into the effects of lithium treatment on patterns of activity in the

mPFC, this does provide promising preliminary evidence that this treatment may be able to reverse deficits imparted by ASD-related genetic alterations.

Ultimately, these findings are specific to layer 5 neurons, and it remains to be studied how Tbr1 loss from L5 alters activity and encoding throughout the mPFC. Many excitatory neurons of the mPFC form local synapses within the microcircuit, including the excitatory L5 neurons studied here (Anastasiades & Carter, 2021). Whether loss of Tbr1 from L5 leads to wider deficits in behavioral encoding within the mPFC requires further investigation. Similarly, our findings here about social ensembles and EPM related signals within L5 of WT mice should also be assessed in other mPFC populations, or in the mPFC as a whole.
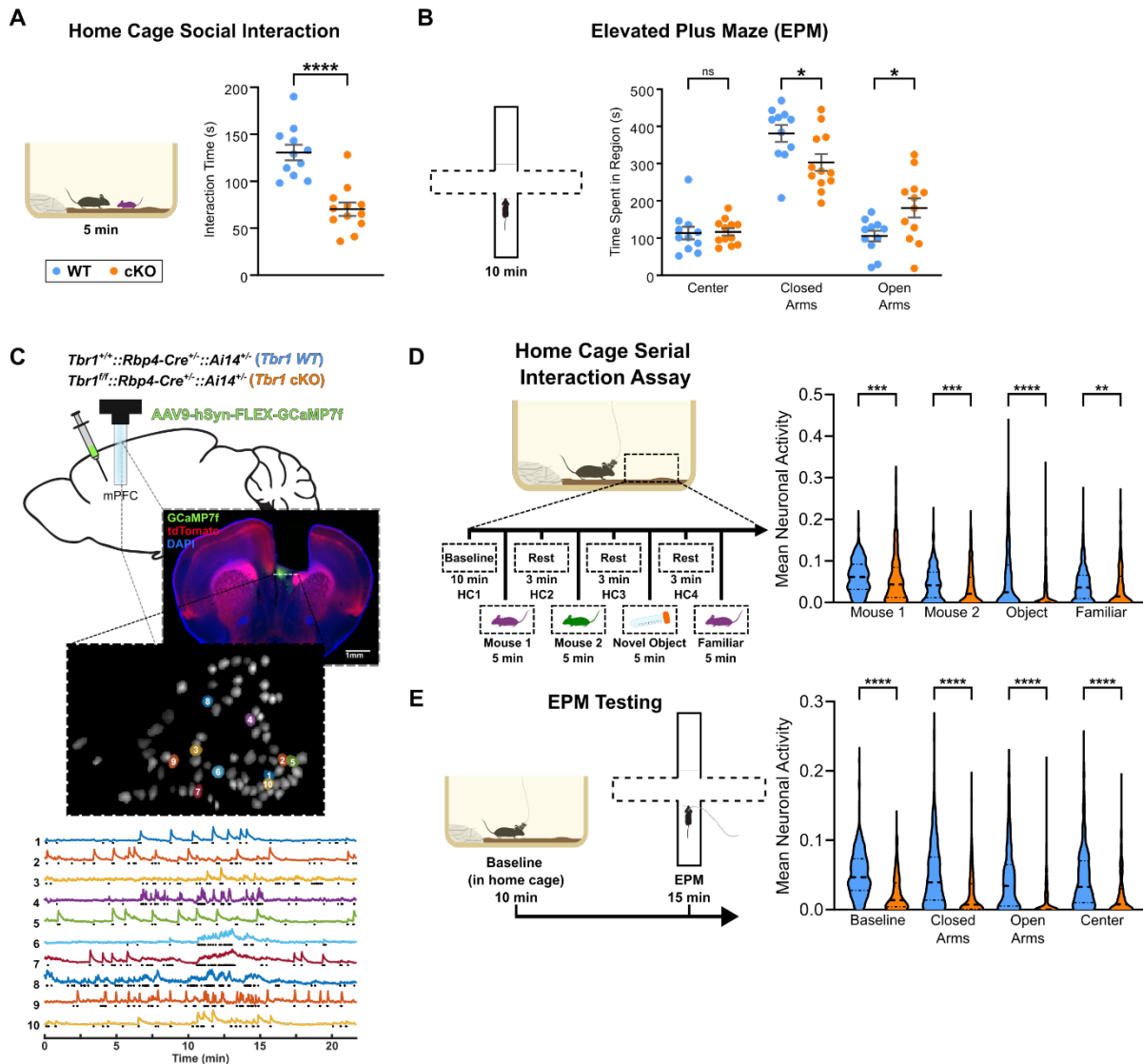
2.6 Figures

Figure 2.1. Tbr1 cKO mice exhibit abnormal socioemotional behavior and reduced activity of L5 mPFC neurons

(A) Interaction time with a novel juvenile conspecific during a 5-minute assay in the home cage. (n= 23 mice: 11 WT, 12 Tbr1 cKO)

(B) Zone time during a 10-minute EPM test. (n= 23 mice: 11 WT, 12 Tbr1 cKO)

(C) Experimental approach for recording mPFC L5 neurons during behavior.

(D) Activity of mPFC L5 neurons during the serial interaction assay. *Left*, timeline of home cage serial interaction assay. *Right,* violin plots of activity of all recorded neurons (n=306 WT, n=278 cKO). Mean activity of each neuron was calculated as the fraction of frames corresponding to active social exploration that had a detected calcium event. P-values shown are from a Kruskal-Wallis test between genotypes.

(Figure caption continued on next page)

(Figure caption continued from previous page)

(E) Activity of mPFC L5 neurons during the EPM assay. *Left*, timeline EPM assay. *Right,* violin plots of activity of all recorded neurons (n=325 WT, n=197 cKO). Mean activity of each neuron was calculated as the fraction of frames corresponding to exploration of each EPM zone that had a detected calcium event. P-values shown are from a Kruskal-Wallis test between genotypes.
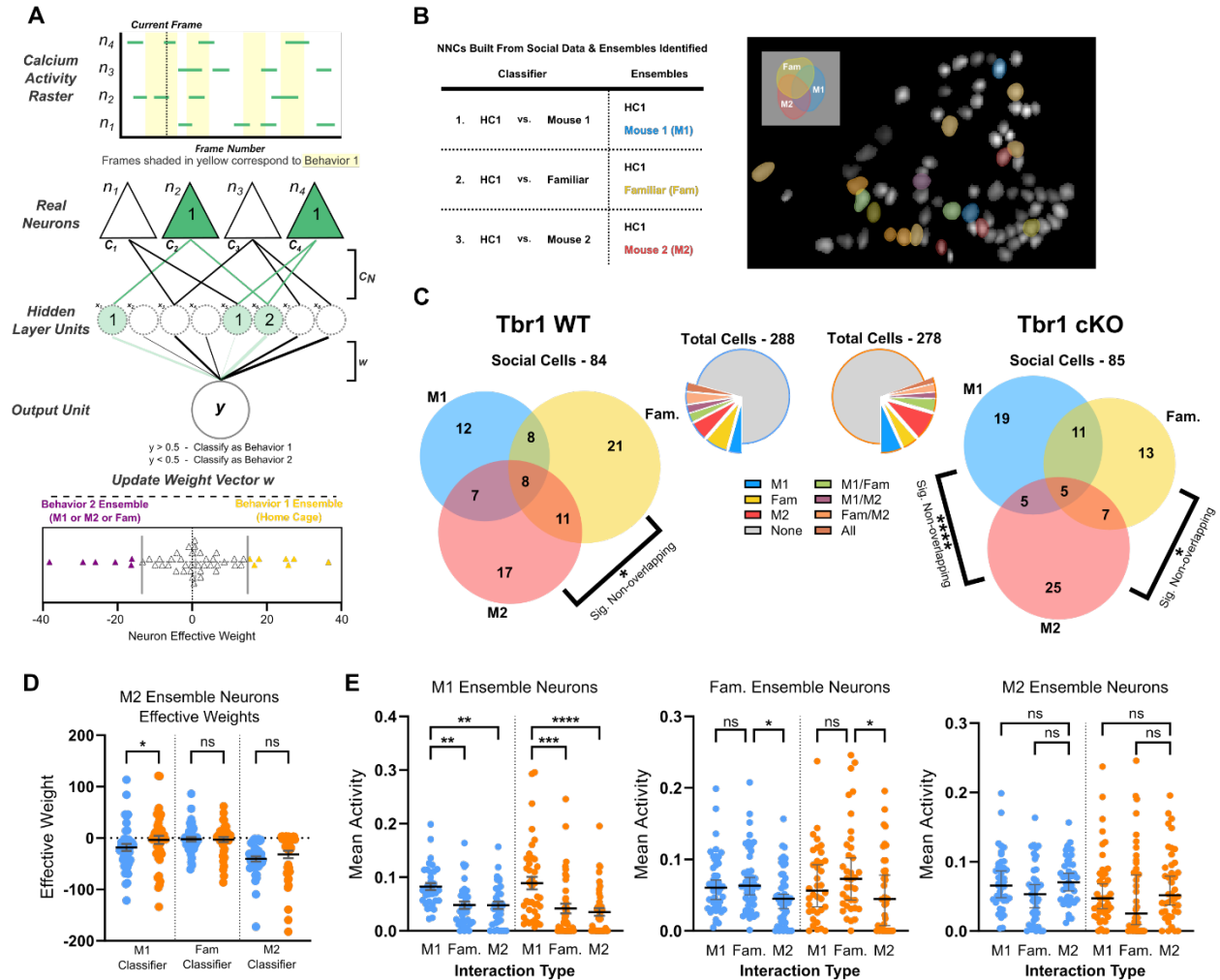
Figure 2.2. Neural ensembles recruited during social behavior are more stable across different interactions in WT than in cKO mice

(A) Overview of NNC architecture and identification of behavioral ensembles from trained NNC weights.

(B) *Left,* summary of NNCs trained for each mouse and ensembles identified from each. *Right,* example FOV from one mouse with the three different social ensembles highlighted.

(C) Summary of identified social cells. *Middle*, pie charts indicating the portion of recorded cells identified as socially encoding for WT (left, blue border) and cKO (right, orange border) mice (n=10 mice: 5 WT, 5 cKO). *Left,* Venn diagram displaying overlaps of WT social ensembles. *Right,* Venn diagram displaying overlaps of cKO social ensembles. In both groups, we found less-than-chance overlap of M2 and Fam neurons. In cKO animals, there was also less-than-chance overlap of M1 and M2 neurons. Significance of ensemble overlaps determined by one-tailed Fisher's Exact tests performed separately for each genotype and pair of ensembles.

(D) Effective weights of M2 neurons (n = 43 for WT, n=42 for cKO) in each of the three trained NNCs from (B). M2 neuron weights in the M1 classifier are significantly lower in WT than cKO mice. P-values shown are from Kruskal-Wallis tests with Dunn's multiple comparison correction between genotypes.
(Figure caption continued on next page)

(Figure caption continued from previous page)

(E) Activity of social ensemble neurons during each type of interaction. Activity of each social ensemble was modulated similarly across interactions in both genotypes. *Left,* M1 neurons (n = 35 for WT, n = 40 for cKO). *Middle*, Fam. neurons (n = 48 for WT, n = 36 for cKO). *Right*, M2 neurons (n = 43 for WT, n=42 for cKO). P-values shown are from Kruskal-Wallis tests with Dunn's multiple comparison correction within genotypes.
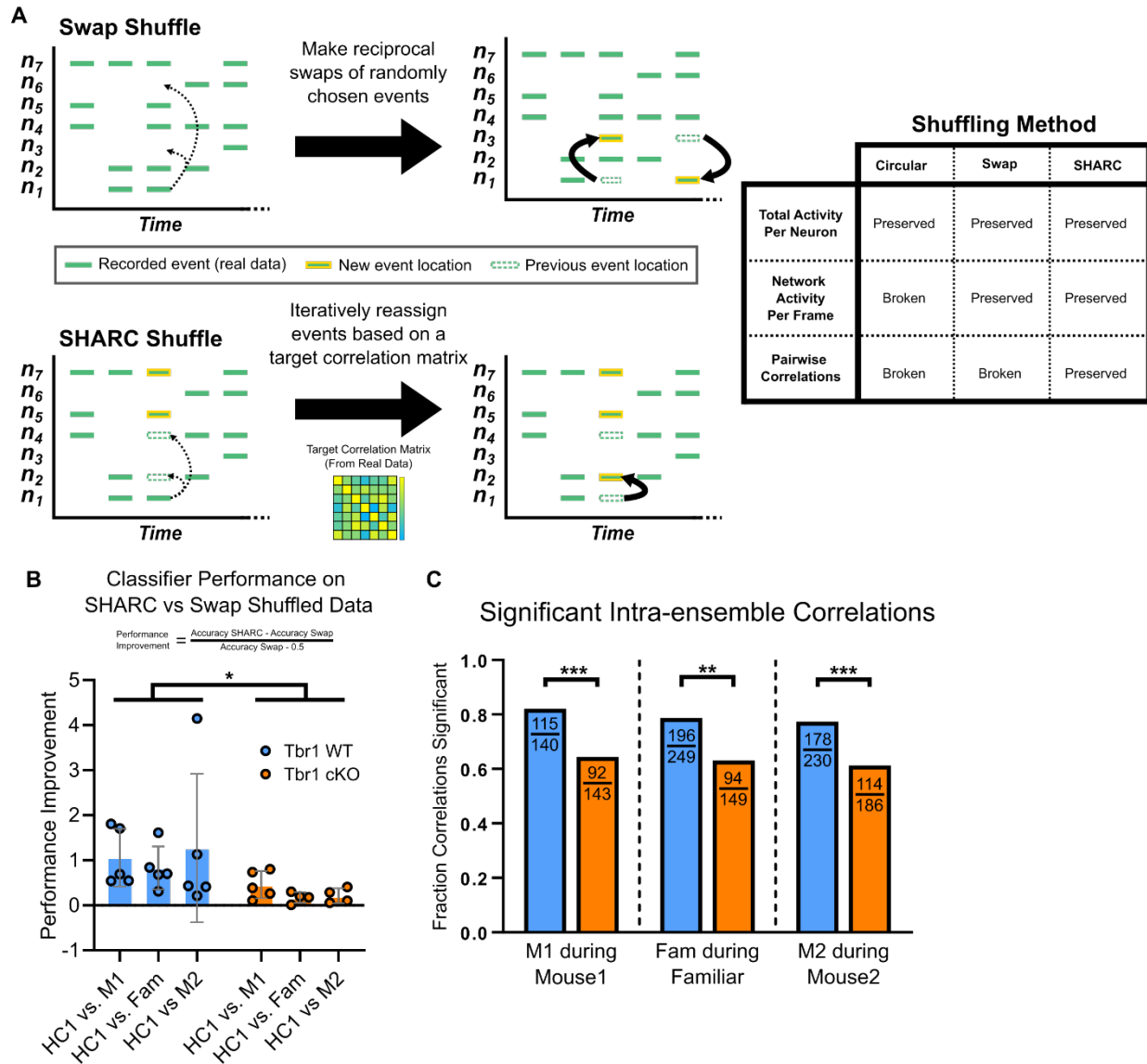
Figure 2.3. Correlated activity of neurons carries more social information in WT mice than in cKO mice

(A) Overview of shuffling methods for binary calcium activity rasters. *Left, top*, schematic of Swap Shuffle method. Reciprocal swaps of calcium events are randomly made, preserving the total number of events in each neuron and in each frame. *Left, bottom,* schematic of SHARC method. Calcium events are iteratively reassigned to other neurons based on a target pairwise correlation matrix, preserving total activity in each neuron and each frame, and preserving the correlation structure of the original data. *Right*, summary table of shuffling methods.

(B) Performance of classifiers trained on real data and tested on shuffled data. Performance of each trained classifier was assessed on Swap Shuffled and SHARC datasets, and a Performance Improvement score was calculated to assess the contribution of correlated activity in encoding social interaction. A 2-way ANOVA with Sidaks correction of Performance Improvement scores revealed a significant effect of genotype (p = 0.0183). *(Figure caption continued on next page)*

(C) The fraction of significant pairwise correlations within each social ensemble during its corresponding interaction epoch is plotted. Pairwise correlations and their p-values were calculated using the Matlab *corr* function. Correlations with a p-value <0.05 were taken as significant. P-values shown are from Fisher's Exact tests performed across genotypes for each ensemble.
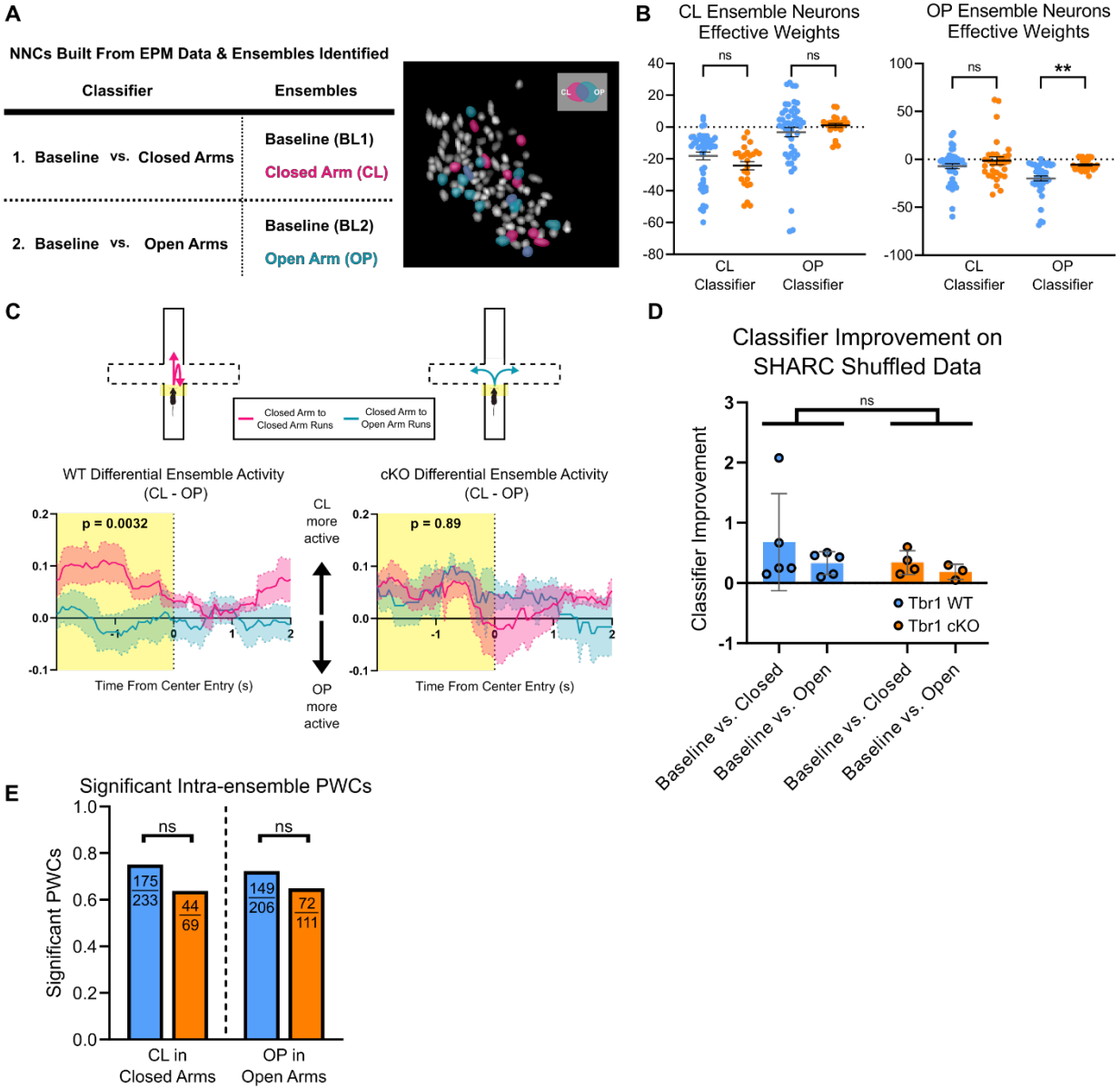
Figure 2.4. Encoding of the open arms of the EPM is disrupted in Tbr1 cKO mice

(A) *Left,* summary of NNCs trained on EPM data for each mouse and ensembles identified from these NNCs. *Right,* example FOV from one mouse with the two EPM ensembles highlighted.

(B) *Left,* effective weights of CL ensemble neurons (n = 48 for WT, n = 25 for cKO) in both trained NNCs from (A). *Right,* effective weights of OP ensemble neurons (n = 43 for WT, n = 31 for cKO) in both trained NNCs from (A). The magnitude of OP neuron weights in the OP classifier are significantly larger in WT than cKO mice. P-values shown are from Kruskal-Wallis tests with Dunn's multiple comparison correction between genotypes.
(Figure caption continued on next page)

(C) Relative activity of CL and OP ensembles is predictive of arm transitions in WT, but not cKO mice. For each ensemble in each mouse, a normalized activity trace was calculated and aligned to Closed-Closed and Closed-Open arm transitions, and mean ensemble activity traces were computed. The Differential Ensemble Activity traces were calculated by subtracting the mean OP activity from the mean CL activity for each transition type, and then these Differential Activity traces were averaged across genotypes (*bottom left* for WT; *bottom right,* for cKO). A multi-factor ANOVA was then completed for each genotype to determine if the Differential Ensemble activity in the 2s before entry to the EPM center (regions highlighted in yellow) was significantly associated with the transition type. Transition type was revealed to be a significant factor in WT ($p = 0.0069$) but not cKO mice ($p = 0.91$).

(D) Performance of NNCs trained on real data and tested on shuffled data. Performance of each trained classifier was assessed on Swap Shuffled and SHARC datasets, and a Performance Improvement score was calculated to assess the contribution of correlated activity in encoding EPM arms. A 2-way ANOVA with Sidaks correction of Performance Improvement scores showed no difference between genotypes ($p = 0.322$).

(E) The fraction of significant pairwise correlations within each EPM ensemble during exploration of its corresponding maze arm is plotted. Pairwise correlations and their p-values were calculated using the Matlab *corr* function. Correlations with a p-value <0.05 were taken as significant. P-values shown are from Fisher's Exact tests performed across genotypes for each ensemble.
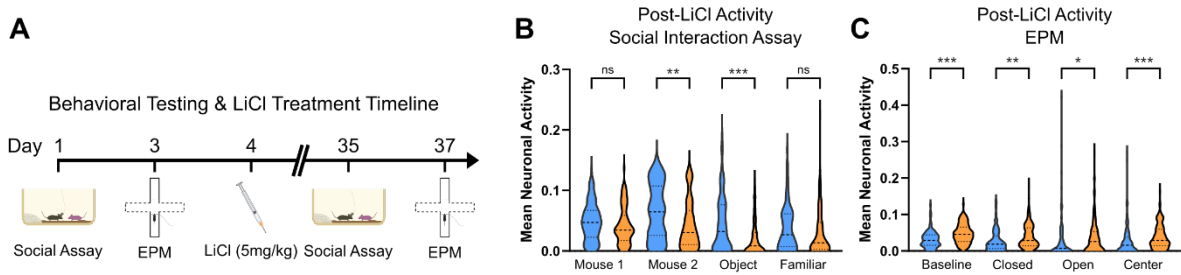
Figure 2.5. Lithium treatment rescues the activity deficit of Tbr1 cKO neurons

(A) Experimental timeline for behavior testing and lithium treatment. All animals, WT and cKO, were given LiCl after initial social and EPM testing. Animals were then retested 30 days after lithium administration.

(B) Violin plots of activity of all recorded neurons during the serial social interaction assay after lithium treatment (n=93 WT, n=98 cKO). Mean activity of each neuron was calculated as the fraction of frames corresponding to active social exploration that had a detected calcium event. P-values shown are from a Kruskal-Wallis test between genotypes.

(C) Violin plots of activity of all recorded neurons during the EPM assay after lithium treatment (n=113 WT, n=118 cKO). Mean activity of each neuron was calculated as the fraction of frames corresponding to exploration of each EPM zone that had a detected calcium event. P-values shown are from a Kruskal-Wallis test between genotypes.

2.7   References

Adhikari, A., Topiwala, M. A., & Gordon, J. A. (2011). Single Units in the Medial Prefrontal Cortex with Anxiety-Related Firing Patterns Are Preferentially Influenced by Ventral Hippocampal Activity. *Neuron*, *71*(5), 898–910. https://doi.org/10.1016/j.neuron.2011.07.027

Anastasiades, P. G., & Carter, A. G. (2021). Circuit organization of the rodent medial prefrontal cortex. *Trends in Neurosciences*, *44*(7), 550–563. https://doi.org/10.1016/j.tins.2021.03.006

Bedogni, F., Hodge, R. D., Elsen, G. E., Nelson, B. R., Daza, R. A. M., Beyer, R. P., Bammler, T. K., Rubenstein, J. L. R., & Hevner, R. F. (2010). Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(29), 13129–13134. https://doi.org/10.1073/pnas.1002285107

Corder, G., Ahanonu, B., Grewe, B. F., Wang, D., Schnitzer, M. J., & Scherrer, G. (2019). An amygdalar neural ensemble that encodes the unpleasantness of pain. *Science*, *363*(6424), 276–281. https://doi.org/10.1126/science.aap8586

Cunniff, M. M., Markenscoff-Papadimitriou, E., Ostrowski, J., Rubenstein, J. L. R., & Sohal, V. S. (2020). Altered hippocampal-prefrontal communication during anxiety-related avoidance in mice deficient for the autism-associated gene pogz. *eLife*, *9*, 1–27. https://doi.org/10.7554/ELIFE.54835

Fazel Darbandi, S., Nelson, A. D., Pai, E. L. lin, Bender, K. J., & Rubenstein, J. L. R. (2022). LiCl treatment leads to long-term restoration of spine maturation and synaptogenesis in adult Tbr1 mutants. *Journal of Neurodevelopmental Disorders*, *14*(1), 1–11. https://doi.org/10.1186/S11689-022-09421-5/FIGURES/4

Fazel Darbandi, S., Robinson Schwartz, S. E., Pai, E. L. L., Everitt, A., Turner, M. L., Cheyette,

B. N. R., Willsey, A. J., State, M. W., Sohal, V. S., & Rubenstein, J. L. R. (2020).

Enhancing WNT Signaling Restores Cortical Neuronal Spine Maturation and

Synaptogenesis in Tbr1 Mutants. *Cell Reports*, *31*(2), 107495.

https://doi.org/10.1016/J.CELREP.2020.03.059

Fazel Darbandi, S., Robinson Schwartz, S. E., Qi, Q., Catta-Preta, R., Pai, E. L. L., Mandell, J.

D., Everitt, A., Rubin, A., Krasnoff, R. A., Katzman, S., Tastad, D., Nord, A. S., Willsey, A.

J., Chen, B., State, M. W., Sohal, V. S., & Rubenstein, J. L. R. (2018). Neonatal Tbr1

Dosage Controls Cortical Layer 6 Connectivity. *Neuron*, *100*(4), 831-845.e7.

https://doi.org/10.1016/j.neuron.2018.09.027

Frost, N. A., Haggart, A., & Sohal, V. S. (2021). Dynamic patterns of correlated activity in the

prefrontal cortex encode information about social behavior. *PLOS Biology*, *19*(5),

e3001235. https://doi.org/10.1371/journal.pbio.3001235

Lee, A. T., Cunniff, M. M., See, J. Z., Ellwood, I. T., Ponnavolu, S., Sohal Correspondence, V. S.,

Wilke, S. A., Luongo, F. J., & Sohal, V. S. (2019). VIP Interneurons Contribute to

Avoidance Behavior by Regulating Information Flow across Hippocampal-Prefrontal

Networks Article VIP Interneurons Contribute to Avoidance Behavior by Regulating

Information Flow across Hippocampal-Prefrontal Networks. *Neuron*, *102*, 1223–1234.

https://doi.org/10.1016/j.neuron.2019.04.001

Levy, D. R., Tamir, T., Kaufman, M., Parabucki, A., Weissbrod, A., Schneidman, E., & Yizhar, O.

(2019). Dynamics of social representation in the mouse prefrontal cortex. *Nature*

Neuroscience, *22*(12), 2013–2022. https://doi.org/10.1038/s41593-019-0531-z

Luongo, F. J., Zimmerman, C. A., Horn, M. E., & Sohal, V. S. (2016). Correlations between

prefrontal neurons form a small-world network that optimizes the generation of

multineuron sequences of activity. *Journal of Neurophysiology*, *115*(5), 2359.

https://doi.org/10.1152/JN.01043.2015

Mukamel, E. A., Nimmerjahn, A., & Schnitzer, M. J. (2009). Automated analysis of cellular

signals from large-scale calcium imaging data. *Neuron*, *63*(6), 747–760.

https://doi.org/10.1016/j.neuron.2009.08.009

Notwell, J. H., Heavner, W. E., Darbandi, S. F., Katzman, S., McKenna, W. L., Ortiz-Londono, C.

F., Tastad, D., Eckler, M. J., Rubenstei, J. L. R., McConnell, S. K., Chen, B., & Bejerano,

G. (2016). TBR1 regulates autism risk genes in the developing neocortex. *Genome*

Research, *26*(8), 1013–1022. https://doi.org/10.1101/GR.203612.115/-/DC1

Sánchez-Bellot, C., AlSubaie, R., Mishchanchuk, K., Wee, R. W. S., & MacAskill, A. F. (2022).

Two opposing hippocampus to prefrontal cortex pathways for the control of approach

and avoidance behaviour. *Nature Communications*, *13*(1), 339.

https://doi.org/10.1038/s41467-022-27977-7

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M.,

Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M.,

Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Buxbaum, J. D. (2020).

Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional

Changes in the Neurobiology of Autism. *Cell*, *180*(3), 568-584.e23.

https://doi.org/10.1016/j.cell.2019.12.036

Selimbeyoglu, A., Kim, C. K., Inoue, M., Lee, S. Y., Hong, A. S. O., Kauvar, I., Ramakrishnan,

C., Fenno, L. E., Davidson, T. J., Wright, M., & Deisseroth, K. (2017). Modulation of

prefrontal cortex excitation/inhibition balance rescues social behavior in CNTNAP2-

deficient mice. *Science Translational Medicine*, *9*(401), eaah6733.

https://doi.org/10.1126/scitranslmed.aah6733

Sohal, V. S., & Rubenstein, J. L. R. (2019). Excitation-inhibition balance as a framework for

investigating mechanisms in neuropsychiatric disorders. *Molecular Psychiatry*, *24*(9),

1248–1257. https://doi.org/10.1038/s41380-019-0426-0

Willsey, A. J., Sanders, S. J., Li, M., Dong, S., Tebbenkamp, A. T., Muhle, R. A., Reilly, S. K., Lin,

L., Fertuzinhos, S., Miller, J. A., Murtha, M. T., Bichsel, C., Niu, W., Cotney, J., Ercan-

Sencicek, A. G., Gockley, J., Gupta, A. R., Han, W., He, X., … State, M. W. (2013).

Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in

the Pathogenesis of Autism. *Cell*, *155*(5), 997–1007.

https://doi.org/10.1016/J.CELL.2013.10.020

Willsey, H. R., Willsey, A. J., Wang, B., & State, M. W. (2022). Genomics, convergent

neuroscience and progress in understanding autism spectrum disorder. *Nature Reviews*

Neuroscience, *23*(6). https://doi.org/10.1038/s41583-022-00576-7

Workman, A. D., Charvet, C. J., Clancy, B., Darlington, R. B., & Finlay, B. L. (2013). Modeling

Transformations of Neurodevelopmental Sequences across Mammalian Species.

Journal of Neuroscience, *33*(17), 7368–7383. https://doi.org/10.1523/JNEUROSCI.5746-

12.2013

Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. J., O'Shea, D. J., Sohal, V. S.,

Goshen, I., Finkelstein, J., Paz, J. T., Stehfest, K., Fudim, R., Ramakrishnan, C.,

Huguenard, J. R., Hegemann, P., & Deisseroth, K. (2011). Neocortical

excitation/inhibition balance in information processing and social dysfunction. *Nature*,

477(7363), 171–178. https://doi.org/10.1038/nature10360

Chapter 3 - Encoding of social behavior by Dopamine Receptor 2 (*Drd2*) expressing neurons in the mPFC

3.1 Introduction

The medial prefrontal cortex (mPFC) is part of the mesocortical dopamine system, and receives dopaminergic input from neurons residing in the ventral tegmental area (VTA) (Björklund & Dunnett, 2007). Dopaminergic neurons in the VTA are heterogeneous with respect to their projection targets and their responses to rewarding or aversive stimuli (Lammel et al., 2014). Many of the critical functions of the mPFC, such as working memory, attention, cognitive flexibility, and emotional processing, have also been shown to depend on this dopamine input (Ellwood et al., 2017; Lammel et al., 2012; Lohani et al., 2019; Sawaguchi & Goldman-Rakic, 1991; Vander Weele et al., 2018). Moreover, numerous neuropsychiatric disorders have been linked to dysregulation of the mesocortical dopamine system, indicating that mPFC dopamine is linked, and may even be causal, to the pathology of these disorders (Grace, 2016; Granon et al., 2000; Howes & Kapur, 2009; Pitman et al., 2012).

Social interaction is a complex behavior that can be simultaneously anxiogenic and rewarding, and depends on the mPFC and dopamine (Bicks et al., 2015; Sotoyama et al., 2022). A recent study used microdialysis to measure dopamine release in the mPFC of rats during social interactions (Sotoyama et al., 2022). This study found a significant positive correlation between the measured mPFC dopamine concentration and the amount of time spent interacting, implying mPFC dopamine levels play a role in promoting social behavior.

Dopamine exerts its influence on downstream target neurons through any of the five G protein-coupled receptors (GPCRs) known to bind dopamine. Dopamine receptors are classified into two main families: "D1-like" (including D1R and D5R) and "D2-like" (including D2R, D3R,

and D4R) (Seamans & Yang, 2004). These two families are divided by their associated G proteins, which influence neuronal activity in opposing ways: the D1R family GPCRs are canonically associated with the stimulatory G proteins Gs and Gq, while the D2R family receptors tend to associate with the inhibitory G protein Gi/o (Seamans & Yang, 2004). Intriguingly, recent work from our lab has shown that for layer 5 D2R+ neurons of the mPFC, dopamine can increase excitability through a mechanism that is dependent on excitatory synaptic input and independent of Gi/o signaling (Robinson & Sohal, 2017).

D1R and D2R are the most highly expressed dopamine receptors in the mPFC, and mostly label non-overlapping populations of neurons (Santana et al., 2009). In a study from our lab using fiber photometry to measure bulk calcium from either D1R+ or D2R+ neurons during social interaction, it was found that the D2R+ neurons are particularly activated when investigating a novel juvenile conspecific (Brumback et al., 2018). While the D1R+ population was also activated at the onset of interaction, the D2R+ population showed an increase in calcium that lasted for tens of seconds. In optogenetic experiments from that same study, continuous excitation of the D2R+ population reduced social interaction time, while inhibition had no effect. These results suggest a possible role for mPFC D2R+ neuron activity in social behavior.

In this study, we sought to further characterize the activity of D2R+ mPFC neurons during social interaction by utilizing microendoscopic imaging to record the calcium activity of individual D2R+ neurons. To probe the role of D2Rs in the activity of these neurons, we also recorded the activity of D2R+ neurons from which D2Rs had been knocked out. Mice performed the 3-chamber social assay, and we assessed neuronal activity during interaction and within different regions of the apparatus. We followed up on these imaging studies by performing optogenetic inhibition experiments which silenced D2R+ mPFC neurons within specific zones of the 3-chamber apparatus that were normally associated with significant D2R+ neuron activity.

3.2 Materials and Methods

Animal Subjects

All animal care procedures and experiments were conducted in accordance with the National Institutes of Health guidelines and approved by the Administrative Panels on Laboratory Animal Care at the University of California, San Francisco. Mice were housed in a temperature-controlled environment (22–24 °C) with ad libitum access to food and water. Mice were reared in normal lighting conditions (12-h light/dark cycle). The following transgenic mouse lines were used for all experiments: Drd2-Cre$^{+/-}$ (line ER44; www.gensat.org) and Drd2$^{fl/fl}$ (https://www.jax.org/strain/020631).

Surgeries

For all experiments, mice underwent surgery when they were 10-12 weeks old. Animals were anaesthetized with isoflurane (3.5% induction, 1-2% maintenance in 95% oxygen, flow rate 0.9L/min) and secured with ear bars in a stereotaxic frame (David Kopf instruments). Analgesics were administered via subcutaneous injection at surgery onset (buprenorphine, 0.1 mg/kg; meloxicam, 2 mg/kg). Body temperature was maintained using a heating pad. The scalp was incised along the rostro-caudal axis to expose the dorsal surface of the skull, which was then aligned using bregma and lambda as references. After surgery, animals were allowed to recover on a heated pad until ambulatory. All animals received an additional dose of meloxicam on the day following surgery.

Microendoscopic Imaging Surgeries

Two distinct strategies were used to image Drd2-labeled mPFC neurons with intact, or ablated, Drd2 expression. To image Drd2+ neurons, Drd2-Cre$^{+/-}$ mice were injected with AAV9-hSyn-FLEX-jGCaMP6m-WPRE (AddGene) diluted 1:3 in sterile saline. To specifically label Drd2-expressing neurons with GCaMP while also knocking out the Drd2 gene, Drd2$^{fl/fl}$ mice were

injected with a cocktail of AAV9-hSyn-FLEX-jGCaMP6m-WPRE (AddGene) and a virus encoding the Cre gene under control of the Drd2 promoter: AAV8-D2SP-Cre-mCherry (Zalocusky et al., 2016). The two viruses were mixed 1:1, and this mixture was further diluted 1:1 in sterile saline immediately before injection.

For all mice, a dental drill was used to create an opening in the skull above the right mPFC (-1.7mm AP, +0.3mm ML relative to bregma). Virus was injected at four different depths, starting with the most ventral, (-2.75mm, -2.5mm, -2.25mm, and -2.0mm DV; 150nl at each depth) at a rate of 100nl/min with a 35-gauge microinjection syringe (World Precision Instruments) connected to a pump (UMP3 UltraMicroPump, World Precision Instruments). The needle was kept at each injection site for 5mins after infusion.

After injections, mice were implanted with a 0.5 mm diameter x 4 mm long integrated GRIN lens (Inscopix), which was slowly lowered to a final depth of 2.3mm. Implants were secured to the skull with Metabond (Parkell). After surgery, mice were singly housed to prevent damage to their implants. Mice were allowed 3 weeks to recover from surgery before behavioral testing and microendoscopic imaging.

Optogenetic Inhibition Surgeries

For optogenetic experiments, Drd2-Cre[+/-] mice along with WT (Drd2-Cre[-/-]) littermates were used. A dental drill was used to create an opening in the skull above each mPFC hemisphere (-1.7mm AP, +/- 0.5mm ML relative to bregma). Mice were injected with AAV5-EF2a-DIO-eNpHR-mCherry (AddGene) bilaterally at two different depths starting with the most ventral, (-2.6mm and -2.3mm DV; 500nl at each depth) at a rate of 150nl/min with a 35-gauge microinjection syringe (World Precision Instruments) connected to a pump (UMP3 UltraMicroPump, World Precision Instruments). The needle was kept at each injection site for 5mins after infusion.

After injections, mice were implanted with a dual fiberoptic cannula (Doric; DFC_200/240-0.22_2.7mm_DF1.0_FLT). Implants were lowered slowly to a final depth of -2.3mm DV, and secured to the skull with Metabond (Parkell). After surgery, mice were returned to their original home cage and group housing arrangement. Mice were given 4 weeks to recover from surgery, and to allow for viral transgene expression, before behavioral testing was conducted.

Behavioral Assays

For all cohorts, mice were habituated to experimenter handling for 3-5 days (5-10min per day) prior to any testing. Additionally, mice were also habituated to the head-mounted microscope (calcium imaging cohorts) or fiberoptic patch cables (optogenetic cohorts) (3-5 days, 45min per day). On testing days, mice were transported to the testing room 45-60min prior to test onset. Behavioral videos were recorded with a video camera, and real-time animal tracking was performed using AnyMaze software.

3-Chamber Social Behavior Assay

The 3-chamber apparatus measured 25 cm length x 58 cm width x 26 cm height. Both side chambers were partially separated from the center chamber by two 7cm-wide walls, leaving a 14cm opening into each side chamber from the center. An inverted wire cup, 10cm in diameter, was placed in each side chamber to hold the target mice for social investigation. The assay consisted of three sequential 10-minute stages in which the test mouse was allowed to freely explore the entire apparatus: (1) habituation, (2) sociability, and (3) social novelty. Between stages, the test mouse was gently guided into the center chamber and contained there by removeable walls until the beginning of the next stage. In the habituation stage, both wire cups remained empty. In the sociability stage, or "SB," a novel sex-matched juvenile mouse (4-6 weeks old) was placed under one of the cups. In the social novelty stage, or "SN," a second

novel sex-matched juvenile mouse was placed under the other cup (the empty cup from the SB stage) while leaving the first novel juvenile in the apparatus.

For calcium imaging cohorts, AnyMaze tracking was used to identify times when the test mice were within 3cm of the wire cups, as well as times when the mice transitioned between chambers. A TTL cable connecting the AnyMaze computer to the Inscopix DAQ box was used to sync calcium imaging and behavior videos.

For optogenetic inhibition cohorts, mice were tested on the 3-chamber assay twice. During both tests, eNpHR stimulation was delivered via a dual fiberoptic patch cord connected to a 532nm laser. Total laser power from the dual fiberoptic patch cord measured 8-10mW. A TTL cable connecting the AnyMaze computer to the laser allowed for light to be delivered when the mice were in certain regions of the apparatus. For the first test, the laser was triggered only when the mice were in the center chamber, and for the second test, the laser was triggered when the mice were in either side chamber. The two tests were performed 10 days apart. All social interactions between test mice and target mice were manually scored. AnyMaze tracking was used to identify transitions between chambers.

Open Field Test (OFT)

Mice from the optogenetic inhibition cohorts performed the open field test (OFT) 4 days after their first 3-chamber social assay. Mice were allowed to explore the apparatus (50cm X 50cm X 50cm) for 15 minutes. During the first 10 minutes, eNpHR stimulation was delivered every other minute, continuously, beginning with minute 2. Minutes 11-15 were stimulation free. AnyMaze tracking was used to evaluate locomotion of animals.

Calcium Imaging

Data Acquisition and Processing

Calcium signals were imaged at 20Hz with 4x spatial downsampling using a head-mounted, one-photon microendoscope (nVoke2; Inscopix Inc.). Calcium activity videos were preprocessed in Inscopix Data Processing Software (IDPS). Videos were spatially band-pass filtered with cut-offs set to 0.015 pixel-1 (low) and 0.300 pixel-1 (high), and then motion corrected. Preprocessed videos were exported from IDPS as tiff files and imported to MatLab for cell segmentation. Putative cellular ROIs were identified using a combined PCA/ICA approach (Mukamel et al., 2009). For each movie, putative neurons were manually sorted to remove ROIs with low SNR or overlapping boundaries. Cell sorting was done with the assistance of the signalSorter GUI from an open-source calcium imaging software package (CIAPKG) (Corder et al., 2019).

Cellular dF/F traces for sorted neurons were extracted and then used to generate a binary calcium event raster for each neuron using a previously described event detection algorithm (Frost et al., 2021). All subsequent analyses were performed using these binary calcium event rasters.

Data Analysis

All analyses of calcium activity data were performed in MatLab using custom scripts. To calculate a cell's mean activity during a particular behavior or set of timepoints, first, all calcium imaging frames corresponding to the designated category were identified. Mean activity was then computed as the fraction of these associated frames with a detected calcium event.

To determine if a cell's activity was significantly associated with a particular behavior, we calculated the normalized cosine similarity between a neuron's binary calcium event raster and a binary behavior vector. These behavior vectors indicated which frames corresponded to the behavior in question. These cosine similarity values were then compared to bootstrap

distributions. Bootstrap distributions were generated by randomly permuting a neuron's calcium

activity raster 250 times and recalculating the cosine similarity with an unshuffled binary

behavior vector. A neuron was considered significantly associated with a behavior if the true

cosine similarity was greater than the 95th percentile of the bootstrap distribution, and it was

considered significantly unassociated if it was below the 5th percentile of the bootstrap

distribution.

Quantification and statistical analysis

All statistical calculations were performed in MatLab or Graphpad Prism. All details of these

statistical analyses are provided in the main text and figure legends.

3.3　Results

Deletion of *Drd2* from mPFC neurons disrupts activity in the 3-Chamber social assay

In order to better understand the role that mPFC Drd2+ neurons play in social behavior, we

utilized single-photon microendoscopic imaging to measure their calcium (Ca2+) activity of

these neurons during a social assay. To specifically image Drd2+ neurons, we utilized a Drd2-

Cre reporter mouse and a Cre-dependent virus containing the genetically encoded calcium

indicator GCaMP6m. We used a different mouse line and labelling strategy to probe the role that

D2Rs played in the activity of these neurons. Specifically, mice harboring a floxed Drd2 allele

(Drd2$^{fl/fl}$) were injected with a viral cocktail consisting of the Cre-dependent GCaMP6m along

with a virus encoding Cre under the control of a Drd2 sufficient promoter (AAV-D2sp-Cre)

(Zalocusky et al., 2016). This approach allowed us to image specifically from neurons that

normally express Drd2 while simultaneously knocking out the receptor from those same

neurons. We refer to the Drd2-Cre mice injected with Cre-dependent GCaMP6m as "Drd2-WT,"

and the Drd2$^{fl/fl}$ mice injected with the D2sp-Cre and Cre-dependent GCaMP6m as "Drd2-KO."

In total, we recorded 429 Drd2-WT neurons from 7 mice, and 536 Drd2-KO neurons from 8 mice.

All mice were tested on the 3-chamber social assay, which consisted of three 10-minute task phases: Habituation, the Sociability Test (SB), and the Social Novelty Test (SN) (see **Fig. 3.1A-C**, left). The 3-chamber apparatus contained an inverted wire cup in each side chamber to house social targets (sex- and strain-matched juvenile mice, 4-6 weeks old). During Habituation, mice explore the apparatus without any social targets. In the SB phase, one social target is placed under one wire cup. Finally, in the SN phase, a second social target is placed under the second cup.

The amount of time the mice spent in proximity (within 3cm) of the cups during all task phases was quantified. During Habituation, since both wire cups remain empty, we arbitrarily designated the cup that would hold the social target during the SB phase as "Cup 1." There were no significant differences between groups in the time spent in proximity to Cup 1, Cup 2, or in the center chamber during Habituation (**Fig. 3.1A**, middle). We also calculated a Preference Index for Cup 1 (Time spent near Cup 1 / Total time spent near either cup). Despite there being no a priori difference between the two cups, and the location of the social target during SB being randomized across mice, the Drd2-WT group showed a slight preference for Cup 1 during Habituation (**Fig. 3.1A,** right). During the SB and SN phases, there were no differences between groups in time spent near the cups or in the center chamber (**Fig. 3.1B-C**, middle). Further, the SB and SN Preference Indices were similar across groups (**Fig. 3.1B-C**, right). The lack of obvious behavioral abnormalities in the Drd2-KO animals may reflect the fact that this receptor knockout was unilateral and confined to a relatively small portion of the mPFC.

To assess whether the loss of the D2R affected neuronal activity, we calculated the mean activity of all recorded neurons during times of behavioral interest. Specifically, for each task phase, we looked at times when the mice were in proximity to the cups, or in the center

chamber. Mean activity was computed as the fraction of behaviorally relevant frames with a detected calcium event. There were no differences between the activity distributions of Drd2-WT and KO neurons when the animals were in proximity to the cups during any task phase (**Fig. 3.2A-C,** left). Surprisingly, though, mean activity of Drd2-KO neurons was significantly lower than their WT counterparts in the center chamber during all task phases. To probe this finding even further, we inspected the relative frequency distributions of mean activity for all task phases. In Drd2-KO neurons, the activity distributions for both cups and the center chamber were nearly identical in all task phases (**Fig. 3.2A-C,** right). Conversely, for Drd2-WT neurons, the activity distribution for the center chamber is clearly shifted right in each task phase, indicating that a portion of these neurons are highly activated in the center (**Fig. 3.2A-C,** middle).

Next, we sought to determine if there were neurons that were selectively activated or silenced during times of cup proximity or center chamber exploration. To determine if a cell's activity was significantly associated with a particular behavior, we calculated the normalized cosine similarity between a neuron's binary calcium event raster and a binary behavior vector. These behavior vectors indicated which frames corresponded to the behavior in question. We then generated a null distribution of cosine similarity values for each neuron/behavior pair by randomly permuting a neuron's calcium activity. Importantly, null distributions were created using activity from the task phase in question, rather than from the entire recording. Neurons with a cosine similarity value above the 95th percentile of the null distribution were deemed an "Up" neuron for that behavior, and those below the 5th percentile were deemed a "Down" neuron for that behavior. The fraction of Up and Down neurons found in each mouse for each behavior was calculated (**Fig. 3.3**). During the Habituation task phase, we find a significantly larger portion of Center Up neurons in Drd2-WT animals compared to Drd2-KO (**Fig. 3.3A**). However,

no significant differences in the portion of Up and Down neurons were found during the SB and SN task phases (**Fig. 3.3B-C**).

Inhibition of Drd2+ neurons of the mPFC during the 3-chamber test disrupts normal social behavior

Due to our observation of elevated Drd2-WT neuron activity in the center chamber, we hypothesized that inhibition of these neurons in the center chamber of a 3-chamber test might impair social behavior. To test this, we injected Drd2-Cre+ mice bilaterally in the mPFC with a Cre-dependent virus encoding the inhibitory opsin eNpHR3.0 (n = 11 mice). As a control, we also injected Drd2-Cre- littermates with this Cre-dependent virus (n = 10 mice). Animals were tested on the 3-chamber assay twice. During Test 1 the stimulation laser was triggered only when the mice were in the center chamber, and during Test 2, performed 10 days later, the laser was triggered when the mice were in either side chamber (**Fig. 3.4A**). All interactions were manually scored.

During Test 1, we find no significant differences in the amount of time spent interacting with any of the social targets (**Fig. 3.4B**). However, looking at the overall number of interactions, there was a significant increase in D2-Inhibition animals compared to control animals (2-Way ANOVA, Main Group Effect, p = 0.0183), despite there being no significant increase for any particular social target (**Fig. 3.4C**). Importantly, there was no difference between groups in the amount of time spent in the center chamber in any task phase, implying that the inhibition was not intrinsically rewarding or aversive (**Fig. 3.4D**). For Test 2, no significant differences between groups were observed for interaction time (**Fig. 3.4E**), number of interactions (**Fig. 3.4F**), or center chamber time (**Fig. 3.4G**).

Finally, to assess any changes in social preferences induced by D2 neuron inhibition, we calculated the SB and SN Preference Indices for each mouse. These preference indices were

calculated as described before, except using manually scored interaction times instead of cup proximity times. For the SB Preference Index, we found no difference between groups in Test 1 or 2. Interestingly, though, both groups showed a significant decrease in SB Preference from Test 1 to Test 2 (**Fig. 3.4H**). For the SN Preference Index, no differences between groups were seen in Test 1 or Test 2. Further, no differences within groups were seen between Test 1 and Test 2 (**Fig. 3.4I**).

Inhibition of Drd2+ neurons of the mPFC increases locomotion in the open field test

Finally, to determine if inhibition of mPFC Drd2+ neurons has a general effect on locomotion, we tested the Control and D2-Inhibition animals on the open field (OFT). Mice were allowed to explore the OFT apparatus for 15 minutes. During the first 10 minutes of the test, mice received laser stimulation every other minute, with no stimulation in the final 5 minutes (**Fig. 3.5A**). Inspecting the distance traveled by minute, there was a clear increase in movement during the periods of laser stimulation in the D2-Inhibition group (**Fig. 3.5B**). To evaluate this, the total distance traveled in the first 10 minutes of the test was summed up separately for Light On and Light Off periods. In the D2-Inhibition group, there was a significant increase in the distance traveled in Light On periods versus Light Off, while in the control animals there was no difference (**Fig. 3.5C**). There was no significant difference between groups with regard to distance traveled in Light On periods.

3.4   Discussion

Here, we explore the role of D2R+ neurons of the mPFC in social behavior by recording calcium activity from individual D2R-labeled neurons with and intact D2R expression and those from which D2Rs had been knocked out. Mice performed the 3-chamber social test, and neuronal activity associated with various regions of the task apparatus was analyzed. From these experiments, we find that D2R+ neurons show significant activity associated with the center

chamber, and that this center chamber associated activity is diminished when the D2R is knocked out.

After identifying these times when the D2R is particularly important for promoting socially related activity, we performed an optogenetic inhibition experiment to inhibit these neurons in the center chamber of the 3-chamber task. This manipulation led to increased numbers of interactions with social targets in the task. Conversely, inhibiting D2R+ neurons in the side chambers had no significant effect on interaction time or number of interactions. Additionally, inhibition of D2R+ neurons in an open field test led to increased locomotion in light on versus light off periods.

While it was surprising to find that D2R deletion primarily effected activity associated with the center chamber and not with ongoing social interaction, there are a couple of potential interpretations of these results. The center chamber can be viewed as a decision point in the 3-chamber task, and both dopamine and the mPFC at-large are well known to support decision making and cognitive functions. Further, most visits to the center chamber are brief and occur between interactions. Increased activity at this location could reflect the intention to explore social targets, or reflect the experience of a recent social interaction. A prominent theory in the field of dopamine in the mPFC states that D1R+ and D2R+ cells play different roles in decision making, with D1R+ activity serving to maintain current representations and behavioral strategies and D2R+ activity serving to promote exploration (Durstewitz & Seamans, 2008). Through this lens, center chamber D2R+ activity could reflect exploratory drive, and disruption of this activity would then decrease social investigation. The increased number of social interactions without an increase in overall interaction time that we saw with center-inhibition could be seen as a diminished capacity to explore. Our finding of increased locomotion with D2R+ inhibition in the open field test, which could be interpreted as increased exploration, highlights the complexities of this system, and the importance of behavioral context. This observation also points to a

potential motor-effect of mPFC D2R+ inhibition and could imply that the increased number of interactions seen with center chamber inhibition is simply due to increased movement. However, we saw no difference in distance traveled in the 3-chamber test with center inhibition (data not shown), indicating this is likely not the case.

Another possible interpretation of these results comes from work demonstrating that the VTA to mPFC dopaminergic inputs are highly activated by aversive stimuli (Lammel et al., 2012, 2014). The 3-chamber apparatus itself is a brightly lit and novel environment that could be anxiogenic to mice, at least initially. Given that mice encounter the center chamber before any other region of the apparatus, elevated D2R+ activity in this region might just reflect these, or other, anxiogenic aspects of this task. More work needs to be done to establish whether or not the D2R+ neurons activated during social exploration are also activated by acutely aversive stimuli, or if there are distinct subpopulations that respond only in one of these contexts.

Given the previous work from our lab describing the elevated activity of the mPFC D2R+ time locked to the first sniff of a social interaction and that optogenetic activation of these neurons leads to reduced social interaction (Brumback et al., 2018), it is surprising that we find the activity of the neurons to be most significantly associated with a non-social chamber. One potential explanation is that this previous study used a different social interaction paradigm: the juvenile social intruder assay, which is performed in the test mouse's home cage. The social intruder assay involves a freely moving social target and a familiar environment, while the three chamber assay is in a novel environment with social targets that are confined. More work needs to be done to disentangle the contributions of novel environments and freely moving conspecifics to D2R+ recruitment in the mPFC. Further, given that the D2R+ population in the mPFC consists of both excitatory and inhibitory neurons with a wide range of projection targets, it is uncertain whether it is a subset of the overall population that is particularly recruited during social behavior. Our labeling strategy for the Drd2-KO animals employed a virus expressing Cre

under a reduced version of the Drd2 promoter. It also remains to be established whether this reduced promoter preferentially expresses in a subpopulation of D2R+ neurons, and how this population overlaps with Cre expression in the D2-Cre transgenic mouse line. Given that similar numbers of neurons were recorded from mice in both groups of animals, and that histology showed similar labeling across mPFC layers, it is unlikely that there are large differences in labeling between the two strategies.
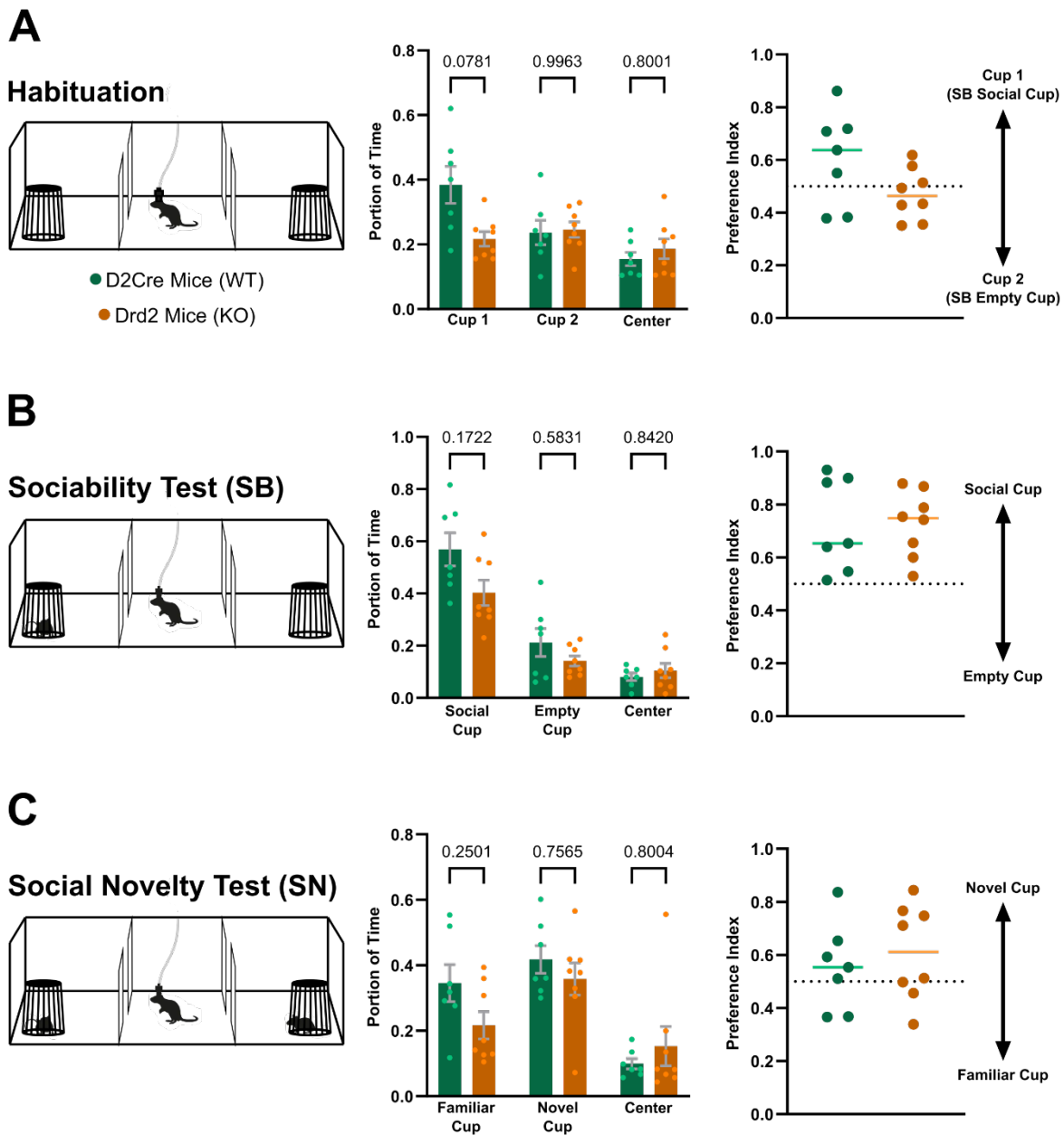
3.5 Figures

A

Habituation

- ● D2Cre Mice (WT)
- ● Drd2 Mice (KO)

B

Sociability Test (SB)

C

Social Novelty Test (SN)

Figure 3.1. Behavior of calcium imaging animals in the 3 Chamber test

(A) *Left,* schematic of the Habituation phase of the 3-chamber test. *Middle,* time spent in proximity (within 3cm) of either cup, or in the center chamber, reported as the portion of entire 10min task phase. P-values shown are from a 2-way ANOVA with Sidaks correction. *Right,* cup preference index, calculated as the time spent near Cup 1 divided by the total time spent near either cup. *(Figure caption continued on next page)*

(A) *(Figure caption continued from previous page)* The line indicates the median of the data. For each mouse, "Cup 1" was designated as the cup that would hold the social target during the SB task phase. Data points represent individual animals (n = 7 Drd2 WT mice, n = 8 Drd2-KO mice).

(B) *Left,* schematic of the Sociability (SB) phase of the 3-chamber test. *Middle,* time spent in proximity (within 3cm) of either cup, or in the center chamber, reported as the portion of entire 10min task phase. P-values shown are from a 2-way ANOVA with Sidaks correction. *Right,* Sociability preference index, calculated as the time spent near the Social Cup divided by the total time spent near either cup. The line indicates the median of the data.

(C) *Left,* schematic of the Social Novelty (SN) phase of the 3-chamber test. *Middle,* time spent in proximity (within 3cm) of either cup, or in the center chamber, reported as the portion of entire 10min task phase. P-values shown are from a 2-way ANOVA with Sidaks correction. *Right,* Social Novelty preference index, calculated as the time spent near the Novel Cup divided by the total time spent near either cup. The line indicates the median of the data.
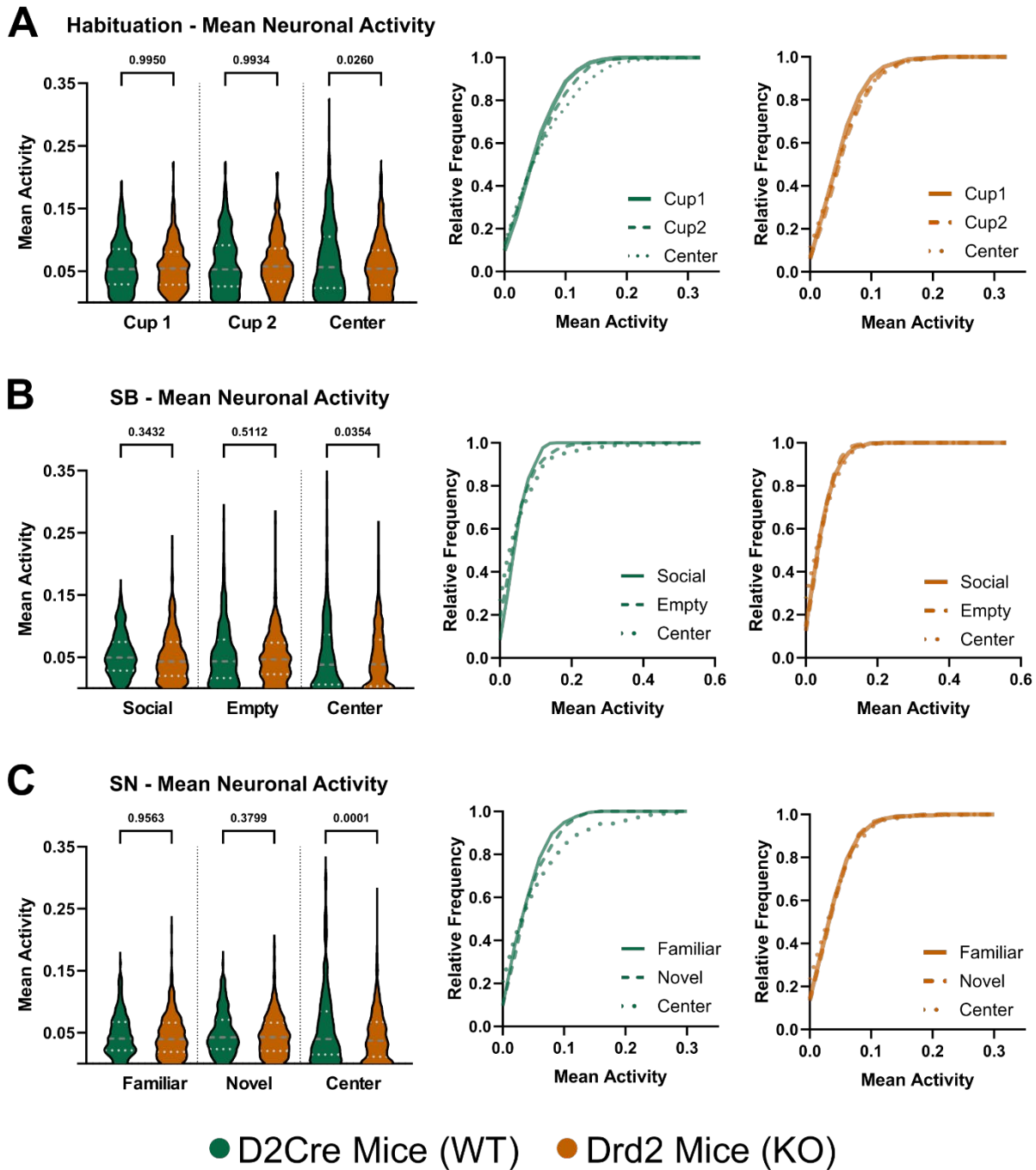
Figure 3.2. Activity associated with the center chamber is weakened in mPFC Drd2-KO neurons

(A) *Left,* violin plot showing the distribution of mean activity of all recorded neurons (n = 429 Drd2+ neurons from 7 mice, n = 536 Drd2-KO neurons from 8 mice) during the habituation task phase. P-values shown are from a Brown-Forsythe ANOVA with Dunnett's T3 multiple comparisons test. *(Figure caption continued on next page)*

(A) *(Figure caption continued from previous page) Middle*, relative frequency distributions of the mean activity of the Drd2+ data shown in the violin plot. *Right*, relative frequency distributions of the mean activity of the Drd2-KO data shown in the violin plot.
(B) *Left,* violin plot showing the distribution of mean activity of all recorded during the Sociability (SB) task phase. P-values shown are from a Brown-Forsythe ANOVA with Dunnett's T3 multiple comparisons test. *Middle*, relative frequency distributions of the mean activity of the Drd2+ data shown in the violin plot. *Right*, relative frequency distributions of the mean activity of the Drd2-KO data shown in the violin plot.
(C) *Left,* violin plot showing the distribution of mean activity of all recorded during the Social Novelty (SN) task phase. P-values shown are from a Brown-Forsythe ANOVA with Dunnett's T3 multiple comparisons test. *Middle*, relative frequency distributions of the mean activity of the Drd2+ data shown in the violin plot. *Right*, relative frequency distributions of the mean activity of the Drd2-KO data shown in the violin plot.
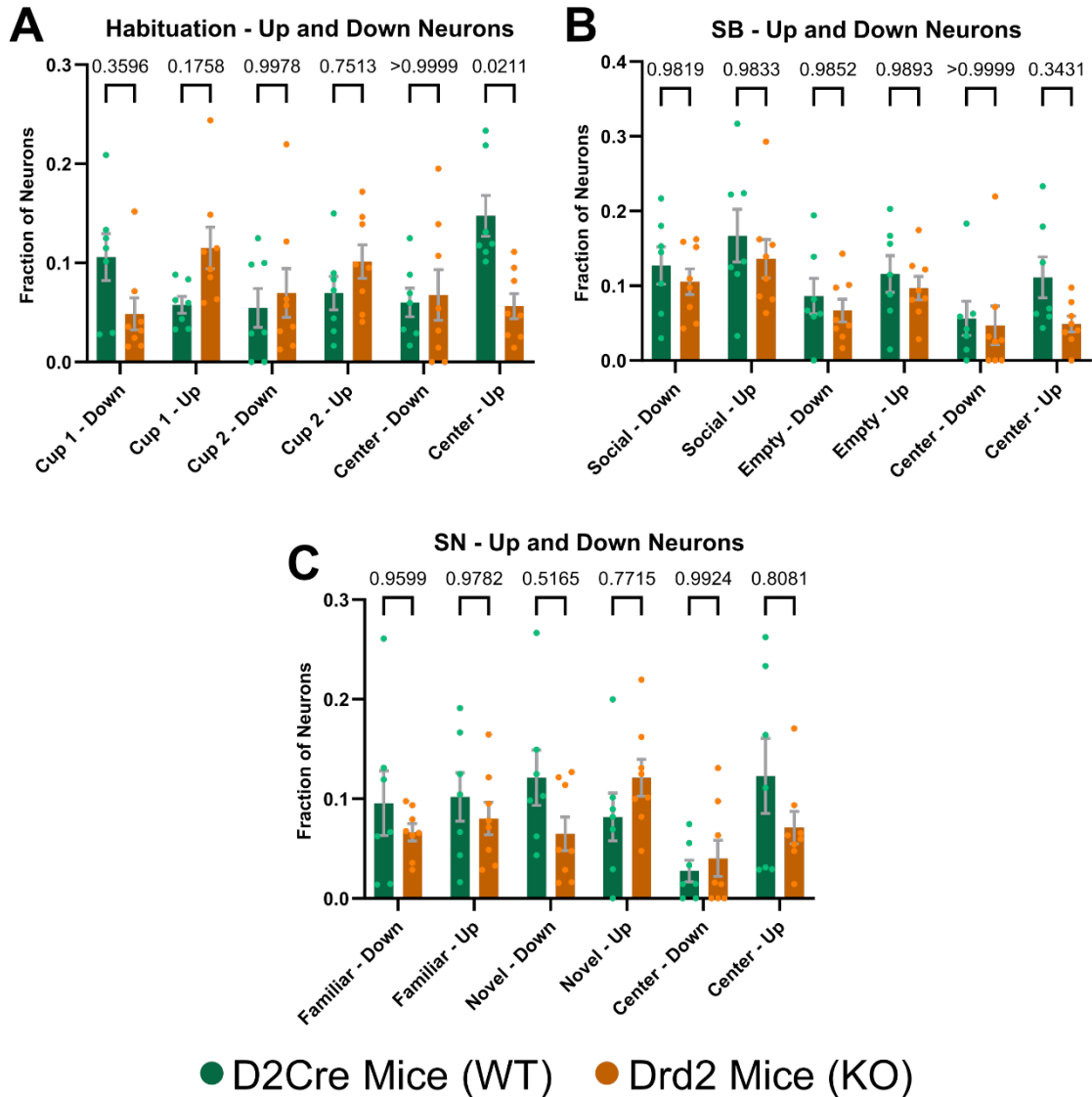
Figure 3.3. Three-chamber behavior-associated neurons are mostly preserved in mPFC Drd2-KO animals

(A) Fraction of neurons from each animal with significantly elevated (Up) or reduced (Down) activity while the animal was in proximity to Cup 1 or Cup 2, or in the center chamber during the Habituation phase. Data points represent individual animals (n = 7 Drd2 WT mice, n = 8 Drd2-KO mice). P-values shown are from a 2-way ANOVA with Sidak's multiple comparisons test.

(B) Fraction of neurons from each animal with significantly elevated (Up) or reduced (Down) activity while the animal was in proximity to the Social or Empty Cup, or in the center chamber during the SB phase. P-values shown are from a 2-way ANOVA with Sidak's multiple comparisons test.

(Figure caption continued on next page)

(Figure caption continued from previous page)

(C) Fraction of neurons from each animal with significantly elevated (Up) or reduced (Down) activity while the animal was in proximity to the Novel or Familiar Cup, or in the center chamber during the SN phase. P-values shown are from a 2-way ANOVA with Sidak's multiple comparisons test.
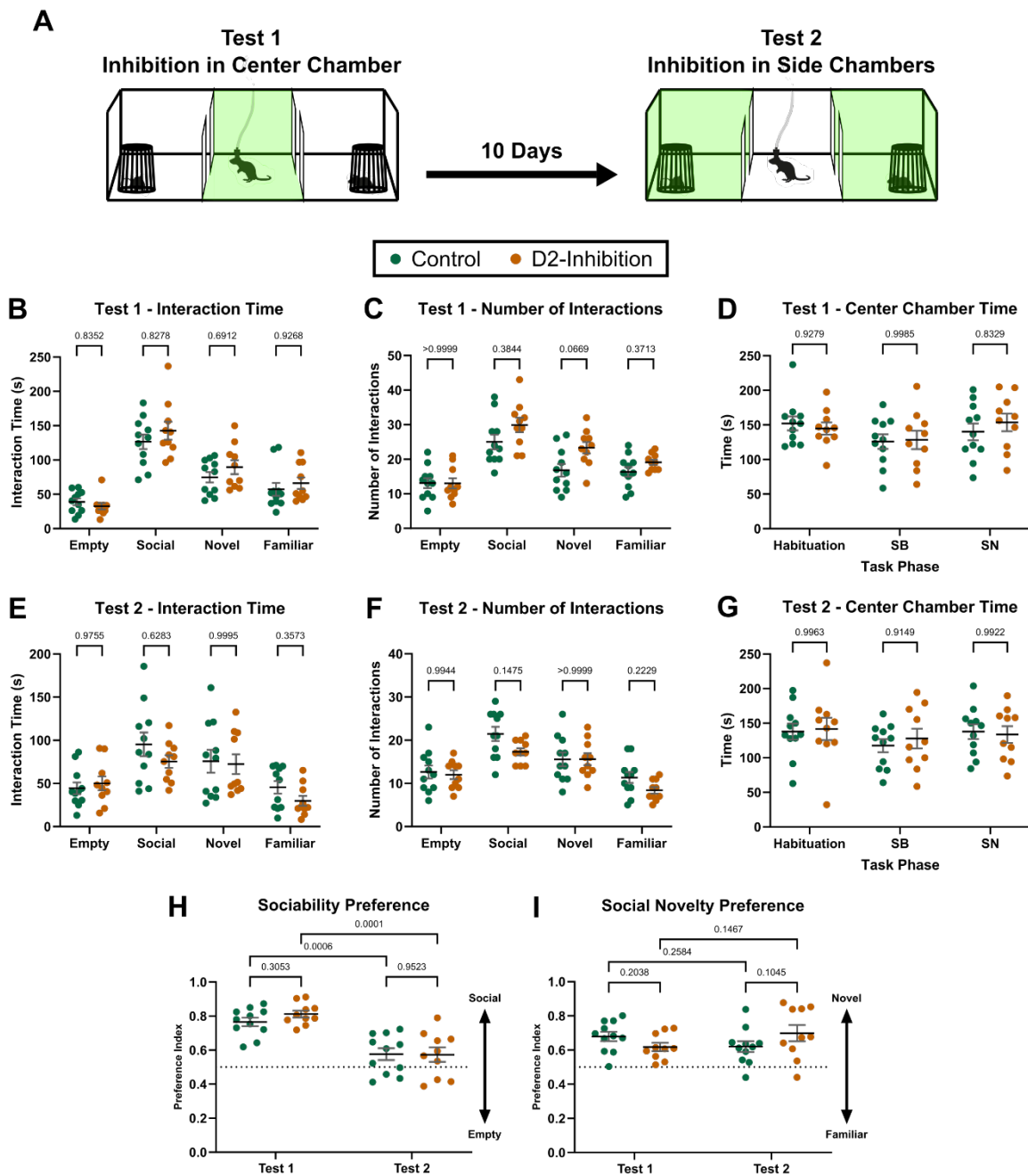
Figure 3.4. Inhibition of mPFC Drd2+ neurons in the center chamber increases the number of interactions

(A) Overview of 3-Chamber testing with optogenetic inhibition of mPFC Drd2+ neurons. In Test 1, the stimulation laser was triggered only in the center chamber during all three task phases. In Test 2, which was conducted 10 days after Test 1, the stimulation laser was triggered in both side chambers during all three task phases.
(Figure caption continued on next page)

(B) Interaction times during SB and SN task phases from Test 1. Data points represent individual animals (n = 11 D2-Cre- control mice, n = 10 D2-Cre+ inhibition mice). P-values shown are from a 2-way ANOVA with Sidak's multiple comparisons test.

(C) Number of interactions during SB and SN task phases from Test 1. Data points represent individual animals (n = 11 D2-Cre- control mice, n = 10 D2-Cre+ inhibition mice). P-values shown are from a 2-way ANOVA with Sidak's multiple comparisons test.

(D) Center chamber time during all task phases from Test 1. Data points represent individual animals (n = 11 D2-Cre- control mice, n = 10 D2-Cre+ inhibition mice). P-values shown are from a 2-way ANOVA with Sidak's multiple comparisons test.

(E) Same as (B), but for 3-chamber Test 2.

(F) Same as (C), but for 3-chamber Test 2.

(G) Same as (D), but for 3-chamber Test 2.

(H) Sociability Preference Index for the SB task phase in Test 1 and Test 2. The Sociability Preference Index was calculated as the interaction time with the social target divided by total interaction time with the social and empty targets. Data points represent individual animals. P-values shown are from a 2-way ANOVA with Fisher's LSD multiple comparisons test.

(I) Social Novelty Preference Index for the SN task phase in Test 1 and Test 2. The Social Novelty Preference Index was calculated as the interaction time with the novel target divided by total interaction time with the novel and familiar targets. Data points represent individual animals. P-values shown are from a 2-way ANOVA with Fisher's LSD multiple comparisons test.
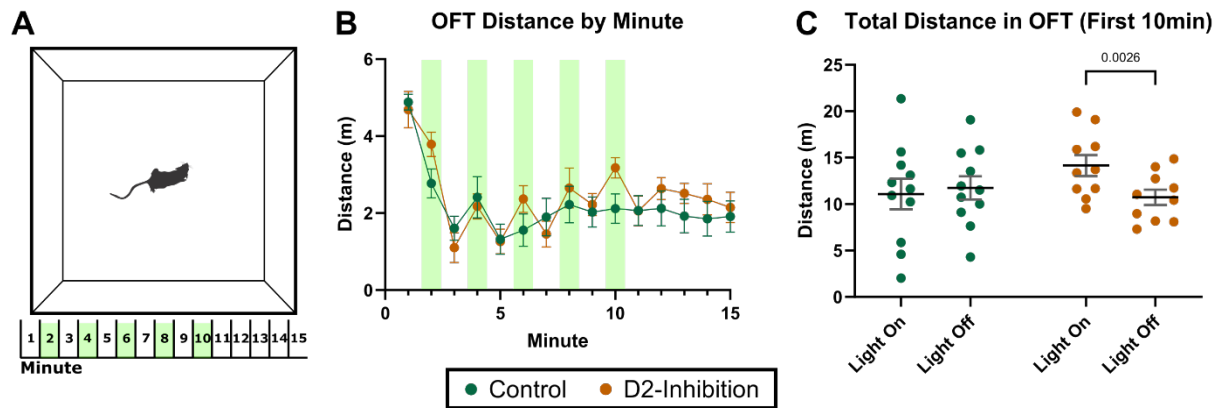
Figure 3.5. Inhibition of mPFC Drd2+ neurons in the open-field test increases locomotion

(A) Overview of open-field test (OFT) with optogenetic inhibition of mPFC Drd2+ neurons. Animals were allowed to explore the OF for 15 minutes. The stimulation laser was turned on every other minute for the first 10 minutes. The final 5 minutes were free of laser stimulation.

(B) Distance traveled in the open field during each minute of the test. Datapoints represent the average of each group (n = 11 D2-Cre- control mice, n = 10 D2-Cre+ inhibition mice) with the SEM. Light green bars indicate minutes with laser stimulation.

(C) Cumulative distance traveled in the OFT during Light Off and Light On timepoints from the first 10 minutes of the test. P-values shown are from a 2-way ANOVA with Fisher's LSD multiple comparisons test.

3.6   References

Bicks, L. K., Koike, H., Akbarian, S., & Morishita, H. (2015). Prefrontal Cortex and Social

    Cognition in Mouse and Man. *Frontiers in Psychology*, *6*, 1805.

    https://doi.org/10.3389/fpsyg.2015.01805

Björklund, A., & Dunnett, S. B. (2007). Dopamine neuron systems in the brain: An update.

    Trends in Neurosciences, *30*(5), 194–202. https://doi.org/10.1016/j.tins.2007.03.006

Brumback, A. C., Ellwood, I. T., Kjaerby, C., Iafrati, J., Robinson, S., Lee, A. T., Patel, T.,

    Nagaraj, S., Davatolhagh, F., & Sohal, V. S. (2018). Identifying specific prefrontal

    neurons that contribute to autism-associated abnormalities in physiology and social

    behavior. *Molecular Psychiatry*, *23*(10), 2078–2089. https://doi.org/10.1038/mp.2017.213

Corder, G., Ahanonu, B., Grewe, B. F., Wang, D., Schnitzer, M. J., & Scherrer, G. (2019). An

    amygdalar neural ensemble that encodes the unpleasantness of pain. *Science*,

    363(6424), 276–281. https://doi.org/10.1126/science.aap8586

Durstewitz, D., & Seamans, J. K. (2008). The Dual-State Theory of Prefrontal Cortex Dopamine

    Function with Relevance to Catechol-*O*-Methyltransferase Genotypes and

    Schizophrenia. *Biological Psychiatry*, *64*(9), 739–749.

    https://doi.org/10.1016/j.biopsych.2008.05.015

Ellwood, I. T., Patel, T., Wadia, V., Lee, A. T., Liptak, A. T., Bender, K. J., & Sohal, V. S. (2017).

    Tonic or Phasic Stimulation of Dopaminergic Projections to Prefrontal Cortex Causes

    Mice to Maintain or Deviate from Previously Learned Behavioral Strategies. *Journal of*

    Neuroscience, *37*(35), 8315–8329. https://doi.org/10.1523/JNEUROSCI.1221-17.2017

Frost, N. A., Haggart, A., & Sohal, V. S. (2021). Dynamic patterns of correlated activity in the

    prefrontal cortex encode information about social behavior. *PLOS Biology*, *19*(5),

    e3001235. https://doi.org/10.1371/journal.pbio.3001235

Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of

    schizophrenia and depression. *Nature Reviews Neuroscience*, *17*(8), 524–532.

    https://doi.org/10.1038/nrn.2016.57

Granon, S., Passetti, F., Thomas, K. L., Dalley, J. W., Everitt, B. J., & Robbins, T. W. (2000).

    Enhanced and Impaired Attentional Performance After Infusion of D1 Dopaminergic

    Receptor Agents into Rat Prefrontal Cortex. *Journal of Neuroscience*, *20*(3), 1208–1215.

    https://doi.org/10.1523/JNEUROSCI.20-03-01208.2000

Howes, O. D., & Kapur, S. (2009). The Dopamine Hypothesis of Schizophrenia: Version III—The

    Final Common Pathway. *Schizophrenia Bulletin*, *35*(3), 549–562.

    https://doi.org/10.1093/schbul/sbp006

Lammel, S., Lim, B. K., & Malenka, R. C. (2014). Reward and aversion in a heterogeneous

    midbrain dopamine system. *Neuropharmacology*, *76*, 351–359.

    https://doi.org/10.1016/j.neuropharm.2013.03.019

Lammel, S., Lim, B. K., Ran, C., Huang, K. W., Betley, M. J., Tye, K. M., Deisseroth, K., &

    Malenka, R. C. (2012). Input-specific control of reward and aversion in the ventral

    tegmental area. *Nature*, *491*(7423), 212–217. https://doi.org/10.1038/nature11527

Lohani, S., Martig, A. K., Deisseroth, K., Witten, I. B., & Moghaddam, B. (2019). Dopamine

    Modulation of Prefrontal Cortex Activity Is Manifold and Operates at Multiple Temporal

    and Spatial Scales. *Cell Reports*, *27*(1), 99-114.e6.

    https://doi.org/10.1016/j.celrep.2019.03.012

Mukamel, E. A., Nimmerjahn, A., & Schnitzer, M. J. (2009). Automated analysis of cellular

    signals from large-scale calcium imaging data. *Neuron*, *63*(6), 747–760.

    https://doi.org/10.1016/j.neuron.2009.08.009

Pitman, R. K., Rasmusson, A. M., Koenen, K. C., Shin, L. M., Orr, S. P., Gilbertson, M. W.,

    Milad, M. R., & Liberzon, I. (2012). Biological studies of post-traumatic stress disorder.

    Nature Reviews Neuroscience, *13*(11), 769–787. https://doi.org/10.1038/nrn3339

Robinson, S. E., & Sohal, V. S. (1523). Dopamine D2 receptors modulate pyramidal neurons in

mouse medial prefrontal cortex through a stimulatory G protein pathway. *J. Neurosci*, *10*,

1893–17. https://doi.org/10.1523/JNEUROSCI.1893-17.2017

Santana, N., Mengod, G., & Artigas, F. (2009). Quantitative Analysis of the Expression of

Dopamine D1 and D2 Receptors in Pyramidal and GABAergic Neurons of the Rat

Prefrontal Cortex. *Cerebral Cortex*, *19*(4), 849–860.

https://doi.org/10.1093/cercor/bhn134

Sawaguchi, T., & Goldman-Rakic, P. S. (1991). D1 Dopamine Receptors in Prefrontal Cortex:

Involvement in Working Memory. *Science*, *251*(4996), 947–950.

https://doi.org/10.1126/science.1825731

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine

modulation in the prefrontal cortex. *Progress in Neurobiology*, *74*(1), 1–58.

https://doi.org/10.1016/j.pneurobio.2004.05.006

Sotoyama, H., Inaba, H., Iwakura, Y., Namba, H., Takei, N., Sasaoka, T., & Nawa, H. (2022).

The dual role of dopamine in the modulation of information processing in the prefrontal

cortex underlying social behavior. *The FASEB Journal*, *36*(2), e22160.

https://doi.org/10.1096/fj.202101637R

Vander Weele, C. M., Siciliano, C. A., Matthews, G. A., Namburi, P., Izadmehr, E. M., Espinel, I.

C., Nieh, E. H., Schut, E. H. S., Padilla-Coreano, N., Burgos-Robles, A., Chang, C.-J.,

Kimchi, E. Y., Beyeler, A., Wichmann, R., Wildes, C. P., & Tye, K. M. (2018). Dopamine

enhances signal-to-noise ratio in cortical-brainstem encoding of aversive stimuli. *Nature*,

563(7731), 397–401. https://doi.org/10.1038/s41586-018-0682-1

Zalocusky, K. A., Ramakrishnan, C., Lerner, T. N., Davidson, T. J., Knutson, B., & Deisseroth, K.

(2016). Nucleus accumbens D2R cells signal prior outcomes and control risky decision-

making. *Nature*, *531*(7596), 642–646. https://doi.org/10.1038/nature17400

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Marc Turner

4E9C3CF8DF2D481...   Author Signature

5/28/2024

Date

76