Title

Circuit inhibition facilitates behavioral strategy updates by creating divergent prefrontal representations for errors

Permalink

https://escholarship.org/uc/item/6122c6xz

Author

Johnson-Cruz, Carlos Allyn

Publication Date

2024

Peer reviewed|Thesis/dissertation

Circuit inhibition facilitates behavioral strategy updates by creating divergent prefrontal representations for errors
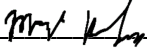
by
Carlos Allyn Johnson-Cruz

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

in

Neuroscience

in the

GRADUATE DIVISION
of the
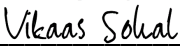UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Mazen Kheirbek
_____
Chair

Saul Kato
_____

Massimo Scanziani
_____

Vikaas Sohal
_____

_____
Committee Members

Circuit inhibition facilitates behavioral strategy updates by creating divergent prefrontal representations for errors

Carlos Allyn Johnson-Cruz

ABSTRACT

Adapting goal-directed behavior to changing environments (cognitive flexibility) depends on the mammalian prefrontal cortex. In mice, Parvalbumin-positive inhibitory interneurons (PVINs) in the medial prefrontal cortex (mPFC) support efficient behavioral adaption in a rule-shifting task (Cho et al., 2015, 2020). How PVINs modulate single-cell activity dynamics and information encoding in mPFC during cognitive flexibility is unclear. Heterozygous knockout of the $Dlx5/6^{+/-}$ transcription factor in mice ($Dlx5/6^{+/-}$ mice) disrupts rule-shifting performance and PVINs electrophysiology— administering a low-dose benzodiazepine (CLNZ) pre-task lastingly rescues these dysfunctions (Cho et al, 2015). We used one-photon calcium imaging in $Dlx5/6^{+/-}$ mice to study mPFC single-cell calcium activity before, during, and after pharmacological rescue. We found single-cell and population abnormalities in mPFC task-associated activity of $Dlx5/6^{+/-}$ mice, predominantly during initial rule learning. CLNZ treatment reversed these abnormalities.

During initial rule learning, pre-treatment $Dlx5/6^{+/-}$ mice mPFC cells had abnormally greater activity following error trials. Cells significantly active during error trials in the initial rule were also abnormally re-activated in all other task stages in $Dlx5/6^{+/-}$ mice. Machine learning classification found increased activity structure similarity during error trials within different rules for pre-treatment $Dlx5/6^{+/-}$ mice. CLNZ treatment lastingly reduced abnormally high initial rule error activity and decreased reactivation of initial rule error cells in other task stages. CLNZ increases the linear separability of mPFC population activity between errors in different rule blocks. These results demonstrate a novel role of PV+ cells in differentiating mPFC neuron activity from previous network states during behavior adaptation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CLNZ- Clonazepam

dlPFC- dorsolateral PFC

Dlx5/6- Transcription factors *Dlx5* and *Dlx6*

GABA- gamma-aminobutyric acid

HET- *Dlx5/6$^{+/-}$* heterozygous knockout mice

IA- Initial Association (1st rule of task session)

MGE- medial ganglionic eminence

mPFC- medial Prefrontal Cortex

mRNA- messenger ribonucleic acid (RNA)

PCA- Principal Component Analysis

PFC- Prefrontal Cortex

PV- Parvalbumin

PVIN- Parvalbumin+ Interneuron

RS- Rule Shift (2nd rule of task session)

TTC- # of trials to complete a rule period

WT- Wild-type mice (*Dlx5/6$^{+/+}$*)

Chapter 1 - General Introduction

1.1 The PFC transforms multi-model input to facilitate cognitive flexibility

The mammalian prefrontal cortex (PFC) integrates sensory, motor, and limbic inputs from cortical and sub-cortical brain regions, generating abstract representations to guide complex processes like executive function and working memory (Miller & Cohen, 2001). Individual PFC neurons in humans, non-human primates, and rodents encode sensory inputs and abstract concepts like decision confidence or reward value (Fusi et al., 2016; Hirokawa et al., 2019; Karlsson et al., 2012; Abe & Lee, 2011; Jamali et al., 2019). This mixed-response selectivity promotes efficient and flexible computation (Tye et al., 2024).

In humans, damage to the PFC impairs executive function, and decreased cognitive acuity is associated with PFC hypoactivity (Milner, 1963; Miller & Cohen, 2001; Pantelis et al., 1999). The ability to appraise changing circumstances and alter goal-directed behavior in response, or cognitive flexibility, is a critical component of executive function (Dajani & Uddin, 2015). Essential for navigating fluidly changing environments, the severity of cognitive flexibility impairment in humans correlates with worse life outcomes (Bowie & Harvey, 2006). In humans, cognitive flexibility depends on dorsolateral PFC (dlPFC) processing (Ferguson and Gao 2018, Waltz., et al. 2017). The rodent mPFC functions akin to the human dlPFC, processing the information necessary for cognitive flexibility (Seamans et al., 2008; Narayanan et al., 2013; Croxson et al., 2014; Ferguson & Gao, 2018). Individual neurons in rodent mPFC encode cognitive variables relevant to adapting behavior to changing environments (Durstewitz et al., 2010; Rikhye et al., 2018). Disrupting mouse mPFC activity impaired mouse cognitive flexibility across multiple experimental paradigms (Marton et al., 2018; Cho et al., 2016).

1.2 Parvalbumin+ cells shape cortical activity

mPFC function requires a tuned balance of excitatory and inhibitory activity (Ferguson & Gao, 2018; Hu et al., 2014). Disrupting this equilibrium causes impairment ranging from seizures to perturbed information encoding (Marin, 2012; Hu et al., 2014; Yizhar et al., 2011). Parvalbumin-positive inhibitory interneurons (PVINs) are critical components of the mPFC microcircuit (Ferguson & Gao, 2018). Parvalbumin is a calcium-binding protein that modulates calcium decay dynamics when expressed in neurons (Collin et al., 2005). The low-input resistance and fast-spiking capabilities of PVINs facilitate high-frequency feedforward inhibition of large numbers of targeted excitatory pyramidal cells, thus tuning general cortical circuit activity in multiple ways (Ferguson & Gao, 2018; Packer & Yuste et al., 2011). For example, PVINs expand the local dynamic range of activity and suppress non-meaningful activity correlations between cortical neurons (Graupner & Reyes, 2013; Cardin, 2018; Pouille et al., 2009). PVINs activated by an ensemble of excitatory neurons can sparsify circuit activity by inhibiting other less-active cells in a 'winner-takes-all' manner (Hu et al., 2014). PVINs also support broader processing-related phenomena like gamma frequency (30-100 Hz) oscillations in activity, which support cross-region information transfer, are abnormal in specific psychiatric disorders, and promote efficient behavioral adaptation (Cho et al., 2020; Mathalon & Sohal, 2015; Sohal, 2016; Sohal et al., 2009; Cardin et al., 2009).

1.3 Deficits in Schizophrenia suggest PVINs support cognitive flexibility

The psychiatric disorder schizophrenia affects approximately 1% of the global population (Velligan & Rao, 2023). Symptoms include delusions and hallucinations, treatable with anti-psychotic medications (Bowie & Harvey, 2006). These medications only minimally improve schizophrenia's cognitive symptoms, which affect up to 80% of patients and lack targeted FDA-approved treatments (Maroney, 2022; Bowie & Harvey, 2006). Cognitive flexibility is one of the 1.5 most severely disrupted faculties, and a patient's degree of cognitive impairment strongly

correlates with their ability to live independently (Bowie & Harvey, 2006; Waltz et al., 2017; Leeson et al., 2009; Caeser et al., 2008).

The "GABAergic hypothesis of schizophrenia" postulates that PVIN dysfunction caused by the disorder may impair cognition (Sohal, 2024). Post-mortem human studies observed average amounts of PV-expressing PFC neurons but decreased per-neuron expression of PV mRNA (Hashimoto et al., 2003; Dienel et al., 2023). PVIN-dependent gamma oscillations are also muted (Senkowski & Gallinat, 2015). Rodent studies demonstrate the causal roles of PFC PVINs in cognitive flexibility. Multiple mouse models of disrupted PVIN development demonstrated cognitive flexibility deficits (Canetta et al., 2016, 2022; Caballero et al., 2020).

1.4 How PVINs modulate neural activity in cognitive flexibility is ill-studied

Little is known about how PVINs shape single-neuron selectivity and information processing in mPFC. Some general evidence of how PVINs shape cortical single-neuron activity exists: in the visual cortex, inhibiting PVINs increased neural response similarity to different visual stimuli (Agetsuma et al., 2018). A recent study probed PV+ cell roles in mPFC single-cell activity during cognitive flexibility. Inhibiting a population of contralaterally-projecting mPFC PVINs increased the similarity of population activity between errors in distinct rule blocks of the task and muted normal post-error activity increases (Cho et al., 2023). Still unknown, though, are the broader activity dynamics occurring while organisms engage in cognitive flexibility, as well as how PVINs shape these dynamics.

The following chapters will describe findings from experiments performed in mutant mice with abnormal PVIN function, recording mPFC neural activity before, during, and after pharmacological restoration of cognitive flexibility. This bi-directional impairment and rescue provide a unique perspective into how modest PVIN function restoration alters mPFC network activity and information encoding. This work elucidates the causal role of PVINs in the mouse

mPFC complex processing during cognitive flexibility. It serves as a first step towards demonstrating how PVIN dysfunction directly causes cognitive deficits.

1.5 References

Abe, H., & Lee, D. (2011). Distributed Coding of Actual and Hypothetical Outcomes in the Orbital and Dorsolateral Prefrontal Cortex. *Neuron*, *70*(4), 731–741. https://doi.org/10.1016/j.neuron.2011.03.026

Agetsuma, M., Hamm, J. P., Tao, K., Fujisawa, S., & Yuste, R. (2018). Parvalbumin-Positive Interneurons Regulate Neuronal Ensembles in Visual Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *28*(5), 1831–1845. https://doi.org/10.1093/cercor/bhx169

Bartolo, R., Saunders, R. C., Mitz, A., & Averbeck, B. B. (2019). Dimensionality, information and learning in the prefrontal cortex. *bioRxiv*, 823377. https://doi.org/10.1101/823377

Bowie, C. R., & Harvey, P. D. (2006). Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatric Disease and Treatment*, *2*(4), 531. https://doi.org/10.2147/nedt.2006.2.4.531

Caballero, A., Flores-Barrera, E., Thomases, D. R., & Tseng, K. Y. (2020). Downregulation of parvalbumin expression in the prefrontal cortex during adolescence causes enduring prefrontal disinhibition in adulthood. *Neuropsychopharmacology*, *45*(9), 1527–1535. https://doi.org/10.1038/s41386-020-0709-9

Canetta, S., Bolkan, S., Padilla-Coreano, N., Song, L. J., Sahn, R., Harrison, N. L., Gordon, J. A., Brown, A., & Kellendonk, C. (2016). Maternal immune activation leads to selective functional deficits in offspring parvalbumin interneurons. *Molecular Psychiatry*, *21*(7), 956–968. https://doi.org/10.1038/mp.2015.222

Canetta, S. E., Holt, E. S., Benoit, L. J., Teboul, E., Sahyoun, G. M., Ogden, R. T., Harris, A. Z., & Kellendonk, C. (2022). Mature parvalbumin interneuron function in prefrontal cortex requires activity during a postnatal sensitive period. *eLife*, *11*, e80324. https://doi.org/10.7554/eLife.80324

Cardin, J. A. (2018). Inhibitory Interneurons Regulate Temporal Precision and Correlations in Cortical Circuits. *Trends in Neurosciences*, *41*(10), 689–700. https://doi.org/10.1016/j.tins.2018.07.015

Ceaser, A. E., Goldberg, T. E., Egan, M. F., McMahon, R. P., Weinberger, D. R., & Gold, J. M. (2008). Set-Shifting Ability and Schizophrenia: A Marker of Clinical Illness or an Intermediate Phenotype? *Biological Psychiatry*, *64*(9), 782–788. https://doi.org/10.1016/j.biopsych.2008.05.009

Cho, K. K. A., Davidson, T. J., Bouvier, G., Marshall, J. D., Schnitzer, M. J., & Sohal, V. S. (2020). Cross-hemispheric gamma synchrony between prefrontal parvalbumin interneurons supports behavioral adaptation during rule shift learning. *Nature Neuroscience*, *23*(7), 892–902. https://doi.org/10.1038/s41593-020-0647-1

Cho, K. K. A., Hoch, R., Lee, A. T., Patel, T., Rubenstein, J. L. R., & Sohal, V. S. (2015). Gamma Rhythms Link Prefrontal Interneuron Dysfunction with Cognitive Inflexibility in Dlx5/6+/− Mice. *Neuron*, *85*(6), 1332–1343. https://doi.org/10.1016/j.neuron.2015.02.019

Cho, K. K. A., Shi, J., Phensy, A. J., Turner, M. L., & Sohal, V. S. (2023). Long-range inhibition synchronizes and updates prefrontal task activity. *Nature*, *617*(7961), Article 7961. https://doi.org/10.1038/s41586-023-06012-9

Croxson, P. L., Walton, M. E., Boorman, E. D., Rushworth, M. F. S., & Bannerman, D. M. (2014). Unilateral medial frontal cortex lesions cause a cognitive decision-making deficit in rats. *European Journal of Neuroscience*, *40*(12), 3757–3765. https://doi.org/10.1111/ejn.12751

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003

Dienel, S. J., Fish, K. N., & Lewis, D. A. (2023). The Nature of Prefrontal Cortical GABA Neuron Alterations in Schizophrenia: Markedly Lower Somatostatin and Parvalbumin Gene Expression Without Missing Neurons. *American Journal of Psychiatry*, *180*(7), 495–507. https://doi.org/10.1176/appi.ajp.20220676

Durstewitz, D., Vittoz, N. M., Floresco, S. B., & Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, *66*(3), 438–448. https://doi.org/10.1016/j.neuron.2010.03.029

Enomoto, T., Tse, M. T., & Floresco, S. B. (2011). Reducing prefrontal gamma-aminobutyric acid activity induces cognitive, behavioral, and dopaminergic abnormalities that resemble schizophrenia. *Biological Psychiatry*, *69*(5), 432–441. https://doi.org/10.1016/j.biopsych.2010.09.038

Enwright Iii, J. F., Huo, Z., Arion, D., Corradi, J. P., Tseng, G., & Lewis, D. A. (2018). Transcriptome alterations of prefrontal cortical parvalbumin neurons in schizophrenia. *Molecular Psychiatry*, *23*(7), 1606–1613. https://doi.org/10.1038/mp.2017.216

Ferguson, B. R., & Gao, W.-J. (2018). PV Interneurons: Critical Regulators of E/I Balance for Prefrontal Cortex-Dependent Behavior and Psychiatric Disorders. *Frontiers in Neural Circuits*, *12*. https://doi.org/10.3389/fncir.2018.00037

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*, 66–74. https://doi.org/10.1016/j.conb.2016.01.010

Haider, B., Duque, A., Hasenstaub, A. R., & McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(17), 4535–4545. https://doi.org/10.1523/JNEUROSCI.5297-05.2006

Hamm, J. P., Peterka, D. S., Gogos, J. A., & Yuste, R. (2017). Altered Cortical Ensembles in Mouse Models of Schizophrenia. *Neuron*, *94*(1), 153-167.e8. https://doi.org/10.1016/j.neuron.2017.03.019

Han, S., Tai, C., Westenbroek, R. E., Yu, F. H., Cheah, C. S., Potter, G. B., Rubenstein, J. L., Scheuer, T., de la Iglesia, H. O., & Catterall, W. A. (2012). Autistic-like behaviour in Scn1a+/- mice and rescue by enhanced GABA-mediated neurotransmission. *Nature*, *489*(7416), 385–390. https://doi.org/10.1038/nature11356

Hashimoto, T., Volk, D. W., Eggan, S. M., Mirnics, K., Pierri, J. N., Sun, Z., Sampson, A. R., & Lewis, D. A. (2003). Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of subjects with schizophrenia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(15), 6315–6326. https://doi.org/10.1523/JNEUROSCI.23-15-06315.2003

Hirokawa, J., Vaughan, A., Masset, P., Ott, T., & Kepecs, A. (2019). Frontal cortex neuron types categorically encode single decision variables. *Nature*, *576*(7787), 446–451. https://doi.org/10.1038/s41586-019-1816-9

Hu, H., Gan, J., & Jonas, P. (2014). Fast-spiking, parvalbumin+ GABAergic interneurons: From cellular design to microcircuit function. *Science*, *345*(6196), 1255263. https://doi.org/10.1126/science.1255263

Jamali, M., Grannan, B., Haroush, K., Moses, Z. B., Eskandar, E. N., Herrington, T., Patel, S., & Williams, Z. M. (2019). Dorsolateral prefrontal neurons mediate subjective decisions and their variation in humans. *Nature Neuroscience*, *22*(6), 1010–1020. https://doi.org/10.1038/s41593-019-0378-3

Kharawala, S., Hastedt, C., Podhorna, J., Shukla, H., Kappelhoff, B., & Harvey, P. D. (2022). The relationship between cognition and functioning in schizophrenia: A semi-systematic review. *Schizophrenia Research: Cognition*, *27*, 100217. https://doi.org/10.1016/j.scog.2021.100217

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis— Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Lee, A. T., Gee, S. M., Vogt, D., Rubenstein, J. L., & Sohal, V. S. (2014). Pyramidal neurons in prefrontal cortex receive subtype-specific forms of excitation and inhibition. *Neuron*, *81*(1), 61–68. https://doi.org/10.1016/j.neuron.2013.10.031

Leeson, V. C., Robbins, T. W., Matheson, E., Hutton, S. B., Ron, M. A., Barnes, T. R. E., & Joyce, E. M. (2009). Discrimination Learning, Reversal, and Set-Shifting in First-Episode Schizophrenia: Stability Over Six Years and Specific Associations with Medication Type and Disorganization Syndrome. *Biological Psychiatry*, *66*(6), 586–593. https://doi.org/10.1016/j.biopsych.2009.05.016

Marín, O. (2012). Interneuron dysfunction in psychiatric disorders. *Nature Reviews Neuroscience*, *13*(2), 107–120. https://doi.org/10.1038/nrn3155

Maroney, M. (2022). Management of cognitive and negative symptoms in schizophrenia. *The Mental Health Clinician*, *12*(5), 282–299. https://doi.org/10.9740/mhc.2022.10.282

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

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

MILNER, B. (1963). Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. *Archives of Neurology*, *9*(1), 90–100. https://doi.org/10.1001/archneur.1963.00460070100010

Namboodiri, V. M. K., Otis, J. M., Heeswijk, K. van, Voets, E. S., Alghorazi, R. A., Rodriguez-Romaguera, J., Mihalas, S., & Stuber, G. D. (2019). Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nature Neuroscience*, 1. https://doi.org/10.1038/s41593-019-0408-1

Narayanan, N. S., Cavanagh, J. F., Frank, M. J., & Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature Neuroscience*, *16*(12), 1888–1895. https://doi.org/10.1038/nn.3549

Packer, A. M., & Yuste, R. (2011). Dense, Unspecific Connectivity of Neocortical Parvalbumin-Positive Interneurons: A Canonical Microcircuit for Inhibition? *Journal of Neuroscience*, *31*(37), 13260–13271. https://doi.org/10.1523/JNEUROSCI.3131-11.2011

Pantelis, C., Barber, F. Z., Barnes, T. R. E., Nelson, H. E., Owen, A. M., & Robbins, T. W. (1999). Comparison of set-shifting ability in patients with chronic schizophrenia and frontal lobe damage. *Schizophrenia Research*, *37*(3), 251–270. https://doi.org/10.1016/S0920-9964(98)00156-X

Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V., & Scanziani, M. (2009). Input normalization by global feedforward inhibition expands cortical dynamic range. *Nature Neuroscience*, *12*(12), 1577–1585. https://doi.org/10.1038/nn.2441

Powell, N. J., & Redish, A. D. (2016). Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12830

Rich, E. L., & Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(22), 7208–7219. https://doi.org/10.1523/JNEUROSCI.6068-08.2009

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Rigotti, M., Ben Dayan Rubin, D. D., Wang, X.-J., & Fusi, S. (2010). Internal Representation of Task Rules by Recurrent Dynamics: The Importance of the Diversity of Neural Responses. *Frontiers in Computational Neuroscience*, *4*. https://doi.org/10.3389/fncom.2010.00024

Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature Neuroscience*, *21*(12), 1753–1763. https://doi.org/10.1038/s41593-018-0269-z

Seamans, J. K., Lapish, C. C., & Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: Insights from electrophysiology. *Neurotoxicity Research*, *14*(2), 249–262. https://doi.org/10.1007/BF03033814

Senkowski, D., & Gallinat, J. (2015). Dysfunctional prefrontal gamma-band oscillations reflect working memory and other cognitive deficits in schizophrenia. *Biological Psychiatry*, *77*(12), 1010–1019. https://doi.org/10.1016/j.biopsych.2015.02.034

Sohal, V. S. (2024). Neurobiology of schizophrenia. *Current Opinion in Neurobiology*, *84*, 102820. https://doi.org/10.1016/j.conb.2023.102820

Tse, M. T., Piantadosi, P. T., & Floresco, S. B. (2015). Prefrontal cortical gamma-aminobutyric acid transmission and cognitive function: Drawing links to schizophrenia from preclinical research. *Biological Psychiatry*, *77*(11), 929–939. https://doi.org/10.1016/j.biopsych.2014.09.007

Tye, K. M., Miller, E. K., Taschbach, F. H., Benna, M. K., Rigotti, M., & Fusi, S. (2024). Mixed selectivity: Cellular computations for complexity. *Neuron*, *0*(0). https://doi.org/10.1016/j.neuron.2024.04.017

Velligan, D. I., & Rao, S. (2023). The Epidemiology and Global Burden of Schizophrenia. *The Journal of Clinical Psychiatry*, *84*(1), 45094. https://doi.org/10.4088/JCP.MS21078COM5

Waltz, J. A. (2017). The Neural Underpinnings of Cognitive Flexibility and their Disruption in Psychotic Illness. *Neuroscience*, *345*, 203–217. https://doi.org/10.1016/j.neuroscience.2016.06.005

Wang, Y., Dye, C. A., Sohal, V., Long, J. E., Estrada, R. C., Roztocil, T., Lufkin, T., Deisseroth, K., Baraban, S. C., & Rubenstein, J. L. R. (2010). Dlx5 and Dlx6 regulate the development of parvalbumin-expressing cortical interneurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(15), 5334–5345. https://doi.org/10.1523/JNEUROSCI.5963-09.2010

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 2 - CLNZ resolves mPFC single-cell activity abnormalities in Dlx5/6+/- mice

2.1 Introduction

Much remains unknown about how PVINs shape information encoding in mPFC neurons. Multiple studies indicate that the mPFC region requires normal PVIN development for efficient cognitive flexibility in rodents. Broadly inhibiting PFC PVINs during a critical post-natal period induced cognitive deficits that were reversed by stimulating PFC PVIN stimulation in adulthood (Canetta et al., 2022). Optogenetic inhibition of the mPFC region PVINs in adult mice impaired cognitive flexibility (Canetta et al., 2016).

This study uses a transgenic mouse line, the $Dlx5/6^{+/-}$ mouse, to study how disrupting PVIN function alters mPFC single-cell activity. This mutant mouse line has previously been used to study how PVINs function in cognitive flexibility. The medial ganglionic eminence (MGE) contains genes regulating the development of most cortical interneurons (Wang et al., 2010). Within the MGE, $Dlx5$ and $Dlx6$ regulate post-natal development of PV+ interneurons (Wang et al., 2010). Heterozygous expression of $Dlx5/6$ in mice ($Dlx5/6^{+/-}$ mice) results in abnormal PVIN electrophysiology but does not significantly alter expression (Cho et al., 2015; Wang et al., 2010). These alterations reduce the magnitude of downstream inhibition in pyramidal neurons during the high-frequency stimulation of upstream interneurons (Cho et al., 2015). These PVIN deficits cause decreased mPFC cross-hemisphere gamma frequency synchrony when $Dlx5/6^{+/-}$ mice engage in cognitive flexibility (Cho et al., 2015, 2020).

Multiple treatments targeting mPFC PVINs reduced cognitive inflexibility in $Dlx5/6^{+/-}$ mice. General optogenetic stimulation of mPFC interneurons in $Dlx5/6^{+/-}$ mice improved cognitive flexibility and the strength of gamma oscillations (Cho et al., 2015). A subsequent study provided

targeted cross-hemispheric mPFC PVIN stimulation, restoring cognitive flexibility (Cho et al., 2020). Pharmacological treatment also improved cognitive flexibility in these mutant mice. Administering the benzodiazepine clonazepam (CLNZ) to these mice restores gamma synchrony and improves cognitive inflexibility (Cho et al., 2020). Clonazepam allosterically modulates GABAa receptors, increasing the magnitude of post-synaptic responses to GABA release but not directly activating GABAa receptors (Han et al., 2012). Finally, both clonazepam and synchronous cross-hemisphere stimulation resulted in lasting improvements in cognitive flexibility when mutant mice were tested multiple days following the acute treatment (Cho et al., 2020).

Altogether, *Dlx5/6*$^{+/-}$ mice provide a within-subjects comparison of the effect of abnormal and boosted GABAergic signaling on network activity in mPFC during cognitive flexibility. We used micro-endoscopic calcium imaging in these mutant mice to observe mPFC single-cell activity dynamics before, during, and after a lasting rescue of cognitive flexibility in mutant mice. We compared it to activity in wild-type controls. We find that single cells in *Dlx5/6*$^{+/-}$ mouse mPFC pre-treatment are significantly more active in most task stages than cells in wild-type mice. The most considerable difference in activity between wild-type and mutant mice pre-treatment was observed after error trials during initial rule learning (IA). Cells significantly active in these error trials also re-activate in other task stages at significantly greater rates than wild-type mice. This re-activation decreased during and after acute CLNZ treatment, suggesting that cognitive inflexibility can result from abnormal re-activation of cells modulated by errors during initial rule learning and that PVIN function is crucial for preventing this.

2.2 Materials and Methods

Animal care and surgical protocol

Guidelines from both the University of California San Francisco's Administrative Panels on Laboratory Animal Care and the National Institutes of Health were followed for animal care. Mice received ad libitum access to food and water and were housed in temperature control conditions with a 12/12 light/dark cycle. *Dlx5/6*$^{+/-}$ mice were generated as described in Cho et al., 2015 and maintained on a mixed CD1 and C57Bl/6 background. All experiments were performed with *Dlx5/6*$^{+/-}$ mice and their age-matched wild-type littermates. Mice underwent stereotactic surgery to implant the virus and lens in preparation for mPFC micro-endoscopic imaging. While anesthetized with isoflurane, mice were injected with 600 μL of AAV9-Syn-jGcAMP7f (1:3 dilution, Addgene) at four depths in mPFC (+1.7 (AP), 0.3 (ML), and -2.75, -2.5, -2.25, -2.0 (DV) in Bregma relative mm). 0.5 mm diameter GRIN lenses (Inscopix Inc) were implanted 0.3mm above the uppermost injection site, and at least four weeks elapsed before behavioral experiments to allow sufficient time for viral expression.

'Rule shifting' task tests cognitive flexibility in mice

Our lab's cognitive flexibility task (or rule-shifting task) has been described previously (Cho et al. 2015, 2020, 2023) but is summarized below. At the start of each trial, mice are moved from a 'rest cage' into the adjacent testing cage containing two small bowls. Each bowl contains a combination of one of two possible odors and one of two possible textures, meaning each trial presents two of four possible odor/texture combinations. A food reward (a peanut butter chip) is buried in bowls based on a rule, where a specific odor or texture is selected as the stimulus cue indicating the presence of a food reward (e.g., a bowl with sand will contain a food reward) (**Fig. 2.1A**).

We categorized trials as 'error' or 'correct' based on the mouse's choice. If mice dug in a bowl for three cumulative seconds, we considered that bowl to be the mouse's choice. Trials were categorized errors if mice chose the non-rewarded bowl. In error trials, the reward bowl was removed immediately after a bowl choice; mice were allowed to dig in the non-reward bowl for at least 15 seconds. After both types of trials, mice were moved to the rest cage after they stopped digging in their chosen bowl (and after eating food reward during correct trials). During error trials, mice were kept in the rest cage for an extended inter-trial interval.

Each testing session is divided into the Initial Association (IA) rule period and the Rule shift (RS). During the IA period, mice first learn an initial rule in sensory modality 1 (e.g., that the bowl with sand will always contain a reward). Once the mouse achieves 80% correct performance over ten consecutive trials, it is deemed to have learned the IA rule. Mice then perform three more trials before the experiment enters the RS phase, and the rule changes to the *other* sensory modality (e.g., from sand signifying reward presence to coriander odor being the signal) (figure 5). This rule change is called the rule shift (**Fig. 2.1B**).

Before being tested on the rule-shifting task, mice are single-housed and food-deprived for several days, with food placed inside the bowls that will later be used to contain reward during the main rule-shifting task. They then undergo a habituation session of 10 trials where reward placement is randomly assigned to one bowl. This teaches mice that only one bowl has a food reward in the testing environment.

Single-cell resolution calcium imaging in mouse mPFC via 1-photon microendoscopy during the rule shifting task

We recorded from the implanted GRIN lens of mice during two consecutive days of them performing the rule-shifting task. On day 2, mice received an intraperitoneal (i.p.) injection (volume: 0.01 ml/g) of clonazepam (at 0.0625 mg/kg) 30 minutes before starting the task (on

day one, mice received i.p Saline injections). This dosage causes no significant locomotor

effects and is sub-sedative and sub-anxiolytic while improving task performance in *Dlx5/6*[+/-]

mice (Cho et al., 2015; Han et al., 2012). After two days of consecutive testing, we waited at

least seven days to guarantee CLNZ cleared, then imaged HET mice (the post-CLNZ condition).

A head-mounted one-photon microendoscope (nVoke2, Inscopix Inc.) records fluorescence

detected from an implanted GRIN lens in mPFC at 20 Hz. This raw fluorescence time-series

data is spatially down-sampled, then spatial bandpass filtered (with a high cutoff of 0.03 pixel-1

and a low cutoff of 0.008 pixel^{-1}) and motion corrected (**Fig. 2.1H**).

ROI detection and time-series creation

Putative ROIs labeling neurons in the recording were found using the EXTRACT algorithm (Inan

et al., 2021). ROIs labeled by EXTRACT were first manually sorted to remove false-positive

cells, then filtered with a custom spatial deduplication algorithm, which leveraged the distribution

of correlations between spatially distant ROIs to remove false-positive cells (**Fig. 2.1H**). When

ROIs were spatially adjacent, we iteratively removed the least active cell within that spatial

radius until no adjacent cells were more correlated than the 95th percentile of spatially distant

cell pairs. Finally, this processed time series was converted to a binary raster of calcium events

via a custom threshold-based event-detection algorithm (Frost et al., 2021).

Ensemble permutation testing

We used permutation testing to identify ensembles of cells significantly activated above chance

in particular stages of the task (e.g., Late RS). We created a null distribution by circularly

shuffling binary event matrices and quantifying mean activity in those shuffled datasets during

the frames in the task stage of interest. We repeat this 300 times to create null distributions of

mean activity per neuron. We then identify those cells active above the 95th percentile of the

distribution of shuffled data's mean activity during that stage.

Behavioral analysis

Video recordings were captured from two vantage points, with one camera mounted above the test cage and one mounted on the test cage(Anymaze). This recording was annotated for behavioral events (e.g., dig start) at a 1-second resolution. Timestamps for behaviors were then synced to the binary raster of calcium events using TTL pulses logged in the calcium data, occurring at the initiation of the behavioral recording. Timestamps were sorted into the pre-decision (time points after the mouse enters the test cage but before the bowl choice), post-decision (time points after the mouse chooses a bowl but before the mouse is removed from the test cage), and inter-trial interval (time points where the mouse is in the rest cage) corresponding to each trial of the recording session.

The analysis focused on the last 5 seconds of the pre-decision before the bowl choice and the first 15 seconds following the mouse's choice of the bowl. Because mice had varying numbers of trials necessary to reach the criterion on each rule, the analysis focused on the first or last five trials in each rule, referred to as the *Early* or *Late* stages. The *Early* stage was further divided into *Error* and *Correct* groups of trials, resulting in a total of 3 unique stages per each rule, yielding six total task stages (**Fig. 2.1B**). Because the late trials occurred when mice demonstrated knowledge of the rule and achieved the preset criteria, most late trials were correct trials. The analysis excluded any Error trials if they were present in the Late stage.

Quantification and statistical analysis

Statistical analysis was performed in MATLAB and Python. The details of the analysis are provided in the main text. We categorize each day's recording into groups designated as a combination of the subject's genotype abbreviation (WT or HET) and the treatment on that day's abbreviation (VEH, CLNZ, and postCLNZ) and refer to this label as the *genotype-treatment*. We analyzed WT-VEH, HET-VEH, HET-CLNZ, and HET-postCLNZ recordings. We excluded HET-

CLNZ from subsequent analysis, as datasets were not fully processed at the time of this publication. Analysis of calcium activity was restricted to the post-decision section, which consisted of the frames in each trial between the start of a mouse digging in the chosen bowl and the mouse being removed from the test cage.

2.3 Results

CLNZ treatment improves cognitive flexibility in HET mice

Following previous studies, *Dlx5/6$^{+/-}$* (HET) mice had impaired rule-shifting performance. (WT N = 7, HET N = 6. *2-way ANOVA on Perseverative errors*: genotype: F=41.9, p = 2.75e^{-7}, treatment day: F = 19.0, p = 3.56e^{-6}, treatment day x genotype: F = 5.17, p = 0.0113. *2-way ANOVA on RS TTC*: genotype: F=129.8, p = 8.48e^{-13}, treatment day: F = 104.1, p = 9.88e^{-15}, treatment day x genotype: F = 19.8, p =2.55e^{-6}) HET mice receiving vehicle injections (HET-VEH) were not impaired in learning the initial rule of a session (IA), but made significantly more perseverative errors during the rule-shift period (selecting the non-reward bowl with a cue that would have contained reward during the first rule), requiring significantly more trials to learn (TTC) the second rule (*Post-hoc 2-sample t-tests* on RS TTC: WT VEH vs. Het VEH: p= 4.38e^{-4}, Het VEH vs. CLNZ: p= 2.59e^{-3}, Het VEH vs. postCLNZ: p= 9.62e^{-3}) (**Fig. 2.1D, E, F**). CLNZ decreased the number of trials required for HET mice to reach 80% accuracy in the rule shift and decreased the number of perseverative errors they made during the rule shift (*Post-hoc 2-sample t-tests* on RS Perseverative errors: WT VEH vs. Het VEH: p= 2.78e^{-3}, Het VEH vs. CLNZ: p= 0.027) (**Fig. 2.1F**).

HET cells activated significantly more in all task stages

Single-cell activity in WT and HET mice was sparse during trials—on average, only 30% of neurons activated during trials. When time series were averaged across all trials in each task stage, WT mice showed similar variation in average peak activation time regardless of the task

stage (**Fig. 2.2B**). The general pattern of activity did not vary by stage. Sorting by the mean activation of all cells for each task stage revealed that cells preferentially activate most in one task stage and activate less strongly in others (**Fig. 2.2C**). Pre-rescue HET-VEH mice had significantly higher levels of single-cell activity than WT mice in Early IA Error and Late RS. CLNZ treatment significantly decreased mean single-cell activity in HET mice in Early IA Correct, Early RS Error, Late IA, and Late RS. Following the 7+ day washout period, HET-postCLNZ cell activity remained significantly lower than HET-VEH levels in Early RS Correct, Early RS Error, Late IA, and Late RS. (2-way ANOVA: *genotype-treatment*, $F=18.3$, $p=7.3e^{-12}$; *stage*: $p=6.3e^{-14}$, $F=14.2$; *genotype-treatment x stage*: $p = 2.29e^{-23}$, $F=9.7$) (*Post-hoc Mann-Whitney U Test* (B-Y FDR correction)- *Early IA Correct*: HET-VEH vs. CLNZ, $p = 3.09e^{-9}$; *Early IA Error*: WT-VEH vs. HET-VEH, $p = 5.348e^{-4}$; *Early RS Correct*: HET-VEH vs. postCLNZ, $p = 1.59e^{-3}$; *Early RS Error*: HET-VEH vs. CLNZ, $p = 1.8e^{-5}$, HET-VEH vs postCLNZ, $p = 2.46e^{-20}$; *Late IA*: HET-VEH vs. CLNZ, $p = 1.66e^{-4}$, HET-VEH vs postCLNZ, $p= 6.6e^{-3}$; *Late RS*: WT-VEH vs. HET-VEH, $p = 0.024$, HET-VEH vs. CLNZ, $p = 6.19e^{-5}$, HET-VEH vs postCLNZ, $p = 2.21e^{-7}$) (**Fig. 2.2D**).

HET cells demonstrated highly correlated activity between task stages, unlike WT

Calculating the mean of Pearson's correlation between cell activity in different task stage pairs of interest, we found a significant effect of genotype-treatment and stage comparison but no interaction factor (2-way ANOVA: *genotype-treatment*, $F=17.7$, $p=1.8e^{-11}$; *stage pair*: $p=8.6e-38$, $F=36.9$; *genotype-treatment x stage*: $p=0.71$, $F=0.7$). The subject-level correlation of cell activation between different task stages was significantly lower in WT mice compared to HET-VEH mice (*Posthoc t-tests*- WT-VEH vs. HET-VEH: IA Correct vs. RS Correct $p=0.011$; IA Error vs. RS Error $p=0.021$; IA Correct vs. IA Error $p=0.007$; RS Correct vs. RS Error $p=0.038$). While CLNZ decreased this correlation, it was not statistically significant at the subject level. (**Fig. 2.2E**).

The activity of cells enriched in each stage was significantly higher in HET mice

We used permutation testing to find the number of cells in each genotype-treatment population that were active above the 95th percentile of their corresponding null shuffled distribution. We refer to the group of cells significantly active in (or *enriched* in) a task stage as that stage's *ensemble*. We observed no significant effect of treatment on the number of cells enriched for each category, suggesting that the overall proportion of cells significantly active in each condition is unaffected by either genotype or treatment (*Chi^2 statistic*: 21.5, p=0.121) (**Fig. 2.4A**). Despite the overall number of cells significantly active per stage not being affected, we observed effects on the magnitude of the significant activation in task stages (*2-way ANOVA*: *genotype-treatment* p = 4.39e^{-3}, F=4.38; *stage* p= 3.6e^{-12}, F 12.6; *genotype-treatment x stage*: p=1.7e^{-4}, F= 2.87).

When examining only the ensembles of cells significantly active in each task stage, we observed that CLNZ decreased activity in HET Early IA Error ensemble cells from levels observed in the HET-VEH condition. Activity in Early IA Error ensemble cells stayed significantly lower in HET-postCLNZ (*Post-hoc Mann-Whitney U Test (B-Y FDR correction)*: Early IA Error HET-VEH vs. CLNZ, p=1.8e^{-5}; HET-VEH vs. postCLNZ, p = 0.0024) (**Fig. 2.4B, C, D**). Early RS Correct enriched cells in HET-VEH had significantly lower activity ensemble than WT-Veh, but this was not reversed by CLNZ treatment (*Post-hoc Mann-Whitney U Test (B-Y FDR correction)*- Early RS Correct WT-VEH vs. HET-VEH, p=0.0076) (**Fig. 2.4B**).

HET mice demonstrated altered patterns of co-enrichment between different stages

We quantified the # of cells enriched in 2 different task stages for all 15 unique stage combinations. We then used Fisher's exact test to quantify if the cells significantly active in both stages overlapped more than expected by chance. WT mice showed no significant overlap in cells activated during error in both rule periods, but HET-VEH showed significant overlap in cells

significantly active in both error stages. This overlap disappeared in HET-CLNZ mice and reappeared after CLNZ washout in the postCLNZ condition (*one-sided Fisher's Exact Test, (BH FDR correction)*- Early IA Error & Early RS Error: WT-VEH p = 0.59, HET-VEH p = 0.024, HET-CLNZ p = 0.085; HET-CLNZ, p = 0.029) (**Fig 2.4F**).

Early IA Error ensemble over-activity re-occurred in other stages

Given that Early IA Error ensemble over-activity was the only phenomenon significantly reversed by CLNZ treatment, we probed whether that ensemble's abnormally high activity spilled over into other task stages. Studying the activity of Early IA Error ensemble neurons, a significant effect of genotype-treatment and stage on the mean ensemble activation was observed, but there was not a significant interaction (*2-way ANOVA*: *genotype-treatment*, p= $8.89e^{-11}$, F=16.8; *ensemble stage*, p= $3.57e^{-89}$, F 93.6; *genotype-treatment x ensemble stage*: p= 0.053, F= 1.65). Early IA Error ensemble activation in all other stages significantly decreased in HET-CLNZ mice. This decrease remained significantly lower in the HET-postCLNZ condition for Early IA Error, Early RS Error and Late RS stages (*Post-hoc testing with Mann-Whitney U Test (B-Y FDR correction)*- *Early IA Correct*: HET-VEH vs. CLNZ, p = 0.0061; *Early IA Error*: HET-VEH vs. CLNZ, p = $1.8e^{-5}$, HET-VEH vs. postCLNZ, p = 0.0024; *Early RS Error*: HET-VEH vs. CLNZ, p = 0.000041, HET-VEH vs postCLNZ, p = 0.0003; *Late RS*: HET-VEH vs. CLNZ p = 0.0034, HET-VEH vs. postCLNZ, p = 0.0027) (**Fig. 2.5C**).

2.4 Discussion

Dlx5/6[+/-] mice had significantly elevated activity in cells enriched after IA rule errors. These enriched cells were re-activated in other task stages at magnitudes significantly higher than wild-type. Low-dose CLNZ significantly decreased the number of errors mutant mice made in trials during the second rule (RS), demonstrating improved cognitive flexibility. CLNZ treatment also decreased abnormally elevated activity, consistent with expected increased circuit

inhibition. In HET mice, CLNZ diminished the overlap of cell enrichment between IA and RS errors. CLNZ decreased the magnitude of activity of Early IA Error enriched cells in other stages, most notably in Early RS Error and Late RS. This abnormally elevated ensemble activity in Early IA Error, Early RS Error, and Late RS remained low after CLNZ cleared. The decreased reactivation of Early IA Error cells in RS Error and Late RS that occurs in HET-CLNZ mice alongside more efficient RS learning suggests that abnormal reactivation of cells significantly active after initial rule errors can cause cognitive inflexibility.

Our results suggest that a task-selective ensemble's activation in other stages contributes more to cognitive flexibility than whether ensemble members are only enriched in a single task stage. WT cells were not more likely than expected by chance to activate after both IA Errors and RS Errors. By contrast, HET-VEH mice showed more cells enriched in both conditions than expected by chance. This disappeared during CLNZ treatment and reappeared post-CLNZ. Because this significant overlap reappears in the cognitively flexible HET-postCLNZ mice, we hypothesize that IA Error ensemble activation *magnitude* in RS Error trials is more important than whether individual neurons in IA Error ensembles are also enriched in RS Error. Our results suggest that efficient cognitive flexibility depends on neural population activity more than single neurons. Chapter 3 will explore CLNZ's effect on neural populations during rule shifting. This study provides evidence to suggest that efficiently engaging cognitive flexibility requires rapid differentiation of neural population activity supported by frontal cortex inhibitory function.

An important caveat to our results is our inability to isolate the effect of interneuron deficits on mPFC only. *Dlx5/6* heterozygosity broadly disrupts cortical PV+ interneuron development (Wang et al., 2010). While interventions only targeting mPFC PVINs in *Dlx5/6*$^{+/-}$ mutants restore cognitive flexibility (Cho et al., 2020), it is possible that our results did not directly result from mPFC interneurons modulating mPFC information processing but rather from indirectly restoring gamma oscillation facilitation of inter-region communication. Future studies can address this by

combining one-photon calcium imaging with targeted mPFC manipulations, e.g., stimulating only mPFC PVINs in *Dlx5/6*[+/-] mice or inhibiting PV cell activity in healthy adult mice.

That a single acute CLNZ injection before mice perform rule shifting can lastingly improve task performance is striking. However, the breadth of cognitive improvement, as well as the specific mechanisms, remain unclear. Various mPFC PVIN populations facilitate cognitive flexibility, but whether CLNZ's pro-cognitive effects arise from modulating inhibition from specific PVIN subpopulations remains unclear. Inhibiting contra-laterally projecting PVINs increased mPFC activity similarity between RS and IA errors during rule shifting (Cho et al., 2023). During cognitive flexibility in head-fixed mice, putative inhibitory interneurons required mediodorsal thalamus input that encoded environmental rules to suppress irrelevant mPFC representations efficiently (Rikhye et al., 2018). Our results support these findings, demonstrating that non-relevant representations are abnormally activated at inappropriate times when PVIN inhibition is impaired. PVIN-dependent activity differentiation between different task stages could be driven by one, both, or other yet unknown groups of mPFC PVINs. Despite this knowledge gap, this study demonstrates that PVINs facilitate differentiating activity as contingencies change in different rule phases. When PVINs are dysfunctional, impaired cognitive flexibility is partly caused by abnormal re-activation of cells associated with previous strategies.

Whether the improvement in cognition is restricted to this task remains unknown, as is whether CLNZ improves cognitive flexibility in other experimental paradigms and if the improvement continues indefinitely. Answers to these questions will illuminate whether the pro-cognitive phenomena described in this study could be used for human interventions in psychiatric disorders involving impaired cognitive flexibility with co-morbid impaired PVIN function. Human trials of benzodiazepines in patients with schizophrenia demonstrated a moderate improvement in cognitive symptoms (Lewis et al., 2008).
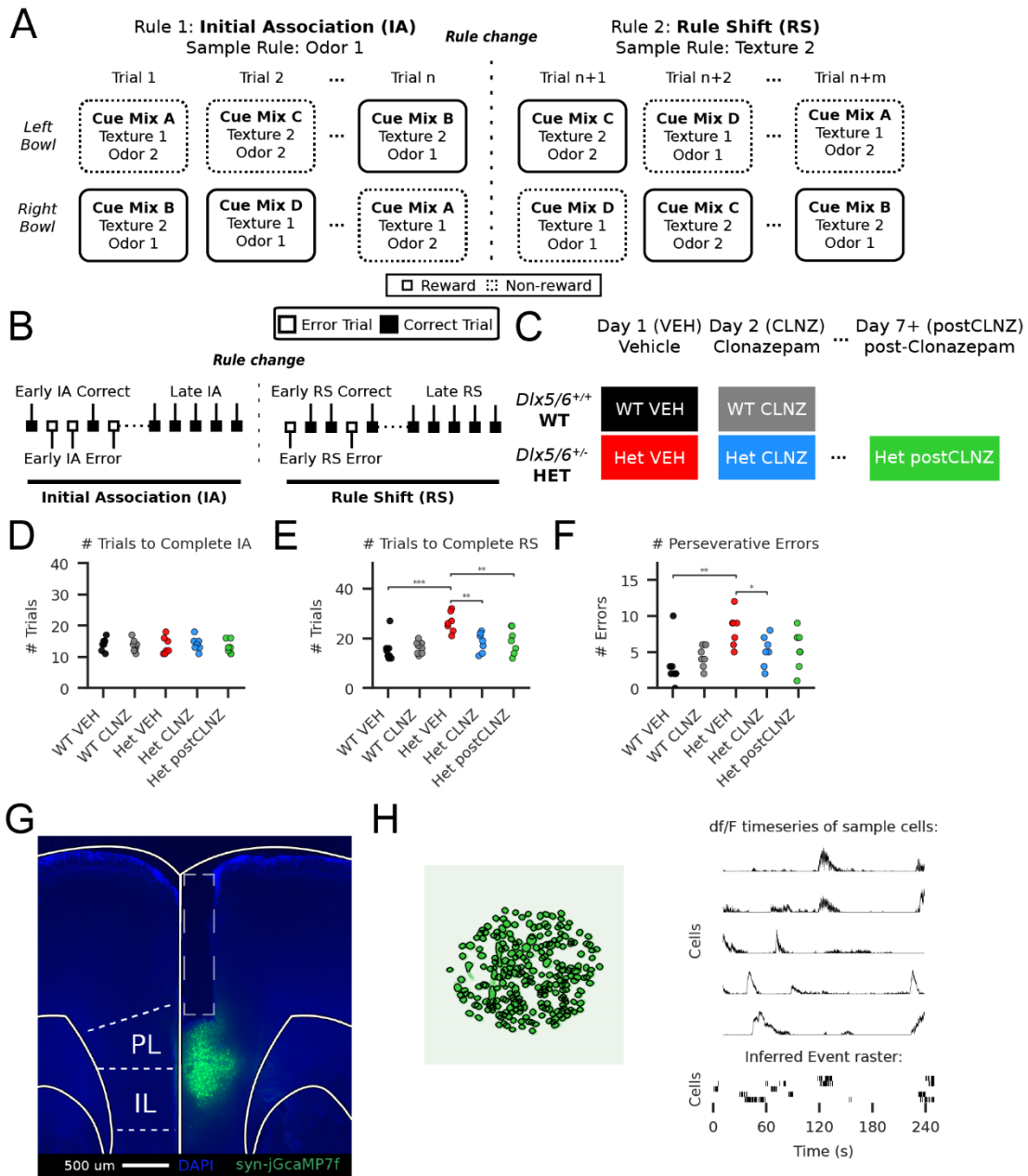
2.5 Figures

Figure 2-1 CLNZ decreases errors made and trials to learn RS in HET mice

a) Schematic showing sample trials from a rule-shifting session. In each trial, a rule governs reward assignment to one of two bowls in the test cag, selecting one possible sensory cue out of 2 odors and two textures. While the rule is in effect, the bowl with the chosen cue always contains a reward. Mice must learn two rules to complete a session. In this diagram, the first rule ("Initial Association"/IA) is Odor 1, so the bowl with Odor 1 always has food reward in IA trials. After mice reach behavioral criteria, the rule changes to a cue of the other sensory modality (the rule shift). Trials within a given task session are grouped into 1 of 6 stages. *(Figure caption continued on next page)*

(Figure caption continued from the previous page) Each rule's first and last five trials are broken into early and late periods. The early period is further categorized into correct and error trials. While rare, if errors occurred during the last five trials of a rule, they were excluded from grouped analysis.

b) Experimental design of drug treatment. Both WT and mutant (HET) mice performed the rule-shifting task after receiving vehicle (VEH) injections and clonazepam (CLNZ) injections on subsequent days. HET mice were tested again 7+ days after CLNZ injection to evaluate performance post-clearout.

c) HET and WT mice take similar numbers of trials to learn the first rule in a task session.

d) HET mice receiving vehicle injections take substantially more trials than WT mice to learn the second rule (RS), but this is reversed when HET mice receive CLNZ.

e) HET mice receiving vehicle injections commit substantially more perseverative errors (where the cues in the bowl erroneously selected were previously the rewarded cues in the IA) trials than WT mice during the second rule (RS). This is reversed when HET mice receive CLNZ.

f) Representative image demonstrating implant location in the mPFC and expression of syn-jGCaMP7f (green) amidst cell bodies stained with DAPI (blue).

g) *Left*: Representative regions of interest (putative cells) detected during a rule-shifting session. *Upper Right*: representative df/F calcium traces from a subset of ROIs. Lower Right: heatmap of calcium events identified from corresponding above calcium traces
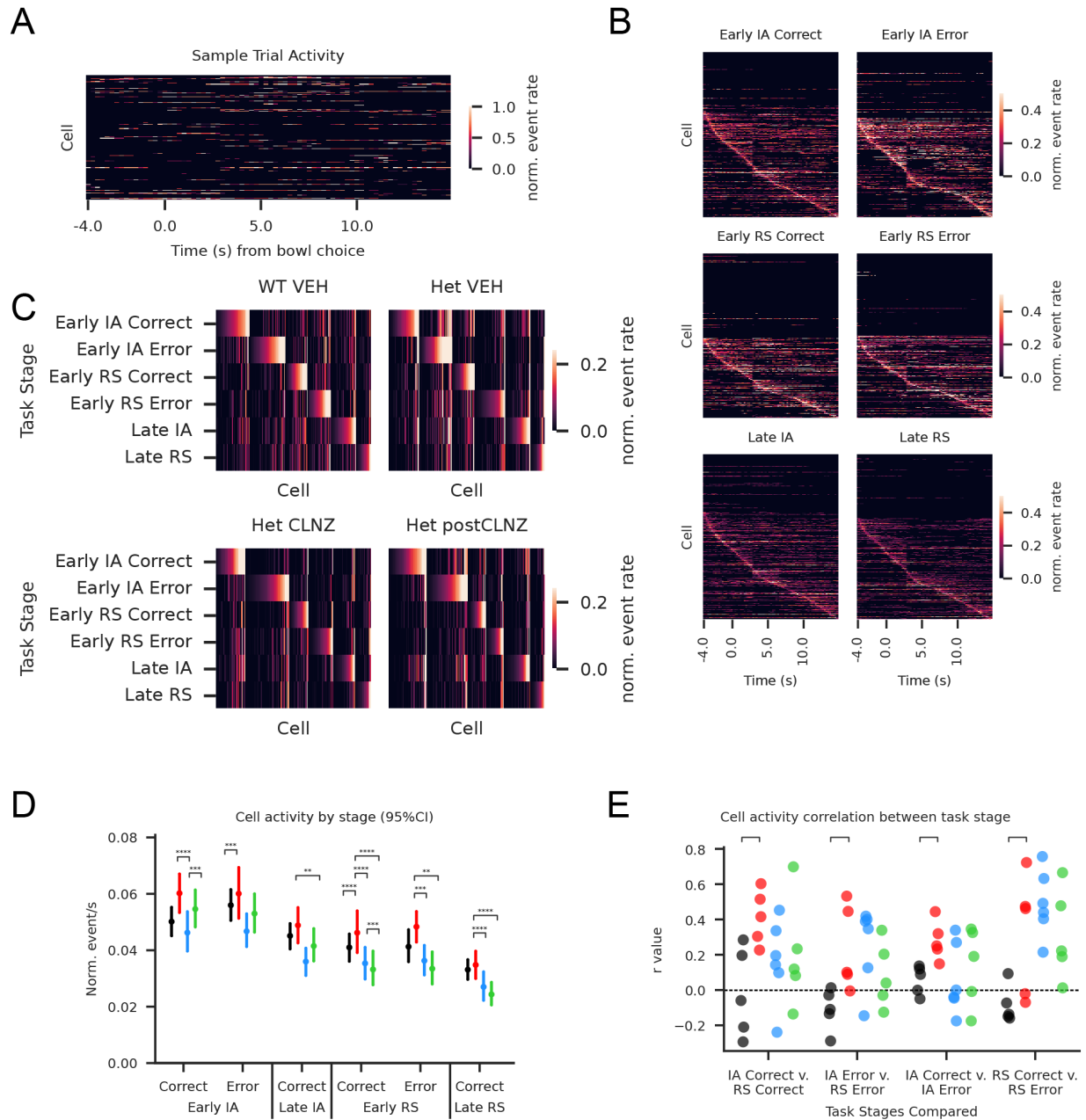
Figure 2-2 CLNZ treatment decreases pre-treatment over-activity in HET mice

a) Representative heatmap of cell activity over time during a trial
b) WT cell averaged activity time series by task stage. Combined across mice
c) Average event rate of each cell in each task stage, tiled by genotype-treatment
d) Mean (+/- SEM) of each genotype-treatment's cell event rates, grouped by task stage.
e) Per-Subject Pearson's correlation of event rates between different trial types (all groups taken from the first five trials of each rule/each early division)
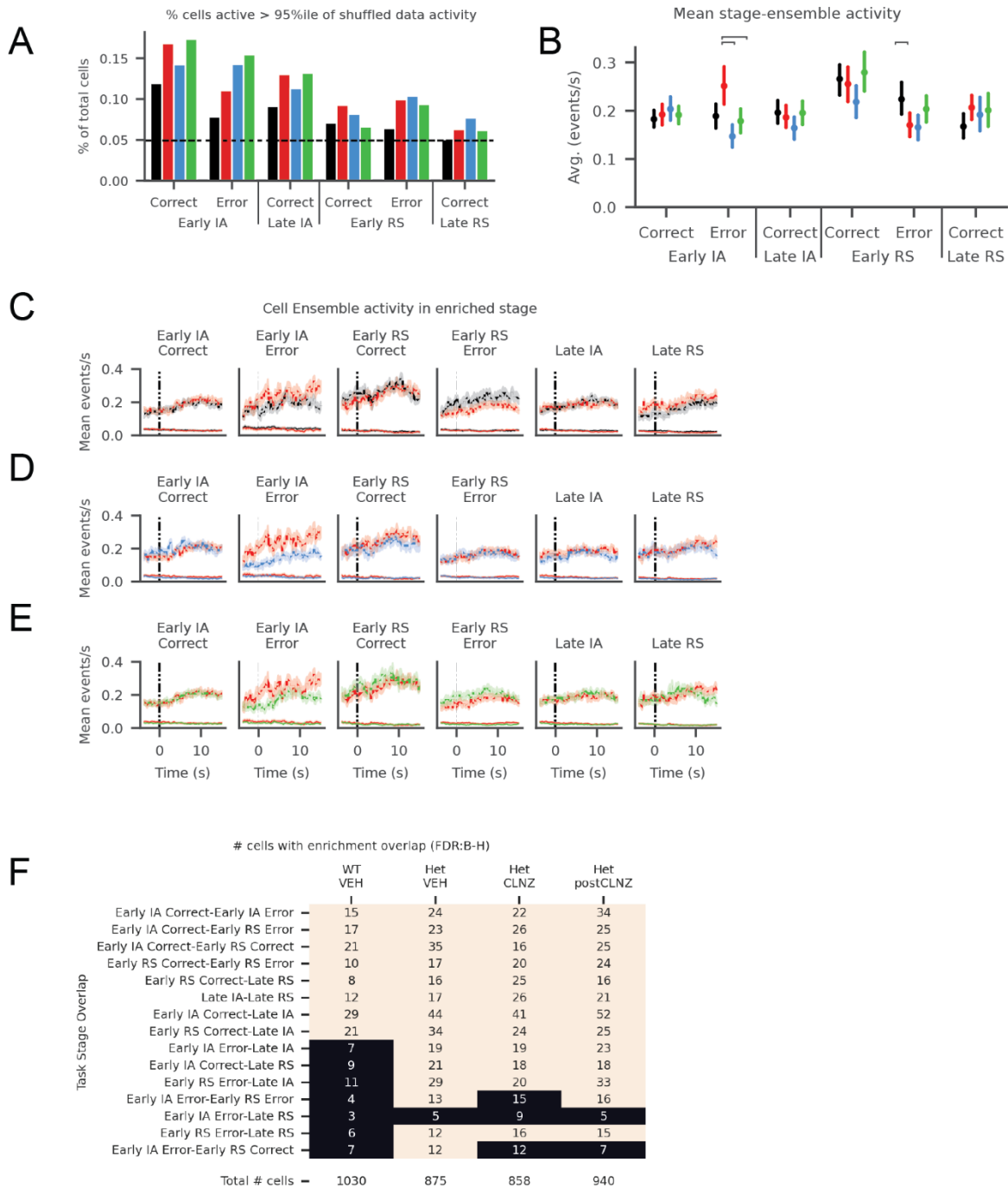
Figure 2-3 CLNZ treatment decreases abnormal Early IA Error activity in HETs

a) % of cells more active than the 95th percentile of shuffled data null distribution (enriched)
b) Mean (+/- SEM) activity of all cells significantly active in a given stage
c) WT-VEH (black) and HET-VEH (red): mean time-series of cells enriched in stage
d) HET-VEH (red) and HET-CLNZ (blue): mean time-series of cells enriched in stage
e) HET-VEH (red) and HET-postCLNZ (green): mean time-series of cells enriched in stage
f) # of cells significantly enriched in both stages by genotype-treatment
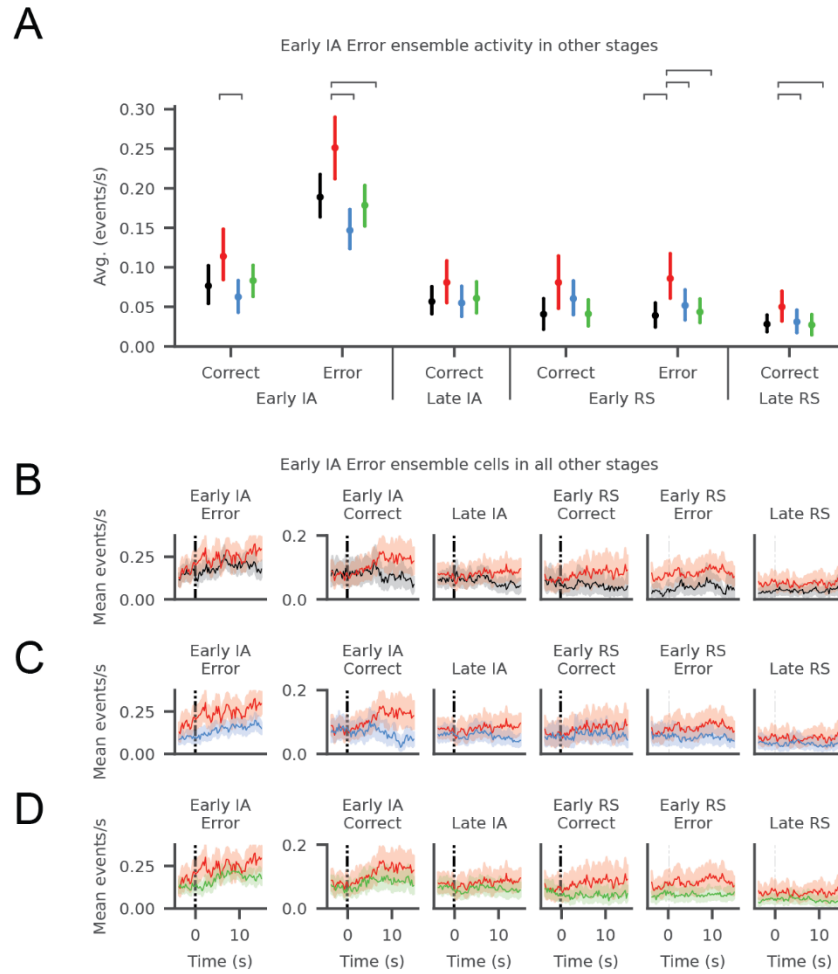
Figure 2-4 CLNZ decreases Early IA Error ensemble re-activation in RS for HETs

a) Mean (+/- SEM) activity in all stages of cells significantly active in Early IA Error
b) WT-VEH (black) and HET-VEH (red): mean time-series of Early IA Error enriched cells in all other stages
c) HET-VEH (red) and HET-CLNZ (blue): mean time-series of Early IA Error enriched cells in all other stages
d) HET-VEH (red) and HET-postCLNZ (green): mean time-series of Early IA Error enriched cells in all other stages

2.6 References

Caballero, A., Flores-Barrera, E., Thomases, D. R., & Tseng, K. Y. (2020). Downregulation of parvalbumin expression in the prefrontal cortex during adolescence causes enduring prefrontal disinhibition in adulthood. *Neuropsychopharmacology*, *45*(9), 1527–1535. https://doi.org/10.1038/s41386-020-0709-9

Canetta, S., Bolkan, S., Padilla-Coreano, N., Song, L. J., Sahn, R., Harrison, N. L., Gordon, J. A., Brown, A., & Kellendonk, C. (2016). Maternal immune activation leads to selective functional deficits in offspring parvalbumin interneurons. *Molecular Psychiatry*, *21*(7), 956–968. https://doi.org/10.1038/mp.2015.222

Canetta, S. E., Holt, E. S., Benoit, L. J., Teboul, E., Sahyoun, G. M., Ogden, R. T., Harris, A. Z., & Kellendonk, C. (2022). Mature parvalbumin interneuron function in prefrontal cortex requires activity during a postnatal sensitive period. *eLife*, *11*, e80324. https://doi.org/10.7554/eLife.80324

Cho, K. K. A., Hoch, R., Lee, A. T., Patel, T., Rubenstein, J. L. R., & Sohal, V. S. (2015). Gamma Rhythms Link Prefrontal Interneuron Dysfunction with Cognitive Inflexibility in Dlx5/6+/− Mice. *Neuron*, *85*(6), 1332–1343. https://doi.org/10.1016/j.neuron.2015.02.019

Cho, K. K. A., Shi, J., Phensy, A. J., Turner, M. L., & Sohal, V. S. (2023). Long-range inhibition synchronizes and updates prefrontal task activity. *Nature*, *617*(7961), Article 7961. https://doi.org/10.1038/s41586-023-06012-9

Enomoto, T., Tse, M. T., & Floresco, S. B. (2011). Reducing prefrontal gamma-aminobutyric acid activity induces cognitive, behavioral, and dopaminergic abnormalities that resemble schizophrenia. *Biological Psychiatry*, *69*(5), 432–441. https://doi.org/10.1016/j.biopsych.2010.09.038

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

Han, S., Tai, C., Westenbroek, R. E., Yu, F. H., Cheah, C. S., Potter, G. B., Rubenstein, J. L.,

Scheuer, T., de la Iglesia, H. O., & Catterall, W. A. (2012). Autistic-like behaviour in Scn1a+/-

mice and rescue by enhanced GABA-mediated neurotransmission. *Nature*, *489*(7416), 385–

390. https://doi.org/10.1038/nature11356

Hirokawa, J., Vaughan, A., Masset, P., Ott, T., & Kepecs, A. (2019). Frontal cortex neuron types

categorically encode single decision variables. *Nature*, *576*(7787), 446–451.

https://doi.org/10.1038/s41586-019-1816-9

Hu, H., Gan, J., & Jonas, P. (2014). Fast-spiking, parvalbumin+ GABAergic interneurons: From

cellular design to microcircuit function. *Science*, *345*(6196), 1255263.

https://doi.org/10.1126/science.1255263

Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., & Seamans, J. K. (2012).

Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proceedings of the

National Academy of Sciences*, *109*(13), 5086–5091. https://doi.org/10.1073/pnas.1114415109

Inan, H., Schmuckermair, C., Tasci, T., Ahanonu, B. O., Hernandez, O., Lecoq, J., Dinç, F.,

Wagner, M. J., Erdogdu, M. A., & Schnitzer, M. J. (n.d.). *Fast and statistically robust cell

extraction from large-scale neural calcium imaging datasets*. 83.

Lewis, D. A., Cho, R. Y., Carter, C. S., Eklund, K., Forster, S., Kelly, M. A., & Montrose, D.

(2008). Subunit-selective modulation of GABA type A receptor neurotransmission and cognition

in schizophrenia. *The American Journal of Psychiatry*, *165*(12), 1585–1593.

https://doi.org/10.1176/appi.ajp.2008.08030395

Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature Neuroscience*, *21*(12), 1753–1763. https://doi.org/10.1038/s41593-018-0269-z

Riss, J., Cloyd, J., Gates, J., & Collins, S. (2008). Benzodiazepines in epilepsy: Pharmacology and pharmacokinetics. *Acta Neurologica Scandinavica*, *118*(2), 69–86. https://doi.org/10.1111/j.1600-0404.2008.01004.x

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in the prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–956. https://doi.org/10.1038/35082081

Wang, Y., Dye, C. A., Sohal, V., Long, J. E., Estrada, R. C., Roztocil, T., Lufkin, T., Deisseroth, K., Baraban, S. C., & Rubenstein, J. L. R. (2010). Dlx5 and Dlx6 regulate the development of parvalbumin-expressing cortical interneurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(15), 5334–5345. https://doi.org/10.1523/JNEUROSCI.5963-09.2010

Chapter 3 – CLNZ differentiates mPFC network activity

3.1 Introduction

The previous chapter demonstrated that PVIN dysfunction in *Dlx5/6*[+/-] mice causes elevated activity in cells significantly active during initial rule errors (Early IA Error enriched cells). Early IA Error-enriched cells in pre-treatment *Dlx5/6*[+/-] mice (HET-VEH) also significantly activated more than similarly enriched WT-VEH cells for all task stages. Administering low-dose benzodiazepine Clonazepam (CLNZ) reduces the activity of Early IA Error enriched cells and their abnormal reactivation across other task stages. Whether mPFC population dynamics are being more broadly reshaped beyond the level of single-cell activity is unknown. This chapter will summarize evidence suggesting that complex mPFC population dynamics encode information beyond the single-cell level and will present results indicating that PVIN dysfunction also affects population encoding.

High-dimensional PFC activity occurs during cognition in primates and rodents

The 'population doctrine' postulates that groups of neurons modulated by multiple inputs sustain representations more efficiently than single neurons (Saxena & Cunningham, 2019; Ebitz & Hayden, 2021). Mixing diverse neural responses increases the dimensionality of activity in networks of neurons (Tye et al., 2024). High dimensional population activity aids in solving complex tasks by facilitating network movement between distinct patterns of network activity (subspaces) that simplifies potential downstream processing (Rigotti et al., 2010; Dubreuil et al., 2022; Tye et al., 2014).

High-dimensional network activity enables the primate and rodent PFC to perform complex tasks like cognitive control (Tye et al., 2014). Prefrontal cortex neurons in non-human primates demonstrate high-dimensional activity and neural mixed selectivity during tasks engaging cognitive processes like working memory and decision-making during reinforcement learning

(Murray et al., 2016; Chiang et al., 2022; Bartolo et al., 2020; Rigotti et al., 2013). Rat mPFC activity moves to distinct network subspaces as rats adapt behavioral strategies to reach rewards (Durstewitz et al., 2010). These network transitions preceded changes in goal-directed behavior, even if identical behaviors occur while following different strategies (Powel & Redish, 2016; Rich & Shapiro, 2009).

Contributions of PVIN activity to population dynamics are under-studied

Chapter 2 demonstrated that PVIN dysfunction caused lasting changes in single-cell activity in mutant mice that were reversed through a low-dose benzodiazepine. The effects of this dysfunction on high-dimensional mPFC population dynamics are unknown. More broadly, how PVINs generally shape information encoding in rodent mPFC during cognitive flexibility is unclear. A recent study by our lab partially addressed this question by inhibiting contra-laterally projecting PVINs in mPFC. Inhibition of their terminals impaired cognitive flexibility and increased the similarity of population activity between Early RS Error and Early RS Correct stages (Cho et al., 2023)**.**

This chapter investigates how PVIN dysfunction in *Dlx5/6$^{+/-}$* mice alters high-dimensional mPFC population dynamics. Machine learning classifiers can classify samples by using high-dimensional dynamics in dataset features (Rigotti et al., 2013). We leverage machine learning classifiers and vector similarity analysis to study high-dimensional activity patterns in wild-type mice and abnormalities observed in mutant mice before, during, and after CLNZ treatment. We discovered that population activity in HET mice was significantly more similar between Early IA Error and Early RS error than in WT mice. This abnormally high degree of activity similarity was reduced after CLNZ treatment. We also found that machine learning classifiers less accurately distinguish Early IA Errors and Early RS Errors in *Dlx5/6$^{+/-}$* mice pre-treatment. This decoding deficit is most prominent when classifying with HET cells enriched in Early RS Error or Early RS Correct stages. This deficit was lastingly reversed with CLNZ treatment. These results suggest

that PVINs support the differentiation of activity between errors in two different rules, a potentially crucial component in engaging cognitive flexibility.

3.2 Materials and Methods

Animal Subjects

The datasets analyzed in Chapter 3 are the same datasets described in Chapter 2. For details on data collection, animal subjects, and the rule-shifting task, see Chapter 2.

Pseudopopulation Creation

To combine cells from various subjects into a single time series, we create a *pseudopopulation* consisting of all cells from a genotype-treatment significantly active in a given stage. As subjects can have different numbers of trials in each task section (e.g., Mouse 1 has 2 Early IA Errors, but Mouse 2 has 1), we resample with replacement to create a new bootstrapped time series for each cell. We combine those individual cell bootstrap samples across subjects into each ensemble pseudopopulation. For a cell A significantly active in stage N, we randomly sample with replacement from all trials that belong to stage N to create a new time series for cell A. We oversample to 400 time-bins (the maximum number of time-bins present in 5 trials, given 250ms time-bins and a maximum trial length of 15 seconds) to ensure all cells had equal frames. We perform this for all cells in the stage N ensemble. We then resample stage N ensemble cell activity from all other stages. This process repeats for all ensembles, all task stages, and all genotype-treatment groups. The stage of origin for each of these datasets is then decoded (linearly classified). We repeat this bootstrapping process 300 times and analyze those distributions of decoding accuracies.

Population Similarity Analysis

We calculated the cosine similarity between the vector of mean activity per cell in each task stage to quantify task stage similarity. Cosine similarity is defined as the dot product of vector A

and vector B, divided by the product of their vector norms. We performed this once per pseudopopulation creation iteration (300 total iterations) and performed statistical analysis on these distributions.

Decoding analysis (SVM)

We used a support vector machine (SVM) linear classifier to classify pseudopopulation frames into 1 of 2 possible task stages per classifier run. As this classification outputs binary classifications, we manually specified a list of pairs of task stages to distinguish between. SVMs using L2 regularization and 5-fold cross-validation (CV) pipelines were built using Python and the scikit-learn module (Pedregosa et al., 2011). For each pair of task stages being differentiated, we generate the resampled activity matrices of the ensemble of interest in each stage. We concatenated the two resampled task stage activity matrices for the ensemble pseudopopulation of interest. We removed cells that were not active in one of the two task stages classified. In each CV fold, SVMs were trained on 50% of the concatenated activity matrix, randomly selected, and tested on the remaining 50%. SVM accuracy was quantified by averaging accuracy over each CV fold for a pseudopopulation iteration (5 folds per iteration, 300 total iterations).

PCA projection

Principal component analysis (PCA) projects matrices onto a set of new orthogonal vectors (Principal Components, PCs) comprised of linear combinations of the original matrix dimensions, capturing the directions of greatest variance in the dataset in successive PCs. We take one random pseudopopulation of the 300 iterations created for decoding and project it into PC space, using the first two principal components for visualization (Pedregosa et al., 2011).

3.3 Results

CLNZ decreases population activity similarity of Early IA Errors and Early RS Errors

We first sampled the activity of a given ensemble pseudopopulation from each task stage, then averaged across frames to create a single vector of the ensemble's mean activity per task stage. We then found the cosine similarity between ensemble activity vectors for different task stages (**Fig. 3.1A**). Because Chapter 2 demonstrated that HET-VEH cells enriched in Early IA Error were abnormally reactivated in Early RS Error, we hypothesized that population activity would be similar between those two stages. We thus quantified the cosine similarity of population activity between those trials.

Significant individual and interaction effects were observed for genotype-treatment and ensembles on the mean cosine similarity between Early IA Error and Early RS Error activity. (*2-way ANOVA*: genotype-treatment: F=206455, $p < 1.05e^{-99}$, ensemble: F = 28264, $p < 1.05e^{-99}$, genotype-treatment x ensemble: F = 15691.8, $p < 1.05e^{-99}$). Het-VEH activity was most similar between Early RS and Early IA Error for all cell ensembles (**Fig. 3.1A**). The greatest gap in cosine similarity between genotype-treatment groups was between WT and HET for the Early RS Correct and Early RS Error ensemble activity vectors (all pairwise Mann-Whitney tests were significant after Benjamini-Yekutileli correction; p-values $< 1.05e^{-99}$) (**Fig 3.1A**). CLNZ administration significantly reduced the similarity of activity for all ensembles (**Fig 3.1A**). WT mice population activity similarity was much lower than HET-VEH, HET-CLNZ, and HET-postCLNZ mice in all conditions.

CLNZ increases decoder discrimination between Early IA Errors and Early RS Errors

High SVM accuracy for binary class decoding means the two classes are easily linearly separable. Mean accuracy is almost 100% for decoding Early IA Error from Early RS Error frames for all WT-VEH ensemble pseudopopulations, indicating major differences in ensemble

activity between the two task stages (**Fig 3.1B**). HET-VEH mice had much lower mean decoding

accuracy than WT-VEH for Early RS Error and Early RS Correct ensembles (**Fig 3.1B**).

Decoding accuracy significantly increased in HET mice following CLNZ injection and after CLNZ

washout (*pairwise Mann-Whitney tests, Benjamini-Yekutieli correction*: Early RS Correct

ensemble- WT VEH vs. HET VEH, p = $1.01e^{-99}$. Early RS Correct ensemble- HET VEH vs.

CLNZ, p = $9.8e^{-100}$. Early RS Correct ensemble: HET VEH vs. postCLNZ, p = $9.9e^{-100}$).

Results do not depend on the amount of ensemble cells enriched in classified stages

If specific ensembles tend to have cells enriched in one or both of the classified stages, this

could affect decoding accuracy and bias results. To evaluate whether this was a factor, we

tabulate how many cells (if any) in each ensemble are significantly enriched in either one of the

two stages being classified for each SVM run decoding between a pair of stages (e.g.,

classifying Early IA Error or Early RS Error activity from the Late RS Ensemble). We then

quantified the relationship between the proportion of cells enriched in the classified stages and

the decoding accuracy. We found that the proportion of cells enriched in at least 1 class being

decoded was not significantly correlated with SVM accuracy (*Ordinary Least Squares*

Regression coefficients and p-values: WT VEH: 0.159, p=0.0167; Het VEH: -0.096, p=0.17; Het

CLNZ: 0.05, p=0.023; Het postCLNZ: -0.01, p=0.77. p-values not significant after Bonf.

Correction).

3.4 Discussion

Summary of Results

Our results indicate that PVIN dysfunction in *Dlx5/6$^{+/-}$* mice, beyond altering single-cell activity,

also alters population activity dynamics and information encoding in mPFC. Early RS Correct

and Early RS Error ensembles had high activity similarity during errors made during different

rules (Early IA Error and Early RS Error). This was significantly higher than in WT mice,

significantly decreased in HET-CLNZ, and stayed low after CLNZ washout. Early IA Error and Early RS Error activity of the Early RS Correct and Early RS Error ensembles in HET-CLNZ and HET-postCLNZ mice did not become as dissimilar as in WT-VEH ensembles. This suggests that relative changes in activity similarity are more important than absolute changes, as HET cortical activity is still likely abnormal even after boosting post-synaptic responses to GABA with CLNZ.

Linear classifiers demonstrate the functional importance of relative dissimilarity between errors in different rules. Linear classification less accurately distinguished activity in Early IA Error from activity in Early RS Error for those same ensembles in HET-VEH mice than in WT mice. SVM decoding accuracy for these ensembles greatly increased in HET-CLNZ and stayed high. Context-specific differentiation of activity improves the efficiency of computations that rely on that information to decode the current environmental context and to guide necessary behavioral alteration. In $Dlx5/6^{+/-}$ mice, neurons downstream of mPFC integrate dysfunctional mPFC output that fails to sufficiently distinguish between changing environmental rules. This disorganized input disrupts downstream computations, which is crucial for altering behavior strategy.

Our results suggest that timely differentiation in mPFC population activity during cognitive flexibility depends on adequately functioning inhibition performed by PVINs. These findings provide deeper insight into how PVINs support mPFC population activity and how PVIN dysfunction can disrupt processing during cognitive flexibility.

mPFC may lack generic error signals during rule-shifting

We note an absent generic error signal in wild-type mice, i.e., a group of cells activate after error trials regardless of the rule period. Generic error signals would possess similar activity structures across different rules, as the sensory stimuli of failing to find a food reward and the neural representation of an abstract reward prediction error ought to be similar across rules.

In WT mice, such a generic error signal could manifest as high cosine similarity between Early IA Error and Early RS Error activity or as relatively low SVM accuracy when classifying frames as belonging to one of those two stages. Neither result was observed; WT mice demonstrate nearly perfect linear separation of pseudopopulation activity between Early IA Errors and Early RS Errors. WT mice also demonstrated much lower cosine similarity in population activity between Early IA Error and Early RS Error. From this, we conclude that while engaging cognitive flexibility during rule shifting, mouse mPFC develops rule-specific error representations. Increased activity similarity between those error periods is associated with impaired cognitive flexibility, meaning that differences in neural representations of errors during changing environmental features are potentially crucial to strategy adaptation.

Results align with previous studies of mPFC PVIN function

Multiple studies from our lab suggest error representation differences are driven by PVIN activity and support efficient cognitive flexibility. Inhibiting contralaterally projecting mPFC PVINs impaired cognitive flexibility and increased activity similarity (of mPFC populations recorded during the rule shifting task) between IA and RS Error (Cho et al., 2023). Previous work also demonstrates that PVINs activity is greater after RS Errors than IA Errors (Cho et al., 2014). This RS Error activation is reduced in *Dlx5/6*$^{+/-}$ mice. A third study found that optogenetic mPFC PVIN stimulation during the first 5 RS trials (error and correct trials) improved cognitive flexibility in HET mice (Cho et al., 2020). This chapter's findings align with all three of these studies by demonstrating that improved cognitive flexibility accompanies decreased similarity between IA Error and RS Error activity and that in PVIN-dysfunctional mPFC, activity in RS errors shows abnormalities that are reversed by augmenting inhibition. Our findings suggest that PVIN activity after RS Errors could crucially dampen IA Error ensemble re-activation during RS Error and thus distinguish network representations of IA Error and RS Error. Future work could inhibit PVINs only after Early RS Errors in WT mice while recording single-cell mPFC activity to explore

whether post-RS Error PVIN activation directly differentiates IA Error representations from RS Error representations.

Caveats and future directions

SVMs decoding task stage used linear kernels, meaning that decision boundaries predicting what subspaces are occupied by each class are generated from linear combinations of ensemble neurons. SVM performance can be a proxy for how hypothetical downstream neural populations could distinguish between task stages using mPFC input. Presumably, impaired decoder performance mirrors impaired in-vivo neural computations downstream to the mPFC. This has not yet been proven; decreased linear separation between different error stages could potentially not alter information processing downstream of mPFC at all. This study cannot speculate on whether local non-linear classifiers embodied in neural subpopulations, utilizing complex and unknown activation functions, can still efficiently separate subspaces occupied by population activity during Early IA error and Early RS Error. While possible, this is unlikely. The mPFC is an essential locus for cognitive flexibility, and deficits in *Dlx5/6$^{+/-}$* mice depend specifically on PVIN dysfunction with mPFC. The observed decrease in the linear separation of Early IA Error and Early RS Error in HET mice disappears after CLNZ treatment. Our results suggest that the observed changes in population encoding are not just epiphenomena and that changes in mPFC network space potentially contribute to impaired cognitive flexibility.

Recording from mPFC projections to other brain regions implicated in cognitive flexibility, like the mediodorsal thalamus, or from cells receiving those inputs could address this question (Rikhye et al., 2018). If a similar decoder analysis on these projection neurons or projection targets also found deficits in classifying error stages, it would support the hypothesis that neural representations of errors are disrupted by PVIN dysfunction and that this is potentially causal in impairing cognitive flexibility.

Linear separability supports downstream processing in various contexts- in simulations,

theoretical work, and empirically. Our results suggest that PVINs are crucial in separating

network activity after errors during rule learning- distinguishing the current rule context. CLNZ

treatment increases the linear separability of activity in state space. It decreases the mean

similarity of mean activity in Early IA Errors and Early RS Errors for ensembles active in 2 key

points of the rule shift- Early RS Error and Early RS Correct trials. These studies demonstrate

that efficient transitions between distinct network states in mPFC depend on proper PVIN

function, suggesting that cognitive deficits in humans associated with PVIN dysfunction might

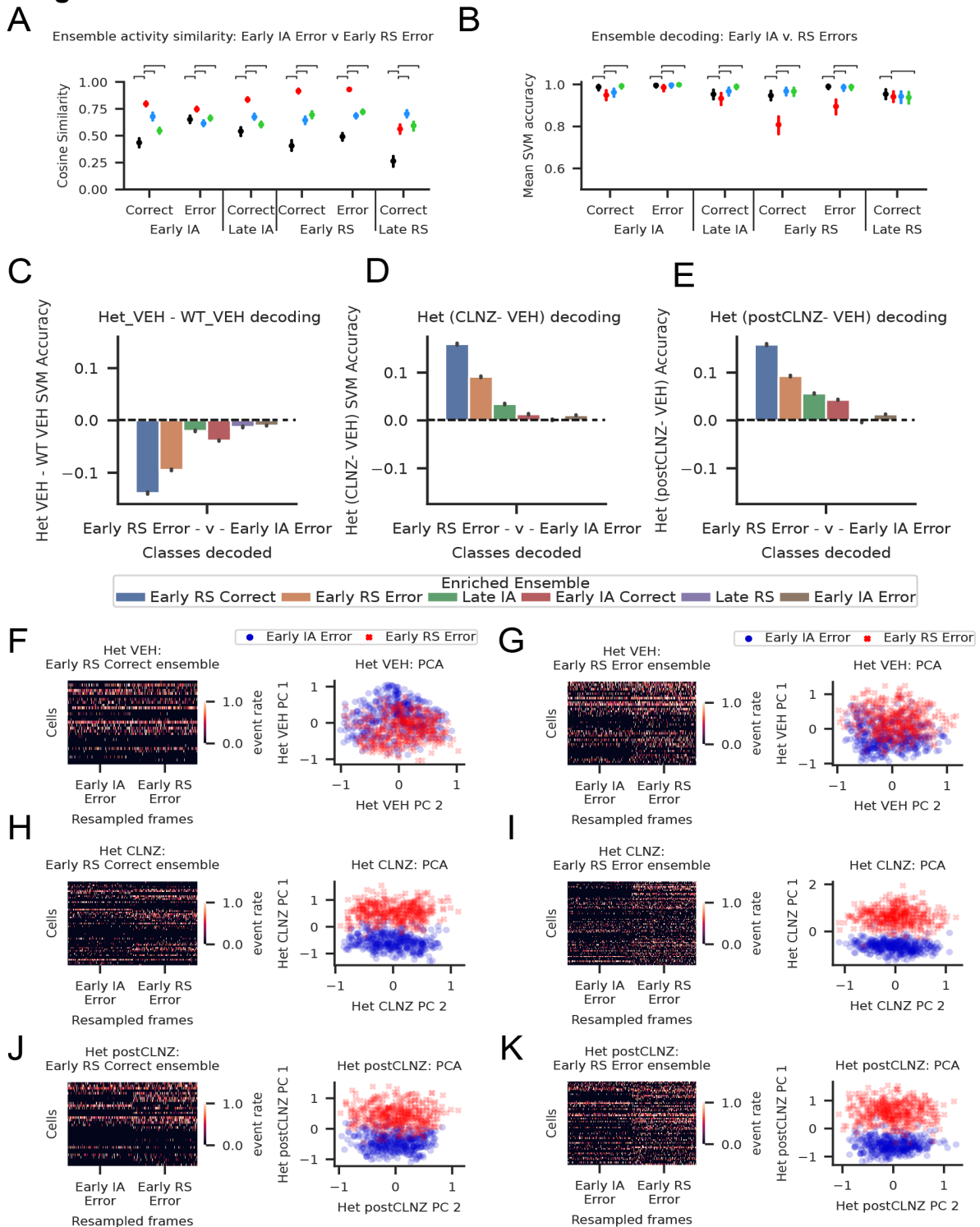arise from mPFC representational rigidity.

3.5 Figures

Figure 3-1 CLNZ separates IA and RS Error activity in HET mice RS Error ensembles

a) Early RS-based ensembles demonstrate lower average SVM decoding accuracy distinguishing between IA and RS Errors. *(Figure caption continued on next page)*

(Figure caption continued from the previous page) This is significantly increased by CLNZ administration (post-hoc comparisons significant after BY correction; $p < 1.05e^{-99}$)

b) Difference between HET VEH and Het CLNZ ensemble decoding accuracy
c) Similar to B), for HET CLNZ and Het VEH ensemble decoding accuracy
d) Similar to C), for HET postCLNZ and Het VEH ensemble decoding accuracy
e) HET VEH sample pseudopopulation and PCA projection of Early RS Correct ensemble activity in Early IA Error and Early RS Error (1st 2 PCs)
f) Similar to E) for Early RS Error ensemble activity in Early IA Error and Early RS Error
g) HET CLNZ sample pseudopopulation and PCA projection of Early RS Correct ensemble activity in Early IA Error and Early RS Error (1st 2 PCs)
h) Similar to G) for Early RS Error ensemble activity in Early IA Error and Early RS Error
i) HET postCLNZ sample pseudopopulation and PCA projection of Early RS Correct ensemble activity in Early IA Error and Early RS Error (1st 2 PCs)
j) Similar to I) for Early RS Error ensemble activity in Early IA Error and Early RS Error

3.6 References

Altan, E., Solla, S. A., Miller, L. E., & Perreault, E. J. (2021). Estimating the dimensionality of the

manifold underlying multi-electrode neural recordings. *PLoS Computational Biology*, *17*(11),

e1008591. https://doi.org/10.1371/journal.pcbi.1008591

Antzoulatos, E. G., & Miller, E. K. (2011). Differences between neural activity in prefrontal cortex

and striatum during learning of novel abstract categories. *Neuron*, *71*(2), 243–249.

https://doi.org/10.1016/j.neuron.2011.05.040

Aoi, M. C., Mante, V., & Pillow, J. W. (2020). Prefrontal cortex exhibits multidimensional dynamic

encoding during decision-making. *Nature Neuroscience*, *23*(11), 1410–1420.

https://doi.org/10.1038/s41593-020-0696-5

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and

computation. *Nature Reviews. Neuroscience*, *7*(5), 358–366. https://doi.org/10.1038/nrn1888

Bartolo, R., & Averbeck, B. B. (2020). Prefrontal Cortex Predicts State Switches during Reversal

Learning. *Neuron*. https://doi.org/10.1016/j.neuron.2020.03.024

Bartolo, R., Saunders, R. C., Mitz, A., & Averbeck, B. B. (2019). Dimensionality, information and

learning in prefrontal cortex. *bioRxiv*, 823377. https://doi.org/10.1101/823377

Bartolo, R., Saunders, R. C., Mitz, A. R., & Averbeck, B. B. (2020a). Dimensionality, information

and learning in prefrontal cortex. *PLOS Computational Biology*, *16*(4), e1007514.

https://doi.org/10.1371/journal.pcbi.1007514

Bartolo, R., Saunders, R. C., Mitz, A. R., & Averbeck, B. B. (2020b). Information-Limiting

Correlations in Large Neural Populations. *The Journal of Neuroscience: The Official Journal of

the Society for Neuroscience*, *40*(8), 1668–1678. https://doi.org/10.1523/JNEUROSCI.2072-

19.2019

Biane, J. S., Ladow, M. A., Stefanini, F., Boddu, S. P., Fan, A., Hassan, S., Dundar, N., Apodaca-Montano, D. L., Zhou, L. Z., Fayner, V., Woods, N. I., & Kheirbek, M. A. (2023). Neural dynamics underlying associative learning in the dorsal and ventral hippocampus. *Nature Neuroscience*, *26*(5), 798–809. https://doi.org/10.1038/s41593-023-01296-6

Bocincova, A., Buschman, T. J., Stokes, M. G., & Manohar, S. G. (2022). Neural signature of flexible coding in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *119*(40), e2200400119. https://doi.org/10.1073/pnas.2200400119

Chelaru, M. I., Eagleman, S., Andrei, A. R., Milton, R., Kharas, N., & Dragoi, V. (2021). High-order interactions explain the collective behavior of cortical populations in executive but not sensory areas. *Neuron*, *109*(24), 3954-3961.e5. https://doi.org/10.1016/j.neuron.2021.09.042

Chiang, F.-K., Wallis, J. D., & Rich, E. L. (2022). Cognitive strategies shift information from single neurons to populations in prefrontal cortex. *Neuron*, *110*(4), 709-721.e4. https://doi.org/10.1016/j.neuron.2021.11.021

Cho, K. K. A., Davidson, T. J., Bouvier, G., Marshall, J. D., Schnitzer, M. J., & Sohal, V. S. (2020). Cross-hemispheric gamma synchrony between prefrontal parvalbumin interneurons supports behavioral adaptation during rule shift learning. *Nature Neuroscience*, *23*(7), 892–902. https://doi.org/10.1038/s41593-020-0647-1

Cho, K. K. A., Hoch, R., Lee, A. T., Patel, T., Rubenstein, J. L. R., & Sohal, V. S. (2015). Gamma Rhythms Link Prefrontal Interneuron Dysfunction with Cognitive Inflexibility in Dlx5/6+/− Mice. *Neuron*, *85*(6), 1332–1343. https://doi.org/10.1016/j.neuron.2015.02.019

Cho, K. K. A., Shi, J., Phensy, A. J., Turner, M. L., & Sohal, V. S. (2023). Long-range inhibition synchronizes and updates prefrontal task activity. *Nature*, *617*(7961), Article 7961. https://doi.org/10.1038/s41586-023-06012-9

Cross, L., Cockburn, J., Yue, Y., & O'Doherty, J. P. (2021). Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron*, *109*(4), 724-738.e7. https://doi.org/10.1016/j.neuron.2020.11.021

Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2022). The role of population structure in computations through neural dynamics. *Nature Neuroscience*, *25*(6), Article 6. https://doi.org/10.1038/s41593-022-01088-4

Durstewitz, D., Vittoz, N. M., Floresco, S. B., & Seamans, J. K. (2010a). Abrupt Transitions between Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule Learning. *Neuron*, *66*(3), 438–448. https://doi.org/10.1016/j.neuron.2010.03.029

Durstewitz, D., Vittoz, N. M., Floresco, S. B., & Seamans, J. K. (2010b). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, *66*(3), 438–448. https://doi.org/10.1016/j.neuron.2010.03.029

Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine in cognitive neuroscience. *Neuron*, *109*(19), 3055–3068. https://doi.org/10.1016/j.neuron.2021.07.011

Ehret, B., Boehringer, R., Amadei, E. A., Cervera, M. R., Henning, C., Galgali, A. R., Mante, V., & Grewe, B. F. (2024). Population-level coding of avoidance learning in medial prefrontal cortex. *Nature Neuroscience*, 1–11. https://doi.org/10.1038/s41593-024-01704-5

Fine, J. M., Maisson, D. J.-N., Yoo, S. B. M., Cash-Padgett, T. V., Wang, M. Z., Zimmermann, J., & Hayden, B. Y. (2023). Abstract Value Encoding in Neural Populations But Not Single Neurons. *Journal of Neuroscience*, *43*(25), 4650–4663. https://doi.org/10.1523/JNEUROSCI.1954-22.2023

Hamm, J. P., Peterka, D. S., Gogos, J. A., & Yuste, R. (2017). Altered Cortical Ensembles in Mouse Models of Schizophrenia. *Neuron*, *94*(1), 153-167.e8. https://doi.org/10.1016/j.neuron.2017.03.019

Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., & Seamans, J. K. (2012). Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, *109*(13), 5086–5091. https://doi.org/10.1073/pnas.1114415109

Inan, H., Schmuckermair, C., Tasci, T., Ahanonu, B. O., Hernandez, O., Lecoq, J., Dinç, F., Wagner, M. J., Erdogdu, M. A., & Schnitzer, M. J. (n.d.). *Fast and statistically robust cell extraction from large-scale neural calcium imaging datasets*. 83.

Karlsson, M. P., Tervo, D. G. R., & Karpova, A. Y. (2012). Network Resets in Medial Prefrontal Cortex Mark the Onset of Behavioral Uncertainty. *Science*, *338*(6103), 135–139. https://doi.org/10.1126/science.1226518

Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature Communications*, *10*(1), 5416. https://doi.org/10.1038/s41467-019-13056-x

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Meyers, E. M. (2018). Dynamic population coding and its relationship to working memory. *Journal of Neurophysiology*, *120*(5), 2260–2268. https://doi.org/10.1152/jn.00225.2018

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(2), 394–399. https://doi.org/10.1073/pnas.1619449114

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., & Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science*, *359*(6374), 463–465. https://doi.org/10.1126/science.aao0284

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Powell, N. J., & Redish, A. D. (2016). Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behavior. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12830

Rich, E. L., & Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(22), 7208–7219. https://doi.org/10.1523/JNEUROSCI.6068-08.2009

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Rigotti, M., Ben Dayan Rubin, D. D., Wang, X.-J., & Fusi, S. (2010). Internal Representation of Task Rules by Recurrent Dynamics: The Importance of the Diversity of Neural Responses. *Frontiers in Computational Neuroscience*, *4*. https://doi.org/10.3389/fncom.2010.00024

Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature Neuroscience*, *21*(12), 1753–1763. https://doi.org/10.1038/s41593-018-0269-z

Saxena, S., & Cunningham, J. P. (2019). Towards the neural population doctrine. *Current Opinion in Neurobiology*, *55*, 103–111. https://doi.org/10.1016/j.conb.2019.02.002

Tye, K. M., Miller, E. K., Taschbach, F. H., Benna, M. K., Rigotti, M., & Fusi, S. (2024). Mixed selectivity: Cellular computations for complexity. *Neuron*, *0*(0). https://doi.org/10.1016/j.neuron.2024.04.017

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:

Carlos Johnson-Cruz

2D93735C9ADF4E0...   Author Signature

8/22/2024

Date