UC Davis

UC Davis Electronic Theses and Dissertations

Title

Developmental Dynamics and Transcriptional Trajectories in Normal and Aberrant Neural Cell Specification During Mouse Corticogenesis

Permalink

https://escholarship.org/uc/item/8m723784

Author

Wang, Yurong

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Developmental Dynamics and Transcriptional Trajectories in Normal and Aberrant Neural Cell Specification During Mouse Corticogenesis

By

YURONG WANG

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Alexander Star Nord

Kyle Fink

Hongwu Chen

Committee in Charge

2024

Acknowledgements

The completion of this dissertation would not have been possible without the strong support and guidance of many kind and knowledgeable individuals. I would like to give special thanks to my Principal Investigator and Major professor, Dr. Alex Nord, for his great support through all these years, giving me countless opportunities on projects, collaboration, teaching opportunities, and encouraging me when I lost confidence. I am truly thankful for his mentorship. I am also thankful towards Dr. John Rubenstein from University of California San Francisco, for his great patience and exceptional expertise. Additionally, many thanks to everyone from the Nord lab, for the peaceful and productive work environment they fostered, as well as to Dr. Susan Lindtner, Dr. Athena Ypsilanti, Dr. Jia Hu, with whom it has been a pleasure to collaborate.

I would also like to express my genuine thanks to my dissertation committee members, Dr. Hongwu Chen and Dr. Kyle Fink, for their prompt and insightful feedbacks. I would also like to express my appreciation to my graduate program, IGG, for providing such supports to its students.

Lastly, I thank my family for their constant support in all ways, and my friends Hanako and Vivi for their support during those challenging moments. I would also like to acknowledge the great advancements on technology, without which the data and tools to complete the analyzing process would be non-existent.

ABSTRACT

Understanding brain development is crucial for advancing knowledge of neurological diseases and improving intervention methods. To understand such processes under normal and perturbed conditions, numerous techniques have been developed, and this pool of possible techniques to choose from is still growing. With the help of these technologies, labs have generated massive datasets that require careful analysis to extract insights that closely reflect actual biological processes. Accurate interpretation of these data is essential for enhancing our knowledge and developing neurological disease treatments effectively. In this dissertation, I employed various computational approaches to analyze data generated to investigate brain development under normal and disturbed conditions, with a focus on datasets related to gene expression and epigenomic changes.

The first project focused on understanding the role of Dlx1 and Dlx2 transcription factors in neurodevelopment post-proliferation. By utilizing single-cell resolution methods, gene expression differences in conditional Dlx1/2 knockouts had been examined, and I computationally identified downstream genes of Dlx1/2 across various subregions of the ganglionic eminence and olfactory bulb. Additionally, analysis of Dlx2 binding regions across different developmental stages and regions revealed significant binding patterns, with enriched GO terms highlighting Dlx2's role on critical processes within GE subregions.

In the second project, we focused on the development of medial ganglionic eminence (MGE)-derived interneurons. We explored chromatin dynamics and gene regulation during the differentiation of embryonic stem cells (ESCs) into MGE-derived GABAergic cortical interneurons (CINs). I computationally integrated open chromatin signals and histone marks across six developmental time points, identifying putative enhancers and linking them to gene expression levels.

The third chapter employed the method similar to the first project to investigate the immediate effects of maternal immune activation (MIA) on brain development in mouse embryos at 12.5 days in utero using a single-cell resolution gene expression profiling approach. I analyzed both neuronal and non-neuronal cell types in this dataset, examining differential gene expression and changes in cell proportions in response to MIA.

In summary, this dissertation further advances our understanding of brain development under both normal and disturbed conditions by integrating computational methods with cutting-edge techniques.

Chapter 1

Introduction

The mammalian brain undergoes a highly orchestrated and complex developmental process, beginning with the formation of the neural tube, which later gives rise to the central nervous system. This process involves multiple stages of patterning and regionalization, including neurogenesis, cell migration, and synaptic refinement. Disruptions to this process, for example following loss of expression of critical developmental genes like Dlx1 and Dlx2 (the focus of Chapter 2) or driven by environmental perturbations such as maternal immune activation (MIA) during pregnancy (the focus of Chapter 3), can significantly impact fetal brain development, potentially disturbing this intricate system. By employing and integrating results from advanced genomic technologies like single-cell RNA sequencing and epigenomic profiling, we are gaining deeper insights into the molecular mechanisms underlying brain development and its disruption by abnormal conditions. Chapter 3 represents a case study in using omics approaches and systems biology to understand neurodevelopment, with a focus on the origination and specification of cortical GABAergic interneurons. The unifying theme of the work presented in this dissertation is the use of functional genomics and systems biology to understand neurodevelopment. Studies such as those presented here are leading to new understanding of the processes directing normal and pathogenic neurodevelopment, give insights into evolution and development, and potentially offer new possibilities for therapeutic interventions for neurodevelopmental and neuropsychiatric disorders.

A. Early mammalian brain development

The development of the mammalian brain is a complex and tightly regulated process that stretches across embryonic and postnatal stages of life. This intricate process is initiated with the formation of the neural tube around embryonic day 8.5 (E8.5) in mice, which serves as the precursor to the entire central nervous system (Greene and Copp, 2009). As the neural tube closes, it forms three primary brain vesicles: the prosencephalon (forebrain), mesencephalon (midbrain), and rhombencephalon (hindbrain). The prosencephalon, which will give rise to the telencephalon and diencephalon, is particularly critical as it forms the basis for the cerebral cortex, basal ganglia, and other major brain structures (Stiles and Jernigan, 2010). The telencephalon, also referred to less specifically as the forebrain is dived into the dorsal and ventral regions, which are also referred to as the pallium and sub-pallium, respectively.

B. Pallial development: glutamatergic neurogenesis and cerebral cortex patterning

During development, the pallium is where embryonic neurogenesis occurs that gives rise to excitatory glutamatergic neurons that make up the layers of the cerebral cortex and play crucial roles in cortical function via their local and long-distance projections. Glutamatergic neurons are generated from neural progenitor cells located in the ventricular zone (VZ) and subventricular zone (SVZ) during neurogenesis (Angevine & Sidman, 1961; Caviness & Takahashi. 1995). Within the VZ, neuroepithelial cells differentiate into *Pax6*-positive radial glial cells (RGCs), which act as both progenitors of new neurons and as structural scaffolds to guide neuronal migration (Englund et al., 2005; Kolk & Rakic, 2022). These radial glial cells undergo expansion and generate the earliest-born neurons, including Cajal-Retzius cells. Cajal-Retzius cells are largely developed from these RGCs, acting as pivotal in organizing the developing cortex and are positioned in the marginal zone, which will become the future layer I of the cortex (Janusonis et al., 2004).

In addition to producing Cajal-Retzius cells, radial glial cells also give rise to intermediate progenitors, which are *Tbr2*/*Eomes*-positive cells found predominantly in the SVZ (Englund et al., 2005; Di Bella et al., 2021). These intermediate progenitors produce neurons that will populate cortical layers 2 through 6, doing so in an inside-out sequence, with deeper layers formed first, followed by more superficial layers (Bayer & Altman, 1991; Molyneaux et al., 2007). Newly generated postmitotic neurons migrate radially from the VZ and SVZ toward the cortical plate, where they settle to form the distinct layers of the cerebral cortex.

C. Subpallial GABAergic neurogenesis and cortical interneuron development

The ventral telencephalon is organized into distinct sub-regions during embryonic development, including the medial ganglionic eminence (MGE), lateral ganglionic eminence (LGE) and caudal ganglionic eminence (CGE). These regions are crucial for the production and migration of GABAergic interneurons, which are essential for the proper functioning of neural circuits in the cerebral cortex and other brain regions (Marín and Rubenstein, 2001; Cai et al., 2013; Roux and Buzsáki, 2015; Wamsley and Fishell, 2017), as well as local GABAergic projection neurons and cholinergic neurons that make up regions of the mature basal ganglia.

The migrating GABAergic interneurons that originate from the MGE are crucial for cerebral cortex development and function. Progenitors within the basal ganglia VZ and SVZ proliferate and, under the influence of key transcription factors like Dlx1 and Dlx2, tangentially migrate toward the neocortex (Marín and Rubenstein, 2001; Anderson et al., 1997; Flames and Marín, 2005; Götz and Huttner, 2005). There, cortical interneurons continue maturation and lineage specification, ultimately establishing connections as mature GABAergic interneuron subtypes, for example parvalbumin, somatostatin, and VIP expressing classes, which together play vital roles in regulating brain function. (Roux and Buzsáki, 2015).

Although the major migration of interneurons from the ganglionic eminences, particularly the medial ganglionic eminence (MGE), occurs predominantly between E11.5 and E16.5, these structures persist into the early postnatal period (Marín and Rubenstein, 2001; Faux et al., 2012; Anderson et al., 2001). During this time, the ganglionic eminences continue to play a role in brain development, although brain shifts toward synaptic maturation and circuit refinement at the time (Kim and Paredes, 2021).

D. Non-neuronal cell types in the telencephalon

The meninges are protective membranes that surround the brain and spinal cord, playing crucial roles in maintaining central nervous system (CNS) homeostasis, and the meningeal structure begins to form at around E9.5 in mice when primary meninx is formed (Dasgupta & Jeong, 2019). These layers consist of the dura mater, the tough outermost layer; the arachnoid mater, the middle layer; and the pia mater, the innermost layer (Vanderah & Gould, 2020). The leptomeninges, composed of the arachnoid and pia mater, forms the major structure holding cerebrospinal fluid and separating the CNS from the outside body environment.

These structures are composed of various types of non-neuronal cells, with several of them derived from mesenchymal cells. Vascular leptomeningeal cells (VLMCs), pericytes, and endothelial cells are important types. VLMC is a specialized type of stromal cell, are key contributors to the structure and function of the brain’s vasculature such as blood vessels, and are present in permeable brain regions such as the choroid plexus and pituitary gland, areas where the blood-brain barrier (BBB) is more permeable than in other regions in adult mice (Bastedo et al., 2024), involved in maintaining the homeostasis of the cerebrospinal fluid and protecting the brain from peripheral immune and inflammatory signals.

Endothelial cells and pericytes are two cell types surrounding blood vessels in brain. Endothelial cells form the inner lining of blood vessels in the brain, while pericytes are cells that surround and support these endothelial cells by wrapping around the blood vessels. Both cell types work together to maintain the structure and function of blood vessels in the brain, especially in forming and regulating the blood-brain barrier (Abbott et al., 2018).

In addition to vasculature associated cell types, there are three major glial cell types: microglia, astrocytes, and oligodendrocytes. Of these, oligodendrocytes and astrocytes are produced from the same multi-potent progenitor populations that produce neurons in the palliums and sub-pallium. In contrast, microglia populate the brain from the yolk sac during embryonic development. While the full functional activity of glial cells is not completely understood, all of these cells play critical roles in brain development and in function. Microglia, which are a focus of Chapter 3, are the brain's resident immune cells, playing a role in both immune response and in neurodevelopment and synaptogenesis via interactions with neurons. Microglia can be activated by immune signaling or neuronal stress.

E. Epigenetics, epigenomics, and gene regulatory biology

Neurodevelopment and the proper function of the brain is dependent on tightly regulated gene expression patterns. The regulation of these patterns is driven by gene regulatory interactions between transcription factors (TFs), co-factor proteins, and regulatory DNA elements like enhancers and promoters. These regulatory elements are regions of the genome that control the expression of genes, playing a vital role in the precise regulation of gene activity across different tissues and developmental stages (Maston et al., 2006; Pennacchio et al., 2013; Friedman and Rando, 2015). To better understand the intricate processes involved in brain development and dysfunction, we can use functional genomics assays that capture gene regulatory interactions and epigenetic state, as well as RNA sequencing. Recently, these

methods have been extended to single-cell level, offering a more detailed perspective on how these gene regulation contributes to both normal development and potential disruptions caused by environmental factors.

Given the difficulty of accurately dissecting regions such as the MGE, CGE, and olfactory bulb and the heterogeneity of cells present in the embryonic brain, utilizing single-cell RNA sequencing (scRNA-seq) instead of bulk RNA-seq offers significant advantages in understanding cell-type specific normal state and response to genetic or environmental perturbation (Kolodziejczyk et al., 2015; Potter et al., 2009). Here, we leverage scRNA-seq to enable detailed analysis of gene expression profiles in individual cell types following Dlx1/2 TF loss in Chapter 2 and MIA in Chapter 3.

Enhancers are DNA sequences that can be located upstream, downstream, or even within the genes they regulate, often functioning over long distances to increase the transcription of specific genes, and modulate the efficiency and rate of transcription, often in a tissue-specific or developmental-stage-specific manner (Pennacchio et al., 2013). When active, the chromosomal region where enhancers reside forms an open-chromatin structure, with bound transcription factors (TFs) and is often marked by histone modifications like H3K27ac; these enhancers then possibly loop over to physically interact with the promoter of the gene they control to assist in transcription (Zhao et al., 2021; Bulger and Groudine, 2010; Splinter et al., 2006; Schoenfelder and Fraser, 2019). Epigenomic marks often serve as indicators of regulatory elements within the genome, and by mapping these marks, researchers can identify potential regulatory regions and gain insights into the mechanisms that regulate gene activity (Visel et al., 2009). For example, Histone modifications such as H3K4Me3 and H3K27Ac are key markers associated with active transcription and enhancer regions. H3K4Me3 is correlated with gene expression and is usually found at the active genes, while H3K27Ac marks active enhancers and promoters, indicating

regions of open chromatin that are transcriptionally active (Creyghton et al., 2010; Howe et al., 2017; Santos-Rosa et al., 2002). In contrast, H3K27me3 is a histone modification associated with repressed chromatin and is typically found at silenced genes and inactive regulatory regions (Cao et al., 2002). Using active enhancer regions as platforms to bind to, TFs control the transcription of target genes by facilitating the recruitment of the transcriptional machinery to the gene's promoter region (Heinz et al., 2015; Schoenfelder and Fraser, 2019). TFs bind to their specific motifs on the DNA within the chromatin and may also interact with other enhancers via indirect interactions via protein-protein interactions or through histone-associated recruitment, collectively regulating condition-specific transcription (Poitras et al., 2007; Schoenfelder and Fraser, 2019). Dlx1 and Dlx2 mentioned in the previous section, for example, are TFs that regulating development-specific genes required for ventral or sub-pallial GABAergic neurogenesis (Price et al., 2022).

Open chromatin structure is another key indicator of potential enhancer regions, as these areas of open chromatin are capable of binding with TFs and regulate gene expression (Heinz et al., 2015). ATAC-seq is an effective method for identifying these regions of accessible chromatin, integrating ATAC-seq data with other epigenomic marks, such as histone modifications, researchers can more accurately predict and validate enhancer activity (Buenrostro et al., 2015; Buenrostro et al., 2013; Heinz et al., 2015). By combining epigenomic marks, we can identify genomic regions that are likely regulatory elements specific to particular developmental time points. In one of the chapters, datasets from these analyses are used to predict putative regulatory elements. Many methods have been developed and are currently used to analyze these epigenomic marks in an integrative way.

By combining epigenomic marks, we can identify genomic regions that are likely regulatory elements specific to particular developmental time points. In Chapter 3, we integrate extensive

epigenomic datasets to predict putative regulatory elements and chromatin state during MGE-derived cortical interneuron specification. Many methods have been developed and are currently used to analyze these epigenomic marks in an integrative way. For example, chromHMM is a widely used tool that identifies and models chromatin states by integrating multiple histone modifications across the genome using a Hidden Markov Model assigning chromatin states to genomic regions, aiding in the prediction of enhancer regions and other regulatory elements that are specific to developmental stages (Ernst et al., 2012; Ernst et al., 2017). Some of the more recent and less widely adopted methods include PALMO, which employs a generalized linear mixed model for integrating longitudinal multi-omics data (Vasaikar et al., 2023). This approach allows for the analysis of variability across different developmental timepoints, enabling the detection of regions whose patterns change over time, potentially identifying them as regulatory elements.

G. Maternal immune activation as a leading NDD/NPD environmental risk factor

In Chapters 1 and 3, my focus is on revealing the gene regulatory processes underlying development of GABAergic cortical interneurons. In contrast, Chapter 2 uses single cell transcriptomics to tackle MIA, a significant risk factor for perturbed neurodevelopment and brain disorders Maternal immune activation (MIA) broadly refers to the activation of the maternal immune system during pregnancy in response to infections, stress, or other inflammatory triggers. This immune response can lead to the production and release of pro-inflammatory cytokines, such as interleukin-6 (IL-6), interleukin-17 (IL-17), and tumor necrosis factor-alpha (TNF-α), which can cross the placental barrier and potentially impact fetal development (Meyer et al., 2006; Smith et al., 2007). Understanding MIA is crucial for identifying potential preventive measures and therapeutic interventions. Given the growing body of evidence linking MIA to neurodevelopmental disorders, researchers are increasingly focused on unraveling the complex interactions between maternal immune responses and fetal brain development.

Research has shown that MIA can disrupt normal brain development, leading to long-term alterations in brain structure and function. These disruptions have been associated with an increased risk of neurodevelopmental disorders, including autism spectrum disorders (ASD), schizophrenia, and bipolar disorder (Brown & Derkits, 2010; Meyer, 2014). The precise mechanisms through which MIA affects fetal brain development are still being explored, but it is believed that the inflammatory environment created by MIA can interfere with critical processes such as neurogenesis, synaptogenesis, and the establishment of neural circuits (Estes & McAllister, 2016). The impact of MIA on the developing fetus can vary depending on the timing, severity, and nature of the immune challenge. Early to mid-gestational periods are particularly sensitive to immune disturbances, as these stages involve key processes in brain development, such as the proliferation and differentiation of neural progenitor cells (Knuesel et al., 2014). The timing of the immune response may determine which brain regions and developmental processes are most affected, leading to the diverse range of neurodevelopmental outcomes observed in MIA-exposed offspring (Boksa, 2010).

Single-cell omics, such as single-cell RNA sequencing (scRNA-seq) and single-nuclei RNA sequencing (snRNA-seq), provides a powerful tool to uncover the complexities of the brain's response to MIA by enabling the examination of gene expression at the individual cell level (Ostrem et al., 2024). By capturing cellular heterogeneity and identifying cell type-specific changes, single-cell omics has the potential to reveal the intricate molecular pathways involved in immune response, neuroinflammation, and brain development under MIA conditions. In the long term, this research holds the potential to inform public health strategies aimed at reducing the risk of neurodevelopmental disorders by managing maternal health and immune responses during pregnancy.

Reference

Abbott, N. J., Pizzo, M. E., Preston, J. E., Janigro, D., & Thorne, R. G. (2018). The role of brain barriers in fluid movement in the CNS: is there a ‘glymphatic’system?. Acta neuropathologica, 135, 387-407.

Anderson, S. A., Eisenstat, D. D., Shi, L., & Rubenstein, J. L. R. (1997). Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes. Science, 278(5337), 474-476.

Anderson, S. A., Marín, O., Horn, C., Jennings, K., & Rubenstein, J. L. (2001). Distinct cortical migrations from the medial and lateral ganglionic eminences. Development, 128(3), 353-363.

Angevine, J. B. J., & Sidman, R. L. (1961). Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. Nature, 192(4804), 766-768.

Bayer, S. A., & Altman, J. (1991). Neocortical development (Vol. 1). New York: Raven Press.

Bastedo, W. E., Scott, R. W., Arostegui, M., & Underhill, T. M. (2024). Single-cell analysis of mesenchymal cells in permeable neural vasculature reveals novel diverse subpopulations of fibroblasts. Fluids and Barriers of the CNS, 21(1), 31.

Boksa, P. (2010). Effects of prenatal infection on brain development and behavior: a review of findings from animal models. Brain, behavior, and immunity, 24(6), 881-897.

Brown, A. S., & Derkits, E. J. (2010). Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. American Journal of Psychiatry, 167(3), 261-280.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods, 10(12), 1213-1218.

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Current protocols in molecular biology, 109(1), 21-29.

Bulger, M., & Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Developmental biology, 339(2), 250-257.

Cai, Y., Zhang, Q., Wang, C., Zhang, Y., Ma, T., Zhou, X., ... & Yang, Z. (2013). Nuclear receptor COUP-TFII-expressing neocortical interneurons are derived from the medial and lateral/caudal ganglionic eminence and define specific subsets of mature interneurons. Journal of Comparative Neurology, 521(2), 479-497.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., ... & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science, 298(5595), 1039-1043.

Caviness Jr, V. S., & Takahashi, T. (1995). Proliferative events in the cerebral ventricular zone. Brain and Development, 17(3), 159-163.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., ... & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences, 107(50), 21931-21936.

Dasgupta, K., & Jeong, J. (2019). Developmental biology of the meninges. Genesis, 57(5), e23288.

Di Bella, D. J., Habibi, E., Stickels, R. R., Scalia, G., Brown, J., Yadollahpour, P., ... & Arlotta, P. (2021). Molecular logic of cellular diversification in the mouse cerebral cortex. Nature, 595(7868), 554-559.

Englund, C., Fink, A., Lau, C., Pham, D., Daza, R. A., Bulfone, A., ... & Hevner, R. F. (2005). Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. Journal of Neuroscience, 25(1), 247-251.

Ernst, J., & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nature methods, 9(3), 215-216.

Ernst, J., & Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nature protocols, 12(12), 2478-2492.

Estes, M. L., & McAllister, A. K. (2016). Maternal immune activation: Implications for neuropsychiatric disorders. Science, 353(6301), 772-777.

Faux, C., Rakic, S., Andrews, W., & Britto, J. M. (2012). Neurons on the move: migration and lamination of cortical interneurons. Neurosignals, 20(3), 168-189.

Flames, N., & Marín, O. (2005). Developmental mechanisms underlying the generation of cortical interneuron diversity. Neuron, 46(3), 377-381.

Friedman, N., & Rando, O. J. (2015). Epigenomics and the structure of the living genome. Genome research, 25(10), 1482-1490.

Götz, M., & Huttner, W. B. (2005). The cell biology of neurogenesis. Nature reviews Molecular cell biology, 6(10), 777-788.

Greene, N. D., & Copp, A. J. (2009). Development of the vertebrate central nervous system: formation of the neural tube. Prenatal Diagnosis: Published in Affiliation with the International Society for Prenatal Diagnosis, 29(4), 303-311.

Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015). The selection and function of cell type-specific enhancers. Nature reviews Molecular cell biology, 16(3), 144-154.

Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. (2017). Is H3K4me3 instructive for transcription activation?. Bioessays, 39(1), 1-12.

Janušonis, S., Gluncic, V., & Rakic, P. (2004). Early serotonergic projections to Cajal-Retzius cells: relevance for cortical development. Journal of Neuroscience, 24(7), 1652-1659.

Kim, J. Y., & Paredes, M. F. (2021). Implications of extended inhibitory neuron development. International journal of molecular sciences, 22(10), 5113.

Knuesel, I., Chicha, L., Britschgi, M., Schobel, S. A., Bodmer, M., Hellings, J. A., ... & Prinssen, E. P. (2014). Maternal immune activation and abnormal brain development across CNS disorders. Nature Reviews Neurology, 10(11), 643-660.

Kolk, S. M., & Rakic, P. (2022). Development of prefrontal cortex. Neuropsychopharmacology, 47(1), 41-57.

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. Molecular cell, 58(4), 610-620.

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. Annu. Rev. Genomics Hum. Genet., 7(1), 29-59.

Marín, O., & Rubenstein, J. L. (2001). A long, remarkable journey: tangential migration in the telencephalon. Nature Reviews Neuroscience, 2(11), 780-790.

Meyer, U., Nyffeler, M., Engler, A., Urwyler, A., Schedlowski, M., Knuesel, I., ... & Feldon, J. (2006). The time of prenatal immune challenge determines the specificity of inflammation-mediated brain and behavioral pathology. Journal of Neuroscience, 26(18), 4752-4762.

Meyer, U. (2014). Prenatal poly (i: C) exposure and other developmental immune activation models in rodent systems. Biological psychiatry, 75(4), 307-315.

Molyneaux, B. J., Arlotta, P., Menezes, J. R., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. Nature reviews neuroscience, 8(6), 427-437.

Ostrem, B. E. L., Domínguez-Iturza, N., Stogsdill, J. A., Faits, T., Kim, K., Levin, J. Z., & Arlotta, P. (2024). Fetal brain response to maternal inflammation requires microglia. Development, 151, dev202252.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. Nature Reviews Genetics, 14(4), 288-295.

Poitras, L., Ghanem, N., Hatch, G., & Ekker, M. (2007). The proneural determinant MASH1 regulates forebrain Dlx1/2 expression through the I12b intergenic enhancer.

Potter, G. B., Petryniak, M. A., Shevchenko, E., McKinsey, G. L., Ekker, M., & Rubenstein, J. L. (2009). Generation of Cre-transgenic mice using Dlx1/Dlx2 enhancers and their characterization in GABAergic interneurons. Molecular and Cellular Neuroscience, 40(2), 167-186.

Price, J. D., Lindtner, S., Ypsilanti, A., Binyameen, F., Johnson, J. R., Newton, B. W., ... & Rubenstein, J. L. (2022). DLX1 and the NuRD complex cooperate in enhancer decommissioning and transcriptional repression. Development, 149(11), dev199508.

Roux, L., & Buzsáki, G. (2015). Tasks for inhibitory interneurons in intact brain circuits. Neuropharmacology, 88, 10-23.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. T., ... & Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature, 419(6905), 407-411.

Sarropoulos, I., Sepp, M., Frömel, R., Leiss, K., Trost, N., Leushkin, E., ... & Kaessmann, H. (2021). Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. Science, 373(6558), eabg4696.

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. Nature Reviews Genetics, 20(8), 437-455.

Smith, S. E., Li, J., Garbett, K., Mirnics, K., & Patterson, P. H. (2007). Maternal immune activation alters fetal brain development through interleukin-6. Journal of Neuroscience, 27(40), 10695-10702.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., ... & de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes & development, 20(17), 2349-2354.

Stiles, J., & Jernigan, T. L. (2010). The basics of brain development. Neuropsychology review, 20(4), 327-348.

Vasaikar, S. V., Savage, A. K., Gong, Q., Swanson, E., Talla, A., Lord, C., ... & Li, X. J. (2023). A comprehensive platform for analyzing longitudinal multi-omics data. Nature Communications, 14(1), 1684.

Vanderah, T. W., & Gould, D. J. (2020). Nolte's The Human Brain E-Book: Nolte's The Human Brain E-Book. Elsevier Health Sciences.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., ... & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature, 457(7231), 854-858.

Wamsley, B., & Fishell, G. (2017). Genetic and activity-dependent mechanisms underlying interneuron diversity. Nature Reviews Neuroscience, 18(5), 299-309.

Zhao, W., Xu, Y., Wang, Y., Gao, D., King, J., Xu, Y., & Liang, F. S. (2021). Investigating crosstalk between H3K27 acetylation and H3K4 trimethylation in CRISPR/dCas-based epigenome editing and gene activation. Scientific reports, 11(1), 15912.

Zhu, J., Adli, M., Zou, J. Y., Verstappen, G., Coyne, M., Zhang, X., ... & Bernstein, B. E. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell, 152(3), 642-654.

Chapter 2

An investigation into the role of Dlx transcription factors in the maturation of cortical and olfactory bulb GABAergic interneuron lineages

a. Preface

This chapter is a part of a larger collaborative study on the function of Dlx transcription factor in maturing GABAergic interneurons with Dr. John Rubenstein's lab in UCSF. The parts of this project presented here are the bioinformatic analysis of scRNA-seq and ChIP-seq data performed for this project, with all the wet lab work performed by the Rubenstein lab. In the final manuscript in preparation now, this data will also include experimental validation of differentially expressed genes discussed below as well as functional analysis of mouse mutants.

b. Acknowledgements

All wet lab work needed to generate the included genomics datasets was performed in Dr. John Rubenstein's lab, mostly by Dr. Susan Lindtner and Dr. Jia Hu. I thank Stephanie Lozano, Nicolas Seban, Nickolas Chu, Ethan Fenton, and Karol Cichewicz for support in snRNA-seq analysis and former Nord Lab Ph.D. Rinaldo Catta-Preta for support with ChIP-seq analysis.

c. Author contributions

Data was generated by experiments performed by Susan Lindtner and Jia Hu. Emily Pai, a former postdoc with John Rubenstein analyzed an early partial version including the first replicate of the scRNA-seq data. I performed all computational analysis included in this chapter, and drafted figures and text with consultation with my PI, Dr. Alex Nord.

d. Declaration of interests

Nothing to declare.

Abstract:

Here, I present a systems-level and cell-type specific analysis of the role of *Dlx1* and *Dlx2* genes in the development of GABAergic neuron types in the mouse forebrain. *Dlx1* and *2* (*Dlx1/2*) are transcription factors (TFs) that in the telencephalon are expressed in early embryonic basal ganglia and remain active in various BG- and cortical-derived GABAergic lineages. We focus on the role and requirement of *Dlx1/2* TFs in GABAergic interneuron lineage specification and maturation via cKO ablation after initial required activity during neurogenesis, and via regulatory target characterization in relevant embryonic and postnatal brain tissues.

For this work, we used I12b-CRE conditional knockout (cKO) and single-cell RNA sequencing (scRNA-seq) of P3 cortex and bulk chromatin immunoprecipitation sequencing (ChIP-seq) of embryonic and postnatal mouse brain tissues. This first part of the study focuses on postnatal day 3 (P3) mouse cerebral cortex, comparing wild-type (WT) to *Dlx1/2* cKO mouse models to elucidate how these TFs influence the GABAergic interneuron subpopulations, particularly those arising in the medial ganglionic eminence (MGE) and caudal ganglionic eminence (CGE), versus cortical-derived olfactory bulb interneurons. We found that the cKO of *Dlx1/2* leads to downregulation of a key set of genes across the MGE, CGE, and olfactory bulb lineages, as well as within specific neuronal cell types. The disruption of these genes underscores the essential role of *Dlx1/2* in GABAergic interneuron maturation and function. In addition, we characterize of impacts on developmental trajectory within the olfactory bulb lineage. Complementing the scRNA-seq findings, the second part of the

study describes use of bulk ChIP-seq analysis to characterize *Dlx2* binding patterns across different brain regions and developmental stages. This work revealed distinct proximal and distal binding patterns, both with and without the Dlx2 motif. The peaks showed differing binding patterns across conditions, with distal binding patterns particularly enriched for biological processes such as neurogenesis, cell proliferation, and migration, highlighting Dlx2's role in regulating key pathways crucial for brain development.

This study provides important insights into the molecular mechanisms underlying normal interneuron maturation and regarding neurodevelopmental disorders, which are associated with impaired interneuron neurogenesis and specification, and sets the stage for further research into potential therapeutic interventions.

Introduction:

The development of the mammalian brain, and in particular the differentiation of specific neuronal lineages, is a highly orchestrated process involving numerous transcription factors (Sur et al., 2005). Among these, the Distal-less homeobox (*Dlx*) gene family, specifically *Dlx1* and *Dlx2*, plays a crucial role in the development of GABAergic interneurons in the embryonic basal ganglia (BG), and specifically the medial ganglionic eminence (MGE) and the caudal ganglionic eminence (CGE), which serve as key proliferative zones for interneuron precursors, as well as the olfactory bulb (OB) (Anderson et al., 2001; Panganiban et al., 2002; Hu al., 2017). *Dlx1* and *Dlx2* (*Dlx1/2*) are arranged in a head-to-tail orientation on mouse chromosome 2 and are co-expressed and partially redundant (Merlo et al., 2000). *Dlx1/2* expression are required for transcriptional activation of *Dlx5* and *Dlx6* (*Dlx5/6*), which have similar gene structure and also play critical roles in maturation of GABAergic lineages. Beyond *Dlx5/6*, *Dlx1/2* regulate a large number of genes via binding to distal enhancers and promoters (Zerucha et al., 2000; Guo et al., 2019). In previous studies, we identified the binding targets of Dlx1/2 in the embryonic mouse BG, finding genes that were both activated and repressed by Dlx1/2 binding (Lindtner et al., 2019; Catta-Preta et al., 2023). Intriguingly, we noted that some binding targets are expressed in more mature cortical interneurons and further investigation indicated that loss of *Dlx1/2* after neurogenesis caused cortical interneuron dysfunction and resulted in increased postnatal apoptosis of these cells (Pla et al., 2018).

The MGE and CGE give rise to a variety of GABAergic interneuron subtypes that populate the cerebral cortex, hippocampus, and other brain structures (Wonders al., 2006). The transcription factors *Dlx1* and *Dlx2* are expressed prominently in these regions and are essential for the proper neurogenesis, differentiation and migration of interneurons (Anderson et al., 2001). The loss of *Dlx1/2* disrupts these developmental processes, leading to a failure in GABAergic interneuron differentiation and tangential migration during forebrain development (Guo et al., 2019). Additionally, *Dlx1/2* directly regulate the expression of GABAergic markers like *Gad1* and *Gad2*, which are essential for the synthesis of GABA, the primary inhibitory neurotransmitter in the brain (Stühmer et al., 2002; Pla et al., 2018).

The olfactory bulb is another brain structure where *Dlx1/2* play a pivotal role. *Dlx1/2* are central components in the transcriptional code required for the generation of olfactory bulb interneurons (Long et al., 2007). These interneurons arise from neural stem/progenitor cells in the subpallial and pallial SVZ and migrate to the olfactory bulb through the rostral migratory stream, where they differentiate into various subtypes, including granule cells and periglomerular cells (Whitman al., 2009; Alvarez-Buylla and Garcia-Verdugo, 2002; Lledo et al., 2006). In *Dlx1/2* cKO models, there is a marked failure in the differentiation of progenitor cells, leading to a significant reduction in the generation of olfactory bulb interneurons (Guo et al., 2019). This finding underscores the essential role of *Dlx1/2* in the development of olfactory circuits, which are critical for olfaction and related behaviors.

The use of conditional knockout (cKO) models, where *Dlx1/2* are selectively deleted in specific brain regions or cell types at particular developmental stages, has been instrumental in uncovering the diverse roles of these genes (McKinsey et al., 2013), and yielded the initial evidence of requirement for Dlx1/2 in maturing cortical interneuron populations (Pla et al., 2018). These studies have raised questions regarding what the molecular and cellular consequences of Dlx1/2 loss are and whether Dlx1/2 direct regulatory targets change later in maturation that we sought to address in the work presented here. The use of I12-b cKO in P3 mice allows for the study of *Dlx1/2* function specifically in postnatal development, avoiding the early embryonic lethality associated with constitutive knockouts (Potter et al., 2009). In this chapter, the identification of *Dlx1/2* cKO effects and regulatory targets on MGE, CGE, and olfactory bulb lineages at P3 contributes to the growing body of knowledge about lasting requirements for these TFs that were previously best understood for their early developmental roles, potentially offering new perspectives on the role of these transcription factors in brain development and disease.

Results:

Impact of Dlx1/2 cKO on GABAergic lineage-specific impacts on P3 interneurons

In this section, we present the results of our single-cell RNA sequencing (scRNA-seq) analysis on wild-type (WT) and conditional knockout (cKO) mice for *Dlx1* and *Dlx2* genes. The data were processed and visualized through various bioinformatics tools to reveal the expression patterns, cellular subpopulations, and dynamic cellular processes

in the brain. Our findings highlight the significant regions influenced by *Dlx1* and *Dlx2*, namely MGE, CGE, and olfactory bulb lineage. Through sub-clustering of these regions, differentially expressed (DE) genes were identified, and via cell velocity and lineage progression analysis we identified impacts across olfactory bulb lineage development.

Identification of DLX-expressing lineages and cell types

Using scRNA-seq data from postnatal day 3 (P3) for two biological replicates from dissected mouse brain samples, 44830 cells passed QC metrics and were carried forward for analysis and the UMAP method was employed to visualize cells based on their gene expression profiles. A k-nearest neighbors graph combined with the Louvain algorithm identified 16 distinct clusters corresponding to neuronal and non-neuronal cell types (Figure 1A).

The FindAllMarkers function was used to identify distinct marker genes for each cluster, allowing annotation of these clusters with their corresponding cell types based on either previous knowledge or comparison with publicly available in situ hybridization (ISH) data (Allen Institute for Brain Science, 2023). After selecting the top marker genes, a dot plot was used to visually represent these markers across clusters, with color indicating expression levels and size reflecting the percentage of cells expressing each gene (Figure 1B). These markers include both general and specific markers that are indicative of various subtypes of neurons and other lineages, with all expected cell types identified. The UMAP highlights the most interesting cell populations with regard to identifying impacts of *Dlx1/2* cKO, those from the MGE, CGE, and olfactory regions,

which are positioned in the top right of the plot. Additionally, scVelo illustrates the developmental trajectories based on RNA velocity, which is a metric that compares spliced to unspliced reads (Bergen et al., 2020), showing neural progenitors progressing towards their respective fates, including a specific trajectory from progenitors towards the olfactory bulb region (Figure 1C).

Our primary focus is on regions directly affected by the conditional knockout (cKO) of the *Dlx1* and *Dlx2* genes (*Dlx1/2* cKO). To visualize these cell types, we used FeaturePlot to highlight the regions expressing *Dlx1* on the UMAP, along with *VGAT* gene expression, which serves as a marker for GABAergic neurons, thus confirming identity of our cells of interest as GABAergic inhibitory interneurons (Figure 1D). The cell types expressing these markers include MGE-derived cortical interneurons (CINs), CGE-derived CINs, olfactory bulb interneurons, and cortical VZ and SVZ progenitors (of which a proportion of are associated with generation of pallial olfactory bulb interneurons). These cell-types/clusters of interest were then further sub-clustered for higher resolution analysis.

Differential Gene Expression in MGE, CGE, and olfactory bulb lineages

Focusing on regions with significant *Dlx1* and *Dlx2* expression, we further refined the original clusters through subclustering, revealing dividing cells on the right side of the UMAP, secondary intermediate progenitors in the SVZ, and differentiating cells including cortical interneuron cell populations that originated in the MGE and CGE, and cells within the olfactory development lineage moving from less to more mature on the

left (Figure 1E). The identification of each cell cluster and trajectory is supported by the heatmap displaying curated, cell-type-specific marker genes (Figure 1F). As anticipated, each cell type demonstrates specific expression of its marker genes, indicated by the yellow color representing high gene expression.

We then conducted an analysis of cell-type-specific differentially expressed (DE) genes in the MGE and CGE derived cortical interneurons, as well as the most mature classes of olfactory bulb interneurons using FindMarkers function in Seurat package with Wilcoxon Rank-Sum Test. For this analysis, cKO samples and wild-type samples were compared to determine the lineage specific impacts of *Dlx1/2* ablation. Using volcano plots, we show DE genes within each sub-cluster, highlighting those with significant upregulation or downregulation (Figure 2A), with a selected list of genes annotated on the graph. As expected *Dlx1* and *2* consistently showed downregulation, particularly within the olfactory lineage. Additionally, in both MGE and CGE the downregulation of *Hbb-bs* and *Hba-a1* was observed, while in OB cell type both genes are upregulated. It is known that *Hbb-bs* and *Hba-a1* are known to be hemoglobin subunits associated with oxygen transport, and studies have also shown that they play a role in the nervous system, particularly under stress conditions, where their downregulation may be involved in neurogenesis and neuroprotection (Weaver et al., 1981; Stankiewicz et al., 2014; Sannino et al., 2016). The downregulation of these genes in the MGE and CGE, coupled with their upregulation in the olfactory bulb, indicates a different influence under *Dlx1/2* cKO among cell types.

There were few neuron-specific genes detected to be differentially expressed only in MGE, as most MGE DE genes are mitochondrial and ribosomal. *Calb1* was one of the genes, though with non-significant adjusted p value (and with raw p value less than 0.05), and picked for its importance in MGE-derived interneurons (Kubota et al., 2011; Bershteyn et al., 2023; Zhao et al., 2008).

VIP+ and *Reln*+ interneurons are both important CGE-derived CINs (Hu et al., 2017), with both genes downregulated in CGE only. Some other genes only DE in the CGE includes *Cnr1*, *Crh*, *Cxcl14*, *Tac2*, and *Cck*, which are which are involved in neuronal development, migration, and signaling. Several genes, including *Kctd13*, *PISD*, *Rgs10*, *Snca*, and *Vsnl1*, were downregulated in both the MGE and CGE. These genes are commonly involved in synaptic function, signaling, and neurodevelopment. In the olfactory bulb, *Cdk8*, *Colec12*, and *Hexb* were downregulated, and are associated with cell cycle regulation, immune response, and lysosomal function, respectively.

These genes, along with *Hbb-bs*, *Hba-a1*, and *Dlx1/2*, are visualized for their differential expression as box plots in Figure 3A. All genes, except *Calb1*, show significant down-regulation, as illustrated in the box plots after excluding zero counts. Figure 3B further examines the proportion of cells that either express or do not express each gene. As shown, a higher percentage of cells do not express these genes under the *Dlx1/2* cKO condition. While *Hbb-bs* and *Hba-a1* are upregulated in the olfactory bulb, their overall trend remains downregulated when combined with the CGE and MGE populations.

The conditional knockout of *Dlx1/2* results in the downregulation of a wide array of genes across various cell types, including those specifically associated with *Dlx* function and GABAergic interneurons, as well as genes with more general roles in cellular processes. This widespread gene downregulation highlights the critical role of *Dlx1/2* in maintaining the proper expression of key developmental genes, and underscore the importance of *Dlx1/2* in the development and function of GABAergic interneurons. Considering the patterns of DE, we found that genes required for maturation and synaptogenesis, such as *Reln* and *Vip* important for more mature form of cortical interneuron development, are not expressed at the expected levels in CGE, which suggests that these neurons will not be able to participate normally in the local cortical circuitry. Other than that, some genes less related to early brain development such as *Cxcl14*, are shown to react to Dlx12/ cKO in CGE.

Analysis of the olfactory bulb lineage trajectory

In the previous result sections, we demonstrated DE for maturing olfactory GABAergic inhibitory interneurons, as well as the MGE- and CGE-derived cortical interneurons. At P3, the Dlx1/2 lineages from the developing basal ganglia (i.e. the MGE and CGE derived cells) have migrated into the cortex, so these lineages are already in a fairly mature state. In contrast, olfactory bulb interneurons are still undergoing neurogenesis from the cortical progenitor zones in the VZ/SVZ. Thus, the full lineage is captured in pseudotime for these cells. In Figure 4A, the single-cell velocity after excluding MGE and CGE regions revealed this clear developmental trajectory, with cells progressing

from the RGC region through G1/S phase SVZ cells into the olfactory bulb cortical interneurons. To further refine the analysis, we removed the G2/M phase SVZ cells and performed reclustering on the remaining clusters. Using Slingshot, we computationally modelled the cell trajectory, represented as pseudotime (Figure 4B). This pseudotime plot illustrates the sequential stages of olfactory bulb interneuron development, providing a temporal framework for understanding cell differentiation along this pathway.

Next, we selected several differentially expressed (DE) genes that were downregulated in the olfactory bulb, and used generalized additive modeling (GAM) to examine how this downregulation occurred across the developmental trajectory. Specifically, we utilized the fitGAM function from the scVelo package, which fits a GAM model to the gene expression data across the pseudotime trajectory, allowing us to visualize the expression trends over time. The selected genes—*Slc32a1*, *Arx*, and *Colec12*—were analyzed, with their expression levels displayed across the trajectory and corresponding feature plots across the UMAP. We did not identify specific pseudotime points showing a sudden or more pronounced downregulation compared to others. Instead, the genes exhibit a smooth, continuous decline in expression throughout the pseudotime, indicating a steady process rather than a sharp transition at any particular stage. This suggests that olfactory bulk interneuron markers are not expressed at appropriate levels in these cells across the developmental trajectory. This is consistent with previous studies that have found that Dlx1/2 cKO in this model does not impact the number of cells produced in these lineages, but rather that cell identity is perturbed (Silberberg et al., 2016).

Bulk ChIP-seq Analysis

In this section, we investigate the genomic binding patterns of the key transcription factor Dlx2 using bulk ChIP-seq analysis. This study complements our scRNA-seq findings by examining how Dlx2 interacts with the genome across different brain regions and developmental stages.

Comparative Dlx2 binding analysis

We followed our previously published methods and used the same Dlx2 antibody to perform ChIP-seq from E16 and P3 olfactory bulb and cortex (Catta-Preta et al., 2023). Tissues were microdissected and then processed for ChIP-seq and the libraries were sequenced following our established protocol. For this analysis we also reanalyzed our previously published E13.5 and E16.5 basal ganglia (BG) Dlx2 ChIP-seq. Following alignment to the mouse genome, we used MACS2 for peak calling, we identified peaks in the E16 BG, cortex, and olfactory bulb, as well as in the P3 cortex and olfactory bulb. Example ChIP-seq coverage at the Gad1 locus for all Dlx2 datasets is shown in Figure 5a. Not unexpectedly considering the differences in developmental stage and Dlx2 expression, the peaks identified in the E16 basal ganglia significantly outnumber those in other regions. This is likely because Dlx1/2 plays a crucial role in both neurogenesis and the migration and maturation of GABAergic interneurons during this developmental period, which predominantly occurs in the ganglionic eminences of the BG.

To refine our analysis, we first filtered out large peaks where it is unclear where the specific binding targets are, defined as those longer than 3 kb. The remaining peaks across all datasets were then merged into a single BED file using Bedtools. We used HOMER's findMotifsGenome and then annotatePeaks tools to classify the peaks into those containing the Dlx2 motif and those that do not and compared peaks identified in OB and cortex datasets to our previously published BG data (Figure 5b). The Dlx motif across samples are correctly identified with intense TAATTA box, with a few other motifs in each sample identified (figure 5C). We identified 11615 peaks with presumed Dlx2 binding to a primary TAATTA motif, and 7688 where no primary motif was identified, and these peaks are further divided based on their distance from any TSS annotated by RefSeq dataset, with those less than 2kb away as proximal and those further away as distal. We identified 1521 proximal and 10094 distal peaks with Dlx2 binding, and 1835 proximal and 5853 distal peaks without Dlx2 binding (figure 5B). There was a substantially higher overlap among peaks featuring a primary motif with peaks identified in embryonic BG reported in Lindtner et al., 2019, suggesting increased role for indirect interactions between Dlx2 and regulatory targets in later stages of GABAergric neuron maturation.

Shared versus distinct Dlx2 targets by developmental stage and brain region

Using the four sets of peaks described above, we generated heatmaps with deepTools, employing computeMatrix and plotHeatmap. In order to assess changes over time and identify regions where binding patterns have remained consistent or have undergone alterations, we incorporated E13 basal ganglia bigwig file from previous studies as an

additional heatmap column. This resulted in four heatmaps representing clustering for distal/proximal and the presence/absence of Dlx2 motif, providing a visual representation of Dlx2 binding patterns across different genomic contexts (Supplemental Figure 1). We selected clusters of the peaks with representative termporal or tissue-specific patterns of interest and displayed these in Figure 5D. Illustrating these pattens at the locus level (Figure 5A), *Gad1* shows strong binding in E13.5 and E16.5 BG and also E16.5 Cx but less in other conditions, while *Gorasp2* has the opposite binding pattern, resulting these Dlx2 binding events to be classified into different clusters. Gene Ontology (GO) analysis showed that functional enrichment for specific GO terms is much stronger for the distal peak clusters, where terns like neurogenesis, cell proliferation, and migration are highly represented (Figure 5E). We highlight a few patterns of interest that offer insights into Dlx2 gene regulatory function (Figure 5D). First, as the bw files are not normalized to the same strength, we assigned Dist_Mot_8 and Dist_Mot_9 as constitutive binding peaks, interpreting them as having Dlx2 bind under all tissue and timepoints. These two clusters are Dlx2 distal targets enriched for primary binding to the TAATTA motif, and include genes associated with GABAergic maturation, suggesting binding by Dlx even before gene expression and then continued binding throughout maturation. Using these sets as standard, we found Dlx2 targets that are biased toward earlier BG binding that are associated more strongly with neurogenesis, again with enrichment for direct binding to the Dlx motif (Prox_Mot_11 and Dist_Mot_14). We identified one distal cluster that is enriched in OB compared to cortex and appears to be enriched for OB-relevant GO terms (Dist_Mot_13). Finally, we identified two sets of P3 specific binding events with and

without the Dlx motif (Dist_Mot_15, Prox_No_12). These peaks did not show as strong of functional enrichment with regard to target genes and also did not feature specific GO terms compared to more broadly bound clusters. Terms related to apoptosis were enriched in stage and tissue specific cortex distal clusters with no Dlx2 motif detected, cluster 14 and 15 with peaks (Dist_No_15, Dist_No_14) highest enriched in e16.5 BG and P3 cortex respectively (figure 5D-E), indicating possible indirect binding of Dlx2 regulates apoptotic processes in these tissues. These analyses provide a deeper understanding of roles of Dlx2 in gene regulation during neurodevelopment and offer insights into how its binding patterns correlate with the transcriptional changes observed in our scRNA-seq data.

Discussion:

Dlx1/2 have been widely recognized as crucial transcription factors during early neurogenesis, particularly in the development, differentiation, and migration of GABAergic interneurons, where they regulate other transcription factors essential for these processes (Rubenstein et al., 2024). However, there is a growing interest in understanding the roles that Dlx1/2 play beyond these initial stages, such as in SVZ progenitor cells and post-mitotic cells in the MZ. Early studies have also highlighted the epigenomic functions of Dlx1/2, focusing on how these genes regulate chromatin state and gene expression during early development. For instance, Lindtner et al. (2019) used a Dlx1/2 double-knockout model and found that disruptions in chromatin structure led to aberrant gene expression in the SVZ and MZ of ganglionic eminence. However, despite the study's focus on non-VZ regions, the double knockout also affected Dlx1/2

in the VZ, leading to a widespread knockout that depleted Dlx5 and other critical factors, resulting an enlarged LGE with smaller MGE and CGE regions (Silberberg et al., 2016).

To address this issue and enable identification of phenotypes after this initial role in neurogenesis, I12b-Cre can be used, where it specifically induces a conditional knockout of the Dlx genes in the SVZ and MZ regions starting from e10.5, without affecting their expression in the VZ. Pla et al. (2018) employed the I12b-Cre line and demonstrated that the absence of Dlx1/2 during these later stages significantly impacts neuronal maturation, survival, synaptic development, and overall circuit integration, while the early tangential migration of interneurons remains unaffected. These findings suggest novel roles for Dlx1/2 in GABAergic interneurons that extend beyond their previously established functions. Through a combination of single-cell RNA sequencing (scRNA-seq) and bulk ChIP-seq on Dlx2, we have uncovered molecular mechanisms on how Dlx1/2 conditional knock may influence later stages of CIN neurodevelopment, providing initial insights into the potential roles those down-regulated genes play in maintaining cellular functions and developmental trajectories.

The findings in this chapter offer insights into the role of *Dlx1* and *Dlx2* in the development of GABAergic interneuron subpopulations across various regions of the mouse brain, including the MGE, CGE, and olfactory bulb. In the DE analysis, many informative genes were identified, including both previously verified and newly discovered loci associated with specific cell subtypes responding to *Dlx1/2* cKO. Though many of the newly-identified genes are known to be important in neuronal

functions, responsiveness towards *Dlx1/2* cKO in CGE and Olfactory bulb GABAergic interneurons are newly discovered. Among these are genes downregulated in both the MGE and CGE including *Kctd13*, *Snca*, and *Vsnl1*, which share a common feature in their involvement in synaptic function and signaling. *Kctd13* and *Snca* are particularly important for synaptic plasticity and vesicle trafficking, which is critical for the formation and maintenance of synapses (Escamilla et al., 2017; Rey et al., 2021; Lin et al., 2015). These and other downregulated synaptic genes suggest decreased synapse-related functions in the two ganglionic regions. *Pisd* was also found to be downregulated in both the CGE and MGE regions, and it is a causal gene for Liberfarb syndrome, a condition that affects brain development. In our *Dlx1/2* cKO model, this downregulation could contribute to the neurological aspects of Liberfarb syndrome (Peter et al., 2015). In the CGE, we observed significant downregulation of *Reln* and *Vip*, which are known to be vital for the maturation of CGE-derived cortical interneurons (reviews Folsom and Fatemi 2012; reviews Williams and Riedemann 2021). Their reduced expression suggests that CGE-derived interneurons in the *Dlx1/2* cKO model may be disturbed for their normal functions. Overall, this scRNA-seq study provides an insight on *how Dlx1/2* cKO on BG impacts GABAergic interneuron development through the disturbed expression on down-stream gene across relevant BG-derived CIN and OB populations.

DLX1/2 transcription factors are known for orchestrating a complex gene regulatory network, with their binding patterns varying across different developmental stages (Lindtner et al., 2019; Catta-Preta et al., 2023). In our analysis, we categorized tissue and stage specific BG, Cortex, and OB Dlx2 binding peaks based on the presence of

the Dlx2 motif and their proximity to the nearest TSS, and further annotated their enriched GO terms. We observed distinct Dlx2 binding patterns that varied by developmental time points and brain regions, with GO terms related to neuronal development enriched in distal regions. Our findings add new insights to existing literature, enhancing our understanding of how Dlx2 orchestrates gene regulation during neurodevelopment.

Limitations of the study

While numerous genes were identified as differentially expressed in the Dlx1/2 cKO model with strong effects and statistical significance, our findings require further experimental validation to verify DE and establish their true biological impact. We used only two replicates, and expansion to more samples would increase power for identifying true DE genes and reducing the potential false positive results. While not included in this dissertation, some of these DE genes have already verified in-situ by our collaborator and one has been further followed up at the functional consequence level, and these experiments will be a part of the final project manuscripts. Another limitation arises from the scRNA-seq analysis, where the use of the KNN machine-learning clustering method may have led to potential misassignment of cells to incorrect clusters, introducing biases or inaccuracies in the interpretation of cellular identities and gene expression patterns. Finally, our scRNA-seq findings represent a statis picture, and there are likely to be further changes at later stages of development and maturation. Regarding the ChIP-seq work, the Dlx2 binding regions lacking the Dlx2 motif are interpreted as secondary binding; however, we currently lack evidence to

confirm either true interactions or the possibility of proximity to direct Dlx2 binding sites, which limits our understanding of these interactions. Further, ChIP-seq experiments can have variable signal to noise, resulting in differences in peak calling and in identification and clustering of clusters based on signal. We were limited here to the use of single replicates for each stage-tissue comparison. Finally, as ChIP-seq requires substantial numbers of cells and we did not have a method to enrich for specific cell populations, we were unable to resolve cell-type specificity within samples. Overall, these limitations suggest that while the study provides valuable insights, further research and validation are necessary to fully understand the roles of Dlx1/2 and the gene regulatory networks they influence.

Summary

This project explored the role of Dlx1/2 TFs in neurodevelopment by combining scRNA-seq and bulk ChIP-seq analyses. The findings revealed that Dlx1/2 are crucial for the proper development, differentiation, and maturation of GABAergic interneurons, with region-specific impacts on gene expression across multiple brain regions, including MGE, CGE, and olfactory bulb, and identified numerous differentially expressed genes under Dlx1/2 cKO. Bulk ChIP-seq analysis further highlighted the dynamic and context-dependent binding patterns of Dlx2, with distinct peaks identified across different developmental stages and brain regions. Our findings give new understanding regarding the continued requirement for Dlx1/2 after neurogenesis and the mechanism of regulation based on genomic interactions.

We propose a model based on our results, where Dlx1/2 binds directly to distal enhancer regulatory targets in the embryonic BG, activating GABAergic lineage transcriptional programs and repressing other fates. As CINs and maturing olfactory bulk interneurons mature further, Dlx1/2 binding is maintained at a small number of likely critical distal enhancer regulatory targets (i.e. the constitutively bound clusters from our ChIP-seq analysis), as well expanded indirect interactions at promoters and frequently within genes. We propose that some of these interactions are critical for maintenance of cell specification and proper CIN/OB interneuron maturation. In addition, Dlx2 gains a small set of new distal targets in P3 CINs which are involved in repression of apoptosis (and likely other processes). Based on the DE findings, loss of Dlx2 binding causes a failure to mature and form productive synapses and a de-repression of apoptosis, which leads to the effects that were previously found in post-neurogenic Dlx1/2 cKO and functional studies (Cobos, et al. 2005; Seybold et al. 2012; Pla et al., 2018; de Melo et al., 2005).

The research underscored the importance of Dlx2 in regulating gene networks associated with neuronal development, particularly when binding with distal regulatory elements. The project not only provided insights into the direct binding targets of Dlx2 but also suggested possible secondary regulatory roles, enriching our understanding of Dlx1/2's influence on brain development and GABAergic interneuron maturation.

Figures:

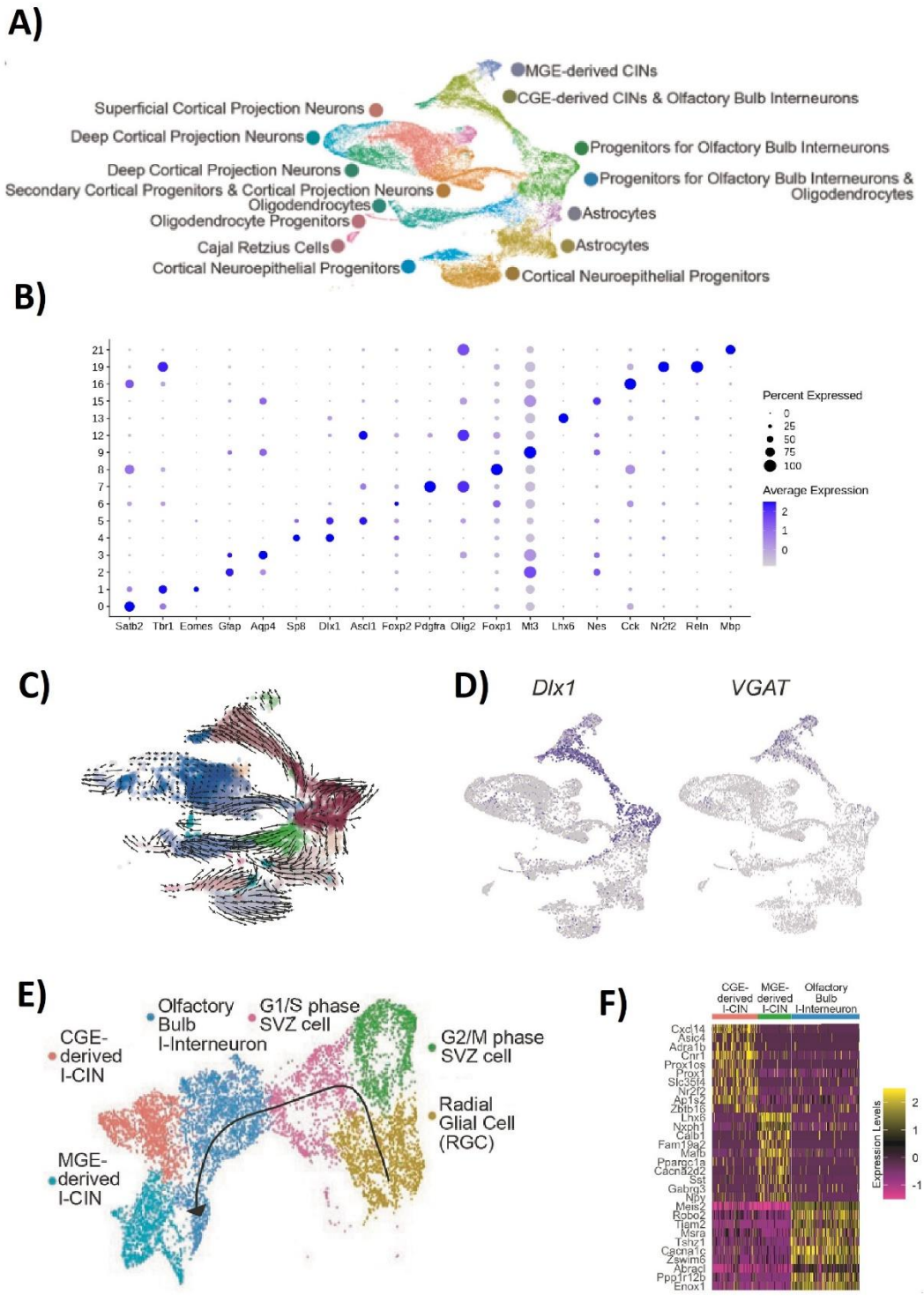

A)
MGE-derived CINs
CGE-derived CINs & Olfactory Bulb Interneurons
Superficial Cortical Projection Neurons
Deep Cortical Projection Neurons
Progenitors for Olfactory Bulb Interneurons
Deep Cortical Projection Neurons
Progenitors for Olfactory Bulb Interneurons & Oligodendrocytes
Secondary Cortical Progenitors & Cortical Projection Neurons
Oligodendrocytes
Astrocytes
Oligodendrocyte Progenitors
Cajal Retzius Cells
Astrocytes
Cortical Neuroepithelial Progenitors
Cortical Neuroepithelial Progenitors
B)
Satb2 Tbr1 Eomes Gfap Aqp4 Sp8 Dlx1 Ascl1 Foxp2 Pdgfra Olig2 Foxp1 Mt3 Lhx6 Nes Cck Nr2f2 Reln Mbp
Percent Expressed
Average Expression
C)
D)
Dlx1
VGAT
E)
Olfactory Bulb I-Interneuron
G1/S phase SVZ cell
CGE-derived I-CIN
G2/M phase SVZ cell
Radial Glial Cell (RGC)
MGE-derived I-CIN
F)
CGE-derived I-CIN
MGE-derived I-CIN
Olfactory Bulb I-Interneuron
Expression Levels

Figure 1: scRNA-seq identifies Dlx1/2 cellular lineages in P3 cortex. Following scRNA-seq performed on WT and Dlx1/2 cKO P3 cortex, cells with reads passing quality control were kept, clustered, and visualized in UMAP, with irrelevant clusters removed, and then subclustered based on Dlx-expressing regions indicative of cortical interneurons. The resulting clusters of cells are shown in A), where the cell type annotation is based on marker genes. Some of these marker genes are highlighted in B) as dot plot, displaying the expression of selected genes (x-axis) across different clusters (y-axis). Cluster numbers in the dot plot correspond to the annotated cell types from the UMAP: 0 and 16: superficial cortical projection neurons; 1: secondary cortical progenitors and cortical projection neurons; 3 and 15: astrocytes; 4: CGE-derived CINs and olfactory bulb interneurons; 5: progenitors for olfactory bulb interneurons; 6 and 8: deep cortical projection neurons; 7: oligodendrocytes; 9: cortical neuroepithelial progenitors; 12: progenitors for olfactory bulb interneurons and oligodendrocytes; 13: MGE-derived CINs; 19: Cajal-Retzius cells; 21: oligodendrocyte progenitors. The size of the dots reflects the percentage of cells within each cluster expressing the gene, and the color indicates the average expression level. C) displays cell velocity, indicating trends in cellular dynamics. D) shows feature plots of *Dlx1* and *VGAT* expression, where clusters expressing these two genes are subclustered in E), which illustrates the subclusters with a manually drawn line indicating the developmental trajectory of olfactory bulb interneurons and presence of CGE- and MGE-derived CINS. The cell annotation is supported by F), which is a heatmap displaying the expression of key marker genes across MGE, CGE, and OB regions.

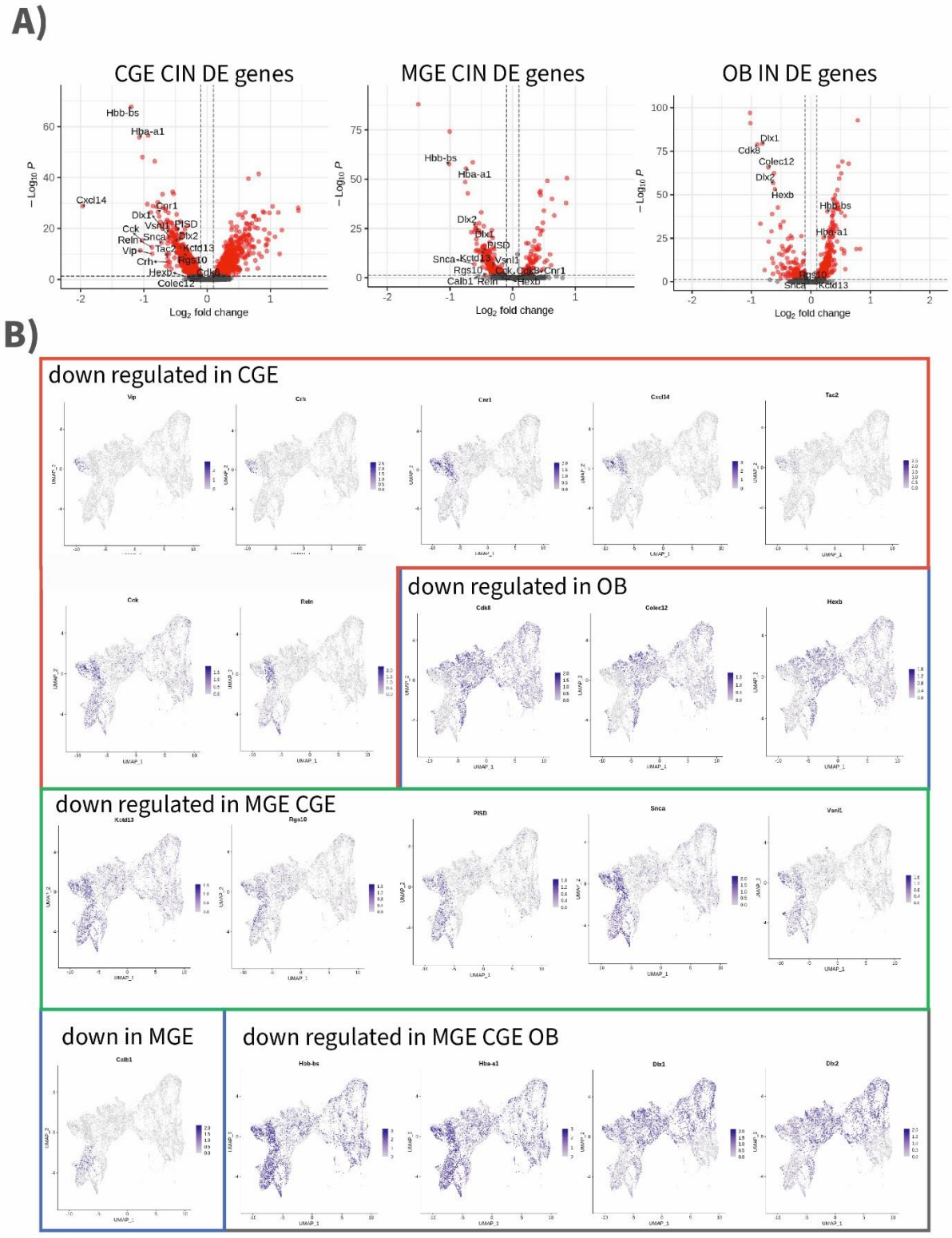

A)
CGE CIN DE genes
MGE CIN DE genes
OB IN DE genes
Hbb-bs
Hba-a1
Cxcl14
Cnr1
Dlx1
Vsnl1
PlSD
Cck
Snca
Dlx2
Reln
Kctd13
Tac2
Vip
Rgs10
Crh
Hexb
Cdk8
Colec12
Calb1
Log2 fold change
B)
down regulated in CGE
down regulated in OB
down regulated in MGE CGE
down in MGE
down regulated in MGE CGE OB

Figure 2: Differential expression in MGE/CGE-derived CINS and OB interneurons. This figure presents the analysis of differentially expressed genes identified from the dataset. A) displays three volcano plots, highlighting DE genes from MGE, CGE, and OB, with a log2 fold change (log2FC) threshold of 0.01 and an adjusted p-value of 0.05. The expression of these genes are further explored in B), where feature plots illustrate their spatial expression patterns across the UMAP clusters.

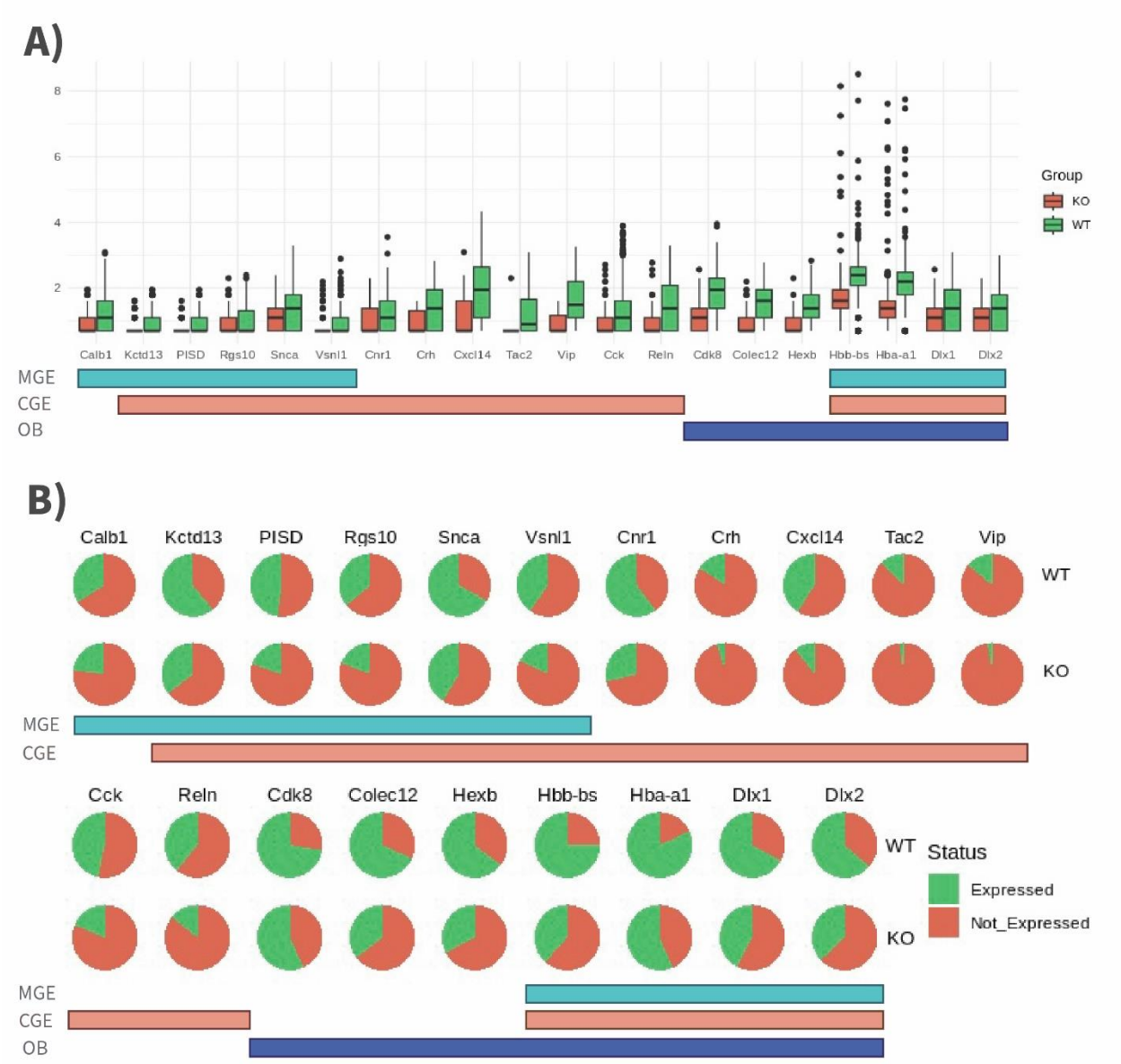

A)
8
6
4
2
Group
KO
WT
Calb1
Kctd13
PISD
Rgs10
Snca
Vsnl1
Cnr1
Crh
Cxcl14
Tac2
Vip
Cck
Reln
Cdk8
Colec12
Hexb
Hbb-bs
Hba-a1
Dlx1
Dlx2
MGE
CGE
OB
B)
Calb1
Kctd13
PISD
Rgs10
Snca
Vsnl1
Cnr1
Crh
Cxcl14
Tac2
Vip
WT
KO
MGE
CGE
Cck
Reln
Cdk8
Colec12
Hexb
Hbb-bs
Hba-a1
Dlx1
Dlx2
WT
KO
Status
Expressed
Not_Expressed
MGE
CGE
OB

Figure 3: DEGs following Dlx1/2 cKO ablation. This figure presents the differential expression analysis and cell expression proportions for selected DEGs. In both panels. for genes associated with multiple cell types, the pie chart aggregates data from all cells in those cell types. The box plot in A) visualizes the expression levels of these genes under the SCT assay, with zero counts removed. The pie chart in B) illustrates the proportion of cells that either express or do not express each gene.

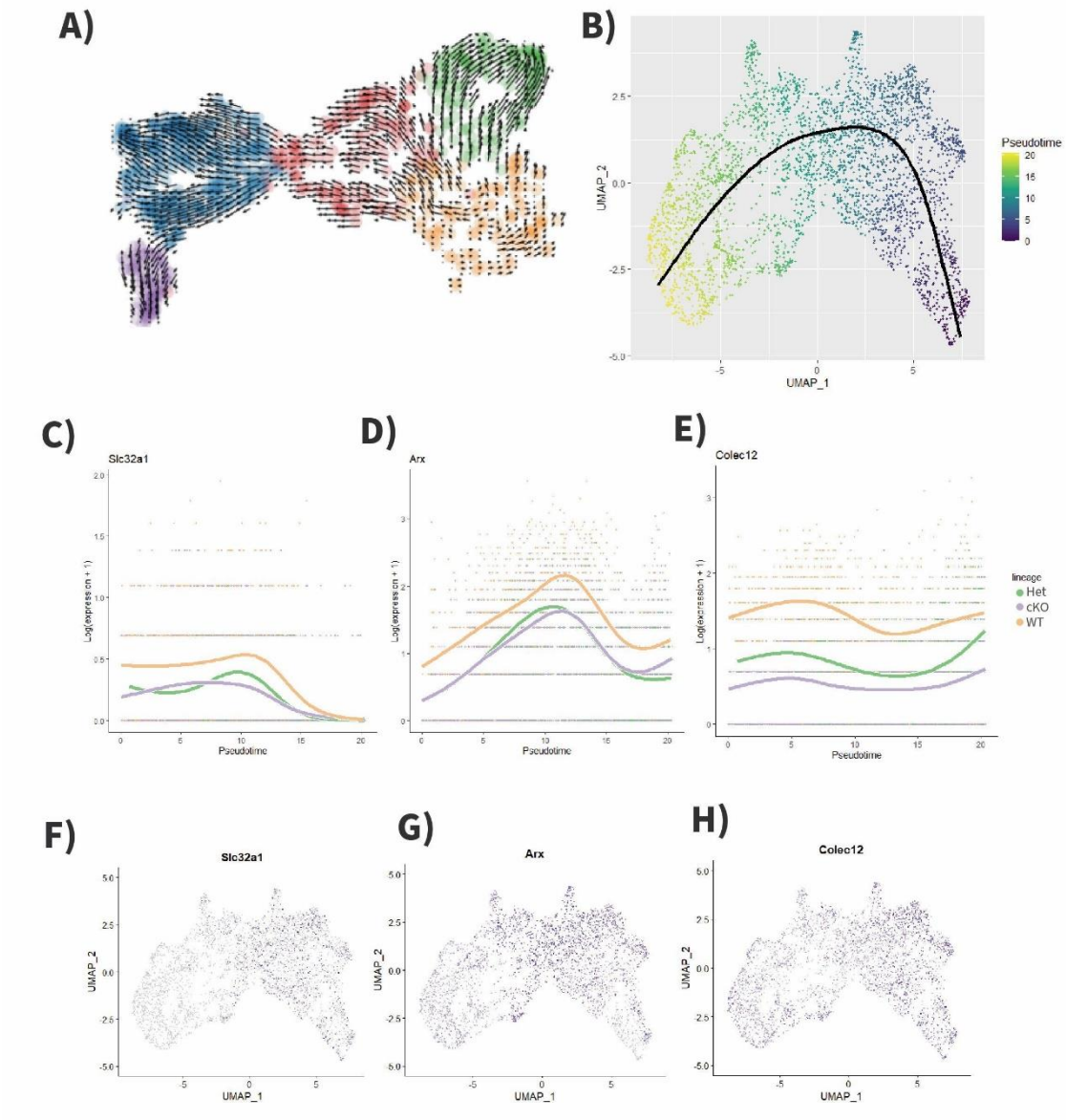

A)
B)
Pseudotime
20
15
10
5
0
UMAP_2
UMAP_1
C)
Slc32a1
D)
Arx
E)
Colec12
Log(expression + 1)
Pseudotime
lineage
Het
cKO
WT
F)
Slc32a1
G)
Arx
H)
Colec12
UMAP_2
UMAP_1

Figure 4: Differential pseudotime trajectory of OB interneuron neurogenesis.

This figure illustrates the developmental trajectory and gene expression analysis of olfactory bulb interneurons in the *Dlx1/2* cKO model. A) highlights the RNA velocity and inferred trajectories of cells as they progress through development. B) presents the pseudotime trajectory, re-clustered, and provides a temporal framework. C)-E) display the gene expression levels of *Slc32a1*, *Arx*, and *Colec12* across pseudotime, with colors indicating the genotype. F)-H) map the spatial expression of these genes on the UMAP.

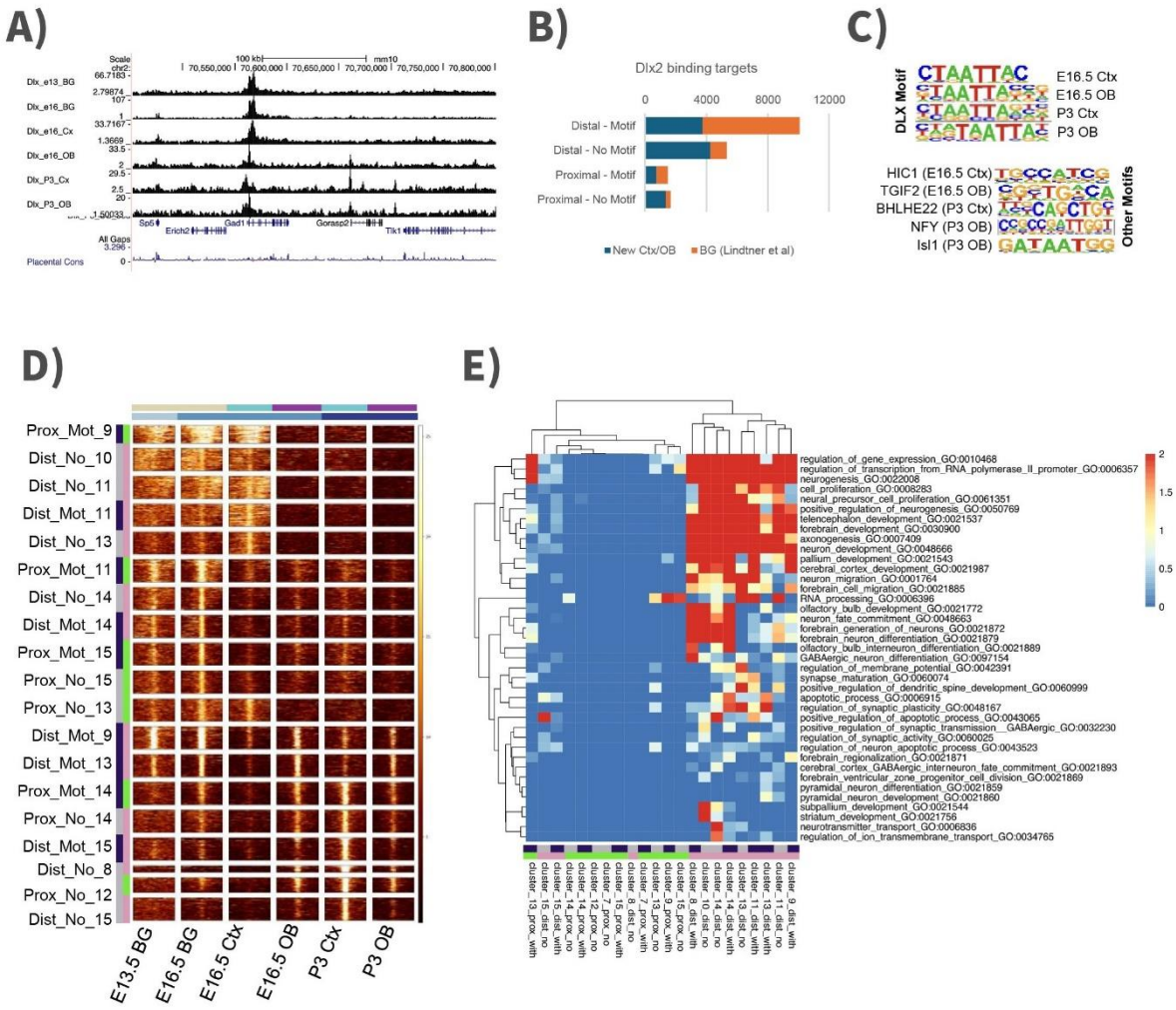
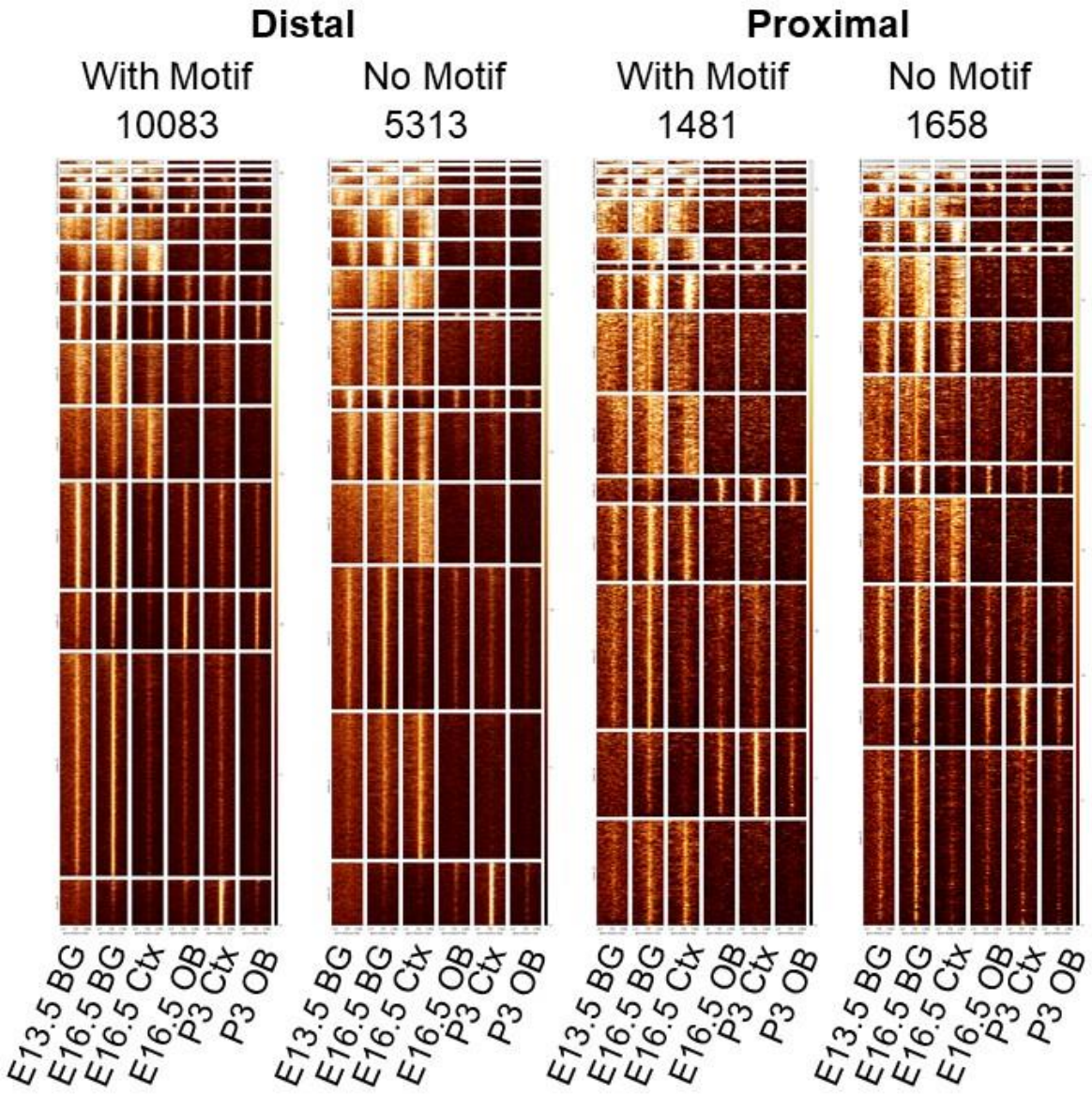

A)
B)
C)
Dlx2 binding targets
0
4000
8000
12000
Distal - Motif
Distal - No Motif
Proximal - Motif
Proximal - No Motif
New Ctx/OB
BG (Lindtner et al)
DLX Motif
E16.5 Ctx
E16.5 OB
P3 Ctx
P3 OB
HIC1 (E16.5 Ctx)
TGIF2 (E16.5 OB)
BHLHE22 (P3 Ctx)
NFY (P3 OB)
Isl1 (P3 OB)
Other Motifs
D)
E)
Prox_Mot_9
Dist_No_10
Dist_No_11
Dist_Mot_11
Dist_No_13
Prox_Mot_11
Dist_No_14
Dist_Mot_14
Prox_Mot_15
Prox_No_15
Prox_No_13
Dist_Mot_9
Dist_Mot_13
Prox_Mot_14
Prox_No_14
Dist_Mot_15
Dist_No_8
Prox_No_12
Dist_No_15
E13.5 BG
E16.5 BG
E16.5 Ctx
E16.5 OB
P3 Ctx
P3 OB
regulation_of_gene_expression_GO:0010468
regulation_of_transcription_from_RNA_polymerase_II_promoter_GO:0006357
neurogenesis_GO:0022008
cell_proliferation_GO:0008283
neural_precursor_cell_proliferation_GO:0061351
positive_regulation_of_neurogenesis_GO:0050769
telencephalon_development_GO:0021537
forebrain_development_GO:0030900
axonogenesis_GO:0007409
neuron_development_GO:0048666
pallium_development_GO:0021543
cerebral_cortex_development_GO:0021987
neuron_migration_GO:0001764
forebrain_cell_migration_GO:0021885
RNA_processing_GO:0006396
olfactory_bulb_development_GO:0021772
neuron_fate_commitment_GO:0048663
forebrain_generation_of_neurons_GO:0021872
forebrain_neuron_differentiation_GO:0021879
olfactory_bulb_interneuron_differentiation_GO:0021889
GABAergic_neuron_differentiation_GO:0097154
regulation_of_membrane_potential_GO:0042391
synapse_maturation_GO:0060074
positive_regulation_of_dendritic_spine_development_GO:0060999
apoptotic_process_GO:0006915
regulation_of_synaptic_plasticity_GO:0048167
positive_regulation_of_apoptotic_process_GO:0043065
positive_regulation_of_synaptic_transmission,_GABAergic_GO:0032230
regulation_of_synaptic_activity_GO:0060025
regulation_of_neuron_apoptotic_process_GO:0043523
forebrain_regionalization_GO:0021871
cerebral_cortex_GABAergic_interneuron_fate_commitment_GO:0021893
forebrain_ventricular_zone_progenitor_cell_division_GO:0021869
pyramidal_neuron_differentiation_GO:0021859
pyramidal_neuron_development_GO:0021860
subpallium_development_GO:0021544
striatum_development_GO:0021756
neurotransmitter_transport_GO:0006836
regulation_of_ion_transmembrane_transport_GO:0034765
2
1.5
1
0.5
0

Figure 5: Dlx2 ChIP-seq identifies differential binding in BG, Cortex, and OB

This figure shows Dlx2 binding patterns across various conditions. A) displays example loci where Dlx2 binding significantly differs between the conditions on UCSC genome browser. B) quantifies the peaks categorized by the presence of the Dlx2 motif and their proximity to known genes. The orange bars represent the number of peaks found in both this dataset and previously detected datasets in Lindtner et al., 2019. C) illustrates the Dlx2 motif detected across different samples and other motifs enriched. D) showcases selected notable patterns of peaks from Supplemental Figure 1, with up to 100 peaks per cluster chosen based on distinct binding patterns, with their different experimental conditions noted on the x and y axes. E) GO terms associated with the peaks assigned to the clusters selected in D, using the entire peak set depicted in Supplemental Figure 1.


Distal
Proximal
With Motif
10083
No Motif
5313
With Motif
1481
No Motif
1658
E13.5 BG
E16.5 BG
E16.5 Ctx
E16.5 OB
P3 Ctx
P3 OB

Supplementary Figure 1: Dlx2 genomic interactions across BG, Cortex and OB.

This figure visualizes the entire peak set, divided into four plots based on the presence or absence of the Dlx2 motif and their proximity to known genes, with the color intensity in each plot represents the strength of the peaks. Selected clusters were plotted in main Figure 5.

Material and Methods:

Sample Collection and Single-Cell RNA Sequencing

Brain samples were collected from P3 mice, including both wild-type and Dlx1/2 conditional knockout (cKO) models (Lindtner et al., 2019), with frontal sections dissected for analysis. Although not included in the final differential expression analysis, heterozygous (het) samples were also collected and included for cell clustering. Single-cell sequencing was prepared from these brain samples and subsequently processed using the 10x Chromium single-cell RNA sequencing platform. Sequencing data will be made available in the GEO repository.

Data Processing

Raw sequencing data was aligned and processed using the 10x Genomics Cell Ranger pipeline (Zheng et al., 2017) with alignment to the mm10 mouse reference genome with intron-mode activated. After alignment, data was filtered based on specific thresholds for gene expression and total RNA counts to exclude low-quality cells and potential doublets. For each experimental condition and replicate, different thresholds were used to tailor the filtering process to the specific characteristics of each dataset. Additionally, the SoupX package (Young et al., 2020) was employed to adjust for ambient RNA contamination, ensuring that the gene expression profiles reflect true intracellular mRNA levels. Filtered datasets from each condition and replicate were normalized and integrated using the SCTransform method (Hafemeister et al., 2019), which accounts for sequencing depth and batch effect. Principal component analysis (PCA) was

performed to reduce the dimensionality of the data and capture the primary sources of variation, with the first 30 principal components selected for further analysis, and these were used to generate a Uniform Manifold Approximation and Projection (UMAP) plot, which provided a two-dimensional visualization of the cells. Finally, a shared nearest neighbor graph was constructed, and cells were clustered using the Louvain algorithm.

Cell Type Identification and Marker Gene Analysis

To identify the distinct cell populations, an unbiased approach was first employed using the FindAllMarkers function from the Seurat package, which compares each cluster against all other clusters. These markers were visualized using heatmaps, where clusters with strongly expressed positive markers were highlighted in yellow, allowing for an initial, unsupervised identification of major cell types based on their unique gene expression profiles.

In addition to the unbiased marker identification, a set of manually selected known marker genes was utilized to refine the identification of specific cell types, particularly in cases where clusters shared similar expression profiles and were difficult to distinguish solely based on the unbiased approach. These known markers were chosen based on their established roles in defining specific cell types in the brain. The expression patterns of these markers were visualized using dot plots, where the size of the dots indicated the percentage of cells expressing each gene within a cluster, and the color gradient represented the average expression level. This combined approach of

unbiased marker discovery and manual marker validation ensured robust and accurate cell type identification across the different clusters.

Subclustering

Dlx1/2 expressing clusters were selected with their raw counts subtracted from the original Seurat object, and normalized and integrated using the SCTransform method (Hafemeister et al., 2019), which accounts for sequencing depth and batch effect. Principal component analysis (PCA) was performed to reduce the dimensionality of the data and capture the primary sources of variation, with the first 30 principal components selected for further analysis, and these were used to generate a Uniform Manifold Approximation and Projection (UMAP) plot, which provided a two-dimensional visualization of the cells. Finally, a shared nearest neighbor graph was constructed, and cells were clustered using the Louvain algorithm.

Differential expression

Differentially expressed genes between wild-type and *Dlx1/2* cKO samples were identified using the Wilcoxon rank-sum test using FindMarkers, implemented within the Seurat package (Hao et al., 2023). The analysis was conducted on data processed using the SCTransform method. The results of the DE analysis were visualized using volcano plots through EnhancedVolcano R package (Blighe et al., 2024).

RNA Velocity Analysis

RNA velocity analysis was conducted using the scVelo method to explore the dynamic developmental trajectories of cells. RNA velocity estimates the future state of cells by modeling the transition from unspliced to spliced mRNA, providing insights into the direction and speed of cellular differentiation processes. The Seurat object with UMAP information were loaded into an AnnData object, and the spliced and unspliced RNA matrices were derived from Loom files generated through velocyto. The data then underwent filtering and normalization using filter_and_normalize from scvelo package to ensure high-quality inputs for velocity estimation, followed by the calculation of moments for each cell. RNA velocities were then estimated using the stochastic mode in scVelo, with the resulting velocity graph embedded in the UMAP space to represent the directional flow of cells through different developmental states.

ChIP-seq analysis

Brain samples from E16.5 and P3 mice are dissected and FAC-sorted into BG and OB and Cx in the lab of John Rubenstein and libraries were prepared and sequenced at the UCSF core. ChIP-seq was performed as described previously using the same antibodies (Lindtner et al., 2019). The ChIP-seq data was aligned to the mouse mm10 reference genome using Bowtie2, with bam files came from the publication from Lindtner et al., 2019. Following alignment, peak calling was conducted using the MACS2 (Zhang et al., 2008) software under default parameters, and bedgraph (bdg) files generated at the same time are used to transform into bw using bedGraphToBigWig (Kent et al., 2010).

Peaks identified by MACS2 were further processed to refine the analysis. Large peaks greater than 3 kb were filtered out. and the remaining peaks were categorized based on the presence or absence of the Dlx2 motif using HOMER and the `findMotifsGenome` and `annotatePeaks` tools, and then merged across different samples using Bedtools merge. These peaks were further classified based on their proximity to the nearest transcription start site (TSS) using the RefSeq gene annotation database. Peaks located within 2 kb of a TSS were designated as proximal, while those further away were classified as distal. Heatmaps were created using deepTools computeMatrix and plotHeatmap (Ramírez et al., 2016) with the four peak sets and bigwig files used as input. Gene Ontology functional enrichment was performed using the GREAT algorithm as implemented in the rGREAT package (Gu et al., 2022).

Data and Code Availability

All scRNA-seq and ChIP-seq datasets will be deposited in GEO, with both raw fastq and processed expression matrix and annotated peak files uploaded. Code is available via the NordLab github site or via request.

References

Anderson, S. A., Marín, O., Horn, C., Jennings, K., & Rubenstein, J. L. (2001). Distinct cortical migrations from the medial and lateral ganglionic eminences. Development, 128(3), 353-363.

Alvarez-Buylla, A., & Garcıa-Verdugo, J. M. (2002). Neurogenesis in adult subventricular zone. Journal of Neuroscience, 22(3), 629-634.

Barrero, F. J., Ampuero, I., Morales, B., Vives, F., de Dios Luna Del Castillo, J., Hoenicka, J., & Garcia Yebenes, J. (2005). Depression in Parkinson's disease is related to a genetic polymorphism of the cannabinoid receptor gene (CNR1). The pharmacogenomics journal, 5(2), 135-141.

Bershteyn, M., Bröer, S., Parekh, M., Maury, Y., Havlicek, S., Kriks, S., ... & Nicholas, C. R. (2023). Human pallial MGE-type GABAergic interneuron cell therapy for chronic focal epilepsy. Cell Stem Cell, 30(10), 1331-1350.

Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nature biotechnology, 38(12), 1408-1414.

Blighe K, Rana S, Lewis M (2024). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.22.0, https://github.com/kevinblighe/EnhancedVolcano.

Calvigioni, D., Máté, Z., Fuzik, J., Girach, F., Zhang, M. D., Varro, A., ... & Harkany, T. (2017). Functional differentiation of cholecystokinin-containing interneurons destined for the cerebral cortex. Cerebral cortex, 27(4), 2453-2468.

Catta-Preta, R., Lindtner, S., Ypsilanti, A., Price, J., Abnousi, A., Su-Feher, L., ... & Nord, A. S. (2023). Combinatorial transcription factor binding encodes cis-regulatory wiring of forebrain GABAergic neurogenesis. bioRxiv.

Cobos I, Calcagnotto ME, Vilaythong AJ, Thwin MT, Noebels JL, Baraban SC, Rubenstein JL. 2005. Mice lacking Dlx1 show subtype-specific loss of interneurons, reduced inhibition and epilepsy. Nat Neurosci. 8:1059–1068.

Colasante, G., Collombat, P., Raimondi, V., Bonanomi, D., Ferrai, C., Maira, M., ... & Broccoli, V. (2008). Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. Journal of Neuroscience, 28(42), 10674-10686.

de Melo, J., Du, G., Fonseca, M., Gillespie, L. A., Turk, W. J., Rubenstein, J. L., & Eisenstat, D. D. (2005). Dlx1 and Dlx2 function is necessary for terminal differentiation and survival of late-born retinal ganglion cells in the developing mouse retina.

Escamilla, C. O., Filonova, I., Walker, A. K., Xuan, Z. X., Holehonnur, R., Espinosa, F., ... & Powell, C. M. (2017). Kctd13 deletion reduces synaptic transmission via increased RhoA. Nature, 551(7679), 227-231.

Folsom T. D. and Fatemi S. H. (2012) The involvement of Reelin in neurodevelopmental disorders. Neuropharmacology 68, 122–135.

Geigerseder, C., Doepner, R., Thalhammer, A., Frungieri, M. B., Gamel-Didelon, K., Calandra, R. S., ... & Mayerhofer, A. (2003). Evidence for a GABAergic system in rodent and human testis: local GABA production and GABA receptors. Neuroendocrinology, 77(5), 314-323.

Gu Z, Huebschmann D (2022). “rGREAT: an R/Bioconductor package for functional enrichment on genomic regions.” Bioinformatics. doi:10.1093/bioinformatics/btac745.

Guo, T., Liu, G., Du, H., Wen, Y., Wei, S., Li, Z., ... & Yang, Z. (2019). Dlx1/2 are central and essential components in the transcriptional code for generating olfactory bulb interneurons. Cerebral cortex, 29(11), 4831-4849.

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome biology, 20(1), 296.

Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R (2023).

Hu, J. S., Vogt, D., Sandberg, M., & Rubenstein, J. L. (2017). Cortical interneuron development: a tale of time and space. Development, 144(21), 3867-3878.

Katona, I., & Freund, T. F. (2008). Endocannabinoid signaling as a synaptic circuit breaker in neurological disease. Nature medicine, 14(9), 923-930.

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., & Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics, 26(17), 2204-2207.

Kubota, Y., Shigematsu, N., Karube, F., Sekigawa, A., Kato, S., Yamaguchi, N., ... & Kawaguchi, Y. (2011). Selective coexpression of multiple chemical markers defines discrete populations of neocortical GABAergic neurons. Cerebral cortex, 21(8), 1803-1817.

Lindtner, S., Catta-Preta, R., Tian, H., Su-Feher, L., Price, J. D., Dickel, D. E., ... & Rubenstein, J. L. (2019). Genomic resolution of DLX-orchestrated transcriptional circuits driving development of forebrain GABAergic neurons. Cell reports, 28(8), 2048-2063.

Lin, C. W., Chang, L. C., Tseng, G. C., Kirkwood, C. M., Sibille, E. L., & Sweet, R. A. (2015). VSNL1 co-expression networks in aging include calcium signaling, synaptic plasticity, and Alzheimer's disease pathways. Frontiers in psychiatry, 6, 30.

Long, J. E., Garel, S., Alvarez-Dolado, M., Yoshikawa, K., Osumi, N., Alvarez-Buylla, A., & Rubenstein, J. L. (2007). Dlx-dependent and-independent regulation of olfactory bulb interneuron differentiation. Journal of Neuroscience, 27(12), 3230-3243.

Lledo, P. M., Alonso, M., & Grubb, M. S. (2006). Adult neurogenesis and functional plasticity in neuronal circuits. Nature Reviews Neuroscience, 7(3), 179-193.

McKinsey, G. L., Lindtner, S., Trzcinski, B., Visel, A., Pennacchio, L. A., Huylebroeck, D., ... & Rubenstein, J. L. (2013). Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons. Neuron, 77(1), 83-98.

Merlo, G. R., Zerega, B., Paleari, L., Trombino, S., Mantero, S., & Levi, G. (2000). Multiple functions of Dlx genes. International Journal of Developmental Biology, 44(6), 619-626.

Panganiban, G., & Rubenstein, J. L. (2002). Developmental functions of the Distal-less/Dlx homeobox genes.

Peter, V. G., Quinodoz, M., Pinto-Basto, J., Sousa, S. B., Di Gioia, S. A., Soares, G., ... & Campos-Xavier, B. (2019). The Liberfarb syndrome, a multisystem disorder affecting eye, ear, bone, and brain development, is caused by a founder pathogenic variant in the PISD gene. Genetics in medicine, 21(12), 2734-2743.

Pla, R., Stanco, A., Howard, M. A., Rubin, A. N., Vogt, D., Mortimer, N., ... & Rubenstein, J. L. (2018). Dlx1 and Dlx2 promote interneuron GABA synthesis, synaptogenesis, and dendritogenesis. Cerebral cortex, 28(11), 3797-3815.

Potter, G. B., Petryniak, M. A., Shevchenko, E., McKinsey, G. L., Ekker, M., & Rubenstein, J. L. (2009). Generation of Cre-transgenic mice using Dlx1/Dlx2 enhancers and their characterization in GABAergic interneurons. Molecular and Cellular Neuroscience, 40(2), 167-186.

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., ... & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research, 44(Web Server issue), W160.

Rey, F., Pandini, C., Messa, L., Launi, R., Barzaghini, B., Zangaglia, R., ... & Carelli, S. (2021). α-Synuclein antisense transcript SNCA-AS1 regulates synapses-and aging-related genes suggesting its implication in Parkinson's disease. Aging Cell, 20(12), e13504.

Rubenstein, J. L., Nord, A. S., & Ekker, M. (2024). DLX genes and proteins in mammalian forebrain development. Development, 151(11), dev202684.

Sannino, G., Pasqualini, L., Ricciardelli, E., Montilla, P., Soverchia, L., Ruggeri, B., ... & Hardiman, G. (2016). Acute stress enhances the expression of neuroprotection-and neurogenesis-associated genes in the hippocampus of a mouse restraint model. Oncotarget, 7(8), 8455.

Sekine K., Kubo K. and Nakajima K. (2014) How does Reelin control neuronal migration and layer formation in the developing mammalian neocortex? Neurosci. Res. 86, 50–58

Seybold BA, Stanco A, Cho KK, Potter GB, Kim C, Sohal VS, Rubenstein JL, Schreiner CE. 2012. Chronic reduction in inhibition reduces receptive field size in mouse auditory cortex. Proc Natl Acad Sci USA. 109:13829–13834.

Silberberg, S. N., Taher, L., Lindtner, S., Sandberg, M., Nord, A. S., Vogt, D., ... & Rubenstein, J. L. (2016). Subpallial enhancer transgenic lines: a data and tool resource to study transcriptional regulation of GABAergic cell fate. Neuron, 92(1), 59-74.

Stankiewicz, A. M., Goscik, J., Swiergiel, A. H., Majewska, A., Wieczorek, M., Juszczak, G. R., & Lisowski, P. (2014). Social stress increases expression of hemoglobin genes in mouse prefrontal cortex. BMC neuroscience, 15, 1-16.

Stühmer, T., Puelles, L., Ekker, M., & Rubenstein, J. L. (2002). Expression from a Dlx gene enhancer marks adult mouse cortical GABAergic neurons. Cerebral cortex, 12(1), 75-85.

Sur, M., & Rubenstein, J. L. (2005). Patterning and plasticity of the cerebral cortex. Science, 310(5749), 805-810

Tao, R., Li, C., Jaffe, A. E., Shin, J. H., Deep-Soboslay, A., Yamin, R. E., ... & Kleinman, J. E. (2020). Cannabinoid receptor CNR1 expression and DNA methylation in human prefrontal cortex, hippocampus and caudate in brain development and schizophrenia. Translational psychiatry, 10(1), 158.

Weaver, S., Comer, M. B., Jahn, C. L., Hutchison III, C. A., & Edgell, M. H. (1981). The adult β-globin genes of the “single” type mouse C57BL. Cell, 24(2), 403-411.

Whitman, M. C., & Greer, C. A. (2009). Adult neurogenesis and the olfactory system. Progress in neurobiology, 89(2), 162-175.

Williams, R. H., & Riedemann, T. (2021). Development, diversity, and death of MGE-derived cortical interneurons. International Journal of Molecular Sciences, 22(17), 9297.

Wonders, C. P., & Anderson, S. A. (2006). The origin and specification of cortical interneurons. Nature Reviews Neuroscience, 7(9), 687-696.

Young, M. D., & Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience, 9(12), giaa151.

Zerucha, T., Stühmer, T., Hatch, G., Park, B. K., Long, Q., Yu, G., ... & Ekker, M. (2000). A highly conserved enhancer in the Dlx5/Dlx6Intergenic region is the site of cross-regulatory

interactions betweenDlx genes in the embryonic forebrain. Journal of neuroscience, 20(2), 709-721.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology, 9, 1-9.

Zhao, Y., Flandin, P., Long, J. E., Cuesta, M. D., Westphal, H., & Rubenstein, J. L. (2008). Distinct molecular pathways for development of telencephalic interneuron subtypes revealed through analysis of Lhx6 mutants. Journal of Comparative Neurology, 510(1), 79-99.

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., ... & Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. Nature communications, 8(1), 14049.

Chapter 3

An epigenetic atlas of *in vitro* human cortical interneuron differentiation

a. Preface

This chapter is part of a study modeling dynamic epigenetic changes and predicting stage-specific enhancers during human embryonic stem cell (ESC) differentiation into MGE-derived interneurons, conducted with Dr. John Rubenstein at UCSF and a private company that is developing interneuron transplantation therapies, Neurona Inc. The work here specifically covers bioinformatic analysis of ATAC-seq, ChIP-seq, and RNA-seq data, as well as enhancer testing performed by my colleagues in the Nord Lab. All wet lab experiments to generate the genomics data were carried out by personnel in the Rubenstein lab and at Neurona.

b. Acknowledgements

Wet lab work for genomics datasets was performed by Dr. Susan Lindtner in Dr. John Rubenstein's lab and Dr. Marina Bershstyn from the Neurona company. I thank the former Nord Lab Ph.D. Rinaldo Catta-Preta for support with ChIP-seq analysis. Cory Ardekani and Ana Santos in the Nord Lab carried out the enhancer assays in the final figure.

c. Author contributions

Data was generated by experiments performed by Susan Lindtner and Marina Bershstyn, and cells provided by the Neurona company. Cory Ardekani and Ana Santos performed the cloning and enhancer testing of the SST1 enhancer. I performed all computational analysis included in this chapter, and drafted figures and text with consultation with my PI, Dr. Alex Nord.

d. Declaration of interests

I have nothing to declare. Dr. John Rubenstein is a co-founder of Neurona. Neurona provided the differentiating cells for this project.

Abstract:

GABAergic cortical interneurons (CINs) arise in embryonic basal ganglia (BG) and play critical roles in cortical circuits. CIN dysfunction is implicated in neurodevelopmental and neurological disorders. The surgical transplantation of immature CINs into the brain represents a promising new therapy, with ongoing clinical studies for treatment-refractory epilepsy. One challenge with this method is deriving pure populations of specific CINs, such as the somatostatin and parvalbumin subtypes that primarily arise from embryonic medial ganglionic eminence (MGE). Here, we sought to characterize dynamic genomic changes in an *in vitro* model of human MGE-derived CIN differentiation. We integrate ATAC-seq, ChIP-seq for H3K27ac, H3K4me3, and H3K27me3, and RNA-seq, to map chromatin states and gene co-expression modules from embryonic stem cells to immature MGE-derived CINs (D0 to D42). We used ChromHMM to segment the genome into dynamic accessibility states from ATAC-seq data and stage-specific chromatin state using histone ChIP-seq data. The results of this temporal epigenomic atlas capture the transcriptional and epigenetic changes differentiating cells go through, revealing stage specific enhancers that are targeted by TFs active in MGE-derived CIN specification and gene co-expression modules underlying specification changes. We verified activity of a newly predicted enhancer of Sst via AAV-mediated enhancer reporter assay, showing activity in Lhx6+ CINs in postnatal mouse cortex. Our results represent a valuable resource for understanding, evaluating, and improving MGE-derived CIN differentiation models and overall understanding of the genomic control of these important cells.

Introduction:

The central nervous system (CNS) is an intricate network composed of a diverse array of neuronal and glial cell types, each playing crucial roles in maintaining the balance between excitation and inhibition that underpins normal brain function. Among these, MGE-derived GABAergic cortical inhibitory interneurons (CINs) hold significant potential for treating neurological disorders and brain injuries, particularly by restoring the balance of excitation and inhibition in damaged neural circuits where they integrate into existing networks, potentially reducing abnormal neural activity, such as that seen in epilepsy (Bershteyn et al., 2023; Fandel et al., 2016). Understanding the development of these CINs is therefore not only fundamental to basic neuroscience but also holds significant therapeutic potential.

Interneurons are a heterogeneous group of inhibitory neuron subtypes that regulate the activity of principal excitatory neurons within the CNS, including in the cortex. The medial ganglionic eminence (MGE), a transient embryonic structure in the basal ganglia, is the primary source of several classes of cortical interneurons, including Somatostatin positive (Sst) and Parvalbumin positive (Pvalb) subtypes that play distinct roles in cortical microcircuits. These interneurons follow a highly orchestrated process of neurogenesis, maturation and tangential migration, and eventual integration into circuits in the developing cortex (Hu et al., 2017). The differentiation of CINs from progenitor cells in the MGE involves a complex interplay of genetic and epigenetic mechanisms and intrinsic and extrinsic controls that orchestrate the precise timing and expression of

genes critical for guiding the differentiation of progenitor cells into specific neuronal subtypes (Allis & Jenuwein, 2016).

Human and mouse CIN specification is thought to follow similar genetic control, with the activity of a shared set of master lineage-defining transcription factors. The mouse MGE starts generating interneuron progenitors around embryonic day 10.5 (E10.5). During this period, Nkx2-1 is essential for specifying MGE progenitors by distinguishing them from those in the lateral ganglionic eminence (LGE). Nkx2-1 activates Lhx6, a transcription factor crucial for interneuron development. By E12.5, Dlx1 and Dlx2 play pivotal roles in promoting the tangential migration of progenitors from the MGE to the developing cortex via established migratory streams through the subventricular zone (SVZ) and intermediate zone (IZ) (Sussel et al., 1999; Vogt et al., 2014; Anderson et al., 1997; Wichterle et al., 2001; Lim et al., 2018). Between E13.5 and E16.5, MGE-derived interneurons continue migrating to the cortical plate. During this time, Lhx6 guides interneurons to their appropriate cortical layers, ensuring the development of parvalbumin (PV) and somatostatin (SST) expressing neurons. These neurons are essential for inhibitory control within cortical circuits. Nkx2-1 continues to maintain MGE identity and represses alternative cell fates. Dlx1 and Dlx2, expressed during migration, help extend dendrites and axons, facilitating integration into cortical circuits (Lim et al., 2018; Paul et al., 2017; Du et al., 2008; Cobos et al., 2007).

The gene regulatory networks that these TFs regulate depend on enhancers and epigenetic state. Enhancers are distal cis-regulatory elements that enhance the

transcription of target genes in a manner that is both cell-type and stage-specific (Heinz et al., 2015). Enhancer activity is determined by the combination of TF binding and epigenetic regulation, with specific histone modifications marking different functional states of the genome and impacting the availability of DNA for interactions TFs and transcriptional co-factors that are critical for transcription. For instance, H3K27ac is associated with active enhancers and is a key marker for regions of the genome that are transcriptionally active (Creyghton et al., 2010), H3K4me3 is strong at promoters and, when combined with H3K27ac, can identify active or transcribed enhancers, H3K4me3 along with H3K27me3 is a signature of bivalent promoters, which occur at developmental gene and can indicate a poised state. H3K27me3 alone is associated with transcriptional silencing and repressed chromatin, highlighting regions of the genome that are epigenetically inactivated. While often used as a proxy for active regulatory elements, genome accessibility is a more general signature of regulatory DNA, as accessibility by itself indicates protein-DNA interactions, but does not differentiate regulatory direction (Tsompana & Buck, 2014; Klemm et al., 2023). Nonetheless, changes in accessibility have proven a reliable if imperfect marker of dynamic regulatory element (e.g. enhancer) activity (Kilpinen et al., 2023).

Neurona is a company that was founded with the mission to develop CIN-based transplantation therapy, and they have developed an optimized proprietary protocol for CIN production from human embryonic stem cells (ESCs) based on the genetic logic described above. In pre-clinical studies in mice and early stages clinical trials in humans, Neurona's MGE-derived CIN transplantation therapy has shown promise in

reducing seizures in treatment-refractory epilepsy (Bershteyn et al., 2023). One limitation of improving treatment using approaches that depend on transplanted derived cells is the production of relatively pure populations of specific defined CINs that might be most effective. Towards this goal, a high-resolution understanding of the CIN differentiation time course could provide new insights into more targeted or effective derivation protocols. More broadly, better understanding of the epigenomic and transcriptomic characteristics of CINs and identification of key regulatory elements and driver genes.

To address these needs, we integrated epigenomic and transcriptomic assays and performed regulatory modeling to profile differentiation dynamics using the Neurona MGE-derived CIN protocol. We generated a gene regulatory atlas for human CIN differentiation, identifying relevant regulatory DNA elements and annotating dynamic epigenetic state. As proof-of-principle, we demonstrate the functional validity of our predictions via AAV-mediated enhancer-reporter assay in postnatal mouse cortex.

Results:

Generation of epigenomic datasets and genome state segregation via ChromHMM

MGE-derived GABAergic inhibitory cortical interneurons derived by Neurona were provided for six developmental stages: D0 (ESCs), D7 (telencephalon-like cells), D14 (VZ-like MGE cells), D28 (MGE VZ/SVZ), D35 (MGE SVZ), and D42 (post-mitotic interneurons), and these stages represent a progression from embryonic stem cells

(ESCs) to more differentiated MGE-derived interneurons (Figure 1A). Using epigenomic methods, including ATAC-seq and histone ChIP-seq targeting H3K27ac, H3K4me3, and H3K27me3 histone modification marks, we sought to uncover the dynamic and complex landscape across these time points. As epigenetic patterns are more informative when multiple representative datasets are modelled together, we used chromHMM which utilizes the combinatorial presence and absence of binarized signal presence/absence to segment the genome into distinct states, each potentially associated with different regulatory functions. To comprehensively map these dynamics, we performed two separate chromHMM analyses. The first analysis integrated three histone modifications (H3K27ac, H3K27me3, H3K4me3) across six time points (D0, D7, D14, D28, D35, D42), while the second analysis utilized ATAC-seq data to define chromatin states based on chromatin accessibility across the six time points. This allowed us to systematically map the dynamics of chromatin states throughout the differentiation process of ESCs into MGE-derived interneurons.

In the histone chromHMM analysis, eight distinct chromatin states were identified based on the presence and intensity of three histone modifications: H3K27ac, H3K4me3, and H3K27me3 as shown in Figure 1B. In these states:

- State 1 displays a weak signal of the repressive H3K27me3 mark, indicating regions that are mildly repressed or transitioning towards repression.
- State 2 has strong H3K27me3 signal, representing highly repressive regions.
- State 3 with strong H3K4me3 and H3K27me3 reflects a bivalent poised state.

- State 4 is marked by strong signals for H3K27ac, H3K4me3, and H3K27me3, suggesting a bivalent active state, often found in developmental genes that are actively transcribed but tightly regulated.
- State 5 with strong H3K4me3 indicates transcriptionally active promoters and enhancers. Notably, we found this state flanking open histone states, possibly caused by overall stronger and wider peak of H3K4me3 compared to H3K27ac.
- State 6 shows strong signals for both H3K27ac and H3K4me3, marking strongly transcribed promoters or enhancers involved in active gene expression.
- State 7 with weak H3K27ac signal suggests weak or poised enhancers.
- Lastly, state 8 is characterized by the absence of significant signals for any of these histone marks, indicating likely quiescent or inactive regions.

The ATAC-seq temporal chromHMM analysis also identified eight distinct chromatin states shown in Figure 1C. States 7 represent null states, where no significant chromatin accessibility was detected. State 8 had weaker accessibility in the middle of the time course. States 1 and 2 correspond to regions that become accessible later in development, while states 5 and 6 are associated with regions that open earlier in the developmental timeline. State 3 is characterized by regions that remain accessible both early and late in the development process, and State 4 represents regions that are constitutively open across all stages of development.

For both ATAC-seq temporal and histone-based chromatin state analyses, the majority of the genome is classified by ChromHMM as the null state, as shown in the genomic

composition of chromHMM ATAC-seq data and a representative example of histone chromHMM (D35) in Figure 1D. The chromHMM histone genomic profiles across different days exhibit a similar composition, hence only one time point is shown as an example. Other regions such as CpG islands, exons, transcription start sites (TSS), and 2kb flanking regions around TSSs show a higher prevalence of open and bivalent histone marks and more constant open chromatin states in the ATAC-seq data.

Dynamic chromosomal states during CIN neuronal development

The epigenomic marks effectively captured dynamic chromosomal states during neuronal development, as evidenced by both histone and ATAC chromHMM analyses. Figure 2A and 2B depict the same region around the *Mafb* gene, shown in the UCSC Genome Browser with two putative enhancers highlighted in blue and green. *Mafb* is a known regulator of later CIN maturation (Pai et al., 2020). The chromHMM states captured expected dynamics and reveal likely *Mafb*-associated enhancers, with the *Mafb* gene region consistently marked by ATAC as constitutively open, the histone marks show a transition from a bivalent state, initially surrounded by strong repressive histone marks, to a fully active histone state as development progresses. Additionally, the two putative enhancers in Figures 2A and 2B, as well as the enhancer highlighted in Figure 2C and the two putative enhancers in Figure 2D, exhibit more subtle chromatin modifications. These regions display a late-open state in chromHMM, indicated by bright green, with histone marks transitioning to more open states over time compared to earlier stages.

As illustrated in Figures 2B-D, the enhancers are located in regions where the genome is in an accessible state at specific timepoints, here later in differentiation marked by ATAC-seq chromHMM state 1. Based on these patterns, we used ATAC-seq peaks and associated temporal accessibility states from chromHMM as anchors for examining the directional specific epigenetic dynamics at CIN-relevant regulatory DNA elements. More specifically, we were interested in the relationship between accessibility and active or repressed chromatin states.

Figure 3B displays the histone chromHMM states across the six time points (with one replicate per time point shown). Overall, the H3K27me3 represents a large proportion of histone modifications in genome, consistent with the majority of regulatory elements in the genome being repressed in any given cell lineage. The coincidence of H3K27me3 and accessible chromatin is not surprising, but nonetheless somewhat notable, as there is a general assumption that accessibility is associated with active regulatory elements.

Because many ATAC-seq peaks are in close proximity, we merged those regions that are directly connected, and set as priority state rank: 1, 2, 4, 3, 5, 6, 8, with 1 being highest priority, meaning if this ATAC-region composed of multiple ATAC states would be assigned to as the state of highest priority. Epigenomic marks around gene bodies appear more complex and unpredictable, complicating interpretation of regulatory element function. Therefore, to simplify the situation, we filtered out a 3kb upstream and downstream region from gene bodies in the GENCODE dataset to reduce this

complexity. This filtering removed the majority of anchor regions, with the remaining sets visualized in orange in Figure 3A. After this exclusion, the number of bivalent and open chromatin states decreased significantly (Figure 3C). Modeling these intergenic ATAC-seq peaks, the constitutive open region (ATAC state 4) and the two late-open (state 1 and state 2) are picked as representative sets. The constitutively open ATAC-seq state contains the largest proportion of open histone chromatin states among all anchor sets, with the lowest proportion of H3K27me3, as expected. There is a notable trend towards increasing proportion of active chromatin states at later stages, suggesting that early accessibility is decoupled from regulatory activity of these elements and therefore that ATAC-seq data is less informative. The "open late" ATAC state, specifically state1 where the four enhancers are anchored, shows fewer histone modifications in ESCs compared to later time points. Also, as time grows, the amount of open state histone and weak histone state grows, while the closed histone state remains relatively constant after Day7. This suggests that many of these late active regulatory elements are in a quiescent state compared to being actively repressed. This is aligned with the overall pattern of low levels of histone post-translational modifications in ESC chromatin and increasing proportion of the genome that is identified with the histone ChIP-seq as differentiation progresses. These findings show that ATAC-seq chromHMM signatures alone have a low sensitivity for enhancer epigenetic state. Some regions that display an open ATAC state may still have a closed histone chromHMM state, even some of them captures regions where histone state jumps to relatively open state (Figure 3D bottom region in heatmap). Therefore, incorporating histone chromHMM states alongside ATAC-seq data improves the reliability of enhancer

predictions, ensuring that regions with open chromatin are more confidently annotated. Overall, integration of ATAC-seq and histone ChIP-seq across the six timepoints yielded a detailed and stage-resolved regulatory DNA atlas, which will be of value to the research community. We will make these results accessible via the Nord Lab github website and via sharing access to the data for visualization on the UCSC genome browser.

Stage-specific gene co-expression modules during CIN differentiation

RNA-seq analysis was conducted on differentiating cells following standard stranded polyA-selection methods with 2-3 replicates per time point, and differential expression data was analyzed using DESeq2. Principal Component Analysis (PCA) of the RNA-seq samples shows that their positions in PCA space align well, with PC1 effectively separating the samples by time points and strong concordance between replicates (Figure 4A). We verified expected changes in genes that have been well-studied in CIN development, observing strong match to expected patterns, for example of sequential rises of *Nkx2.1*, *Dlx2*, *Lhx6*, *Mafb*, and finally *Sst*. These findings are consistent with the initial publication describing this differentiation protocol (Bershteyn et al. 2023). To further explore gene expression patterns, we used Weighted Gene Co-expression Network Analysis (WGCNA) to assign genes to 17 modules based on the covariance of expression patterns, with the cluster dendrogram shown in Figure 4B.

Most modules exhibited patterns of developmental change as visualized through bar plots (Figure 4C). Two prominent modules, lightcyan and blue, exhibited consistent

directional change across the timepoints, with lightcyan expression increasing as development progresses and blue expression decreasing (Figure 4C). All timepoints with the exception of D35 had a module with peak expression at the specific timepoint, indicating that these timepoints capture dynamic processes that might have been lost with less frequent sampling. Module based clustering was generally capitulated using a k-means approach (Figure 4E). Functional analysis testing for Gene Ontology (GO) enrichment showed that gene modules/k-means clusters were associated with shared biological processes and pathways. As example relevant terms associated with the blue module (Figure 4D) and lightcyan module (Figure 4F) are plotted using dot plots. The blue module shows gene expression decreasing overtime is enriched for processes related to translation, such as tRNA synthesis and energy generation. In contrast, the cluster with genes overlapping with lightcyan module shows increased expression level as development progresses and is enriched with GO terms related to synapse formation and axon development, suggesting that these genes are involved in key neural developmental processes. Many other modules captured similarly distinct and interesting functional enrichment and represent module and gene-level markers for in vitro differentiation stage.

Enhancer verification

We selected a set of predicted enhancers for functional testing both In vitro using the same MGE-derived CIN protocol as well as in vivo in postnatal mouse cortex. Putative enhancer sequences were cloned from the human genome into testing backbones. While testing is ongoing, we have completed proof-of-principle work confirming a series

of four enhancers associated with the Sst locus (Figure 5A). For these experiments, we packaged the enhancer reporter construct, which included a betaGlobin minimal promoter and GFP, using the PhP.eB AAV serotype and delivered the packaged virus via neonatal intracranial injection. Example results are shown for one of these enhancers which at P14 had the highest specificity and sensitivity for Lhx6+ cells, a marker of MGE-derived CINs. This enhancer is located 223 kb downstream of the SST gene, highlighted in red in Figure 2D, and depicted in Figure 5A-B, the co-localization of GFP and Lhx6 suggests that this enhancer may play a role in regulating SST+ interneurons, confirming the reliability of our predictions. We expect that validation of additional enhancers, both active in later stage maturing CINs as well as other timepoints, will be useful for labeling and manipulating stage-specific differentiating CIN populations, e.g. via driving a recombinase or biosensor in a restricted subset of cells during CIN differentiation.

Discussion

The findings in this chapter offer key insights into the regulatory dynamics of MGE-derived interneuron development, focusing on the use of integrated epigenomic modeling to predict putative enhancers and map gene co-expression modules that are transiently active during this process. Through a combination of ATAC-seq and histone ChIP-seq, we have mapped chromatin states across six distinct developmental time points, revealing regions of the genome potentially involved in regulatory processes. Our gene regulatory atlas expands on previous similar work in several ways. First, the

fidelity of the MGE-derived CIN protocol is more specific compared to other publications that generate less pure populations based on markers such as *Sst*. Second, we sampled at shorter intervals and with both ATAC-seq and histone ChIP-seq, yielding a more granular and comprehensive perspective. While single cell methods (e.g. snRNA-seq and snATAC-seq) can provide additional information and better capture heterogeneity, these methods do not yet extend to capture chromatin state at the level of histone post-translational modifications, which we show provide critical information regarding regulatory element activity and chromatin state. Finally, we demonstrate validity of enhancer activity predictions. With SST#1 enhancer validated in Lhx6-expressing neurons, and further validation across other putative enhancers is necessary Our results build upon established knowledge of chromatin state transitions and gene expression patterns during CIN differentiation (de la Torre-Ubieta et al., 2018; Lewis et al., 2022) and provide a foundation for further exploration of enhancer activity in MGE-derived CIN subtypes, with additional enhancers verified or under verification.

A few patterns emerged that illustrate the need for integrated epigenomic analysis rather than reliance on one biochemical signature (e.g. ATACs-eq or a single histone mark). First, the chromatin states around genomic regions are complex with significant local structure that can expand tens of kilobases. This makes activity scoring for genes much easier than for discrete regulatory elements within these regions. Due to this, our functional predictions for regulatory DNA that is neat the TSS or within certain gene bodies is likely to reflect the gene-level regulation more than regulatory element level. Therefore, we focused on distal regions to refine our analysis and focus on enhancer

regions that are more likely to be functional enhancers. We anticipate that this issue is widespread both for ATAC-seq and histone ChIP-seq and so we suggest treating intergenic enhancer predictions, particularly within the first intron near the TSS with caution. Second, while ATAC-seq is a good indicator of putative enhancer elements in the genome, accessibility alone cannot predict the chromatin state level activity of putative enhancers. This is evidenced by the decoupling of accessibility and histone-indicated activity/repression across all ChromHMM temporal ATAC-seq states. Our data shows the value of combining ATAC-seq with histone marks, which goes against recent trends due to the ease of ATAC-seq and extension of this method to the single cell level. Our results capture how much is lost by this perspective and the histone chromHMM-based activity predictions

However, for a more reliable dataset, further analyses are essential, such as integrating RNA-seq and transcription factor (TF) binding data. While ATAC-seq and histone marks reveal chromatin accessibility and potential enhancer regions, TF binding data would directly link enhancers to gene regulation. By combining the TF ChIP-seq and RNA-seq, we could identify which putative enhancers are functional based on TF recruitment and associated gene expression changes, improving enhancer predictions. This integration would also help reveal regulatory networks, showing how TFs interact with enhancers to control gene expression during interneuron development. Additionally, identifying combinatorial TF binding events could provide deeper insights into complex transcriptional control mechanisms. Overall, combining RNA-seq and TF data would enhance the precision of enhancer identification and better elucidate the molecular

networks driving MGE-derived interneuron differentiation. To accomplish this, some other more novel but less well-established methods might be used, for example, GRaNIE, which links regulatory elements to gene expression by combining ATAC-seq, RNA-seq, and TF data, would allow for precise enhancer-gene interactions, improving enhancer predictions. PALMO could be used to analyze longitudinal multi-omics data, helping to track temporal enhancer dynamics and TF activity during interneuron development. Combining these methods with our existing analysis would enhance the accuracy of identifying functional enhancers and provide deeper insights into the transcriptional networks controlling MGE-derived interneuron differentiation. There is also a number of machine learning methods that might further output enhancer sets with higher reliability.

In summary, this study represents an attempt to model the gene regulatory networks driving human MGE-derived CIN differentiation through chromHMM analysis of ATAC-seq and histone ChIP-seq data. RNA-seq analysis provided additional layers of information, showing genes active at different developmental stages. Our findings highlight the dynamic nature of chromatin states during development and underscore the importance of integrating multiple epigenomic datasets for reliable enhancer prediction. While these results provide a strong foundation, further validation and refinement are needed to fully understand the regulatory mechanisms governing CIN development.

Figures

A)

Day	Week	Proximate mouse embryonic day	Stage
0	0	ESC	ESC
7	1	e8	Telencephalon
14	2	e9-10	Ventral telencephalon/MGE (VZ-like)
28	4	e10	MGE (VZ/SVZ), immature migrating iN
35	5	e12	MGE (SVZ-like), immature migrating iN
42	6	e12-e13	post-mitotic migrating iN

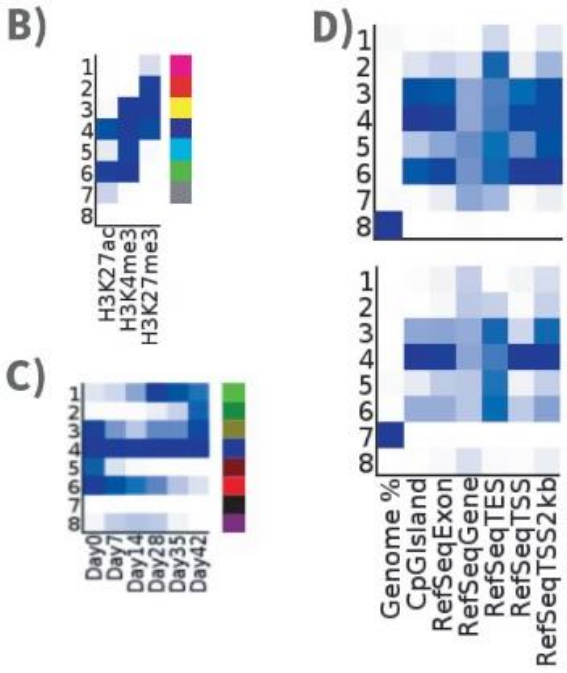


Figure 1: Overview of Cell Developmental Time Points, ChromHMM States, and Genomic Enrichment

(A) Timeline of cell developmental stages used, with corresponding human and mouse fetal development stages and associated cell types.

(B-C) ChromHMM states showing the histone mark or accessibility as x-axis and corresponding enriched states as y-axis, with histone marks (B) and ATAC-seq (C), and color use for each state is consistent in the paper.

(D) Enrichment of each chromHMM state across different genomic components.

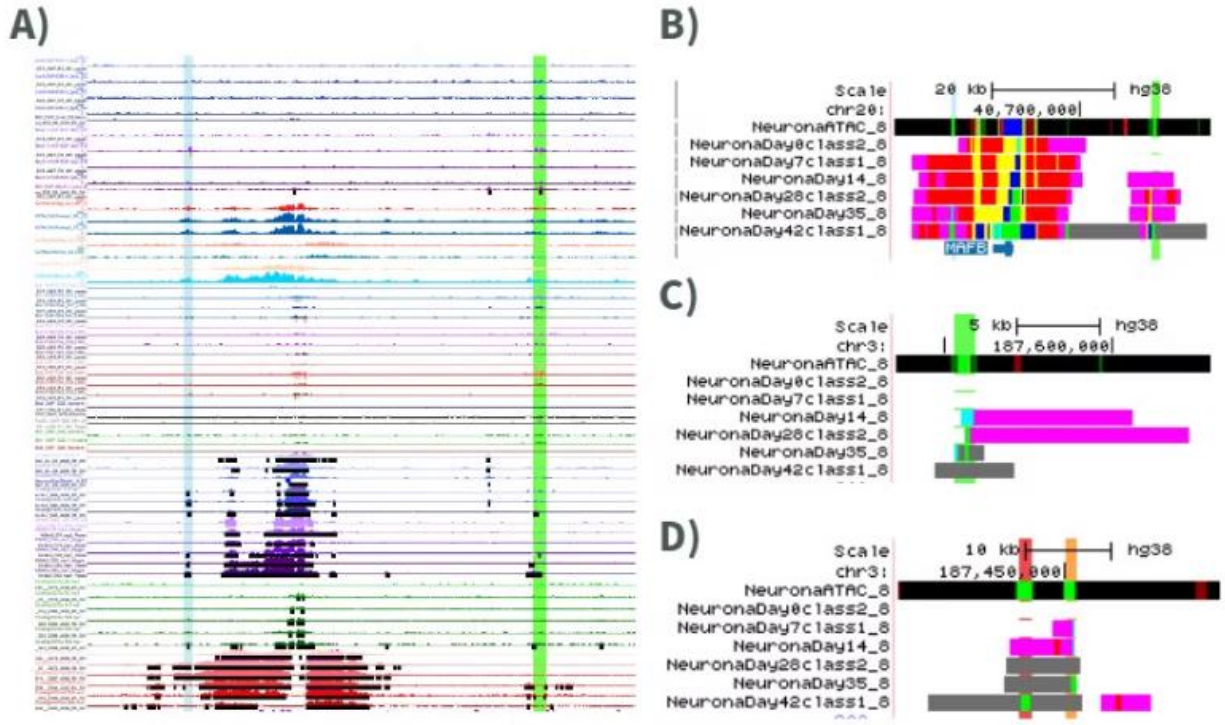


Figure 2: Epigenomic Marks and ChromHMM States Around the MafB Region and Putative Enhancers

(A) Various epigenomic marks illustrating the complexity of the region surrounding MafB.

(B-D) Epigenomic marks simplified using chromHMM states, showing the ATAC signal as the top track followed by histone chromHMM states tracks for each date. (B) same region as depicted in (A), and (C-D) as two example regions with putative enhancers highlighted.

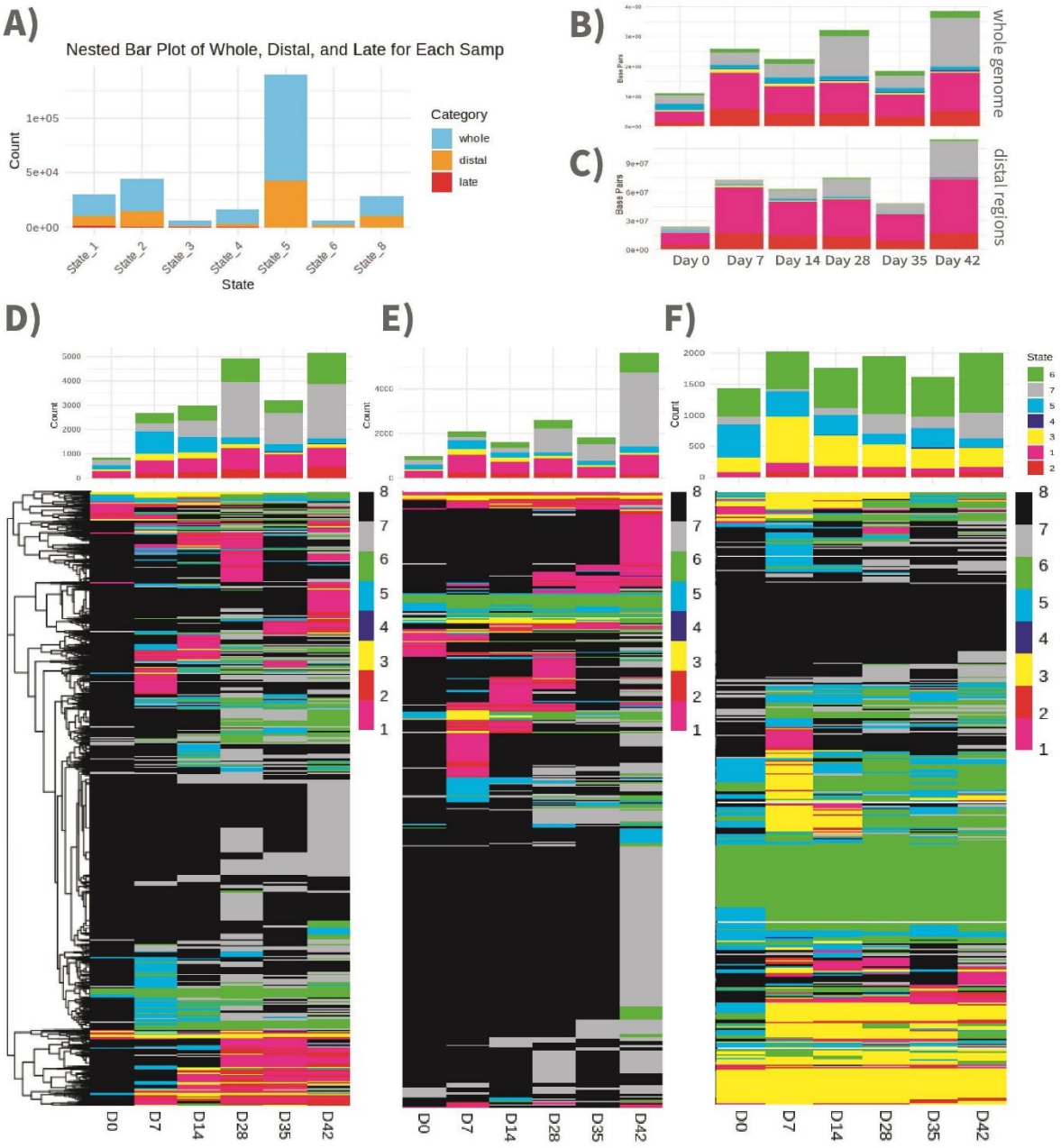

A)
Nested Bar Plot of Whole, Distal, and Late for Each Samp
Count
1e+05
5e+04
0e+00
State_1
State_2
State_3
State_4
State_5
State_6
State_8
State
Category
whole
distal
late
B)
whole genome
C)
distal regions
Day 0
Day 7
Day 14
Day 28
Day 35
Day 42
D)
E)
F)
D0
D7
D14
D28
D35
D42
State
8
7
6
5
4
3
2
1

Figure 3: Temporal Dynamics of Histone ChromHMM States and Their Relationship with ATAC-seq chromHMM Regions

(A) Distribution of non-NULL ATAC chromHMM regions with each bar as one ATAC chromHMM state, excluding the NULL state 7. Orange represents filtered regions outside of 3kb upstream or downstream of protein-coding genes, red being those regions open at Day35/42, while blue shows the remaining non-NULL regions.
(B-C) Stacked bar plots showing histone chromHMM state distributions from Day 0 to Day 42, with colors following the chromHMM scheme in Figure 1B, and y axis being the number of bp in each state. (B) States across the entire genome. (B) States after excluding regions 3kb upstream and downstream of ENCODE-annotated genes.
(D-F) Temporal pattern of histone modification activation under selected regions excluding regions where histone states remain NULL in all time points, with y-axis of bar plots being number of regions. regions selected are (D-E) ATAC chromHMM late open (state 1 and state 2), (F) ATAC chromHMM constitutive open (state 4)

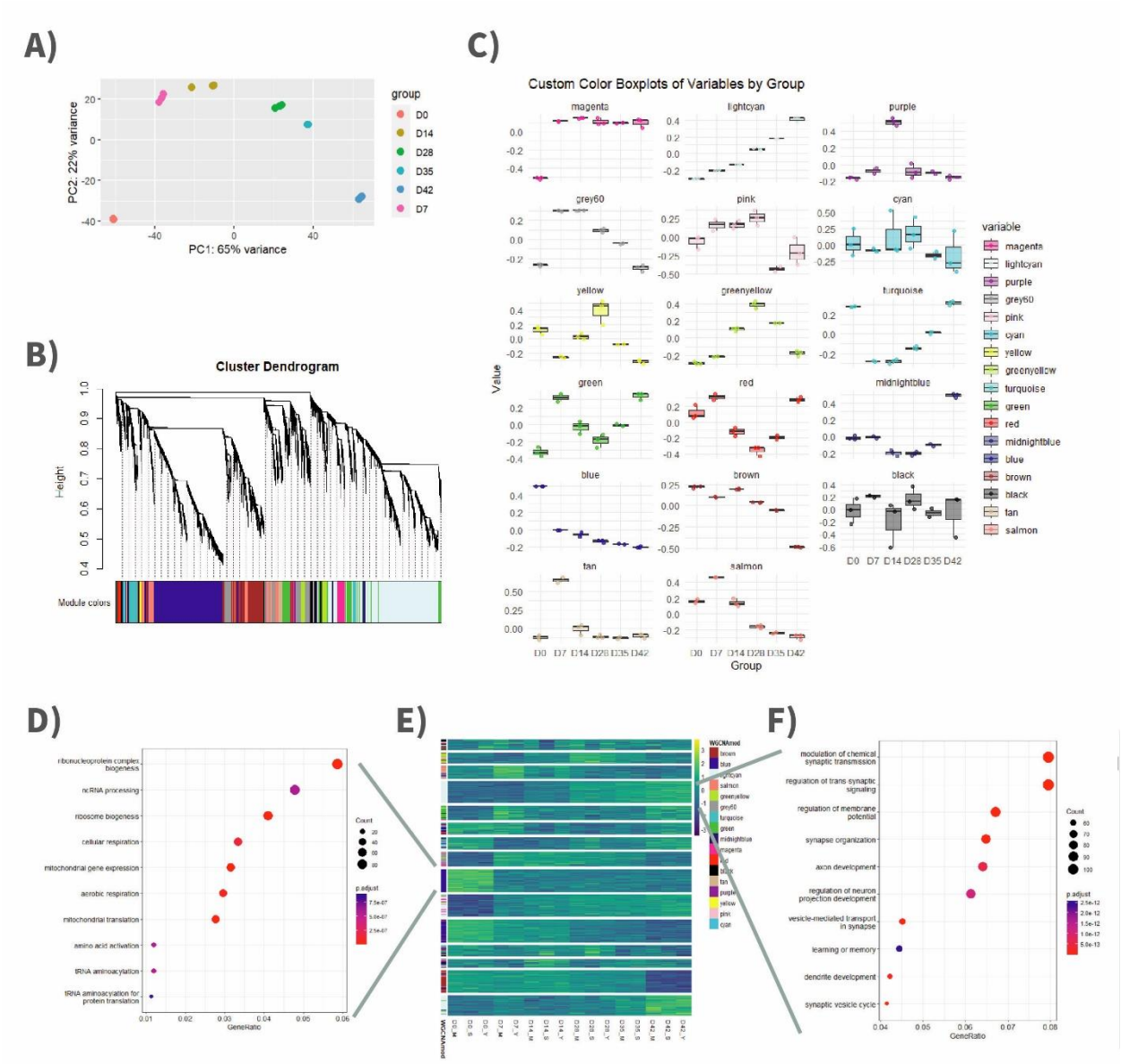

A)
PC2: 22% variance
PC1: 65% variance
group
D0
D14
D28
D35
D42
D7
B)
Cluster Dendrogram
Height
Module colors
C)
Custom Color Boxplots of Variables by Group
magenta
lightcyan
purple
grey60
pink
cyan
yellow
greenyellow
turquoise
green
red
midnightblue
blue
brown
black
tan
salmon
Value
Group
variable
D0 D7 D14 D28 D35 D42
D)
ribonucleoprotein complex biogenesis
ncRNA processing
ribosome biogenesis
cellular respiration
mitochondrial gene expression
aerobic respiration
mitochondrial translation
amino acid activation
tRNA aminoacylation
tRNA aminoacylation for protein translation
GeneRatio
Count
p.adjust
E)
WGCNAmod
F)
modulation of chemical synaptic transmission
regulation of trans synaptic signaling
regulation of membrane potential
synapse organization
axon development
regulation of neuron projection development
vesicle-mediated transport in synapse
learning or memory
dendrite development
synaptic vesicle cycle
GeneRatio
Count
p.adjust

Figure 4: RNA-seq Analysis and Gene Co-expression Network in Neuronal Development

A) Principal Component Analysis (PCA) of RNA-seq samples, with each point representing an RNA-seq sample, and colored by developmental days.

B) Weighted Gene Co-expression Network Analysis (WGCNA) dendrogram, with gene modules represented as color blocks at the bottom.

C) Mean gene expression levels for each WGCNA module, with each column within the box plots to be a timepoint, from Day 0 to Day 42.

E) K-means clustering of RNA-seq gene expression data, with each column to be one RNA-seq sample, and are ordered by time point. The color annotation bar on the left side indicates the WGCNA module that each gene gets assigned to. Two of the blocks showing representative patterns had their GO term enrichment depicted in D) and F).

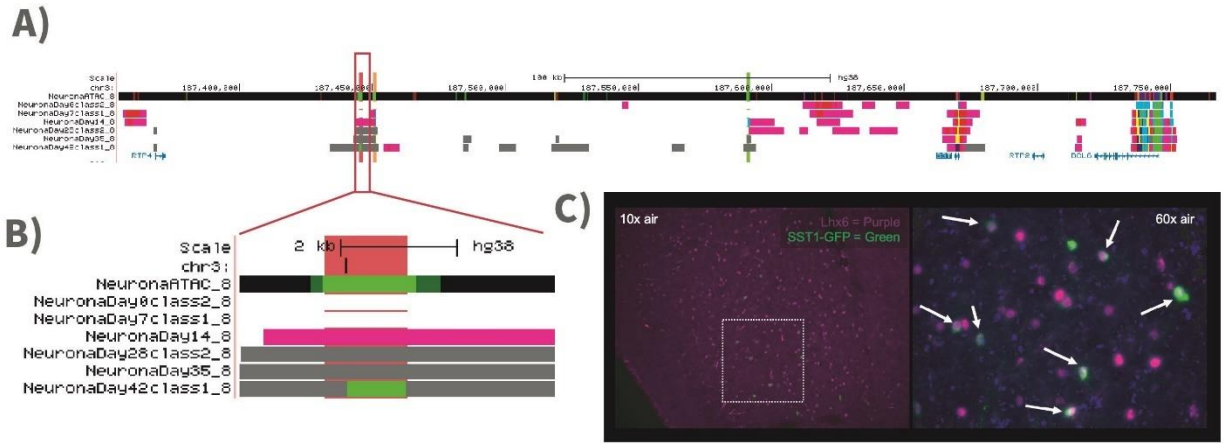


Figure 5: SST#1 Enhancer Activity in Lhx6-Expressing Neurons
A) UCSC Genome Browser view of the SST gene region, highlighting the predicted SST#1 enhancer (highlighted in red). The zoom-in view on the SST#1 enhancer region is shown in B).

C): Enhancer reporter assay results for the SST#1 enhancer in P14 mice, with GFP driven by the SST#1 enhancer in green and Lhx6 expression in purple. The arrows indicate cells co-expressing GFP and Lhx6.

Material and Methods:

Cell Line Generation and Differentiation

MGE-derived GABAergic inhibitory cortical interneurons were differentiated from a human embryonic stem cell (hESC) line provided by Neurona Therapeutics. Cells were cultured and differentiated into MGE-derived cortical interneurons using an established protocol, and got harvested under Day 0, Day 7, Day 14, Day 28, Day 35, and Day 42. Cells harvested at each stage were processed for subsequent epigenomic and transcriptomic analysis.

ChromHMM Analysis

To map the chromatin landscape onto genome, we employed ChromHMM, a hidden Markov model-based tool that segments the genome into distinct chromatin states based on combinatorial patterns of histone modification and chromatin accessibility (Ernst & Kellis, 2017). Two distinct ChromHMM analyses were performed: one integrating three histone modifications (H3K27ac, H3K4me3, and H3K27me3) and the other using ATAC-seq data to assess chromatin accessibility across six developmental time points (D0, D7, D14, D28, D35, and D42).

Raw sequencing reads were aligned to the mm10 mouse genome using BWA (Li & Durbin, 2009), which generates input BAM files for chromHMM modeling. We binarized the aligned data into 200 bp intervals using ChromHMM's BinarizeBam function, and

then trained the ChromHMM model using the LearnModel function and identified eight distinct chromatin states. The binarization and training models was applied separately for the histone modification and ATAC-seq datasets. These Chromatin state annotations were then visualized using UCSC Genome Browse.

Data Integration and Filtering

To refine the enhancer prediction, we integrated the ATAC-seq chromHMM states with the histone-based states. Consecutive regions with non-NULL ATAC-seq states were merged, and if multiple states existed in a region, we assigned the dominant state based on the following priority: 1, 2, 4, 3, 5, 6, 8. We filtered out regions within 3kb of annotated genic regions defined by GENCODE hg38 main annotation file under protein_coding, Mt_rRNA, and Mt_tRNA to focus on intergenic regulatory elements.

Gene Expression and RNA-seq

RNA was extracted from cells at each time point (D0 to D42), and RNA-seq libraries were prepared using stranded polyA-selection protocols. RNA-seq was conducted on the Illumina platform with 2-3 replicates per time point. Differential gene expression was analyzed using DESeq2 (Love et al., 2014), and principal component analysis (PCA) was used to visualize sample clustering. To explore gene co-expression modules, Weighted Gene Co-expression Network Analysis (WGCNA) was performed (Langfelder et al., 2008), and genes were clustered into 17 modules based on expression covariance. Gene Ontology (GO) enrichment analysis was conducted using enrichGO

(Yu et al., 2012) for each module to identify biological processes associated with co-expressed genes.

Cloning, packaging, and Delivery of SST1 Reporter Assay:

We cloned our selected top somatostatin candidate enhancer (SST1) sequence of interest from human DNA via polymerase chain reaction (PCR) and Gibson assembly directly into a DNA plasmid backbone with EGFP fluorescence reporter. Our cloned construct was then sent to the Stanford Gene Vector and Virus Core (GVVC) for packaging into a PHP.eB AAV9 serotype. We then delivered the packaged reporter virus into postnatal day 0 mice via intraventricular injections at titer of 2.00E+09. Mice were than collected at postnatal day 14 days for a total of 14 days of an incubation period.

Immunohistochemistry:

After 14 days incubation, mice were perfused with 4% Paraformaldehyde for sample collection. Brains were then sliced on the Cryostat at 30um. For our IHC protocol, antigen retrieval was performed for a 1 hour at 60°C. After cooling for 20 minutes sections were permeabilized for 20 min at room temperature in PBS containing 0.5 % Triton X-100. slices were then transferred to into a blocking solution that consists of 3% goat serum ,5% Bovine Serum, and were incubated at room temperature for 1 hour. Slices were then transferred to blocking solution containing primary antibodies and incubated for 24 hours days at 4 °C. Sections were washed 5 times in 0.1 % Triton X-100 PBS for 20 min each time. Sections were then transferred to blocking solution

containing secondary antibodies and were incubated for 50 minutes at room temperature. Sections were then washed 3-5 times at room temperature for 20 mins, stained with Hoechst Solution for 20 minutes, and mounted on slides with ProLong Gold Antifade Mountant. Primary antibodies included chicken anti-GFP (ThermoFisher #A10262, 1:1000) and mouse anti-LHX6 (Santa Cruz Biotechnology #sc-271433, 1:200) and secondary antibodies included a Donkey anti-chicken AlexaFluor-488 (Jackson ImmunoResearch #703-545-155, 1:500) and a Donkey anti-mouse AlexaFlour-647 (Jackson ImmunoResearch #715-605-151, 1:500).

Data Availability

All datasets generated during the study, including ATAC-seq, ChIP-seq, RNA-seq, and ChromHMM segmentation data, will be made available via the Nord Lab GitHub repository.

References

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature Reviews Genetics, 17(8), 487-500.

Anderson, S. A., Marín, O., Horn, C., Jennings, K., & Rubenstein, J. L. R. (1997). Distinct cortical migrations from the medial and lateral ganglionic eminences. Development, 124(13), 2633-2644.

Bershteyn, M., Bröer, S., Parekh, M., Maury, Y., Havlicek, S., Kriks, S., ... & Nicholas, C. R. (2023). Human pallial MGE-type GABAergic interneuron cell therapy for chronic focal epilepsy. Cell Stem Cell, 30(10), 1331-1350.

Cobos, I., Long, J. E., Thwin, M. T., & Rubenstein, J. L. R. (2007). Cellular patterns of transcription factor expression in developing cortical interneurons. Cerebral Cortex, 17(Suppl 1), i82-i88.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences, 107(50), 21931-21936.

de la Torre-Ubieta, L., Stein, J. L., Won, H., Opland, C. K., Liang, D., Lu, D., & Geschwind, D. H. (2018). The dynamic landscape of open chromatin during human cortical neurogenesis. Cell, 172(1), 289-304.

Du, T., Xu, Q., Ocbina, P. J., & Anderson, S. A. (2008). NKX2.1 specifies cortical interneuron fate by activating Lhx6. Development, 135(8), 1559-1567.

Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015). The selection and function of cell type–specific enhancers. Nature Reviews Molecular Cell Biology, 16(3), 144-154.

Ernst, J., & Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nature protocols, 12(12), 2478-2492.

Hu, J. S., Vogt, D., Sandberg, M., & Rubenstein, J. L. (2017). Cortical interneuron development: a tale of time and space. Development, 144(21), 3867-3878.

Kilpinen, S., Heliölä, H., & Achim, K. (2023). Range of chromatin accessibility configurations are permissive of GABAergic fate acquisition in developing mouse brain. BMC genomics, 24(1), 725.

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics, 20(4), 207-220.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics, 9, 1-13.

Lewis, E. M. A., Chapman, G., Kaushik, K., Determan, J., Antony, I., Meganathan, K., ... & Kroll, K. L. (2022). Regulation of human cortical interneuron development by the chromatin remodeling protein CHD2. Scientific reports, 12(1), 15636.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760.

Lim, L., Mi, D., Llorca, A., & Marín, O. (2018). Development and functional diversification of cortical interneurons. Neuron, 100(2), 294-313.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15, 1-21.

Pai, E. L. L., Chen, J., Fazel Darbandi, S., Cho, F. S., Chen, J., Lindtner, S., ... & Rubenstein, J. L. (2020). Maf and Mafb control mouse pallial interneuron fate and maturation through neuropsychiatric disease gene regulation. Elife, 9, e54903.
Paul, A., Crow, M., Raudales, R., He, M., Gillis, J., & Huang, Z. J. (2017).

Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. Cell, 171(3), 522-539.

Sussel, L., Marin, O., Kimura, S., & Rubenstein, J. L. R. (1999). Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: Evidence for a transformation of the pallidum into the striatum. Development, 126(15), 3359-3370.

Tsompana, M., & Buck, M. J. (2014). Chromatin accessibility: a window into the genome. Epigenetics & chromatin, 7, 1-16.

Wichterle, H., Turnbull, D. H., Nery, S., Fishell, G., & Alvarez-Buylla, A. (2001). In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. Development, 128(21), 4213-4226.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 16(5), 284-287.

Chapter 4

Characterizing the acute response in fetal cerebral cortex to mid-gestational maternal immune activation via single nucleus RNA Sequencing

a. Preface

The data presented in this chapter constitutes the majority of a manuscript that is in preparation, product of an ongoing collaboration between the Nord lab and the lab of Prof. Kim McAllister at the UC Davis Center for Neuroscience. This study was motivated by earlier research presented in the publication “Sequential perturbations to mouse corticogenesis following in utero maternal immune activation” (Canales et. al, PMID: 33666173). In this work, we sought to define at cell-type resolution the acute changes in mouse brain development observed after 6h of Maternal Immune Activation (MIA) which is induced on embryonic day 12.5 (e12.5) via injection of polyIC.

b. Acknowledgements

I would like to acknowledge the McAllister lab for their expertise with the MIA model and general guidance and performing the polyIC screening and MIA injections. I am grateful to Dr. Canales from the Nord lab for performing the all the dissections. Fellow Nord lab Ph.D. students Stephanie Lozano and Nicolas Seban performed the nuclei prep and worked with the DNA Technologies core to run the 10X platform. I thank Stephanie Lozano, Nicolas Seban, Nickolas Chu, Ethan Fenton, and Karol Cichewicz for support in snRNA-seq data analysis.

c. Author contributions

The McAllister lab performed polyIC testing and MIA prescreening and challenge during pregnancy. Dr. Canales performed embryonic dissections. Nuclei preparation and 10x

Genomics analysis were carried out by lab mates Nicolas Seban and Stephanie Lozano, and the DNA technologies core. I conducted all data analysis and generated all figures, with guidance and coding contributions from my mentor Dr. Nord, postdoctoral fellow Dr. Karol Cichewicz and Ph.D. students Stephanie Lozano, Nicolas Seban, Nickolas Chu, Ethan Fenton (all from the Nord Lab, who, with myself also, form our working group on snRNA-seq). I drafted initial versions of figures and text, with input and editing support from Dr. Nord and Dr. Canales.

d. **Declaration of interests**

Nothing to declare

Abstract:

Epidemiological studies in humans and modelling studies in rodents and non-human primates have shown a link between maternal infection or immune challenge during pregnancy and various neurodevelopmental and neuropsychiatric disorder outcomes in offspring. Multiple studies have tracked the effects of MIA on embryonic and postnatal brain, but the immediate acute changes in fetal brain have not been systematically characterized. This study investigates the acute effects (6 hours after treatment) of maternal immune activation (MIA) on the developing E12.5 mouse forebrain using single-cell RNA sequencing (snRNA-seq). We identified significant changes in the proportional make up and transcriptional profiles in key cell populations, including microglia, VLMC (vascular leptomeningeal cells), and pericytes, as well as more broad impacts on neurogenesis. Notably, genes involved in immune activation (e.g. *Mrc1* and *Adgre1*) and extracellular matrix remodeling (e.g. *Col3a1* and *Col5a2*) were upregulated following to maternal PolyIC treatment. Conversely, genes related to neuronal maturation and function (e.g. *Grm5* and *Grin2b)* were downregulated, indicating potential disruptions in maturation and signaling of glutamatergic and GABAergic neurons. Our results highlight microglia and neuro-vasculature associated cell types as the likely dominant initial responders following MIA, with changes in neurogenesis also emerging. In addition to providing new insights into the earliest responses across cell types in the fetal forebrain, our results expand understanding beyond previous focus on microglia and immune response, highlighting significant initial impact on the populations of cells and molecular pathways associated with vascularization and extracellular matrix formation in the embryonic forebrain.

Introduction:

Maternal immune activation (MIA) refers to the physiological response triggered by an immune challenge, such as a viral or bacterial infection, during pregnancy. This immune response in the pregnant mother can have effects on fetal development, particularly in the brain, where the risk of neurodevelopmental disorders such as autism spectrum disorder (ASD) and schizophrenia have been reported to increase (Yamashita et al., 2003; Hornig et al., 2018; Hameete et al., 2020). When a pregnant female experiences an infection, the immune system responds by releasing cytokines which are believed to cross the placenta and drive fetal neuroimmune signaling pathways. For examples, interleukin-6 (IL-6) plays a critical role in cell signaling during immune response and has been linked to changes in fetal brain structure and function (Smith et al., 2007; Estes et al., 2016). Studies with viral or bacterial mimic challenge instead of true infection have demonstrated that the maternal immune response is sufficient to produce negative outcomes in animals (Canetta et al., 2014; Machado et al., 2015).

Animal studies, particularly in rodents, have demonstrated that inducing an immune response in pregnant animals can lead to offspring with behavioral and neurological abnormalities that resemble human neurodevelopmental disorders through polyinosinic:polycytidylic acid PolyIC injection (Meyer et al., 2012; Hameete et al., 2020). When administered at embryonic day 12.5 (E12.5) at mid-gestation, a critical period of rapid neurogenesis and differentiation, polyIC induces a robust immune

response that leads to significant alterations in the developing brain (Smith et al., 2007). These alterations can include changes in cytokine levels, microglial activation, and disruptions in neurotransmitter systems, all of which are implicated in the pathophysiology of neurodevelopmental disorders (Smith et al., 2007; Ostrem et al., 2024; Bergdolt et al., 2019).

Studies using both bulk RNA sequencing (RNA-seq) and single-cell RNA sequencing (scRNA-seq) have been conducted to explore the early effects of maternal immune activation (MIA) on brain development. In an earlier study, we used bulk RNA-seq to identify significant transcriptional changes at various developmental stages, including 6 hours after MIA treatment at e12.5 (Canales et al., 2021). Our findings, validated at neuroanatomy level at E17.5, indicated that MIA disrupts neurodevelopmental timing by altering cortical development. Gene Ontology (GO) analysis highlighted pathways involved in immune signaling, extracellular matrix organization, cell adhesion, and phagocytosis, emphasizing the extensive impact of MIA on these critical biological processes. However, as we used bulk methods, mapping impacts to specific neuronal and non-neuronal cell populations was not possible, leaving major questions regarding which cell populations exhibited which changes. Thus, building on this study and utilizing the same acute experimental time point and design, here we aimed to use snRNA-seq to uncover changes in cellular make-up and cell-type specific transcriptional signatures.

Results:

Identification of cell types and developmental trajectories in E12.5 forebrain

Initial validation of immunogenic response to PolyIC was performed in female mice before inclusion in this study. Pre-screened pregnant dams were administered either PolyIC or saline at E12.5 and embryos were collected 6 hours later. Microdissected forebrain (pallial and subpallium) was collected for individual males and female embryos from two litters per treatment. We used two pools consisting of equal numbers of male and female samples from each litter for each treatment, performing four 10X 3' snRNA-seq assays representing two replicates per treatment for both male and female samples and sequenced the resulting libraries.

Sequence data were analyzed via application of standard computational methods for alignment, quality control, dimensionality reduction, and cell clustering using the Seurat package (Stuart et al., 2019). Male and female cells that were pooled in the same sequencing lane were separated via analysis ox sex-linked gene expression using the sxreveal R package (Chu, 2024). 31764 total cells passed QC, including 16325 from PolyIC and 15439 from saline treated dams. Using k-nearest neighbors and the Louvain algorithm, 21 distinct cell clusters were identified. Distinct cell types were annotated using the unbiased FindAllMarkers function from the Seurat package. A Wilcoxon rank-sum test was applied to compare each cluster against all others, generating heatmaps for visualization. Figure 1A illustrates the UMAP plot of the whole forebrain snRNA-seq

data, showcasing the identification of all major cell types. Non-neuronal clusters consist of distinct cell types such as microglial cells (cluster 19), pericytes (cluster 16), endothelial cells (cluster 8), and VLMC (cluster 1 and cluster 18). Progenitor and early postmitotic pallial and subpallial neuronal cells were more interconnected, with cluster 2, 4, 10 being progenitor cell types, and the trajectory towards cluster 5 represents the development of glutamatergic neurons and the trajectory towards cluster 7 and 11 representing the development of GABAergic neurons (Figure S1). This developmental trajectory for the two neuronal cell types is further elucidated via analysis by scVelo (Bergen et al., 2020), which models RNA velocity at the single-cell level (Figure 1B). In this model, the upward trajectory represents the progression of glutamatergic neurons, and the downward-left trajectory indicates the development of GABAergic neurons, with two distinct paths of maturation that appear clearly delineated.

As shown in Figure 1C, all clusters exhibited strong expression of their respective positive gene markers (highlighted in yellow rectangles). Some clusters, such as clusters 1 and 18, shared a subset of positive gene markers, specifically collagen-related genes associated with VLMC. To improve the accuracy of cell type identification, we further examined known marker genes that were manually curated. While unbiased marker genes are often sufficient to distinguish certain clusters, such as microglia, in cases where cell types are closely related, the known marker genes provide more informative distinctions. In order to visualize gene expression patterns across different cell clusters, a dotplot was used with the size of each dot representing the percentage of cells within a cluster that express a particular gene, and the color indicating the

gene's average expression level. This visualization helps to distinguish clusters that may share similar marker genes. For instance, while clusters 1 and 18 share several markers, cluster 18 uniquely expresses Mki67, a gene associated with active cell cycling, helping to differentiate it from cluster 1. In summary, all major expected cell populations were identified and sufficiently represented to enable overall and cell-type specific comparison of transcriptional differences across treatments.

Pseudobulk and principal component analyses identify treatment specific transcriptional differences between PolyIC and saline.

To overcome the heterogeneity and sampling variance of single cell level data, we used pseudobulk methods (Zimmerman et al., 2022), pooling all reads from all cells that come from each sample. This approach facilitates the use of various statistical techniques for RNA-seq and differential gene expression analysis that otherwise have poor performance when applied at the single cell level. Thus, we could generate a total of eight pseudobulk samples, corresponding to two male and female samples from each treatment. With our design, each male/female replicate represents pools of different individual male/female embryonic forebrains pulled from two separate litters, which should reduce contributions arising from differences between litters or dissection/sample preparation. These samples were then analyzed using a standard bulk RNA-seq methods to identify the transcription profile changes under MIA treatment.

We first used the pseudobulk approach across all cell clusters, yielding expression data that would be akin bulk RNA-seq preparation for unsorted nuclei, in order to test if generalized transcriptomic signatures distinguished polyIC treated samples. To compare the similarity and relationships between samples based on their expression profile, the RNA counts were summed and then transformed using a regularized-log transformation to stabilize variance (Love et al., 2014), and via Principal Component Analysis (PCA) was subsequently performed. As illustrated in fig 2A-2B, the first principal component (PC1) separated the samples based on sequencing batch effects, while the second principal component (PC2) distinguished MIA treatment, separating PolyIC-treated samples and the saline controls, hence PC2 captures the biological impact of the PolyIC treatment on gene expression. To investigate which genes are informative for separating treatment conditions, we examined gene loadings for PC2 (Figures 2C-2D). Notable genes driving the positive loading on PC2 include *Mrc1* and *Adgre1*, which are key markers of microglia, and *Col3a1* and *Col5a2*, indicative of collagen-expressing cells. Conversely, the negative loadings for PC2 were characterized by genes such as *Neurod2* and *Grin1*, which are associated with neuronal function and differentiation.

To characterize PC2 expression, we employed the AddModuleScore_UCell (Andreatta et al., 2021) function on the top 100 loading genes on PC2 for both directions, and used feature plots to display the spatial expression of these genes after excluding *Tsix* and *Xist* to avoid the effects from these gender-related genes (Figures 2E-F). The analysis revealed that genes contributing positively to PC2 were highly and largely specifically

expressed in microglia and various neurovasculature and collagen-expressing cell clusters. Conversely, genes with negative loadings on PC2 were predominantly expressed in the left half of the neurogenesis large group on the UMAP, which corresponds mainly to relatively mature neurons (both pallial and subpallial origin). When examining the mean expression levels of the top loading genes instead of using the AddModuleScore_UCell function, the epithelial cluster and pericytes also showed high expression in the feature plots, indicating a subset of genes with high expression in those cell types (Figure S2).

To test for functional enrichment of genes with strong contributions to each PC, we used enrichGO function from the R package clusterProfiler. For this gene ontology (GO) enrichment analysis the top 100 genes from each of the eight groups (four principal component analyses, both positive and negative loadings) were included. Given a vector of genes, this function will return the enrichment GO categories after FDR control. GO terms from the output were selected after filtering for a q-value and adjusted P value less than 0.05 ($P<0.05$), and their significance was color-coded accordingly. Notably, no relevant GO enrichment was observed for genes associated with positive and negative loadings of PC1, which is associated with sequencing batch. In contrast, genes in the PC2 positive group were significantly enriched for GO terms related to extracellular matrix organization and endothelial/epithelial development and vasculature and immune response, and those in the PC2 negative group were associated with formation and maturing of neurons such as synapse organization and neurotransmitter transport, corroborating the observations in PCA shown in Figure 2B-F. Genes

associated with negative loadings for PC3 were primarily related to cell cycle, while positive loading genes were similar to positive loadings in PC2. The classification of genes in PC4 did not follow a distinct pattern, indicating a less definitive functional grouping in this component. Overall, this analysis identified systematic changes in gene expression that distinguish the polyIC treatment group which could arise from changes in overall cell proportions or in cell-type specific transcriptional changes.

MIA causes increased proportion of immune and vasculature-associated cells in fetal forebrain 6 hours after treatment

According to previous finding from our lab (Canales et al. 2021) and others (Laighneach et al., 2021), maternal immune activation (MIA) treatment results in alterations in gene expression linked to both cell proliferation and differentiation and other changes reflected in our PCA findings. However, as mentioned before, effects in less well represented cell types such as microglia and vasculature-associated cells were likely masked and distinguishing whether changes were due to expression effects or changes in cell composition was not possible in bulk studies. Moving to cell-type specific analysis, we first tested for changes in the proportion of the cell cluster representation that were reproducible in the polyIC treatment samples.

For this analysis, cell type proportions was identified for each sample and compared via Wilcoxon rank-sum test. MIA treatment was associated with a significant increase (p value < 0.05) in cell proportion on VLMC (clusters 1 and 18), microglia (cluster 19), and

pericytes (cluster 16), with replicate level proportions shown in Figure 3. These findings align with the high expression of PC2-associated genes within these populations, and indicate that at least some of the overall transcriptional differences in our pseudobulk data are driven by changes in the representation of these cell classes. Marginal significance (p value = 0.057) was observed in glutamatergic neurons (cluster 5) and developing neurons (cluster 9), both of which displayed a decrease in abundance. As part of the cluster 9 neuronal cells are gathered at the Glutamatergic neuron-located peninsula, we further checked the proportion of the peninsula containing cluster 5, 3, and the portion of cluster 9 where UMAP_2 is higher than 4. The resulting Glutamatergic peninsula did not show statistical cell proportion change under MIA condition. Cells in cluster 14, identified as Reln-expressing neurons, with marginal significance (p value = 0.057) showed an increase on cell composition proportion under MIA conditions, suggesting selective, albeit subtle effects on specific neuronal subtypes (Figure 3). These findings suggest that MIA treatment induces recruitment or population expansion of collagen-expressing cells, microglia, and pericytes, manifested by an increase in these populations, while potentially decreasing the proportion of relatively mature neurons, possibly by delaying their maturation process. An exception to this trend appears to be the *Reln*-expressing Layer 1 Cajal-Retzius cells, which exhibit higher proportions under MIA conditions.

Cluster and gene level differential signatures associated with MIA

We next applied differential expression analysis to specific cell populations to identify the transcriptional response to MIA at cellular level resolution. We applied the

pseudobulk approach both to all cells and also at cluster level, combining all reads from cells within a specific cluster and then performing differential expression analysis with DESeq2. We identified differentially expressed genes following MIA conditions across all cell clusters, but focus our presentation on cell clusters that had significant expression of either the positive or negative gene loadings associated with PC2. Differentially expressed genes were further analyzed for Gene Ontology (GO) terms. Pericytes, Cajal-Retzius cells, and neuronal cells grouped in cluster 9 did not show any significant Gene Ontology (GO) terms, and as a result, their differential expression results are not presented here. Figure 4 includes volcano plots representing the DE results for all-cells, as well as microglia, VLMC, and postmitotic glutamatergic cells (glut). Figure 5 lists the top 20 DEGs identified from relevant comparisons filtered for consistency in logFC across sequencing batches and for genes with no overlap in expression distribution between PolyIC and saline samples, with single gene feature UMAP plots comparing polyIC and saline included for a selected set of representative DEGs.

As discussed earlier, changes in cell proportions suggest that a significant factor in response to MIA is the alteration in cell numbers. This is supported by the observation that the highest number of differentially expressed (DE) genes was found in the full pseudobulk, when including all cells from all clusters. In the PolyIC-treated condition, 1,413 genes were downregulated and 1,535 genes were upregulated at adjusted P-value < 0.05. Many of these upregulated genes were associated with immune activation and extracellular matrix remodeling, consistent with the patterns observed in the PC2

loading genes. Among the upregulated genes were *Csf1r* and *Adgre1* (log2FC 1.2, 1.89, 1.61), which are key markers of activated microglia, indicating an increased microglial activation and potential involvement in neuroinflammation. Interestingly, although these genes are highly differentially expressed in MIA-treated samples when considering all cells passing QC, their adjusted p-values (0.79 and 0.45, respectively) did not reach significance when differential expression was analyzed specifically within the microglia subset. This discrepancy is likely due to the significant DE detected for these specific genes in the all-cells analysis being driven by the increased number of microglia under MIA conditions, rather than changes in the expression levels within the microglia population itself. Similarly, genes related to the extracellular matrix, such as *Col4a5* and *Col5a2* (log2FC 0.88 and 0.99), were also upregulated when taking all cells that pass the QC into account, but are not when only checking for DE genes within the VLMC population, suggesting enhanced collagen synthesis and extracellular matrix remodeling is at least in part driven by increased number of cells in the VLMC population.

Conversely, several genes were significantly downregulated in the PolyIC-treated condition, particularly those associated with neuronal function. *Grm5* and *Grin2b* were among the downregulated genes (0.52, 0.37), pointing to potential disruptions in synaptic signaling and plasticity. Additionally, the downregulation of *Snap25* (logFC 0.38), a gene essential for synaptic vesicle fusion and neurotransmitter release, suggests a broader impact on synaptic transmission and neuronal communication. Other notable downregulated genes include *Camkv* and *Snap91* (logFC 0.68, 0.38),

which are involved in neuronal development and synaptic function, indicating a potential delay or disruption in neuronal maturation and connectivity (Liang et al., 2016; Sun et al., 2021). Of note, we identified down regulation of genes with high neuronal gene expression in non-neuronal cell clusters. Considering we used a nuclei-based method, we believe these effects are most likely due to ambient RNA contamination from lysed neuronal cells based on estimations of contributions of these genes to ambient RNA using SoupX.

Moving to specific cell populations, we highlight transcriptional response within cell types. These analyses are normalized for total read counts, and thus DE effects will not be driven by changes in the number of cells, but rather capture changes in gene expression. There are a substantial number of genes that are DE in both microglia and all-cell DE results, indicating that microglia likely undergo both proportional increase as well as exhibiting relatively robust transcriptional response. For example, the top two microglia upregulated genes *Lyve1* and *Dab2* in both overall DE and microglia cell specific DE (logFC 1.10, 0.76). Of note, the top two all-cell upregulated genes, *Mrc1* and *Runx1* are both also significantly upregulated in microglia (logFC 0.69, 0.64), with many other genes of interest also identified suggesting increased microglial activation following MIA. Comparing VLMC cells, genes related to cell adhesion such as *Flrt2* and *Vcan* show increased gene expression (log2FC 0.41, 0.49). A relatively larger set of genes were strongly DE within VLMCs versus other non-neuronal cell types, and as such the VLMC DEGs reached significance for GO enrichment associated with extracellular matrix (Figure 4).

As for neuronal development, there were few genes strongly differentially expressed. Interestingly, there doesn't seem to be a consistent pattern based on these DE gene's functions. For example, *Cdh4*, which is higher expressed in less mature Glutamatergic neurons, get down-regulated. On the other hand, *Cdh11*, which is higher expressed in Glutamatergic neuronal cluster, is goes higher expressed under MIA condition. On the other hand, genes related to cell junction assembly such as *Ptprt*, *Cdh4*, *Nrp1* are downregulated under MIA condition (logFC 0.79, 0.48, 0.72), yet genes of the same function such as *Cdh11*, *Prkca*, *Cadm1* (log2FC 0.53, 0.33, 0.36) are upregulated under MIA condition.

Collectively, these findings suggest that MIA impacts brain development not only by modulating microglial numbers and neuroimmune response, but also by significantly altering the developing brain's structural framework, with potential implications for both neurodevelopment, neuronal maturation and brain function.

Discussion:

Maternal immune response during pregnancy is associated with acute and long-lasting effects on the developing offspring, increasing the risk of neurodevelopmental disorders like autism spectrum disorder (ASD) and schizophrenia in the offspring (Han et al., 2021). In our previous study using bulk RNA-sequencing, we found major shifts in neurogenesis across cortical development and evidence of initial changes to immune, stress, and extracellular matrix gene expression (Canales et al., 2021). However, we did not have the ability to determine which cells were impacted and expression of cell-type specific marker genes for populations such as microglia and neuro-vasculature indicated that effects in these cells were unlikely to be captured. Unlike bulk RNA sequencing, which averages gene expression across all cells in a sample, single-cell RNA sequencing allows for the resolution of gene expression at the level of individual cells. This granularity is especially important in the context of MIA, where different cell types in the brain may respond uniquely to the immune challenge. By applying differential expression analysis at the single-cell level, we identified specific cell populations that are most affected immediately at an acute stage following MIA, uncovering distinct gene expression changes within neurons and non-neuronal cells. The strongest changes featured both changes in cell proportions, highlighted by increased numbers of immune and vasculature-associated populations, as well as robust gene expression changes within cell types. Our results yield an understanding of the acute molecular mechanisms following MIA that is both more granular, with cell-type specific resolution, and more comprehensive as it encompasses all major cell

populations. Our findings reveal potential new targets for therapeutic intervention in neurodevelopmental disorders linked to immune challenges during pregnancy.

While most studies have focused on the effects of MIA at later stages, our analysis at 6 hours post-polyIC injection at e12.5 has uncovered several potential immediate effects of the MIA response in embryonic tissue. This early time point allows us to capture the initial molecular and cellular changes that occur shortly after immune activation, providing valuable insights into the early mechanisms through which MIA may influence brain development and function. This analysis was organized to systematically address the research objectives, beginning with the identification and characterization of cell types using single-nuclei RNA sequencing (snRNA-seq) data, followed by an analysis of differential gene expression between MIA and control conditions, highlighting key genes and pathways affected by MIA. Finally, I conducted statistical tests to assess significant changes in cell type proportions, offering a comprehensive view of how MIA impacts cellular composition and development at this critical stage. These results collectively enhance our understanding of the molecular and cellular mechanisms underlying the impact of MIA on neurodevelopment.

Comparison to previous transcriptomic studies of MIA

While there are numerous publications that use RNA-seq to test for the impacts of MIA in mouse brain (Tsivion-Visbord et al., 2020; Laighneach et al., 2021), our results are particularly relevant to three recently published studies. The first is our own study that

leveraged a temporal time course bulk RNA-seq approach to characterize impacts using the same exact MIA design, featuring comparisons at E13.5, E14.5, E17.5, and P0 (Canales et al., 2021). The other two studies by different groups also used mid-gestational PolyIC for immune challenge in mice, but in contrast used single cell RNA-seq methods and combined transcriptomics with knockout mouse work to show necessity and sufficiency of *Il17* (Kalish et al., 2021) and *Csfr1* (Ostrem et al., 2024) signaling pathways. The latter study was published while we were preparing our results, and focuses on the microglial axis of MIA response.

Compared to our initial bulk RNA-seq study, we identified the same overall induced signature of immune/stress and extracellular matrix pathways and reduction of pathways associated with more mature neurons. While similar at the pathway/GO level, somewhat surprisingly, overlap between specific DEGs from the bulk cortex with our all-cell pseudobulk is relatively weak, albeit with some strong shared hits (e.g., *Runx1*, *Il20rb*, and *Flt1* induction). There are a number of explanations for differences, from technical (bulk tissue/whole cell versus prepped nuclei, more specific cortical dissections versus forebrain in the current study) to biological (variation in environment or strength of MIA immune response induction). Considering the differences in methods and the overall similarity at the GO/pathway level, we believe that these studies replicate the overall acute impact of MIA on fetal forebrain. The current study captures effects in non-neuronal populations that were not present or captured well in the bulk, most critically the strong microglial response. Analysis using snRNA-seq at later embryonic timepoints would be needed to evaluate whether the much larger scale

effects on neurogenesis and cortical patterning we identified previously later in gestation (at E14.5 and particularly E17.5) reproduce.

Compared to the two studies that used single cell transcriptomics (Kalish et al., 2021 and Ostrem et al., 2024) and the same mid-gestational E12.5 polyIC exposure model, our findings are distinct from both these studies and greatly expand understanding of the acute impacts immediately following MIA (as the other studies focused on later timepoints, with the earliest data from either E13.5 or E14.5) and on cell lineages beyond neurons and microglia. Kalish et al. (2021) profiled E14.5 and E18.5 single cells following E12.5 MIA exposure, with the primary findings that there was a male-specific IL17a-dependent induction of the integrated stress response that results in decreased global translation. Our results show very little overlap with the findings from this study. This is most likely due to changes that occur form the acute timepoint to their E14.5 and E18.5 timepoints, particularly given their focus on neuronal lineages at these timepoints.

The paper that was published during preparation of results for our study (Ostrem et al., 2024) used microglial enrichment to profile the MIA induced changes in microglial gene expression and testing for necessity of microglia in the MIA response via study of *Csf1r* null mice, which do not have microglia in the developing cortex. First, they showed strong impacts of MIA across E15.5 cell populations, with findings that generally aligned with our bulk study and the previously described single cell work, and showed that these transcriptional signatures are abrogated in the *Csf1r* null mice. Aligned with their

findings, we saw *Csf1r*, *Tenm4*, and *Hes1* DE in our data under the analysis on all cell types, with *Csf1r* and *Hes1* upregulated (1.21, 0.58) and *Tenm4* downregulated (0.23). *Tenm4* is also reported to be downregulated in the VLMC cluster. Ostrem et al., profiled unstimulated microglia across development to determine subtypes, identifying one subtype that expressed *Ms4a7* dominates at early timepoints, including E12.5. We found increased expression of *Ms4a7* in our all-cell pseudobulk, but not in our microglia-specific analysis. We also found increased *Mki67* and *Top2a* in MIA-exposed microglia, markers of proliferation. This is consistent with increased colonization and proliferation of microglia 6hrs following MIA. We also identified one other subtype marker they identified as increased (*Hmox1*). Somewhat surprisingly, microglia-specific DEGs identified here at E12.5 and in Ostrem et al. at E13.5 did not overlap. There are two major reasons why could be the case. First, there could be a change over time, even in the ~18-24hr time difference between our data and theirs. This is supported by their reported weak overlap between E13.5 and E15.5 microglial DEGs following MIA. Second, we used a nuclei-based protocol while they used full intact sorted microglial cells. Regardless, of overlap among DEGs in microglia, our results are consistent with increased populations of the *Ms4a7*-expressing microglial subtype and our findings further implicate *Dab2*-associated activation of these cells following MIA.

An integrated cellular perspective of the acute MIA response in fetal forebrain

The differential expression signatures observed in VLMCs, vascular endothelial cells, and other non-neuronal cell types in response to maternal immune activation (MIA) likely reflect acute recruitment and changes in cell proportions within these populations.

During embryonic development, extensive intercellular communication is critical for coordinating the maturation and differentiation of various cell types, including neural progenitors, microglia, and vascular cells. The presence and functionality of these non-neuronal cells, such as microglia and vascular cells, are essential for proper brain development, as they actively communicate with neural cells to guide differentiation, migration, and overall neural architecture (Haigh et al., 2003; Crouch et al., 2015; Hellstrom et al., 1999).

The early changes we observed in the expression patterns of VLMCs and other vascular cells suggest that MIA may trigger a rapid recruitment or activation of these cells to respond to the inflammatory environment. This recruitment is consistent with the known roles of vascular and immune cells in modulating neural development, such as endothelial cells influencing neural stem cell behavior and microglia shaping synaptic structures (Nikolakopoulou et al., 2013; Paolicelli et al., 2011). Given the critical role of vascular cells in forming the blood-brain barrier and supporting neural progenitors, their altered proportions could reflect an adaptive response to maintain neural homeostasis under MIA conditions. Additionally, the upregulation of immune and extracellular matrix remodeling genes within these cell types suggests a coordinated response to inflammation, potentially affecting neurodevelopmental trajectories. These findings highlight the importance of understanding how acute changes in non-neuronal cell populations contribute to the broader impact of MIA on brain development and underscore the need to map out the specific molecular mediators of these intercellular interactions more systematically.

Notably, the cell proportions of pericytes, microglia, and VLMCs increased under MIA condition, accompanied by significant upregulations of genes related to immune activation including *Lyve1* and *Dab2* in microglial population, which are both important markers for different types of microglia (Zelco et al., 2021), and VLMC also includes genes related to adhesion such as *Ednra*, *Flrt2*, and *Vcan* upregulated under MIA condition. This result enhances previous result of increased microglia population under MIA (Tansley et al., 2022), as well as the effect of IL-6 function as growth factor on mesenchymal cells, which are progenitor cells for VLMC.

Finally, when examining the impact of MIA on neuronal development, we observe that only a few genes are differentially expressed. Interestingly, there does not appear to be a consistent pattern in terms of the functions of these differentially expressed (DE) genes. This lack of a clear pattern suggests that the response to MIA within neuronal populations may be more heterogeneous or context-dependent than initially expected. The DE genes do not seem to cluster around specific developmental pathways or processes, indicating that MIA might affect various neuronal functions and stages of maturation in a more complex manner. This variability could be due to differences in neuronal subtypes, regional brain-specific responses, or varying susceptibilities of neuronal cells to immune challenges. Further analysis of the individual DE genes could provide deeper insights into how these scattered changes might contribute to broader impacts on neuronal development and function. Given that other studies described above identified significant perturbations in neurogenesis at later timepoints, future work

would benefit by modeling longer term developmental dynamics of neurogenesis and neuronal maturation, thereby clarifying the temporal progression of gene expression changes and their effects on the development of specific neuronal lineages.

Limitations of the study

The short 6-hour time frame post-injection may not be sufficient to fully capture the extent of acute MIA changes effects or may even be too long after challenge and miss initial changes, potentially limiting our ability to observe more profound or widespread changes in gene expression and cellular behavior. This brief window may only reflect the initial responses to maternal immune activation (MIA), and more subtle or secondary changes could emerge later, beyond this early phase. Consequently, the current findings should be interpreted with caveats, as they represent an early snapshot of the brain's response to MIA. Extending the observation period in future studies (as we did previously with bulk RNA-seq) could provide a more comprehensive understanding of how initial gene expression changes evolve over time and contribute to long-term alterations in brain development and function. Furthermore, our reliance on nuclei-based methods may miss critical changes that are exaggerated or only present in mature mRNA that has been exported from the nucleus. This is offset by the potential for biased cell dropout when using intact cell preparations. Finally, we were limited to two replicates per sex and used only two litters per treatment, and increasing the sample size and number of independent litters would be likely to increase the power to identify more subtle or heterogeneous changes.

Conclusion

This study represents an initial detailed exploration of the acute effects of MIA on fetal forebrain at a single-cell level using an MIA mouse model. By examining the embryonic brain just 6 hours after PolyIC injection at e12.5, we provided a snapshot of the initial molecular and cellular changes that occur in response to immune activation. While our findings highlight some immediate effects, such as changes in cell proportions and differential gene expression among non-neuronal cell types like microglia and VLMCs, they also underscore the complexity of MIA's impact on the developing brain. The presence of neuronal gene signatures in non-neuronal clusters and the lack of consistent patterns among differentially expressed genes related to neuronal development suggest a nuanced response to MIA, possibly reflecting both immediate and adaptive changes to the inflammatory environment.

This work lays the groundwork for future studies aimed at capturing the dynamic and evolving effects of MIA over time. Extending our observation period beyond this early 6-hour window will be crucial for uncovering the full spectrum of changes in gene expression, cell behavior, and cellular interactions, thereby providing a more comprehensive understanding of how MIA influences brain development. Additionally, further investigation into the specific molecular mechanisms driving these initial responses will help clarify whether the changes we observed are transient or indicative of long-term shifts in developmental trajectories, the latter more possible as per

previous longitudinal time course developmental findings in previous bulk RNA-sequencing experiments. Despite limitations of this work, our results identified novel pathways and sensitive cellular populations through which MIA initiates downstream impacts that increase the risk of neurodevelopmental and neuropsychiatric disorders, and provide insights that may inform strategies for early intervention and prevention.

Figures

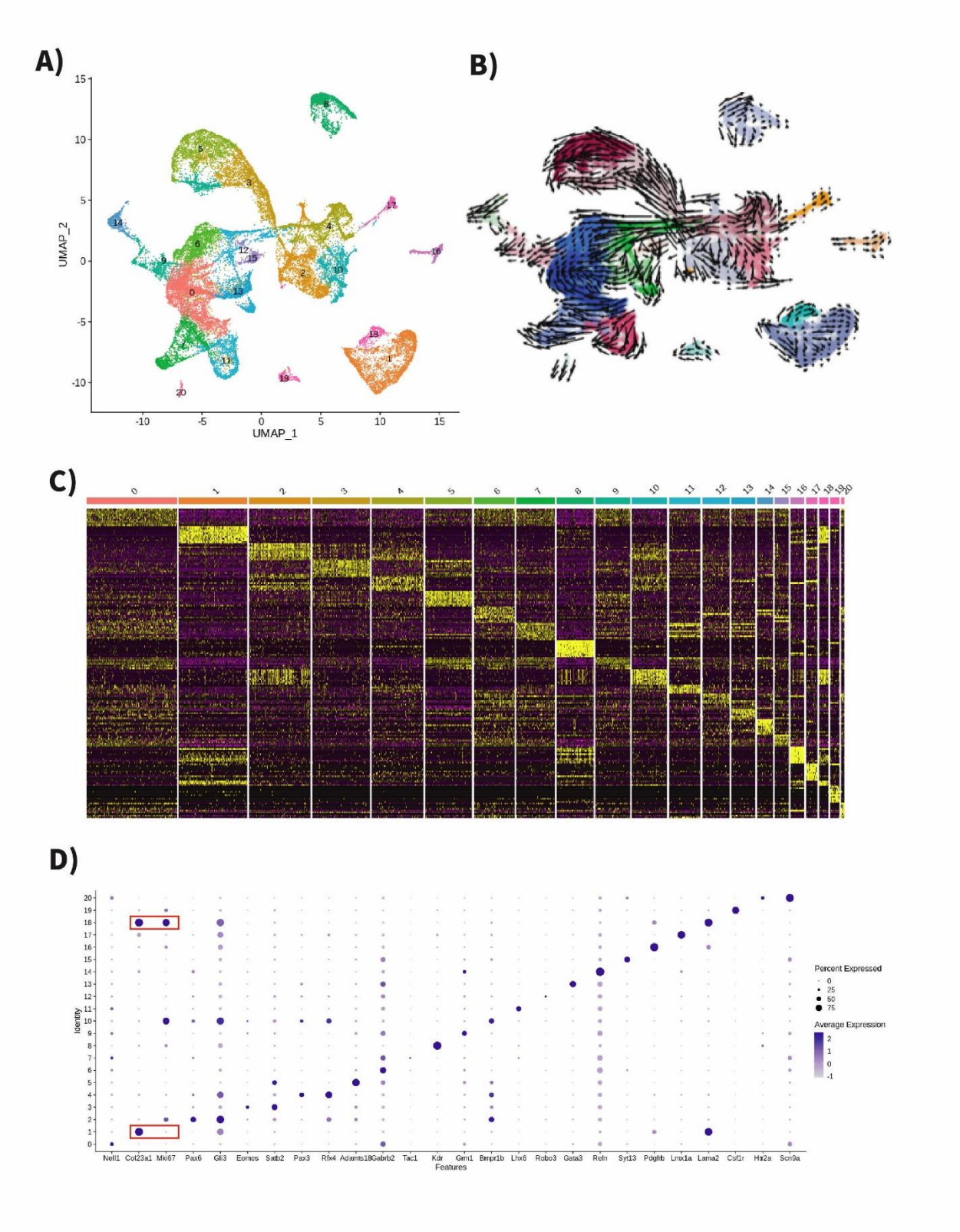

A)
B)
C)
D)
UMAP_1
UMAP_2
Nell1
Col23a1
Mki67
Pax6
Gli3
Eomes
Satb2
Pax3
Rfx4
Adamts18
Gabrb2
Tac1
Kdr
Grm1
Bmpr1b
Lhx6
Robo3
Gata3
Reln
Syt13
Pdgfrb
Lmx1a
Lama2
Csf1r
Htr2a
Scn9a
Features
Identity
Percent Expressed
Average Expression

Figure 1: **Visualization of cell clusters, RNA velocity, and marker gene expression in mouse E12.5 forebrain 6 hours after maternal PolyIC or saline treatment.**

(A) UMAP plot showing distinct clusters, each represented by different colors and numbers. (B) Cell velocity analysis with arrows indicating local directional changes. Note generalized progenitor cells in central cluster serve as origin for directional change towards IPCs and maturing pallial (left and up) and subpallial (left and down) neurons, as well as astrocytes and oligodendrocytes (right). For smaller clusters that do not arise from these generalized progenitors, directionality captures progenitor substructure. (C) Heatmap of unbiased marker genes displaying the top genes for each cluster with higher expression levels highlighted in yellow blocks. D) Dot plot of curated cell marker genes for each cluster, where the size of the dots indicates the percentage of cells expressing the gene, and the color reflects the mean expression level.

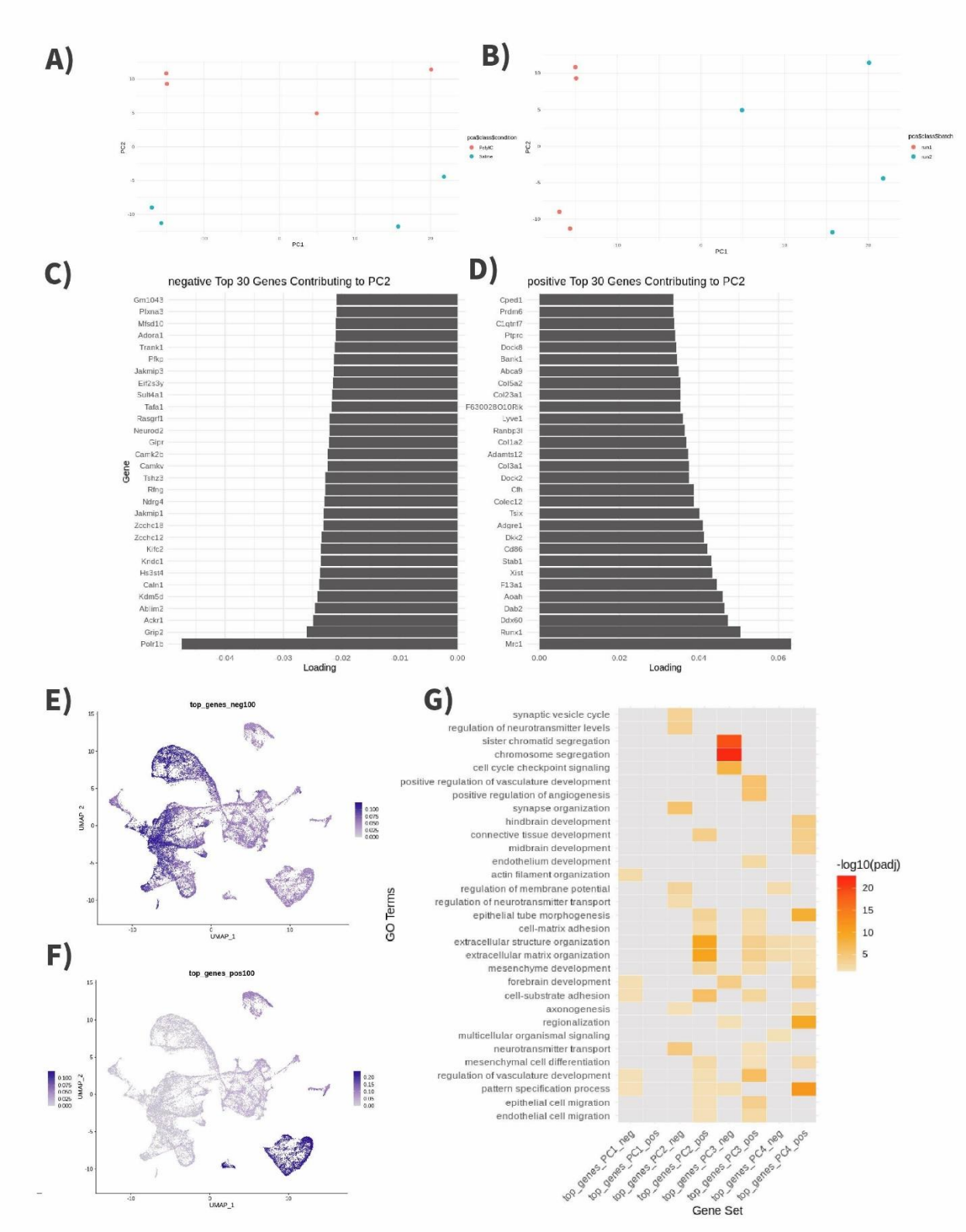

A)
B)
C)
negative Top 30 Genes Contributing to PC2
Gene
Loading
D)
positive Top 30 Genes Contributing to PC2
Loading
E)
top_genes_neg100
F)
top_genes_pos100
G)
GO Terms
synaptic vesicle cycle
regulation of neurotransmitter levels
sister chromatid segregation
chromosome segregation
cell cycle checkpoint signaling
positive regulation of vasculature development
positive regulation of angiogenesis
synapse organization
hindbrain development
connective tissue development
midbrain development
endothelium development
actin filament organization
regulation of membrane potential
regulation of neurotransmitter transport
epithelial tube morphogenesis
cell-matrix adhesion
extracellular structure organization
extracellular matrix organization
mesenchyme development
forebrain development
cell-substrate adhesion
axonogenesis
regionalization
multicellular organismal signaling
neurotransmitter transport
mesenchymal cell differentiation
regulation of vasculature development
pattern specification process
epithelial cell migration
endothelial cell migration
-log10(padj)
Gene Set

Figure 2: **Principal components analysis captures systematic MIA signatures.** (A-B) Top loading genes on PC1 are shown to separate the data by batches in (A), while PC2 differentiates the MIA condition in (B). (C-D) Display of top negative and positive loading genes on PC2. (E-F) UMAP plots showing the spatial distribution of negative loading genes in (E) and positive loading genes in (F) on PC2. (G) Gene Ontology (GO) term enrichment analysis for the top 100 loading genes from PC1 to PC4, with color representing the adjusted p-value for the enrichment.

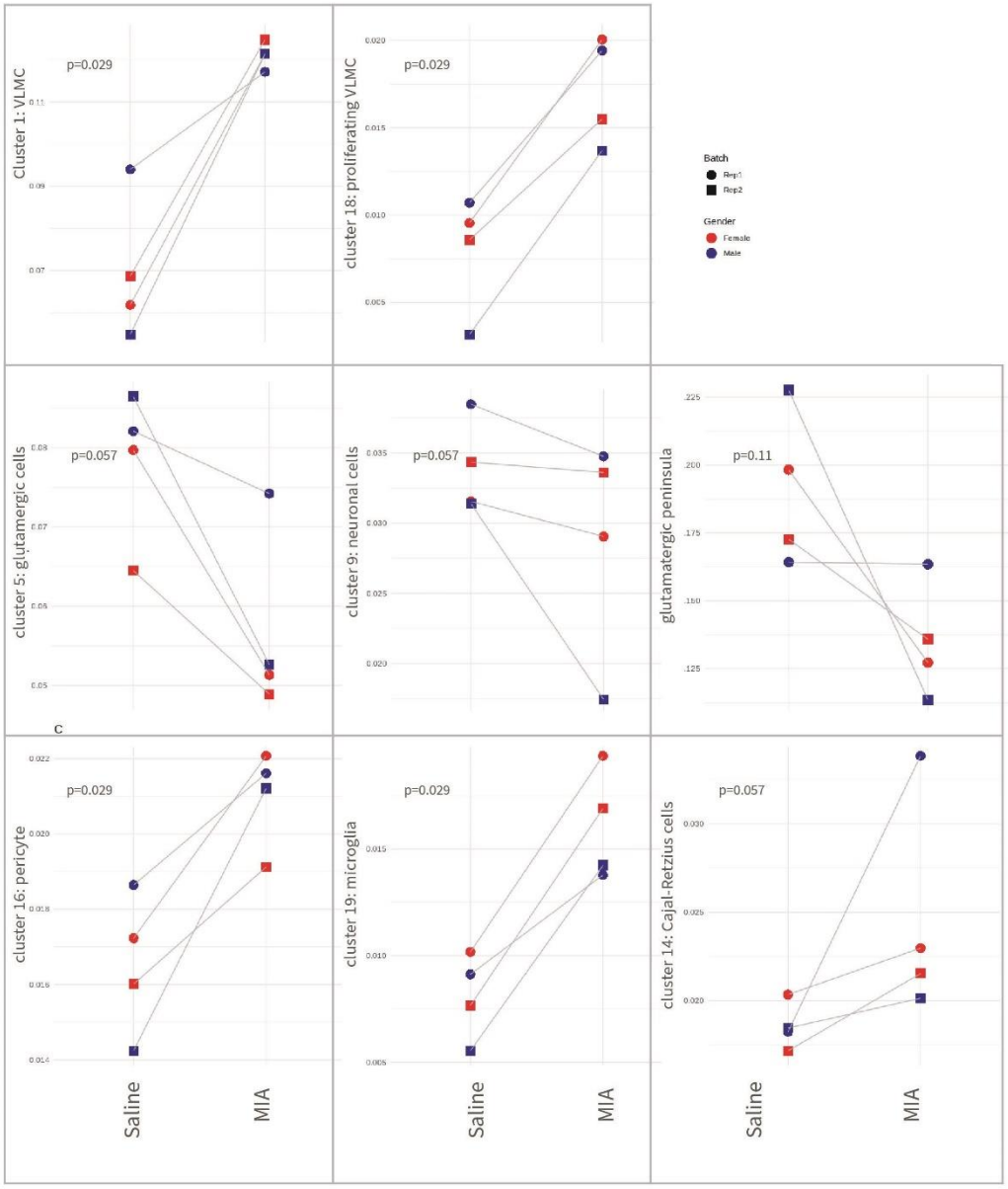

Cluster 1: VLMC
p=0.029
cluster 18: proliferating VLMC
p=0.029
Batch
Rep1
Rep2
Gender
Female
Male
cluster 5: glutamergic cells
p=0.057
cluster 9: neuronal cells
p=0.057
glutamatergic peninsula
p=0.11
C
cluster 16: pericyte
p=0.029
cluster 19: microglia
p=0.029
cluster 14: Cajal-Retzius cells
p=0.057
Saline
MIA
Saline
MIA
Saline
MIA

Figure 3: **Specific cell clusters exhibit changes in proportional representation following MIA.** Proportion of each cell type within their respective pseudobulk samples are on y axis, representing seven distinct cell types. All clusters with significant proportion changes based on Wilcoxon test are displayed here, as well as clusters worth noting. Clusters belonging to "Glutamatergic Peninsula" (3, 5, and a portion of cluster 9 that aligns with the Glutamatergic cell lineage) are shown. The displayed portion of cluster 9 has been adjusted by excluding the segment adjacent to the Glutamatergic region included in the Glutamatergic Peninsula. The columns in the figure are separated by MIA treatment condition, and colors indicate the gender of the pseudobulk samples, while shapes differentiate the batches.

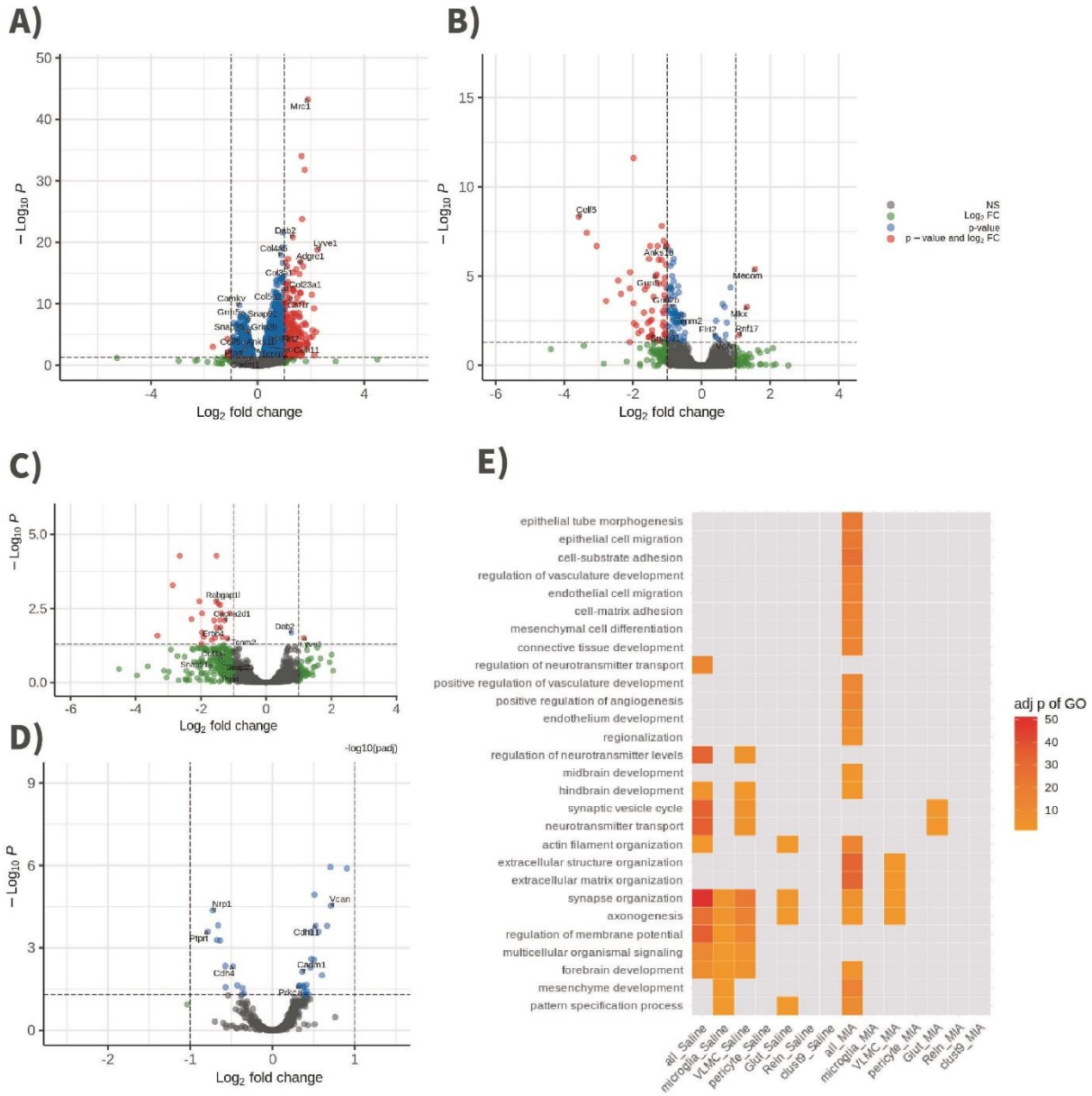

A)
B)
C)
D)
E)
− Log10 P
Log2 fold change
NS
Log2 FC
p-value
p − value and log2 FC
-log10(padj)
adj p of GO
epithelial tube morphogenesis
epithelial cell migration
cell-substrate adhesion
regulation of vasculature development
endothelial cell migration
cell-matrix adhesion
mesenchymal cell differentiation
connective tissue development
regulation of neurotransmitter transport
positive regulation of vasculature development
positive regulation of angiogenesis
endothelium development
regionalization
regulation of neurotransmitter levels
midbrain development
hindbrain development
synaptic vesicle cycle
neurotransmitter transport
actin filament organization
extracellular structure organization
extracellular matrix organization
synapse organization
axonogenesis
regulation of membrane potential
multicellular organismal signaling
forebrain development
mesenchyme development
pattern specification process
all_Saline
microglia_Saline
VLMC_Saline
pericyle_Saline
Glut_Saline
Rein_Saline
clus9_Saline
all_MIA
microglia_MIA
VLMC_MIA
pericyle_MIA
Glut_MIA
Rein_MIA
clus9_MIA

Figure 4: **MIA-associated differential gene expression and GO enrichment at the combined and cell-type specific level.** (A-D) Volcano plots of the differentially expressed genes under the MIA condition on all cells passed QC (A), VLMC cell type (B), microglia (C), and Glutamatergic neurons (D), with upregulated genes displaying positive log2 fold change (log2FC) values. Each gene is represented by a dot, with colors indicating whether they passed the threshold of adjusted p-value and log2FC.

(E) GO terms associated with DE genes detected using FindMarkers Wilcoxon test filtered by p-adj < 0.05, where terms labeled with "_Saline" indicate downregulated genes under MIA, and those labeled with "_MIA" represent upregulated genes. The GO results was further filtered by p-adj < 0.1, q-value < 0.1, and inclusion of more than one gene per GO term. Color scale: gray indicates absence of genes from that GO term in the DE gene set, while an orange-red gradient reflects increments of the adjusted p-value.

A

All Down	Glut. Down	All Up	Micro. Up	VLMC Up	Peri. Up	Glut. Up
Camkv	*Zfp385b*	*Mrc1*	*Dab2*	*Mecom*	*Dgkb*	*Sv2b*
Trank1	*Nrp1*	*Runx1*	*Lyve1*	*Rnf17*	*Rgs7bp*	*St18*
Grm5	*Cdh18*	*Ddx60*	*Nfatc1*	*Flrt2*	*Brip1*	*Slco3a1*
Jpt1	*Atf4*	*Aoah*	*Rgl1*	*Slc1a3*	*Rad51b*	*Vcan*
Hs3st4	*Bnc2*	*Itga11*	*Dock2*	*Cacna2d3*	*Diaph3*	*Cntn5*
Lrrtm4	*Zdhhc14*	*Dab2*	*Adgrv1*	*Vcan*	*Ncapd2*	*Cdh11*
Zcchc18	*Gnb2*	*Slc1a3*	*Mrc1*	*Grm7*	*Itga4*	*Pcsk2*
Jakmip1	*Zbtb20*	*Lyve1*	*Slc9a9*	*Adamts12*	*Brip1os*	*Csmd3*
Fbxo41	*Lingo2*	*Col4a5*	*Psd3*	*Arrdc3*	*Efcab11*	*Fam13a*
Cdh12	*Cdh4*	*Agmo*	*Rasal2*	*Eda*	*Cacnb2*	*Zc3h6*
Rbfox1	*Gm49678*	*Adgre1*	*Tgfbr2*	*Fbxl4*	*Tmem132c*	*Prkcb*
Snap91	*Eif4a1*	*Mertk*	*Gm42047*	*B130024G19Rik*	*Sema5a*	*Arrdc3*
Unc5d	*Pard3*	*Frem1*	*Runx1*	*Agmo*	*AY036118*	*Gm14697*
Sult4a1	*Pou6f2*	*Cd86*	*Mitf*	*9630014M24Rik*	*Cep128*	*Trpm3*
Mir124a-1hg	*Meis2*	*Col3a1*	*Hdac9*	*Fbn2*	*Ncapd3*	*Gm29478*
Ackr1	*Ivns1abp*	*Il16*	*Ddx60*	*Itga9*	*Itpr1*	*Ebf3*
Syngr3	*Rfng*	*Stab1*	*Sash1*	*St6galnac3*	*Atad2*	*Lcorl*
Lrrtm3	*Gas5*	*Ptprc*	*Ect2*	*Arhgap20*	*Pcnt*	*Lgi1*
Adora1	*Ntm*	*Ehd4*	*Vwa5a*	*Acss3*	*Sntb1*	*Cadm1*
Zcchc12	*Gm36560*	*Lama2*	*Diaph2*	*Colec12*	*Gpc4*	*Pde1a*

B

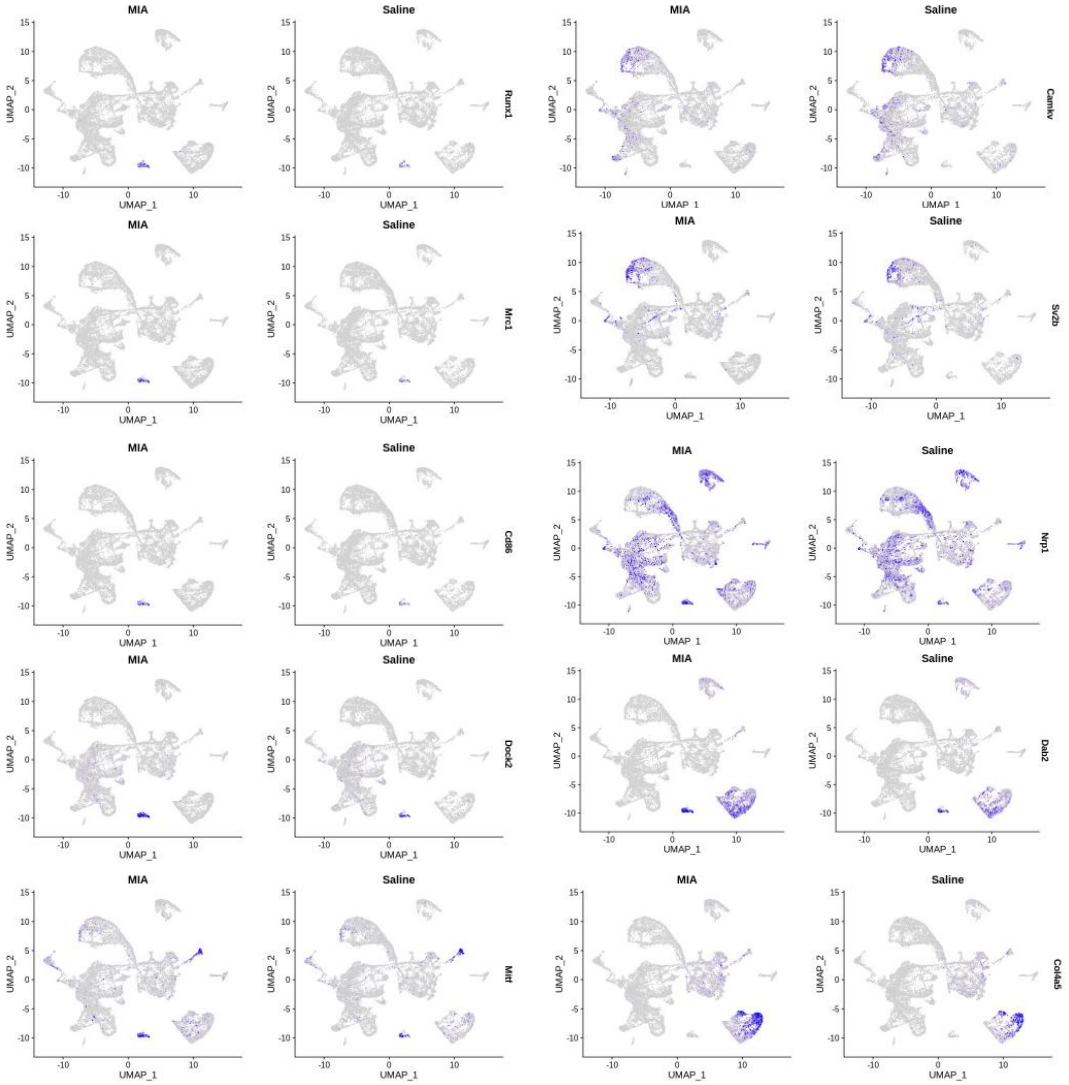


Figure 5: Differentially expressed gene list and representative expression plots. (A) Top 20 differentially expressed genes following MIA treatment, showing both up-regulated and down-regulated genes across selected cell populations. (B) Relative expression levels of representative DEGs, with MIA-treated samples shown on the left and saline control samples on the right. The color intensity reflects gene expression levels.

Material and Methods:

Maternal immune activation

Assessment of females baseline immunoreactivity before pregnancy and maternal immune activation were performed as previously described (Estes et al., 2020). Since female mice display a wide range of baseline immunoreactivity to poly(I:C) that dictates susceptibility and resilience of offspring from later pregnancies to MIA, virgin females in the lowest 25th percentile of immune responsiveness to poly(I:C) (IL-6 serum levels) were excluded from the study to reduce variability and ensure sufficient immune responsiveness of all included dams. E12.5 pregnancies were determined by visualizing vaginal seminal plug (noted as E0.5) and by body weight increase. IL-6 serum levels were measured at 2.5 hr post-injection. Temperature, weight, and sickness behavior were tracked as previously described (Estes et al., 2020).

Tissue dissections for RNA-seq and immunoblotting

Dissections of dorsal telencephalon/pallium or developing cortex were performed at E12.5 + 6 hr control (saline) and poly(I:C)-exposed MIA litters. At E12.5, due to the smaller embryo size, dissections were made via a diagonal cut that maximized collection of dorsal telencephalon, but could have also included some subpallium/ventral telencephalon and superficial tissues. Potential variance in the

individual dissections should be mitigated by the inclusion of multiple replicates across both groups. Dissections included both hemispheres from males and females.

Sample Collection and Single-Cell RNA Sequencing

Whole-brain samples were collected from both male and female mice that had been injected with either PolyIC or saline at e12.5, and samples were collected 6 hours post-injection to capture the immediate effects of the treatment. Single-cell sequencing was prepared from these whole-brain samples and subsequently processed using 10x Chromium single-cell RNA sequencing. Raw sequencing data was aligned and processed using the Cell Ranger pipeline (10x Genomics) and got alignment to the mm10 mouse reference genome with intron-mode activated.

Data Preprocessing

After alignment through CellRanger (Zheng et al., 2017), data has been filtered based on specific thresholds for gene expression and total RNA counts to exclude low-quality cells and potential doublets. For each experimental condition and replicate, different thresholds were used to tailor the filtering process to the specific characteristics of each dataset. Additionally, the SoupX package (Young et al., 2020) was employed to adjust for ambient RNA contamination, ensuring that the gene expression profiles reflect true intracellular mRNA levels. Filtered datasets from each condition and replicate were normalized and integrated using the SCTransform method (Hafemeister et al., 2019), which accounts for sequencing depth and batch effect.

Principal component analysis (PCA) has been performed to reduce the dimensionality of the data and capture the primary sources of variation, with the first 30 principal components selected for further analysis, and these were used to generate a Uniform Manifold Approximation and Projection (UMAP) plot, which provided a two-dimensional visualization of the cells. Finally, a shared nearest neighbor graph was constructed, and cells were clustered using the Louvain algorithm.

Cell Type Identification and Marker Gene Analysis

To identify the distinct cell populations, an unbiased approach was first employed using the FindAllMarkers function from the Seurat package, which compares each cluster against all other clusters. These markers were visualized using heatmaps, where clusters with strongly expressed positive markers were highlighted in yellow, allowing for an initial, unsupervised identification of major cell types based on their unique gene expression profiles.

In addition to the unbiased marker identification, a set of manually selected known marker genes was utilized to refine the identification of specific cell types, particularly in cases where clusters shared similar expression profiles and were difficult to distinguish solely based on the unbiased approach. These known markers were chosen based on their established roles in defining specific cell types in the brain. The expression patterns of these markers were visualized using dot plots, where the size of the dots

indicated the percentage of cells expressing each gene within a cluster, and the color gradient represented the average expression level. This combined approach of unbiased marker discovery and manual marker validation ensured robust and accurate cell type identification across the different clusters.

Pseudobulk Analysis and differential expression

Pseudobulk has been applied to investigate the differential gene expression across different treatment conditions. This approach involves aggregating single-cell RNA-seq data into bulk-like profiles, allowing the use of standard bulk RNA-seq analytical techniques. The four snRNA-seq datasets from two samples treated with PolyIC and two treated with Saline had been divided into male and female groups, resulting in eight pseudobulk samples.

For each of these pseudobulk samples, RNA counts for each gene across cells were summed and filtered to retain genes with a minimum expression threshold of 100 counts in total. The pseudobulk data were subsequently analyzed using the DESeq2 package (Love et al., 2014), which accounted for variables such as treatment condition (PolyIC vs. Saline), sex (male vs. female), and batch effects. Differential expression analysis was then performed to identify genes that were significantly upregulated or downregulated in response to PolyIC treatment.

PCA was performed on the pseudobulk samples after the gene count matrix was variance-stabilized using regularized log (rlog) transformation, which normalizes the data and mitigates the effects of variability in sequencing depth.

RNA Velocity Analysis

RNA velocity analysis was conducted using the scVelo method to explore the dynamic developmental trajectories of cells. RNA velocity estimates the future state of cells by modeling the transition from unspliced to spliced mRNA, providing insights into the direction and speed of cellular differentiation processes.

The Seurat object with UMAP information were loaded into an AnnData object, and the spliced and unspliced RNA matrices were derived from Loom files generated through velocyto.

The data then underwent filtering and normalization using filter_and_normalize from scvelo package to ensure high-quality inputs for velocity estimation, followed by the calculation of moments for each cell. RNA velocities were then estimated using the stochastic mode in scVelo, with the resulting velocity graph embedded in the UMAP space to represent the directional flow of cells through different developmental states.

Statistical Analysis of Cell Type Proportions

To assess the impact of MIA on cell type proportions, the Wilcoxon rank-sum test was performed. Proportions of each cell type were calculated by dividing the number of cells in each cluster by the total number of cells in each condition. This data was further categorized by batch and gender information to ensure a comprehensive analysis. The Wilcoxon rank-sum test was applied using the function wilcox.test to assess differences in cell type proportions across three comparisons: PolyIC-treated versus saline controls, male versus female, and batch 1 versus batch 2.

Reference

Andreatta, M., & Carmona, S. J. (2021). UCell: Robust and scalable single-cell gene signature scoring. Computational and structural biotechnology journal, 19, 3796-3798.

Bergen, V., et al. (2020). Generalizing RNA velocity to transient cell states with dynamical modeling. Nature Biotechnology, 38(12), 1408-1414.

Bergdolt, L., & Dunaevsky, A. (2019). Brain changes in a maternal immune activation model of neurodevelopmental brain disorders. Progress in neurobiology, 175, 1-19.

Bifari, F., Berton, V., Pino, A., Kusalo, M., Malpeli, G., Di Chio, M., ... & Decimo, I. (2015). Meninges harbor cells expressing neural precursor markers during development and adulthood. Frontiers in cellular neuroscience, 9, 383.

Canales CP, Estes ML, Cichewicz K, Angara K, Aboubechara JP, Cameron S, Prendergast K, Su-Feher L, Zdilar I, Kreun EJ, Connolly EC, Seo JM, Goon JB, Farrelly K, Stradleigh TW, van der List D, Haapanen L, Van de Water J, Vogt D, McAllister AK, Nord AS. Sequential perturbations to mouse corticogenesis following in utero maternal immune activation. Elife. 2021 Mar 5;10:e60100. doi: 10.7554/eLife.60100. PMID: 33666173; PMCID: PMC7979158.

Chu, N. (2024) sxreveal: A Python Package for SXRF Data Analysis. GitHub repository. Available at: https://github.com/nickolas-chu/sxreveal (Accessed: 18 August 2024)

Estes ML, McAllister AK. Maternal immune activation: Implications for neuropsychiatric disorders. Science. 2016 Aug 19;353(6301):772-7. doi: 10.1126/science.aag3194. PMID: 27540164; PMCID: PMC5650490.

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biology, 20(1), 296.

Hameete, B. C., Fernández-Calleja, J. M. S., de Groot, M. W. G. D. M., Oppewal, T. R., Tiemessen, M. M., Hogenkamp, A., de Vries, R. B. M. and Groenink, L. (2020). The poly(I:C)-induced maternal immune activation model; a systematic review and meta-analysis of cytokine levels in the offspring. Brain Behav. Immun. Health 11, 100192. doi:10.1016/j.bbih.2020.100192

Hornig, M., Bresnahan, M. A., Che, X., Schultz, A. F., Ukaigwe, J. E., Eddy,M. L., Hirtz, D., Gunnes, N., Lie, K. K., Magnus, P. et al. (2018). Prenatal fever and autism risk. Mol. Psychiatry 23, 759-766. doi:10.1038/mp.2017.119

Knuesel, I., Chicha, L., Britschgi, M., Schobel, S. A., Bodmer, M., Hellings, J. A., ... & Prinssen, E. P. (2014). Maternal immune activation and abnormal brain development across CNS disorders. Nature Reviews Neurology, 10(11), 643-660.

Laighneach, A., Desbonnet, L., Kelly, J. P., Donohoe, G., & Morris, D. W. (2021). Meta-analysis of brain gene expression data from mouse model studies of maternal immune activation using poly (I: C). Genes, 12(9), 1363.

Liang, Z., Zhan, Y., Shen, Y., Wong, C. C., Yates III, J. R., Plattner, F., ... & Ip, N. Y. (2016). The pseudokinase CaMKv is required for the activity-dependent maintenance of dendritic spines. Nature communications, 7(1), 13282.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550.

Meyer U, Feldon J. To poly(I:C) or not to poly(I:C): advancing preclinical schizophrenia research through the use of prenatal immune activation models. Neuropharmacology. 2012

Mar;62(3):1308-21. doi: 10.1016/j.neuropharm.2011.01.009. Epub 2011 Jan 14. PMID: 21238465.

Nakagomi, T., Molnár, Z., Taguchi, A., Nakano-Doi, A., Lu, S., Kasahara, Y., ... & Matsuyama, T. (2012). Leptomeningeal-derived doublecortin-expressing cells in poststroke brain. Stem cells and development, 21(13), 2350-2354.

Ostrem, B. E. L., Domínguez-Iturza, N., Stogsdill, J. A., Faits, T., Kim, K., Levin, J. Z., & Arlotta, P. (2024). Fetal brain response to maternal inflammation requires microglia. Development, 151, dev202252.

Smith SE, Li J, Garbett K, Mirnics K, Patterson PH. Maternal immune activation alters fetal brain development through interleukin-6. J Neurosci. 2007 Oct 3;27(40):10695-702. doi: 10.1523/JNEUROSCI.2178-07.2007. PMID: 17913903; PMCID: PMC2387067.

Stuart, T., & Satija, R. (2019). Integrative single-cell analysis across multimodalities with Seurat. Nature Biotechnology, 37(5), 459-467.

Sun, Y., Chen, G., He, J., Huang, Z. G., Li, S. H., Yang, Y. P., ... & Wu, H. (2021). Clinical significance and potential molecular mechanism of miRNA-222-3p in metastatic prostate cancer. Bioengineered, 12(1), 325-340.

Tansley, S., Uttam, S., Ureña Guzmán, A., Yaqubi, M., Pacis, A., Parisien, M., ... & Khoutorsky, A. (2022). Single-cell RNA sequencing reveals time-and sex-specific responses of mouse spinal cord microglia to peripheral nerve injury and links ApoE to chronic pain. Nature communications, 13(1), 843.

Tsivion-Visbord, H., Kopel, E., Feiglin, A., Sofer, T., Barzilay, R., Ben-Zur, T., ... & Levanon, E. Y. (2020). Increased RNA editing in maternal immune activation model of neurodevelopmental disease. Nature communications, 11(1), 5236.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83. doi:10.2307/3001968

Yamashita, Y., Fujimoto, C., Nakajima, E., Isagai, T. and Matsuishi, T. (2003). Possible association between congenital cytomegalovirus infection

and autistic disorder. J. Autism Dev. Disord. 33, 455-459. doi:10.1023/A:1025023131029

Young, M. D., & Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience, 9(12), giaa151.

Zelco, A., Börjesson, V., De Kanter, J. K., Lebrero-Fernandez, C., Lauschke, V. M., Rocha-Ferreira, E., ... & Wang, X. (2021). Single-cell atlas reveals meningeal leukocyte heterogeneity in the developing mouse brain. Genes & Development, 35(15-16), 1190-1207.

Zheng, G. X. Y. et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications 8: 1-12, doi:10.1038/ncomms14049

Zimmerman, K. D., Evans, C., & Langefeld, C. D. (2022). Reply to: A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. Nature communications, 13(1), 7852.

Chapter 5

Conclusion

The dissertation focused on the intricate regulation of gene expression during neurodevelopment and in disease states, combining transcriptomic and functional genomic approaches. The regulation of gene expression, particularly in neurodevelopment, is governed by the precise spatio-temporal action of transcription factors (TFs), which interact with regulatory elements like promoters and enhancers. Chapters 2, 3, and 4 build on this by exploring how gene regulatory networks, controlled by TFs and other epigenomic marks, shape brain development and respond to disease-related perturbations.

We analyzed the brain cells using scRNA-seq and epigenomic datasets, focusing on early developmental time points. We analyzed for the importance of Dlx1/2 as regulatory factors during the development of GABAergic interneurons, highlighting the complex network of TFs to modulate gene expression differently under in different sub areas of ganglionic eminence. As Dlx1/2 are transcription factors functioning with enhancer elements to regulate developmental processes, we further analyzed the putative regulatory elements under various developmental timepoints under normal condition, specifically focusing on MGE cells developed from ESC through multiple epigenomic marks. We also looked at the cell-level effects of maternal immune

activation during development, checking on how MIA alters the developmental processes of a variety of brain cells.

The findings in these chapters highlights the complexity of gene regulatory networks in both development and disease, with a special focus on ganglionic eminence derived interneurons, particularly the role of Dlx1/2 and genomic regions acting as stage-specific putative enhancers in their development and specification. Additionally, we proposed novel hypothesis on how environmental disruptions such as MIA can perturb the cell population proportions.

Together, these studies provide critical insights into how both internal (genetic) and external (environmental) factors regulate neurodevelopment and how disruptions to these processes can lead to disorders like autism, schizophrenia, or neurodegenerative diseases.