UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Advancing Models of Transcriptional Diversity in Human Brain Development

Permalink
https://escholarship.org/uc/item/9t10q1ww

Author
Pebworth, Mark-Phillip L

Publication Date
2020

Supplemental Material
https://escholarship.org/uc/item/9t10q1ww#supplemental

Peer reviewed|Thesis/dissertation

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

Advancing Models of Transcriptional Diversity in Human Brain Development:
Mapping Cell-type Specific Enhancers and Neuronal Progenitor Diversity

by
Mark-Phillip Pebworth

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

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Arturo Alvarez-Buylla
A98A3A270B5D492...

Arturo Alvarez-Buylla
Chair

DocuSigned by:
[signature]
[illegible]

Arnold Kriegstein

DocuSigned by:
Daniel Lim
[illegible]

Daniel Lim

DocuSigned by:
Michael Oldham
566EEE787F97430...

Michael Oldham

Committee Members

Acknowledgements: ***Posuimus Enim Resurgemus***
I could not have finished this work without the support of the community around me.

I want to first thank my thesis advisor, **Arnold Kriegstein**, for his support and kindness over the last several years. My thesis work also benefited from the input of my thesis committee members, **Dan Lim**, **Arturo Alvarez-Buylla**, and **Michael Oldham**.

Within the Kriegstein lab, **Shaohui Wang** was invaluable in collecting, and processing tissue samples, a key laborious step for much of the data I generated. **William Walantus**, the lab manager of lab managers, was always around when microscopes went down, or key orders need expediting, rain or shine. One key individual, **Jayden Ross**, was everything I was looking for in an intern – consistent, responsible, and a quick study on protocols. Much staining and imaging for Chapter 2 were conducted by them, and much more remains unpublished. I expect it will be very useful for future work in the lab.

I also want to thank the members of the **Yin Shen lab**, with whom I very closely collaborated for Chapter 1. **Yin Shen** opened up a whole new field to me, and I appreciated the opportunity to work closely with **Michael Song**, and **Xiaoyu Yang**, who co-authored the manuscript with me. We worked together through much successful and failure to generate the datasets behind the manuscript.

Finally, this acknowledgement would not be complete without mentioning the difficulties I’ve experienced during my PhD. My son was born, my wife diagnosed with cancer, a very dear grandmother passed away, and the COVID19 epidemic shutdown the world. **Aparna Bhaduri**, and **Madeline Andrews** really stepped in with mentorship. They

brought clarity and hope in my work when things were looking grim personally. Their support helped me stay afloat, and make progress despite what was going on. Likewise, Arnold was incredibly understanding, and offered many valuable connections around my wife's cancer. A soft touch can be difficult to quantify. Nevertheless, I found his concern and empathy for my wife's situation so reassuring, and I believe it greatly helped my forward momentum to have a PI such as him. I also want to thank my parents, who despite knowing little about how graduate school and PhDs worked, always urged me on. Despite all the difficulties, they pushed me to never stop in pursuit of my doctoral education. Lastly, I would be remiss to not thank my dear wife, who is currently battling cancer while I finish this thesis. She has had to live with the stress of my PhD, the long hours in the lab, and the uncertainty of my graduation timeline. I cannot thank her enough.

Contributions:

Figures and results from Part 1 were previously published as cited below:

Song, Michael[1], Mark Phillip Pebworth[1], Xiaoyu Yang[1], Armen Abnousi, Changxu Fan, Jia Wen, Jonathan D. Rosen, et al. 2020. "Cell-Type-Specific 3D Epigenomes in the Developing Human Cortex." *Nature* 587 (7835). https://doi.org/10.1038/s41586-020-2825-4.

[1] = Co-first author

Advancing Models of Transcriptional Diversity in Human Brain Development: Mapping Cell-type Specific Enhancers and Neuronal Progenitor Diversity

Mark-Phillip Lee Pebworth

Abstract:

The human cerebral cortex is comprised of a diverse set of neurons that are essential to its functions in perception, judgment, and sensory integration. The complexity of this organ emerges during development with the aid of tightly coordinated gene expression. Modern techniques now allow us to map the transcriptional diversity of cell types in the developing brain and the epigenetic alterations that generate them. Here, we explore two projects that advance our understanding of transcriptional diversity in the developing human cortex. The first maps cell-type specific enhancers that might drive transcriptional diversity, and the second identifies subtypes of neuronal progenitors in the dorsal cortex. To map enhancers in part 1, we ran a PLAC-Seq, ATAC-seq, and RNA-seq on four FACS-sorted populations of human neural cell types and validated our cell-type specific results using a CRISPRi system. For part 2, the identification of neuronal progenitors relied on single cell RNA-seq results, paired with histology and viral labeling to identify 5 subpopulations. One of these populations stained for DLX5 and was related to a late neurogenic shift in neuronal progenitor location, while the other four populations dominated early neurogenesis through the germinal zone of the developing cortex. Interestingly, we did not see a clear relationship between morphology and cell type identity.

Table of Contents

INTRODUCTION 1

Models of Cortical Development 1

Enhancers 2

Significance and Thesis Overview 3

Contributions 5

CHAPTER 1: CELL TYPE-SPECIFIC 3D EPIGENOMES IN THE DEVELOPING HUMAN CORTEX 6

Introduction 7

Results 8

Sorting specific cell types from the developing human cortex 8
Characterizing cell type-specific 3D epigenomes 9
Chromatin interactions influence cell type-specific transcription 11
Transposable elements in SIP formation 15
Developmental trajectories from RG to eNs 18
Human-specific aspects of cortical development 18
Partitioning SNP heritability for complex disorders and traits 19
Characterizing distal interacting regions in primary cells 21

Discussion 22

Acknowledgements 24

Author contributions 24

Data availability statement 25

Methods 25

Supplemental Figures 42

CHAPTER 2: HUMAN INTERMEDIATE PROGENITOR DIVERSITY DURING CORTICAL DEVELOPMENT 58

Abstract 58

Introduction 59

Results 61

Intermediate Progenitors follow distinct spatiotemporal patterns in early vs late neurogenesis 61
Analysis of Single Cell RNA Sequencing Identifies Subtypes of Intermediate Progenitor Cells 64
Validation of Intermediate Progenitor Cell Subtypes throughout Human Neurogenesis 67
nIPC morphology does not correlate with transcriptomic subtypes 72

Discussion 76

Methods 82

Supplemental Figures 87

REFERENCES 94

Table of Figures

Figure 0.1. Models of human cortical development and transcriptional regulation by enhancers. 6

Figure 1.1. Experimental design and features of 3D epigenomes during human corticogenesis. 9

Figure 1.2. H3K4me3-mediated chromatin interactions influence cell type-specific transcription.. 11

Figure 1.3. Super interactive promoters are enriched for lineage-specific genes. 14

Figure 1.4. Features of cortical development and partitioning SNP heritability for complex disorders and traits. 17

Figure 1.5. Validation of cell-type specific distal regulatory elements using CRISPRview. 21

Extended Data Figure 1.1. Representative contour plots depicting FACS gating strategy. 42

Extended Data Figure 1.2. Reproducibility between RNA-seq, ATAC-seq, and PLAC-seq replicates..... 44

Extended Data Figure 1.3. Identification of significant H3K4me3-mediated chromatin interactions. 45

Extended Data Figure 1.4. Chromatin interactions influence cell type-specific transcription. 47

Extended Data Figure 1.5. Super interactive promoters are enriched for lineage-specific genes..... 49

Extended Data Figure 1.6. Transposable elements in SIP formation. 51

Extended Data Figure 1.7. Developmental trajectories and mapping complex

disorder- and trait-associated variants to their target genes. 53

Extended Data Figure 1.8. Partitioning SNP heritability for complex disorders and traits using alternative epigenomic annotations. 54

Extended Data Figure 1.9. Enriched biological processes for genes interacting with non-coding variants for each disease and cell type. 55

Extended Data Figure 1.10. Characterization of RG- and eN-specific loci using CRISPRview. .. 56

Figure 2.1. Spatiotemporal Diversity. 62

Figure 2.2. Bioinformatic Controls and Reproducibility 66

Figure 3. RG-Like nIPC Validation. 68

Figure 2.4. Neuron-like nIPC Validation. 70

Figure 2.5. DLX5+ nIPC Validation. 72

Figure 2.6. nIPC Morphology does not correlate well with transcriptomic subtypes. 75

Supplemental Figure 2.1: IPCs in the oSVZ appear more proliferative than those in the iSVZ during early neurogenesis. 87

Supplemental Figure 2.2: EOMES & TBR1 immunohistochemistry reveals a laminar pattern of transcription factor expression in the iSVZ during early neurogenesis. 88

Supplemental Figure 2.3: Further verification of EOMES staining pattern during late neurogenesis. 89

Supplemental Figure 2.4: Staining for neuronal markers support disappearance of neurogenesis from the oSVZ during late neurogenesis. 90

Supplemental Figure 2.5: The original dataset and its comparison with the clusters and markers from 1000 scrambled iterations. 91

Supplemental Figure 2.6: Subclusters of RG-Like IPCs do not appear to correlated well with RG-subtypes. 92

Supplemental Figure 2.7. nIPC Markers in other cell types. 94

Introduction

Models of Cortical Development

As human beings, our cognition depends on the diverse set of neurons and glia of the cerebral cortex. It defines who we are and what we observe around us. This incredibly complex organ begins in the developing embryo as the simple tube of epithelial cells known as the neural tube. In studying this developing structure, Wilhelm His discovered that rapidly dividing cells lay on the interior (aka ventricular) surface of this structure, and that “neuroblasts” migrate out from this rapidly dividing cell population, through radial structures formed by “spongioblast” cells (Bentivoglio and Mazzarello, 1999). In modern parlance, he discovered that the stem cell niche lay on the interior ventricular surface of the developing central nervous system. Influenced by His’s work, Cajal, Golgi, and Magini showed that the stem cell niche contained both nonradial proliferative cells and radial spongioblasts. As neuroanatomical terminology became standardized, these spongioblasts became known as radial glia (RG), since they were considered non-neuron cells (glia) with cell bodies at the ventricle and long fibers that extend to the outer edge of the developing cortex, known as the pial surface. As a postdoctoral fellow in the lab of Fernando Nottebohm, Arturo Alvarez-Buylla discovered that RG act as a neurogenic stem cell for the adult bird brain (Alvarez-Buylla, Theelen and Nottebohm, 1988, 1990). Further work in rodents validated this concept for the developing mammalian brain (Malatesta, Hartfuss and Götz, 2000; Miyata *et al.*, 2001; Noctor *et al.*, 2001). Around the same time as Alvarez-Buylla was studying radial glia, Pasko Rakic published the radial unit hypothesis (Rakic, 1988). Part of his hypothesis proposed that RG fibers guide neurons, born near the ventricular zone (VZ), to the cortical plate, the

layer of neurons that reside on the outer edge of the cerebral cortex (Rakic, 1971, 1972). This mechanistic framework predicted that excitatory neurons from a given RG were guided to the same columnar region of the cortical plate. As a result, there is theoretically a 1-to-many positional map between RG in the ventricle and excitatory neurons in the cortex.

However, this hypothesis did not discuss the non-radial proliferative cells of Golgi, Cajal, and Magini. This specific observation began to be explained by the live-imaging work of Noctor et al. in 2004, in which RG were seen to produce a second cell type. Unlike the typical asymmetric divisions of RG, these daughter cells divide symmetrically to produce two neurons. As result, these neuronal progenitors are a continually produced and depleted population, in contrast to the more self-renewing RG. Lastly, it was noticed that these cells migrated from the VZ into the subventricular zone (SVZ) (Noctor *et al.*, 2004). Subsequent work validated the existence of a neuronal progenitor that divided away from the ventricular edge of the dorsal cortex, remarkably unlike radial glia (Miyata *et al.*, 2001; Haubensak *et al.*, 2004). This second cell type is currently known as a neural intermediate progenitor cell (nIPC). When first proposed, nIPCs were defined in contrast to radial glia, and terminology varied between different groups, with names varying from intermediate progenitor and intermediate radial glia to non-surfacing dividing cell, and basal progenitor (Miyata *et al.*, 2001; Haubensak *et al.*, 2004; Reillo *et al.*, 2011). This novel neurogenic cell type would escape a specific molecular definition until Englund et al. elucidated that TBR2 in mice, also known as EOMES in humans, was specifically expressed by nIPCs. This is in contrast to RG in the mouse cortex that express PAX6, and postmitotic neurons that express TBR1 (Englund and others, 2005;

Arnold *et al.*, 2008; Sessa *et al.*, 2008; Mihalas *et al.*, 2016). nIPCs were further implicated in the production of neurons for all cortical layers (Mihalas *et al.*, 2016). More importantly, nIPCs altered our models of human brain development. While the radial unit hypothesis remained intact, the number of neurons produced was now decoupled from the number of RG along the ventricle. Instead, the size of the nIPC population and the number of symmetric divisions undergone became factors directly influencing the number of neurons in the cortex. The nIPC population was now seen as a key modulator of neuron production, and the corresponding increase in surface area along the cortical plate. At this point, the literature suggested a relatively simple model of dorsal neurogenesis. Most of the time, RG produce nIPCs. nIPCs then generate the vast majority of neurons, which migrate along radial fibers to their final destination in the cortical plate.

However, this pattern only holds true for the dorsal regions of the developing neocortex. Ventrally, RG produce EOMES-negative nIPCs, which create inhibitory neurons. The migration of these neurons is initially guided by chemotactic signals and not ventral RG fibers (Stumm *et al.*, 2003; Wang *et al.*, 2011; Hansen *et al.*, 2013). A portion of these inhibitory neurons migrate tangentially to the dorsal cortex, passing through the germinal zones as well as the marginal zone, a thin band of cells on the superficial layers of the cortical plate (Miller, 1985; Fairén, Cobas and Fonseca, 1986; Anderson *et al.*, 1997). These interneurons are then partially guided by radial fibers to migrate into the cortical plate (Elias *et al.*, 2010). The developing dorsal cortex thus contains ventrally-derived interneurons, dorsally-derived excitatory neurons. all intermixed with

the local population of RG, and nIPCs in early stages of human development, as well as other support cells.

Our working model of human brain development changed significantly when a second neurogenic radial glia was discovered in the dorsal cortex (Hansen *et al.*, 2010). Born from ventricular RG (vRG) from about gestational week (GW) 15, this second RG subtype, known as an outer RG (oRG), migrates into the outer SVZ (oSVZ), losing contact with the ventricle but maintaining a radial fiber. This oRG helps create a large outer SVZ (OSVZ), which is divided from the inner SVZ by a thin acellular region classically interpreted to be a fiber layer (Zecevic, Chen and Filipovic, 2005; Fish *et al.*, 2008; Yang and others, 2020). These observations of the developing human brain were later replicated in other mammal species including non-human primates and carnivores (Reillo *et al.*, 2011; Gertz *et al.*, 2014).

Historically, oRGs were considered to be translocating RG, which were differentiating into astrocytes (Rakic, 1972; Choi and Lapham, 1978; Schmechel and Rakic, 1979). At the time of these observations, neurogenesis was considered a ventricular property not specifically associated with RG, and so such non-ventricular radial cells were not considered to be neurogenic. While oRGs do not appear to be gliogenic before GW18 or so, there is evidence that they are likely gliogenic after the peak of neurogenesis (Kostović, Išasegi and Krsnik, 2018). Nevertheless, early studies using thymidine labeling in primates showed that the proliferation in the oSVZ coincided with cortical neurogenesis, hinting at the presence of a neurogenic source in the oSVZ (Rakic, 1974, Lukaszewicz et al., 2005). Likewise, the presence of RG markers such as nestin (NES), vimentin (VIM), PAX6, and GFAP, was observed in oSVZ cells (Zecevic, Chen and

Filipovic, 2005; Bayatti *et al.*, 2008; Mo and Zecevic, 2008), though the relationship between the oSVZ and vRG was not clear. It had been previously hypothesized that RG-like epithelial cells or nIPC-like progenitors inhabited the oSVZ (Kriegstein, Noctor and Martínez-cerdeño, 2006; Fish *et al.*, 2008). These questions were settled when oRGs were observed to produce neurogenic EOMES+ cells (Hansen *et al.*, 2010). This second RG subtype provided a second means for rapid cortical expansion. RG were no longer limited to the two-dimensional surface of the VZ epithelium but could self-renew and proliferate throughout the oSVZ, a much larger three-dimensional germinal zone in the developing brain. Given that RG cells along the ventricle were no longer the only neurogenic units, a ventricle-to-cortical plate position map of the radial unit hypothesis was now in question. Furthermore, it was also noted that nIPCs derived from oRGs could divide multiple times, providing further amplification for the generation of neuronal progeny (Hansen *et al.*, 2010). This new model of brain development meant that two neurogenic lineages coexist for a period (Fig 0.1 A), and the means and rate of neuron production were far greater than previously imagined. Additionally, these co-existent neurogenic sources meant that transcriptional variation was likely to distinguish these two RG subtypes, and a number of oRG marker genes were soon identified (Pollen *et al.*, 2015).

The complexity of our model only grew with the advent of single-cell sequencing. Studies of developing human telencephalon have found up to three transcriptomic subtypes of RG, and many neuron subtypes (Nowakowski *et al.*, 2016; Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent, Kriegstein, *et al.*, 2017;

Bhaduri *et al.*, 2020). In chapter two, I apply similar techniques to identify nIPC subtypes for the first time. Additionally, with the proliferation of cell types comes a proliferation of transcriptional markers. Each cell type or subtype requires its own set of active genes, creating a massive amount of transcriptional variation.

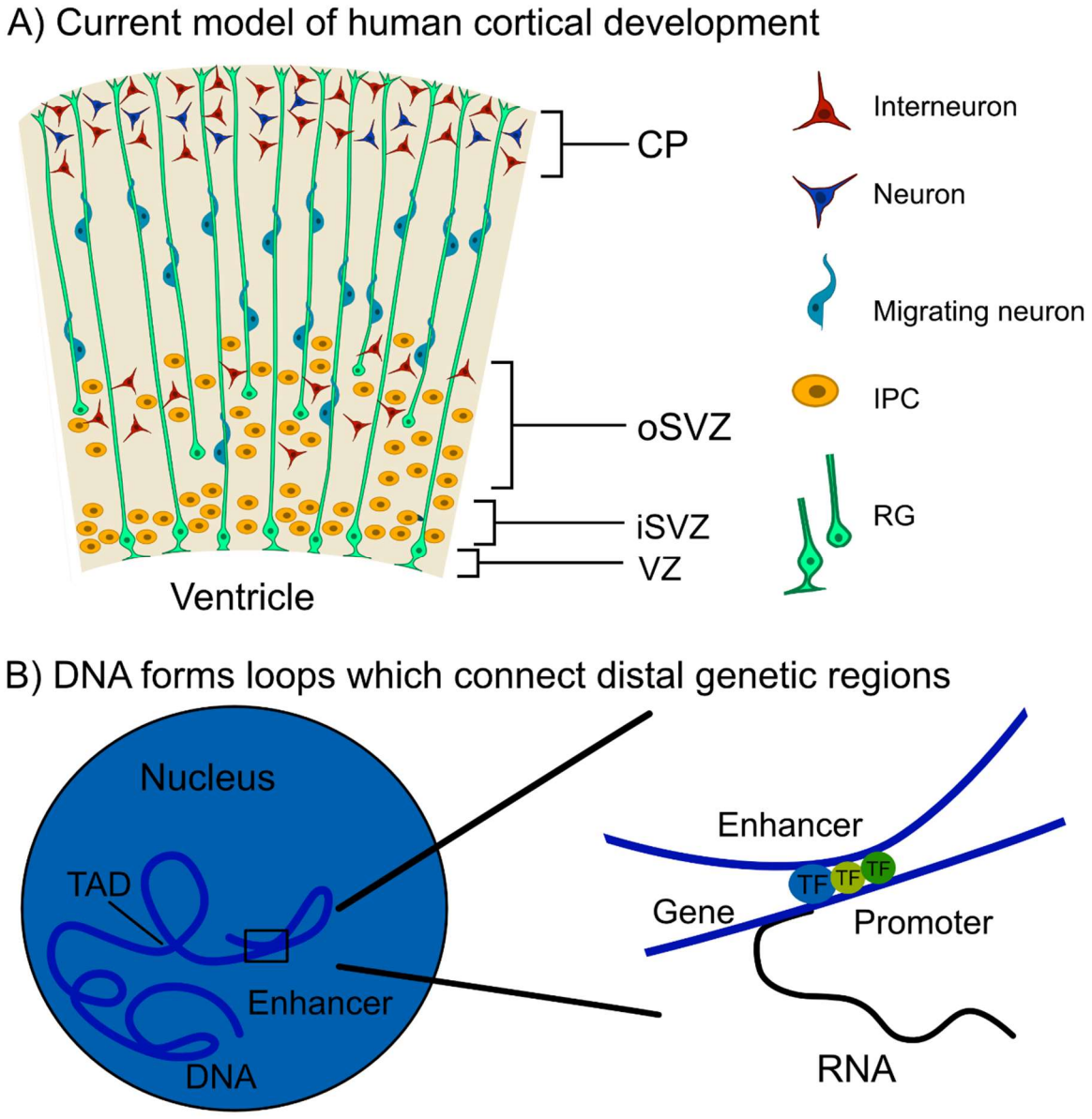


Figure 0.1 Models of human cortical development and transcriptional regulation by enhancers. A) The current model of human cortical development. VZ (ventricular zone), iSVZ (inner subventricular zone), oSVZ (outer subventricular zone), cortical plate (CP), IPC (intermediate progenitor cell), RG (radial glia, includes both outer radial glia and ventricular radial glia). **B)** DNA forms loops which connect distal genetic regions.

Some connections are structural, like TADs (topologically-associated domains), and others impact transcription, such as enhancer-promoter interactions.

Enhancers

Some of this transcriptional variation is explained by enhancers. Genes may be broadly expressed across different cell types, but at different levels. For example, SOX2 is known to be highly expressed in radial glia and some intermediate progenitors but it is also expressed at low levels in interneurons (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent and Kriegstein, 2017). Where does this fine-scale variation in expression come from? One possible mechanism involves the physical structure of transcriptional regulation. DNA forms physical loops within each nucleus (Zheng and Xie, 2019). These loops bring distal regions of the genome into close proximity (Fig. 0.1 B). It has been shown that this close proximity is important for the regulation of some genes and could play a critical role in development and cell fate commitment(Zheng and Xie, 2019). Previous work discovered that certain regions, known as enhancers, influence the transcription of distal genes based on physical interactions (Blackwood and Kadonaga, 1998). The dysregulation of these enhancer-gene interactions is also linked to complex disorders and traits (Li, Hu and Shen, 2018). However, these enhancers are difficult to identify from pure sequence analysis, because they do not follow common sequence-specific patterns as observed in promoters. Additionally, enhancers are known to act in lineage and cell-type specific manners (Zheng and Xie, 2019). As a result, enhancers must be observed by identifying an actual physical interaction in the tissue or cell type of interest. Furthermore, distal

genetic regions may interact structurally to form topologically associated domains (TAD), which are regions of co-regulated genes and enhancers (Li, Hu and Shen, 2018). These structural interactions, as well as enhancers and gene bodies, can be discovered by a technique called ATAC-seq, which identifies accessible regions of the genome (Li, Hu and Shen, 2018). These studies provide a useful dataset for understanding the structure of DNA within the nucleus, but cannot easily discriminate between structural interactions and enhancer elements (Li, Hu and Shen, 2018). Furthermore, past studies have relied on bulk brain tissue from the cortical plate or germinal zones (VZ, SVZ) to look for long distance physical interactions. The use of bulk tissue means that all measurements are an average across many cell types. With the exception of cell-type specific genes, cell-type specific promoter-enhancer information is difficult to glean from this bulk data. However, cell-type resolution is highly important in understanding the connection between a disease and potentially causal mutations in enhancer elements. This information permits clinicians and scientists to identify the cell type or lineage affected by a given enhancer mutation, and thus the potential etiology of disease symptoms. These cell-type specific characterizations are noticeably absent from the literature. In Chapter 1 of this thesis, I and others worked together to build a cell-type-specific map of enhancers during human brain development.

Significance and Thesis Overview

This thesis attempts to address two gaps in our current understanding of transcriptional diversity in the developing human cortex: 1) What enhancers are present and

influencing transcription within the major cell types of the developing human brain? 2). What transcriptional diversity exists within the nIPC population?

Chapter One lays out the first cell-type specific enhancer map in the developing human brain. To access cell type information, I developed a novel flow-assisted sorting strategy to isolate four major cell types from the human brain; radial glia, IPCs, excitatory neurons, and interneurons. For each population, we then used an emerging technique known as PLAC-seq to sequence the distal regions that are physical interactions with active promoters. In conjunction with ATAC-seq and RNA-seq measurements, we mapped enhancer interactions across the genome and validated several. Using this information, we were able to do preliminary studies into neural disease etiology, transcriptional regulation, and human evolution.

Chapter Two explores the diversity of nIPCs in the human dorsal cortex. We identified five novel and reproducible populations of nIPCs in single-cell data and related them to periods of neurogenesis as well as morphology. We identified an uncharacterized and rapid shift in nIPC distribution across the germinal zone around GW19 and GW20. Before GW19, three subtypes of nIPCs predominate, and after GW20, a subtype of novel DLX5+ nIPC appears. Unlike RG that show temporal shifts in nIPC subtypes occur, nIPC morphology does not change with subtype, unlike RG. Lastly, we observed nIPC morphologies that were reminiscent of RG. In the light of past observations, this evidence suggests that nIPC morphology may change dynamically as they proliferate.

Contributions

Chapter One is work published as a manuscript in Nature with myself, Michael Song, and Xiaoyu Yang as co-first authors. Authors of this published manuscript also include Armen Abnousi, Changxu Fan, Jia Wen, Jonathan D. Rosen, Mayank NK Choudhary, Xiekui Cui, Ian R. Jones, Seth Bergenholtz, Ugomma C. Eze, Ivan Juric, Bingkun Li, Lenka Maliskova, Jerry Lee, Weifang Liu, Alex A. Pollen, Yun Li, Ting Wang, Ming Hu, Arnold R. Kriegstein, and Yin Shen. I processed and cultured all the human samples, developed and performed the flow cytometry, and was either involved in or leading much of the downstream analysis and figure generation. I identified enhancers for validation and conducted much of the microscopy needed for validation. I was also heavily involved in bioinformatic pipeline development, reviewing results to ensure the pipeline was thresholding enhancers appropriately. Chapter two is in review at PNAS with me as a first author. Other authors include Jayden Ross, Madeline Andrews, Aparna Bhaduri, and Arnold Kriegstein.

Chapter 1:

Cell type-specific 3D epigenomes in the developing human cortex

Abstract

Lineage-specific epigenomic changes during human corticogenesis have remained elusive due to challenges with sample availability and tissue heterogeneity. For example, previous studies used single-cell RNA sequencing to identify at least nine major cell types and up to 26 distinct subtypes in the dorsal cortex alone (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent and Kriegstein, 2017). Here, we characterize cell type-specific cis-regulatory chromatin interactions, open chromatin peaks, and transcriptomes for radial glia, intermediate progenitor cells, excitatory neurons, and interneurons isolated from mid-gestational human cortex samples. We show that chromatin interactions underlie multiple aspects of gene regulation, with transposable elements and disease-associated variants enriched at distal interacting regions in a cell type-specific manner. In addition, promoters with significantly increased levels of chromatin interactivity, termed super interactive promoters, are enriched for lineage-specific genes, suggesting that interactions at these loci contribute to the fine-tuning of transcription. Finally, we develop CRISPRview, a novel technique integrating immunostaining, CRISPRi, RNAscope, and image analysis

for validating cell type-specific cis-regulatory elements in heterogeneous populations of primary cells. Our study presents the first cell type-specific characterization of 3D epigenomes in the developing human cortex, advancing our understanding of gene regulation and lineage specification during this critical developmental window.

Introduction

The human cortex is a complex, heterogeneous structure that undergoes extensive expansion during development, a process that is markedly different and features distinct cell types from mouse cortical development. Much of its diversity arises from cortical stem cells known as radial glia (RG), which give rise to intermediate progenitor cells (IPCs) and excitatory neurons (eNs) that undergo radial migration until they reach the cortical plate (CP) (Pontious *et al.*, 2008; Hansen *et al.*, 2010). Meanwhile, interneurons (iNs) migrate tangentially into the dorsal cortex through the marginal and germinal zones (GZ) (Anderson *et al.*, 1999). These processes result in a CP composed primarily of eNs and iNs, and a GZ where all four cell types are intermixed. Dynamic changes in the epigenomic landscape have been shown to play a critical role in development and cell fate commitment, for instance through the rewiring of physical chromatin loops between promoters and distal regulatory elements including enhancers (Zheng and Xie, 2019). These interactions are of particular interest as their dysregulation has been linked to alterations in gene expression and complex disorders and traits (Li, Hu and Shen, 2018; Schoenfelder and Fraser, 2019) Although previous studies have investigated bulk tissues including the CP and GZ (Won and others, 2016), detailed characterizations are still missing for specific cell types in the developing human cortex. Here, we describe a novel

approach for isolating RG, IPCs, eNs, and iNs from mid-gestational human cortex samples, enabling a comparison of their 3D epigenomes. Furthermore, we develop CRISPRview, a sensitive technique for validating cell type-specific distal regulatory elements in single cells. Our results identify key mechanisms underlying gene regulation and lineage specification during human corticogenesis, providing a framework for the understanding of diverse processes in development and disease.

Results

Sorting specific cell types from the developing human cortex

To isolate RG, IPCs, eNs, and iNs from mid-gestational human cortex samples between gestational weeks (GW) 15 to 22 (**Supplementary Table 1.1**), we expanded upon an established approach for isolating RG from human cortical samples using fluorescence-activated cell sorting (FACS) (Thomsen *et al.*, 2015). Microdissected GZ and CP samples were dissociated, stained using antibodies for EOMES, SOX2, PAX6, and SATB2, and partitioned into their constituent populations using FACS (**Fig. 1.1a; Extended Data Fig. 1.1**). IPCs were isolated as the EOMES+ population, while eNs were isolated from the EOMES- and SOX2- population based on high SATB2 expression, which marks both newborn and mature eNs at the samples' ages (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent and Kriegstein, 2017). RG were isolated based on high SOX2 and high PAX6 expression, and iNs were isolated based on medium SOX2 and low PAX6 expression. The gene expression profiles of the sorted cell populations were both highly consistent with cellular identity and reproducible between individuals

(**Fig. 1.1b; Extended Data Fig. 1.2a, b**).

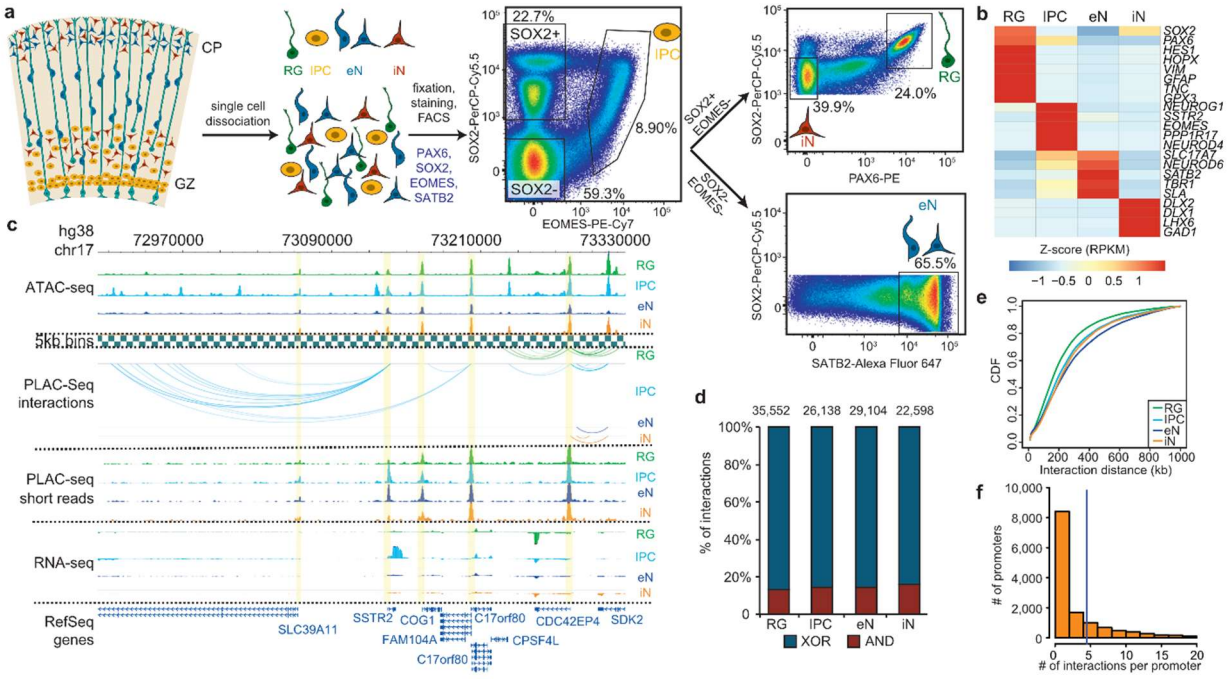


Figure 1.1. Experimental design and features of 3D epigenomes during human corticogenesis. (**a**) Schematic of the sorting strategy. Microdissected GZ and CP samples were dissociated into single cells prior to being fixed, stained with antibodies for PAX6, SOX2, EOMES, and SATB2, and sorted using FACS. (**b**) Heatmap displaying the expression of key marker genes for each cell type. (**c**) WashU Epigenome Browser snapshot displaying a region (chr17: 72,970,000-73,330,000) with interactions linked to *SSTR2* expression in IPCs. (**d**) Bar graph of interaction counts for each cell type, with the proportions of anchor to anchor (red) and anchor to non-anchor (blue) interactions highlighted. (**e**) Cumulative distribution function (CDF) plots of interaction distances for each cell type. (**f**) Histogram displaying the numbers of interactions for interacting promoters across all cell types.

Characterizing cell type-specific 3D epigenomes

We used H3K4me3 proximity ligation-assisted ChIP-seq (PLAC-seq) (Fang and others, 2016) to identify chromatin interactions at active promoters and assay for transposase-accessible chromatin using sequencing (ATAC-seq) to profile open chromatin peaks for the sorted cell populations (**Fig. 1.1c; Supplementary Table 1.2**). After confirming that

the samples cluster by cellular identity (**Extended Data Fig. 1.2c, d**), we applied the Model-based Analysis of PLAC-seq (MAPS) pipeline (Juric and others, 2019) to call significant H3K4me3-mediated chromatin interactions at a resolution of 5 kb. We identified 35,552, 26,138, 29,104 and 22,598 interactions in RG, IPCs, eNs, and iNs, respectively, with approximately 85% of the interactions classified as anchor to non-anchor, and the remaining interactions classified as anchor to anchor (**Fig. 1.1d**; **Extended Data Fig. 1.3a, b**). The median interaction distance was between 170 kb to 230 kb (**Fig. 1.1e**), with an average of 4-5 interactions per promoter (**Fig. 1.1f**), and the majority of interactions occurred within topologically associated domains (TADs) in the GZ or CP (Won and others, 2016) (**Extended Data Fig. 1.3c**).

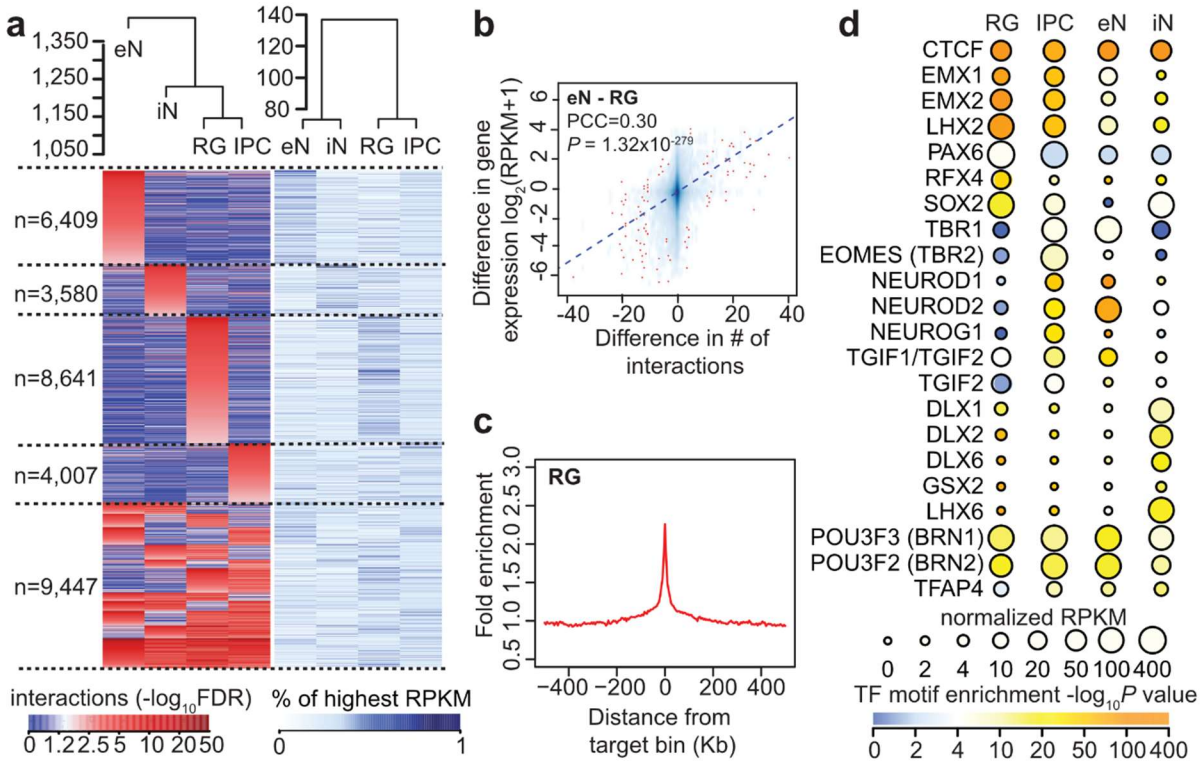


Figure 1.2. H3K4me3-mediated chromatin interactions influence cell type-specific transcription. (a) Heatmaps showing interaction strengths (left) and gene expression (right) for anchor to non-anchor interactions grouped according to their cell type specificity. Interaction strengths are based on the -log$_{10}$FDR from the MAPS pipeline. (b) Scatterplot showing the correlation between the difference in the number of interactions for each promoter and the difference in the expression of the corresponding genes for RG and eNs (Pearson product-moment correlation coefficient, two-tailed, $P = 1.32*10^{-279}$, n = 13,996 anchor bins with promoters). The trendline from linear regression is shown. (c) Fold enrichment of open chromatin peaks over distance-matched background regions in 1 Mb windows around distal interacting regions for RG. (d) TF motif enrichment analysis for open chromatin peaks at cell type-specific distal interacting regions in each cell type. We analyzed 4,203, 1,412, 3,088, and 949 regions in RG, IPCs, eNs, and iNs, respectively. Colors represent enrichment scores based on the p-value from HOMER, while sizes represent the gene expression of the corresponding TFs.

Chromatin interactions influence cell type-specific transcription

Since H3K4me3 is a mark associated with active promoters, we characterized the extent to which H3K4me3-mediated chromatin interactions influence cell type-specific

transcription. First, the sorted cell populations cluster by developmental age based on their interaction strengths across all interacting loci (**Fig. 1.2a**). This is consistent with iNs at this age possessing progenitor-like characteristics including high *SOX2* expression. Meanwhile, genes participating in cell type-specific interactions are enriched for biological processes linked to their respective cell types, including cell proliferation for RG and IPCs, neuron projection development for IPCs and eNs, and synaptogenesis for eNs (**Extended Data Fig. 1.4a; Supplementary Table 1.3**). Interaction strength and gene expression are positively correlated across all cell types (**Fig. 1.2a, b; Extended Data Fig. 1.4b**), suggesting that chromatin interactions orchestrate transcription in a manner that is distinctly cell type-specific. Next, we leveraged the enrichment of open chromatin peaks at distal interacting regions (**Fig. 1.2c; Extended Data Fig. 1.4c**) and performed transcription factor (TF) motif enrichment analysis for distal interacting regions in each cell type(Heinz and others, 2010) (**Fig. 1.2d; Supplementary Table 1.4**). The motifs for PAX6, EOMES, and TBR1 are enriched in RG, IPCs, and eNs, respectively, recapitulating their sequence of expression along this developmental trajectory (Englund and others, 2005). Meanwhile, the motifs for DLX1, DLX2, DLX6, GSX2, and LHX6 are enriched in iNs, in accordance with their roles in iN maturation and function (Lim *et al.*, 2018). Finally, we detect motifs that are enriched in distal interacting regions for co-expression modules in the developing human cortex (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent and Kriegstein, 2017) (**Supplementary Table 1.5**). Our results identify key lineage-specific TFs while linking them to their interacting genes, enabling novel insights into gene regulatory networks during human corticogenesis.

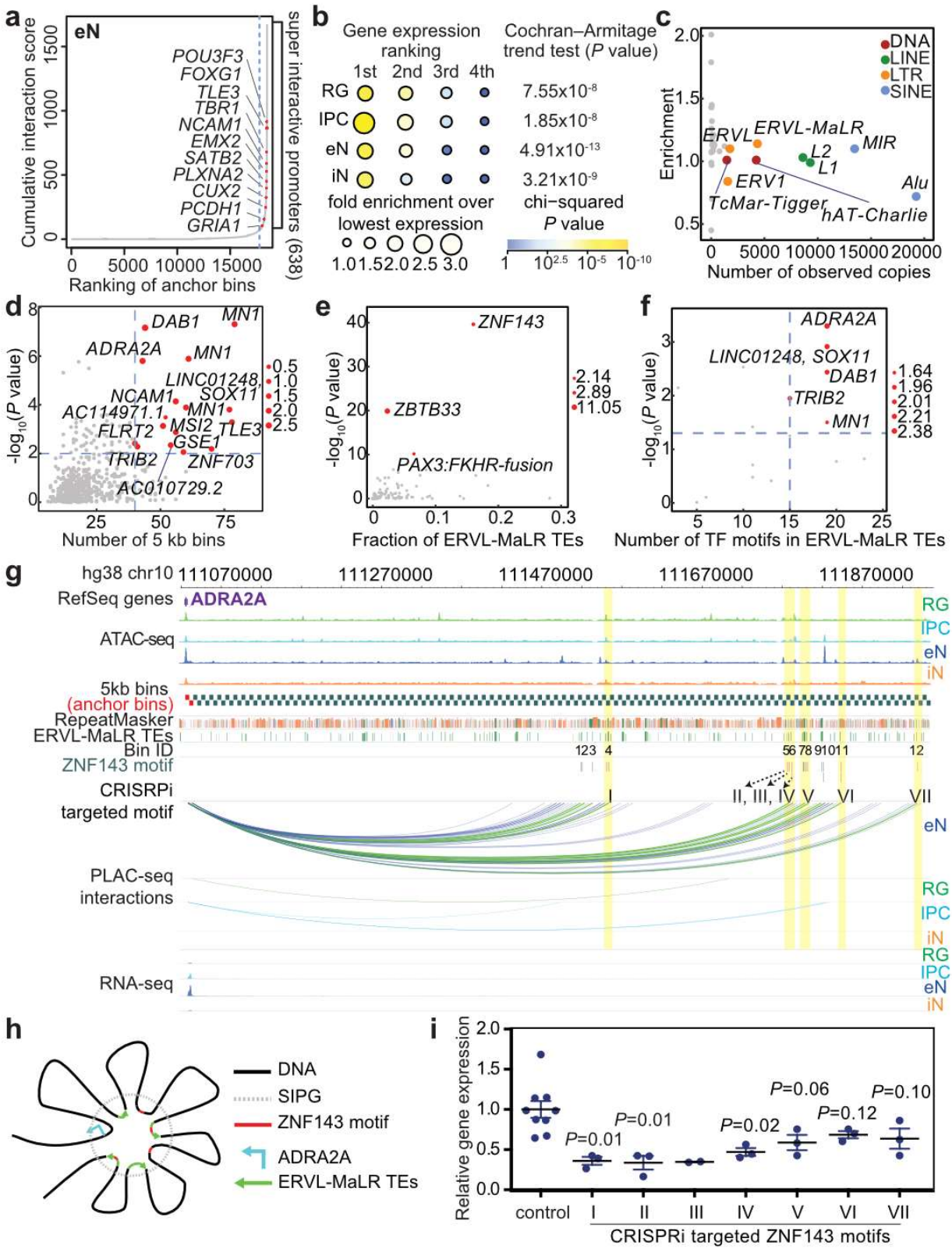

a
eN
Cumulative interaction score
super interactive promoters (638)
POU3F3
FOXG1
TLE3
TBR1
NCAM1
EMX2
SATB2
PLXNA2
CUX2
PCDH1
GRIA1
Ranking of anchor bins
b
Gene expression ranking
1st 2nd 3rd 4th
RG
IPC
eN
iN
Cochran–Armitage trend test (P value)
7.55x10-8
1.85x10-8
4.91x10-13
3.21x10-9
fold enrichment over lowest expression
chi−squared P value
c
DNA
LINE
LTR
SINE
Enrichment
ERVL
ERVL-MaLR
MIR
L2
L1
ERV1
Alu
TcMar-Tigger
hAT-Charlie
Number of observed copies
d
DAB1
MN1
ADRA2A
LINC01248
NCAM1
SOX11
AC114971.1
FLRT2
MSI2
TLE3
GSE1
TRIB2
ZNF703
AC010729.2
-log10(P value)
Number of 5 kb bins
e
ZNF143
ZBTB33
PAX3:FKHR-fusion
2.14
2.89
11.05
Fraction of ERVL-MaLR TEs
f
ADRA2A
LINC01248, SOX11
DAB1
TRIB2
MN1
1.64
1.96
2.01
2.21
2.38
Number of TF motifs in ERVL-MaLR TEs
g
hg38 chr10
111070000
111270000
111470000
111670000
111870000
RefSeq genes
ADRA2A
ATAC-seq
RG
IPC
eN
iN
5kb bins
(anchor bins)
RepeatMasker
ERVL-MaLR TEs
Bin ID
ZNF143 motif
CRISRPi targeted motif
I
II, III, IV
V
VI
VII
PLAC-seq interactions
RNA-seq
h
DNA
SIPG
ZNF143 motif
ADRA2A
ERVL-MaLR TEs
i
Relative gene expression
P=0.01
P=0.01
P=0.02
P=0.06
P=0.12
P=0.10
control
CRISPRi targeted ZNF143 motifs

Figure 1.3. Super interactive promoters are enriched for lineage-specific genes. Anchor bins were ranked according to their cumulative interaction scores in eNs. Super interactive promoters (SIPs) are located past the point in each curve where the slope is equal to 1. (b) The number of SIPs was divided by the total number of anchor bins (both SIPs and non-SIPs) associated with genes with the 1st, 2nd, 3rd, and 4th highest expression among all four cell types (n = 13,996 anchor bins with promoters). Fold enrichment was calculated relative to the group with the lowest expression among all four cell types. (c) Scatterplot showing the enrichment and numbers of observed copies for TE families in SIPGs for eNs. TE families occupying more than 1% of the genome are colored. (d) Scatterplot showing the enrichment and numbers of distal interacting regions for ERVL-MaLR TEs in SIPGs for eNs (n = 638 SIPGs). The 16 SIPGs with significant enrichment (hypergeometric test, one-tailed, $P < 0.01$) and 40 or more distal interacting regions are highlighted. (e) Scatterplot showing the enrichment of TF motifs in ERVL-MaLR TEs for the 16 SIPGs highlighted in (d). Enrichment *P* values are from HOMER. (f) Scatterplot showing the enrichment of ZNF143 motifs in ERVL-MaLR TEs for the 16 SIPGs highlighted in (d) (Poisson distribution, see methods). (g) Interactions between the *ADRA2A* promoter and 12 distal interacting regions containing ERVL-MaLR TE-localized ZNF143 motifs. (h) Proposed mechanism for the contribution of TEs to SIP formation. (i) *ADRA2A* expression was significantly downregulated for 3 of 7 regions relative to control sgRNAs (two-sample t-test, two-tailed, $P < 0.05$, n = 3 for all regions except region III, which has n = 2). Means are indicated and error bars represent the SEM.

Super interactive promoters are enriched for lineage-specific genes

The number of chromatin interactions at H3K4me3-mediated anchor bins is only modestly correlated with gene expression (**Extended Data Fig. 1.5a**). One potential explanation is that individual genes are expressed to varying degrees in the contexts of their diverse cellular functions, and a subset of regulatory elements may be better described as fine-tuning rather than independently inducing or silencing the expression of their interacting genes. Multiple regulatory interactions can also exert synergistic or nonlinear effects on transcription. We first demonstrate that cell-type specific genes tend to harbor more chromatin interactions than shared genes across all four cell types (**Extended Data Fig. 1.5b)**. Next, by ranking anchor bins according to their cumulative interaction scores, we delineate a subset of promoters with significantly increased levels of chromatin

interactivity, termed super interactive promoters (SIPs) (**Fig. 1.3a; Extended Data Fig. 1.5c**). We identify 755, 765, 638, and 663 SIPs in RG, IPCs, eNs, and iNs, respectively (**Extended Data Fig. 1.5d; Supplementary Table 1.6**). SIPs are enriched for key lineage-specific genes including *GFAP* and *HES1* for RG, *EOMES* for IPCs, *SATB2* for eNs, and *DLX5*, *DLX6*, *GAD1*, *GAD2*, and *LHX6* for iNs. We also observe forebrain-specific SIPs including *FOXG1* in all four cell types, progenitor-specific SIPs including *SOX2* in RG, IPCs, and iNs, and cortical neuron-specific SIPs including *TBR1* in IPCs and eNs. Numerous promoters for lincRNAs including *LINC00461* and *LINC01551* are annotated as SIPs, consistent with their expression in the developing human cortex (Liu and others, 2016). In general, SIPs are enriched in cell types with the highest expression of their linked genes, supporting their putative roles in lineage specification (**Fig. 1.3b**). Moreover, super-enhancers and DNA methylation valleys (DMVs) (Luo and others, 2016), two epigenetic features associated with developmental genes and cellular identity, both exhibit enrichment at SIPs (**Extended Data Fig. 1.5e, f**). Finally, SIPs based on promoter capture Hi-C data in neutrophils, naive CD4+ T cells, monocytes, megakaryocytes, and erythroblasts (Javierre and others, 2016) are analogously enriched for cell type-specific over shared genes (**Extended Data Fig. 1.5g**), implying that SIPs present a generalized mechanism for maintaining the expression of key genes underlying cellular identity and function.

Transposable elements in SIP formation

To explore potential mechanisms underlying SIP formation, we evaluated the contributions of transposable elements (TEs), which have been shown to influence 3D

chromatin architecture and propagate regulatory elements throughout the genome (Sundaram and Wang, 2018; Zhang and others, 2019; Choudhary and others, 2020) We analyzed the enrichment of TEs at the class, family, and subfamily levels in sequences defined by SIPs and their distal interacting regions, termed super interactive promoter groups (SIPGs) (**Fig. 1.3c; Extended Data Fig. 1.6a-c**). We first observe that ERVL-MaLR TEs are enriched in SIPGs across all four cell types. Next, we identify 16 SIPGs in eNs that exhibit significant enrichment for ERVL-MaLR TEs and have 40 or more distal interacting regions (hypergeometric test, one-tailed, $P < 0.01$) (**Fig. 1.3d**). TF motif enrichment analysis for ERVL-MaLR TEs in these SIPGs reveals the most enriched motif to be that for ZNF143, an architectural protein that has been demonstrated to mediate physical chromatin looping between promoters and distal regulatory elements (Bailey and others, 2015) (**Fig. 1.3e**). Moreover, ERVL-MaLR TE subfamilies have been linked to ZNF143 binding in 3T3 and HeLa cells (Ngondo-Mbongo *et al.*, 2013). We find that ZNF143 motifs are broadly enriched in ERVL-MaLR TEs, SIPGs, and ERVL-MaLR TEs in SIPGs (**Extended Data Fig. 1.6d-f**). The *ADRA2A* SIPG in particular exhibits the strongest enrichment of ERVL-MaLR TE-localized ZNF143 motifs (hypergeometric test, one-tailed, $P = 1.59 \times 10^{-6}$) (**Fig. 1.3f**) and is linked to elevated *ADRA2A* expression in eNs (**Extended Data Fig. 1.6g**). The *ADRA2A* SIPG spans 42 distal interacting regions, 25 of which contain ERVL-MaLR TEs, and 12 of which contain ERVL-MaLR TE-localized ZNF143 motifs (**Fig. 1.3g; Extended Data Fig. 1.6h**). Many of these ZNF143 motifs can be mapped back to the consensus sequences of their corresponding ERVL-MaLR TE subfamilies (**Extended Data Fig. 1.6i, j**). This supports a model in which ZNF143 motifs are coordinately expanded by ERVL-MaLR TE insertion, promoting increased binding site

redundancy and strengthened assembly of the *ADRA2A* regulatory unit (**Fig. 1.3h**). Using CRISPRi to target ERVL-MaLR TE-localized ZNF143 motifs in the *ADRA2A* SIPG resulted in the significant downregulation of *ADRA2A* expression for 3 of 7 regions in eNs (two-sample t-test, two-tailed, $P < 0.05$) (**Fig. 1.3i**), implying that TEs are capable of mediating the formation of higher order chromatin features including SIPs (Sundaram and Wang, 2018).

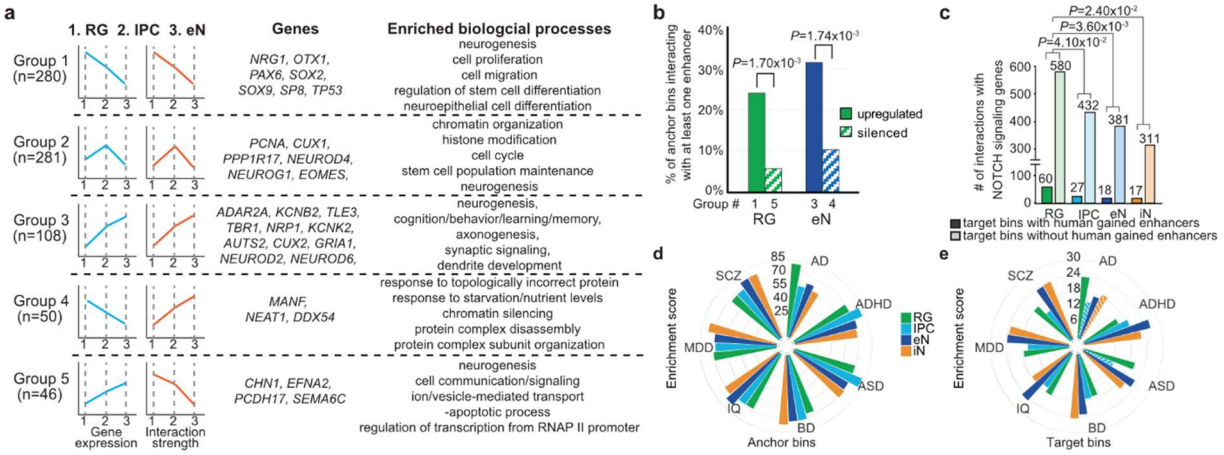


Figure 1.4. Features of cortical development and partitioning SNP heritability for complex disorders and traits. (a) Genes categorized based on their gene expression and chromatin interactivity from RG to eNs. Groups 1-5 represent RG-upregulated, IPC-upregulated, eN-upregulated, eN-silenced, and RG-silenced genes, respectively. Representative genes and biological processes are shown for each group. (**b**) Groups 1 (75 of 312 bins) and 3 (40 of 127 bins) are enriched for interactions with enhancers relative to groups 4 (6 of 58 bins) and 5 (3 of 52 bins) (chi-squared test, two-tailed). Only bins with at least one interaction were considered. (**c**) Bar graph of interaction counts from Notch signaling genes to regions with and without HGEs in each cell type (chi-squared test, two-tailed). We observed 2,541, 1,854, 1,869, and 1,610 interactions with HGEs in RG, IPCs, eNs, and iNs, respectively. (**d-e**) LDSC enrichment scores for each disease and cell type, stratified by PLAC-seq anchor and target bins. Non-significant enrichment scores are shown as striped bars.

Developmental trajectories from RG to eNs Since RG, IPCs, and eNs represent a developmental trajectory from dorsal cortical progenitors to mature functional neurons, we grouped genes based on their gene expression and chromatin interactivity along this axis and identified genes linked to cell type-specific processes in RG, IPCs, and eNs (groups 1-3) (**Fig. 1.4a; Extended Data Fig. 1.7a; Supplementary Table 1.7**). We similarly identified genes with anticorrelated gene expression and chromatin interactivity from RG to eNs (groups 4-5), which represent eN-silenced and RG-silenced genes, respectively. eN-silenced genes are enriched for biological processes linked to chromatin remodeling and epigenetic regulation, while RG-silenced genes are enriched for eN-specific signatures. Furthermore, genes in these two groups are depleted for interactions with enhancers annotated using ChromHMM in the germinal matrix (Davis and others, 2018) while exhibiting enrichment for interactions with TFs containing domains associated with transcriptional repression (**Fig. 1.4b; Extended Data Fig. 1.7b; Supplementary Table 1.8**). Our results demonstrate that cell type-specific 3D epigenomes are capable of identifying distinct modes of epigenetic regulation during development. *Human-specific aspects of cortical development* Human corticogenesis is dramatically distinct from other mammals, driven largely by the increased diversity and proliferative capacity of cortical progenitors which contribute to the enhanced size and complexity of the human brain (Miller *et al.*, 2019). Notch signaling genes in particular have been implicated in the clonal expansion of RG, which represent the major subtype of cortical progenitors in the developing human cortex (Rani and others, 2016; Suzuki and others, 2018). Here, RG are enriched relative to other cell types for interactions involving Notch signaling genes(Carbon and others, 2009) (**Fig. 1.4c**). Furthermore,

compared to other cell types, interactions in RG target a significantly higher proportion of human-gained enhancers (HGEs) identified through comparative analyses of human, rhesus macaque, and mouse brains (Reilly and others, 2015). This suggests that epigenetic modifications surrounding Notch signaling genes in RG contribute to significant neurological differences between humans and other species. Additional biological processes exhibiting enrichment for interactions with HGEs include forebrain neuron fate commitment in RG, neuroblast proliferation in IPCs, forebrain neuron development in eNs, and GABAergic interneuron development in iNs (**Supplementary Table 1.9)**. We provide detailed annotations of genes interacting with HGEs and in vivo-validated enhancer elements (Visel *et al.*, 2007) in **Supplementary Table 1.10**.

Partitioning SNP heritability for complex disorders and traits

Chromatin interactions present a unique resource for mapping complex disorder- and trait-associated variants to their target genes (**Extended Data Fig. 1.7c, d; Supplementary Table 1.11**). This is important as regulatory variants can modulate transcription through the formation or disruption of physical chromatin loops. For instance, both fetal (Walker and others, 2019) and adult (Hoffman and others, 2019) brain eQTLs are enriched at chromatin interactions across all four cell types (**Extended Data Fig. 1.8a-c**). Chromatin interactions additionally enable the assessment of cell-type specific patterns of SNP heritability enrichment. Therefore, we leveraged linkage disequilibrium score regression (LDSC) (Bulik-Sullivan and others, 2015; Finucane and others, 2015) using our 3d epigenomes to partition SNP heritability for seven complex neuropsychiatric disorders and traits: Alzheimer's disease (AD), attention deficit hyperactivity disorder

(ADHD)(Demontis and others, 2019), autism spectrum disorder (ASD) (Phane Ory *et al.*, 2003), bipolar disorder (BD)(Kelemenova *et al.*, 2010), intelligence quotient (IQ) (Savage and others, 2018), major depressive disorder (MDD) (Howard and others, 2019), and schizophrenia (SCZ) (Pardinas and others, 2018). First, conditioned on a baseline model (Gazal and others, 2017), PLAC-seq anchor and target bins exhibit significant enrichment for all of the disorders and traits we tested, except for AD and ASD (**Extended Data Fig. 1.8d**). Anchor and target bins are also more informative relative to distal open chromatin peaks and cell-type specific genes (**Extended Data Fig. 1.8e, f**). This can be attributed to the utility of chromatin interactions for linking genes to functional regulatory sequences over long genomic distances. Next, we utilized a joint model incorporating all four cell types to investigate cell-type specific patterns of SNP heritability enrichment (**Fig. 1.4d, e**). First, target bins exhibit significantly more variability than anchor bins in terms of their enrichment scores, reflecting the increased cell-type specificity of distal regulatory elements compared to promoters. Furthermore, eNs and iNs present higher enrichment scores at target bins relative to RG and IPCs, suggesting the increased relevance of neuronal cell types for the disorders and traits in our study. We used H-MAGMA (Yang *et al.*, 1998) to identify enriched biological processes for genes interacting with non-coding variants for each disease and cell type (**Extended Data Fig. 1.9; Supplementary Table 1.12**). Our results recapitulate the roles of lipoprotein metabolism and transport in AD pathophysiology (Andersen and Willnow, 2006), mechanisms our results imply are RG-specific. Meanwhile, IPCs and eNs are enriched across all diseases for interactions between SNPs and genes linked to neural precursor cell proliferation, axon guidance, and axonogenesis. Finally, our results for SCZ align with extensive evidence that the

disruption of chromatin and epigenetic regulators is a significant contributor to disease risk (Akbarian, 2014; Won and others, 2016).

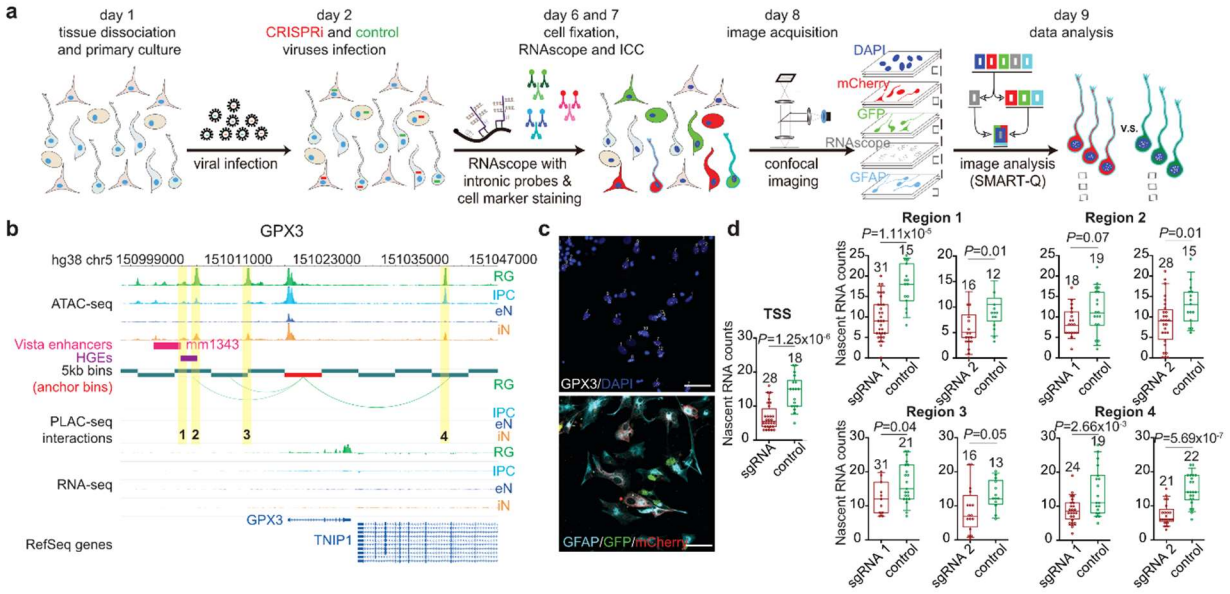


Figure 1.5. Validation of cell-type specific distal regulatory elements using CRISPRview. (**a**) Schematic of the CRISPRview workflow. Image analysis was performed using the SMART-Q pipeline. (**b**) Interactions between the *GPX3* promoter and distal interacting regions containing open chromatin peaks that were targeted for silencing are highlighted. Notably, region 1 overlaps both an HGE and Vista enhancer element (mm1343), supporting its function as a putative enhancer. (**c**) Representative images show staining for intronic RNAscope probes (white), DAPI (blue), GFAP (light blue), GFP (green), and mCherry (red). The scale bar is 50 μm. (**d**) Box plots show results for experimental (red) and control (green) sgRNA-treated cells for each region (two-sample t-test, two-tailed). The median, upper and lower quartiles, and 10% to 90% range are indicated. Open circles represent single cells. Sample sizes are indicated above each box plot.

Characterizing distal interacting regions in primary cells

Validating distal regulatory elements in primary cells has proved challenging in the past, with most experiments performed using cell lines or iPSC-derived cells. A major obstacle lies in the robust detection of transcriptional changes resulting from epigenetic perturbations in complex, heterogeneous samples. Therefore, we developed

CRISPRview to validate cell-type specific distal regulatory elements in single cells (**Fig. 1.5a**). Specifically, primary cultures of GZ or CP samples are first infected with lentivirus expressing mCherry, dCas9-KRAB, and sgRNAs targeting open chromatin peaks interacting with a gene of interest, in addition to lentivirus expressing GFP, dCas9-KRAB, and sgRNAs targeting non-human sequences. Next, the cells are fixed and stained using antibodies for mCherry, GFP, cell-type specific markers, DAPI, and intronic RNAscope probes targeting the gene of interest. Finally, we leverage SMART-Q (Yang and others, 2020) to compare the number of nascent RNA transcripts between experimental and control sgRNA-treated cells. We validated four regions interacting with the *GPX3* promoter, all of which exhibited significant downregulation in terms of *GPX3* expression upon silencing (**Fig. 1.5b-d**). Meanwhile, silencing three regions interacting with the *IDH1* promoter in RG and eNs resulted in the significant downregulation of *IDH1* expression in the respective cell types (**Extended Data Fig. 1.10a, b**). Finally, we characterized two additional RG-specific loci in *TNC* and *HES1*, both of which are annotated as SIPs (**Extended Data Fig. 1.10c-h**). The observation of small but significant changes in gene expression supports the hypothesis that multiple interactions frequently work in concert to titrate the expression of key genes underlying cellular identity and function.

Discussion

Recent publications leveraging single-cell RNA sequencing have highlighted the heterogeneity of the developing human cortex, with RG, IPCs, eNs, iNs, microglia, endothelial cells, and subplate neurons present in the dorsal cortex alone. Despite significant differences in lineage and maturation state, many of these cell types share

intriguing similarities in their transcriptional landscapes. For example, iNs express genes for TFs that are typically associated with RG proliferation, including SOX2, as well as with eN differentiation, including ASCL1 and NPAS3[1]. By isolating specific cell types in the developing human cortex, we are able to distinguish nuanced regulatory programs driving cell-type specific differences during human corticogenesis. We identify SIPs which are enriched for key lineage-specific genes and represent distinct chromatin features from A/B compartments(Lieberman-Aiden and others, 2009), TADs(Dixon and others, 2012), frequently interacting regions (FIREs)(Schmitt and others, 2016), and highly interacting regions (HIRs) (Nagashima *et al.*, 2015). Furthermore, we uncover a mechanism in which TEs propagate binding sites for architectural proteins such as ZNF143, facilitating the formation of multi-interaction clusters that function to sustain transcription. Lastly, by developing CRISPRview, we achieve several emergent advantages for validating distal regulatory elements in primary cells. First, we are able to focus our analysis on specific cell types, circumventing averaging effects associated with bulk measurements in complex samples. Next, we are able to directly compare experimental and control sgRNA-infected cells within the same population. Finally, we achieve enhanced sensitivity and statistical power based on the detection of nascent RNA transcripts in single cells. Future experiments leveraging CRISPRview in live tissue cultures should continue to reveal regulatory relationships in a manner that is truly representative of the complex in vivo environment during human corticogenesis.

Acknowledgements

This work was supported by the UCSF Weill Institute for Neuroscience Innovation Award (to Y.S. and A.R.K.), the National Institutes of Health (NIH) grants R01AG057497, R01EY027789, and UM1HG009402 (to Y.S.) and R35NS097305 (to A.R.K.), the Hillblom Foundation, and the American Federation for Aging Research New Investigator Award in Alzheimer's Disease (to Y.S). This work was also supported by the NIH grants R01HL129132, U544HD079124, and R01MH106611 (to Y.L.), R01HG007175, U24ES026699, and U01HG009391 (to T.W.), and the American Cancer Society grant RSG-14-049-01-DMC (to T.W). M.S. is supported by T32GM007175. M.P. is supported by the National Science Foundation Graduate Research Fellowship Program grant 1650113. U.C.E. is supported by 5T32GM007618-42. This work was made possible in part by the NIH grants P30EY002162 to the UCSF Core Grant for Vision Research, P30DK063720, and S101S10OD021822-01 to the UCSF Parnassus Flow Cytometry Core.

Author contributions

Y.S., A.A.P., and A.R.K. conceived the study. Y.S., M.H., A.A.P., and A.R.K. supervised the study. M.S., M.P., X.Y., I.R.J., X.C., U.C.E., L.M., and J.L. performed experiments. M.S., A.A., S.B., J.D.R., B.L., I.R.J., and M.H. performed computational analysis. C.F. and M.N.C. performed TE element analysis under the supervision of T.W. J.W. and W.L. performed SNP heritability enrichment analysis under the supervision of Y.L. M.S., M.P., X.Y., and Y.S. analyzed and interpreted the data. M.S., M.P., X.Y., M.H, and Y.S. prepared the manuscript with input from all other authors.

Data availability statement

All datasets used in this study (PLAC-seq, ATAC-seq, RNA-seq) are available at the Neuroscience Multi-Omic Archive (NeMO Archive) under controlled access. Chromatin interactions, open chromatin peaks, and gene expression profiles for each cell type can be downloaded from the NeMO Archive using the following link: https://assets.nemoarchive.org/dat-uioqy8b

Cell type-specific 3D epigenomes can be visualized on the WashU Epigenome Browser using the datahub at the following link: http://epigenomegateway.wustl.edu/browser/?genome=hg38&position=chr17:72918238-73349675&hub=https://shen-msong.s3-us-west-1.amazonaws.com/hfb_submission/hfb_datahub.json

Methods

Ethics statement

Deidentified tissue samples were collected with prior informed consent in strict observance of legal and institutional ethical regulations. All protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee (GESCR) and Institutional Review Board (IRB) at the University of California, San Francisco.

Tissue dissociation

The tissue dissociation protocol was adapted from Nowakowski et al, 2017. Briefly, samples were first cut into small pieces in artificial cerebrospinal fluid before being added to pre-warmed papain dissociation media (Worthington #LK003150). The samples were incubated in dissociation media for 45 minutes at 37°C. Next, they were triturated, filtered through a 70 μM nylon mesh, and centrifuged for 8 minutes at 300 g. For individual germinal zone (GZ) and cortical plate (CP) cultures, samples were first cut coronally into thin slices. As previously described, cell density drops dramatically past the outer subventricular zone, enabling the clear identification of the outer filamentous zone and subplate. Samples were dissected along this boundary to separate the GZ from the CP prior to dissociation.

Sample fixation

Mid-gestational human cortex samples between GW15 and GW22 were fixed in 2% paraformaldehyde prepared in PBS with gentle agitation for 10 minutes at room temperature. Glycine was added to a final concentration of 200 mM to quench the reactions, and the samples were centrifuged for 5 minutes at 4°C and 500 g. The samples were washed twice with PBS before being frozen at -80°C for further processing.

Permeabilization and staining

The cell pellet was thawed on ice and resuspended in PBS containing 0.1% Triton-X-100 for 15 minutes. The cells were then washed twice with PBS and resuspended in 5% BSA

in PBS for staining. Staining proceeded for at least one hour with FcR Blocking Reagent (Miltenyi Biotech, 1/20 dilution), EOMES PE-Cy7 (Invitrogen, Cat 25-4877-42, Clone WD1928, Lot 1923396, 1/10 dilution), PAX6 PE (BD Biosciences, Cat 561552, Clone O18-1330, Lot 8187686, 1/10 dilution), SOX2 PerCP-Cy5.5 (BD Biosciences, Cat 561506, Clone O38- 678, Lot 8165744, 1/10 dilution), and SATB2 Alexa Fluor 647 (Abcam, Cat ab196536, Clone EPNCIR130A, Lot GR3208103-I and GR228747-2, 1/100 dilution). After staining, the cells were centrifuged for 5 minutes at 500 g, and the pellet was diluted into PBS. When sorting cells for RNA-seq, 1% RNasin Plus RNase Inhibitor (Promega) was added to all buffers, and acetylated BSA was used to prepare 5% BSA in PBS for staining.

FACS

AbC Total Antibody Compensation Beads (Thermo Fisher) were used to generate single color compensation controls prior to sorting. Sorting was conducted on either the FACSAria II, FACSAria IIu, or FACSAria Fusion instruments using a 70 μM nozzle, and cells were collected in 5 ml tubes pre-coated with FBS. A sample of each sorted cell population was reanalyzed on the same machine to assess purity. Cells were collected by centrifuging for 10 minutes at 500 g, and the cell pellet was frozen at -80°C for further processing. When sorting cells for RNA-seq, cells were collected in 5 ml tubes pre-coated with both FBS and RNAlater (Thermo Fisher).

Primary cell culture

Following dissociation, cells were plated onto Matrigel-coated coverslips in 48 well plates or chamber slides at a density of approximately $0.7x10^6$ cells per well. All cell culture was handled in sterile conditions. The cells were infected with lentivirus the day after plating, and media was changed every two days. Media was composed of 96% DMEM/F-12 with GlutaMAX, 1% N-2, 1% B-27, and 1% penicillin/streptomycin. The cells were grown in 8% oxygen and 5% carbon dioxide and harvested four days post-infection for CRISPRview. For qPCR at the *ADRA2A* locus, the cells were harvested six days post-infection.

PLAC-seq

PLAC-seq was performed according to Fang et al., 2016. 1 to 5 million cells were used to prepare each library. Digestion was performed using 100 U MboI for 2 hours at 37°C, and chromatin immunoprecipitation was performed using Dynabeads M-280 sheep anti-rabbit IgG (Invitrogen #11203D) superparamagnetic beads bound with 5 µg anti-H3K4me3 antibody (Millipore 04-745). Sequencing adapters were added during PCR amplification. Libraries were sent for paired-end sequencing on the HiSeq X Ten or NovoSeq 6000 instruments (150 bp paired-end reads). fastp was applied to trim reads to 100 bp for all downstream analysis.

MAPS

We used the MAPS pipeline to call significant H3K4me3-mediated chromatin interactions at a resolution of 5 kb based on our PLAC-seq data. First, bwa mem was used to map

raw reads to hg38. Unmapped reads and reads with low mapping quality were discarded, and the resulting read pairs were processed as previously reported (Juric and others, 2019). To define PLAC-seq anchor bins, we took the union of peaks identified by MACS2 using the options "--nolambda --nomodel --extsize 147 --call-summits -B --SPMR" and an FDR cutoff of 0.0001 for all read pairs with interaction distance < 1 kb in each cell type. Next, we classified read pairs as AND, XOR, or NOT interactions based on whether both, one, or neither of the interacting 5 kb bins overlapped anchor bins (**Extended Data Fig. 1.3a**). Since we were specifically interested in identifying long-range H3K4me3-mediated chromatin interactions, we retained only read pairs corresponding to intrachromosomal XOR and AND interactions with interaction distances between 10 kb and 1 Mb. We downsampled the number of read pairs separately for each chromosome to ensure that we started with the same number of read pairs for each cell type.

To call significant interactions, we employed a Poisson regression-based approach to normalize systematic biases from restriction sites, GC content, sequence repetitiveness, and ChIP enrichment. We fitted models separately for AND and XOR interactions and calculated FDRs for interactions based on the expected and observed contact frequencies between interacting 5 kb bins. We grouped interactions whose ends were located within 15 kb of each other into clusters and classified all other interactions as singletons. We defined our significant H3K4me3-mediated chromatin interactions as interactions with 12 or more reads, normalized contact frequency (defined as the ratio between the observed and expected contact frequency) ≥ 2, and FDR < 0.01 for clusters and FDR < 0.0001 for singletons. This was based on the reasoning that biologically

meaningful interactions are more likely to appear in clusters, while singletons are more likely to represent false positives.

Reproducibility analysis

PCA was performed based on the normalized contact frequencies for interacting 5 kb bins from our PLAC-seq data. We first extracted AND and XOR interactions based on cell-type specific anchor bins for each of the 11 replicates. Next, we applied zero-truncated Poisson regression adjusting for the same biases as the MAPS pipeline. We derived normalized contact frequencies based on the ratios between the observed and expected contact frequencies for interacting 5 kb bins, with the expected contact frequencies being the fitted values from the zero-truncated Poisson regression. Normalized contact frequencies were then log-transformed and merged across the 11 replicates. The merged data was used to generate the PCA plots. We restricted our analysis to interacting 5 kb bins in both 300 and 600 kb windows for **Extended Data Fig. 1.2d**.

ATAC-seq

ATAC-seq was performed as previously described using the Nextera DNA Library Prep Kit (Illumina #FC-121-1030). Briefly, fixed cells were washed once with ice-cold PBS containing 1x protease inhibitor before being resuspended in ice-cold nuclei extraction buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Igepal CA630, and 1x protease inhibitor) for 5 minutes. 50,000 cells were aliquoted, exchanged into 50 µL 1x

Buffer TD, and incubated with 2.5 µL TDE1 enzyme for 45 minutes at 37°C with shaking. Following transposition, 150 µL reverse crosslinking solution (50 µL 1 M Tris pH 8.0, 100 µL 10% SDS, 2 µL 0.5 M EDTA, 10 µL 5 M NaCl, 800 µL water, and 2.5 µL 20 mg/mL Proteinase K) was added to each tube and incubated at 65°C overnight. DNA was column purified, PCR amplified, and size-selected for fragments between 300 and 1000 bp. Libraries were sent for paired-end sequencing on the NovaSeq 6000 instrument (150 bp paired-end reads). Raw reads were trimmed to 50 bp, mapped to hg38, and processed using the ENCODE pipeline (https://github.com/kundajelab/atac_dnase_pipelines) running the default settings. The optimal naive overlap peaks for each cell type were used for all downstream analysis.

RNA-seq

We extracted total RNA from the sorted cell populations using the RNAstorm™ FFPE RNA extraction kit (Cell Data Sciences #CD501) starting with $5x10^5$ to $1.5x10^6$ cells. The quality of the extracted RNA was checked by determining the percentage of RNA fragments with size > 200 bp (DV200) from the Agilent 2100 Bioanalyzer. RNA samples with DV200 >= 40% were used for library construction. First, samples were depleted of ribosomal RNA using the KAPA RNA HyperPrep Kit with RiboErase (HMR #KK8560). Next, we performed first and second strand synthesis, dA-tailing, and sequencing adapter ligation. cDNA was cleaned up and sequencing adapters were added via PCR amplification. Libraries were sent for paired-end sequencing on the NovaSeq 6000 instrument (150 bp paired-end reads). Raw reads were trimmed using Trim Galore and aligned to hg38 using STAR running the standard ENCODE parameters, and transcript

quantification was performed in a strand-specific manner using RSEM with the GENCODE 29 annotation. The edgeR package in R was used to calculate TMM-normalized RPKM values for each gene, and the mean values across all replicates were used for all downstream analysis.

GO enrichment analysis

Protein coding and non-coding RNA genes participating in cell-type specific XOR interactions were used for GO enrichment analysis. Only interactions with open chromatin peaks overlapping promoters (defined as the 1 kb region centered around a gene's TSS) in their anchor bins and distal open chromatin peaks (defined as open chromatin peaks not overlapping promoters) in their target bins were used. A minimum RPKM of 0.5 was used to retain only genes that were expressed, and the resulting genes were input into DAVID 6.8 running functional annotation clustering using the "GOTERM_BP_ALL" ontology. Group enrichment scores based on the geometric mean of EASE scores for terms in each group are reported. To report enriched biological processes for genes interacting with non-coding variants for each disease and cell type, we assigned non-coding SNPs for each disorder and trait to genes based on interactions with the 5 kb bins containing their promoters. Next, we ran H-MAGMA using our annotations to generate ranked lists of gene-level association statistics which were used to perform functional enrichment analysis using the gprofiler2 package in R (Ri Reimand *et al.*, 2007): gost(ranked.list, organism="hsapiens", ordered_query=T, significant=F, correction_method="fdr", sources="GO:BP").

TF motif enrichment analysis

We used 200 bp windows centered around open chromatin peaks participating in cell-type specific XOR interactions for TF motif enrichment analysis using HOMER. We used the complete set of vertebrate motifs from the JASPAR database, specifying the "-float" option to adjust the degeneracy threshold, and the entire genome was used as the background. The binomial distribution was used to calculate p-values. For the analysis of co-expression modules in the developing human cortex, we downloaded co-expression modules from Nowakowski et al. 2017. Specifically, we used the "all" network set for all four cell types, as well as network sets matched to individual cell types as follows: "page.rg" for RG, "page.ipc" for IPCs, "page.n" for eNs, and "vage.in" for iNs. This was to capture biological variation both between and within cell types, respectively. We used HOMER to perform TF motif enrichment analysis for the set of open chromatin peaks interacting with promoters of genes assigned to each co-expression module. For ranking TFs according to the number of co-expression modules they were enriched for in each network set and cell type, an FDR threshold of 0.05 was applied.

Super interactive promoters

We used an approach similar to calling super-enhancers(Hnisz *et al.*, 2013) to annotate super interactive promoters (SIPs) in each cell type. For each anchor bin, we calculated the cumulative interaction score, defined as the sum of the $-\log_{10}$FDR for interactions overlapping each anchor bin. We used this metric as it accounts for noise and is directly

associated with the interaction strength in PLAC-seq data. Next, we prepared plots of ranked cumulative interaction scores for anchor bins in each cell type and defined SIPs to be anchor bins located past the point in each curve where the slope is equal to 1.

Cell-type specific versus shared genes

We classified each gene as cell-type specific or shared according to its Shannon entropy score across all four cell types. Specifically, for each gene, we calculated its relative expression value in each cell type, defined as its RPKM in that cell type divided by the sum of its RPKMs across all four cell types. Next, we calculated the Shannon entropy score for each gene based on its relative expression values across all four cell types. We classified a gene as specific for a cell type if met the following conditions: its Shannon entropy score was < 0.01, its RPKM was > 1 in that cell type, and its RPKM in that cell type was the highest across all four cell types. All other genes with RPKM > 1 were classified as shared.

TE enrichment in SIPGs

TE enrichment in SIPGs was evaluated as follows. The foreground enrichment was defined as the number of TEs at the class, family, or subfamily levels overlapping SIPGs in each cell type. The background enrichment was defined as the number of TEs overlapping all interacting 5 kb bins (both SIPGs and non-SIPGs). At least 50% of a TE had to overlap a 5 kb bin for it to be considered overlapping. The overall enrichment was

defined as the foreground enrichment divided by the background enrichment multiplied by the proportion of interacting 5 kb bins that were assigned to SIPGs.

For the enrichment of SIPGs for ERVL-MaLR TEs, the foreground enrichment for each SIPG was defined as the number of distal interacting regions containing one or more ERVL-MaLR TEs for that SIPG. The background enrichment for each SIPG was defined as the number of randomly shuffled distal interacting regions containing one or more ERVL-MaLR TEs for that SIPG. We computed the background enrichment over 100 permutations. The overall enrichment was defined as the foreground enrichment divided by the background enrichment. The significance for each SIPG was calculated using the hypergeometric distribution as follows:

$$P = \frac{\left(\frac{m}{q}\right)\left(\frac{n}{k-q}\right)}{\left(\frac{m+n}{k}\right)}$$

where “q” is the number of distal interacting regions containing one or more ERVL-MaLR TEs for that SIPG, “m” is the number of 5 kb bins containing one or more ERVL-MaLR TEs on the same chromosome, “n” is the number of 5 kb bins containing no ERVL-MaLR TEs on the same chromosome, and “k” is the size of the SIPG.

ZNF143 motif enrichment

For the enrichment of SIPGs for ERVL-MaLR TE-localized ZNF143 motifs, the foreground enrichment for each SIPG was defined as the number of ERVL-MaLR TE-localized ZNF143 motifs in its distal interacting regions. FIMO[53] was used to detect ZNF143 motifs within ERVL-MaLR TEs. The background enrichment was defined as the total number of ZNF143 motifs in the SIPG. The overall enrichment was defined as the foreground enrichment divided by the background enrichment multiplied by the proportion of the SIPG that is occupied by ERVL-MaLR TEs. The significance for each SIPG was calculated using a Poisson distribution where the number of events (k) is the foreground enrichment and the rate parameter (λ) is the background enrichment multiplied by the proportion of the SIPG that is occupied by ERVL-MaLR TEs.

For evaluating the genome-wide enrichment of ZNF143 motifs in ERVL-MaLR and THE1C TEs, we first used FIMO to scan all ERVL-MaLR and THE1C TEs for instances of ZNF143 motifs. As a background, we scanned 100 sets of chromosome- and length-matched, non-overlapping sequences randomly sampled to avoid gaps and blacklisted regions in the human genome. We used a similar approach to evaluate the enrichment of ZNF143 motifs in ERVL-MaLR TEs in SIPGs. For evaluating the enrichment of ZNF143 motifs in SIPGs, we compared the mean numbers of ZNF143 motifs per 5 kb bin for distal interacting regions across all SIPGs to 100 sets of chromosome- and length-matched, non-overlapping sequences randomly sampled to avoid gaps and blacklisted regions in the human genome. For comparing the distributions of the mean numbers of ZNF143 motifs per 5 kb bin for actual versus shuffled SIPGs, we sampled distal interacting regions for each SIPG 100 times on the same chromosome in a non-overlapping manner.

Target gene annotation for enhancers and GWAS SNPs

To determine whether a human-gained enhancer, Vista enhancer element, or GWAS SNP interacted with a gene, we determined whether any of its promoters participated in interactions with the element of interest on the other end. All human-gained enhancers and Vista enhancer elements were expanded to a minimum width of 5 kb, and all GWAS SNPs were expanded to a minimum width of 1 kb to account for potential functional sequences around each element. Furthermore, we determined the proportion of GWAS SNPs interacting with their nearest and more distal genes, except when all the promoters for the nearest gene fell within the same 5 kb bin as the GWAS SNAP and could not be resolved for interactions.

Partitioning SNP heritability for complex disorders and traits.

We leveraged linkage disequilibrium score regression (LDSC) to partition SNP heritability separately for each complex neuropsychiatric disorder and trait based on joint models incorporating PLAC-seq anchor or target bins across all cell types. We also ran LDSC using a baseline model (Gazal and others, 2017) consisting of coding, UTR, promoter, and intron regions, histone marks, DNase I hypersensitive sites, ChromHMM/Segway predictions, regions that are conserved in mammals, super-enhancers, FANTOM5 enhancers, and LD-related annotations (recombination rate, nucleotide diversity CpG content, etc.) that are not specific to any cell type. This informs us whether our epigenomic annotations for a given cell type are informative for SNP heritability enrichment compared

to a comprehensive set of genomic features that has been widely adopted in the field. To compare different epigenomic annotations for each cell type, we used both distal open chromatin peaks and 100 kb windows around the transcription start and end sites of cell-type specific genes according to their Shannon entropy scores and RPKM > 1.

Validating ERVL-MaLR-localized ZNF143 motifs

CRISPRi and qRT-PCR were used to validate ERVL-MaLR TE-localized ZNF143 motifs at distal interacting regions in the *ADRA2A* SIPG. Of the 12 distal interacting regions containing ERVL-MaLR TE-localized ZNF143 motifs, we were able to design sgRNAs to target ZNF143 motifs overlapping open chromatin peaks for 7 of the regions. ZNF143 motifs were extended by 100 bp in both directions for designing sgRNAs. To maximize CRISPRi efficiency, we designed two sgRNAs for each region and cloned them into the dual expression cassette in the CRISPRi vector as described for CREST-seq (Diao and others, 2017). sgRNA sequences were confirmed by Sanger sequencing and packaged into lentivirus. Primary cell cultures enriched for eNs based on SATB2 staining were infected with lentivirus for 24 hours, and mRNA was extracted on day 7. qRT-PCR was used to quantify *ADRA2A* expression using the following primers: TCGTCATCATCGCCGTGTTC (forward) and AAGCCTTGCCGAAGTACCAG (reverse). All sgRNA sequences used for validation can be found in **Supplementary Table 1.13**.

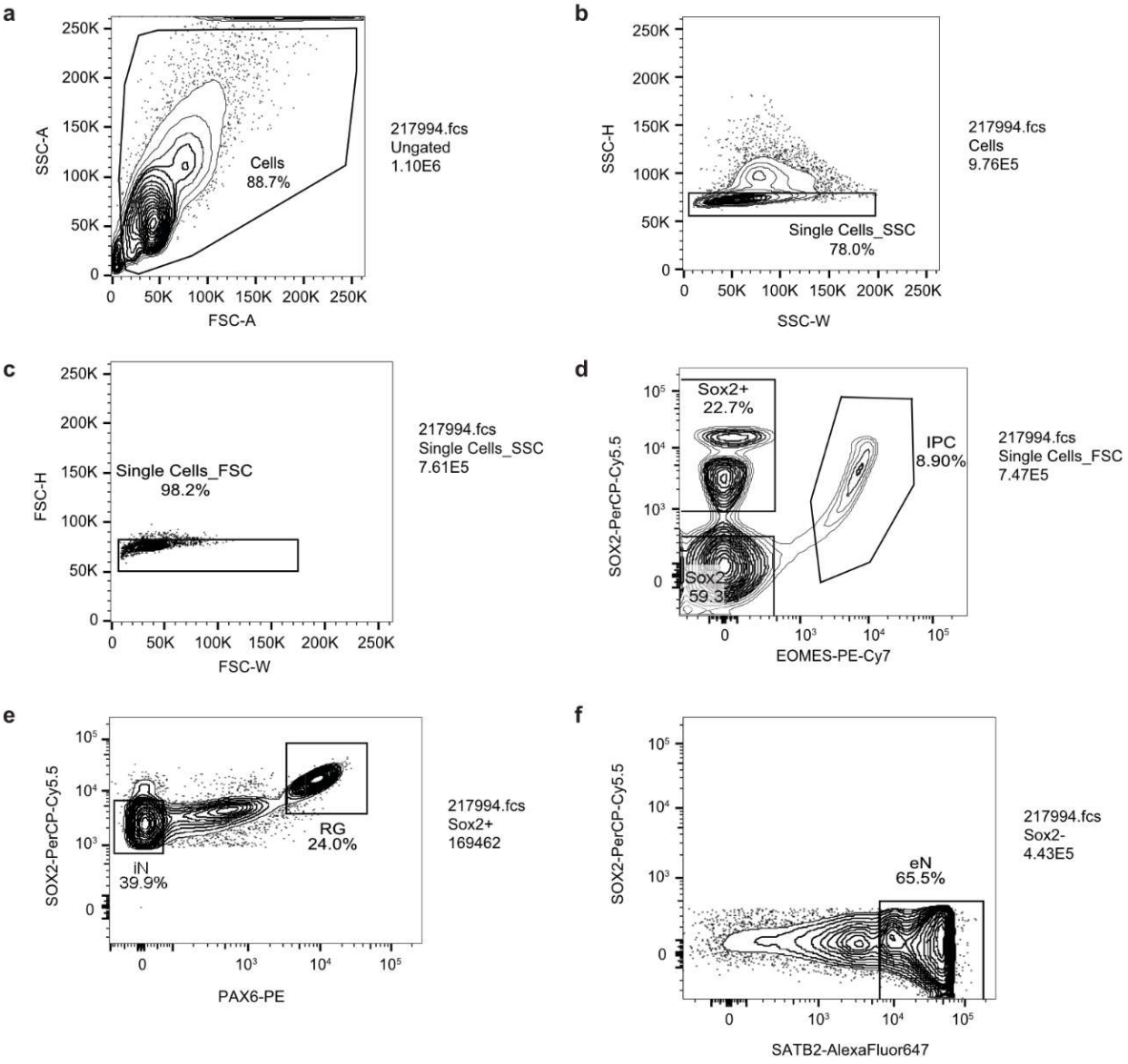
Validating distal interacting regions using CRISPRview

The CRISPRi vector was modified from the Mosaic-seq (Xie *et al.*, 2017)v and CROP-seq vectors (Datlinger and others, 2017). The hU6-sgRNA expression cassette from the CROPseq-Guide-Puro vector (Addgene #86708) was cloned and inserted downstream of the WPRE element in the Lenti-dCas9-KRAB-blast vector (Addgene #89567). The blasticidin resistance gene was replaced with either mCherry or EGFP. sgRNAs targeting open chromatin peaks in distal interacting regions were designed using CHOPCHOP (Labun and others, 2019). Single-stranded DNA was annealed and ligated into the CRISPRi vector at the BsmBI cutting locus. Single clones were picked following transformation, and the sgRNA sequences were confirmed by Sanger sequencing. For lentiviral packaging, the CRISPRi vector, pMD2.G (Addgene #12259), and psPAX (Addgene #12260) were transformed into 293T cells using PolyJet (SignaGen Laboratories #SL100688) according to the manufacturer's instructions. Virus-containing media was collected three times over 16 to 20 hours and concentrated using Amicon 10K columns. All lentivirus was immediately stored at -80°C. Primary cell cultures were infected with virus (MOI < 1) 24 hours after plating, and cells were fixed with 4% PFA four days post-infection for FISH and immunostaining. All sgRNA sequences used for validation can be found in **Supplementary Table 1.13**.

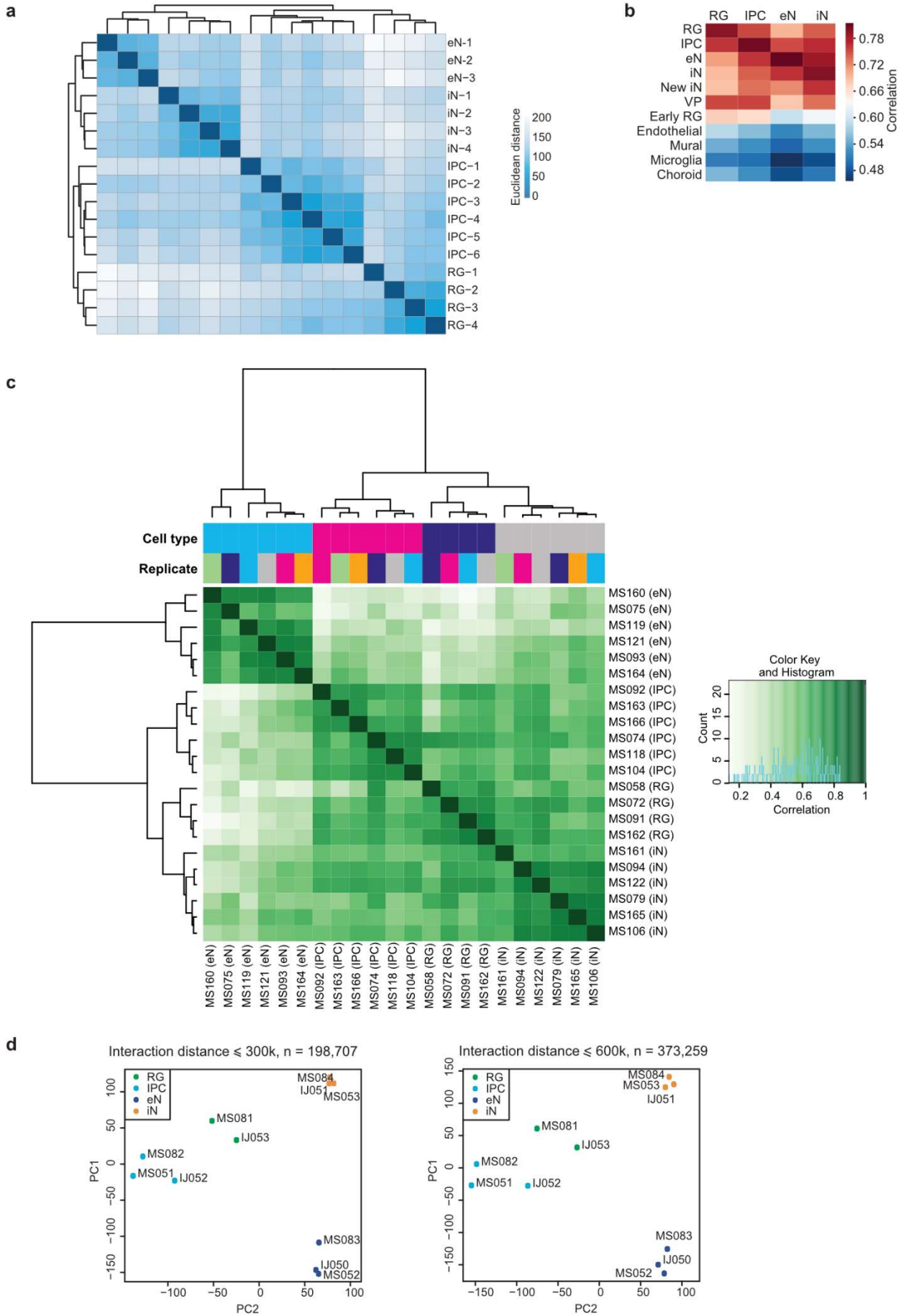
FISH experiments were performed using the RNAScope Multiplex Fluorescent V2 Assay kit (ACDBio #323100). Probes targeting intronic regions for *GPX3* (ACDBio #572341), *IDH1* (ACDBio #832031), *TNC* (ACDBio #572361), and *HES1* (ACDBio #560881) were custom-designed, synthesized, and labeled with TSA Cyanine 5 (Perkin Elmer #NEL705A001KT, 1:1000 dilution). Fixed cells were pretreated with hydrogen peroxide

for 10 minutes and Protease III for 15 minutes, and probes were hybridized and amplified according to the manufacturer's instructions. Slides were washed with PBS before blocking with 5% donkey serum in PBS for 30 minutes at room temperature. Next, slides were incubated with primary antibodies against mCherry (Abcam ab205402, 1/200), GFP (Abcam ab1218, 1/500), and GFAP (Abcam ab7260, 1/400) for RG or SATB2 (Abcam ab92446, 1/300) for eNs overnight at 4°C, followed by incubation with Alexa Fluor 488 donkey anti-mouse IgG (Thermo Fisher Scientific #A21202, 1/800), Alexa-546 nm donkey anti-rabbit IgG (Thermo Fisher Scientific #A10040, 1/500), and Alexa-594 nm goat anti-chicken IgG (Thermo Fisher Scientific #A11042, 1/500) for 1 hour at room temperature. 3D confocal microscopy images were captured using a Leica TCS SP8 with a 40x oil-immersion objective lens (NA = 1.30). The z-step size was 0.4 μm. For five color multiplexed imaging, three sequential scans were performed to avoid overlapping spectra. The excitation lasers were 405 nm and 594 nm, 488 nm and 633 nm, and 561 nm. All images were obtained using the same acquisition settings. For FISH analysis, we developed a Python-based pipeline called Single-Molecule Automatic RNA Transcription Quantification (SMART-Q) for quantifying nascent RNA transcripts in single cells. Briefly, the RNAscope channel was first filtered and fitted in three dimensions using a Gaussian model. Next, segmentation was performed in two dimensions on the DAPI channel to ascertain the location of each nucleus. Finally, segmentation was performed on the remaining channels to identify experimental and control sgRNA-infected RG or eNs for nascent RNA transcript quantification.

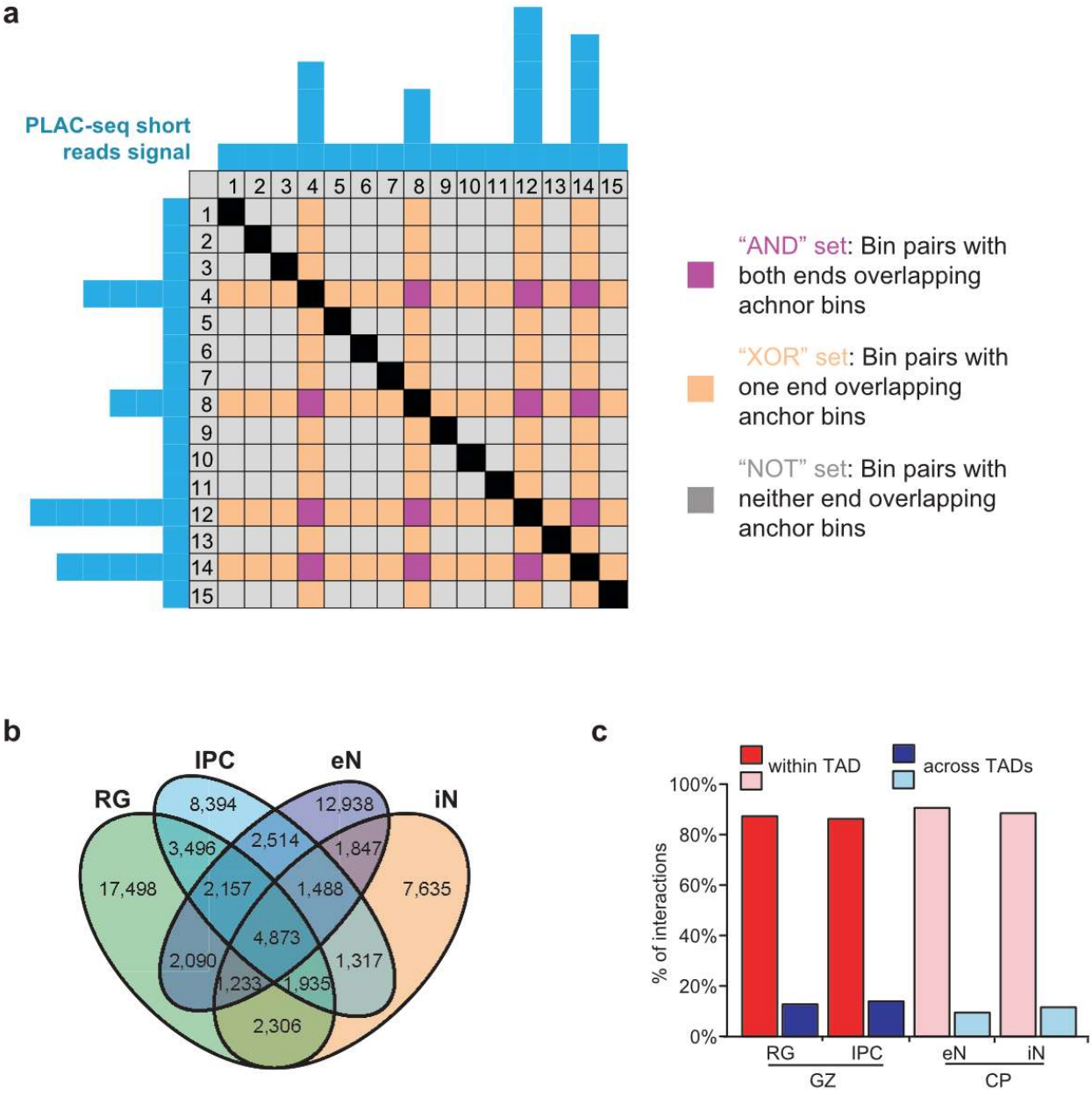
Supplemental Figures



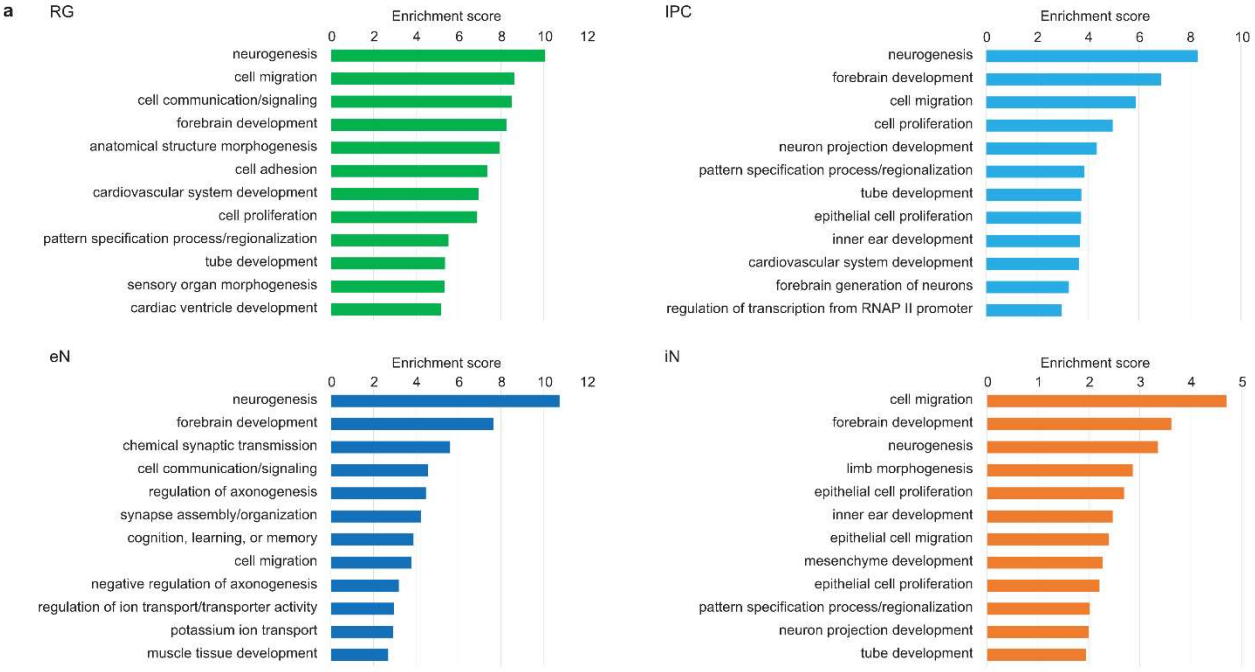
Extended Data Figure 1.1. Representative contour plots depicting FACS gating strategy. (a) Cells were separated from debris of various sizes based on the forward scatter area (FSC-A) and side scatter area (SSC-A). Specifically, they were passed through two singlet gates using the width and height metrics of the (**b**) side scatter (SSC-H versus SSC-W) and (**c**) forward scatter (FSC-H versus FSC-W). (**d**) SOX2+, and SOX2-, and intermediate progenitor (IPC) populations were isolated by gating on EOMES-PE-Cy7 and SOX2-PerCP-Cy5.5 staining. (**e**) Radial glia (RG) and interneurons (iNs) were isolated based on high PAX6/high SOX2 and medium SOX2/low PAX6 staining, respectively. (**f**) Excitatory neurons (eNs) were isolated from the SOX2- population by gating on SATB2-Alexa Fluor 647 staining.

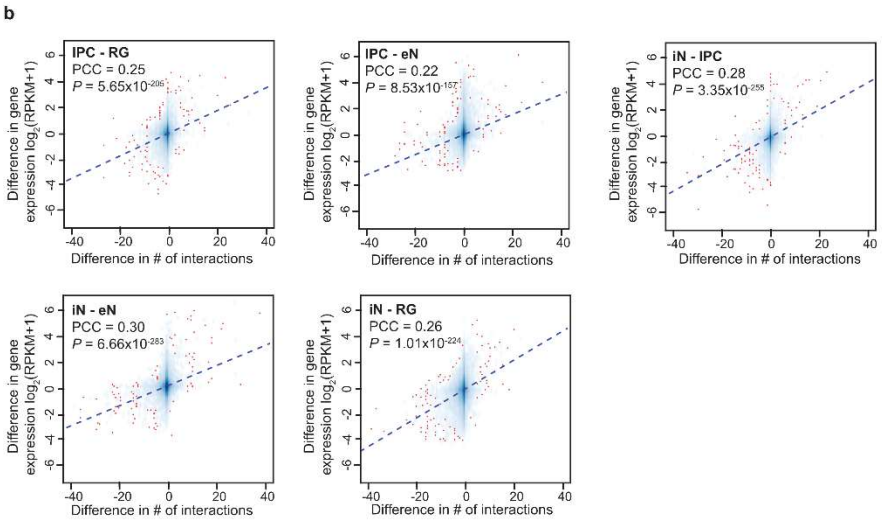

a
eN-1
eN-2
eN-3
iN-1
iN-2
iN-3
iN-4
IPC-1
IPC-2
IPC-3
IPC-4
IPC-5
IPC-6
RG-1
RG-2
RG-3
RG-4
Euclidean distance
200
150
100
50
0
b
RG
IPC
eN
iN
RG
IPC
eN
iN
New iN
VP
Early RG
Endothelial
Mural
Microglia
Choroid
0.78
0.72
0.66
0.60
0.54
0.48
Correlation
c
Cell type
Replicate
MS160 (eN)
MS075 (eN)
MS119 (eN)
MS121 (eN)
MS093 (eN)
MS164 (eN)
MS092 (IPC)
MS163 (IPC)
MS166 (IPC)
MS074 (IPC)
MS118 (IPC)
MS104 (IPC)
MS058 (RG)
MS072 (RG)
MS091 (RG)
MS162 (RG)
MS161 (iN)
MS094 (iN)
MS122 (iN)
MS079 (iN)
MS165 (iN)
MS106 (iN)
Color Key and Histogram
Count
0
5
10
15
20
0.2
0.4
0.6
0.8
1
Correlation
d
Interaction distance ⩽ 300k, n = 198,707
RG
IPC
eN
iN
MS084
IJ051
MS053
MS081
IJ053
MS082
MS051
IJ052
MS083
IJ050
MS052
PC1
PC2
Interaction distance ⩽ 600k, n = 373,259
RG
IPC
eN
iN
MS084
MS053
IJ051
MS081
IJ053
MS082
MS051
IJ052
MS083
IJ050
MS052
PC1
PC2

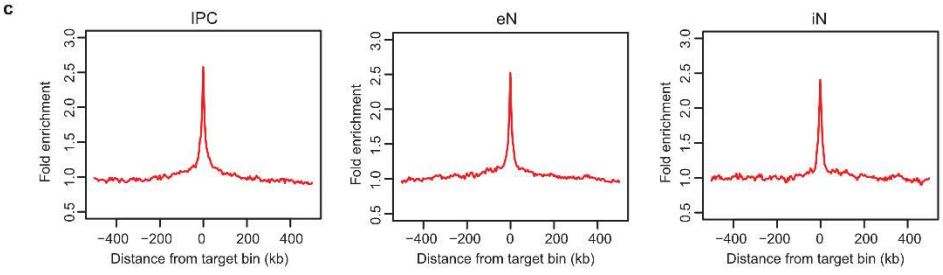
Extended Data Figure 1.2. Reproducibility between RNA-seq, ATAC-seq, and PLAC-seq replicates. (**a**) RNA-seq replicates were hierarchically clustered according to gene expression sample distances using DESeq2. (**b**) Heatmap showing correlations between gene expression profiles for the sorted cell populations and single-cell RNA sequencing (scRNA-seq) data in the developing human cortex. The sorted cell populations exhibited the highest correlation with their corresponding subtypes while exhibiting reduced correlation with the endothelial, mural, microglial, and choroid plexus lineages. (**c**) Heatmap showing correlations and hierarchical clustering for read densities at open chromatin peaks across all ATAC-seq replicates. (**d**) Principle component analysis (PCA) was performed based on normalized contact frequencies across all PLAC-seq replicates (see methods). PCA was performed using interacting 5 kb bins in both 300 and 600 kb windows.



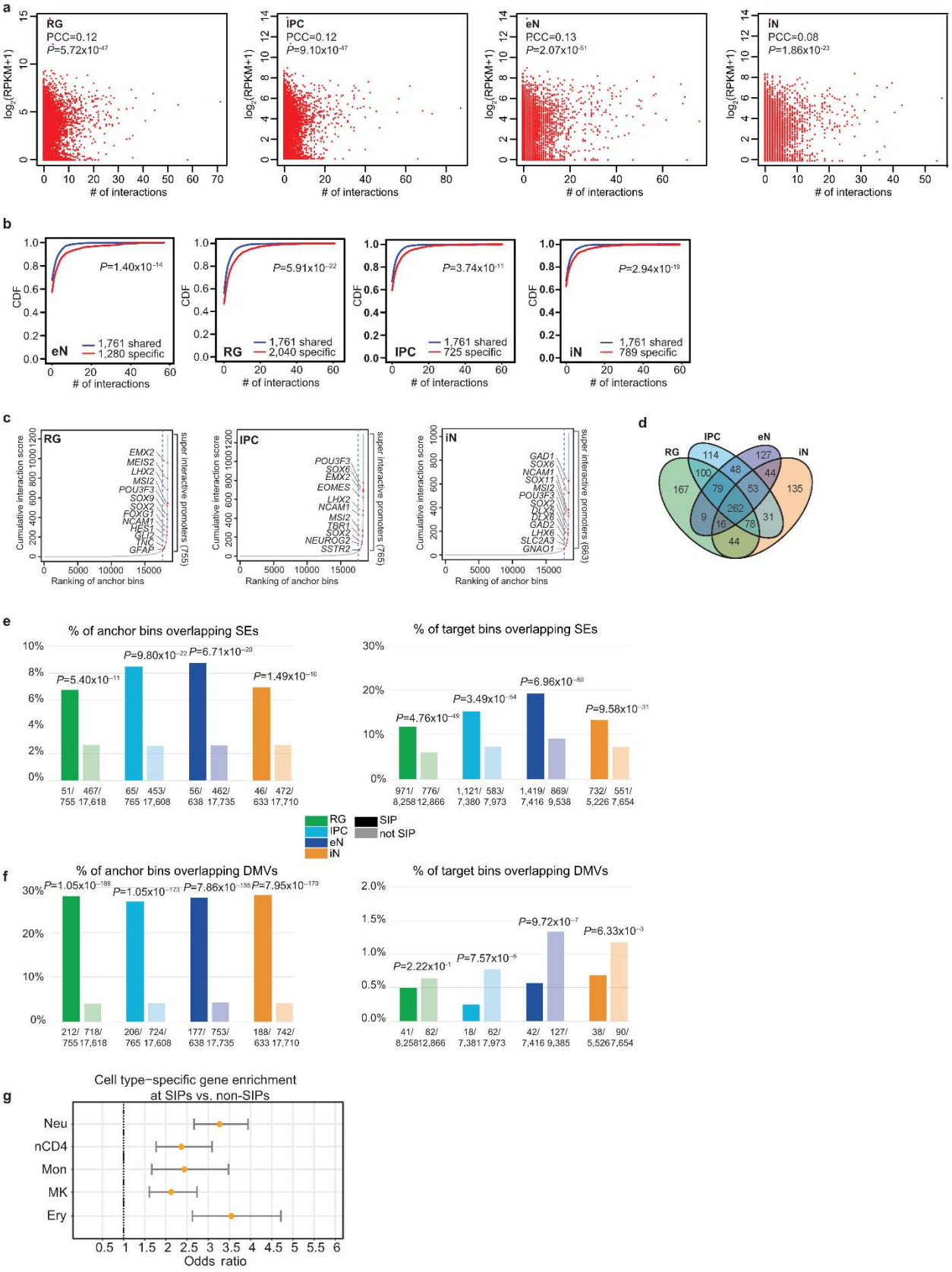
Extended Data Figure 1.3. Identification of significant H3K4me3-mediated chromatin interactions. (**a**) Illustration of XOR and AND interactions in a representative PLAC-seq contact matrix. The blue tracks represent H3K4me3 peaks at anchor bins. Purple cells represent AND interactions where both of the interacting bins are anchor bins. Orange cells represent XOR interactions where only one of the interacting bins is an anchor bin. Grey cells represent NOT interactions where neither of the interacting bins are anchor bins. (**b**) Venn diagram displaying cell-type specificity for interactions in each cell type. (**c**) Proportions of interactions occurring within and across TADs in the GZ and CP for matching cell types.


a
RG
Enrichment score
0 2 4 6 8 10 12
neurogenesis
cell migration
cell communication/signaling
forebrain development
anatomical structure morphogenesis
cell adhesion
cardiovascular system development
cell proliferation
pattern specification process/regionalization
tube development
sensory organ morphogenesis
cardiac ventricle development
IPC
Enrichment score
0 2 4 6 8 10
neurogenesis
forebrain development
cell migration
cell proliferation
neuron projection development
pattern specification process/regionalization
tube development
epithelial cell proliferation
inner ear development
cardiovascular system development
forebrain generation of neurons
regulation of transcription from RNAP II promoter
eN
Enrichment score
0 2 4 6 8 10 12
neurogenesis
forebrain development
chemical synaptic transmission
cell communication/signaling
regulation of axonogenesis
synapse assembly/organization
cognition, learning, or memory
cell migration
negative regulation of axonogenesis
regulation of ion transport/transporter activity
potassium ion transport
muscle tissue development
iN
Enrichment score
0 1 2 3 4 5
cell migration
forebrain development
neurogenesis
limb morphogenesis
epithelial cell proliferation
inner ear development
epithelial cell migration
mesenchyme development
epithelial cell proliferation
pattern specification process/regionalization
neuron projection development
tube development

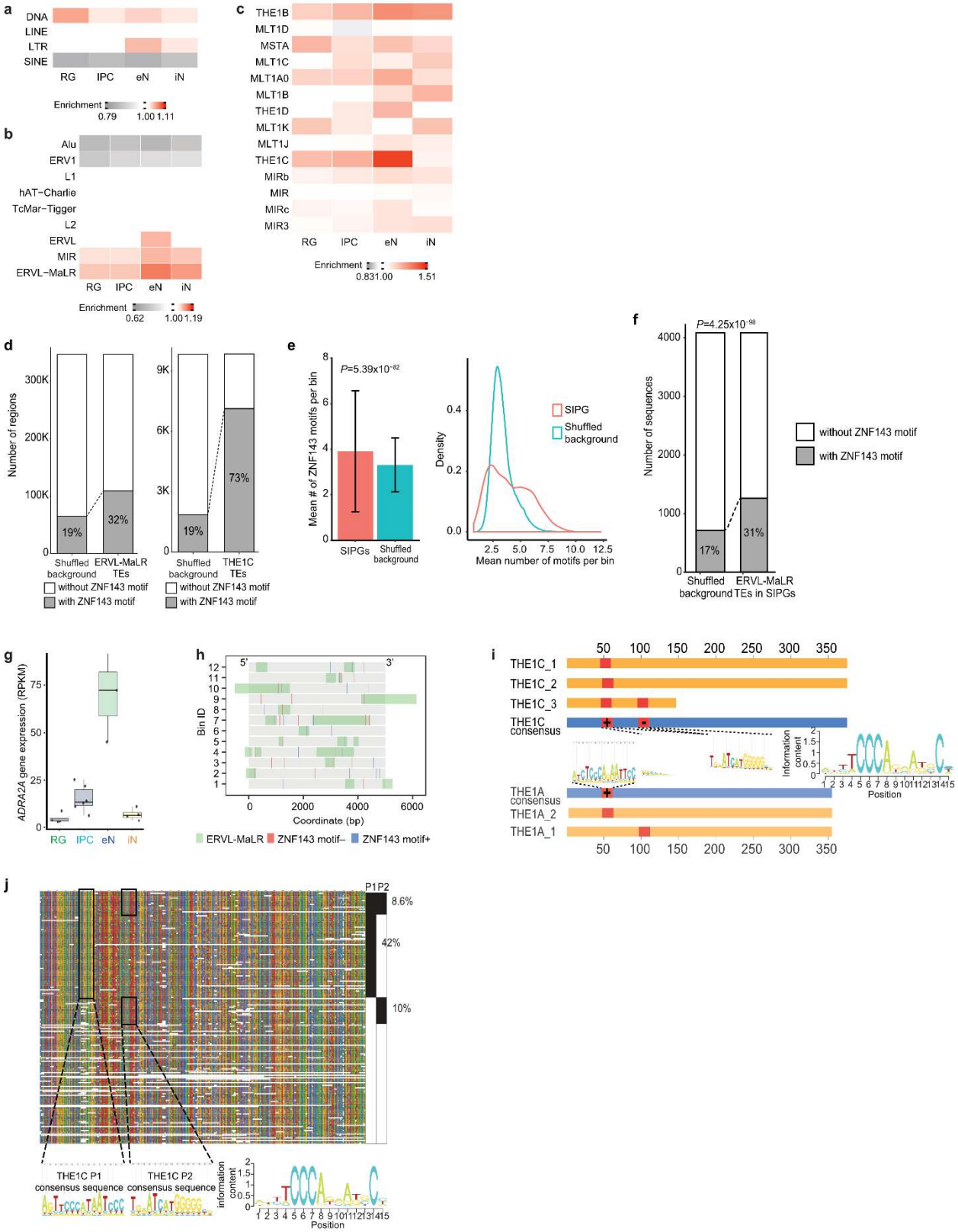

b
IPC - RG
PCC = 0.25
P = 5.65x10-205
IPC - eN
PCC = 0.22
P = 8.53x10-157
iN - IPC
PCC = 0.28
P = 3.35x10-255
iN - eN
PCC = 0.30
P = 6.66x10-283
iN - RG
PCC = 0.26
P = 1.01x10-224
Difference in gene expression log2(RPKM+1)
Difference in # of interactions


c
IPC
eN
iN
Fold enrichment
Distance from target bin (kb)

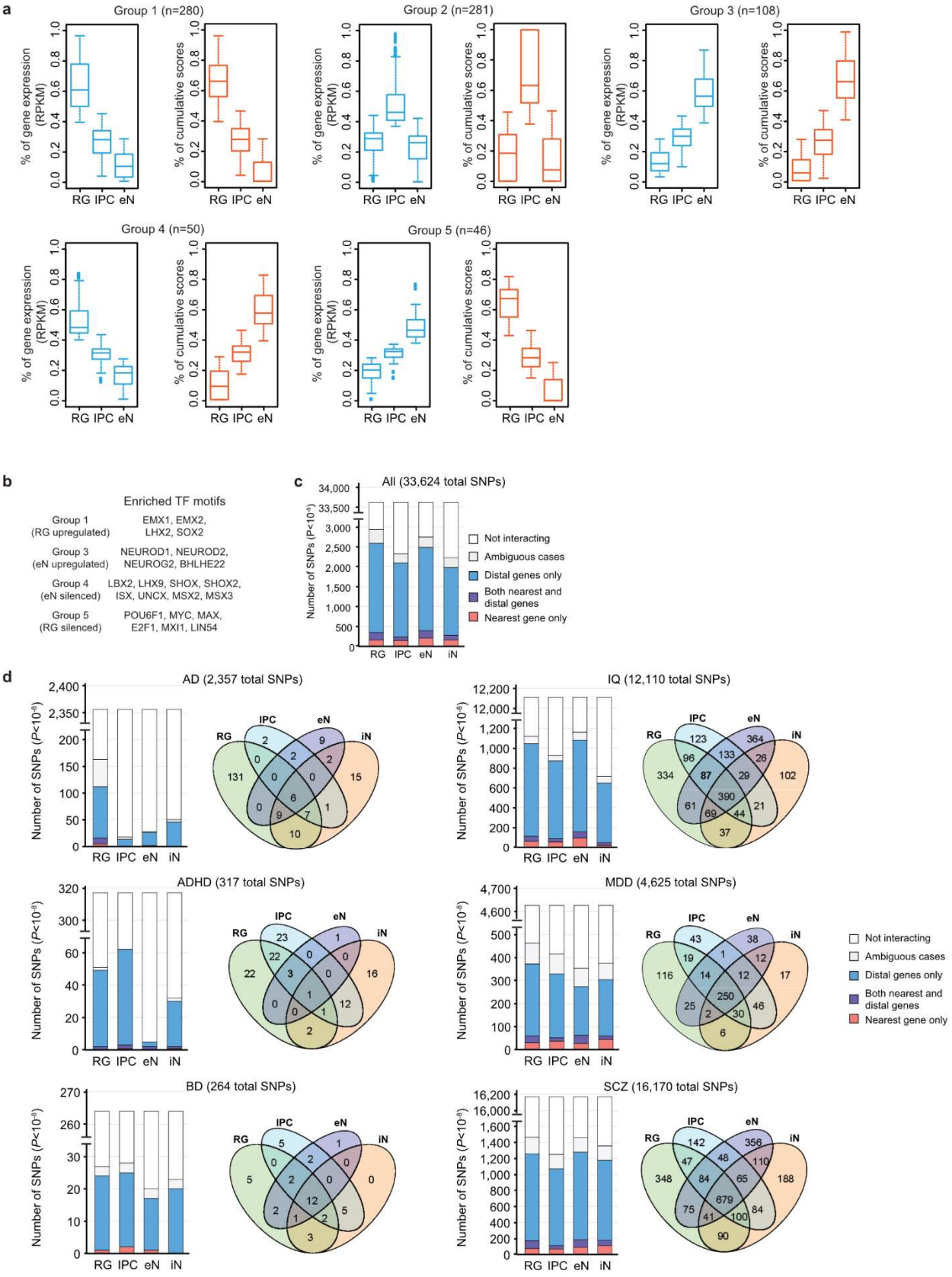
Extended Data Figure 1.4. Chromatin interactions influence cell type-specific transcription. (**a**) GO enrichment analysis for genes participating in cell type-specific interactions. The top annotation clusters from DAVID are reported along with their group enrichment scores for each cell type (see methods). (**b**) Scatterplots showing the correlation between the difference in the number of interactions for each promoter and the difference in the expression of the corresponding genes across all cell types (Pearson product-moment correlation coefficient, two-tailed, n = 13,996 anchor bins with promoters). The trendline from linear regression is shown. (**c**) Fold enrichment of open chromatin peaks over distance-matched background regions in 1 Mb windows around distal interacting regions for IPCs, eNs, and iNs.


a
RG
PCC=0.12
P=5.72x10-47
IPC
PCC=0.12
P=9.10x10-47
eN
PCC=0.13
P=2.07x10-51
iN
PCC=0.08
P=1.86x10-23
log2(RPKM+1)
# of interactions
b
CDF
P=1.40x10-14
eN
1,761 shared
1,280 specific
P=5.91x10-22
RG
1,761 shared
2,040 specific
P=3.74x10-11
IPC
1,761 shared
725 specific
P=2.94x10-19
iN
1,761 shared
789 specific
c
RG
IPC
iN
Cumulative interaction score
Ranking of anchor bins
super interactive promoters (755)
super interactive promoters (663)
EMX2
MEIS2
LHX2
MSI2
POU3F3
SOX9
SOX2
FOXG1
NCAM1
HES1
GLI2
TNC
GFAP
POU3F3
SOX6
EMX2
EOMES
LHX2
NCAM1
MSI2
TBR1
SOX2
NEUROG2
SSTR2
GAD1
SOX6
NCAM1
SOX11
MSI2
POU3F3
SOX2
DLX5
DLX6
GAD2
LHX6
SLC2A3
GNAO1
d
RG
IPC
eN
iN
114
127
48
100
44
167
79
53
135
262
9
16
78
31
44
e
% of anchor bins overlapping SEs
P=5.40x10-11
P=9.80x10-22
P=6.71x10-20
P=1.49x10-10
51/ 755
467/ 17,618
65/ 765
453/ 17,608
56/ 638
462/ 17,735
46/ 633
472/ 17,710
% of target bins overlapping SEs
P=4.76x10-49
P=3.49x10-94
P=6.96x10-60
P=9.58x10-31
971/ 8,258
776/ 12,866
1,121/ 7,380
583/ 7,973
1,419/ 7,416
869/ 9,538
732/ 5,226
551/ 7,654
RG
IPC
eN
iN
SIP
not SIP
f
% of anchor bins overlapping DMVs
P=1.05x10-189
P=1.05x10-173
P=7.86x10-155
P=7.95x10-170
212/ 755
718/ 17,618
206/ 765
724/ 17,608
177/ 638
753/ 17,735
188/ 633
742/ 17,710
% of target bins overlapping DMVs
P=2.22x10-1
P=7.57x10-6
P=9.72x10-7
P=6.33x10-3
41/ 8,258
82/ 12,866
18/ 7,381
62/ 7,973
42/ 7,416
127/ 9,385
38/ 5,526
90/ 7,654
g
Cell type−specific gene enrichment
at SIPs vs. non-SIPs
Neu
nCD4
Mon
MK
Ery
Odds ratio

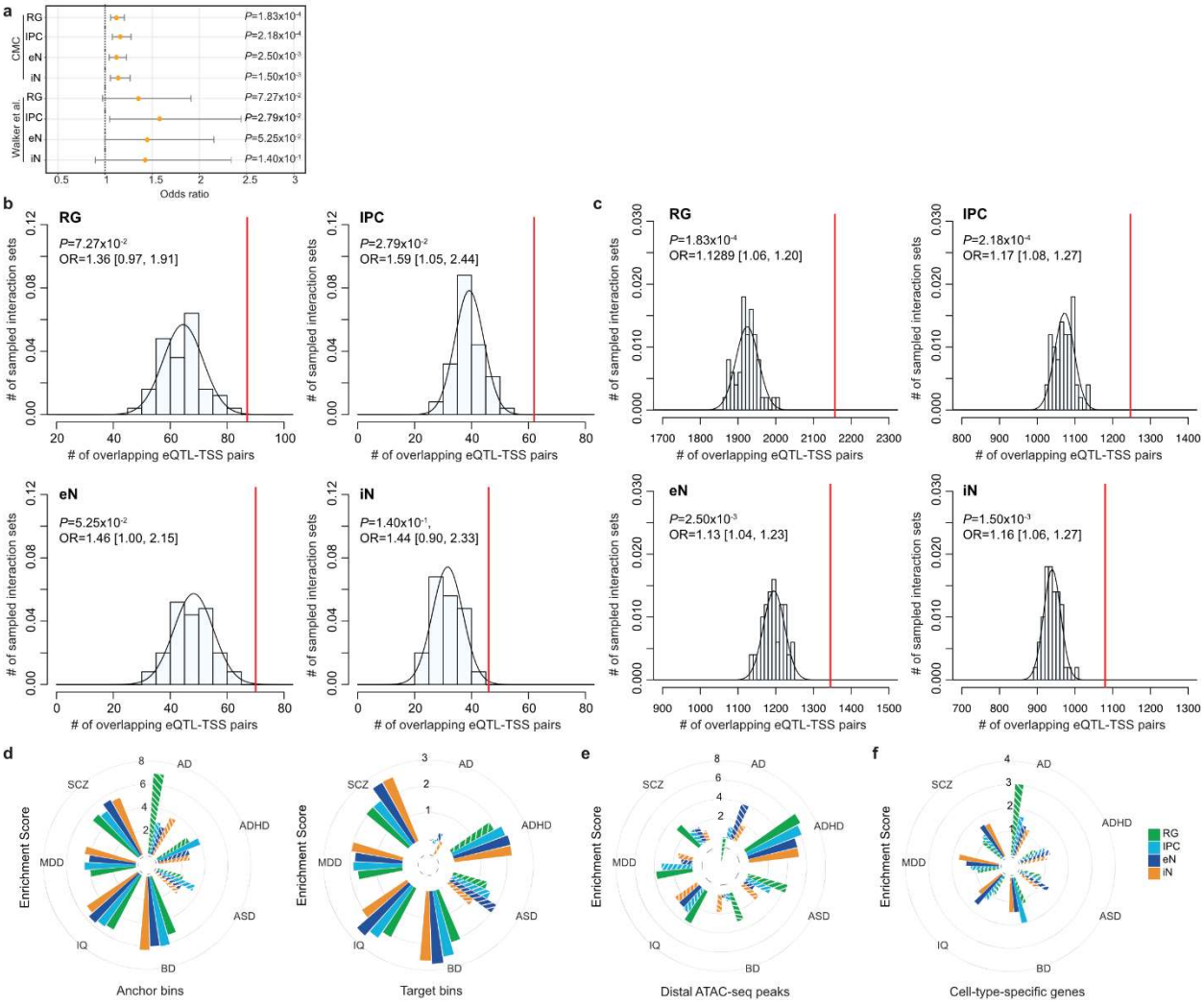
Extended Data Figure 1.5. Super interactive promoters are enriched for lineage-specific genes. (**a**) Scatterplots showing the correlation between interaction counts and gene expression at promoters for each cell type (Pearson product-moment correlation coefficient, two-tailed, n = 13,996 anchor bins with promoters). (**b**) CDF plots of the numbers of interactions for shared versus cell type-specific genes for each cell type (two-sample t-test, two-tailed). (**c**) Anchor bins were ranked according to their cumulative interaction scores in RG, IPCs, and iNs. Super interactive promoters (SIPs) are located past the point in each curve where the slope is equal to 1. (**d**) Venn diagram displaying cell type-specificity for SIPs in each cell type. (**e-f**) Enrichment of super-enhancers and DMVs at SIPs versus non-SIPs (left) and distal interacting regions for SIPs versus non-SIPs (right) (Fisher's exact test, two-tailed). Super-enhancers were based on data in the fetal brain and adult cortex, while DMVs were based on data in 40 and 60 day cerebral organoids with closely matched gene expression profiles to mid-fetal cortex samples. (**g**) Forrest plot showing that SIPs identified in hematopoietic cells are analogously enriched for cell type-specific over shared genes. Odds ratios and 95% confidence intervals are shown. We identified 554, 709, 460, 712, and 401 SIPs in neutrophils, naive CD4+ T cells, monocytes, megakaryocytes, and erythroblasts, respectively.


a
DNA
LINE
LTR
SINE
RG IPC eN iN
Enrichment
0.79 1.00 1.11
b
Alu
ERV1
L1
hAT−Charlie
TcMar−Tigger
L2
ERVL
MIR
ERVL−MaLR
RG IPC eN iN
Enrichment
0.62 1.00 1.19
c
THE1B
MLT1D
MSTA
MLT1C
MLT1A0
MLT1B
THE1D
MLT1K
MLT1J
THE1C
MIRb
MIR
MIRc
MIR3
RG IPC eN iN
Enrichment
0.83 1.00 1.51
d
Number of regions
300K
200K
100K
0
19%
32%
Shuffled background
ERVL-MaLR TEs
without ZNF143 motif
with ZNF143 motif
9K
6K
3K
0
19%
73%
Shuffled background
THE1C TEs
without ZNF143 motif
with ZNF143 motif
e
P=5.39x10-82
Mean # of ZNF143 motifs per bin
SIPGs
Shuffled background
Density
SIPG
Shuffled background
Mean number of motifs per bin
f
P=4.25x10-98
Number of sequences
4000
3000
2000
1000
0
17%
31%
Shuffled background
TEs in SIPGs
without ZNF143 motif
with ZNF143 motif
g
ADRA2A gene expression (RPKM)
RG IPC eN iN
h
Bin ID
5'
3'
Coordinate (bp)
ERVL-MaLR
ZNF143 motif–
ZNF143 motif+
i
THE1C_1
THE1C_2
THE1C_3
THE1C consensus
THE1A consensus
THE1A_2
THE1A_1
Information content
Position
j
P1P2
8.6%
42%
10%
THE1C P1 consensus sequence
THE1C P2 consensus sequence
Information content
Position

Extended Data Figure 1.6. Transposable elements in SIP formation. (a-c) Enrichment of TEs at the class (**a**), family (**b**), and subfamily (**c**) levels in SIPGs for each cell type. Only TE families occupying more than 1% of the genome are shown in (**b**). Only TE subfamilies from the MIR and ERVL-MaLR TE families occupying more than 0.1% of the genome are shown in (**c**). (**d**) Both ERVL-MaLR TEs (left, 32% versus 19% of sequences, $P < 2.2*10^{-16}$, binomial test, two-tailed) and THE1C TEs (right, 73% versus 19% of sequences, $P < 2.2*10^{-16}$, binomial test, two-tailed) are enriched over background sequences for ZNF143 motifs in eNs. (**e**) ZNF143 motifs are enriched at SIPGs in eNs (left, $P = 5.39x10^{-82}$, two-sample t-test, two-tailed, n = 8,894 distal interacting regions). Means are indicated and error bars represent the SEM. Distributions comparing the number of ZNF143 motifs per bin for actual versus shuffled SIPGs are shown (right, $P < 2.2*10^{-16}$, Kolmogorov-Smirnov test, two-tailed, n = 638 SIPGs). (**f**) ERVL-MaLR TEs in SIPGs are enriched over background sequences for ZNF143 motifs in eNs (31% versus 17% of sequences, $P = 4.3x10^{-98}$, binomial test, two-tailed). (**g**) Box plots showing elevated *ADRA2A* gene expression in eNs. The median, upper and lower quartiles, minimum, and maximum are indicated. (**h**) Illustration of the 12 distal interacting regions containing ERVL-MaLR TE-localized ZNF143 motifs in the *ADRA2A* SIPG. ZNF143 motifs are colored by strand. The bin numbers correspond to **Fig. 1.3j**. (**i**) Conservation of ERVL-MaLR TEs in the *ADRA2A* SIPG. Blue bars indicate consensus sequences, yellow bars indicate ERVL-MaLR TEs, and red bars indicate ZNF143 motifs. (**j**) Alignment of THE1C TEs in the human genome to their consensus sequence. The THE1C subfamily contains two ZNF143 motifs, one at positions 47-61 (P1), and another at positions 96-110 (P2).

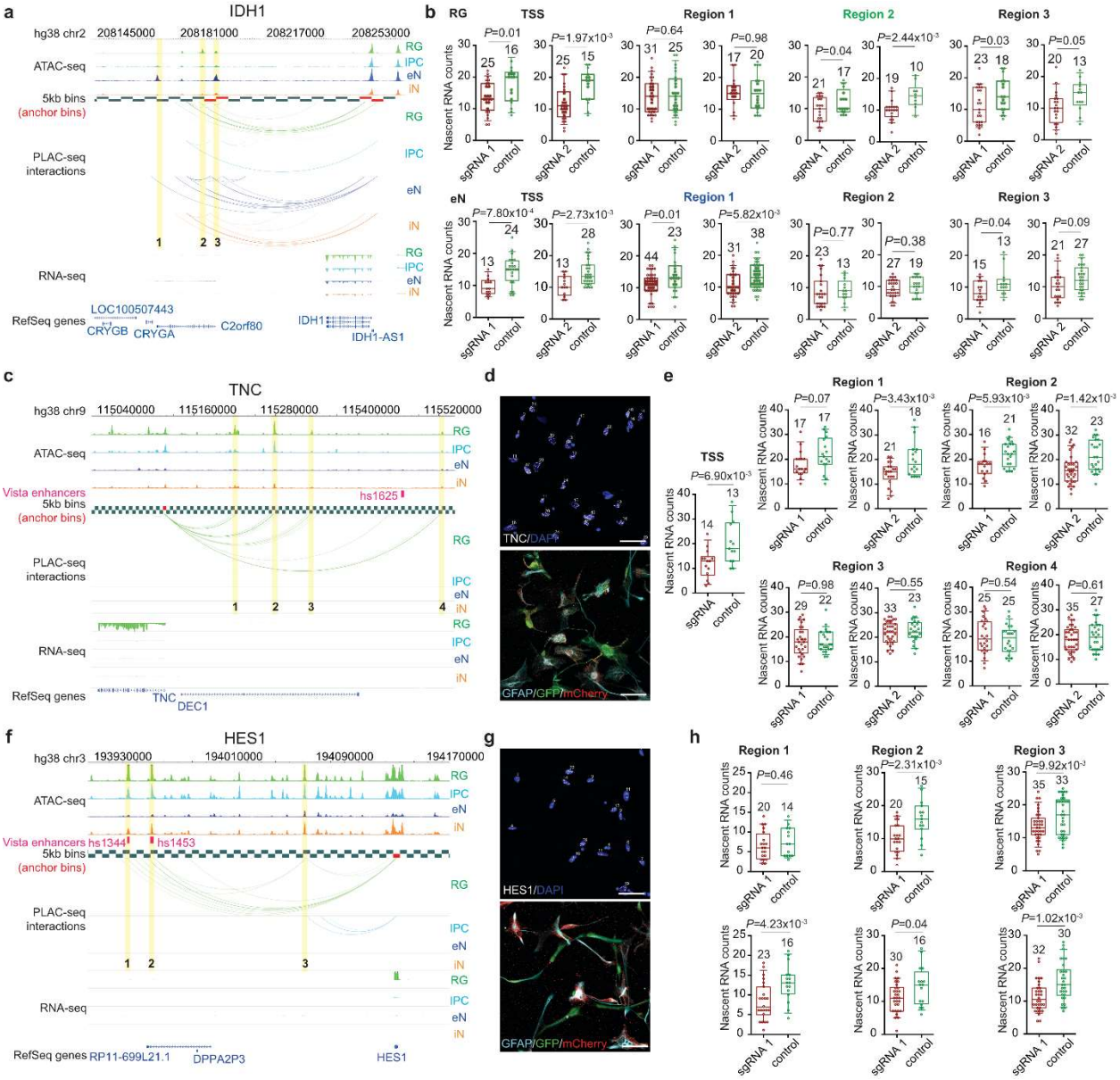

a
Group 1 (n=280)
Group 2 (n=281)
Group 3 (n=108)
Group 4 (n=50)
Group 5 (n=46)
% of gene expression (RPKM)
% of cumulative scores
RG IPC eN
b
Enriched TF motifs
Group 1 (RG upregulated) EMX1, EMX2, LHX2, SOX2
Group 3 (eN upregulated) NEUROD1, NEUROD2, NEUROG2, BHLHE22
Group 4 (eN silenced) LBX2, LHX9, SHOX, SHOX2, ISX, UNCX, MSX2, MSX3
Group 5 (RG silenced) POU6F1, MYC, MAX, E2F1, MXI1, LIN54
c
All (33,624 total SNPs)
Number of SNPs (P<10-6)
RG IPC eN iN
Not interacting
Ambiguous cases
Distal genes only
Both nearest and distal genes
Nearest gene only
d
AD (2,357 total SNPs)
IQ (12,110 total SNPs)
ADHD (317 total SNPs)
MDD (4,625 total SNPs)
BD (264 total SNPs)
SCZ (16,170 total SNPs)
Number of SNPs (P<10-6)
RG IPC eN iN

Extended Data Figure 1.7. Developmental trajectories and mapping complex disorder- and trait-associated variants to their target genes. (**a**) Box plots showing the distributions of gene expression and cumulative interaction scores for the groups identified in **Fig. 1.4a**. The median, upper and lower quartiles, minimum, and maximum are indicated. (**b**) Groups 4 and 5 are enriched for interactions with TFs containing domains associated with transcriptional repression. (**c-d**) Counts of the numbers of GWAS SNPs ($P < 10^{-8}$) interacting with their nearest gene only, with both their nearest and more distal genes, and with more distal genes only across all diseases (**c**) and specific disorders and traits (**d**).



Extended Data Figure 1.8. Partitioning SNP heritability for complex disorders and traits using alternative epigenomic annotations. (**a**) Forrest plot showing the enrichment of fetal and adult brain eQTL-TSS pairs in our interactions compared to n = 50 sets of distance-matched control interactions (Fisher's exact test, two-tailed). Odds ratios and 95% confidence intervals are shown. The increased significance of adult brain eQTLs can be attributed to the larger sample size of the CommonMind Consortium (CMC) study (n = 1,332,863), while larger odds ratios were observed for the more closely matched fetal brain eQTLs (n = 6,446). (**b-c**) Histograms displaying the numbers of adult and fetal brain eQTL-TSS pairs recapitulated by n = 50 sets of distance-matched control interactions in each cell type. The numbers of eQTL-TSS pairs recapitulated by our interactions are indicated by red lines (Fisher's exact test, two-tailed). (**d**) LDSC enrichment scores for each disease and cell type, conditioned on the baseline model from Gazal et al. 2017 and stratified by PLAC-seq anchor and target bins. Non-significant enrichment scores are shown as striped bars. (**e-f**) LDSC enrichment scores for each disease and cell type, conditioned on the baseline model from Gazal et al. 2017 and using either distal open chromatin peaks (**e**) or cell type-specific genes (**f**). Non-significant enrichment scores are shown as striped bars.

Extended Data Figure 1.9. Enriched biological processes for genes interacting with non-coding variants for each disease and cell type. GO enrichment analysis for genes interacting with non-coding variants for each disease and cell type using H-MAGMA and gProfileR (Fisher's exact test, two-tailed, BH method). The full results can be found in **Supplementary Table 1.12**.



Extended Data Figure 1.10. Characterization of RG- and eN-specific loci using CRISPRview. (**a-b**) Validation of distal interacting regions at the *IDH1* locus in RG and eNs. Silencing region 1, which interacts with the *IDH1* promoter only in eNs, results in the significant downregulation of *IDH1* expression in eNs but not in RG. Silencing region 2, which interacts with the *IDH1* promoter only in RG, results in the significant downregulation of *IDH1* expression in RG but not in eNs. Silencing region 3, which interacts with the *IDH1* promoter in both RG and eNs, results in the significant downregulation of *IDH1* expression in both cell types. Interactions between the promoter of *IDH1* and distal interacting regions containing open chromatin peaks that were targeted for silencing are highlighted. Box plots show results for experimental (red) and control (green) sgRNA-treated cells for each region (two-sample t-test, two-tailed). The median, upper and lower quartiles, and 10% to 90% range are indicated. Open

circles represent single cells. Sample sizes are indicated above each box plot. (**c-h**) Validation of distal interacting regions at the *TNC and HES1* loci in RG. Interactions between the promoters of *TNC and HES1* and distal interacting regions containing open chromatin peaks that were targeted for silencing are highlighted. Representative images show staining for intronic RNAscope probes (white), DAPI (blue), GFAP (light blue), GFP (green), and mCherry (red). The scale bar is 50 μm.

Chapter 2:
Human Intermediate Progenitor Diversity during Cortical Development

Abstract

Studies of the spatiotemporal, transcriptomic, and morphological diversity of radial glia (RG) have spurred our current models of human corticogenesis (Betizeau, Cortay, D. Patti, *et al.*, 2013; Pollen *et al.*, 2015; Thomsen *et al.*, 2015; Nowakowski *et al.*, 2016; Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent, Kriegstein, *et al.*, 2017). In the developing cortex, neural intermediate progenitor cells (nIPCs) are a neuron-producing transit-amplifying cell type born in the germinal zones of the cortex from RG. The potential diversity of the nIPC population, that produces a significant portion of excitatory cortical neurons, is understudied, particularly in the developing human brain. Here we explore the spatiotemporal, transcriptomic, and morphological variation that exists within the human nIPC population and provide a resource for future studies. We observe that the spatial distribution of nIPCs in the cortex changes abruptly around gestational week (GW) 19/20, marking a distinct shift in cellular distribution and organization during late neurogenesis. We also identify five transcriptomic subtypes, one of which appears at this spatiotemporal transition. Finally, we observe a diversity of nIPC morphologies that do not correlate with specific transcriptomic subtypes. These results provide an analysis of the spatiotemporal,

transcriptional, and morphological diversity of nIPCs in developing brain tissue and provide an atlas of nIPC subtypes in the developing human cortex that can benchmark human organoid models and help inform future studies of how nIPCs contribute to cortical neurogenesis.

Introduction

The human cerebral cortex is comprised of a diverse set of neurons that are essential to its functions in perception, judgment, and sensory integration. Furthermore, it is three times larger than that of our closest living relative, the chimpanzee. The abundance and diversity of progenitor cells that emerge during development contribute to this expansion. The cerebral cortex arises from the daughter cells produced by radial glia (RG) that serve as neural stem cells. RG reside in the ventricular zone (VZ), a progenitor region that lines the ventricles of the developing brain. In mammals, RG can produce neurons directly, but they more often produce neurons indirectly, via a transit amplifying cell, the neural intermediate progenitor cell (nIPC). nIPCs can divide one or more times before producing neurons (Noctor *et al.*, 2004; Hevner, 2006; Hansen *et al.*, 2010). In rodents, RG produce nIPCs that at early stages of neurogenesis divide in the VZ, but at later stages migrate out of the VZ to divide in a secondary proliferative zone, the subventricular zone (SVZ)(Weissman *et al.*, 2004; Hevner, 2006). In addition to ventricular RG (vRGs), the developing brains of humans and large brain mammals contain an abundant second type of RG, the outer radial glia (oRG) cell, that also generates nIPCs (Hansen *et al.*, 2010). In the dorsal cortex of the mouse, the identity of the SVZ-dividing nIPCs is driven by EOMES, a transcription factor enriched in nIPCs that is sufficient for their generation (Arnold *et al.*, 2008; Sessa *et al.*, 2008; Mihalas *et*

al., 2016). Within the human brain, multiple single cell studies have shown that EOMES expression also marks nIPCs (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent and Kriegstein, 2017; Polioudakis *et al.*, 2019; Fan *et al.*, 2020). In mice, manipulation of EOMES strongly influences cortical neuron number (Arnold *et al.*, 2008; Sessa *et al.*, 2008) and loss of function mutations in EOMES are typically embryonic lethal (Lek *et al.*, 2016). Likewise, patients with partial EOMES deletions have developmental disorders including intellectual disability and autism (Firth *et al.*, 2009). Despite the relevance of nIPCs to the evolutionary and developmental expansion of human cortex, human nIPCs remain poorly characterized.

The heterogeneity of other progenitors in the developing human and primate cortex, including RG and oligodendrocyte precursor cells, has been tied to temporal, spatial, morphological, and transcriptomic characteristics (Hansen *et al.*, 2010; Betizeau, Cortay, D. Patti, *et al.*, 2013; Pollen *et al.*, 2015; Rash *et al.*, 2019; Kalebic and Huttner, 2020). However, similar descriptions of nIPC features are scarce, and previous descriptions of neural progenitor cells in post-mortem pre-term infant tissue are subject to concerns of tissue preservation (Malik *et al.*, 2013). Here we analyze the spatiotemporal, transcriptomic, and morphological diversity of EOMES^{+} nIPCs within developing human cortex.

Results

Intermediate Progenitors follow distinct spatiotemporal patterns in early vs late neurogenesis

To verify the prevalence and distribution of nIPCs at different stages of human cortical neurogenesis during the second trimester, we performed immunostaining for EOMES, which serves as a marker of nIPCs. From gestational week (GW) 14 to GW18, the majority of EOMES+ nIPCs form a dense band in the inner SVZ (iSVZ), with increasing numbers in the oSVZ **(Fig 2.1 B-E)**. Consistent with previous observations that oRGs produce highly proliferative nIPCs (Hansen *et al.*, 2010), we observed that the nIPCs in the oSVZ expressed KI67 more often than those in the iSVZ **(Fig 2.1 C, Sup Fig 2.1)**. Nevertheless, most nIPCs were found within the iSVZ from GW14 to GW18 **(Fig 2.1 C).** Furthermore, staining with SOX2, EOMES and TBR1 between GW14 and GW18 revealed laminar transcription factor expression in the VZ and iSVZ. $SOX2^{+}$ cells were present in the VZ, with $EOMES^{+}$ cells in the adjacent 'inner' iSVZ, and TBR1+ cells inset from the $EOMES^{+}$ cells in the 'outer' iSVZ **(Sup Fig 2)**. No such pattern was evident in the oSVZ; instead we observed SOX2, EOMES, and TBR1 expressing cells mixed throughout the oSVZ. The gradient of transcription factors in the iSVZ, ranging from RG to nIPCs to newborn neurons, suggests that most nIPCs in the iSVZ differentiate into neurons locally and do not migrate into the oSVZ to differentiate. Thus, nIPCs born from the iSVZ may not contribute significantly to the oSVZ nIPC population.

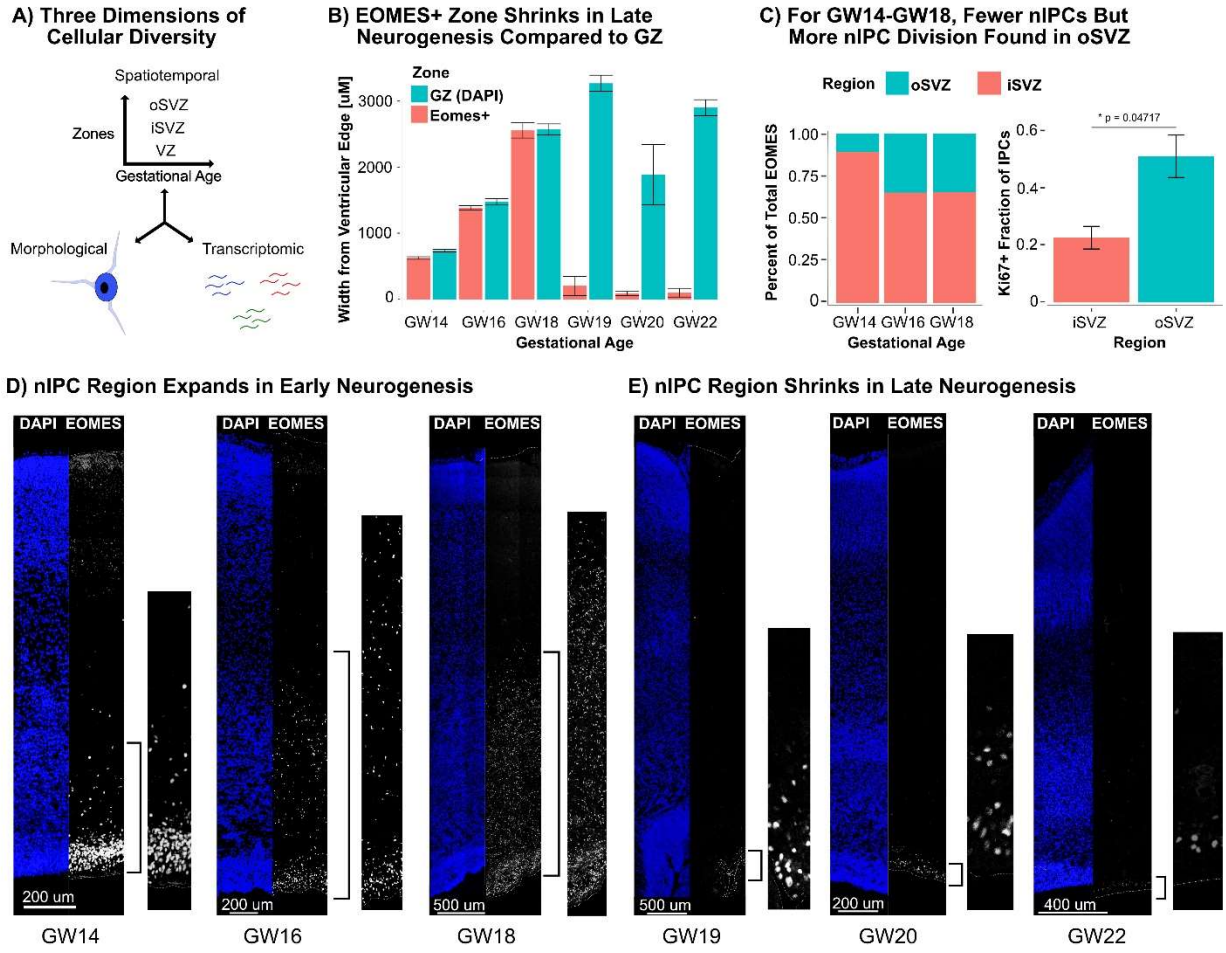


Figure 2.1. Spatiotemporal Diversity. a) The spatiotemporal, transcriptomic, and morphological diversity of nIPCs remains almost entirely unexplored. B) The germinal zones (GZ) expand in thickness during neurogenesis. The EOMES$^+$ area expands until GW19, after which it virtually disappears in the oSVZ and diminishes in the iSVZ, as shown in panel E. Distance measurements were taken from the ventricle to the outermost EOMES$^+$ cell. Error bars represent technical variations from 9 independent measurements of one biological sample. The GW20 measurements represent an average of nine values from both the occipital and prefrontal cortex. Error bars are standard deviation. C) From GW14 to GW18, most nIPCs are found within the iSVZ, not the oSVZ. However, in the oSVZ, nIPCs co-stain with KI67 at a higher rate, indicating that these nIPCs are more proliferative **(Fig 2.1 C)**. The percentage of total EOMES in the oSVZ or iSVZ is the number of EOMES+ nuclei in that region over the total EOMES$^+$ nuclei for a given length along the ventricle. The KI67 fraction represents the percentage of EOMES$^+$ nuclei that also stain for KI67 in a given region for a given length along the ventricle. Error bars are standard deviation. D) The number of nIPCs expand from GW14 to GW18, during early neurogenesis. E) Around GW19, the number of nIPCs abruptly declines. A thin band of EOMES$^+$ cells remains in the iSVZ, and few or no nIPCs are found in the oSVZ. This later distribution was evident in three out of four samples from GW19-20 and all three samples after GW20.

Between GW19-21, we observed an abrupt change in both the number and distribution of nIPCs **(Fig 2.1 B, E)**. While EOMES^{+} nuclei largely disappeared from the oSVZ, a thin band of EOMES+ cells remained in the iSVZ **(Fig 2.1C)**. This change corresponds to approximately the time period when gliogenesis becomes dominant (Semple *et al.*, 2013). We found this pattern in two out of three samples at GW19 and GW20, and all three samples at GW21 and GW22. Because this pattern had not been observed before (Malik *et al.*, 2013), we sought to confirm our observations using a second EOMES antibody which replicated the same pattern **(Sup Fig 2.3 A)**.

Because nIPCs are hypothesized to give rise to the majority of neurons during human cortical development (Baala *et al.*, 2007; Lui, Hansen and Kriegstein, 2011), we expected that excitatory neuron markers in the germinal zones (GZ) would follow similar spatiotemporal trends and mostly disappear from the oSVZ at the same time as EOMES+ cells. To verify this concordance, we stained for transcription factors NEUROD1, TBR1, and NHLH2. We found that these markers followed the expected spatiotemporal pattern **(Sup Fig 2.4)**, with substantial labeling throughout the germinal zone during peak neurogenesis (GW18) and few cells remaining in the iSVZ afterwards. Little to no marker expression was observed in the oSVZ after GW19/GW20. Since five antibodies for EOMES and neuron markers across five samples showed little to no staining in the oSVZ during late neurogenesis, we sought to verify whether the RG scaffold had changed in the oSVZ at this time. Across our samples, markers for RG (HOPX & VIM) as well as intermediate zone fiber tracts (AChE) matched previous

descriptions and revealed the expected cytoarchitecture of developing human brain (Nowakowski *et al.*, 2016; Krsnik *et al.*, 2017; Žunić Išasegi *et al.*, 2018) **(Sup Fig 2.3 B)**. Most neuron marker staining at later ages was in the cortical plate **(Sup Fig 2.4)**, indicating that most newborn excitatory neurons have migrated to the cortical plate at this time. These dynamics indicate that late neurogenesis is marked by a relatively abrupt shift in both the distribution of EOMES expressing cells and excitatory neurogenesis across the germinal zones.

Analysis of Single Cell RNA Sequencing Identifies Subtypes of Intermediate Progenitor Cells

To identify potential transcriptomic subtypes of nIPCs in the developing human cortex, we examined a subclustering of 5,959 Eomes$^+$ cells from a dataset we recently published (Bhaduri *et al.*, 2020) collected from five independent samples of GW8 to GW22 human cortex **(Fig 2.2)**. Noticeably, no EOMES$^+$ cells were detected at GW8. This refined dataset was then normalized, batch corrected, and clustered to identify putative sub-populations of nIPCs. Given the filtering criteria and the comparative homogeneity of nIPCs, we sought to thoroughly verify our clusters and cluster markers by generated statistical thresholds from scrambled data **(Sup Fig 2.5)**. By comparing our clusters to this significance threshold and heuristically combining clusters based on similar marker genes, five clear subtypes or states of nIPCs emerged: three neuron-like nIPC clusters that expressed neurogenic genes; one radial glia-like nIPC (RG-Like nIPCs) cluster; and one *DLX5*$^+$ nIPC cluster **(Figure 2.2A)**. The distribution of p-values for all these clusters was also significantly different from similarly sized clusters of

scrambled cells **(Sup Table 2.5, Sup Fig 2.5)**. Unbiased clustering can also be influenced by the percentage of reads in each cell related to important cellular processes, like cell cycle phase and transcription, especially when other sources of variation are absent (Acosta *et al.*, 2017; Freytag *et al.*, 2018). When we removed other cell types, several clusters formed solely around these processes, and we did not include these clusters in our nIPC subtype/state analysis. One cluster, labeled Archetypic, only had one marker gene that passed our significance threshold and no evident similarity to another cluster to enable combination. Cell cycle, mitochondrial and ribosomal genes did not appear to influence identity of this cluster, but the p-values of its marker genes were not significantly different from scrambled clusters, leading us to believe that this cluster may be forming because of technical noise. We report it here to inform future single cell studies.

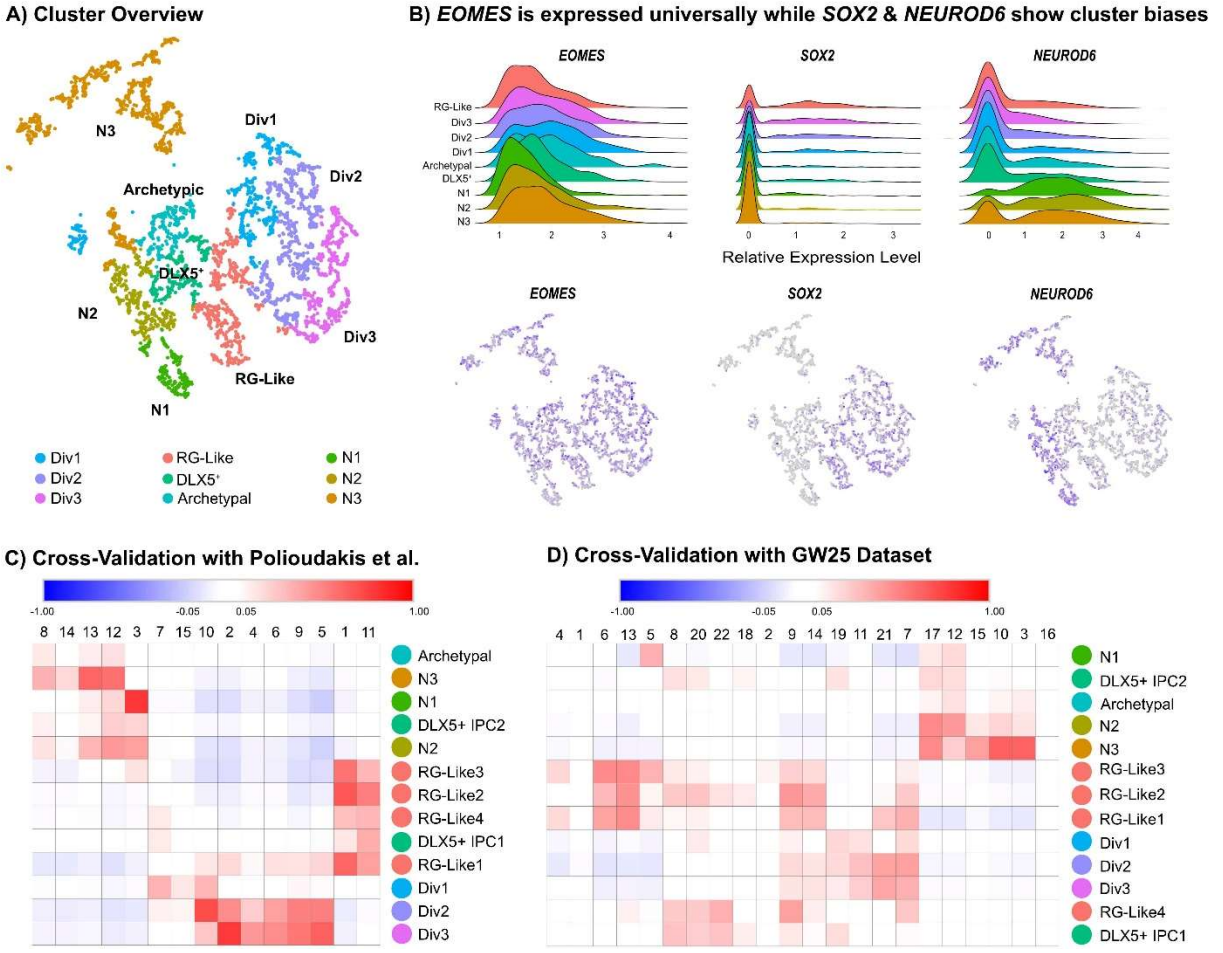


Figure 2.2. Bioinformatic Controls and Reproducibility A) Overview of high-quality clusters in tSNE space. “Div” clusters are the dividing clusters. nIPC clusters (N1, N2, & N3) are biased for neuron genes. Archetypal nIPCs express classic nIPC genes, but do not express unique identifying markers. This cluster did not pass the statistical threshold for a biologically meaningful cluster. B) Within *EOMES*-expressing nIPCs, *SOX2* and *NEUROD6* clearly differentiate clusters and guide interpretation. *SOX2* is biased towards RG-Like and Dividing Clusters, while *NEUROD6* is biased towards neuron-like (N1-N3 nIPC) clusters. C). Many clusters cross-validated well with specific clusters of *EOMES*-expressing cells from a previously published GW17-GW18 dataset. N3 and N1 nIPCs validated particularly well. D) Several clusters showed a near 1-to-1 relationship with nIPC clusters from a separate validation dataset.

All five nIPC subtype clusters expressed roughly similar amounts of *EOMES* **(Fig 2.2B)**.

SOX2 and *NEUROD6* were inversely related across the clusters as expected **(Fig 2.2B)**. RG-Like nIPCs specifically expressed *SOX2*, while neuron-likenIPC clusters

were marked broadly by neuron genes, including *NEUROD6* **(Fig 2.2B)**. When we analyzed other single cell datasets (Polioudakis *et al.*, 2019), we found similar nIPC subtypes as measured by marker gene expression **(Fig 2.2 C-D)**. The reproducibility of individual clusters helped validate cluster identities. If clusters represent several unique transcriptomic subtypes, we would expect each cluster to correlate with a different set of clusters in other datasets. Instead, we observed near one-to-one correspondence across datasets, highlighting the robustness of our nIPC subtype classification across two other single-cell datasets from developing human cortex.

Validation of Intermediate Progenitor Cell Subtypes throughout Human Neurogenesis

The RG-like nIPCs maintained expression of many RG progenitor genes, including *VIM* (Pixley and de Vellis, 1984) and SOX2 **(Fig 2.4)**. Immunostaining for VIM and SOX2 revealed protein co-expression in a subset of *EOMES*$^{+}$ nIPCs, validating that these genes identify a transcriptomic subtype of RG-like nIPCs. Because of this persistent RG gene expression, we hypothesized that if the nIPCs were uniquely derived from one specific RG subtype they might also retain subtype-specific gene expression. For example, oRGs are marked by HOPX, while truncated RG are marked by NR4A1. However, while marker genes for RG subtypes were expressed by nIPCs, they did not uniquely mark subclusters of RG-Like nIPCs **(Sup Fig 2.6)**. This suggests that nIPCs derived from different RG subtypes may not have significantly different transcriptomes. Additionally, these RG-like nIPCs could be found in both the iSVZ and oSVZ across

multiple ages, further suggesting that multiple RG subtypes can produce RG-like nIPCs **(Fig 2.4)**.

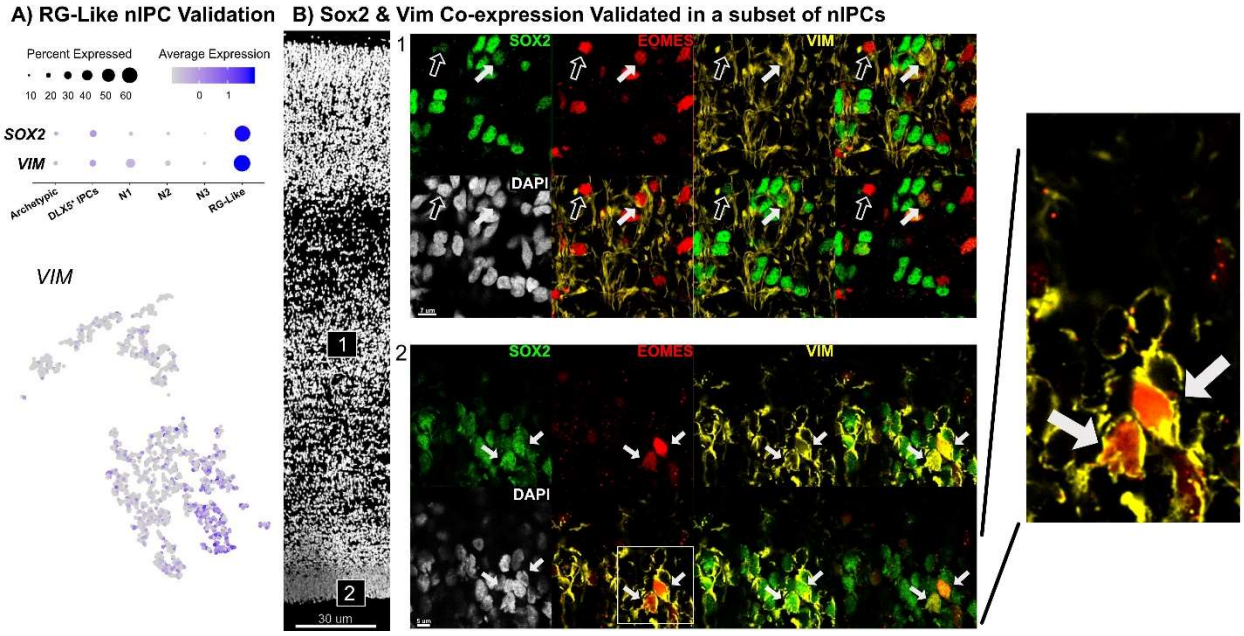


Figure 3. RG-Like nIPC Validation. A) RG-Like nIPCs co-express several classic RG genes including *SOX2 a*nd *VIM*. B) VIM and SOX2 protein are colocalized with a subset of nIPCs in the iSVZ and the oSVZ. VIM, SOX2, and EOMES triple positive cells are marked by filled-in arrows. EOMES nuclei are marked by empty arrows.

In contrast to the uniform cluster of RG-like nIPCs, three discrete types of nIPCs emerged from the dataset, each with a near one-to-one relationship with nIPC clusters from other published datasets (Polioudakis *et al.*, 2019) **(Fig 2.2 C-D)**. These three types, N1, N2, and N3 nIPCs, were uniquely enriched for DNM3, NHLH2, and PPP1R17 respectively. These markers, along with many others, formed gene sets unique to each cluster **(Sup Table 2.2)**. We validated each subtype via immunohistochemistry and found that each marker was expressed in a subset of EOMES^{+} cells **(Fig 2.3)**. PPP1R17 marked the cell bodies of some nIPCs. DNM3, a dynamin motor protein,

marked a subset of EOMES^{+} cells. NHLH2, a neuronal transcription factor, was highly elevated in a subset of nIPC nuclei. Taken together, these analyses suggest that each of these three nIPC clusters could be considered as a discrete biological subtype. All three subtypes were broadly visible across the iSVZ and the oSVZ **(Fig 2.4)**. However, within the iSVZ, DNM3 and PPP1R17 nIPCs showed a laminar distribution **(Fig 2.4 B, F)**. nIPCs closer to the ventricle were low or negative for both markers. In contrast, EOMES^{+}/NHLH2^{+} cells were distributed across the iSVZ **(Fig 2.4 C)**. Some of the markers of nIPC subtypes are also expressed by neurons **(Sup Fig 2.4 & 2.7 C)**, consistent with the differentiation of neurons from nIPCs. For example, neurons in the cortical plate also express DNM3 (N1) and NHLH2 (N2). One nIPC subtype marker, PPP1R17 (N3), was uniquely expressed in the germinal zone **(Sup Fig 2.7 A)** where it was also found in cells that did not co-stain with EOMES **(Sup Fig 2.7 B)**. The presence of TBR1 in these PPP1R17^{+}/EOMES^{-} cells suggested that this population may represent neurons derived from N3 nIPCs **(Sup Fig 2.7 D)**. These results suggest that some nIPC subtype markers, like PPP1R17, may remain expressed in newborn neurons prior to migration out of the germinal zone.

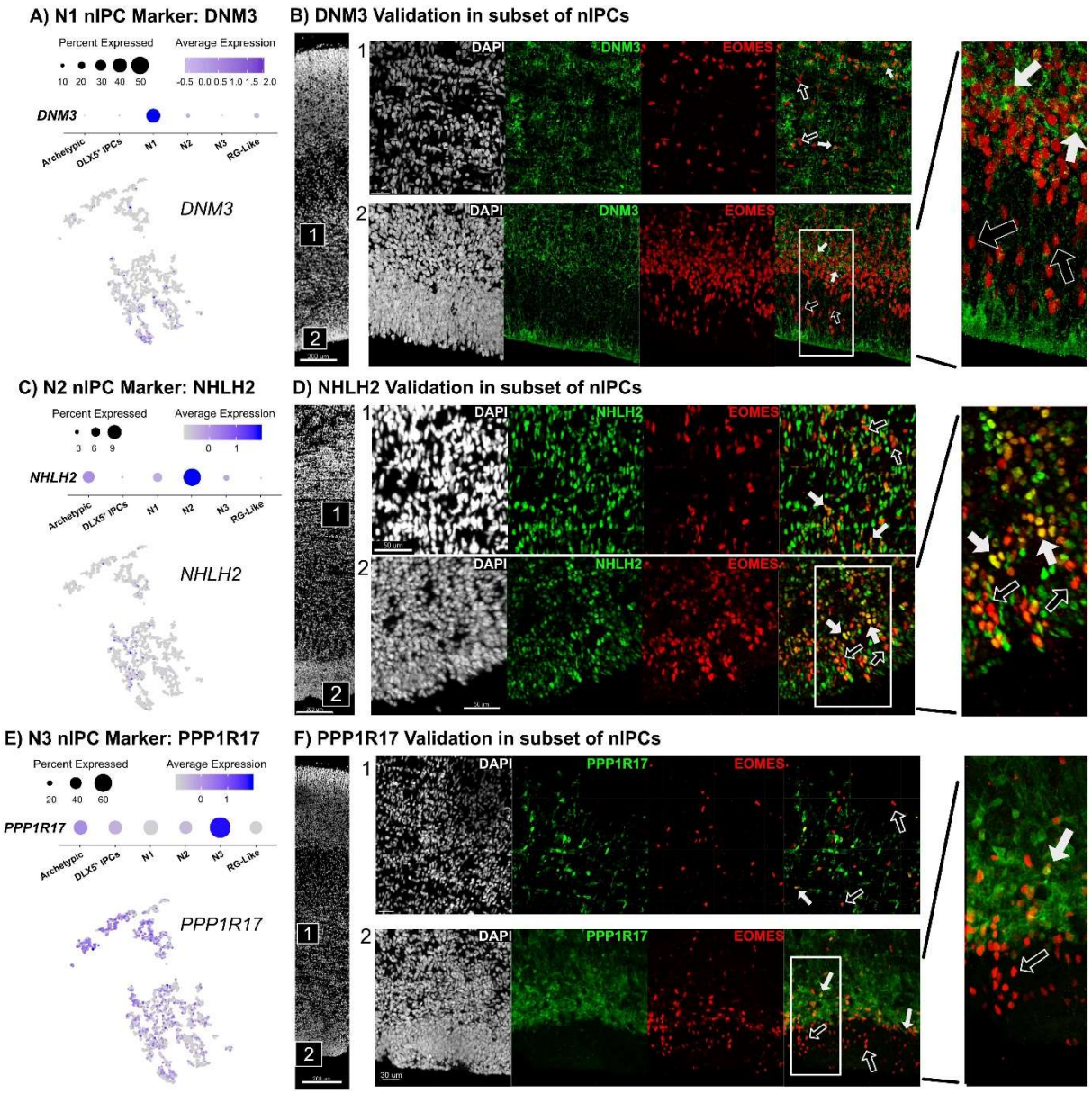


Figure 2.4. Neuron-like nIPC Validation. A) N1 nIPCs are marked by *DNM3*. B) DNM3 is expressed in a subset of nIPCs within both the iSVZ and oSVZ at GW14. Filled-in arrows represent colocalization, and empty arrows represent nuclei with only EOMES. This theme carries through the rest of the figure. C). N2 nIPCs are marked by strong *NHLH2* expression. D). NHLH2 is expressed in a subset of EOMES$^+$ nuclei in both the iSVZ and the oSVZ. E) N3 nIPCs are marked by strong *PPP1R17* expression. F). PPP1R17 can also be seen in a specific subset of EOMES$^+$ cells in both the iSVZ and the oSVZ.

While most nIPC subtypes were prevalent during early neurogenesis, DLX5^{+} nIPCs appeared in late neurogenesis **(Figure 2.5)**. When each age within our dataset was clustered independently, a *DLX5*$^{+}$ cluster was not evident before GW18, but was clearly distinct at GW22 **(Fig 2.5B)**. Likewise, we found that DLX5^{+} cells are present in a very small cluster in a previously published GW17-GW18 dataset (Polioudakis *et al.*, 2019) **(Fig 2.5 C)** while a large DLX5^{+} nIPC cluster is evident in a separate GW25 dataset **(Fig 2.5 C)**. When we evaluated DLX5 and EOMES co-expression, our samples from GW20 (n=3) and later had significant co-labeling **(Fig 2.5 D)**. We did not note such co-labeling, or similar ventricular staining between GW14 and GW18 (n=3). Noticeably, at GW20, we also observed a proportion of EOMES+ cells labeled by **GABA (Fig 2.5 E)**, suggesting that these progenitors may give rise to GABAergic neurons (see discussion). These results suggest that unique molecular programs may be active during late neurogenesis that may not be expressed at earlier stages.

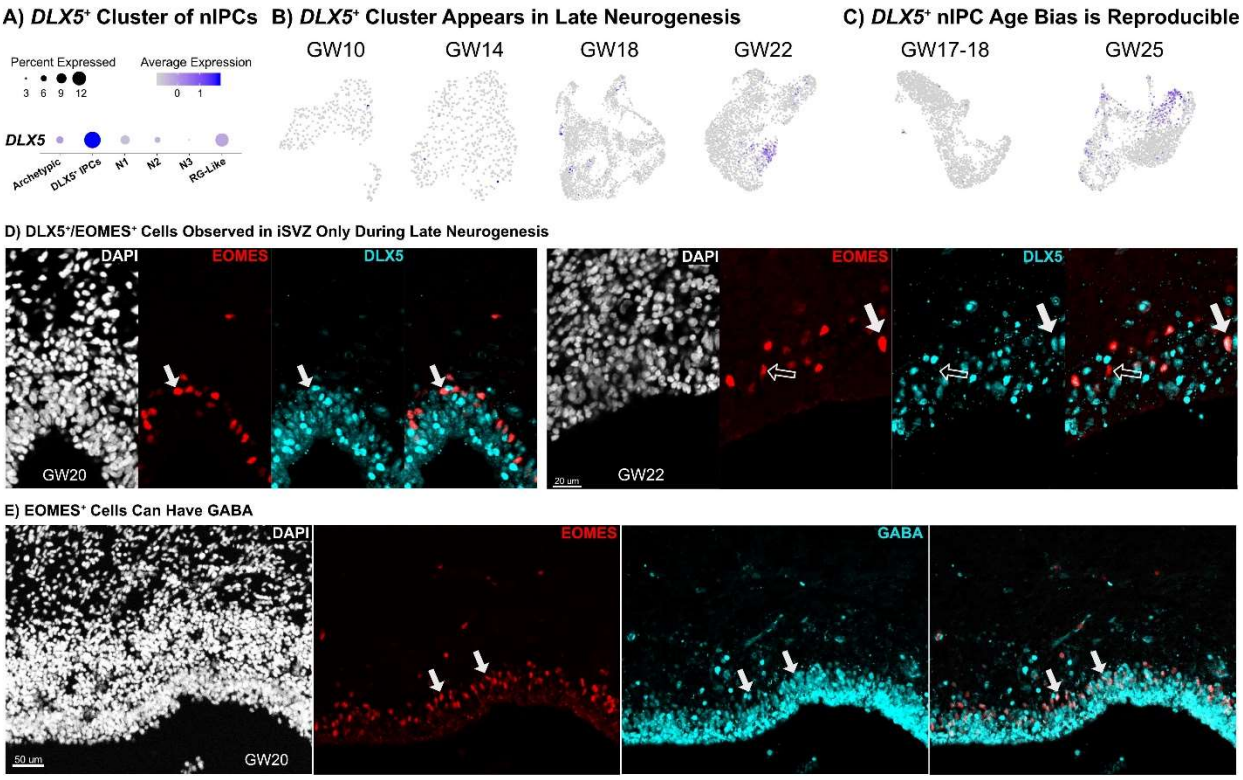


Figure 2.5. DLX5+ nIPC Validation. A) One cluster of nIPCs express *DLX5*. B) This DLX5$^+$ cluster appears at later ages in our dataset. The DLX5$^+$ cluster is absent at GW10 and GW14. By GW18, this is a small cluster to the left, and by GW22, there is a large cluster of DLX5$^+$ cells. C). This pattern repeats with other datasets. For the GW17-GW18 Polioudakis et al. dataset, a small cluster of DLX5$^+$ cells is visible. Our validation dataset (GW25) confirms a DLX5$^+$ population. C). DLX5 and EOMES co-stains a subset of EOMES$^+$ cells at GW20 and GW22. Filled-in arrows mark co-staining, and empty arrows mark EOMES staining. D) EOMES$^+$ cells at GW20 can also be GABA+, suggesting that inhibitory neurons may be born from these nIPCs at this time. Co-labeled cells are marked with filled-in arrows.

nIPC morphology does not correlate with transcriptomic subtypes

Many features of the developing nervous system were initially identified based on observations of cellular morphology as typified by the work of Ramón y Cajal (Ramón y Cajal, 1909). RG have been defined by their long radial extensions while subtypes of ventricular and outer RG are often differentiated by the presence or absence of apical

extensions. Transcriptomic differences have recently been related to RG morphology, and these features together help define RG subtypes (Nowakowski *et al.*, 2016; Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent, Kriegstein, *et al.*, 2017). In contrast, the morphology of nIPCs remains largely undefined. We sought to characterize the morphology of human nIPCs through viral labeling and immunohistochemical approaches **(Figure 2.6)**. We began by infecting organotypic human brain slices with Adenoviral-CMV-GFP to label dividing progenitors and identified nIPCs by post-staining with EOMES. Reconstruction of $EOMES^{+}/GFP^{+}$ cells revealed a wide range of morphologies across the iSVZ and oSVZ. Our results clearly indicate the $EOMES^{+}$ cells can have very long processes, up to 175 mm in length **(Sup Table 2.4)**. Cells with processes less than 30 mm we labeled as either apically-biased, basal-biased, or short multipolar, depending on the direction and number of extensions. Noticeably, we did not observe nIPCs with only a single long basal process, and only one cell out of 53 was basal-biased. However, we did observe several nIPCs with long apical or bipolar extensions **(Fig 2.6 B)**, similar to cells that have been labeled as RG in previous publications (Betizeau, Cortay, D. D. E. Patti, *et al.*, 2013). Out of 53 total nIPCs, five had no processes, and we observed similar numbers of apical or bipolar nIPCs, suggesting that nIPCs are just as likely to have long processes as they are to have none (spherical morphology). In contrast, short multipolar and horizontal nIPCs constituted more than half of nIPCs **(Fig 2.6 C)**. To control for alterations that could result from organotypic slice culture or viral labeling, we sought to identify nIPC morphologies in thin sections of fixed tissue stained for EOMES and

PPP1R17, a marker that labels the cell bodies of N3 nIPCs. Among EOMES^{+}/PPP1R17^{+} cells we observed horizontal as well as apical and bipolar morphologies, corroborating our previous observations **(Fig 2.6 D)**. As with our viral labeling approach, we did not find obvious examples of basal PPP1R17+ nIPCs. Our PPP1R17 staining also suggests that morphologies do not correspond to transcriptomic subtypes, since N3 nIPCs could be observed with multiple different morphologies **(Fig 2.6 D)**. The large morphological diversity across individual nIPCs suggests that caution should be exercised when identifying cell types based on morphology alone.

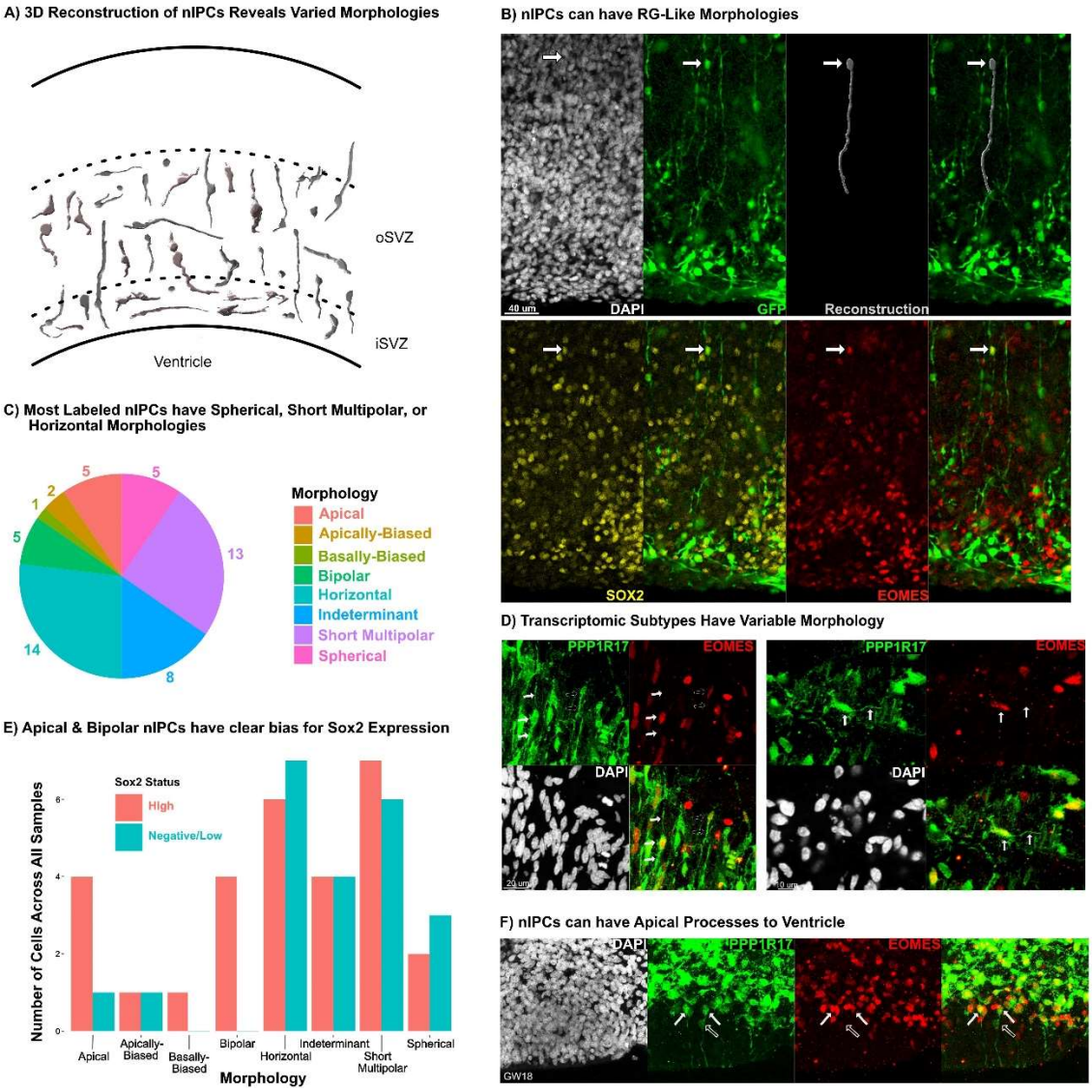


Figure 2.6. nIPC Morphology does not correlate well with transcriptomic subtypes. A) A representation of reconstructed nIPCs, based on GFP viral labeling and EOMES co-staining. B). nIPCs can have very long extensions that resemble radial glia. C) Most labeled nIPCs either had no extensions (spherical), had minimal extensions (less than 30 um), or had horizontal morphologies with long extensions that ran parallel to the VZ. D) PPP1R17^{+}/EOMES^{+} cells have varied morphologies with both apical and basal extensions (filled-in arrows) and with only apical extensions (empty arrows). Horizontal morphologies can also be found. F) nIPCs can contact the ventricle. EOMES^{+} cells with low PPP1R17 expression can have extensions to the ventricular edge. We observed this staining in all samples observed from GW14 to GW18.

To further investigate the possible correspondence between transcriptomic subtypes and morphology we quantified the SOX2 expression status of our reconstructed nIPCs shown in Figure 2.6A. SOX2 marks RG-like nIPCs in humans, and we found that apical and bipolar nIPCs had a clear positive bias for high SOX2 expression. However, Sox2-high cells also had other morphologies, and at similar numbers to Sox2-low/negative cells **(Fig 2.6 E)**. This wide range of observed morphologies suggests that transcriptomic nIPC subtypes may not be reliably identified based on their morphology alone. Lastly, we identified EOMES^{+} cells with low levels of PPP1R17 in the VZ/iSVZ. These cells were particularly common between GW14 and GW18, and had apical projections extending to the ventricle, a feature characteristic of vRGs **(Fig 2.6 F)**. These nIPCs could potentially be influenced by factors within the ventricle or the ventricular niche. Our findings suggest that apical surface contact, long apical processes, or long bipolar processes cannot be used as the sole basis for identifying radial glia as these features are also observed across multiple nIPC types.

Discussion

Studies of the progenitor cells underlying human cortical neurogenesis have largely focused on RG cells. Indeed, models of corticogenesis center around spatiotemporal, transcriptomic, and morphological diversity of the RG population, of which there appear to be at least three subtypes in the developing human brain (Betizeau, Cortay, D. Patti, *et al.*, 2013; Pollen *et al.*, 2015; Thomsen *et al.*, 2015; Nowakowski *et al.*, 2016; Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent,

Kriegstein, *et al.*, 2017). However, neuronal output during neurogenesis also depends significantly on the role of nIPCs. Defining the identity and diversity of nIPCs will be critical to understanding the production of neuronal diversity during normal development as well as alterations to neural populations associated with neurodevelopmental diseases. A comprehensive atlas of human nIPC cell subtypes can help resolve lineage relationships of neuronal and glial cell types and may clarify the intermediate steps by which RG influence neurogenesis and corticogenesis. In our study, we observe RG-like and N1-N3 nIPC subtypes in both vRG and oRG niches, the iSVZ and oSVZ respectively, during early neurogenesis. After GW19/GW20, we observe that the oRG niche becomes depleted of nIPCs, with only small numbers remaining in the iSVZ. We find a clear difference in EOMES between the iSVZ and the oSVZ in all samples at all ages, with a clear laminar pattern in the iSVZ, regardless of age, and an abrupt decrease in EOMES across the oSVZ at the end of neurogenesis. Previous studies have also found a shift at this time, but with a different pattern of EOMES staining. Malik et al reported a diffuse, and relatively unstructured pattern across the SVZ, with little to no structural difference between the iSVZ and oSVZ (Malik *et al.*, 2013). Several experimental details may explain the difference in observation, as Malik et al relied on tissue collected after a much longer post-mortem interval, from patients suffering from serious premorbid conditions including sepsis, heart failure, respiratory failure, and necrotizing enterocolitis. In contrast, our samples were collected with a post-mortem interval of 10-20 minutes. It is difficult to ascertain whether the observed differences result from greater post-mortem interval, a biologically important stress response, or other experimental differences. We did not observe the diffuse, late neurogenic staining

pattern of Malik et al. using two different EOMES antibodies, and multiple neuronal markers that were expressed earlier in the germinal zone. Additionally, we see a concordance between our staining pattern and the single cell data, with N1, N2, and N3 nIPC clusters enriched for early nIPCs and depleted of late neurogenic nIPCs (after controlling for sample size). Since we observe a dramatic shift in the distribution of nIPCs, late neurogenesis may be a developmentally distinct period from early neurogenesis. We find four nIPC subtypes that dominate early neurogenesis across both the iSVZ and oSVZ and a fifth transcriptomic subtype of nIPC unique to late neurogenesis. These five nIPC subtypes were validated using three independent approaches: internal statistical controls, bioinformatic cross-validation with two other datasets, and immunohistochemistry. The primary transcriptional features appeared to be maturity, with one RG-like IPC, and four more neuron-like nIPC subtypes . The transcriptional features of more neuron-likenIPCs may be evidence that neuron subtype identity starts emerging at the nIPC stage. For example, NHLH2 marks both N2 nIPCs, and a specific subset of excitatory neurons in the cortical plate (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent, Kriegstein, *et al.*, 2017). However, the neuronal lineages of N1 and N3 are not as clear, since DNM3 appears to be broadly expressed in the cortical plate, and PPP1R17, an excitatory lineage gene, is only expressed by IPCs and newborn excitatory neurons in the germinal zone **(Sup Fig 2.7 A,D)** (Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent, Kriegstein, *et al.*, 2017). It is tempting to speculate that these two broader subtypes (N1,

N3) may relate to the broader categories of upper layer intracortical and deep layer subcortical projection neurons. It is also possible that the RG-like nIPCs produce the more neuron-likeN1, N2 or N3 nIPC subtypes. However, significant follow-up work will be needed to test this hypothesis and elucidate the exact lineage of N1 and N3 nIPCs. Unlike RG, the transcriptomic nIPC subtypes do not appear to be correlated with a particular morphological feature. However, nIPC morphologies do overlap with some morphological patterns typically seen in RG. Our observations help clarify the range of molecular and morphological properties of nIPCs and can be used to identify the functions of nIPC subtypes in future studies.

While the majority of nIPCs are found in the iSVZ, a significant number of nIPCs appear in the oSVZ in early neurogenesis, where we observed higher rates of Ki67+ labeling suggesting high proliferative activity. These results are consistent with previous observations of nIPCs dividing multiple times unlike their mouse counterparts that usually undergo one division (Noctor *et al.*, 2004; Hevner, 2006). These observations suggest that oSVZ nIPCs may be more highly proliferative than their iSVZ counterparts. Surprisingly, this observation was the only clear molecular difference between nIPCs in the iSVZ and the oSVZ, despite evidence that these nIPCs arise from molecularly and morphologically distinct RG subtypes. The patterns of gene expression did not suggest a relationship between a particular nIPC subtype and an RG subtype, though we cannot definitively rule out a possible lineage association. Most transcriptomic nIPC subtypes (RG-like and N1-N3 nIPCs) were found across both the iSVZ and oSVZ. The only spatial bias appears to be with N1 and N3 nIPCs, both of which showed preferential

location within the iSVZ. This pattern changed after GW19, when the spatial distribution of nIPCs dramatically shifts. At this point, a novel DLX5-expressing nIPC appears **(Fig 2.5 B-C)**. The absence of nIPCs in the oSVZ and the minimal number of nIPCs in the iSVZ at these later timepoints suggests a change in neurogenesis and coincides temporally with the onset of gliogenesis (Semple *et al.*, 2013). The near absence of excitatory neuron markers in the proliferative zones at this age supports this notion **(Sup Fig. 2.2)**. These observations are also consistent with the model that excitatory neurogenesis is largely dependent on EOMES but suggests that the inverse may not be true. EOMES might be capable of promoting inhibitory neuronal identity as well. The presence of GABA within the iSVZ, and within EOMES$^+$ cells specifically, suggests that GABAergic neurons may arise from nIPCs in the cortex at late stages of neurogenesis. These results are consistent with a previous report of rare DLX$^+$ and GABA$^+$ radial glia in the human dorsal cortex during late neurogenesis (GW20) (Yu and Zecevic, 2011). The exact nature of this late neurogenic period, whether locally produced GABAergic neurons are destined for the cortex or possibly for the olfactory bulb, and the noncanonical role of EOMES and DLX5 will be interesting to explore.

Here we have conducted an analysis of nIPC morphotypes and attempted to relate them to molecular subtypes. Unlike RG, we find that nIPC morphologies do not relate well to transcriptomic subtypes and may exhibit RG-like shapes, such as long apical and bipolar extensions. However, long range basal processes may be unique to oRG cells, since we did not observe this morphological feature in nIPCs. We frequently observed EOMES+ cells with these morphologies, suggesting caution when classifying

progenitor cells by these features (Betizeau, Cortay, D. D. E. Patti, *et al.*, 2013). The clear bias for high SOX2 expression in morphotypes with long processes indicates that RG-like nIPCs may preserve a RG-like morphology. However, roughly half of our spherical EOMES$^+$ cells were RG-like nIPCs based on SOX2 expression. The diversity of RG-like nIPC morphologies may help explain observations from Betizeau et al, where RG-like cells with long bipolar or apical extensions could be produced by cells without processes that were defined as nIPCs (Betizeau, Cortay, D. D. E. Patti, *et al.*, 2013). The authors also observed cells with bipolar morphology that retracted their processes entirely. Importantly, the above patterns were only true for bipolar and apical morphologies, not basal ones. These observations correlate well with our observations of nIPC morphologies. It is highly likely that RG-like nIPCs with apical and bipolar processes can dynamically change shape. RG-like nIPCs may convert from bipolar or apical to spherical and produce additional RG-like nIPCs which then extend long, RG-like processes. The potential for such dynamic activities requires future study. Additionally, the fact that we observed long horizontal processes in Sox2 low/negative IPCs suggests that long processes are not unique to RG-like nIPCs. To validate that nIPC subtypes could take on a wide variety of shapes *in vivo* we stained for PPP1R17, which marks the cell bodies of nIPCs and their processes. PPP1R17 is also not highly expressed by RG-Like nIPCs **(Sup Fig. 2.5)**. We observed PPP1R17$^+$ nIPCs with a wide range of morphologies, corroborating our viral-labeling in organotypic cultures. Additionally, we found that nIPCs can have extensions to the ventricle, a trait traditionally linked to RG. These results indicate that RG morphological properties

alone, including apical extensions and long basal processes, cannot be used to differentiate between RG and nIPCs.

Within the RG population, spatial, molecular, and morphological diversity converge upon clear cellular subtypes. For nIPCs however, our results indicate that these defined axes of variation do not similarly converge. While transcriptomic subtypes can be found, they do not relate well with germinal zone niche or morphological features. We also found a dramatic temporal shift of nIPC diversity as neurogenesis shifts to gliogenesis, altering the distribution of nIPCs in the germinal zones and correlating with the appearance of a novel nIPC subtype. These results provide a dynamic roadmap of changing nIPC diversity and characteristics during human brain development and will help inform future studies of the mechanisms by which nIPCs influence neurogenesis.

Methods

Bioinformatics analysis

Our main dataset was published by Bhaduri *et al.*, 2020. All initial clustering and marker identification were conducted in Seurat version 2. When noted, some subsequent analysis was conducted in Seurat version 3. Cells were QC by less than 5% mitochondrial input. We filtered for all cells that expressed *EOMES* above 0.25 after normalization. The comparison dataset was a GW17-GW18 dataset from Polioudakis *et al.*, 2019. After filtering, the cells were scaled and batch corrected according to methods previously described (Bhaduri *et al.*, 2020). We ran the FindAllMarkers command and

correlated the specificity Using this correlation matrix and our clustering of scrambled data, we combined these initial clusters heuristically. The final clusters included one defined entirely by ribosomal genes, and two clusters defined by mitochondrial genes. Three clusters were defined by cell cycle genes and scored as either G2/M or S phase cells by Seurat's cell cycle scoring function. The remaining clusters were considered high quality G1-phase clusters, and tentatively representative of biological subtypes.

Clusters and Cluster Marker Generated on Scrambled Data

Both technical noise and biological variance can drive single cell clustering, which *a priori* biases differential gene expression calculations. We found that even scrambled data could generate statistically significant differentially expressed cluster markers, and this p-value may vary according to cluster size. To evaluate the significance of differentially expressed marker genes, we created a significance threshold of each cluster according to size by generated a distribution of marker p-value scrambled data. Cluster markers were evaluated by their corresponding cluster's p-value threshold. In more detail, the original gene expression matrix was scrambled 1000 times by gene across single cells. This scrambling preserved the distribution of individual gene expression but scrambled any coordinated gene expression that might mark biological nIPC subpopulations. For each iteration, clustering was performed at varying resolution to generate a range of cluster sizes. For each iteration and resolution, we recorded the cluster sizes, and identified positively expressed cluster markers using the Seurat command FindAllMarkers. To generate our comparison distribution of p-values, we identified all clusters that were within 50 cells of each original cluster size, and for each iteration, we found the resolution for which the most clusters were generated. Using this

filtered maker list, we generated a null distribution of false positives using significant ($p < 0.05$) Benjamini-Hochberg corrected p-values. After a negative-log transform, we used the 95th quantile of these p-values as our significance threshold. A robust number of significant markers were found in all our final clusters, with exception of Archetypic nIPCs **(Sup Table 2.2)**. Given that clustering could be influenced by technical noise, we validated the significance of each cluster by comparing its p-value distribution to p-values generated from similarly-sized clusters of scrambled data, as discussed previously. After conducting an unpaired Mann Whitney U test without continuity correction, all but one cluster (Archetypic) had marker p-values that were significantly different from scrambled dataset **(Sup Table 2.3)**. Code for this work can be found at https://github.com/MPPebworth/HumanIPCs.

Viral Labeling of Slice Culture & Analysis

Organotypic slices of developing human brain tissue were prepared and infected with CMV::GFP adenovirus (Vector Biolabs) as previously described(Hansen *et al.*, 2010; Bhaduri *et al.*, 2020). Four to five days after viral infection, the slices were fixed in 4% PFA for 1 hour at room temperature, and washed 2x in PBS. Slices were then stained according to previously described protocol with antibodies for GFP (1:500, AB13970 Abcam, RRID: AB_300798), Sox2 (1:300, sc-365823 Santa Cruz Biotechnologies RRID: AB_10842165), and EOMES (1:250 AF6166 R&D Systems RRID: AB_10569705) (Hansen *et al.*, 2010).

Morphological Analysis

For morphological analysis, we identified nIPC morphologies based on the number, direction, and length of extensions. When we could track extensions farther than 30 um towards the basal surface, apical surface, or both surfaces, we designated the nIPC morphology as either basal, apical, or bipolar, respectively. If extensions were less than 30 um apically, basally, or directionally, morphologies were labeled as either apically-biased, basally-biased, or as short multipolar. Cellular extensions were measured until they passed by other GFP+ cells and extensions and could not be reliably tracked further. As such, our measurements represent a conservative estimate of nIPC projection length. Some nIPCs had no evident extensions and were labeled as spherical. Finally, for some nIPCs, we could only reconstruct the morphology in one direction because of the presence of other cells, and these cells were labeled as indeterminant. 3D cell reconstruction were done with the Imaris image analysis software.

Immunohistochemistry

Developing human brain tissue was prepared for sectioning as previously described (Hansen *et al.*, 2010; Nowakowski, Bhaduri, Pollen, Alvarado, Mostajo-Radji, Di Lullo, Haeussler, Sandoval-Espinosa, Liu, Velmeshev, Ounadjela, Shuga, Wang, Lim, West, Leyrat, Kent, Kriegstein, *et al.*, 2017). Slices were stained with primary antibodies for SOX2 (1:300, sc-365823 Santa Cruz Biotechnologies, RRID: AB_10842165), PPP1R17

(1:150 HPA047819 Atlas Antibodies/Sigma Aldrich, RRID: AB_2680166), EOMES (1:250 AF6166 R&D Systems, RRID: AB_10569705) , DNM3 (1:100 GTX23458 GeneTex, RRID: AB_385058), NHLH2 (1:100 HPA055238 Sigma Aldrich, RRID: AB_2682752), DLX5 (1:100 HPA005670 Sigma Aldrich, RRID: AB_1078681), VIM (1:250 AB5733 EMD Millipore, RRID: AB_11212377), SATB2 (1:250 AB92446 Abcam, RRID: AB_10563678), TBR1 (1:100 20932-1-AP Proteintech, RRID: AB_10695502) and NEUROD1 (1:100 AB60704 Abcam, RRID: AB_943491). For TBR1, Perkin Elmer's TSA kit (RRID: AB_2572409) was used to amplify signal, following methods from previously published work (Raju *et al.*, 2018).

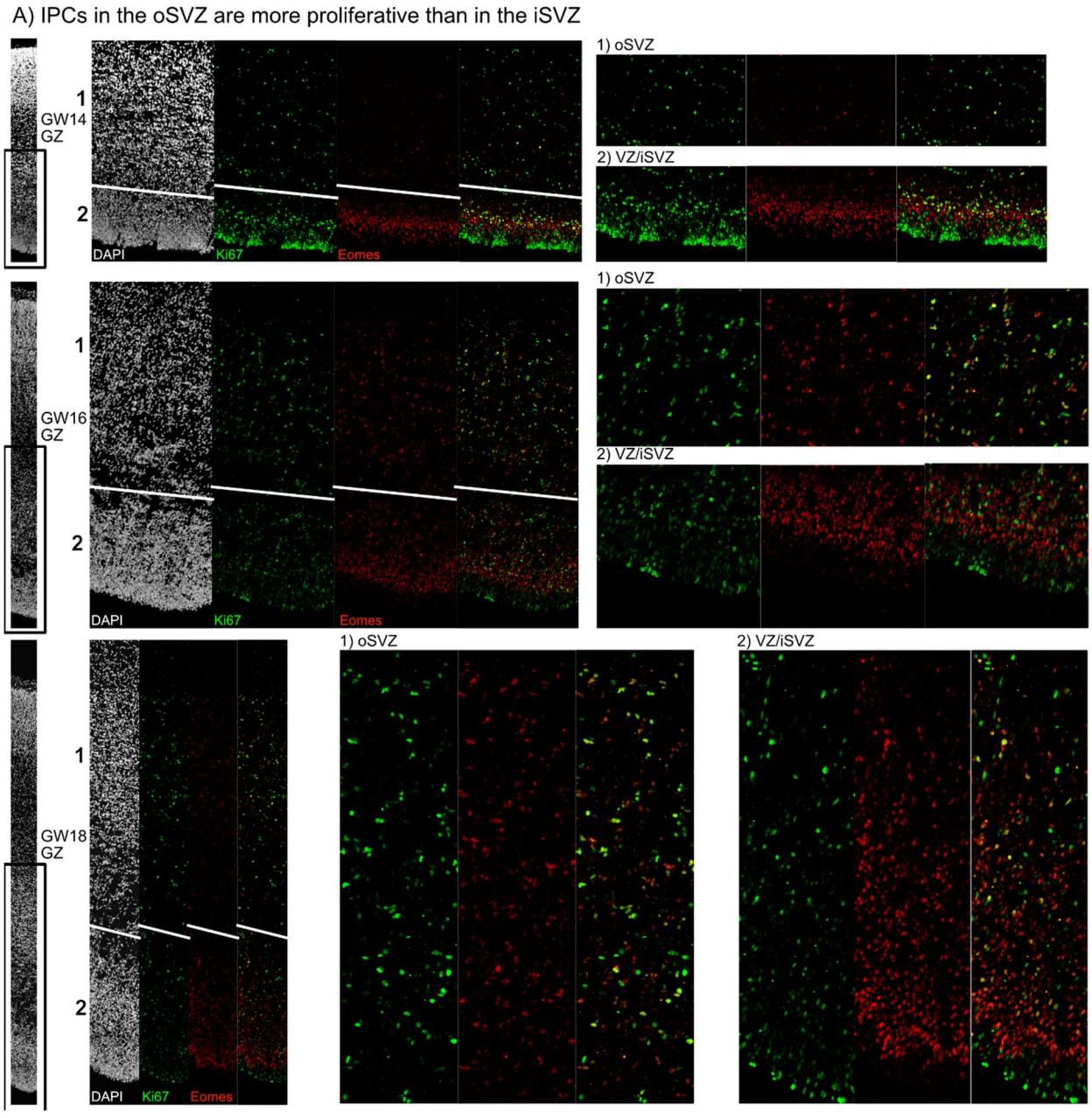
Tissue Collection

As previously published, developing human brain tissue was obtained and processed according to guidelines approved by the UCSF Human Gamete, Embryo and Ste Cell Research Committee (GESCR, approval 10-05113) (Bhaduri *et al.*, 2020). Patient consent was obtained before the collection of all samples used in this study. Samples were collected within 10-20 minutes of elective termination at the San Francisco General Hospital, and placed in fresh, pre-oxygenated artificial cerebral spinal fluid on ice for preservation until fixation later that day (<8 hours).

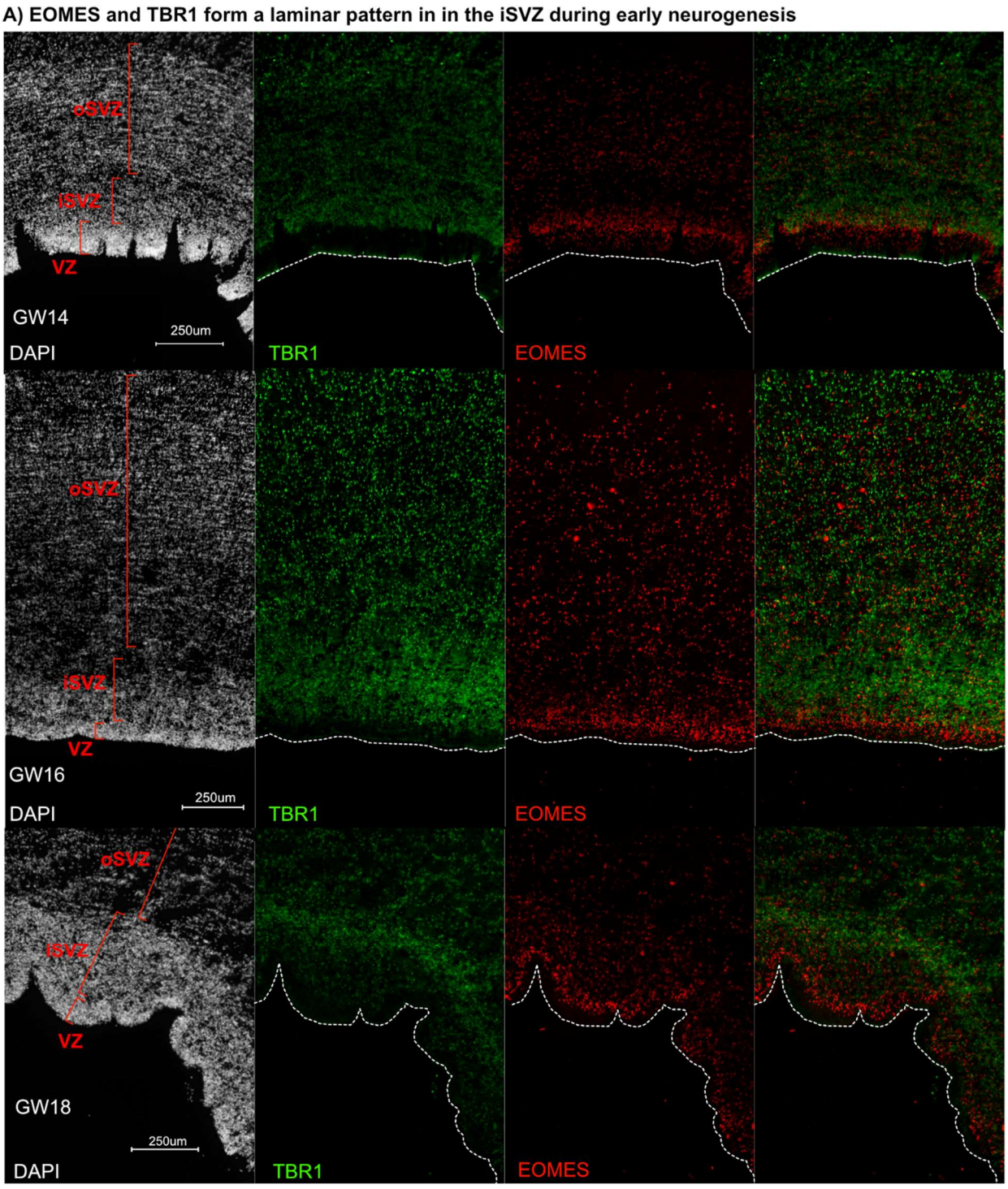
Microscopy

Confocal imaging was conducted using Leica SP5 and Leica SP8s from the microscopy core for the Eli & Edythe Broad Center for Regeneration Medicine and Stem Cell Research. Large tile scans were either taken using similar confocal microscopes, or a Keyence BZ-X from the same core facility.

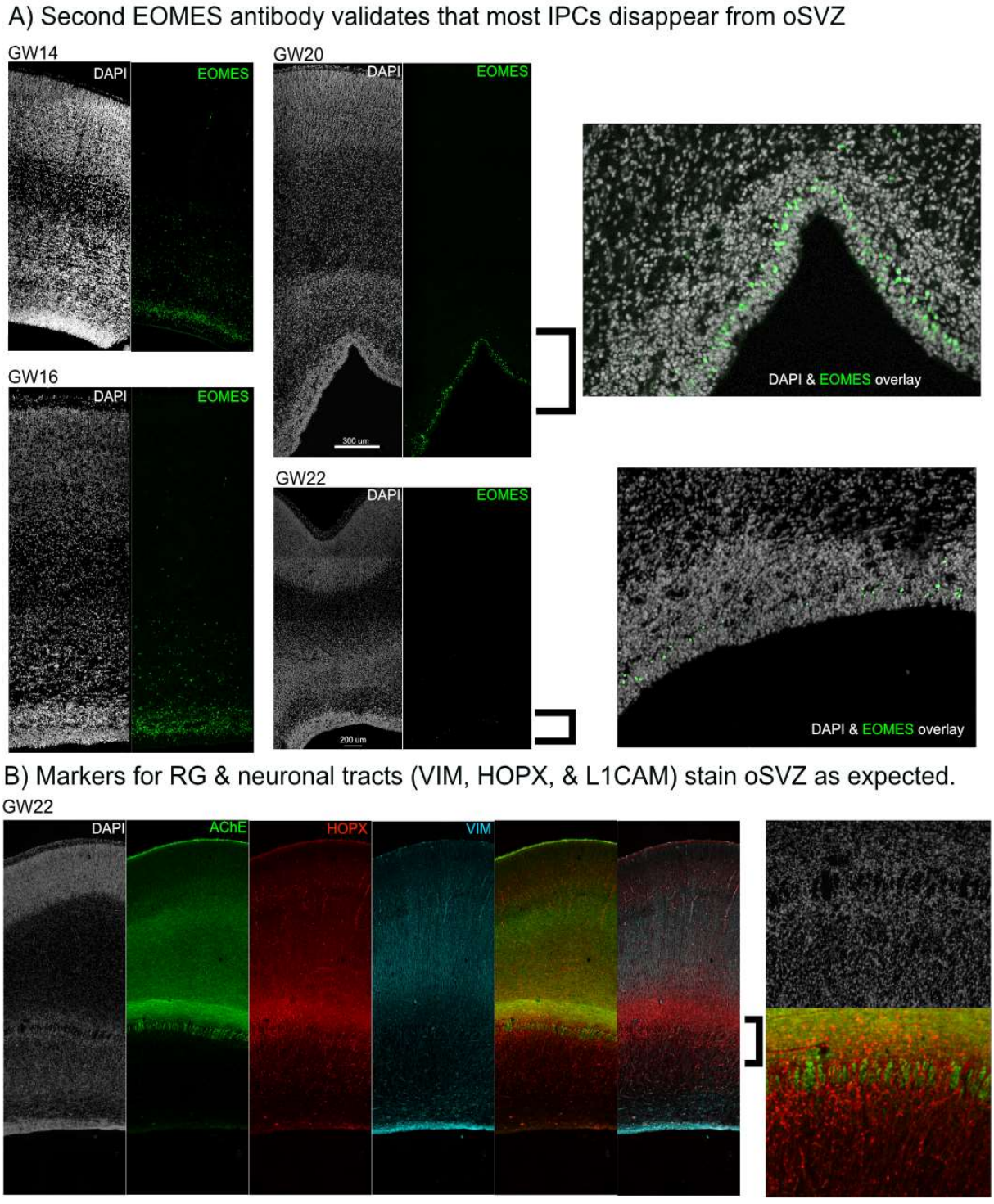
Supplemental Figures



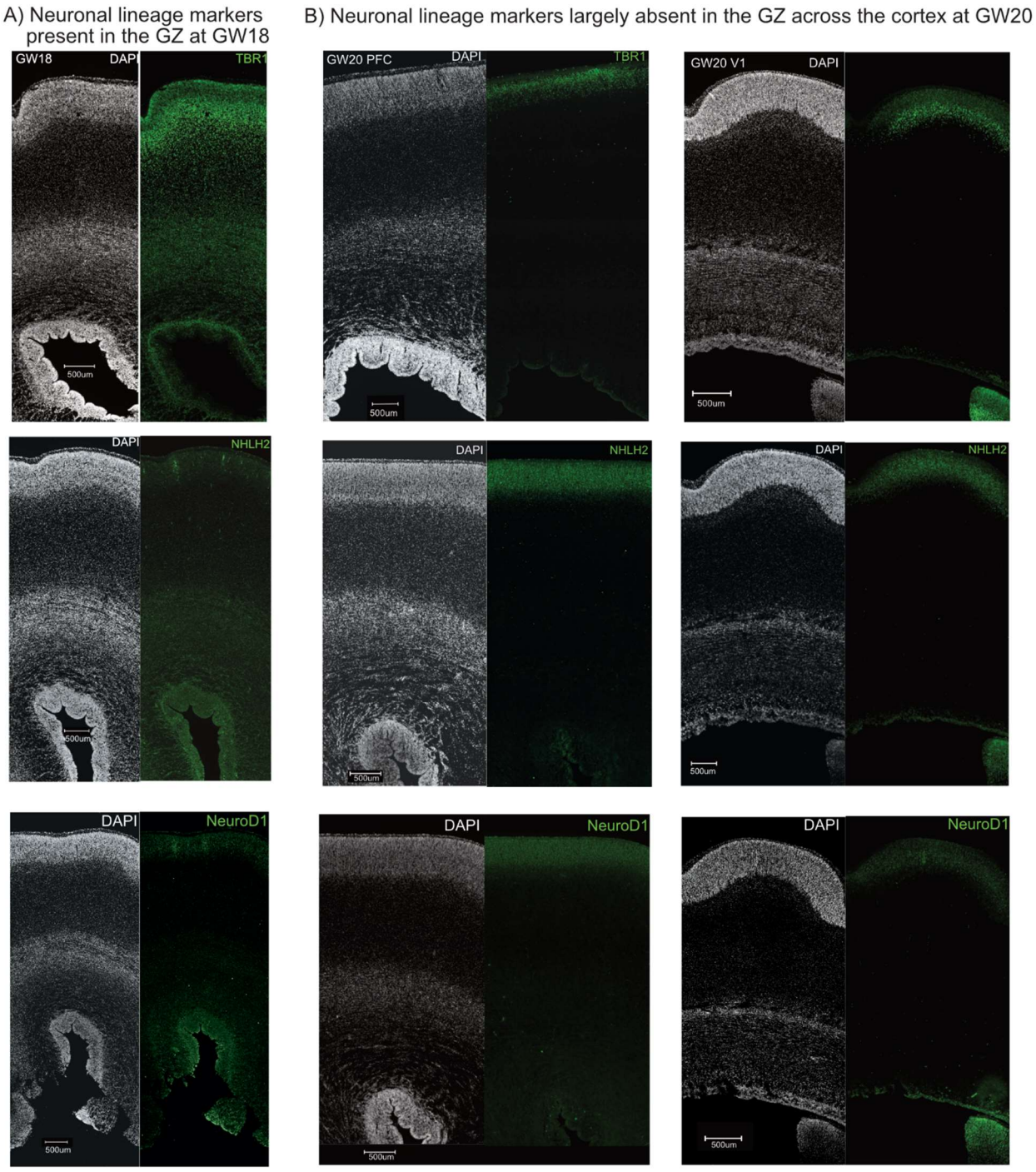
Supplemental Figure 2.1: IPCs in the oSVZ appear more proliferative than those in the iSVZ during early neurogenesis. EOMES nuclei co-stain with KI67 more often in the oSVZ than the iSVZ. Images are from GW14, GW16, and GW18 and are the images quantified in Fig 1 C of the main text.



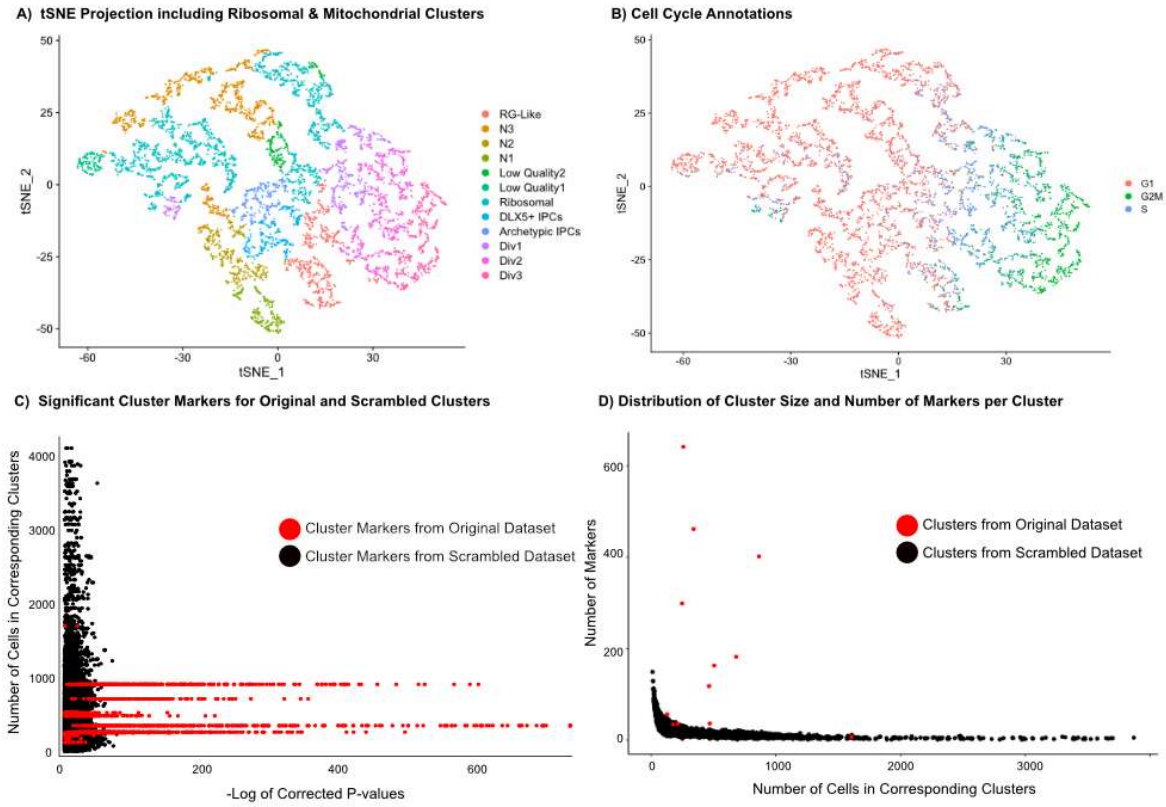
Supplemental Figure 2.2: EOMES & TBR1 immunohistochemistry reveals a laminar pattern of transcription factor expression in the iSVZ during early neurogenesis. This evidence suggests that most iSVZ IPCs differentiate without migrating into the oSVZ.



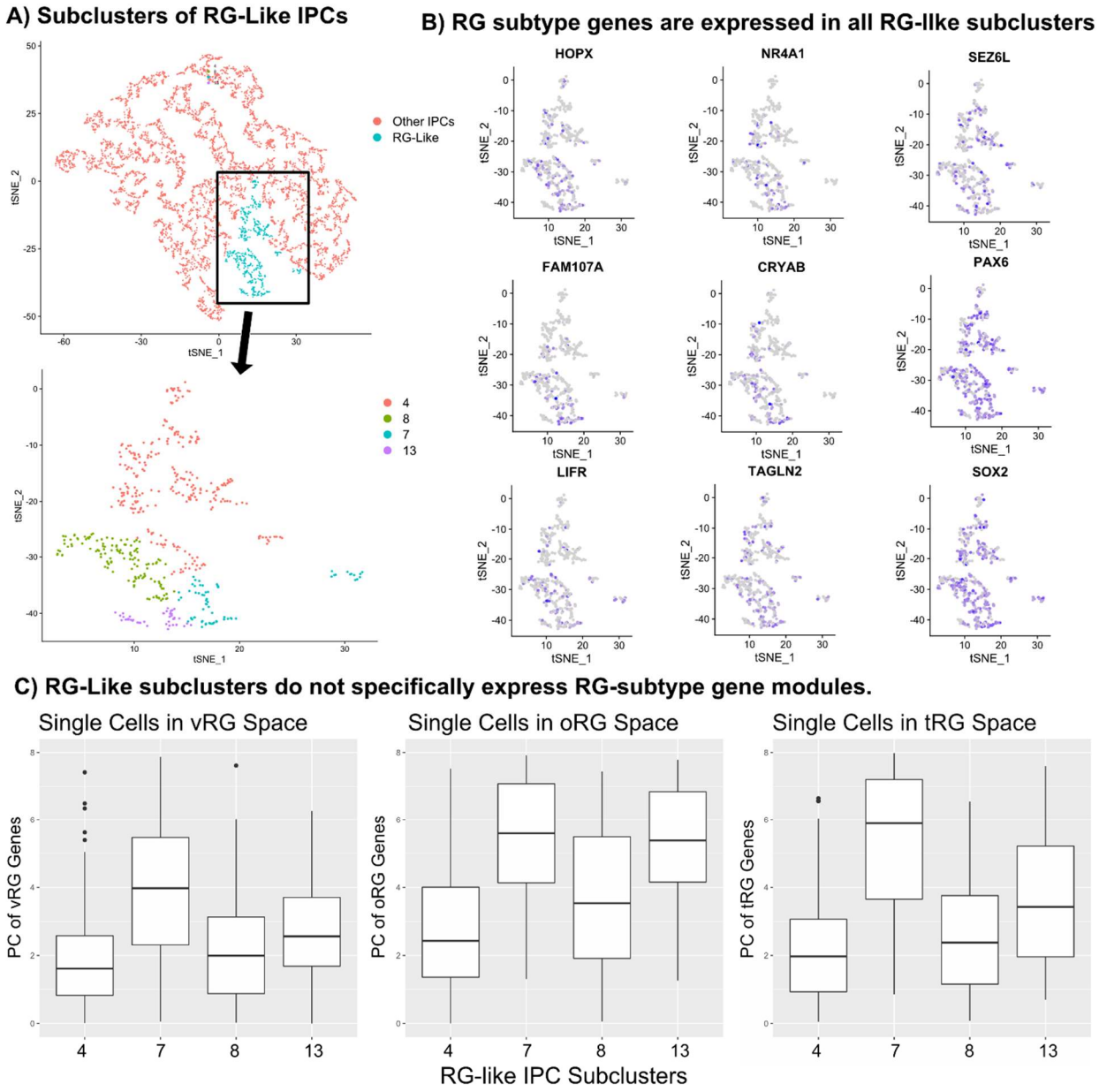
Supplemental Figure 2.3: Further verification of EOMES staining pattern during late neurogenesis. A) Second EOMES antibody validates that most IPCs disappear from oSVZ. Staining shows GW14, GW16, GW20 and GW22 staining with a second EOMES antibody. B) Markers for RG & neuronal tracts (VIM, HOPX, & L1CAM) stain oSVZ as expected. Cytoarchectiture of oSVZ confirmed to be normal for samples.



Supplemental Figure 2.4: Staining for neuronal markers support disappearance of neurogenesis from the oSVZ during late neurogenesis. A) Markers expressed in the neuronal lineages (NeuroD1, NHLH2, TBR1), and which begin in IPCs, are expressed throughout the GZ at GW18. B) By GW20, the same neuronal markers are no longer expressed throughout the germinal zones, and tend to be found in the cortical plate, or in the iSVZ.

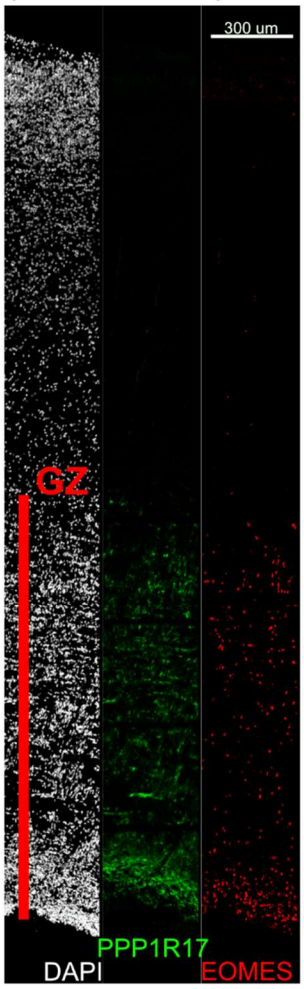


Supplemental Figure 2.5: The original dataset and its comparison with the clusters and markers from 1000 scrambled iterations. A) Full clustering and annotations from the original dataset. Several clusters were marked by mitochondria genes and were annotated as low quality. One cluster was dominated by ribosomal genes. One other cluster was Archetypic, and did not have a significant difference from our scrambled dataset. B) Cell Cycle annotations, conducted through Seurat, identified the three dividing clusters. C) Significant Cluster markers against the number of cells found in the cluster for our original dataset and our scrambled datasets. Significant markers had a p-value of less than 0.05 after Bonferroni correction. The vast majority of clusters had a significant number of markers far outside the range of p-values attained through scrambled data. However, Archetypic, and other low-quality clusters did not. D). Cluster size vs Number of Markers for clusters from the original and scrambled data. As the number of cells decreases, the number of significant markers increases. Most, but not all, clusters have a high number of marker genes compared to the number of cells in the clusters, as compared to the scrambled data.

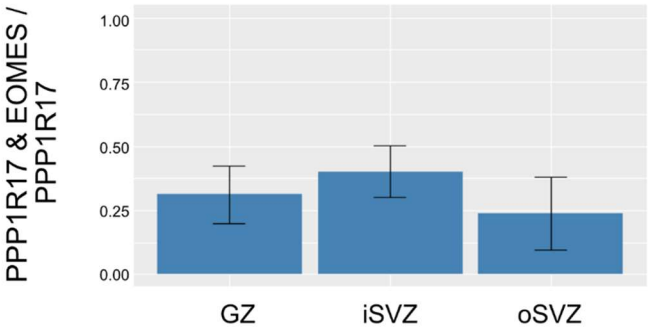


Supplemental Figure 2.6: Subclusters of RG-Like IPCs do not appear to correlated well with RG-subtypes. A) RG-Like IPCs can be clustered further bioinformatically. B) RG-subtypes genes do not appear to uniquely distinguish one subclusters from another. *PAX6* and *SOX2* included to show high levels of RG-like progenitor genes. C). Gene sets for each RG subtype also do not appear to distinguish subclusters. Gene set expression generated by projecting the marker gene set for each RG subtype into PCA space.

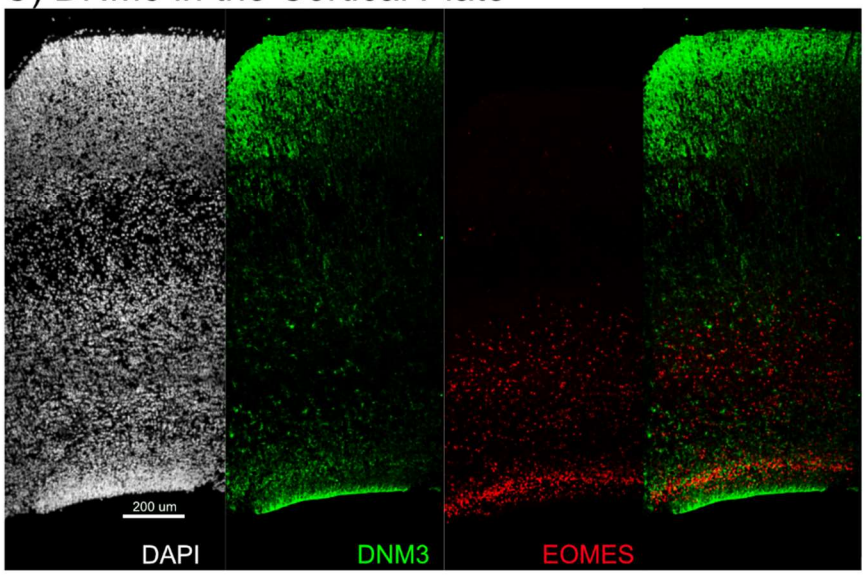
A) PPP1R17 only in GZ



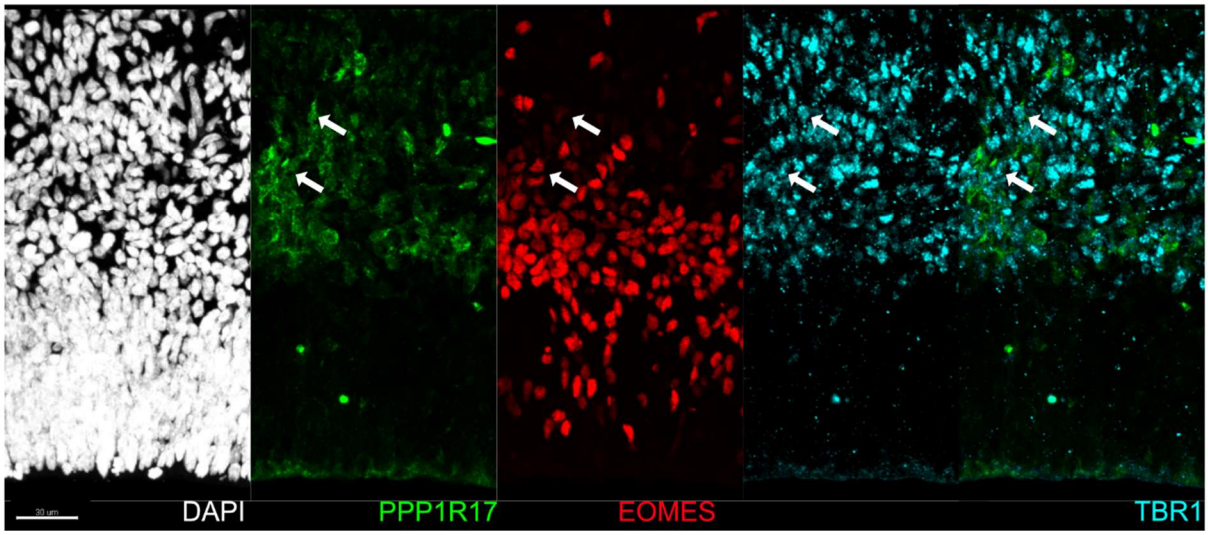
B) Most PPP1R17+ cells are not EOMES+



C) DNM3 in the Cortical Plate



D) PPP1R17+/EOMES- Cells Express TBR1



Supplemental Figure 2.7. nIPC Markers in other cell types. A) PPP1R17 is restricted to the GZ. Image taken at GW18. B) Most PPP1R17+ cells do not co-express EOMES. Graph represents the fraction of PPP1R17 cells that co-express EOMES. Values represent the average from an image at GW14, GW16, & GW18. Images were quantified to generated values for the germinal zone, before the same image was divided up into iSVZ & oSVZ. Errors bars represent standard deviation. C) DNM3 is expressed in the cortical plate at GW14. D) PPP1R17+/EOMES- Cells can also express TBR1, marking them as potentially newborn neurons.

References

Acosta, J. R. *et al.* (2017) 'Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population', *Stem Cell Research and Therapy*. BioMed Central Ltd., 8(1), p. 250. doi: 10.1186/s13287-017-0701-4.

Akbarian, S. (2014) 'Epigenetic mechanisms in schizophrenia', *Dialogues Clin Neurosci*, 16, pp. 405–417.

Alvarez-Buylla, A., Theelen, M. and Nottebohm, F. (1988) *Mapping of Radial Glia and of a New Cell Type in Adult Canary Brain*, *The Journal of Neuroscience*.

Alvarez-Buylla, A., Theelen, M. and Nottebohm, F. (1990) 'Proliferation "hot spots" in adult avian ventricular zone reveal radial cell division', *Neuron*. Elsevier, 5(1), pp. 101–109. doi: 10.1016/0896-6273(90)90038-H.

Andersen, O. M. and Willnow, T. E. (2006) 'Lipoprotein receptors in Alzheimer's disease', *Trends Neurosci*, 29, pp. 687–694. doi: 10.1016/j.tins.2006.09.002.

Anderson, S. *et al.* (1999) 'Differential origins of neocortical projection and local circuit neurons: role of Dlx genes in neocortical interneuronogenesis', *Cereb Cortex*, 9, pp.

646–654. doi: 10.1093/cercor/9.6.646.

Anderson, S. A. *et al.* (1997) ‘Interneuron migration from basal forebrain to neocortex: Dependence on Dlx genes’, *Science*. American Association for the Advancement of Science, 278(5337), pp. 474–476. doi: 10.1126/science.278.5337.474.

Arnold, S. J. *et al.* (2008) ‘The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone’, *Genes and Development*, 22(18), pp. 2479–2484. doi: 10.1101/gad.475408.

Baala, L. *et al.* (2007) ‘Homozygous silencing of T-box transcription factor EOMES leads to microcephaly with polymicrogyria and corpus callosum agenesis’, *Nature Genetics*. Nature Publishing Group, 39(4), pp. 454–456. doi: 10.1038/ng1993.

Bailey, S. D. and others (2015) ‘ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters’, *Nat Commun*, 2, p. 6186. doi: 10.1038/ncomms7186.

Bayatti, N. *et al.* (2008) ‘A Molecular Neuroanatomical Study of the Developing Human Neocortex from 8 to 17 Postconceptional Weeks Revealing the Early Differentiation of the Subplate and Subventricular Zone’, *Cerebral Cortex July*, 18, pp. 1536–1548. doi: 10.1093/cercor/bhm184.

Bentivoglio, M. and Mazzarello, P. (1999) ‘The history of radial glia’, *Brain Research Bulletin*. Elsevier Inc., 49(5), pp. 305–315. doi: 10.1016/S0361-9230(99)00065-9.

Betizeau, M., Cortay, V., Patti, D., *et al.* (2013) ‘Precursor Diversity and Complexity of Lineage Relationships in the Outer Subventricular Zone of the Primate’, *Neuron*, 80(2),

pp. 442–457. doi: 10.1016/j.neuron.2013.09.032.

Betizeau, M., Cortay, V., Patti, D. D. E., *et al.* (2013) ‘Precursor Diversity and Complexity of Lineage Relationships in the Outer Subventricular Zone of the Primate’, *Neuron*, 80(2), pp. 442–457. doi: 10.1016/j.neuron.2013.09.032.

Bhaduri, A. *et al.* (2020) ‘Cell stress in cortical organoids impairs molecular subtype specification’, *Nature*. Nature Research, 578(7793), pp. 142–148. doi: 10.1038/s41586-020-1962-0.

Blackwood, E. M. and Kadonaga, J. T. (1998) ‘Going the distance: A current view of enhancer action’, *Science*, pp. 60–63. doi: 10.1126/science.281.5373.60.

Bulik-Sullivan, B. K. and others (2015) ‘LD Score regression distinguishes confounding from polygenicity in genome-wide association studies’, *Nat Genet*, 47, pp. 291–295. doi: 10.1038/ng.3211.

Carbon, S. and others (2009) ‘AmiGO: online access to ontology and annotation data’, *Bioinformatics*, 25, pp. 288–289. doi: 10.1093/bioinformatics/btn615.

Choi, B. H. and Lapham, L. W. (1978) ‘Radial glia in the human fetal cerebrum: A combined golgi, immunofluorescent and electron microscopic study’, *Brain Research*. Elsevier, 148(2), pp. 295–311. doi: 10.1016/0006-8993(78)90721-7.

Choudhary, M. N. and others (2020) ‘Co-opted transposons help perpetuate conserved higher-order chromosomal structures’, *Genome Biol*, 21, p. 16. doi: 10.1186/s13059-019-1916-8.

Datlinger, P. and others (2017) ‘Pooled CRISPR screening with single-cell

transcriptome readout', *Nat Methods*, 14, pp. 297–301. doi: 10.1038/nmeth.4177.

Davis, C. A. and others (2018) 'The Encyclopedia of DNA elements (ENCODE): data portal update', *Nucleic Acids Res*, 46, pp. D794–D801. doi: 10.1093/nar/gkx1081.

Demontis, D. and others (2019) 'Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder', *Nat Genet*, 51, pp. 63–75. doi: 10.1038/s41588-018-0269-7.

Diao, Y. and others (2017) 'A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells', *Nat Methods*, 14, pp. 629–635. doi: 10.1038/nmeth.4264.

Dixon, J. R. and others (2012) 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*, 485, pp. 376–380. doi: 10.1038/nature11082.

Elias, L. A. B. *et al.* (2010) 'Connexin 43 mediates the tangential to radial migratory switch in ventrally derived cortical interneurons', *Journal of Neuroscience*, 30(20), pp. 7072–7077. doi: 10.1523/JNEUROSCI.5728-09.2010.

Englund, C. and others (2005) 'Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex', *J Neurosci*, 25(1), pp. 247–251. doi: 10.1523/JNEUROSCI.2899-04.2005.

Fairén, A., Cobas, A. and Fonseca, M. (1986) 'Times of generation of glutamic acid decarboxylase immunoreactive neurons in mouse somatosensory cortex', *Journal of Comparative Neurology*. John Wiley & Sons, Ltd, 251(1), pp. 67–83. doi:

10.1002/cne.902510105.

Fan, X. *et al.* (2020) *Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development*, *Sci. Adv*. Available at: http://advances.sciencemag.org/ (Accessed: 25 August 2020).

Fang, R. and others (2016) 'Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq', *Cell Res*, 26, pp. 1345–1348. doi: 10.1038/cr.2016.137.

Finucane, H. K. and others (2015) 'Partitioning heritability by functional annotation using genome-wide association summary statistics', *Nat Genet*, 47, pp. 1228–1235. doi: 10.1038/ng.3404.

Firth, H. V. *et al.* (2009) 'DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources', *American Journal of Human Genetics*. Elsevier, 84(4), pp. 524–533. doi: 10.1016/j.ajhg.2009.03.010.

Fish, J. L. *et al.* (2008) 'Making bigger brains - The evolution of neural-progenitor-cell division', *Journal of Cell Science*, 121(17), pp. 2783–2793. doi: 10.1242/jcs.023465.

Freytag, S. *et al.* (2018) 'Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data', *F1000Research*. NLM (Medline), 7, p. 1297. doi: 10.12688/f1000research.15809.2.

Gazal, S. and others (2017) 'Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection', *Nat Genet*, 49, pp. 1421–1427. doi: 10.1038/ng.3954.

Gertz, C. C. *et al.* (2014) 'Diverse Behaviors of Outer Radial Glia in Developing Ferret

and Human Cortex', *Journal of Neuroscience*, 34(7), pp. 2559–2570. doi: 10.1523/JNEUROSCI.2645-13.2014.

Hansen, D. V. *et al.* (2010) 'Neurogenic radial glia in the outer subventricular zone of human neocortex', *Nature*. Nature Publishing Group, 464(7288), pp. 554–561. doi: 10.1038/nature08845.

Hansen, D. V. *et al.* (2013) 'Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences', *Nature Neuroscience*. NIH Public Access, 16(11), pp. 1576–1587. doi: 10.1038/nn.3541.

Haubensak, W. *et al.* (2004) 'Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 101(9), pp. 3196–3201. doi: 10.1073/pnas.0308600100.

Heinz, S. and others (2010) 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38, pp. 576–589. doi: 10.1016/j.molcel.2010.05.004.

Hevner, R. F. (2006) 'From radial glia to pyramidal-projection neuron: transcription factor cascades in cerebral cortex development.', *Molecular neurobiology*, 33(1), pp. 33–50. doi: 10.1385/MN:33:1:033.

Hnisz, D. *et al.* (2013) 'Super-enhancers in the control of cell identity and disease', *Cell*. Cell Press, 155(4), p. 934. doi: 10.1016/j.cell.2013.09.053.

Hoffman, G. E. and others (2019) 'CommonMind Consortium provides transcriptomic

and epigenomic data for Schizophrenia and Bipolar Disorder', *Sci Data*, 6, p. 180. doi: 10.1038/s41597-019-0183-6.

Howard, D. M. and others (2019) 'Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions', *Nat Neurosci*, 22, pp. 343–352. doi: 10.1038/s41593-018-0326-7.

Javierre, B. M. and others (2016) 'Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters', *Cell*, 167, pp. 1369–1384. doi: 10.1016/j.cell.2016.09.037.

Juric, I. and others (2019) 'MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments', *PLoS Comput Biol*, 15(E2816044). doi: 10.1371/journal.pcbi.1006982.

Kalebic, N. and Huttner, W. B. (2020) 'Basal Progenitor Morphology and Neocortex Evolution', *Trends in Neurosciences*. Elsevier Ltd. doi: 10.1016/j.tins.2020.07.009.

Kelemenova, S. *et al.* (2010) 'Polymorphisms of candidate genes in Slovak autistic patients.', *Psychiatric genetics*, 20(4), pp. 137–9. doi: 10.1097/YPG.0b013e32833a1eb3.

Kostović, I., Išasegi, I. Ž. and Krsnik, Ž. (2018) 'Sublaminar organization of the human subplate: developmental changes in the distribution of neurons, glia, growing axons and extracellular matrix', *Journal of Anatomy*. John Wiley & Sons, Ltd (10.1111), p. joa.12920. doi: 10.1111/joa.12920.

Kriegstein, A., Noctor, S. and Martínez-cerdeño, V. (2006) 'Evolutionary Cortical

Expansion', *Neuroscience*, 7(November), pp. 883–890. doi: 10.1038/nrn2008.

Krsnik, Ž. *et al.* (2017) 'Growth of Thalamocortical Fibers to the Somatosensory Cortex in the Human Fetal Brain.', *Frontiers in neuroscience*. Frontiers Media SA, 11, p. 233. doi: 10.3389/fnins.2017.00233.

Labun, K. and others (2019) 'CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing', *Nucleic Acids Res*, 47, pp. W171–W174. doi: 10.1093/nar/gkz365.

Lek, M. *et al.* (2016) 'Analysis of protein-coding genetic variation in 60,706 humans', *Nature*. Nature Publishing Group, 536(7616), pp. 285–291. doi: 10.1038/nature19057.

Li, Y., Hu, M. and Shen, Y. (2018) 'Gene regulation in the 3D genome', *Hum Mol Genet*, 27, pp. R228–R233. doi: 10.1093/hmg/ddy164.

Lieberman-Aiden, E. and others (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*, 326, pp. 289–293. doi: 10.1126/science.1181369.

Lim, L. *et al.* (2018) 'Development and Functional Diversification of Cortical Interneurons', *Neuron*, 100, pp. 294–313. doi: 10.1016/j.neuron.2018.10.009.

Liu, S. J. and others (2016) 'Single-cell analysis of long non-coding RNAs in the developing human neocortex', *Genome Biol*, 17, p. 67. doi: 10.1186/s13059-016-0932-1.

Lui, J. H., Hansen, D. V. and Kriegstein, A. R. (2011) 'Development and evolution of the human neocortex', *Cell*. Elsevier Inc., 146(1), pp. 18–36. doi:

10.1016/j.cell.2011.06.030.

Luo, C. and others (2016) ‘Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain’, *Cell Rep*, 17, pp. 3369–3384. doi: 10.1016/j.celrep.2016.12.001.

Malatesta, P., Hartfuss, E. and Götz, M. (2000) ‘Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neural lineage’, *Development*, 127(24), pp. 5253–5263.

Malik, S. *et al.* (2013) ‘Neurogenesis continues in the third trimester of pregnancy and is suppressed by premature birth’, *Journal of Neuroscience*, 33(2), pp. 411–423. doi: 10.1523/JNEUROSCI.4445-12.2013.

Mihalas, A. B. *et al.* (2016) ‘Intermediate Progenitor Cohorts Differentially Generate Cortical Layers and Require Tbr2 for Timely Acquisition of Neuronal Subtype Identity’, *Cell Reports*, 16(1), pp. 92–105. doi: 10.1016/j.celrep.2016.05.072.

Miller, D. J. *et al.* (2019) ‘Shared and derived features of cellular diversity in the human cerebral cortex’, *Curr Opin Neurobiol*, 56, pp. 117–124. doi: 10.1016/j.conb.2018.12.005.

Miller, M. W. (1985) ‘Cogeneration of retrogradely labeled corticocortical projection and GABA-immunoreactive local circuit neurons in cerebral cortex’, *Developmental Brain Research*. Brain Res, 23(2), pp. 187–192. doi: 10.1016/0165-3806(85)90040-9.

Miyata, T. *et al.* (2001) ‘Asymmetric inheritance of radial glial fibers by cortical neurons’, *Neuron*. Cell Press, 31(5), pp. 727–741. doi: 10.1016/S0896-6273(01)00420-2.

Mo, Z. and Zecevic, N. (2008) ‘Is Pax6 Critical for Neurogenesis in the Human Fetal Brain?’, *Cerebral Cortex*, 18(6), pp. 1455–1465. doi: 10.1093/cercor/bhm181.

Nagashima, S. *et al.* (2015) ‘MAGI2/S-SCAM outside brain’, *Journal of Biochemistry*, 157(4), pp. 177–184. doi: 10.1093/jb/mvv009.

Ngondo-Mbongo, R. P. *et al.* (2013) ‘Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11’, *Nucleic Acids Res*, 41, pp. 4000–4014. doi: 10.1093/nar/gkt088.

Noctor, S. C. *et al.* (2001) ‘Neurons derived from radial glial cells establish radial units in neocortex’, *Nature*. Nature Publishing Group, 409(6821), pp. 714–720. doi: 10.1038/35055553.

Noctor, S. C. *et al.* (2004) ‘Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases’, *Nat Neurosci*, 7(2), pp. 136–144. doi: 10.1038/nn1172.

Nowakowski, T. J. *et al.* (2016) ‘Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development’, *Neuron*, 91(6), pp. 1219–1227. doi: 10.1016/j.neuron.2016.09.005.

Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshev, D., Ounadjela, J. R., Shuga, J., Wang, X., Lim, D. A., West, J. A., Leyrat, A. A., Kent, W. J., Kriegstein, A. R., *et al.* (2017) ‘Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex’, *Science*. American Association for the Advancement of Science, 358(6368), pp. 1318–1323. doi: 10.1126/science.aap8809.

Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshev, D., Ounadjela, J. R., Shuga, J., Wang, X., Lim, D. A., West, J. A., Leyrat, A. A., Kent, W. J. and Kriegstein, A. R. (2017) 'Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex', *Science*. American Association for the Advancement of Science, 358(6368), pp. 1318–1323. doi: 10.1126/science.aap8809.

Pardinas, A. F. and others (2018) 'Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection', *Nat Genet*, 50, pp. 381–389. doi: 10.1038/s41588-018-0059-2.

Phane Ory, S. *et al.* (2003) 'Protein Phosphatase 2A Positively Regulates Ras Signaling by Dephosphorylating KSR1 and Raf-1 on Critical 14-3-3 Binding Sites', *Current Biology*, 13, pp. 1356–1364. doi: 10.1016/S.

Pixley, S. K. R. and de Vellis, J. (1984) 'Transition between immature radial glia and mature astrocytes studied with a monoclonal antibody to vimentin', *Developmental Brain Research*. Elsevier, 15(2), pp. 201–209. doi: 10.1016/0165-3806(84)90097-X.

Polioudakis, D. *et al.* (2019) 'A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation', *Neuron*, 103(5), pp. 785-801.e8. doi: 10.1016/j.neuron.2019.06.011.

Pollen, A. A. *et al.* (2015) 'Molecular Identity of Human Outer Radial Glia during Cortical Development', *Cell*, 163(1), pp. 55–67. doi: 10.1016/j.cell.2015.09.004.

Pontious, A. *et al.* (2008) 'Role of intermediate progenitor cells in cerebral cortex development', *Dev Neurosci*, 30, pp. 24–32. doi: 10.1159/000109848.

Raju, C. S. *et al.* (2018) ‘Secretagogin is Expressed by Developing Neocortical GABAergic Neurons in Humans but not Mice and Increases Neurite Arbor Size and Complexity’, *Cerebral Cortex*. Oxford University Press, 28(6), pp. 1946–1958. doi: 10.1093/cercor/bhx101.

Rakic, P. (1971) ‘Guidance of neurons migrating to the fetal monkey neocortex’, *Brain Research*. Elsevier, 33(2), pp. 471–476. doi: 10.1016/0006-8993(71)90119-3.

Rakic, P. (1972) ‘Mode of cell migration to the superficial layers of fetal monkey neocortex’, *Journal of Comparative Neurology*, 145(1), pp. 61–83. doi: 10.1002/cne.901450105.

Rakic, P. (1988) ‘Specification of cerebral cortical areas’, *Science*. American Association for the Advancement of Science, 241(4862), pp. 170–176. doi: 10.1126/science.3291116.

Ramón y Cajal, S. (1909) *Histologie du système nerveux de l’homme & des vertébrés.* Paris: Maloine. Available at: https://archive.org/details/histologiedusyst01ram.

Rani, N. and others (2016) ‘A Primate lncRNA Mediates Notch Signaling during Neuronal Development by Sequestering miRNA’, *Neuron*, 90, pp. 1174–1188. doi: 10.1016/j.neuron.2016.05.005.

Rash, B. G. *et al.* (2019) ‘Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(14), pp. 7089–7094. doi: 10.1073/pnas.1822169116.

Reillo, I. *et al.* (2011) ‘A Role for Intermediate Radial Glia in the Tangential Expansion of the Mammalian Cerebral Cortex’, *Cerebral Cortex July*, 21, pp. 1674–1694. doi: 10.1093/cercor/bhq238.

Reilly, S. K. and others (2015) ‘Evolutionary changes in promoter and enhancer activity during human corticogenesis.’, *Science*, 347, pp. 1155–1159. doi: 10.1126/science.1260943.

Ri Reimand, J. *et al.* (2007) ‘g:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments’, *Nucleic Acids Research*, 35, pp. 193–200. doi: 10.1093/nar/gkm226.

Savage, J. E. and others (2018) ‘Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence’, *Nat Genet*, 50, pp. 912–919. doi: 10.1038/s41588-018-0152-6.

Schmechel, D. E. and Rakic, P. (1979) ‘A golgi study of radial glial cells in developing monkey telencephalon: Morphogenesis and transformation into astrocytes’, *Anatomy and Embryology*. Springer-Verlag, 156(2), pp. 115–152. doi: 10.1007/BF00300010.

Schmitt, A. D. and others (2016) ‘A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome’, *Cell Rep*, 17, pp. 2042–2059. doi: 10.1016/j.celrep.2016.10.061.

Schoenfelder, S. and Fraser, P. (2019) ‘Long-range enhancer-promoter contacts in gene expression control’, *Nat Rev Genet*, 20, pp. 437–455. doi: 10.1038/s41576-019-0128-0.

Semple, B. D. *et al.* (2013) ‘Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species’, *Progress in Neurobiology*. NIH Public Access, pp. 1–16. doi: 10.1016/j.pneurobio.2013.04.001.

Sessa, A. *et al.* (2008) ‘Tbr2 Directs Conversion of Radial Glia into Basal Precursors and Guides Neuronal Amplification by Indirect Neurogenesis in the Developing Neocortex’, *Neuron*, 60(1), pp. 56–69. doi: 10.1016/j.neuron.2008.09.028.

Stumm, R. K. *et al.* (2003) ‘CXCR4 regulates interneuron migration in the developing neocortex’, *Journal of Neuroscience*. Society for Neuroscience, 23(12), pp. 5123–5130. doi: 10.1523/jneurosci.23-12-05123.2003.

Sundaram, V. and Wang, T. (2018) ‘Transposable Element Mediated Innovation in Gene Regulatory Landscapes of Cells: Re-Visiting the “Gene-Battery” Model’, *Bioessays*, 40. doi: 10.1002/bies.201700155.

Suzuki, I. K. and others (2018) ‘Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation’, *Cell*, 173(e1316), pp. 1370–1384. doi: 10.1016/j.cell.2018.03.067.

Thomsen, E. R. *et al.* (2015) ‘Fixed single-cell transcriptomic characterization of human radial glial diversity’, *Nature Methods*. doi: 10.1038/nmeth.3629.

Visel, A. *et al.* (2007) ‘Enhancer Browser--a database of tissue-specific human enhancers’, *Nucleic Acids Res*, 35, pp. D88-92. doi: 10.1093/nar/gkl822.

Walker, R. L. and others (2019) ‘Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms’, *Cell*, 179(e722), pp. 750–771.

doi: 10.1016/j.cell.2019.09.021.

Wang, Y. *et al.* (2011) 'CXCR4 and CXCR7 Have Distinct Functions in Regulating Interneuron Migration', *Neuron*. Neuron, 69(1), pp. 61–76. doi: 10.1016/j.neuron.2010.12.005.

Weissman, T. A. *et al.* (2004) 'Calcium Waves Propagate through Radial Glial Cells and Modulate Proliferation in the Developing Neocortex', *Neuron*, 43, pp. 647–661. Available at: http://ac.els-cdn.com/S0896627304004970/1-s2.0-S0896627304004970-main.pdf?_tid=ce5239bc-7008-11e7-982c-00000aab0f01&acdnat=1500857058_276f66a8fbff0591a10eeb860139e423 (Accessed: 23 July 2017).

Won, H. and others (2016) 'Chromosome conformation elucidates regulatory relationships in developing human brain', *Nature*, 538, pp. 523–527. doi: 10.1038/nature19847.

Xie, S. *et al.* (2017) 'Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells', *Mol Cell*, 66, pp. 285–299. doi: 10.1016/j.molcel.2017.03.007.

Yang, L. *et al.* (1998) *Synthesis and biological activities of potent peptidomimetics selective for somatostatin receptor subtype 2*. Merck & Co., Inc. Available at: www.pnas.org. (Accessed: 30 November 2018).

Yang, X. and others (2020) 'SMART-Q: An Integrative Pipeline Quantifying Cell Type-Specific RNA Transcription', *PLoS One*, 15(e0228760). doi: 10.1371/journal.pone.0228760.

Yu, X. and Zecevic, N. (2011) ‘Dorsal radial glial cells have the potential to generate cortical interneurons in human but not in mouse brain’, *Journal of Neuroscience*. Society for Neuroscience, 31(7), pp. 2413–2420. doi: 10.1523/JNEUROSCI.5249-10.2011.

Zecevic, N., Chen, Y. and Filipovic, R. (2005) ‘Contributions of cortical subventricular zone to the development of the human cerebral cortex’, *Journal of Comparative Neurology*, 491(2), pp. 109–122. doi: 10.1002/cne.20714.

Zhang, Y. and others (2019) ‘Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells’, *Nat Genet*. doi: 10.1038/s41588-019-0479-7.

Zheng, H. and Xie, W. (2019) ‘The role of 3D genome organization in development and cell differentiation’, *Nat Rev Mol Cell Biol*, 20, pp. 535–550. doi: 10.1038/s41580-019-0132-4.

Žunić Išasegi, I. *et al.* (2018) ‘Interactive histogenesis of axonal strata and proliferative zones in the human fetal cerebral wall’, *Brain Structure and Function*. Springer Berlin Heidelberg, (0123456789). doi: 10.1007/s00429-018-1721-2.

Publishing Agreement

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

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

DocuSigned by:
D160C7B818D9434...
Author Signature

12/11/2020
Date