Maintaining Subtype-Specific Neuronal Identity in the Mammalian Cerebral Cortex

Maintaining Subtype-Specific Neuronal Identity in the Mammalian Cerebral Cortex

A dissertation presented

by

Aurora Claudia Zhang

to

The Division of Medical Science

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

March 2020

Dissertation Advisor: Paola Arlotta Aurora Claudia Zhang

Maintaining Subtype-Specific Neuronal Identity in the Mammalian Cerebral Cortex

Abstract

In the mammalian cerebral cortex, distinct classes of neurons are generated embryonically and maintain their subtype-specific identities throughout the lifetime of the organism. While their developmental programs have begun to be elucidated, much less is known about the mechanisms that preserve their identity through time. We showed previously that *Fezf2* is a selector gene for corticospinal motor neurons (CSMNs) and the first selector gene known for any class of projection neurons in the cortex. Here, we hypothesize that *Fezf2* is a terminal selector gene and is therefore needed to maintain CSMN identity throughout the life of the animal. We employed a genetic strategy to conditionally delete *Fezf2* specifically from CSMNs at distinct postnatal time points, when CSMN fate specification and connectivity have already been established. Our data shows that while Fezf2 is needed early postnatally to maintain the molecular identity of CSMNs, it is not necessary in the adult. By exploiting *Fezf2* function in mature CSMNs, our work suggests that other mechanisms are used for neurons to maintain their identity and function in the mature cerebral cortex.

Table of Contents

Acknowledgements

Getting through 23rd grade is hard (22nd? 23rd? 24th? What *is* time in graduate school, anyway?), but I am grateful to have made it out (mostly intact) with a lot of help. I would like to take this opportunity to thank my community of supporters over the past few years.

To start, I am grateful to have received funding from the National Institutes of Health, National Institute of Neurological Diseases and Stroke's National Research Service Award (5F31NS098539).

And now, for the more mushy acknowledgements: I want to thank my graduate advisor, Paola Arlotta, for the opportunity to grow in her lab. Paola gave me the freedom to dream big and follow my ideas. She always had confidence in me and urged me to stay positive, even when I could not see the light at the end of the tunnel. Thank you, Paola, for inspiring me, pushing me, and believing in me.

I want to thank the Arlotta lab, past and present, for their endless support both personally and scientifically. Thank you to Simona Lodato for taking me on as a rotation student and showing me the ropes. Very special thank you's go to Juliana Brown, May Chen, Vahbiz Johki, JoJo Nguyen, Giorgia Quadrato, John Sherwood, Ashwin Shetty, Jeff Stogsdill, Zachary Trayes-Gibson, and Wen Yuan for their advice, encouragement, and friendship over the years. Muchas gracias to Daniella DiBella and obrigada to Bruna Paulsen for their infectious energy; please stay "estupid" forever. I also want to thank fellow (now former) graduate students in the lab, Ryoji Amamoto and James Harris, for always attending my "TED Talks" (and for your friendship!). Thank you also to the wonderful research technicians over the years: Nathan Curry, Paul Oyler, and Helen Zhang, all of whom have contributed toward my project. I am incredibly grateful. Thank you, Arlotta lab, for all of the memories.

I also want to thank Joshua Levin and his group for their support, especially Amanda Kedaigle and Sean Simmons, for helping me sort through all of the sequencing data. Thank you for your patience and teaching me how to fish.

I want to also thank my Dissertation Advisory Committee members: Connie Cepko, Chinfei Chen, and Jesse Gray for their scientific and career support throughout the years. I would like to give a special thank you to Connie for serving as the chair of my defense committee and to the rest of my defense committee, Corey Harwell, Michela Fagiolini, and Myriam Heiman.

A big thank you to the Biological and Biomedical Sciences program, the Division of Medical Sciences, and the Stem Cell and Regenerative Biology Department for all the behind-the-scenes work that has supported this research. Thank you to the Bauer Flow Cytometry Core, specifically Jeff Nelson and Zach Niziolek for their amazing contribution toward my projects. Thank you to the Harvard Center for Biological Imaging Core, especially Doug Richardson, as well as the Simmons family, for their generous gift to the graduate student community at Harvard.

Thank you to the wonderful people who were on the graduate school struggle bus with me the past few years. I honestly cannot imagine going through graduate school without all of you by my side. Thank you, Kelly Biette and Kelsey Tyssowski, for your unwavering support through my most questionable ideas, like bringing frozen corn dogs and pizza rolls to camp in July. Thank you, Jeff Gerold, for being my beard and fellow RedVolution and cheese aficionado. Thank you, Emma Sedivy, for all of our dumb misadventures together, starting all the way back in Kimball and Bio Core. Dan Fuentes and Cameron Lee, thank you for our special food adventures, from mislabeled cakes to gas station tacos. Thank you to Carrie Mills and Zach Sherman, my wonderful colloidal neighbors, for literally always being right next to me with beers and seltzer ready to go. Thank you to my adopted Program in Neuroscience class for taking me in. I especially want to thank Andrea Yung, my almost twin, who has been my classmate and dear friend through high school, undergrad, and grad school. It's an honor to be confused for you, even among our so-

called friends who claim to be able to tell us apart. I also want to thank Ben Andreone and Brian Chow, for the endless stream of (good) mess. You goobers always know how to put a smile on my face.

Although I am an only child, I am lucky to have found some great people that I consider as close as kin across the years. Thank you to Pooja Bakhai, Laura Ceng, Kelly Ding, Paul Iona, Gea Kang, Teresa Ortega, and Romance Explosion (Will Bassett, Joyce Chai, Max Kim, and Katie Stanley) for reminding me that there is life outside of the academic bubble and always accompanying me in eating everything in sight. I especially want to thank Joyce and Max for always letting me third wheel and the never-ending lifelines they've sent over the years. Thank you to Kenny Wong, for always going above and beyond and making sure we take advantage of every Costco run. Thank you to Aaron Garg, my roommate for five years, for all of the wine and feels, the H-Mart surprises, and the wonderful memories we made in the Bat Cave. Thank you to Brenda Stella "Bobo Bread" Ou, my best friend, for staying by my side for over a decade, despite the thousands of miles between us. YDB, but also WAYSD.

Finally, I want to thank my parents. My parents raised me to be critical, resilient, and independent, which turned out to be incredibly useful for this PhD. Thank you for always asking if I already ate, and thank you for supporting me in everything.

List of Figures and Tables

Chapter I.

Introduction: Generation and Maintenance of Neurons in the Mammalian Cerebral Cortex

This chapter is adapted from an as-yet-unpublished manuscript I wrote with guidance from Paola Arlotta.

Introduction

Mammalian central nervous system neurons are a textbook example of a stable and terminally differentiated cell type. Understanding how these neurons develop and function in proper circuits for the lifetime of the organism have been longstanding questions for neurobiologists, as their genesis and maintenance are the foundation for the functional capacity and complex behaviors driven by the brain. These questions are particularly interesting in the cerebral cortex, which has undergone the most expansion during evolution, and is involved in a staggering assortment of functions, including sensory perception, cognition, and fine motor function (Florio and Huttner, 2014; Lui et al., 2011). The functional output of the cortex is underpinned by the tremendous diversity of embryonically-born neurons which work in concert with many other cell types across the entire lifetime of the organism (Greig et al., 2013; Harris et al., 2015; Harris and Shepherd, 2015; Lodato et al., 2015). Considering the size of the mammalian genome and the wide diversity of cell types needed to orchestrate cortical function, how are all of these cell types generated and subsequently maintained? Do all of the neuronal subtypes then share maintenance programs, once identity is established? If not, what are the different mechanisms employed?

Much of our knowledge of cortical neuron development and insight into identity maintenance comes from comparative gene expression analyses between neuronal subtypes (Arlotta et al., 2005; Molyneaux et al., 2009; Molyneaux et al., 2015; Saunders et al., 2018; Tasic et al., 2016; Tasic et al., 2018). Because of the wide diversity of cell types within the cortex, defining neuronal subtypes in the cortex and decoding their developmental mechanisms is complicated. A preliminary view of the cellular diversity of the cortex can be gauged from its

unique cytoarchitecture. The mammalian cortex is organized into six horizontal layers, which can be broadly grouped by function: supragranular (layers I-III), granular (layer IV), and infragranular (layers V-VI). Generally speaking, the supragranular layers process intracortical inputs, while the granular layer receives inputs from noncortical regions. The infragranular layers then send information out of the cortex. However, defining cell types by their layer localization is not sufficient, as cellular composition is not homogenous within the layers and is further complicated by molecular and physiological differences between sensory areas (Custo Greig et al., 2013; Harris et al., 2015; Harris and Shepherd, 2015; Lodato et al., 2015).

Further classification of cortical neurons has traditionally considered a variety of functional and structural traits. At the broadest level, cortical neurons can be roughly separated by neurotransmitter identity: excitatory (glutamatergic) projection neurons (PNs) or inhibitory (GABAergic) interneurons. The somas of these two populations reside in the stereotyped six-layered cortex, but their projection targets vary (Figure 1.1A). The excitatory projection neurons encompass about 80% of the cortical neuronal population, and they extend their axons to distal targets either within or outside of the cortex; the inhibitory neurons, on the other hand, represent about 20% of the population and connect locally within the cortex. Projection neurons are classically further subdivided by hodology: callosal projection neurons (CPNs, also known as intertelencephalic, or IT neurons) reside in all layers and project within the cortical hemispheres (Figure 1.1B), while the corticofugal projection neurons (CFuPNs) reside in the deep layers (layer V-VI) and project out of the cortex. The CFuPNs can be further subdivided into subcerebral projection neurons (SCPNs, also known as extratelencephalic, or ET neurons), which send projections to subcerebral targets such as the spinal cord and pons (Figure 1.1C), and the corticothalamic projection neurons (CThPNs), which connect to the thalamus. However, these superficial forms of classification alone do not truly capture their full biological diversity, as there are exceptions to these rules, such as the corticostriatal PNs that also project contralaterally

(Shepherd, 2013). Furthermore, morphologically similar subtypes can have distinct molecular features. For example, while all CPNs express *Satb2* and reside in layers II-VI in the cortex, they can be further subdivided molecularly: upper layer CPNs express *Cux1* and *Cux2*, whereas both upper layer CPNs and layer Va CPNs express *Plexind1* (Molyneaux et al., 2009).

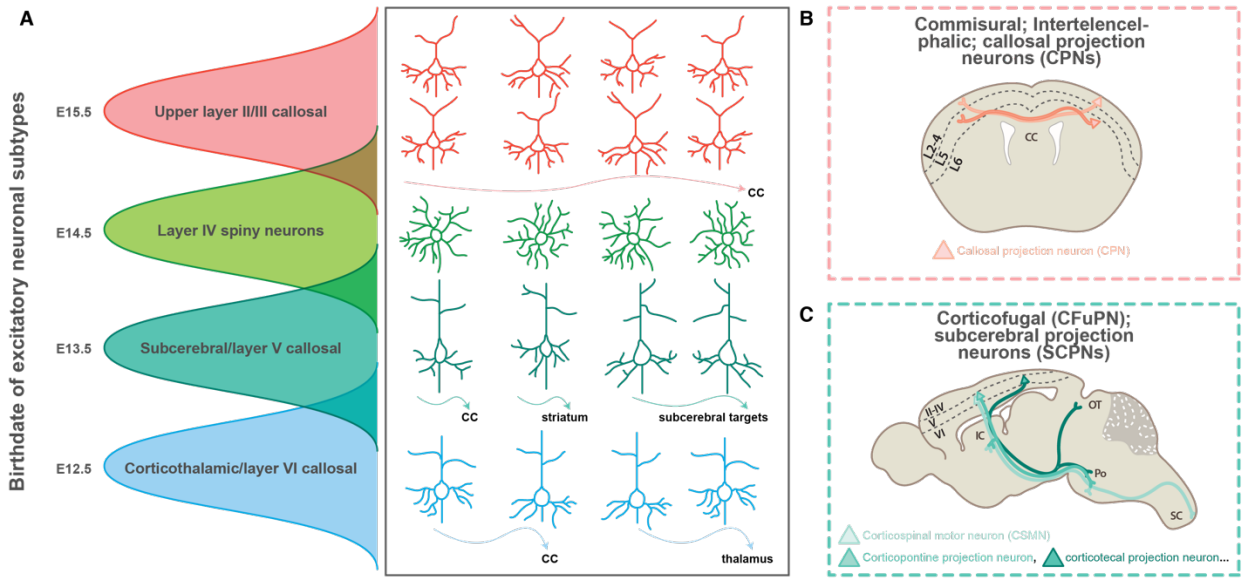


FIGURE 1.1. THE CORTEX CONTAINS A GREAT DIVERSITY OF EXCITATORY NEURON SUBTYPES

A. Cortical neurons are generated sequentially during embryonic development, in an inside-out manner, and eventually migrate to their appropriate cortical layers. Only some of the major classes of PNs with their corresponding axonal projections are depicted here; the larger diversity is not shown. CC: corpus callosum. ***B.*** Callosal projection neurons (CPNs) are a population of commissural neurons which cross the two hemispheres via the CC, also known as intertelencephalic (IT) neurons. They populate all layers. ***C.*** Layer V subcerebral projection neurons (SCPNs) are a subset of corticofugal projection neurons (CFuPN). SCPNs send their projections down the internal capsule (IC), and are further subdivided by axonal targets, including corticospinal motor neurons (CSMNs) that project to the spinal cord (SC), corticopontine projection neurons that project to the pontine nuclei (Po), and corticotectal projection neurons that project to the optic tectum (OT).

Technological advances over the last decade have propelled classification and our understanding of neuronal diversity forward tremendously. Pioneering work that isolated subtypes by hodology allowed the in-depth molecular characterization of several classically-defined cell types. This paved the way for complementary deep profiling through methods such as immunopanning (Foo et al., 2011), genetic labeling (Tasic et al., 2016; Tasic et al., 2018), and

intranuclear immunostaining (Amamoto et al., 2020; Molyneaux et al., 2015), which all ultimately converge to identify the molecular signatures of various classes of neurons. Single-cell sequencing technology now allows the molecular dissection of these cell types in an unbiased, high-throughput manner. However, this can complicate defining and isolating neuronal subtypes for transcriptional analysis. For example, although *Cux1* is a reliable upper layer CPN marker during early development, *Cux1* expression becomes more promiscuous with age and can be found in a subset of adult layer V SCPNs (Tasic et al., 2016). Thus, with higher resolution comes the need to examine the full molecular profile of neurons, instead of relying on single marker genes to identify and classify neurons. These transcriptional studies nevertheless have revealed several mechanisms behind the generation of the many cortical neuronal subtypes in the cortex.

Terminal Selector Genes as a Method for Identity Maintenance

Despite our progress in understanding the development of cortical neuronal subtypes, much less is known about how their identity is maintained. Mechanisms behind maintenance have been uncovered by studying the development of invertebrates such as *C. elegans*, where highly-invariant and well-characterized lineage decisions produce 302 individual neurons. In these systems, a key component of the regulatory logic behind neuronal diversity lies in terminal selector genes, transcription factors that directly govern the expression of numerous downstream genes necessary for the development and maintenance of a specific neuronal class (Hobert, 2008). The terminal selector concept can be illustrated with the *C. elegans* gene, *unc-86,* a powerful transcription factor in *C. elegans*. *Unc-86* is necessary to initiate 31 neuron classes in the developing *C. elegans*, and its expression in maintained in 30 classes of neurons (Serrano-Saiz et al., 2018). The post-developmental removal of *unc-86* disrupts the neurotransmitter identity of a smaller subset of neurons, including PVD, PHC, and PLN sensory neurons, without losing pan-neuronal markers (Serrano-Saiz et al., 2018). Thus, *unc-86* is a terminal selector for a

small subset of sensory neurons; it is able to initiate and maintain their subtype-specific properties in the post-development animal. Intriguingly, this mechanism is conserved in mammalian neurons. *Unc-86*'s mouse ortholog, *Brn3a,* was also found as necessary to maintain the neurotransmitter identity of medial habenula neurons (Serrano-Saiz et al., 2018). Thus, despite the diversity of cell types in the mammalian cortex, it is conceivable that neurons use terminal selector genes as part of the regulatory logic behind their diversity.

To this date, we have yet to find a true terminal selector, despite our progress in identifying selector genes for mammalian cortical neuronal subtypes. A terminal selector gene needs to meet three main criteria. First, a true terminal selector gene is necessary for the specification of a specific neuronal subtype, where the gene actively selects for subtype-specific features. Second, it is sufficient to induce subtype-specific identity, to promote its downstream effector genes and repress those of other lineages. Finally, its continued expression in adult neurons is necessary to maintain the identity established during development; removal of the terminal selector would cause a loss of subtype-specific genes, such as those involved in its neurotransmitter identity.

Several transcription factors have been shown as necessary to acquire subtype-specific identity in the cortex, such as Satb2 in CPNs. Satb2-deficient neurons have abnormal expression of subtype-specific genes throughout the cortex, including the de-repression of CTIP2, a SCPN marker (Alcamo et al., 2008; Britanova et al., 2008). This loss of the CPN genetic program contributes to the physiological phenotype in *Satb2* knockout neurons: instead of projecting across the corpus callosum, Satb2-deficient neurons largely project aberrantly to the septum or subcortically (Alcamo et al., 2008; Britanova et al., 2008), and they acquire the electrophysiological properties of layer V SCPNs, including the distinct voltage sag (Leone et al., 2015). Satb2 is also sufficient to repress other identities: its embryonic overexpression in neural progenitors at a time when they predominantly produce SCPNs, is sufficient to reduce the number of CTIP2 positive cells and projections to the cerebral peduncle, suggesting that the

overexpression of *Satb2* can cause a loss of SCPN identity at the molecular and morphological levels (Britanova et al., 2008). However, Satb2's role in maintaining CPN identity in post-mitotic neurons is still unknown, and it therefore cannot be considered a terminal selector yet.

Addressing this final criterion of maintenance has been challenging in the mammalian cortex because of the lack of tools that grant spatiotemporal specificity in the postnatal animal. Most of the studies claiming maintenance have done their experiments during embryonic development, before the neurons have fully formed their connections within the cortex. A notable example is *Tbr1*, a transcription factor necessary for the generation of corticothalamic projection neurons (Bedogni et al., 2010; Hevner et al., 2001). It is only shown as necessary to maintain CThPN identity with late embryonic deletion (Fazel Darbandi et al., 2018), where its deletion leads to layer VI neurons losing CThPN marker genes in favor of layer V SCPN marker genes. Similarly, *Lhx2*, a transcription factor for necessary for corticogenesis (Mangale et al., 2008) and proper subtype-speciifcation in both deep and upper layer neurons (Muralidharan et al., 2017; Shetty et al., 2013; Wang et al., 2017), has not been investigated beyond embryonic development. These examples only illustrate that cortical neurons are transcriptionally responsive and amenable to molecular changes, or plastic, during embryonic development. Whether the same molecular changes would occur with deletion from neurons postnatally is still unexplored.

Fezf2 as a Selector Gene for CSMN Identity

Of the transcription factors that have been discovered as necessary for subtype-specific identity in the cortex, Fezf2 is the strongest candidate to reveal key insights into maintaining cortical neuron identity in the cerebral cortex, because of its clear role as a selector gene for CSMNs.

First, Fezf2 is necessary for the birth and specification of CSMNs. In *Fezf2* constitutive knockout (KO) mice, the only projection neuron subtypes lost in the cortex are CSMNs and other SCPNs; cells normally fated to become CSMNs in Fezf2 KO mice generate neurons similar to deep layer CPNs (Chen et al., 2005a; Chen et al., 2005b; Molyneaux et al., 2005). They do not acquire the stereotypically large soma of layer V CSMN and have a smaller cell body, which send projections contralaterally through the anterior commissure instead of subcerebrally through the internal capsule (Chen et al., 2005a; Chen et al., 2005b; Chen et al., 2008; Molyneaux et al., 2005). They also acquire a diminished dendritic morphology with fewer terminal branches as compared to controls (Chen et al., 2005a; Chen et al., 2008). The layer V CPN-like neurons in Fezf2 KO mice also lose markers of SCPNs and gain those of CPNs (Chen et al., 2005b; Molyneaux et al., 2005). Finally, recordings from layer V neurons in Fezf2 KO animals versus controls show that KO animals have more neurons in layer V that acquire CPN-like electrophysiological traits. Upon current injection in layer V of Fezf2 KO mice, there was a greater proportion of neurons that showed spike frequency adaptation, a feature of CPNs which CFuPN typically do not have (Chen et al., 2008). These observations point toward Fezf2 as necessary for the generation of CSMNs.

Second, Fezf2 is sufficient to impose a deep layer, CFuPN identity in many cell types. *Fezf2* overexpression in dorsal progenitors at E13.5, when the progenitors largely produce layer IV neurons, causes the newborn neurons to prematurely differentiate and exhibit migration defects (Molyneaux et al., 2005). These neurons also ectopically express a battery of CFuPN signature genes and extend their projections subcortically through the internal capsule (Molyneaux et al., 2005), like endogenous CSMN. *Fezf2* overexpression can also produce CFuPN from progenitors of the embryonic lateral ganglionic eminence (Rouaux and Arlotta, 2010), as well as from the neural stem cells of the adult subventricle zone (Zuccotti et al., 2014), which normally are lineage restricted to produce short-range, GABAergic inhibitory interneurons (Lois and

Alvarez-Buylla, 1994). While these reprogrammed neurons do not integrate into the cortex, they gain the morphology of pyramidal neurons, express CFuPN markers, and change their GABAergic neurotransmitter identity to glutamatergic (Rouaux and Arlotta, 2010; Zuccotti et al., 2014). Intriguingly, Fezf2 can also reprogram post-mitotic neurons, which were previously thought to be immutable. *Fezf2* overexpression in early post-mitotic layer IV spiny neurons or layer II/III CPNs can reprogram them into a CFuPN identity at the molecular, morphological, and electrophysiological levels (De la Rossa et al., 2013; Rouaux and Arlotta, 2013, Ye et al., 2015). Reprogrammed neurons lose upper layer cortical markers such as *Cux1*, and gain CFuPN markers such as *Crym*, *Tle4*, and *Er81* (De la Rossa et al., 2013; Rouaux and Arlotta, 2013; Ye et al., 2015); they change their intrinsic electrophysiological parameters, such as firing pattern, and acquire of the trademark layer V neuron voltage sag, to resemble endogenous layer V CFuPN (De la Rossa et al., 2013, Ye et al., 2015). Finally, reprogrammed neurons have an increased soma size and a dramatic increase in arborization, again comparable to endogenous layer V CFuPN (De la Rossa et al., 2013; Ye et al., 2015). Although the reprogramming capacity, as seen by the number of cells capable of changing its identity, of post-mitotic neurons drops precipitously during the first postnatal week (Rouaux and Arlotta, 2013; Ye et al., 2015), these data altogether have demonstrated Fezf2's sufficiency to generate CFuPN. To date, no other selector gene has been shown to be sufficient to reprogram post-mitotic neurons into that of another neuronal subtype.

Fezf2 actively specifies CSMN identity by directing downstream effector genes and repressing that of other neuronal lineages. Analysis of Fezf2 promoter binding and downstream transcriptional changes showed that Fezf2 binds to the promoter and induces the expression of, *Vglut1*, endowing CSMNs with its glutamatergic neurotransmitter identity (Lodato et al., 2014). Fezf2 also directly activates CSMN's distinctive axonal targeting to the spinal cord, by binding to the promoter region of EphB1, an axon guidance receptor; in Ephb1 constitutive knockout mice,

subcerebral axons were found to pass through the anterior commissure instead of projecting subcortically through the internal capsule (Lodato et al., 2014). Fezf2 can also directly repress a GABAergic identity, as Fezf2 binds the promoter of *Gad1*, which is necessary for the synthesis of GABA for inhibitory INs (Lodato et al., 2014). Indeed, *Fezf2* overexpression dramatically reduces Gad1 expression both *in vitro* and *in vivo* (Lodato et al., 2014; Rouaux and Arlotta, 2010; Zuccotti et al., 2014). Altogether, these data illustrate that Fezf2 works both to activate CSMN-specific genes and repress genes characteristic of other identities.

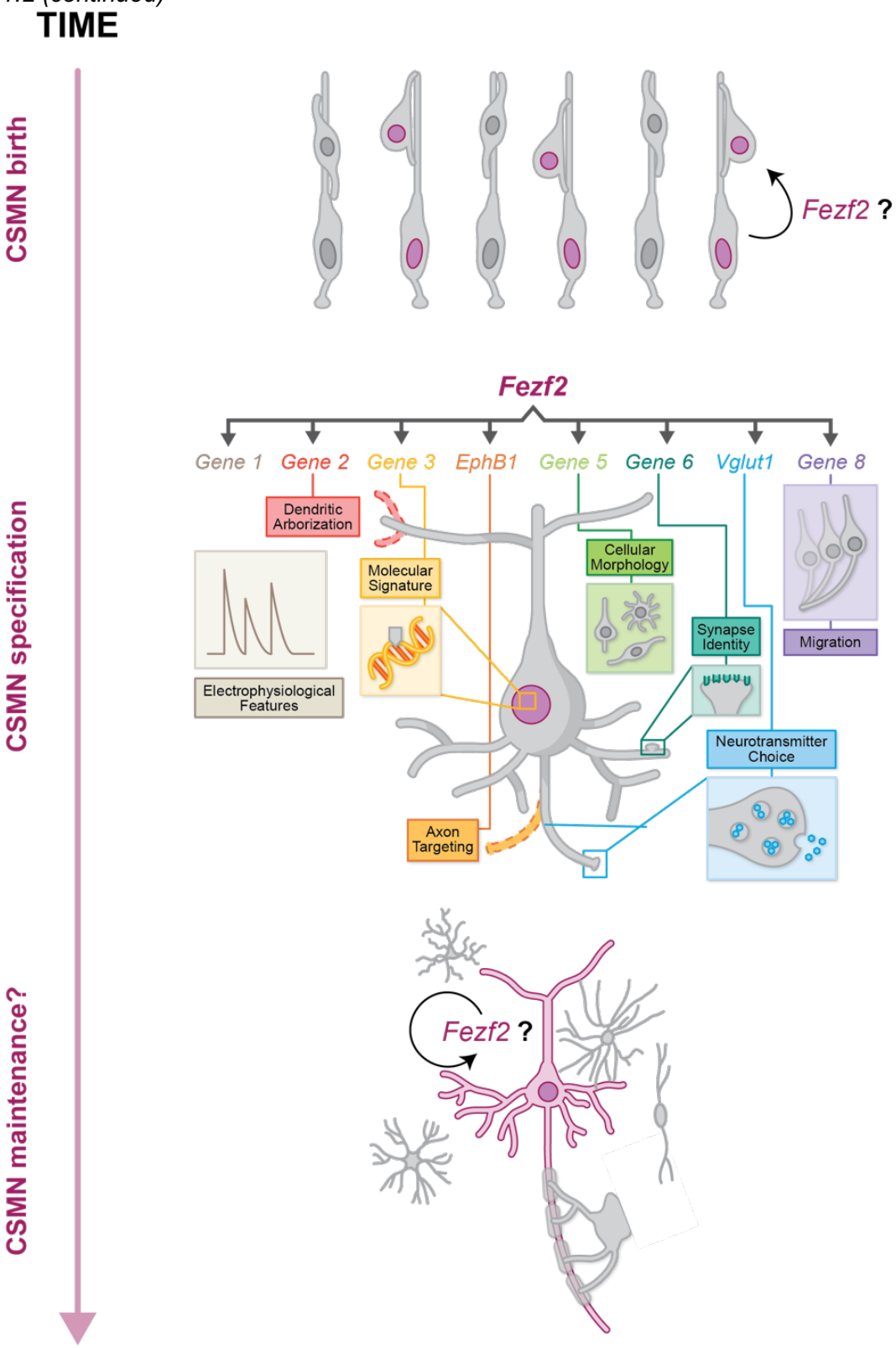
Given these observations and the maintained expression of *Fezf2* in CSMNs throughout adulthood, we hypothesized that Fezf2 is also a terminal selector gene for CSMNs, where postnatal deletion of Fezf2 would cause CSMNs to lose or change subtype-specific features, but maintain pan-neuronal traits (Figure 1.2). Should terminal selector genes exist in cortical neurons, it would demonstrate an active and reversible mechanism for neuronal maintenance, and would highlight a new form of plasticity in cortical neurons and helps us understand a component of the logic used to build neuronal diversity in the mammalian cortex.

In the adult cortex, there is preliminary evidence of molecular changes upon reduction of *Fezf2*. RNA-sequencing of bulk motor cortex tissue after lentiviral-mediated global knockdown of *Fezf2* detected many differentially expressed genes (Clare et al., 2017), suggesting that Fezf2 is not entirely dispensable in the adult. This type of transcriptional resolution, unfortunately, is not fine enough to determine whether *Fezf2* is necessary to maintain CSMN identity. Sequencing bulk, unfractionated tissue dilutes the putative molecular changes in CSMNs with the transcriptomes of all cell types, including CThPN, which also express Fezf2 and would confound the effects of Fezf2 deletion in CSMNs. It would also include any changes from other cell types that interact with CSMNs, such as the surrounding interneurons, which are capable of responding to molecular changes in surrounding projection neurons (Lodato et al., 2011; Wester et al., 2019;

FIGURE 1.2. HYPOTHESIZED ROLES OF FEZF2 IN CSMNS.

Fezf2 is expressed in the ventricular zone during CSMN birth. Constitutive loss of *Fezf2* prevents the birth of CSMNs. Fezf2 directly activates downstream effector genes, such as *EphB1* and *Vglut1*, to specify CSMN identity. Fezf2's continuous expression in adult CSMNs suggests it may be needed to maintain CSMN identity and mediate CSMN function in the adult cortex.

Figure 1.2 (continued)



Ye et al., 2015). Thus, the actual biological relevance of Fezf2 in adult CSMNs is still unknown and requires further study.

Goals for this Dissertation

Here, I aim to understand if Fezf2 is necessary to maintain the identity of mammalian CSMNs at the molecular and morphological levels. I establish strategies to effectively delete *Fezf2* at two distinct time points. I then use single-cell and single-nuclei RNA-sequencing to deeply examine molecular changes in CSMNs upon loss of *Fezf2* and understand the degree of necessity of *Fezf2* to maintain identity. I also examine a key morphological feature of CSMNs, cell body size, upon loss of *Fezf2*. I find that Fezf2 is necessary to maintain the molecular identity of CSMNs in early postnatal development, but not later in life. Furthermore, Fezf2 is not necessary to maintain the stereotypically large cell body size of CSMNs. Thus, Fezf2, despite being a selector gene for CSMNs, is surprisingly not necessary to maintain the full identity of SCPNs throughout the lifetime of the organism.

Chapter II.

Data: Fezf2 maintains the molecular identity of CSMNs in early postnatal life

This chapter is adapted from an as-yet-unpublished manuscript I wrote with guidance from Paola Arlotta.

Author contributions: I designed the study and wrote the manuscript with Paola Arlotta, with input from other authors. I did all experiments and analysis except: Fezf2 *constitutive KO FACS sorting and library preparation (Hsu-Hsin Chen, Simona Lodato);* Fezf2 *constitutive KO electrophysiology (John L. Sherwood);* Fezf2 *conditional knockout line generation (Juliana Brown); early* Fezf2 *deletion: library preparation, sequencing, and initial data cleanup (Xian Adiconis, Sean Simmons, Joshua Levin). I made all figures with input from the authors.*

Abstract

In the mammalian cerebral cortex, distinct classes of neurons are generated embryonically and maintain their subtype-specific identities throughout the lifetime of the organism. While their developmental programs have begun to be elucidated, less is known about identity preservation mechanisms. Here, we hypothesize that *Fezf2*, the first selector gene known for any class of projection neurons in the cortex, is also a terminal selector gene and is therefore needed to maintain corticospinal motor neuron (CSMN) identity throughout the life of the animal. We used two strategies to conditionally delete *Fezf2* specifically from CSMNs at distinct postnatal times. Our data shows that Fezf2 is only necessary to maintain CSMN identity in early postnatal animals, and not in the adult, unveiling a distinct window of nuclear plasticity. By exploiting *Fezf2* function in mature CSMNs, our work suggests that other mechanisms are used for neurons to maintain their identity and function in the mature cerebral cortex.

Introduction

We have made tremendous progress in recent years to understand the developmental mechanisms behind cortical neuron development (Arlotta et al., 2005; Molyneaux et al., 2009;

Molyneaux et al., 2015; Saunders et al., 2018; Tasic et al., 2016; Tasic et al., 2018), but much less is known about how these neurons, born embryonically, are subsequently maintained for the lifetime of the organism. Part of the difficulty in studying identity maintenance in mammalian cortical neurons is the diversity of neuronal subtypes. The mammalian neocortex is organized into six horizontal layers, with glutamatergic projection neurons (PNs) encompassing about 80% of the cortical neuronal population. PNs are classically further subdivided by hodology; callosal projection neurons (CPNs, also known as intertelencephalic, or IT neurons) reside in all layers and project within the cortical hemispheres, while the corticofugal projection neurons (CFuPNs) reside in the deep layers (layers V and VI) and project out of the cortex. CFuPNs can be further subdivided into subcerebral PNs (SCPNs, also known as extratelencephalic, or layer V ET neurons), which send projections to subcerebral targets such as the spinal cord and pons, and the corticothalamic PNs (CThPNs), which connect to the thalamus (see Lodato et al., 2015 for a comprehensive review). Morphological distinctions have allowed the isolation and transcriptional profiling of PNs during early time points which revealed many insights into developmental processes, but how mammalian subtype-specific PN identity is maintained in the adult is still largely unexplored.

We hypothesize here that terminal selector genes as used as part of the regulatory logic behind the diversity in the mammalian cerebral cortex, where a transcription factor is both necessary and sufficient for the specification of a cortical neuronal subtype, as well as necessary to maintain its subtype-specific neuronal identity. We believe Fezf2 is the strongest candidate to reveal key insights into maintaining cortical neuron identity in the cerebral cortex, because of its role as a selector gene for CSMNs.

First, Fezf2 is necessary for the birth and specification of CSMNs. Cells normally fated to become CSMNs in *Fezf2* constitutive knockout (KO) mice generate neurons similar to deep layer CPNs at morphological, molecular, and electrophysiological levels (Chen et al., 2005a; Chen et

al., 2005b; Chen et al., 2008; Molyneaux et al., 2005). Second, Fezf2 is sufficient to impose a CFuPN identity in many cell types, including post-mitotic neurons, which were previously thought to be immutable (De la Rossa et al., 2013; Rouaux and Arlotta, 2010; Rouaux and Arlotta, 2013, Ye et al., 2015). Finally, ChIP-seq and transcriptional data reveal that Fezf2 has an active role in specifying CSMN identity by activating CSMN downstream effector genes and repressing that of other neuronal lineages (Lodato et al., 2014). Notably, although the reprogramming capacity, as seen by the number of cells capable of changing any aspect of neuronal identity, of post-mitotic neurons drops precipitously during the first postnatal week (Rouaux and Arlotta, 2013; Ye et al., 2015), no other selector gene has been shown to be sufficient to reprogram post-mitotic neurons into that of another neuronal subtype.

Here, we established strategies to effectively delete *Fezf2* at two distinct time points and evaluated molecular and morphological changes in identity. We find that Fezf2, despite being a selector gene for CSMNs, is only necessary to maintain the molecular identity of CSMNs in early postnatal development, but not later in life.

Results

Fezf2 KO neurons lose their SCPN identity for a CPN identity

Previous work have established *Fezf2*'s role as a selector gene (Chen et al., 2005a; Chen et al., 2005b; Chen et al., 2008; Lodato et al., 2014; Molyneaux et al., 2005), but assessments of changes in molecular identity have only been done at the level of single markers, and the full extent of molecular changes have not been completely resolved. We therefore wanted to first fully understand *Fezf2*'s role in establishing CSMN identity before approaching the question of maintenance. To do this, we purified and compared the transcriptomes of control CSMNs from animals heterozygous for *Fezf2* (Fezf2 HZ) and layer Vb neurons in *Fezf2* full knockout animals (Fezf2 KO; Hirata et al., 2004) at postnatal day 1 (P1). We took advantage of the expression

pattern of two subtype-specific markers, CTIP2 (BCL11B) and SATB2, as well as Beta-Galactosidase in the *Fezf2* locus of the Fezf2 null allele, to intracellularly label layer Vb neurons in control and KO animals. We then used fluorescence activated cell sorting (FACS) and gated on the relative expression of the markers to collect our populations of interest for bulk RNA-sequencing (Supplementary Figure A2.1A-B).

As expected, we found that the biological replicates clustered together (Supplementary Figure A2.1C). Subtype-specific markers that were previously found to change in Fezf2 KO neurons were indeed found to change in our bulk RNA-sequencing dataset, as well as subtype-specific genes previously not associated with the loss of *Fezf2* (Supplementary Figure A2.1D, Supplementary Table A2.1). To get a comprehensive view of the degree of change, we tested for enrichment of cortical subtype-specific genes using pre-ranked Gene Set Enrichment Analysis (GSEA; Subramanian et al., 2005) in our ranked list of differentially expressed genes (DEGs) between Fezf2 HZ and Fezf2 KO neurons. The expression profile of Fezf2 KO neurons showed a significant enrichment of SCPN genes over CPN genes (Supplementary Figure A2.1E). Together, these data suggest that Fezf2 KO lose their SCPN molecular identity for that of CPNs.

Previous work suggests that layer Vb Fezf2 KO neurons convert their electrophysiological properties to that of CPNs (Chen et al., 2008), and we wanted to build on this work and have a complete picture of the intrinsic differences in electrophysiological properties between layer Vb neurons in Fezf2 WT and Fezf2 KO animals. We assayed for several intrinsic electrical properties. Measured parameters included action potential (AP) properties (threshold, rate of change during AP rise, amplitude, width, rheobase), resting membrane potential, membrane resistance, voltage sag, and adaptation ratio (Supplementary Figure A2.2). We see a significant increase in the input resistance of layer Vb Fezf2 KO neurons (Supplementary Figure A2.2F), which is in line with reduced dendritic arborization and smaller soma size in layer Vb Fezf2 KO neurons (Arlotta et al., 2005, Chen et al., 2005a; Chen et al., 2008). We also see a significant reduction in voltage sag

in layer Vb Fezf2 KO neurons (Supplementary Figure A2.2H), a known trademark of SCPNs (Mason and Larkman, 1990). Consistent with prior published work (Chen et al., 2008), we also see a difference in adaptation ratios between genotypes, with layer Vb Fezf2 KO neurons decelerating their firing rate during their action potential train (Supplementary Figure A.2.2I). Altogether, with our additional molecular and electrophysiology data, we have a comprehensive reference to benchmark how Fezf2 could maintain CSMN identity in postnatal life.

Fezf2 *can be conditionally ablated*

To address whether Fezf2 is necessary to maintain CSMN identity, we needed to temporally control the deletion of Fezf2. We generated animals with Fezf2 exons 3-5 flanked by loxP sites (Fezf2 cKO animals) to mediate Cre-dependent excision. To validate the functionality of the conditional allele, we crossed the Fezf2 cKO animals to Emx1-Cre, which express Cre recombinase in dorsal telencephalon progenitors (Gorski et al., 2002). We asked whether Fezf2 deletion in the progenitors of Emx1-Cre; Fezf2 cKO animals could phenocopy full Fezf2 KO animals. As expected, Emx1-Cre; Fezf2 cKO animals lost Fezf2 and other CSMN markers such as CTIP2 and *Ldb2* (Arlotta et al., 2005; Supplementary Figure A2.3), suggesting we had bred functional Fezf2 cKO animals.

Early deletion of Fezf2 *causes SCPNs to lose their molecular identity*

With an established Fezf2 cKO line, we needed a strategy to gain spatial control over *Fezf2* deletion and subsequent labeling. We crossed Fezf2 cKO animals to Tcerg1l-CreERT2, a CSMN-labeling CreERT2 knock-in driver line, along with Ai14, a Cre-inducible tdTomato reporter line (Madisen et al., 2010; Figure 2.1A-C). 4-hydroxytamoxifen (4OHT) injection at P3 in Tcerg1l-CreERT2; Ai14 mice indeed labels CTIP2 positive CSMNs with a small subset of other neuronal subtypes in other layers of the cortex, including CPNs and interneurons (Figure 2.1D).

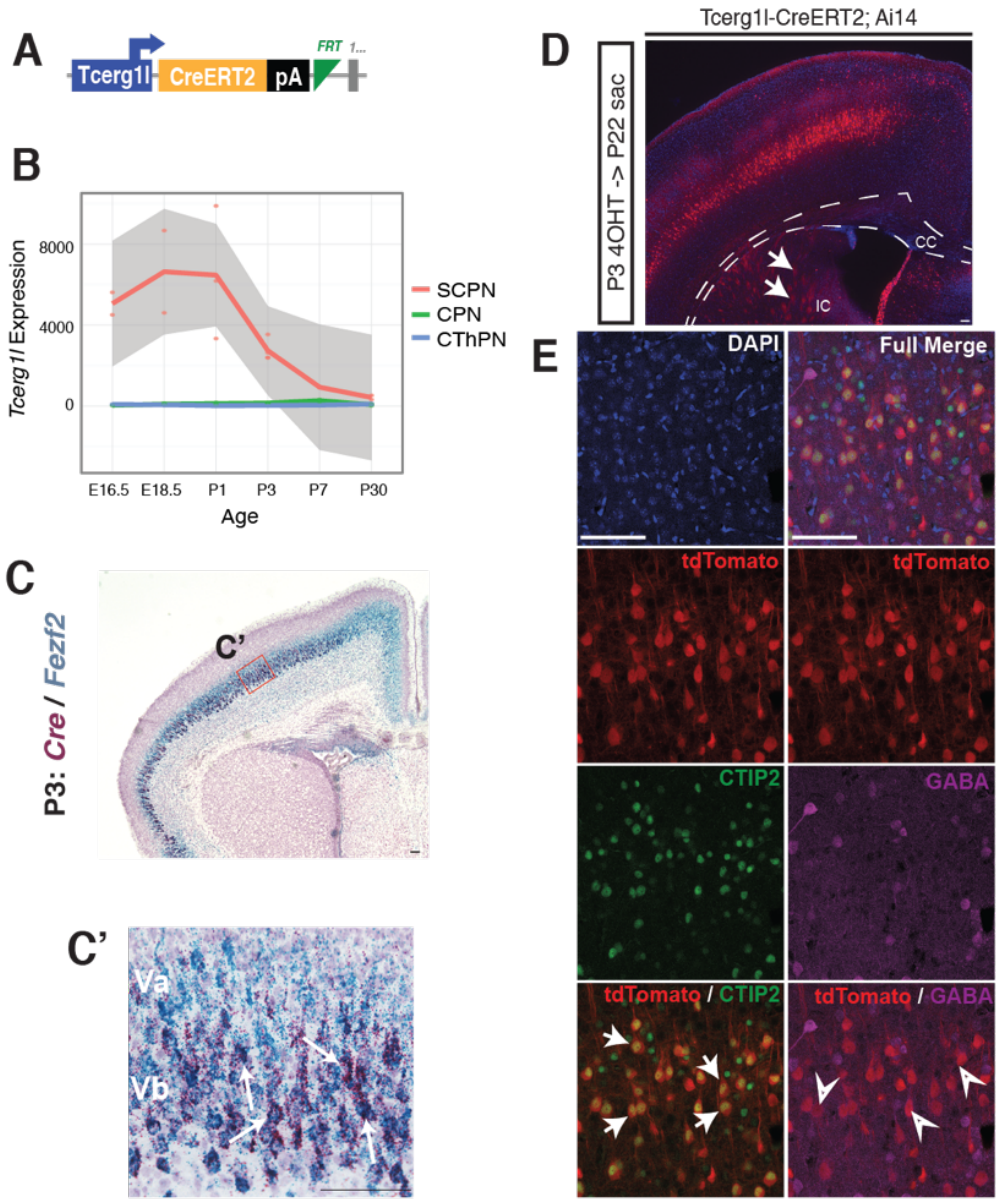


FIGURE 2.1. A GENETIC APPROACH TO DELIVER CRE TO CSMNS

A. Schematic of the tamoxifen-inducible Cre driver, Tcerg1l-CreERT2. ***B.*** *Tcerg1l* expression in subcerebral projection neurons (SCPNs), callosal projection neurons (CPNs), and corticothalamic projection neurons (CThPNs) over time shows *Tcerg1l*'s specificity in SCPNs in early development. ***C.*** RNAScope *in situ* hybridization shows co-expression of *Cre* and *Fezf2* in layer Vb at P3. ***D.*** 4-hydroxytamoxifen (4OHT) injection in Tcerg1l-CreERT2; Ai14 mice at postnatal day 3 (P3) labels projection neurons in layer Vb. Arrows point to axons that project through the internal capsule (IC). CC: corpus callosum. ***E.*** tdTomato+ cells in layer Vb are virtually all CTIP2+, GABA- neurons. Arrows: CTIP2 and tdTomato double positive cells; arrowheads: GABA negative and tdTomato positive cells. Scale bars are all 100 microns.

We wanted to first ask if Fezf2 was necessary to maintain CSMN identity at a stage when post-mitotic neurons still retain nuclear plasticity and are amenable to changing their identity (Rouaux and Arlotta, 2013). We therefore injected Tcerg1l-CreERT2; Fezf2 WT; Ai14 control animals and Tcerg1l-CreERT2; Fezf2 cKO; Ai14 animals with 4OHT at P3 and used FACS to purify recombined, tdTomato+ neurons from the whole cortex at P22 for bulk RNA-seq (Figure 2.2A, Supplementary Figure A2.4). In total, 532 genes were differentially expressed, with 215 genes downregulated in and 317 genes upregulated in Tcerg1l-CreERT2; Fezf2 cKO mice (FDR<0.05; Figure 2.2B, Supplementary Table A2.2). *Fezf2* is significantly downregulated in Tcerg1l-CreERT2; Fezf2 cKO; Ai14 animals (Figure 2.2C).

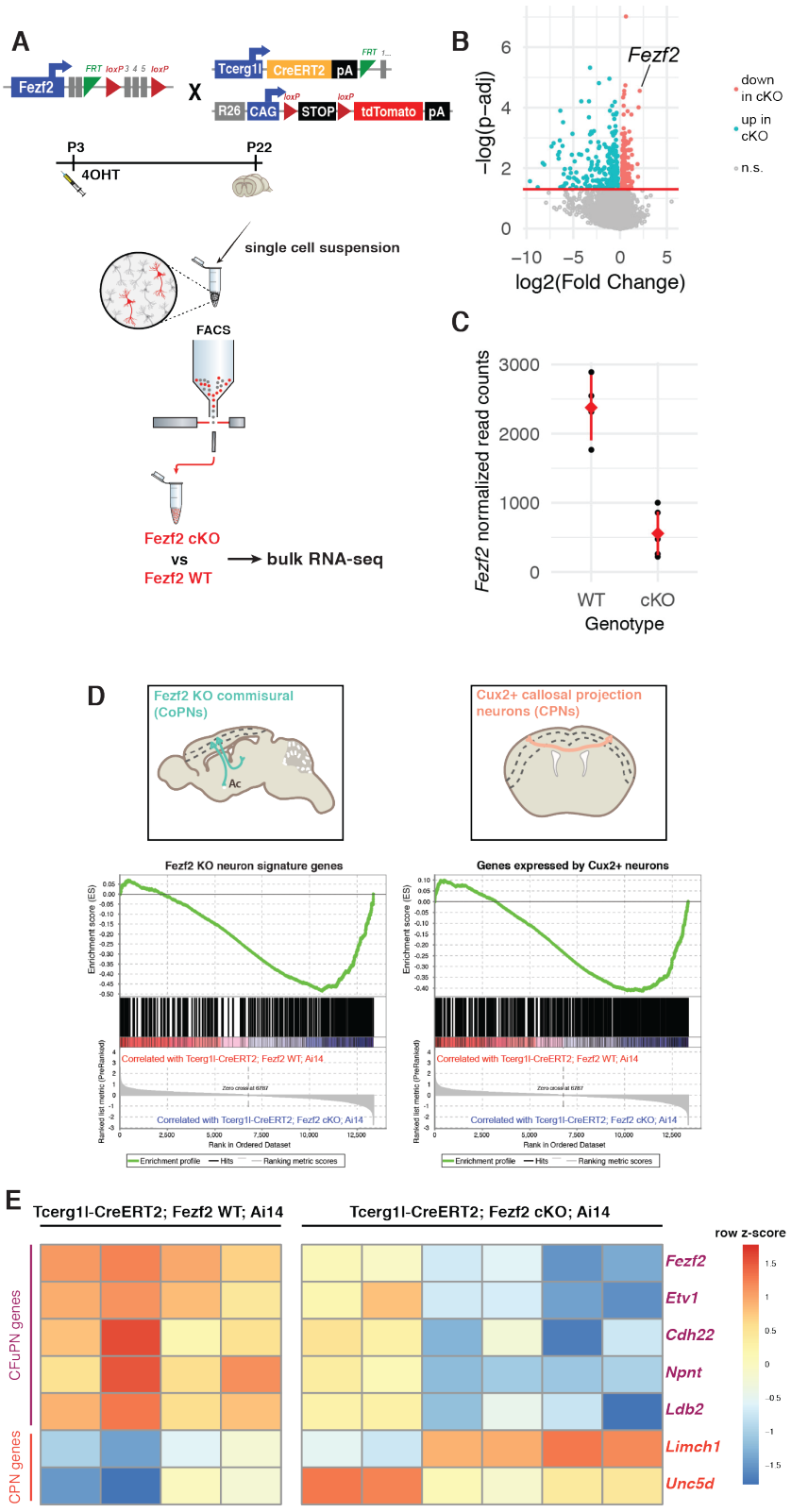
To directly assess changes in subtype-specific identity, we again used GSEA (Subramanian et al., 2005) to directly compare the transcriptional identity of Tcerg1l-CreERT2; Fezf2 cKO; Ai14 tdTomato positive neurons to that of known neuronal subtypes. As expected, these neurons strongly resemble the neurons generated in lieu of CSMNs in Fezf2 KO mice (Figure 2.2D-E). Specifically, they lose their CFuPN identity in favor of that of Cux2 positive, upper layer CPNs (Figure 2.2D-E). In addition, we also unbiasedly looked for pathways that change upon *Fezf2* deletion. We found that Tcerg1l-CreERT2; Fezf2 cKO; Ai14 neurons significantly lose nuclear-encoded mitochondrial genes and gain genes related to migration (Supplementary Figure A2.5; Supplementary Table A2.3).

The shift in subtype-specific identity upon early *Fezf2* deletion could be caused by three reasons. It could be from a subpopulation that either completely converted or partially converted identity, or contamination of non-SCPN, tdTomato positive cells from the Tcerg1l-CreERT2 line (Figure 2.1D). We therefore proceeded with single-cell RNA-seq to distinguish among the three possibilities.

FIGURE 2.2. BULK SEQUENCING OF TCERG1L-CREERT2; FEZF2 WT; AI14 VERSUS TCERG1L-CREERT2; FEZF2 CKO; AI14 NEURONS SHOWS A REDUCTION OF SCPN IDENTITY FOR CPN IDENTITY

A. Schematic of experimental approach to assess molecular changes with early postnatal *Fezf2* deletion. ***B.*** Volcano plot of differential expression. 532 genes are differentially expressed between Tcerg1l-CreERT2; Fezf2 WT; Ai14 and Tcerg1l-CreERT2; Fezf2 cKO; Ai14 neurons. Significance was determined with the Wald test as implemented in the DESeq2 R package, FDR <0.05. ***C.*** *Fezf2* normalized read counts in tdTomato positive sorted cells shows significant reduction in Tcerg1l-CreERT2; Fezf2 cKO; Ai14 neurons. ***D.*** Gene set enrichment analysis demonstrates that Tcerg1l-CreERT2; Fezf2 cKO; Ai14 neurons shift toward the molecular signatures of Fezf2 KO commuisural neurons and Cux2 positive CPNs. ***E.*** Selected cortical projection neuron signature genes and their expression in Tcerg1l-CreERT2; Fezf2 WT; Ai14 versus Tcerg1l-CreERT2; Fezf2 cKO; Ai14 neurons

Figure 2.2 (continued)



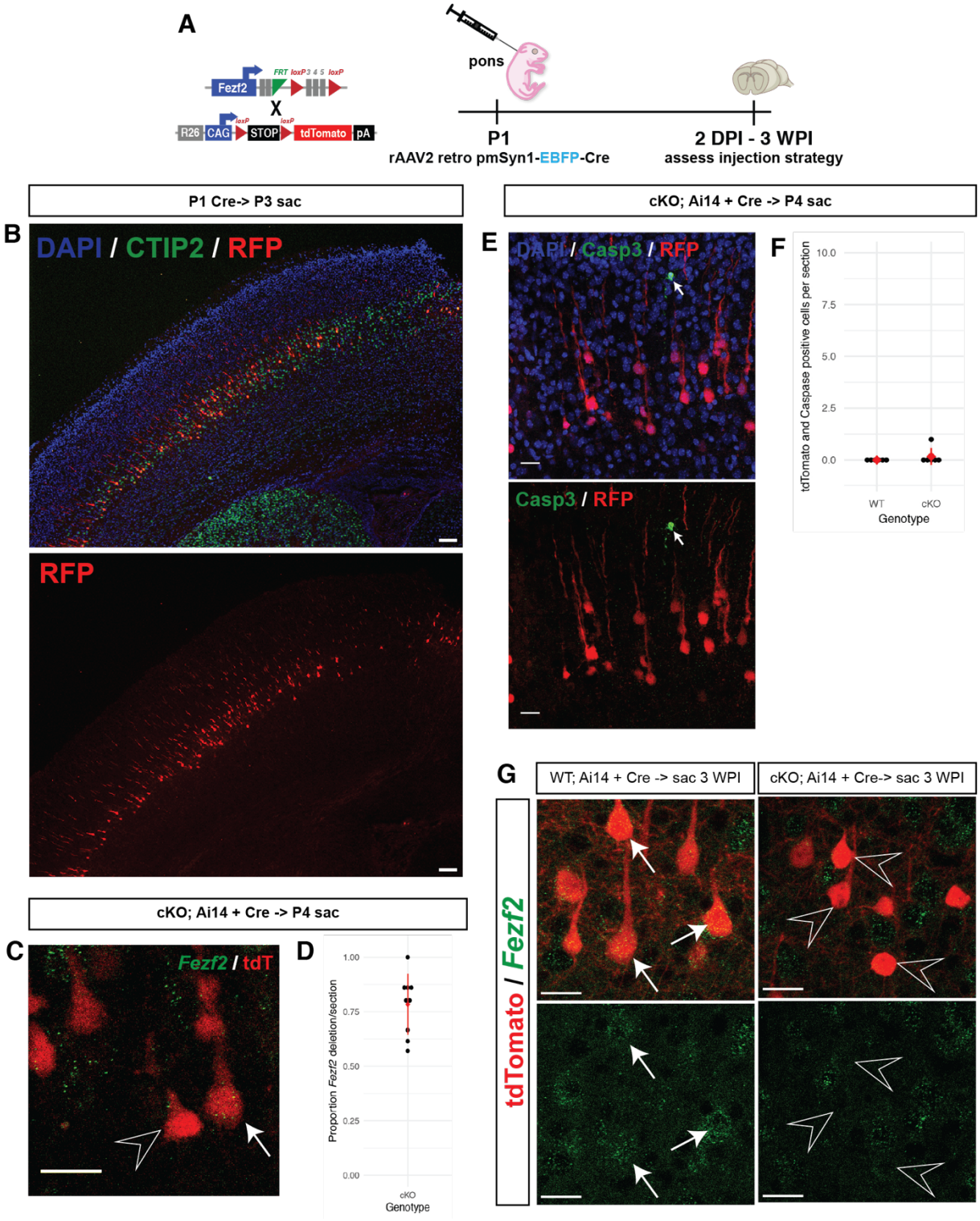
For the single-cell experiments, we established a cleaner method to conditionally ablate *Fezf2* from early post-mitotic CSMN. We took advantage of the long, stereotyped projections of CSMN and used an AAV pmSyn1-EBFP-Cre virus (Madisen et al., 2015) packaged with an rAAV-retro helper (Tervo et al., 2016) to retrogradely deliver Cre into CSMNs. To label neurons that express Cre, Fezf2 WT (WT) and Fezf2 cKO (cKO) animals were again crossed to Ai14, a Cre-dependent tdTomato reporter line (Figure 2.3A; Madisen et al., 2015). We injected Cre at P1 into the pons of WT; Ai14 or cKO; Ai14 animals which delivered Cre into CSMNs, as well as other SCPNs. tdTomato protein was present as early as two days post injection (DPI; Figure 2.3B), with 75% *Fezf2* deletion at 3 DPI (Figure 2.3C-D). Of note, terminal selector removal can cause cell death (Ninkovic et al., 2010; Serrano-Saiz et al., 2018), and we did not notice SCPN death with early *Fezf2* deletion. Cleaved Caspase3 staining at 3 DPI shows almost no tdTomato labeled cells dying (Figure 2.3E-F). By 3 weeks post injection (WPI), *Fezf2* was gone from SCPNs (Figure 2.3G). with no cleaved Caspase3 staining at 3 WPI (data not shown). Furthermore, a TUNEL stain at 3 WPI did not show signs of cell death upon *Fezf2* deletion (Supplementary Figure A2.6), and tdTomato labeled cells persist up to 40 WPI (data not shown).

We proceeded to clarify whether Fezf2 was necessary to maintain CSMN molecular identity in early postnatal animals. We injected Cre in the pons of P1 WT; Ai14 or cKO; Ai14 animals and waited 3 WPI to sacrifice animals for transcriptional analysis (Figure 2.4A) to follow an analogous timeline as with the bulk sequencing experiment (Figure 2.2). Live, tdTomato positive neurons were individually collected with fluorescent activated cell sorting (FACS) into 96 well plates for sequencing (Figure 2.4A, Supplementary Figure A2.7). We sequenced approximately 180 neurons per genotype, with over 140 neurons per genotype that passed quality control (Figure 2.4B). As expected, *Fezf2* is significantly downregulated in cKO; Ai14 neurons, and we proceeded with further analysis (Figure 2.4C).

FIGURE 2.3. P1 INJECTION OF RAAV2-RETRO-PSYN1-CRE DELIVERS CRE TO SUBCEREBRAL PROJECTION NEURONS (SCPNS).

A. Experimental design and timeline of characterization. DPI: days post injection; WPI: weeks post injection. ***B.*** tdTomato is expressed as early as 2 DPI in the cortex. CTIP2 (BCL11B) delineates layer Vb in the cortex. Scale bars are 100 micron. ***C.*** *Fezf2* is deleted as early as 3DPI by RNAscope *in situ* hybridization. Empty arrowhead: SCPN without *Fezf2*; arrow: SCPN with *Fezf2*. Scale bar is 25 microns. ***D.*** Quantification of *Fezf2* deletion in Figure 2.3C. ***E.*** Representative image of cleaved Caspase 3 staining 3 DPI. Arrow: Caspase 3 positive cell. Scale bars are 25 microns. ***F.*** Quantification of tdTomato and Caspase3 positive cells 3 DPI. Red dot: mean; red bars: standard deviation. There is no significant difference (two-sided t-test). ***G.*** *Fezf2* expression is gone from tdTomato+ cells at 3 WPI by RNAscope *in situ* hybridization. Empty arrowhead: SCPN without *Fezf2*; arrow: SCPN with *Fezf2*. Scale bar is 25 microns.

Figure 2.3 (continued)



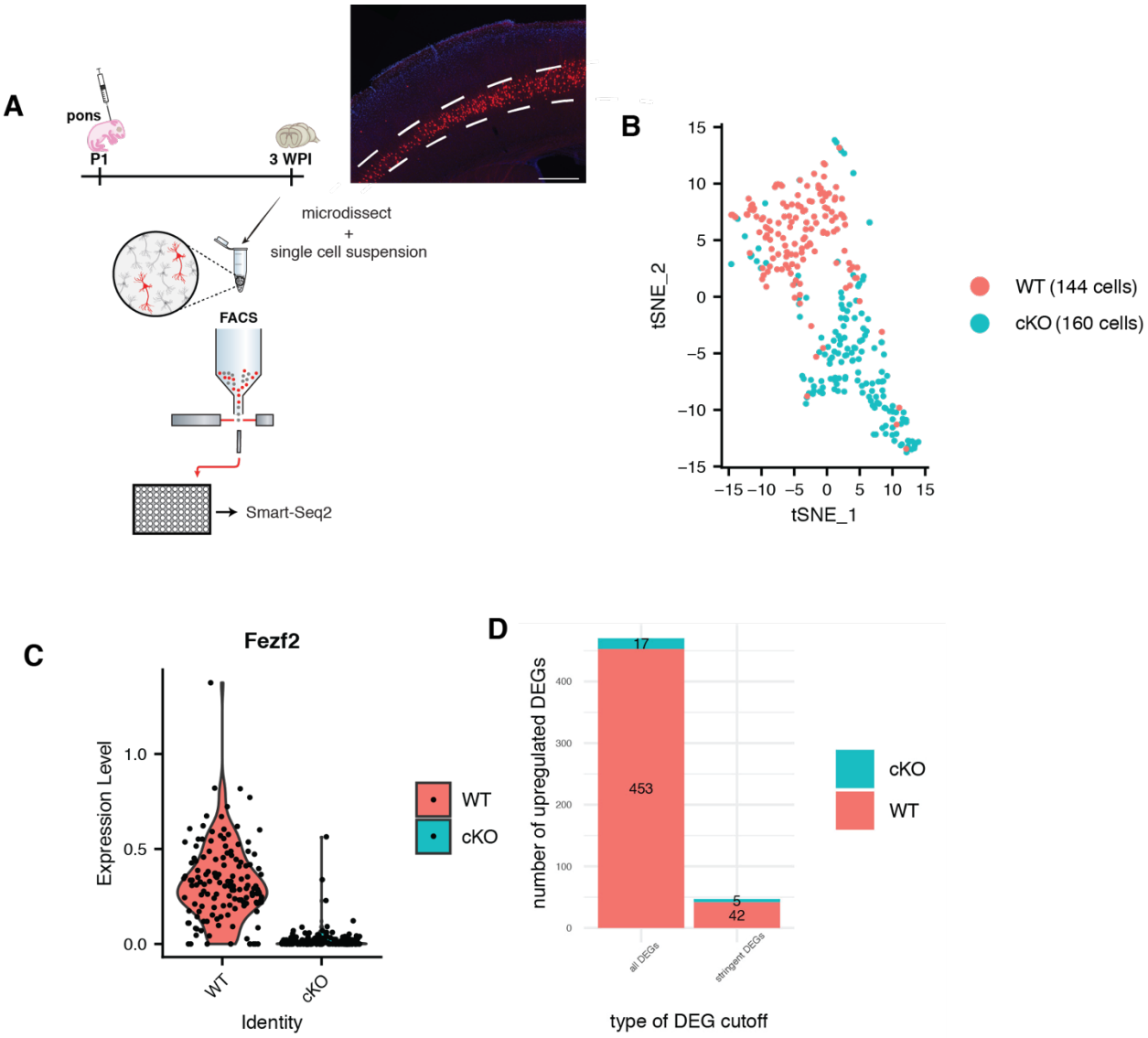


FIGURE 2.4. EARLY *FEZF2* DELETION IN SUBCEREBRAL PROJECTION NEURONS (SCPNS) SHOWS MOLECULAR CHANGES.

A. Experimental design and timeline of characterization. Layer Vb of control and experimental animals were microdissected for FACS. WPI: weeks post injection. Scale bar is 500 microns. ***B.*** Visualization of single cKO; Ai14 (cKO) and WT; Ai14 (WT) neurons by t-distributed stochastic neighbor embedding plot (TSNE) plot. ***C.*** Violin plot of *Fezf2* expression between genotypes. ***D.*** Quantification of number of differentially expressed genes (DEGs) between genotypes. All DEGs refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype.

To be thorough, we analyzed these data in four different ways to catch subtle changes in cKO; Ai14 neurons: with all sequenced cells, cells with only the most and least amount of *Fezf2,* only cKO; Ai14 neurons, and with adult neurons from the Allen Brain Institute as a reference

(Tasic et al., 2016). Single-cell RNA-seq gives us the ability to filter differentially expressed genes (DEGs) based on the percentage of cells that express the genes of interest between conditions as well as by levels of expression between conditions. Therefore, for all analyses, in addition to looking at DEGs that are widely expressed across all cells between genotypes, we further filtered our DEGs by imposing a minimum 10% difference in the number of expressing cells between genotypes, which we call our "stringent DEGs."

In our first analysis, we analyzed all cells from both genotypes together. As compared to WT; Ai14 neurons, cKO; Ai14 neurons predominantly downregulate genes (Figure 2.4D, Supplementary Table A2.4). The number of downregulated genes in cKO; Ai14 neurons decreases over 90% when we apply our stringent cutoff (Figure 2.4D), suggesting that cKO; Ai14 neurons are more likely to lose aspects of, rather than convert, their molecular identity to that of another subtype.

We first evaluated the DEGs by either searching for enrichment of known biological processes, cellular components, and molecular function with EnrichR (E. Y. Chen et al., 2013; Kuleshov et al., 2016). Given the few genes upregulated in cKO; Ai14 neurons, it is unsurprising that no pathways are significantly enriched in cKO; Ai14 neurons. Interestingly, of the genes downregulated in cKO; Ai14 neurons, we see a substantial enrichment decrease of pathways involved in mitochondrial function in cKO; Ai14 neurons (Figure 2.5A), in agreement with our previous bulk sequencing data (Supplementary Figure A2.5; Supplementary Table A2.3). This enrichment, however, disappears with our stringent cutoff (Supplementary Table A2.5).

We also directly tested whether cKO; Ai14 neurons gained or lost subtype-specific genes using hypergeometric testing for enrichment of known subtype-specific or developmental stage-specific genes in the DEGs (see Appendix 1 for how gene sets were generated). cKO; Ai14 neurons predominantly reduce expression of SCPN-enriched genes and gain expression of a few

upper layer cortical neuron-enriched genes (Figure 2.5B; Supplementary Figure A2.8;

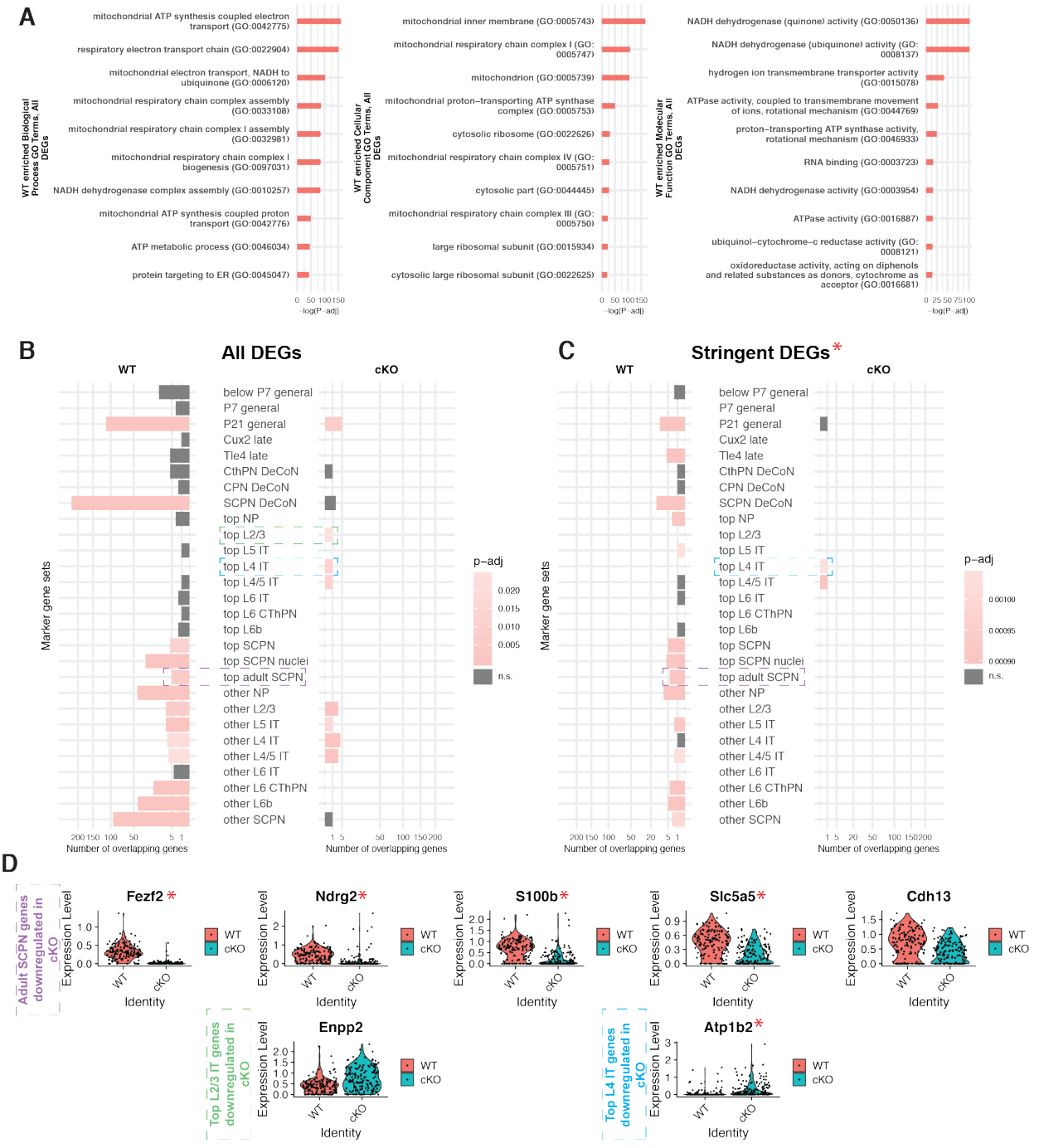


FIGURE 2.5. DEGS FROM EARLY *FEZF2* DELETION SHOWS A LOSS OF NUCLEAR-ENCODED MITOCHONDRIAL GENES AND A REDUCTION OF SCPN IDENTITY.

A. The top 10 most strongly associated gene ontology (GO) groups in the DEGs enriched in WT; Ai14 (WT) neurons as defined by biological process, cellular component, molecular function, ranked by log adjusted p-value. DEGs enriched in WT neurons show strong enrichment of nuclear-encoded mitochondrial genes. P-values are from the Fisher exact test, and adjusted p-values are reported with

Figure 2.5 (continued) Benjamini-Hochberg multiple hypothesis correction. ***B.*** All DEGs enriched in various cortical subtype-specific or cortical neuron development gene sets. DEGs enriched in WT neurons show an enrichment of SCPN-related genes as compared to DEGs enriched in cKO; Ai14 (cKO) neurons. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***C.*** Stringent DEGs enriched in various cortical subtype-specific or cortical development gene sets. DEGs enriched in WT neurons still show an enrichment of SCPN-related genes as compared to DEGs in cKO neurons, but the number of genes is drastically reduced. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***D.*** Violin plots of DEGs enriched in the 'top adult SCPN,' 'top L2/3,' and 'top L4 IT' gene sets highlighted previously. Red asterisks denote stringent DEGs.

Supplementary Table A2.6). This trend persists when we apply our stringent DEG cutoff, but the number of overlapping genes in each gene set decreases (Figure 2.5C; Supplementary Figure A2.8; Supplementary Table A2.6). Of note, nuclear-encoded mitochondrial genes are highly represented in the 'SCPN DeCoN' gene set and many are downregulated in cKO; Ai14 neurons (Supplementary Figure A2.9), consistent with the unbiased pathway analysis with EnrichR (Figure 2.5A).

When we search for well-recognized, known subtype-specific genes and those related to neurotransmitter identity, we do not see a significant difference between genotypes (Supplementary Figure A2.10). These data suggest that while there is a reduction in deep-layer identity genes and gain of select upper-layer identity genes in cKO; Ai14 neurons, these cells do not fully lose or gain expression of the full battery of subtype-specific genes, and only some subtype-specific genes remain amenable to change.

Our second analysis examined only the top 50% of WT; Ai14 cells and bottom 50% of cKO; Ai14 cells as ranked by *Fezf2* expression (Supplementary Figure A2.11A). Prior studies have shown a reduction of nuclear plasticity over time (Rouaux and Arlotta, 2013; Ye et al., 2015), and it is therefore possible that only a small population of cKO; Ai14 neurons change identity upon *Fezf2* deletion. Because our previous line of analysis of combining the two genotypes could mask any subtle changes in a rare cKO; Ai14 population, we wanted to also actively search for a potential rare population. This analysis also addresses technical limitations of *Fezf2* deletion,

because it is possible that residual *Fezf2* RNA can maintain the expression of its downstream genes and skew our previous analysis.

After subsetting our WT; Ai14 and cKO; Ai14 populations (Supplementary Figure A2.11A-B), we looked for DEGs between the two genotypes (Supplementary Figure A2.11C; Supplementary Table A2.7), as described previously. We still see a significant de-enrichment of nuclear encoded mitochondrial genes in cKO; Ai14 neurons, but this enrichment disappears when we apply our stringent cutoff (Supplementary Figure A2.12A-C; Supplementary Table A2.8).

When we test for enrichment of subtype-specific genes across all DEGs with this analysis, we again see that cKO; Ai14 neurons predominantly reduce expression of the top SCPN specific genes and gain expression of a few intertelencephalic (IT)-specific genes (Supplementary Figure A2.12D-F). This analysis suggests that even in our best-case scenario with neurons that have completely lost *Fezf2*, most differentially expressed SCPN-specific genes are only reduced in cKO; Ai14 neurons, instead of completely eliminated with *Fezf2* deletion.

Our third analysis, analyzing only cKO; Ai14 neurons, also tests the hypothesis in which only a rare subset of cKO; Ai14 neurons responded to *Fezf2* deletion. Should there be a rare population of “reprogrammed” cKO; Ai14 neurons, they would molecularly look significantly different than most cKO; Ai14 neurons. The would have a signature of either CPN identity, reminiscent of Fezf2 full KO neurons, or a mixed SCPN/CPN identity. Because in the opposite scenario, when we overexpress *Fezf2* in CPNs, we see a reprogramming efficiency of less than 10%, we predicted that about 10% of cells would be amenable to change without *Fezf2*, and therefore used a resolution that produced subclusters with cells of approximately 10% of the total cKO population. This subclustered our cKO population into five populations, with the smallest subcluster containing only 18 cells (Supplementary Figure A2.13A-B). Notably, *Fezf2* is not differentially expressed between clusters (Supplementary Figure A2.13C), suggesting that the separation of the clusters is not driven by amounts of *Fezf2* in cKO; Ai14 neurons.

We then looked into the marker genes of each subcluster, to see whether any subclusters separated based on expression of subtype-specific identity genes or unique molecular pathways. When we looked into the marker genes that separated the smallest cluster, c4, we found largely unannotated genes (Supplementary Table A2.9). Clusters c0 and c3 had no significantly upregulated DEGs (Supplementary Figure A2.13D). These data suggest that we overclustered some neurons and went beyond the detection capacity of Smart-Seq2.

For the larger clusters, c1 and c2, we looked for enrichment of unique molecular pathways in their DEGs. With only four upregulated genes as compared to all other clusters (Supplementary Figure A2.13D), cluster c1 had no significant enrichment of known pathways and minimal enrichment of known subtype-specific genes (Supplementary Table A2.10). When compared to all other clusters, cluster c2 had an enrichment of genes involved in neuron projection morphogenesis, regulation of synaptic vesicle exocytosis, and the endoplasmic reticulum (Supplementary Figure A2.14A; Supplementary Table A2.11). On the other hand, clusters c0, c1, c3, and c4 had a reduction in subtype-specific genes belonging to deep layer neurons as compared to cluster c2 (Supplementary Figure A2.14B). This corroborates our finding from previous analyses that *Fezf2* deletion reduces the expression of deep-layer subtype-specific genes. Furthermore, we conclude that there are subsets of cKO; Ai14 neurons that reduce the expression of identity genes more so than others.

Finally, we wanted to directly compare our three week old WT; Ai14 and cKO; Ai14 neurons to all adult wildtype, glutamatergic cortical neurons, as our previous analysis only looked at single DEGs and lacked other transcriptional context. We addressed this in two ways. First, we integrated our data with the Allen Brain's dataset of single glutamatergic neurons from adult mouse cortex. In this analysis, our cKO; Ai14 neurons separate more from endogenous SCPNs as compared to their WT; Ai14 counterparts, which is in agreement with our previous analysis (Figure 2.6A). In a second parallel method, we projected the Allen Brain's dataset onto ours to

classify our cells based on their reference data. With this projection, cKO; Ai14 neurons have a higher proportion of cells classified as “L4/5 IT cells” (intertelencephalic, also known as CPNs) instead of “Layer 5 extratelencephalic cells” (ET, also known as SCPNs; Figure 2.6B). When we examine the distribution of prediction scores, we see that cKO; Ai14 neurons have a wider spread of prediction scores for L5 ET and L4/5 IT cells as compared to that of other cell types. cKO; Ai14 neurons have more cells with lower L5 ET prediction scores and higher L4/5 IT prediction scores (Figure 2.6C-D), but they do not look completely like L4/5 IT cells, as seen by the lack of cKO; Ai14 neurons with a predicted score of 1 (Figure 2.6D). These data clarify our previous single-cell analyses and bulk sequencing data: without *Fezf2*, there is a gradient in terms of identity change, where a subset of cKO; Ai14 neurons drift away from their SCPN identity toward that of a intertelecephalic neuron or CPN, but do not necessarily fully change subtype-specific identity. Together, these data suggest that *Fezf2* is necessary to maintain SCPN identity in early postnatal neurons.

We wondered how this gradient of identity loss occurs. A gradient effect could be because of varying amounts of *Fezf2* in cKO; Ai14 neurons; however, the predicted L5 ET score is not correlated with *Fezf2* expression, suggesting the amount of *Fezf2* is not causing the gradient of molecular change (Supplementary Figure A2.15A). Another hypothesis is that the biological age of the neuron during recombination affects the capacity of the neuron to lose its identity, with neurons that received *Cre* earlier more likely to lose SCPN identity. A proxy of when neurons receive Cre is of *Cre* expression*,* with neurons that received the Cre virus earlier theoretically expressing more *Cre*. There is a slight negative correlation between the levels of *Cre* and predicted L5 ET score (Supplementary Figure A2.15B), suggesting that with more Cre and a younger biological age of recombination, cKO neurons look less like L5 ET neurons. This corresponds to neurons losing their nuclear plasticity with age. We therefore speculate that the recombination timing is one reason for the gradient of change in identity.

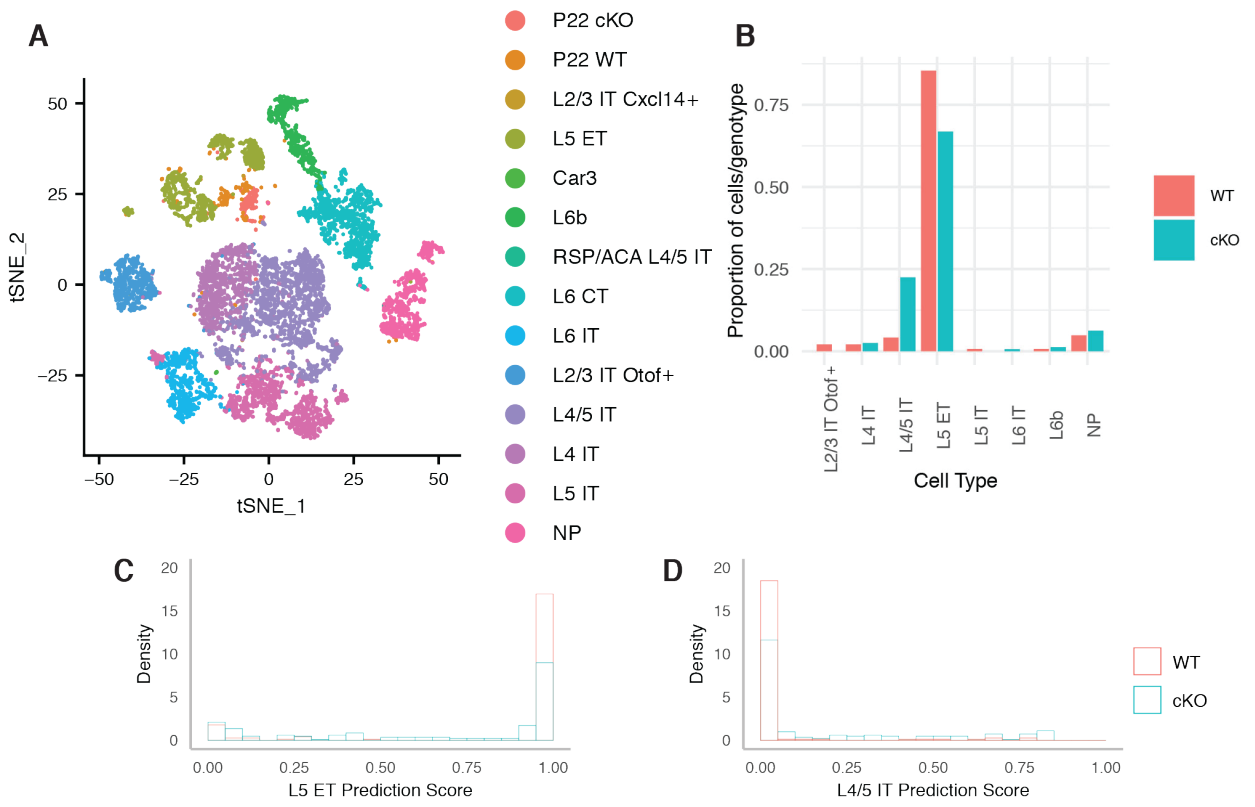


FIGURE 2.6. COMPARISON OF WT; AI14 AND CKO; AI14 NEURONS TO SINGLE, ADULT CORTICAL NEURONS REVEALS A SHIFT AWAY FROM SCPN IDENTITY IN CKO; AI14 NEURONS.

A. Visualization of merged Fezf2 WT; Ai14 (P22 WT) and Fezf2 cKO; Ai14 (P22 cKO) neurons with P56 adult cortical neurons. ***B.*** Proportion of P22 WT and P22 cKO neurons and their predicted adult cortical neuron cell type identification. ***C.*** Distribution of L5 ET neuron prediction scores by genotype. ***D.*** Distribution of L4/5 IT neuron prediction scores by genotype.

Early Fezf2 deletion does not change cell body size

Because of the shift in molecular identity, we were curious if early loss of *Fezf2* would cause morphological changes. Upon constitutive loss of *Fezf2*, we see that the neurons born in the place of CSMNs exhibit a different morphology, with a smaller cell body and less dramatic dendritic arborization (Chen et al., 2005b; Chen et al., 2008; Molyneaux et al., 2005). We only focused on cell body size upon early *Fezf2* deletion. Surprisingly, we did not see a significant difference in cell body size between genotypes (Figure 2.7).

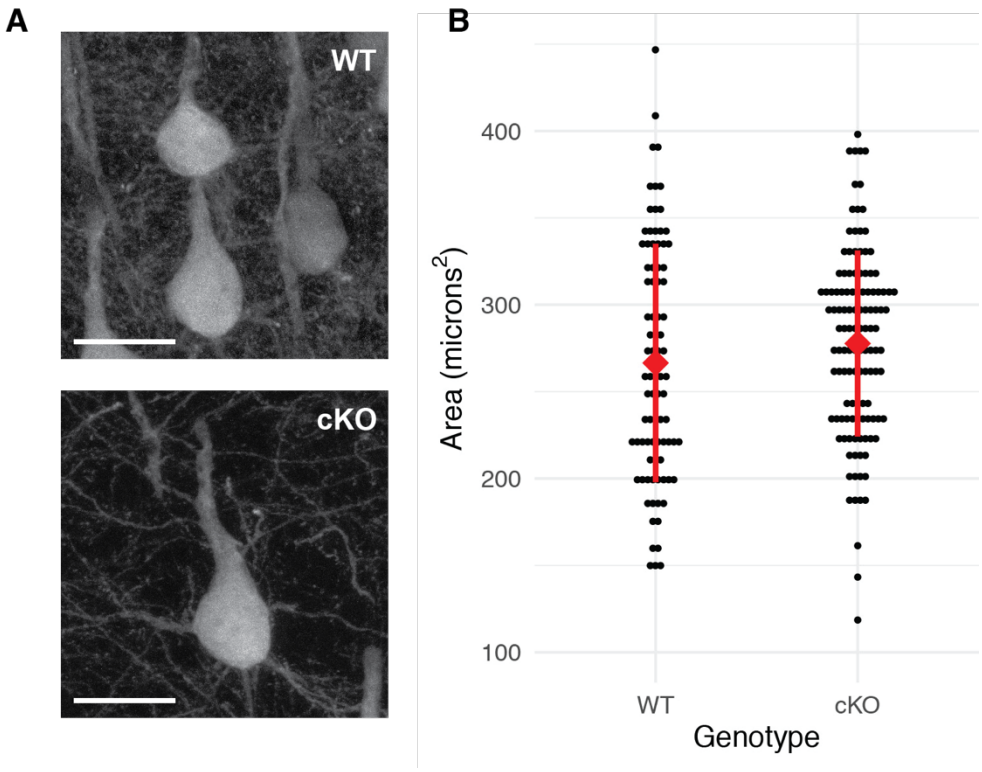


FIGURE 2.7. EARLY *FEZF2* DELETION DOES NOT CHANGE CELL BODY SIZE.

A. Representative z-stack images of tdTomato positive neurons. Scale bars are 25 microns. ***B.*** Quantification of cell body size. Each point represents one cell. Red diamond indicates mean, red bars indicate one standard deviation. There is no significant difference between genotypes (nested ANOVA test).

Late deletion of Fezf2 *does not change subtype-specific identity*

To address whether Fezf2 is necessary to maintain CSMN identity throughout the lifetime of the organism, we wanted to delete *Fezf2* at a later time point. We first optimized our injection strategy by injecting into the pons stereotaxically at P30 in WT and cKO mice crossed to Ai14 tdTomato reporter mice (Supplementary Figure A2.16A). Injected animals showed Cre-mediated recombination of the tdTomato and Fezf2 loci in layer V SCPN (Supplementary Figure A2.16B-C), and we again do not see cell death according to a TUNEL assay (Supplementary Figure A2.17).

Because of the difficulty in isolating live neurons at late timepoints, we adapted our strategy to collect nuclei instead of live neurons. We therefore crossed WT and cKO mice to Sun1-GFP mice, which have Cre-dependent expression of nuclear membrane-bound GFP (Mo et al., 2015). Using stereotaxic injection, we injected the virus into the pons of WT; Sun1-GFP and cKO; Sun1-GFP mice at P30, and collected GFP positive nuclei from the whole cortex for single nuclei sequencing with 10x-Chromium sequencing five weeks after (Figure 2.8A; Supplementary Figure A2.18). We sequenced 10,124 and 7,609 nuclei in cKO; Sun1-GFP and WT; Sun1-GFP animals, with 9,575 and 6,382 nuclei that were used for downstream analysis, respectively (Figure 2.8B). As expected, *Fezf2* is significantly downregulated in cKO; Sun1-GFP nuclei (Figure 2.8C).

We decided to analyze our data in two ways: with all SCPNs and with only putative CSMNs. We again further filtered our DEGs by imposing a minimum 10% difference in the number of expressing cells between genotypes, which we call our “stringent DEGs.”

In our first analysis, we analyzed all SCPNs from both genotypes together. As compared to WT; Sun1-GFP nuclei, cKO; Sun1-GFP nuclei predominantly upregulate genes (Figure 2.8D; Supplementary Table A2.12); this in contrast with early *Fezf2* deletion, where cKO neurons predominantly downregulate genes (Figure 2.4D). The number of upregulated genes in cKO; Sun1-GFP nuclei decreases from 58 genes to 42 genes when we apply our stringent cutoff (Figure 2.8D).

We first evaluated the DEGs by either searching for enrichment of known biological processes, cellular components, and molecular function with EnrichR (E. Y. Chen et al., 2013; Kuleshov et al., 2016). Given the few genes that change, it is unsurprising that no pathways are significantly enriched, except for the enrichment of 3’, 5’-cyclic-AMP phosphodiesterase activity in cKO; Sun1-GFP nuclei without the stringent DEG cutoff (Figure 2.9A; Supplementary Table A2.13).

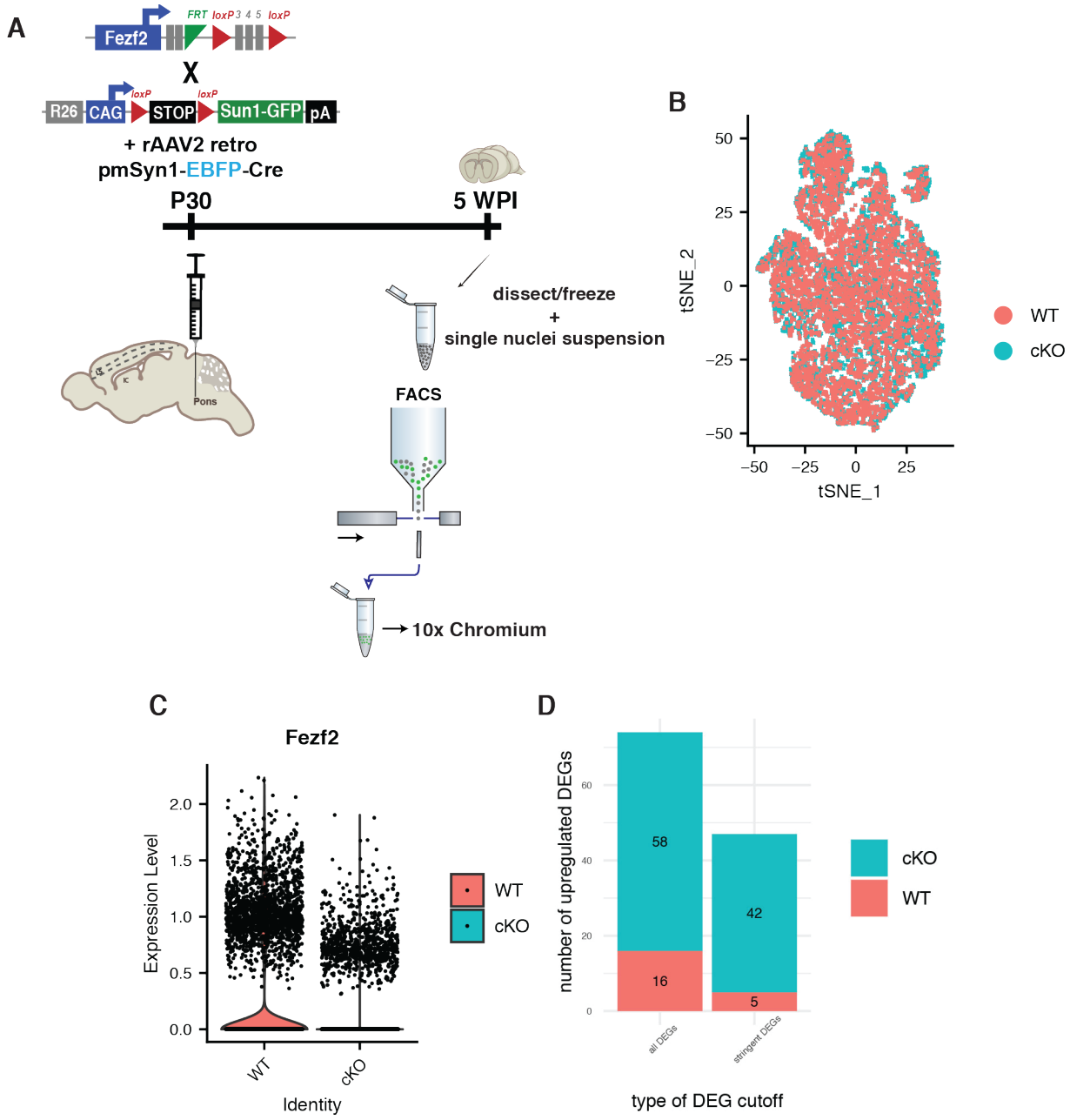


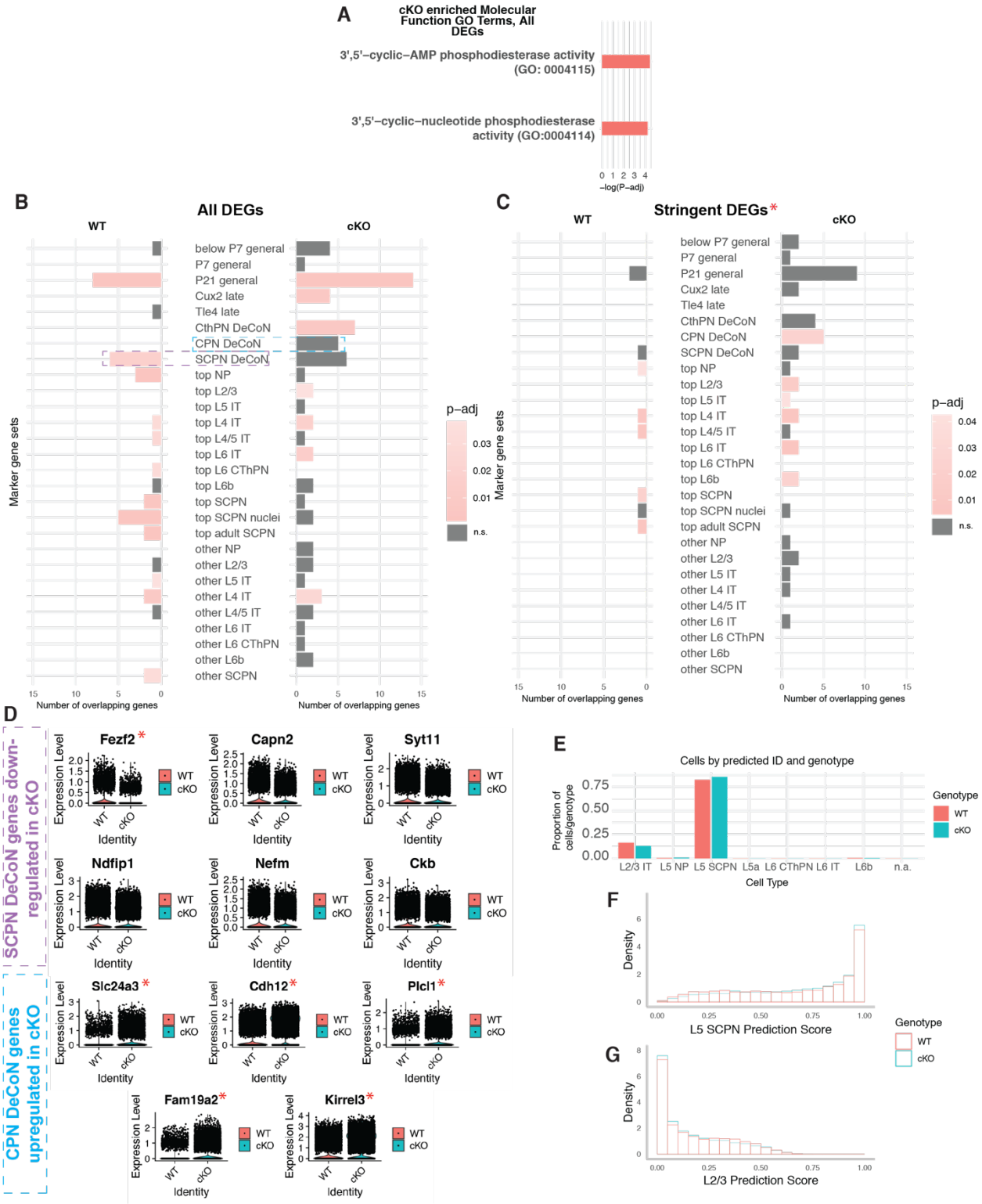
FIGURE 2.8. LATE *FEZF2* DELETION IN SUBCEREBRAL PROJECTION NEURONS (SCPNS) SHOWS MOLECULAR CHANGES.

A. Experimental design and timeline of characterization. WPI: weeks post injection. ***B.*** Visualization of single Fezf2 cKO; Sun1-GFP (cKO) and Fezf2 WT; Sun1-GFP (WT) nuclei by t-distributed stochastic neighbor embedding plot (TSNE) plot. ***C.*** Violin plot of *Fezf2* expression between genotypes. ***D.*** Quantification of differentially expressed genes (DEGs) between genotypes. All DEGs refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype.

FIGURE 2.9. LATE *FEZF2* DELETION SHOWS MOLECULAR CHANGES, BUT NOT A LOSS OF SCPN IDENTITY.

A. The only significantly associated GO groups in the DEGs are enriched in cKO; Sun1-GFP (cKO) neurons. Gene sets are ranked by log adjusted p-value. P-values are from the Fisher exact test, and adjusted p-values are reported with Benjamini-Hochberg multiple hypothesis correction. ***B.*** All DEGs enriched in various cortical subtype-specific or cortical neuron development gene sets. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***C.*** Stringent DEGs enriched in various cortical subtype-specific or cortical development gene sets. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***D.*** Violin plots of DEGs enriched in the 'SCPN DeCoN' and 'CPN DeCoN' gene sets highlighted previously. Red asterisks denote stringent DEGs. ***E.*** Proportion of 5 WPI nuclei and their predicted adult cortical neuron cell type identification (Allen Brain Institute, unpublished data). ***F.*** Distribution of L5 SCPN neuron prediction scores by genotype. ***G.*** Distribution of L2/3 IT neuron prediction scores by genotype.

Figure 2.9 (continued)



We also directly tested whether cKO; Sun1-GFP nuclei gained or lost subtype-specific genes using hypergeometric testing for enrichment of known subtype-specific or developmental stage-specific genes in the DEGs. cKO; Sun1-GFP nuclei reduce expression of at most six SCPN genes, including *Fezf2*, and gain expression of at most 5 CPN genes (Figure 2.9B-D; Supplementary Table A2.14). When we search for well-recognized, known subtype-specific genes, we do not see a significant difference between genotypes (Supplementary Figure A2.19). Indeed, when we project the Allen Brain's dataset of single, frozen glutamatergic nuclei from adult mouse primary motor cortex (unpublished) onto ours to classify our nuclei, WT; Sun1-GFP and cKO; Sun1-GFP nuclei have similar proportions of nuclei that are classified as L5 SCPN (Figure 2.9E-F). There are also nuclei classified as L2/3 IT cells, but the distribution of prediction scores and proportion of nuclei is similar between genotypes (Figure 2.9G). Together, these data suggest the late deletion of *Fezf2* does not cause SCPNs to lose their subtype-specific identity.

Because we injected the virus in the pons, we are capturing other SCPNs in addition to CSMNs which could confound our previous analysis. Therefore, for our second analysis, we only analyzed putative CSMNs, as defined by the expression of *Bcl11b* and low expression of *Satb2* and *Tshz2* (Harb et al., 2016; Tasic et al., 2016), and asked whether there were substantial differences between genotypes (Supplementary Figure A2.20A). As compared to our previous analysis with all SCPNs, we see more upregulated genes in cKO; Sun1-GFP nuclei (Supplementary Figure A2.20B; Supplementary Table A2.15). Specifically, 69 DEGs are shared between both analyses, with 10 DEGs unique to the "CSMN only" analysis, and 5 DEGs unique to the "all SCPN analysis" (Supplementary Figure A2.20C-E).

We again evaluated the DEGs by either searching for enrichment of known biological processes, cellular components, and molecular function with EnrichR (E. Y. Chen et al., 2013; Kuleshov et al., 2016). Similar to our previous analysis, we only see GO enrichment in DEGs upregulated in cKO; Sun1-GFP nuclei without the stringent cutoff, with a total of 7 unique genes

encompassing 6 GO terms (Supplementary Figure A2.21A; Supplementary Table A2.16). Notably, we do not have a significant enrichment of genes related to nuclear-encoded mitochondrial genes, which we saw previously with early *Fezf2* deletion.

We also tested for the gain or loss of subtype-specific or developmental stage-specific genes in the DEGs. Similar to our previous analysis, cKO; Sun1-GFP nuclei reduce expression of at most five SCPN genes, including *Fezf2*, and gain expression of at most 5 CPN genes (Supplementary Figure A2.21B-C; Supplementary Table A2.17). Finally, we again project the Allen Brain's dataset of single, frozen glutamatergic nuclei from adult mouse primary motor cortex onto our data, and find WT; Sun1-GFP and cKO; Sun1-GFP nuclei have similar proportions of nuclei that are classified as L5 SCPN (Supplementary Figure A2.21D-E), as we saw in the previous analysis (Figure 2.9E-F). There are also nuclei classified as L2/3 IT cells, but the distribution of prediction scores and proportion of nuclei is again similar between genotypes and analyses (Supplementary Figure A2.21F, Figure 2.9G). We therefore conclude that the late deletion of *Fezf2* does not change subtype-specific identity, and that Fezf2's capacity to maintain CSMN identity decreases over time.

Late Fezf2 *deletion does not change cell body size*

We wanted to examine the morphology of neurons with late Fezf2 deletion, and again focused on cell body size. Instead of the Sun1 nuclear GFP reporter mice, WT and cKO mice were again crossed to Ai14 tdTomato reporter mice to resolve the morphology of the neuron. Using stereotaxic injection, we injected the virus into the pons at P30, and examined the morphology 6 WPI. Unsurprisingly, we did not see a significant difference in cell body size between genotypes (Figure 2.10), similar to the results of early *Fezf2* deletion (Figure 2.7). Altogether, these results indicate that the deletion of *Fezf2* in adult corticospinal motor neurons is not necessary to maintain neuronal class identity.

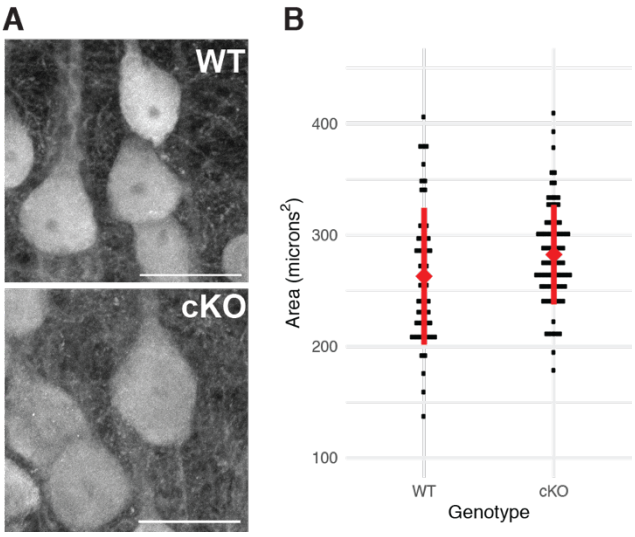


FIGURE 2.10. LATE *FEZF2* DELETION DOES REDUCE CELL BODY SIZE.

A. Representative z-stack images of tdTomato positive neurons. Scale bars are 25 micron. ***B.*** Quantification of cell body size. Each point represents one cell. Red diamond indicates mean, red bars indicate one standard deviation.

Chapter III.

Discussion and Future Directions

This chapter is adapted from an as-yet-unpublished manuscript I wrote with guidance from Paola Arlotta.

Summary of Findings and Significance

The mammalian cerebral cortex contains a multitude of neuronal subtypes, each with their own developmental trajectory, and a functional brain results from the interactions between these diverse neuronal populations and other neighboring cell types. Despite our knowledge behind the mechanisms responsible for the generation of all of the subtypes, we still do not know how mammalian neurons, born embryonically, maintain their identity throughout the entire lifetime of the organism. Here, we wanted to understand the underlying principles behind maintaining neuronal identity in the mammalian cerebral cortex. We found that *Fezf2*, a selector gene for CSMNs, is necessary to maintain CSMN molecular identity during early development, but not later in life.

Although the concept of terminal selector genes was first coined in *C. elegans*, this mechanism of neuronal identity maintenance is conserved in some neuronal subtypes in the mammalian brain, from the habenula (Serrano-Saiz et al., 2018) to the olfactory bulb (Zhang et al., 2016). However, to this date, we have yet to find a true terminal selector gene in the cerebral cortex. We only have evidence that early post-mitotic neurons are responsive to removal of selector genes (Bedogni et al., 2010; Fazel Darbandi et al., 2018; Zembrzycki et al., 2015), in alignment with the observation of neurons being reprogrammable and plastic early in life (De la Rossa et al., 2013; Rouaux and Arlotta, 2013; Ye et al., 2015).

An attempt was made previously to understand the role of *Fezf2* in the adult motor cortex, using lentiviral-mediated knockdown of *Fezf2* (Clare et al., 2017). However, there was not enough resolution to detect molecular changes in CSMNs. First, there was a confounding effect of *Fezf2* knockdown in other cell types, such as in CThPN and deep layer CPN, which also express *Fezf2*.

Second, bulk RNA-sequencing includes non-cell autonomous transcriptional changes from other cell types. Third, bulk RNA-sequencing may not necessarily have the resolution to assess the degree of molecular changes in a subpopulation as small as CSMNs. The strategies employed here overcome the limitations of the previous study, as we selectively labeled and collected layer Vb neurons for single-cell and nuclei sequencing.

Neuronal identity encompasses many features, including molecular, morphological, and electrophysiological traits. We predominantly focused on the molecular identity of CSMNs, to understand which genes, if any, change upon *Fezf2* deletion. With early postnatal *Fezf2* deletion, many SCPN-enriched genes were downregulated in Fezf2 conditional knockout (cKO) neurons as compared to wild type (WT) neurons, suggesting that Fezf2 is necessary in early postnatal animals to fully maintain CSMN identity. However, delineating one subtype from another with non-discrete phenotypes, such gene expression, is complicated by the number of genes and degree of expression. Now, with the abundance of single-cell datasets, we are beginning to see that single genes alone are no longer sufficient to distinguish one subtype from another, especially across developmental time. Therefore, in addition to looking at known subtype-specific gene sets, we took a comprehensive approach to integrate our data with published single-cell datasets, to see if and how much cKO neurons transcriptionally diverge from endogenous SCPNs. We found that a subset of cKO neurons globally diverge away from SCPN molecular identity when compared to WT control neurons. Thus, the net reduction in the expression of a battery of SCPN-specific genes in cKO neurons, as well as the global transcriptional divergence of a subset of cKO neurons, suggest that *Fezf2* is necessary to maintain the molecular identity of CSMNs early in life. This phenomenon, however, is not recapitulated with late *Fezf2* deletion. Therefore, while *Fezf2* is necessary to maintain CSMN molecular identity early in life, we conclude it does not meet the full criteria to be a terminal selector gene.

In our study, we also evaluated one morphological aspect of identity, cell body size, and found no significant differences between genotypes, regardless of when *Fezf2* was removed. It is

possible that Fezf2 is necessary for other CSMN features across time, including electrophysiological features or other morphological features, such as dendritic morphology and axonal targeting. Whether these aspects of identity change with late *Fezf2* deletion are potential topics of future study.

Nuclear plasticity of neurons decreases over time

While *Fezf2* does not fit the definition of a terminal selector gene, our data are further evidence for neurons reducing their nuclear plasticity over time. This was first established with reprogramming studies, where the capacity to change molecular identity severely declines with age (Rouaux and Arlotta, 2013). This phenomenon is also seen outside the cortex, where Nurr1 ablation in ventral midbrain dopamine (DA) neurons is more effective at reducing genes in the dopamine synthesis pathway in P0 animals than in five week old animals (Kadkhodaei et al., 2013). Similarly, Pet-1 deletion in serotonergic neurons also exhibits a time-sensitive effect. Pet-1 constitutive knockout neurons remain functionally immature and do not acquire their serotonergic neurotransmitter identity. However, Pet-1 became dispensable for serotonin synthesis genes as neurons matured, and instead became only necessary for genes modulating serotonergic neuron excitability (Wyler et al., 2016). Our data is in agreement with these aforementioned experiments, as we only see substantial molecular changes in CSMNs with early *Fezf2* deletion. Furthermore, only a subset of neurons is responsive to early *Fezf2* deletion, and neurons do not fully convert to a CPN identity as they do in constitutive Fezf2 KO animals. We suspect that this reprogramming heterogeneity is due to the timing of *Fezf2* deletion, with the neurons that recombine later less likely to show a molecular phenotype. The window of nuclear plasticity and capacity to switch identities in CSMNs can be tested further by systematically deleting *Fezf2* across early development with viral strategies similar to those described here.

Distinct forms of plasticity are differentially regulated across neuronal subtypes

Our data and that of others demonstrate a decline in the capacity of cortical neurons to change subtype-specific identity within the first postnatal week, as well as the existence of separate plasticity windows for different features within subtypes. For example, within CPNs, early post-mitotic overexpression of *Fezf2* changes dendritic morphology and axonal projections, but does not necessarily change the CPN-specific molecular signature of CPNs (Rouaux and Arlotta, 2013; Ye et al., 2015), suggesting the morphological features of CPNs retain their plasticity longer than the molecular features associated with subtype-specific identity. In our data, we found that subtype-specific nuclear plasticity is largely closed within the first postnatal week, as cKO neurons reduced their expression of CSMN-specific genes only with early, not late, *Fezf2* deletion. Interestingly, while cKO neurons have varying degrees of reduction of CSMN molecular identity, there is a population decrease in the expression of nuclear-encoded mitochondrial genes upon early *Fezf2* deletion. This discrepancy in differential gene expression of the two gene sets suggests a delayed window of plasticity in metabolic features, where metabolic features are malleable later in life than subtype-specific molecular features. Intriguingly, adult Nurr1 deletion in DA neurons can still reduce the expression of nuclear encoded mitochondrial genes (Kadkhodaei et al., 2013), suggesting that distinct features, such as metabolism, can have different windows of plasticity within, as well as between, neuronal subtypes.

Alternative Mechanisms to Maintain Neuronal Identity

To date, Fezf2 is the only transcription factor in the cortex that is both necessary and sufficient to impose subtype-specific identity, yet it is still not necessary later in life. Going forward, we believe there are two large categories of investigation to explore in understanding identity maintenance: non-cell autonomous versus autonomous.

Throughout the history of cortical development, we have learned that while genetics contribute to brain development, proper environmental input is necessary as well. A classic example is the development of the visual cortex, where balanced visual input is necessary for the structural formation of ocular dominance columns (Hensch, 2005). External inputs can induce molecular changes as well; somatosensory layer IV cortical neurons can change their molecular identity in response to which thalamocortical afferent input it receives (Pouchelon et al., 2014). Furthermore, cortical neurons are surrounded by other non-neuronal cell types, which contribute many extracellular cues. All of these environmental inputs could work in concert with the genetic programs to help maintain neuronal identity. However, the role of non-cell autonomous inputs is hard to dissect in the postnatal cortex, given the complexity in terms of the cell types and diversity of molecules that could contribute to the extracellular space. We nevertheless speculate that may play some role in maintaining subtype-specific identity in the adult brain (see Appendix 3).

Within the potential autonomous mechanisms, Fezf2 may require co-regulators to maintain CSMN identity, where CSMN terminal features are regulated by a combination of transcription factors, rather than by Fezf2 alone. This is seen in *C. elegans*, where *unc-86* can work with three separate LIM homeobox genes to generate distinct classes of neurons (Duggan et al., 1998; Serrano-Saiz et al., 2016; Zhang et al., 2014). Given the diversity of neuronal subtypes in the mammalian cortex, combinatorial co-regulation with several transcription factors seems more feasible than single terminal selector genes per subtype. It is possible that enhancer regions which regulate *Fezf2* attract other cofactors responsible for maintenance, or that Fezf2 directly activates its own maintenance cofactors during fate specification. *In vivo* CRIPSR screens can be used to investigate this further, by identifying candidate cofactors involved in the generation of CSMNs. These candidates can then further tested for necessity in maintaining CSMN identity. Alternatively, proteomic approaches can be used to pull down Fezf2 co-regulators across development. These co-regulators may not necessarily be subtype-specific, and could also be a shared maintenance mechanism with other cortical neuronal subtypes.

A related method could be through the establishment of stage specific enhancers for downstream effector genes. In nascent ESC-derived motor neurons, Lhx3 and Isl1 specify motor neuron identity by establishing motor neuron enhancers to drive effector gene expression. However, during maturation, the original enhancers are deactivated, and Isl1 moves to a different enhancer region bound by a separate transcription factor to maintain the same effector genes (Rhee et al,. 2016). Enhancer switching occurs *in vivo* as well. Coincidentally, this occurs with endogenous *Fezf2* regulation during development. The E4 downstream enhancer drives *Fezf2* expression in progenitors (Shim et al., 2012), while the upstream enhancer 1316 is utilized in post-mitotic deep layer neurons (Eckler et al., 2014). It is therefore possible that after CSMN-specification, a separate set of enhancers and other co-factors are recruited to maintain expression of downstream effector genes, rendering Fezf2 redundant later in life.

Finally, we see a potential role for epigenetic mechanisms to lock in neuronal identity, where during postnatal development, epigenetic modifications promote the expression of effector genes and actively close loci needed for other lineages. In *C. elegans*, genetic screens revealed several factors, including DOT1, a histone methyltransferase, which restrict cellular plasticity (Rahe and Hobert, 2019). Interestingly, some candidates found in this screen are ubiquitously expressed across cell types and development, suggesting there can be converging methods across cell types to restrict molecular identity. Building on this, we believe comparative studies between cortical neuron subtypes using RNA-seq and ATAC-seq could reveal shared and divergent mechanisms for maintaining identity within the mammalian cortex.

While a shared epigenetic mechanism among neurons is the most parsimonious method to maintain neuronal identity, selector gene recruitment of epigenetic modifiers can still occur in the cortex. For example, Lhx2, a marker of upper layer cortical neurons, interacts with nucleosome remodeling and histone deacetylase histone remodeling complex subunits (Muralidharan et al., 2017). Without Lhx2, the transcriptional start sites of Fezf2 and its regulator, Sox11, have an increase of active histone marks H3K4me3 and H3K9Ac, suggesting that Lhx2

can recruit epigenetic remodelers to repress SCPN identity in CPNs. To investigate a similar possibility in CSMNs, proteomic studies assessing for epigenetic modifiers that associate with endogenous Fezf2 during and after CSMN fate specification could reveal a Fezf2-dependent mechanism for epigenetic maintenance of CSMN identity.

The diversity in neuronal subtypes in the mammalian cortex is unparalleled, and understanding how these neurons are born and develop is crucial to understanding cortical function. Fezf2 has demonstrated that solely intrinsic, transcription factor-based mechanisms can be utilized to confer class-specific traits in mammalian cortical neurons. We tested whether Fezf2 was also necessary to maintain CSMN identity, and we found that Fezf2 was only necessary to maintain molecular identity in young postnatal animals. The capacity of Fezf2 to maintain identity decreases with age, unveiling a distinct window of nuclear plasticity. Although Fezf2 is not necessary to maintain CSMN identity in late postnatal animals, it has been a powerful tool in dissecting basic principles behind plasticity and identity maintenance behind building and maintaining the cerebral cortex.

Appendix 1.

Methods for Chapter II.

This appendix includes the material and methods related to the results in Chapter II and is adapted from an as-yet-unpublished manuscript I wrote with guidance from Paola Arlotta.

Author contributions are the same as in Chapter II.

Mice

All procedures were approved by the Institutional Animal Care and Use Committee of Harvard University and were performed in accordance with institutional and federal guidelines.

Ai14. Mice were generated by Madisen et al., 2010, and purchased from Jackson Laboratory (JAX stock #007908).

Emx1-Cre. Mice were generated by Gorski et al., 2002, and acquired from Jackson Laboratory (JAX stock #005628).

Fezf2 KO. Mice were generated by Hirata et al., 2005.

Fezf2 cKO. Mice were generated with the Genome Modification Facility at Harvard University. Briefly, embryonic stem cell clones with the Fezf2 conditional allele were purchased from EUCOMM (HEPD0722_1_G06) and injected into albino C57BL/6 embryos to create chimeric mice. Positive mice were bred further with C57BL/6 mice to generate the colony. LacZ Neo cassette was removed by breeding to Actin-FlpE mice (Jackson labs, JAX stock #005703).

Sun1-GFP. Mice were generated by Mo et al., 2015, and acquired from Jackson Laboratory (JAX stock #021039).

Tcerg1l-CreERT2. Mice were newly generated in collaboration with Z. Josh Huang's laboratory; it is a knock-in line, with a 2A peptide linker sequence between *Tcerg1l* and *CreERT2*. It will be described fully elsewhere. Animals were bred to be homozygous for CreERT2 for experiments. Experimental P3 pups received a single dose of 0.5 mg of 4-hydroxytamoxifen (4-OHT) intraperitoneally.

Day of birth was designated postnatal day 0 (P0). All mice were maintained in standard housing conditions on a 12-h light/dark cycle with food and water *ad libitum*.

Electrophysiology

Slice preparation. Coronal brain slices (350 micron thick) were prepared from P21-P24 mice, using a vibrating blade microtome (Leica VT1200S). All slicing was conducted in ice cold cutting medium composed of: D-glucose (Sigma, 49158, 10 mM), $NaHCO_3$ (Sigma, S5761, 26.2 mM), KCl (Sigma, 60130, 2.5 mM), NaH_2PO_4 (Sigma, S8282, 1.3 mM), L-Ascorbic acid (Sigma A5960, 3 mM), Na-pyruvate (Sigma, P5280, 1.85 mM), Glycerol (Sigma, G7757, 207 mM), $MgCl_2$ (Sigma, M1028, 7 mM), and $CaCl_2$ (Sigma, 21115, 0.5 mM), saturated with 95% O_2 / 5% CO_2. After slicing, the brain slice was transferred to a standard artificial cerebrospinal fluid (aCSF) solution,and maintained at 35°C for 30 minutes. Following the period of incubation and prior to recording, slices were transferred to rest at room temperature for a minimum of 30 minutes. The standard aCSF solution was composed of: NaCl (Sigma, S7653, 125 mM), D-glucose (Sigma, 49158, 10 mM), $NaHCO_3$ (Sigma, S5761, 26.20 mM), KCl (Sigma, 60130, 2.5 mM), NaH_2PO_4 (Sigma, S8282, 1.3 mM), L-Ascorbic acid (Sigma A5960, 1 mM), Na-pyruvate (Sigma, P5280, 1 mM), $MgCl_2$ (Sigma, M1028,1 mM), and $CaCl_2$ (Sigma, 21115, 2 mM), saturated with 95% O_2 / 5% CO_2. To avoid osmotic shock and maximize slice health, the osmolality of the cutting medium and recording aCSF was measured and matched on a daily basis (Advance Instruments, Model 3320 Osmometer).

Data acquisition and digital signal processing. Signal amplification and current/voltage clamp were controlled via the Axon Multiclamp 700B (Molecular Devices). The 700Bs inbuilt four-pole Bessel filter served as a low pass anti-aliasing filter, and was set to no less than 4 times the digitisation frequency. Analog-to-digital/digital-to-analog conversion was via National Instruments

DAQ boards (PCIe-6321 and PCIe-6323). Recordings were performed using open source Ephus software (Suter et al., 2010).

Electrophysiological recordings. During recordings, slices were placed in a submerged chamber, held at 32 °C (temperature controller and in-line solution heater; Warner instruments, TC-344C) and perfused with aCSF at a rate of 3.6 mL/min by a peristaltic pump (Ismatec Reglo). Visually guided patch clamp recordings of layer Vb pyramidal cells was achieved using infrared illumination and differential interference contrast optics (Scientifica SliceScope Pro 6000; 40x submersion objective, Olympus LUMPlanFL-N; CMOS camera, Q Imagining Rolera Bolt; and Scientifica PatchStar micromanipulators). Cells were identified by their anatomical location. Whole cell recordings were made with patch electrodes (3-6 MΩ) pulled from borosilicate glass (1.5mm ODx 0.86 mm ID x 100 mm L; Harvard Apparatus, part # GC150F-10) using 2 stage vertical puller (Narishige, PC-10), filled with: potassium D-gluconate (Sigma, G4500, 120 mM), ATP magnesium salt (Sigma, A9187, 4 mM), GTP sodium salt hydrate (Sigma, G8877, 0.3 mM), HEPES (Sigma, H7523, 10 mM), EGTA (Sigma, E0396, 0.5 mM), sodium chloride (Sigma, S7653, 6 mM), creatine phosphate, dipotassium salt (EMD, 237911, 7 mM). Solution was adjusted to pH 7.2 with potassium hydroxide and 290 mOsm/kg with sucrose. Holding potential was at -70 mV. Recordings were made in the presence of NBQX (Abcam, ab120046, 10 μM), AP5 (Abcam, 120003, 50 μM), picrotoxin (Abcam, ab120315, 50 μM), and CGP-55845 (Abcam, ab120337, 2 μM, Abcam).

Data analysis. All post-processing and data analysis were performed using MATLAB R (MathWorks) using custom built algorithms developed in-house. The parameters were measured as follows: action potential (AP) threshold, measured at the time of the first peak in d^3V/dt^3 (the 3rd derivative of the voltage response; max dV/dt, maximum rate of change during the AP rise; AP amplitude, measured as the difference between the AP peak and the AP threshold; AP width, the width of the AP (in seconds) at half the maximum amplitude; Vrest, membrane resting potential

at break-in; resistance, input resistance; rheobase, the minimal current needed of infinite duration to generate an AP; voltage sag, calculated as the proportional difference between maximum and steady-state voltage transients resulting from hyperpolarizing current injections; adaptation ratio: measured as the ratio of the second inter-spike interval and last inter-spike interval.

Virus Production

AAV pmSyn1-EBFP-Cre (Madisen et al., 2015) was packaged with an rAAV-retro helper (Tervo et al., 2016) and purchased from AddGene and the UPenn Vector core with a titer greater than 1e13 GC/mL. Multiple thaw and refreeze cycles were avoided.

Early Postnatal (P1) Pons Injection

Pups were anesthetized by hypothermia at P1. Six injections (69 nL per injection) of virus were delivered to the pons under ultrasound guidance (Vevo 660, VisualSonics) as previously described (Arlotta et al., 2005).

Adult (P30) Pons Injection

Mice were anesthetized with isoflurane (~2% by volume in oxygen). A small hole was drilled at: A/P: 0.4, M/L 0.4 mm. 100 nL injections were made slowly at D/V: -6.0, -5.75, and -5.5 mm. Injections were done with a Hamilton syringe (701 RN) mounted on a stereotaxic frame with a microinjection unit (Kopf, Model 5000). A/P and M/L coordinates are given with respect to lambda, and D/V coordinates are given with respect to the pia.

FACS Purification

Fixed cells, bulk sequencing. Fezf2 HZ or Fezf2 KO pups were deeply anesthetized, and their sensorimotor cortices were dissociated into single cell suspension as previously described (Arlotta et al., 2005). Cells were stained with anti-CTIP2, anti-SATB2, anti-TLE4, and anti-Beta Galactosidase (1:2000 dilution) as previously described (Molyneaux et al., 2015), Gates were set based on relative levels of CTIP2, SATB2, TLE4, and Beta-Galactosidase.

Live cells, bulk sequencing. Tcerg1l-CreERT2; Fezf2 WT; Ai14 (n=4) and Tcerg1l-CreERT2; Fezf2 cKO; Ai14 (n=6) animals were deeply anesthetized on separate days, and their sensorimotor cortices were enzymatically digested into a single-cell suspension using a modified protocol from the Worthington Papain Dissociation System kit (Worthington Biochemical) (Velasco et al., 2019). Digested tissues were filtered with a 40 micron strainer before the addition of DAPI (Thermo Fisher, D1306, 1:500 dilution). tdTomato positive and DAPI negative cells were sorted on a Beckman Coulter MoFlo XDP Cell sorter with a 100 micron nozzle and sorted directly into Trizol LS (Thermo Fisher Scientific) and frozen for future RNA extraction and library preparation.

Live single cells. Injected animals were anesthetized, and a fluorescence dissecting microscope was used to microdissect layer V from sensorimotor sections. Microdissected pieces were enzymatically digested into a single-cell suspension using a modified protocol from the Worthington Papain Dissociation System kit (Worthington Biochemical) (Velasco et al., 2019). Fezf2 cKO; Ai14 and Fezf2 WT; Ai14 animals were processed on separate days, and 4 animals per genotype were used. Digested tissues were filtered with a 40 micron strainer before Vybrant DyeCycle Green (Thermo Fisher, V35004, 1:500 dilution) and SYTOX Red (Thermo Fisher, S34859, 1:500 dilution) were added to stain for live and dead cells, respectively. tdTomato positive, Vybrant DyeCycle Green positive, and SYTOX Red negative cells were sorted on a Beckman Coulter MoFlo XDP Cell sorter with a 100 micron nozzle. Single cells were sorted

directly into 5 uL of lysis buffer, composed of Buffer TCL (Qiagen, 1031576) and 1% 2-Mercaptoethanol (Millipore Sigma, M6250) for single-cell sequencing. Cells were spun down and frozen for future downstream RNA extraction and library preparation.

Frozen nuclei. Injected animals were anesthetized, and their sensorimotor cortices were microdissected and frozen on separate days. Two lanes were sequenced per genotype, and a WT sample was always processed in parallel with a cKO sample for sequencing. Subsequent nuclei isolation was adapted from a protocol previously described (Habib et al., 2017). Briefly, frozen brain tissue was homogenized a 2mL glass dounce tissue grinder (Sigma, #D8939) with EZ Lysis Buffer from the Nuclei EZ Prep kit (Sigma, #NUC-101), and resuspended in Nuclei Suspension Buffer (NSB): RNAse-free molecular-biology grade PBS, 0.1% BSA at 100 µg/ml (NEB, B9000S), and 0.2U/µl RNase inhibitor (Clontech, #2313A), with Vybrant DyeCycle Ruby (Thermo Fisher, V10309, 1:500 dilution). The suspension was filtered through a 35 micron cell strainer, and GFP positive nuclei were sorted on a Beckman Coulter MoFlo XDP Cell sorter with an 85 micron nozzle into NSB prior to 10x Genomics Chromium single-nucleus sequencing.

Bulk RNA-seq Analysis, Fixed Cells

RNA isolation, library preparation, and sequencing. RNA was recovered and subsequently prepared for sequencing as previously described (Molyneaux et al., 2015).

Differential analysis. Reads were clipped and mapped onto the mouse genome (mm10, UCSC) using STAR (Dobin et al., 2013). The resulting matrix of read counts from the STAR alignment were analyzed for differential expression by DESeq2 (Love et al., 2014). Cutoff for non-expressed genes were defined as the lowest five percent of genes expressed in transcripts per million using Kallisto (Bray et al., 2016), and genes that did not meet this cutoff in all samples were discarded.

Differentially expressed genes were determined using the Wald test, with Benjamini-Hochberg false discovery rate (FDR) less than 0.1, as implemented in the DESeq2 R package.

Gene Set Enrichment Analysis. GSEA analysis was performed on a ranked gene list of expressed genes from our Fezf2 HZ versus Fezf2 KO RNA-Seq data using GSEA v3.0 (Subramanian et al., 2005). Genes were ordered by log2 fold change. Subtype-specific projection neuron gene set databases were identified previously (Molyneaux et al., 2015) and were tested for enrichment or depletion in Fezf2 KO neurons using the preranked GSEA method. FDR<0.01 was used to determine significance, as implemented by the GSEA package.

Bulk RNA-seq Analysis, Live Cells

RNA isolation, library preparation, and sequencing. Cells were collected in Trizol LS (Thermo Fisher Scientific), and RNA was recovered using their published protocol. 1 ng of RNA to generate cDNA libraries with the SMART-Seq v4 Ultra Low Input RNA kit (Takara, 634888), which served as input for the standard Nextera XT DNA library preparation kit (Illumina, FC-131-1024). Libraries from different samples were pooled based on molar concentrations and sequenced on a HiSeq 2500 instrument (Illumina) with 101 base paired-end reads.

Differential analysis and Gene Set Enrichment Analysis. Analysis was done as previously described with fixed cells. Cutoff for differential expression was set at FDR<0.05.

Single Cell RNA-seq Analysis

Library preparation and sequencing. RNA-seq libraries were prepared as described previously (Ding et al., 2019). Libraries were sequenced on a NextSeq 500 instrument (Illumina), with 50 base pair reads for read 1, 25 base pair reads for read 2, and 8 base pair reads for the indices.

Single-cell analysis. RSEM (Li and Dewey, 2011) was used to align reads from RNA-seq to an adapted mm10 mouse reference genome (UCSC) with pmSyn1-EBFP-Cre and tdTomato to produce the associated cell-by-gene count matrix. Unique molecular identifier (UMI) counts were analysed using the Seurat R package v.3 (Stuart et al., 2019). Cells expressing over 2000 genes and less than 10% mitochondrial genes were kept. Seurat's default method for identifying variable genes was used with nFeatures = 3000, and the ScaleData function was used to regress out variation due to differences in UMIs, number of genes per cell, and percent mitochondrial RNA. Principal component analysis (PCA) was performed on the scaled data for the variable genes, and top principal components were chosen. Cells were clustered in PCA space. Differentially expressed genes between genotypes were identified using Seurat's implementation of the model-based analysis of single cell transcriptomics (MAST) algorithm (Finak et al., 2015). Differentially expressed genes were calculated using FindAllMarkers in two ways. "All" differentially expressed genes had a Bonferroni-adjusted P value less than 0.05 and a log fold change of at least 0.25, but those that were considered "stringent" differentially expressed genes used min.diff.pct = 0.1, indicating a minimum 10% difference in the number of cells expressing the gene between genotypes. t-SNE plots were used to visualize variation in the data.

EnrichR analysis. Gene Ontology (GO) enrichment was computed through Enrichr through its R interface (Chen et al., 2013; Kuleshov et al., 2016). Differentially expressed genes enriched in either Fezf2 WT; Ai14 cells or Fezf2 cKO; Ai14 cells were compared to known databases, specifically "GO Biological Process 2018," "GO Cellular Component 2018," and "GO Molecular Function 2018." Gene sets were ranked by their Benjamini-Hochberg method adjusted p-value with the Fisher exact test. Further analysis on nuclear-encoded mitochondrial genes used gene lists from MitoCarta 2.0 (Calvo et al., 2015; Pagliarini et al., 2008).

Hypergeometric testing of enrichment of cortical subtype-specific genes. Differentially expressed genes enriched in either Fezf2 WT; Ai14 cells or Fezf2 cKO; Ai14 cells were compared to gene

sets compiled from several published and unpublished sources. Unpublished gene sets include: “below P7 general”, “P7 general”, “P21 general”, “Cux2 late”, and “Tle4 late.” Briefly, these datasets were generated using the Cux2-CreERT2; Ai14 and Tle4-CreERT2; Ai14 lines that broadly label upper layer CPNs and CthPNs, respectively. Pregnant dams received 4-OHT at E17.5 and labeled neurons were collected at various postnatal timepoints. “Below P7 general”, “P7 general”, “P21 general” refer to genes that are shared between the two neuronal populations during the given time points, whereas “Cux2 late” and “Tle4 late” refer to subtype-speciifc genes that retain specificity at P48 (Yuan et al., in preparation). “CThPN DeCoN,” “CPN DeCoN”, and “SCPN DeCoN” were gene sets from the DeCoN database (Molyneaux et al., 2015). The rest are derived from the Allen Brain Atlas and their subtype classification (Tasic et al., 2016). We took their dataset and used Seurat v.3 to identify differentially expressed genes with their subtype classification. Differentially expressed marker genes were found using FindAllMarkers in Seurat v.3, and had a Bonferroni-adjusted P value less than 0.05 and a log fold change of at least 0.25. Gene sets with 'top' preceding it means the genes are significantly enriched in that subtype with min.pct.diff = 25%. Gene sets with 'other' preceding it means the genes are significantly enriched in that subtypes with the default of min.pct.diff = -Inf. With the exception of gene sets ending with ‘SCPN nuclei,’ the data are derived from their adult, live cell data instead of from frozen adult nuclei. Of note, because of the differences in how all sequencing data were generated, the gene sets are not mutually exclusive. Finally, enrichment of known subtype-specific markers within differentially expressed genes was computed using the hypergeometric test in R, with the number of all expressed genes as the main population from which to sample. Significant gene sets have a Bonferroni-adjusted p-value less than 0.05.

Merged analysis with Allen Brain Dataset. Our single cell data was merged with the adult, live neuron dataset from the Allen Brain Atlas (Tasic et al., 2016) using the standard data integration method in Seurat v.3. Briefly, we performed standard preprocessing, found variable features in

both datasets, and identified anchors with a dimensionality of 30. Anchors were passed to the 'IntegrateData' function, which created our merged Seurat object from both datasets. t-SNE plots were used to visualize variation in the data.

Cell type classification using the Allen Brain dataset. We followed the standard cell type classification pipeline in Seurat v.3. Briefly, we found transfer anchors with the adult Allen Brain dataset as our reference (Tasic et al., 2016), with our single cell data as the query, and made predictions using the 'TransferData' function.

Single Nuclei RNA-seq Analysis

Library preparation and sequencing. Sorted, GFP positive nuclei were resuspended into RNAse-free molecular-biology grade PBS and 0.04% BSA (NEB, B9000S) and were loaded onto a Chromium Single Cell v3 Chip (10x Genomics) and processed through the Chromium controller to generate single-cell gel beads in emulsion. Single nuclei libraries were prepared with the Chromium Single Cell 3′ Library & Gel Bead Kit v.3 (10x Genomics). Libraries from different samples were pooled based on molar concentrations and sequenced on a NextSeq 500 instrument (Illumina) with 35 bases for read 1, 100 bases for read 2 and 8 bases for index 1. Cells were sequenced until mean reads per cell exceed 20000 and median genes per cell exceeded 3000.

Single-nuclei analysis. The Cell Ranger 3 pipeline (10x Genomics) was used to align reads from RNA-seq to an adapted mm10 mouse reference genome with pmSyn1-EBFP-Cre and Sun1-GFP to produce the associated cell-by-gene count matrix. Default parameters were used, except for the '–cells' argument. Unique molecular identifier (UMI) counts were analysed using the Seurat R package v.3 (Stuart et al., 2019). Cells expressing between 1500 and 6000 genes and less than 2.5% mitochondrial genes were kept. Seurat's default method for identifying variable genes was

used with nFeatures = 3000, and the ScaleData function was used to regress out variation due to differences in UMIs, number of genes per cell, sex, and sequencing batches. Principal component analysis (PCA) was performed on the scaled data for the variable genes, and top principal components were chosen. Cells were clustered in PCA space. Clusters with a CPN signature and equal representation from both genotypes were removed. Differentially expressed genes between genotypes were identified using Seurat's implementation of the model-based analysis of single cell transcriptomics (MAST) algorithm, with assay = "RNA". Differentially expressed genes had a Bonferroni-adjusted P value less than 0.05 and a log fold change of at least 0.25. t-SNE plots were used to visualize variation in the data.

EnrichR analysis. Analysis was done as previously described for single-cell RNA-seq.

Hypergeometric testing of enrichment of cortical subtype-specific genes. Analysis was done as previously described for single-cell RNA-seq.

Cell type classification using the Allen Brain dataset. We followed the standard cell type classification pipeline in Seurat v.3 (Stuart et al., 2019). Briefly, we found transfer anchors with the adult Allen Brain dataset from frozen primary mouse cortex nuclei as our reference (unpublished), with our single nuclei data as the query, and made predictions using the 'TransferData' function.

Immunocytochemistry

Brains were processed and stained using standard methods, as previously described (Arlotta et al., 2005). Primary antibodies and dilutions used were as follows: rabbit Beta-Galactosidase, 1:2000 (gift from J. Sanes); rabbit anti-cleaved Caspase-3, 1:300 (Cell Signaling), rat anti-CTIP2, 1:100 (Abcam); mouse anti-SATB2, 1:50 (Abcam); rabbit anti-RFP, 1:500 (Rockland). Goat anti-rabbit IgG Alexa Fluor 488, 546 and 647 (Life Technologies A11070, A-11071, A21246), goat

anti-rat 488, 546, 647 (Life Technologies A11006, A11081, A21247), goat anti–mouse IgG Alexa Fluor 488, 546, 647 (Life Technologies A11017, A11018, A21237). Secondary antibodies were diluted 1:750. Sections were imaged with either a Nikon 90i fluorescence microscope equipped with a Retiga Exi camera (Q-Imaging) and analyzed with Volocity image analysis software v4.0.1 (Improvision), or with a Zeiss AxioImager Z2 Upright Microscope and analyzed with Zen Blue image-processing software. Confocal images were taken with a Zeiss LSM 700 Inverted Confocal Microscope and processed using Zen Blue image-processing software.

In situ Hybridization (ISH)

ISH with alkaline phosphatase, anti-digoxigenin probes. ISH was done as described previously (Lodato et al., 2014). The Etv1 riboprobe was a gift from M. Studer and was generated as described in Arlotta, 2005.

RNAscope. Fixed, frozen brains were sectioned on a cryostat at 14 microns. Double *in situ* fluorescence hybridization was performed using the RNAscope Fluorescent Multiplex assay according to their protocol (Advanced Cell Diagnostics, 320850). The following probes were used: *Bcl11b* (413271-C3), *Crym* (466131-C3), *Etv1* (442901), *Fezf2* (313301, 313301-C2), *Ldb2* (466061-C2), *Slc17a7* (416631-C2). Chromogenic double in situ hybridization was performed for embryonic tissue using the RNAscope 2.5 HD Duplex Assay (Advanced Cell Diagnostics, 322500) according to their protocol. The following probes were used: *Fezf2* (313301), *Ldb2* (466061-C2).

Tissue sections were imaged using a Nikon 90i fluorescence microscope equipped with a Retiga Exi camera (Q-Imaging) and analyzed with Volocity image analysis software v4.0.1 (Improvision). Confocal images were taken with a Zeiss LSM 700 Inverted Confocal Microscope and processed using Zen Blue image-processing software.

Cell Body Area Quantification

Neurons in primary sensorimotor cortex from matching sections were imaged on a Zeiss LSM 700 Inverted Confocal Microscope and a 40X objective. For early Fezf2 deletion, 4 Fezf2 cKO; Ai14 animals were used, and 3 Fezf2 WT; Ai14 were used. For late *Fezf2* deletion, 2 Fezf2 cKO; Ai14 animals were used, and 1 Fezf2 WT; Ai14 animal was used. Z-stacks were acquired at the optimal step size (every 0.46 micron) to span the depth of the neurons. Image processing afterward was done using the Fiji implementation of the ImageJ software. Images were flattened using the "Z-stack" tool to create a 2-dimensional projection of each neuron. Neurons were identified by thresholding the images, followed by watershed segmentation. Size was calculated using the "Analyze Particles" tool. Neurons that were under- or over-segmented were removed from analysis by investigators blinded to the genotype. Significance was calculated using a nested ANOVA with a mixed effects model in R.

TUNEL Assay

TUNEL assay was completed with the *In Situ* Cell Death Detection Kit, Fluorescein (Roche, 11684795910). Protocol was adapted from the kit. Brains were processed as previously described for immunocytochemistry and permeabilized with ethanol and acetic acid in a 2:1 ratio, respectively, prior to continuing with the assay as described by the manufacturer. For early Fezf2 deletion, 3 Fezf2 cKO; Ai14 animals were used, and 2 Fezf2 WT; Ai14 were used. For late *Fezf2* deletion, 2 animals per genotype were used. The positive control was a 90 minute treatment with DNAseI. The DNAseI control was reconstituted from lyophilized powder (Millipore Sigma, DN25) to a working solution of 3000 U/mL in 50mM Tris-HCL, pH 7.5 (Teknova, T1075), 1 mg/mL BSA (Sigma, A4161), and 10mM MgCl2 (Thermo Fisher, AM9530G). Images were acquired

immediately after on the Zeiss AxioImager Z2 Upright Microscope and processed using Zen Blue image-processing software.

Data Availability

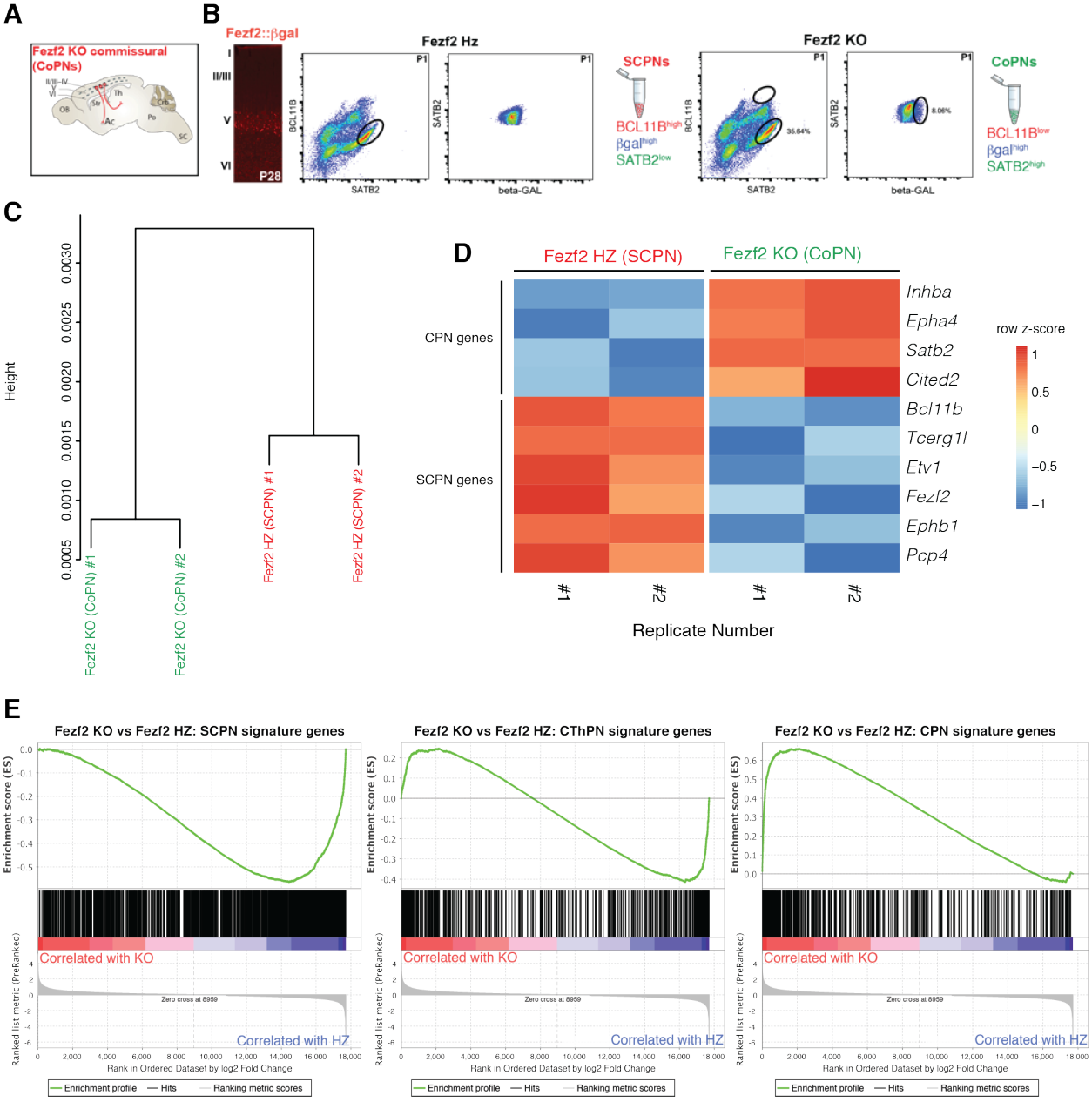
The RNA-seq data will be deposited in GEO.

Appendix 2.

Supplemental Materials Related to Chapter II

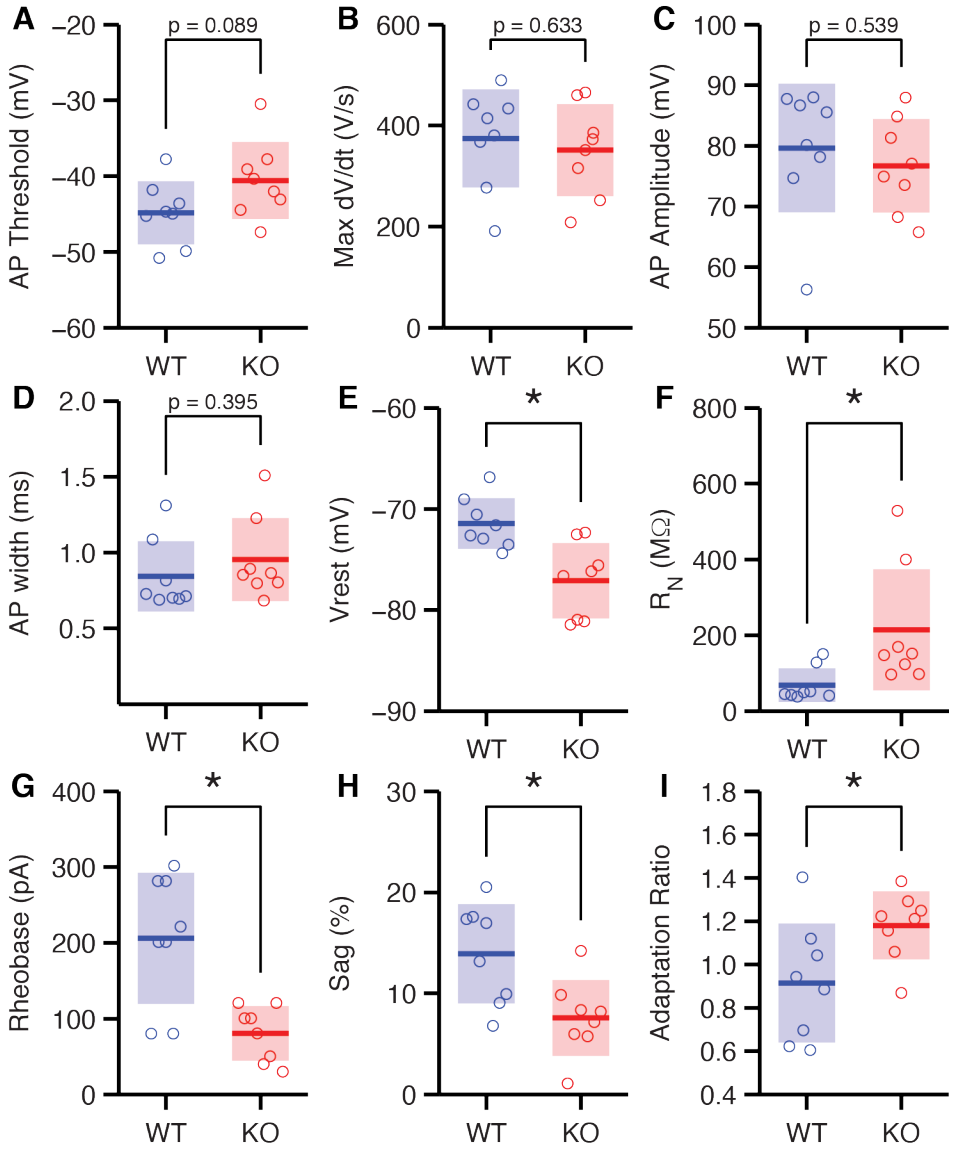
This Appendix includes the supplementary figures and tables related to the results in Chapter II.

Author contributions same as in Chapter II.



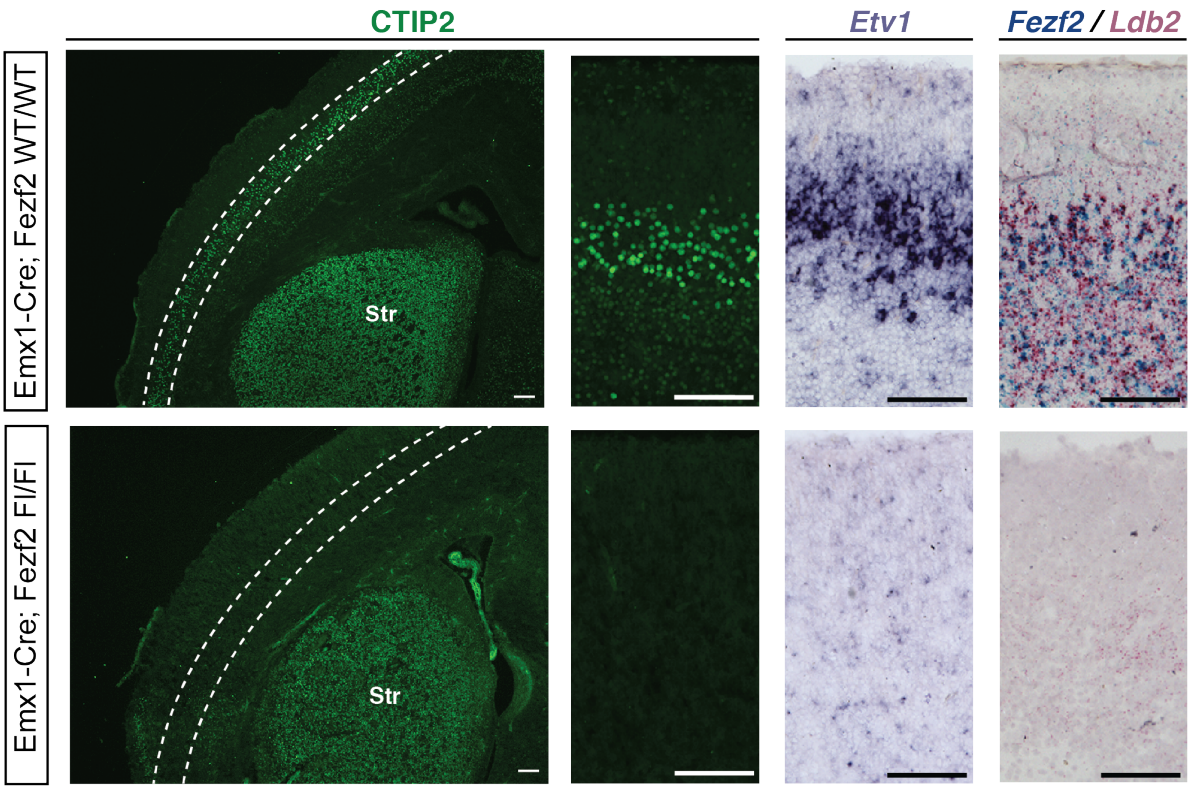
SUPPLEMENTARY FIGURE A2.1. FEZF2 KO NEURONS EXCHANGE A SCPN MOLECULAR IDENTITY FOR CPN IDENTITY

A. Schematic of Fezf2 KO commissural neurons, the layer Vb neurons in born in lieu of SCPNs in Fezf2 KO animals. ***B.*** Schematic of FACS gating strategy. Control SCPN have high expression of BCL11B, Beta-Galactosidase, and low expression of SATB2, while Fezf2 KO neurons have low expression of BCL11B and high expression of beta-galactosidase and SATB2. ***C.*** Dendrogram of rlog transformed read counts of biological replicates. Distance was calculated with Pearson correlation. ***D.*** Heatmap of selected CPN and SCPN genes and their relative expression in Fezf2 HZ and Fezf2 KO samples. ***E.*** GSEA for cortical subtype-specific gene sets. Fezf2 KO neurons have significant enrichment of CPN genes instead of SCPN genes, Kolmogorov-Smirnov test, FDR<0.01.



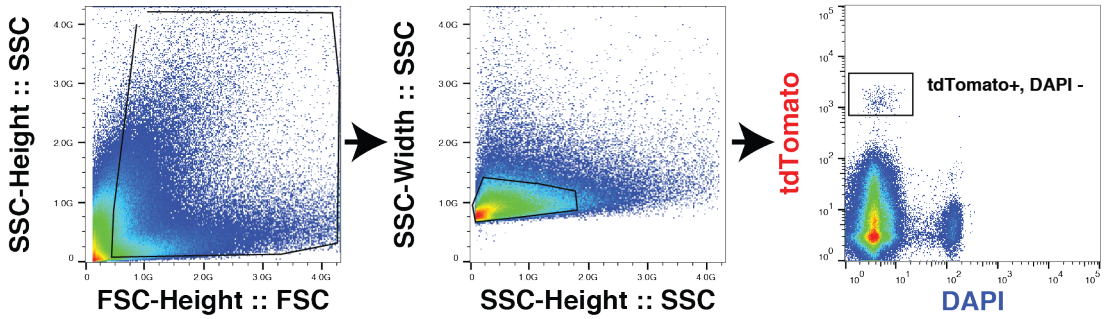
SUPPLEMENTARY FIGURE A2.2. FEZF2 KO NEURONS CHANGE INTRINSIC ELECTRICAL PROPERTIES

A. Action potential (AP) threshold between layer Vb control (WT) and layer Vb Fezf2 KO neurons (KO). ***B.*** Maximum change in voltage over time (dV/dt) during AP. ***C.*** Amplitude of AP from baseline. ***D.*** Half width of AP. ***E.*** Resting membrane potential. ***F.*** Input resistance. ***G.*** Rheobase, the minimal current needed of infinite duration to generate an AP. ***H.*** Voltage sag. ***I.*** Adaptation ratio. Colored boxes indicate mean +/- 2 standard deviations. Asterisks indicate p-values of less than 0.05, and significance was calculated with a two-way unpaired Student's t-test.



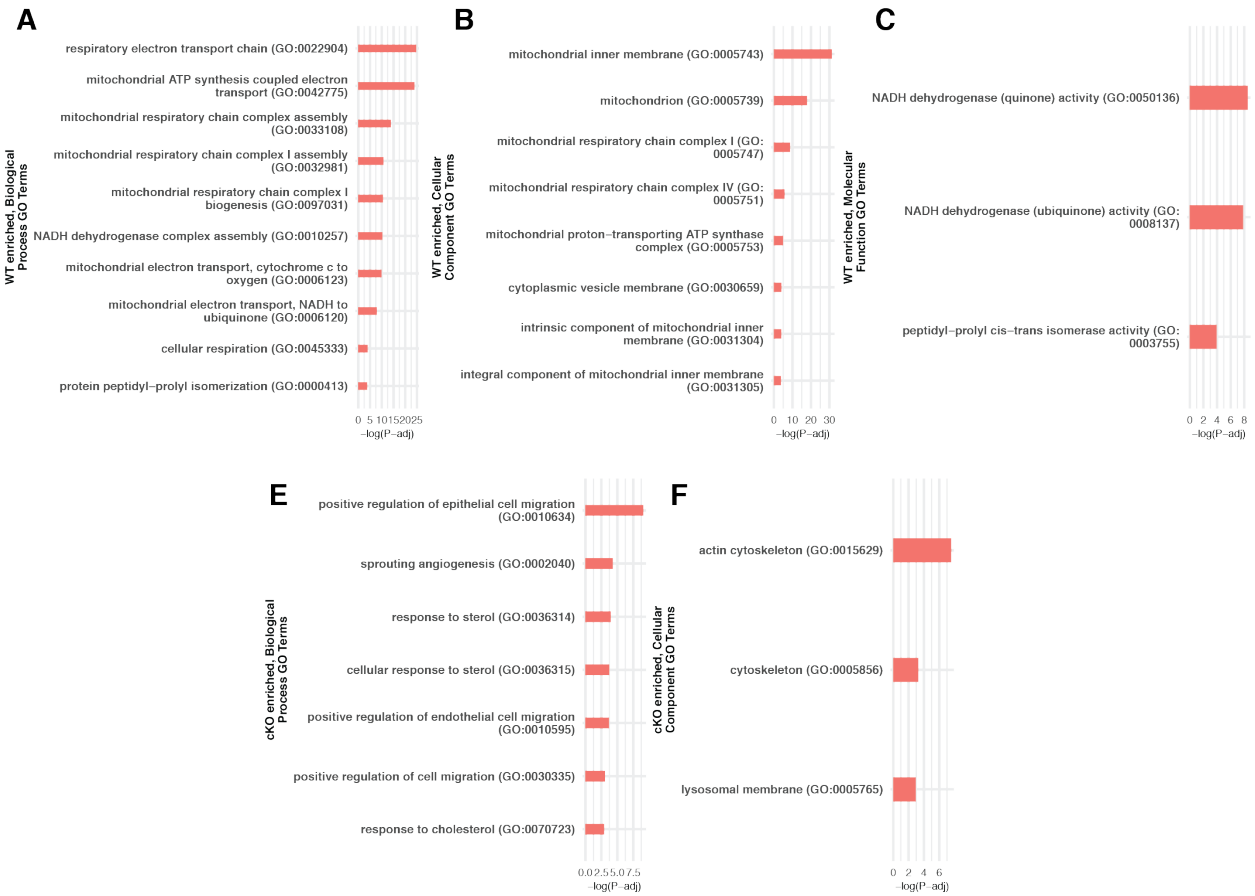
SUPPLEMENTARY FIGURE A2.3. *FEZF2* DELETION IN DORSAL NEURAL PROGENITORS PHENOCOPIES FEZF2 CONSTITUTIVE KNOCKOUT ANIMALS

Emx1-Cre; Fezf2 cKO animals lose CTIP2 (BCL11B), *Fezf2*, and *Ldb2*, markers of CSMNs, in the cortex, as compared to Emx1-Cre; Fezf2 WT animals. Scale bars are 100 microns.



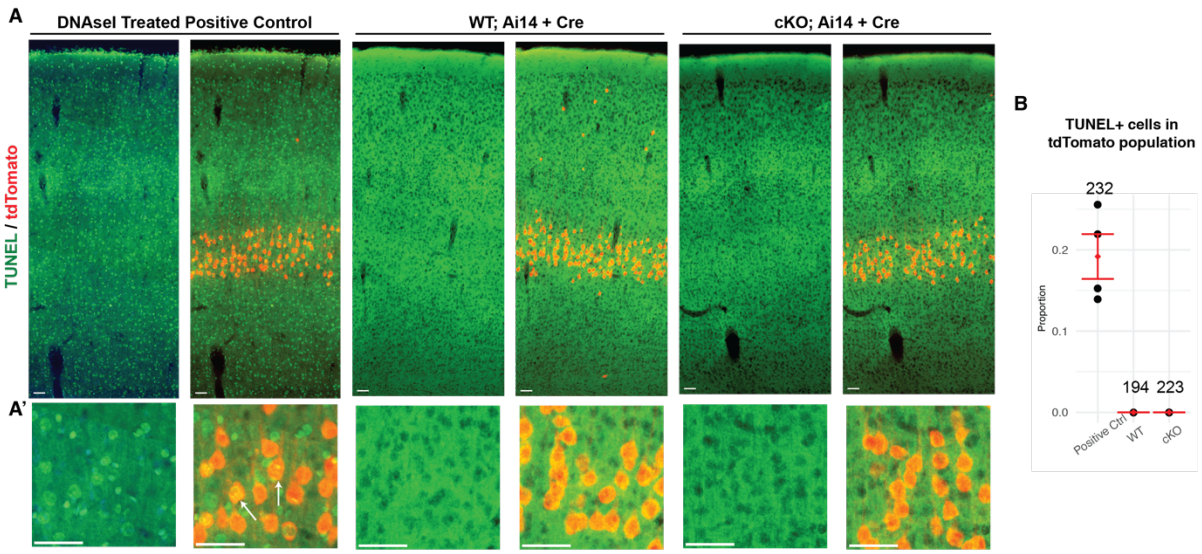
SUPPLEMENTARY FIGURE A2.4. FACS GATING STRATEGY OF TCERG-CREERT2; FEZF2 WT; AI14 AND TCERG-CREERT2; FEZF2 CKO; AI14 ANIMALS

Schematic of representative FACS gating strategy. tdTomato positive, DAPI negative cells are collected for bulk RNA-sequencing.



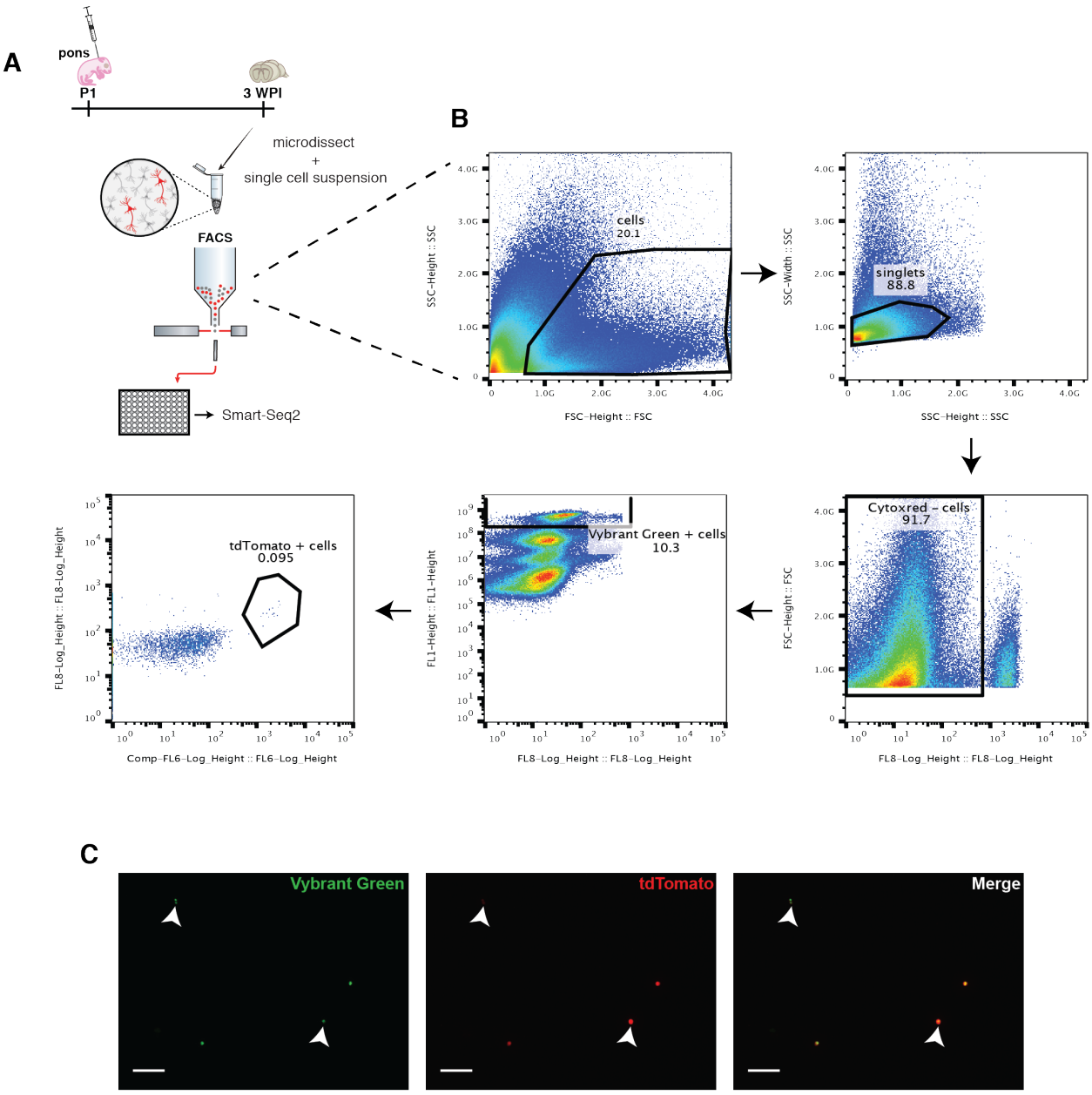
SUPPLEMENTARY FIGURE A2.5. TCERG-CREERT2; FEZF2 CKO; AI14 NEURONS LOSE EXPRESSION OF NUCLEAR-ENCODED MITOCHONDRIAL GENES AND GAIN MIGRATION-RELATED GENES AS COMPARED TO TCERG-CREERT2; FEZF2 WT; AI14 NEURONS

A. The top 10 most strongly associated gene ontology (GO) groups in the differentially expressed genes (DEGs) upregulated in Tcerg-CreERT2; Fezf2 WT; Ai14 (WT) neurons, as defined by biological process. ***B.*** The top 10 most strongly associated GO groups in the DEGs enriched in WT neurons, as defined by cellular component. ***C.*** The only significantly associated GO groups in the DEGs enriched in WT neurons, as defined by molecular function. ***D.*** The only significantly associated GO groups in the DEGs enriched in Tcerg-CreERT2; Fezf2 cKO; Ai14 (cKO) neurons, as defined by biological process. ***E.*** The only significantly associated GO groups in the DEGs enriched in cKO neurons, as defined by cellular component. There are no significantly enriched gene sets as defined by molecular function. All are ranked by log adjusted p-value of the Fisher's exact test, reported with Benjamini-Hochberg multiple hypothesis correction as implemented in the EnrichR R package.



SUPPLEMENTARY FIGURE A2.6. EARLY *FEZF2* DELETION DOES NOT CAUSE CELL DEATH

A. TUNEL stain of P1 injected animals and sacrificed at 3 WPI, with magnified insets (***A'***). Arrows point to tdTomato, TUNEL double positive cells. Scale bars are 50 micron. ***B.*** Quantification of tdTomato and TUNEL double positive cells. Numbers above bars indicate total tdTomato cells counted per condition. Red diamond indicates mean, red bars indicate standard error of the mean.



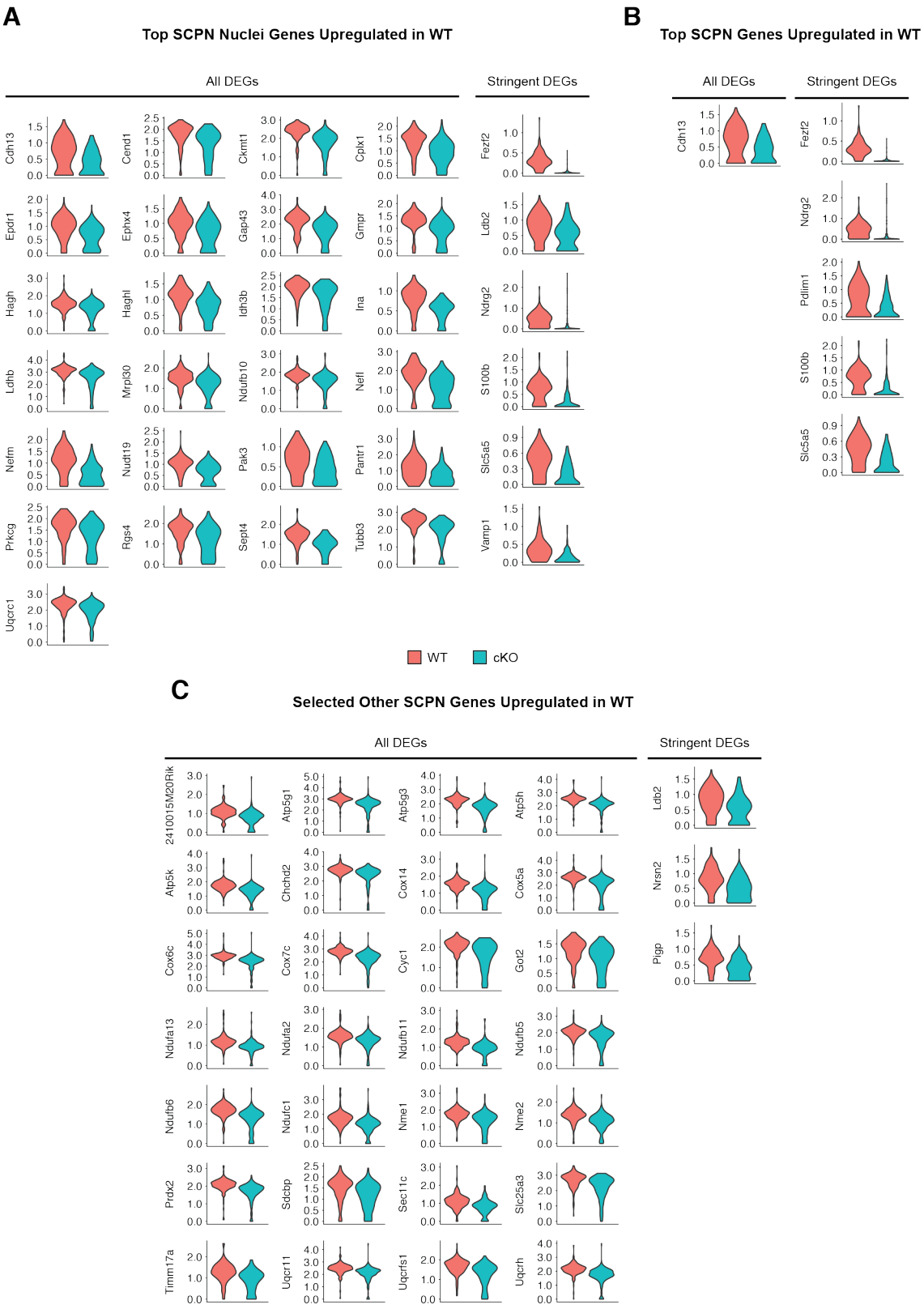
SUPPLEMENTARY FIGURE A2.7. FACS GATING STRATEGY OF FEZF2 WT; AI14 + CRE AND FEZF2 CKO; AI14 + CRE ANIMALS

A. Experimental design. ***B.*** Representative FACS gating strategy. tdTomato positive, Cytoxred negative, and Vybrant Green positive cells are collected for single-cell RNA-seq in 96 well plates. ***C.*** Sorted cells under the microscope. Arrowheads point to Vybrant Green and tdTomato positive cells. Scale bars are 100 micron.

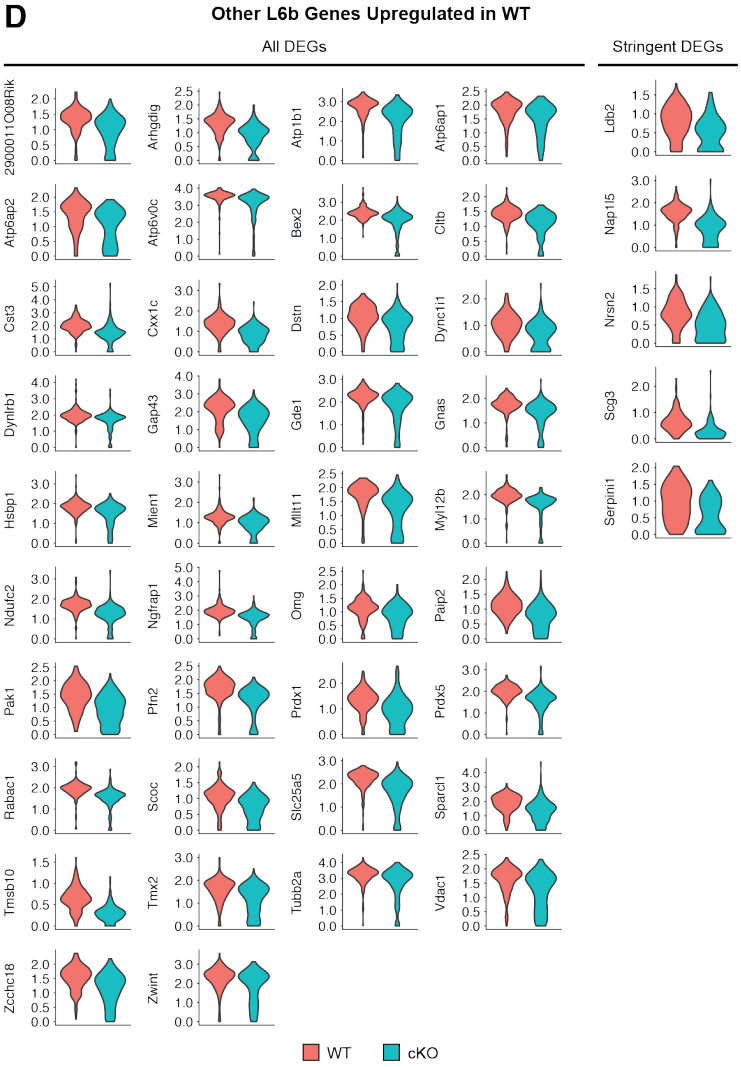
SUPPLEMENTARY FIGURE A2.8. SUBTYPE-SPECIFIC GENES THAT DECREASE UPON EARLY *FEZF2* DELETION

A. Violin plots of every DEG belonging to the 'top SCPN nuclei' gene set. ***B.*** Violin plots of every DEG belonging to the 'top SCPN' gene set. ***C.*** Violin plots of selected DEGs belonging to the 'other SCPN' gene set. ***D.*** Violin plots of all DEGs belonging to the other layer 6b gene set. ***E.*** Violin plots of every DEGs belonging to the 'layer 6 CThPN' gene set. ***F.*** Violin plots of every DEG belonging to the 'other layer V non-projecting (NP) neuron' gene set. All differentially expressed genes (DEGs) refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype.

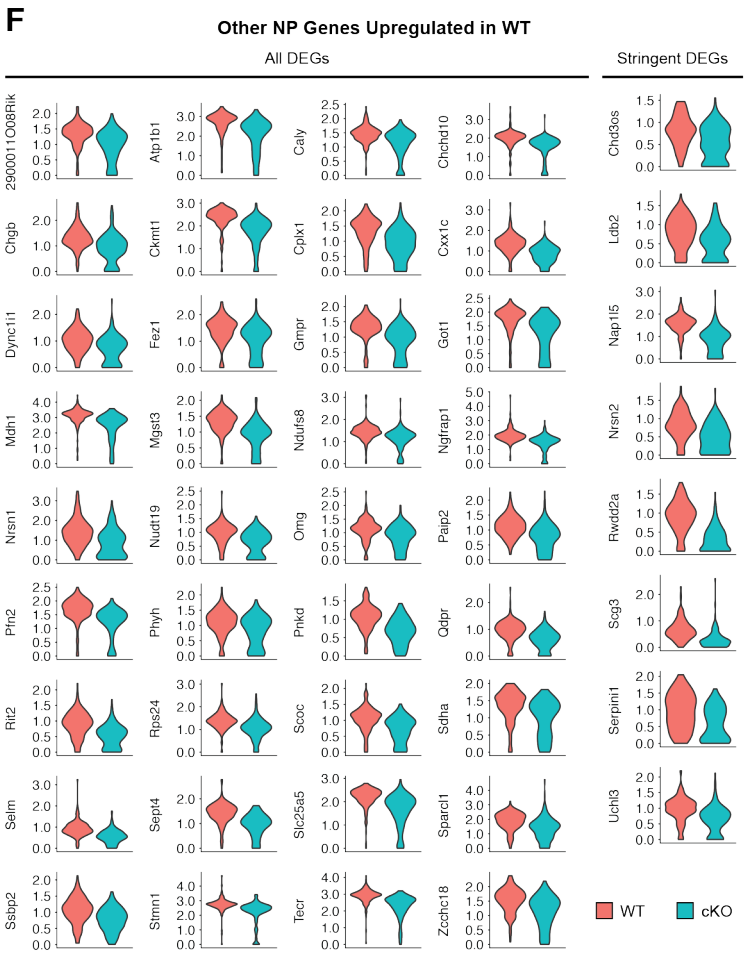
Supplementary Figure A2.8 (continued)



Supplementary Figure A2.8 (continued)



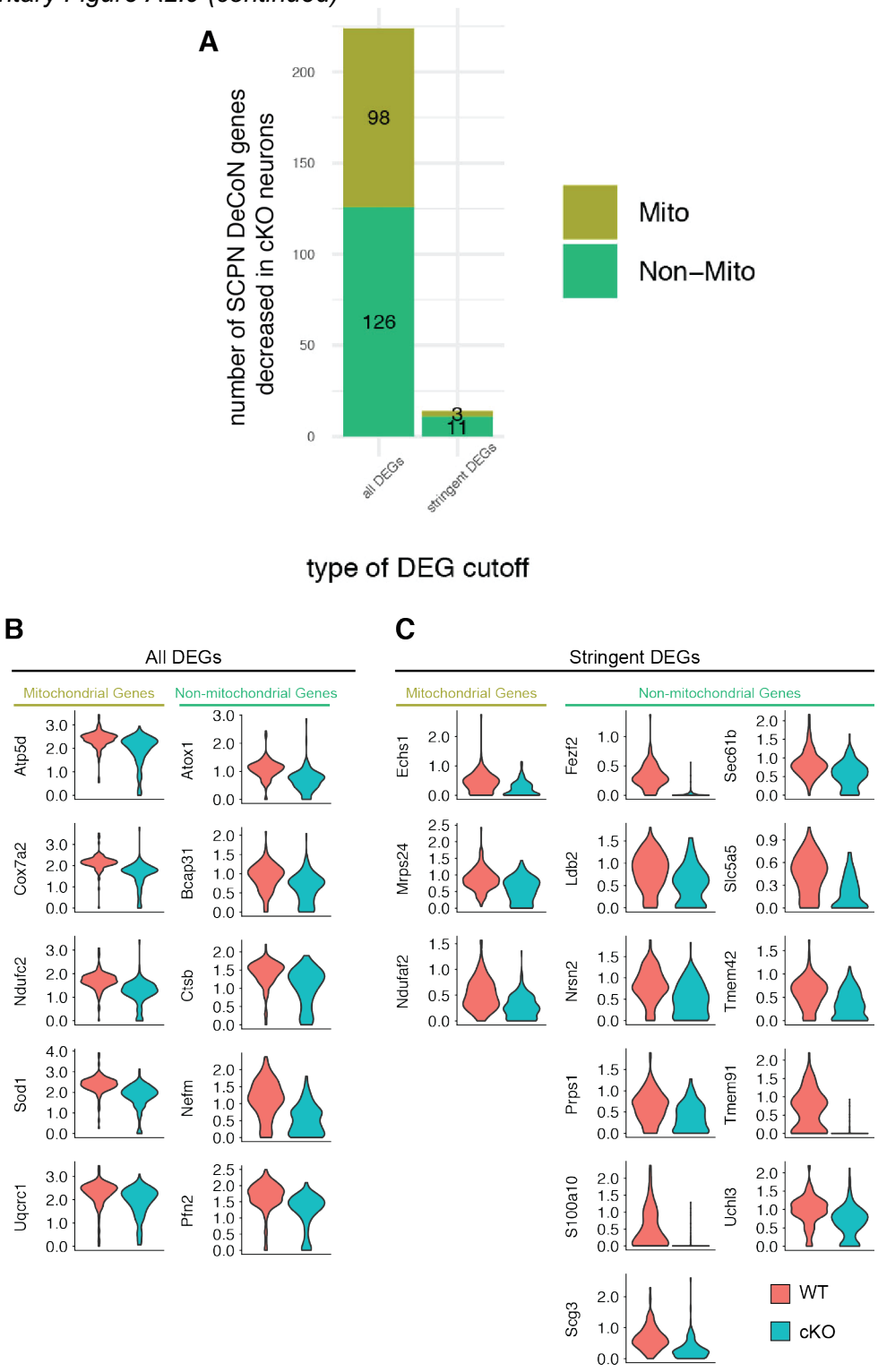
Supplementary Figure A2.8 (continued)



SUPPLEMENTARY FIGURE A2.9. EXPRESSION OF SCPN-ENRICHED NUCLEAR ENCODED MITOCHONDRIAL GENES IS REDUCED IN CKO; AI14 NEURONS WITH EARLY *FEZF2* DELETION

A. Quantification of significantly reduced nuclear-encoded mitochondrial genes in the 'SCPN DeCoN' marker gene set (Mito) versus all other genes in the 'SCPN DeCoN' marker gene set (Non-Mito) in Fezf2 cKO; Ai14 (cKO) neurons as compared to Fezf2 WT; Ai14 (WT) neurons. All differentially expressed genes (DEGs) refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype. ***B.*** Violin plots of selected mitochondrial and non-mitochondrial genes in the 'SCPN DeCoN' gene set that are significantly reduced in cKO neurons. ***C.*** Violin plots of all mitochondrial and non-mitochondrial genes in the 'SCPN DeCoN' gene set that are significantly reduced in cKO neurons and meet our stringent DEG cutoff.

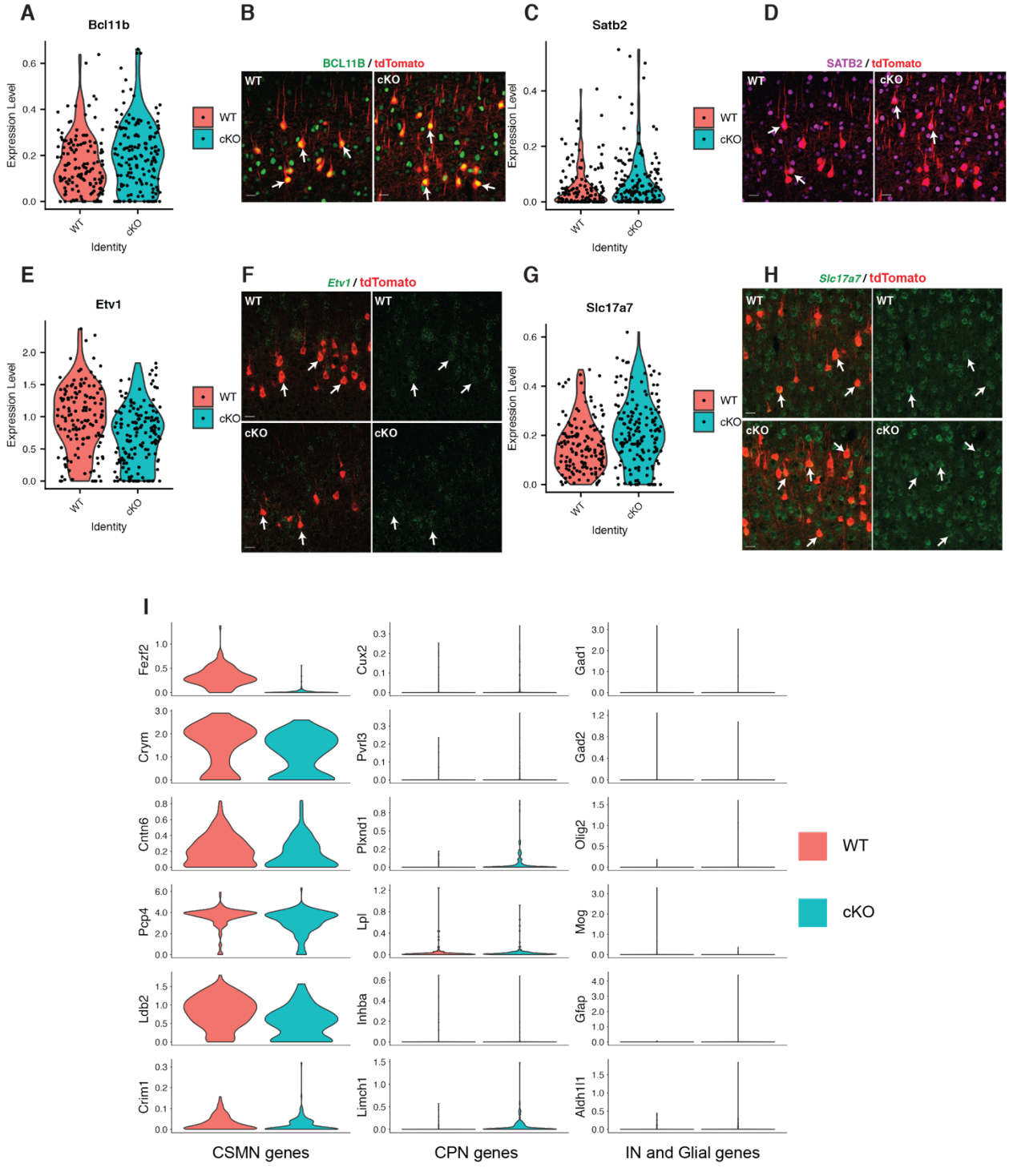
Supplementary Figure A2.9 (continued)

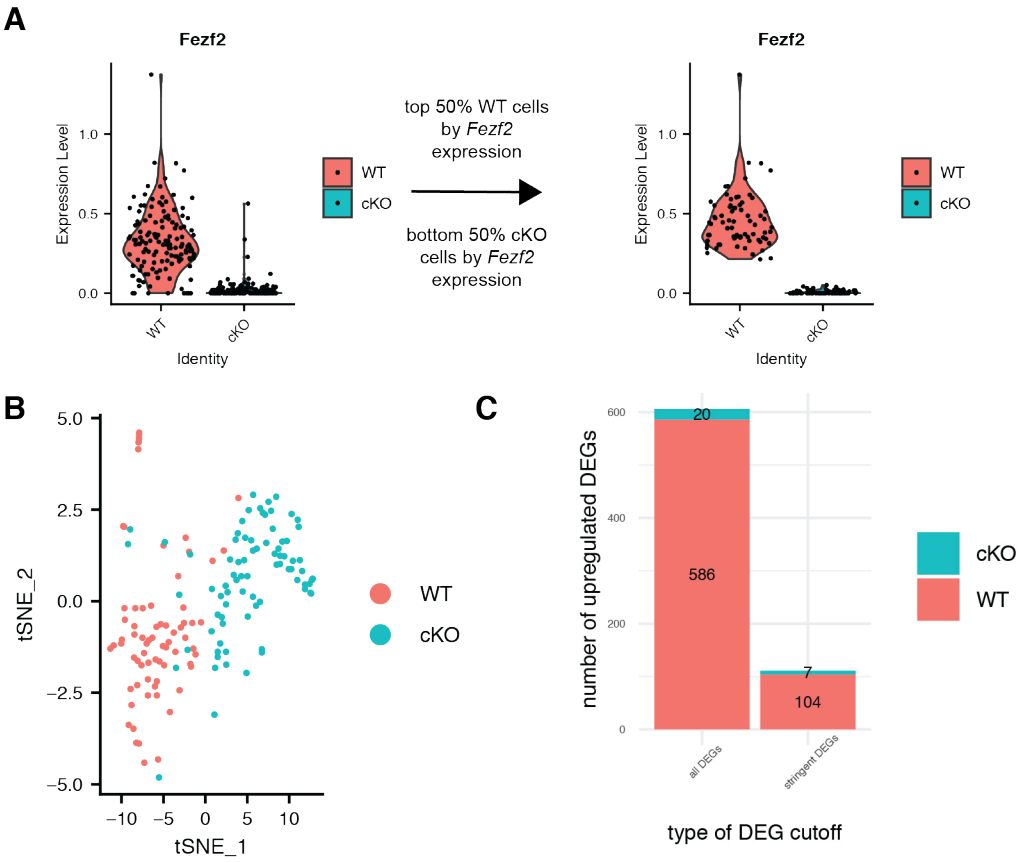


SUPPLEMENTARY FIGURE A2.10. EARLY *FEZF2* DELETION DOES NOT CHANGE THE EXPRESSION OF CLASSICAL SUBTYPE-SPECIFIC GENES

A. Violin plot of *Bcl11b*, a SCPN marker, between genotypes. ***B.*** Immunohistochemistry on Fezf2 WT; Ai14 + Cre (WT) and Fezf2 cKO; Ai14 +Cre (cKO) sections shows no difference in BCL11B. Arrows: tdTomato, Bcl11b double positive cells. ***C.*** Violin plot of *Satb2*, an early CPN marker, between genotypes. ***D.*** Immunohistochemistry on WT and cKO sections shows no difference in SATB2. Arrows: tdTomato, SATB2 double positive cells. ***E.*** Violin plot of *Etv1*, a layer V projection neuron marker, between genotypes. ***F.*** *In situ* hybridization with RNAscope on WT and cKO sections shows no difference. Arrows: tdTomato, *Etv1* double positive cells. ***G.*** Violin plot of *Slc17a7*, the vesicular glutamate transporter, between genotypes. ***H.*** *In situ* hybridization with RNAscope on WT and cKO sections shows no difference. Arrows: tdTomato, *Slc17a7* double positive cells. All scale bars are 25 micron. ***I.*** Violin plots of selected corticospinal motor neuron (CSMN), callosal projection neuron (CPN), interneuron (IN), and glial genes. There is no significant difference between genotypes.

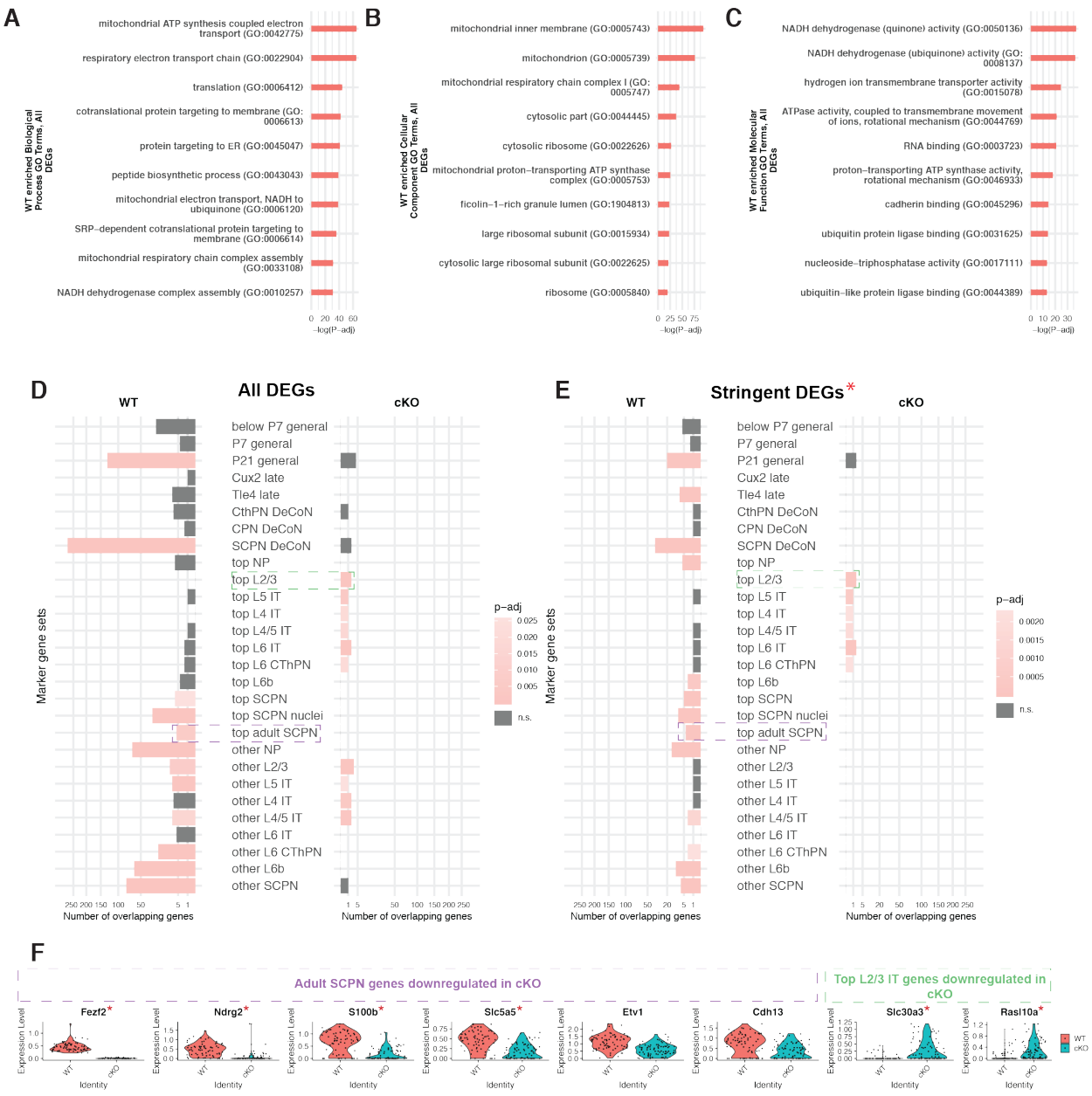
Supplementary Figure A2.10 (continued)





SUPPLEMENTARY FIGURE A2.11. SUBSETTING THE EARLY *FEZF2* DELETION DATASET BASED ON *FEZF2* EXPRESSION FOR SUBANALYSIS

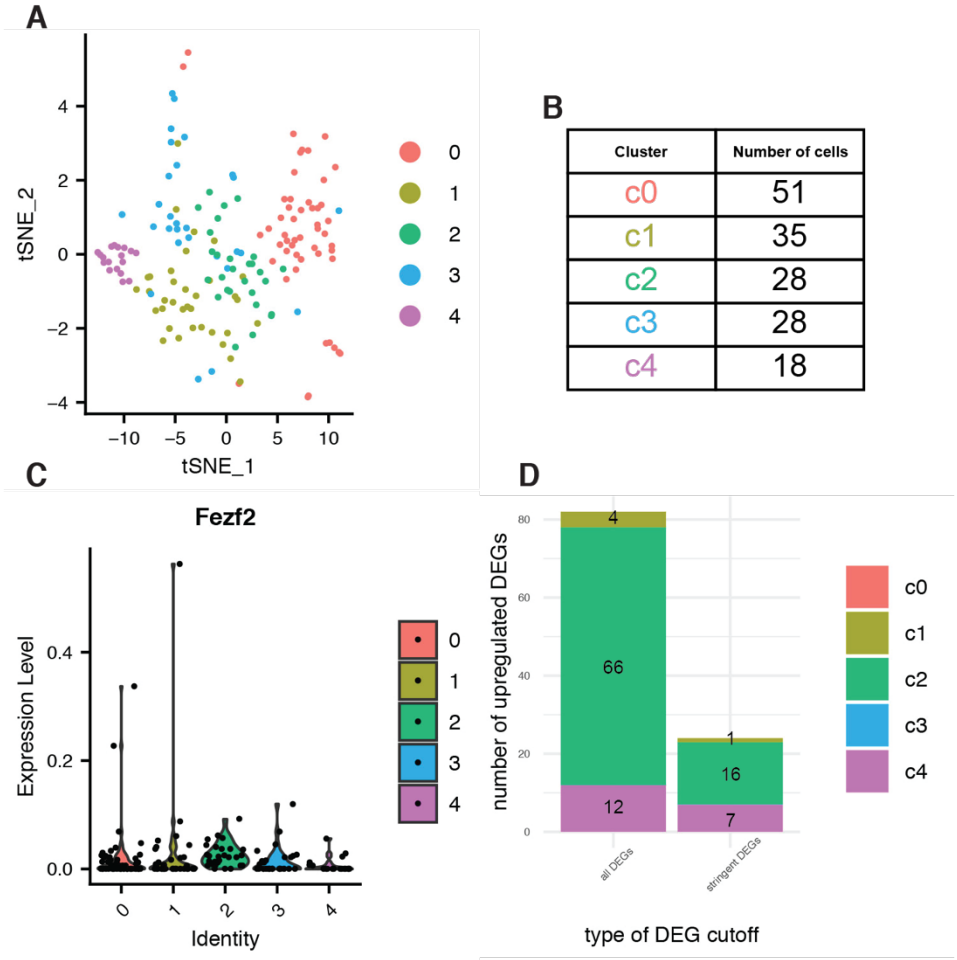
A. Summary of how data were subsetted. Top 50% of Fezf2 WT; Ai14 + Cre (WT) neurons and bottom 50% of Fezf2 cKO; Ai14 + Cre (cKO) neurons as ranked by *Fezf2* expression were taken for subanalysis. ***B.*** Visualization of the subsetted data, showing single cKO and WT neurons by t-distributed stochastic neighbor embedding plot (TSNE) plot. ***C.*** Quantification of number of differentially expressed genes (DEGs) between genotypes. All DEGs refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype.



SUPPLEMENTARY FIGURE A2.12. DATA SUBSETTED ON *FEZF2* EXPRESSION SHOWS A LOSS OF NUCLEAR-ENCODED MITOCHONDRIAL GENES AND A REDUCTION OF SCPN IDENTITY

A. The top 10 most strongly associated GO groups in the DEGs enriched in subsetted Fezf2 WT; Ai14 + Cre (WT) neurons as defined by biological process, ***B.*** cellular component, ***C.*** molecular function, ranked by log adjusted p-value. DEGs enriched in subsetted WT neurons again show strong enrichment of nuclear-encoded mitochondrial genes. P-values are from the Fisher exact test, and adjusted p-values are reported with Benjamini-Hochberg multiple hypothesis correction. ***D.*** All DEGs enriched in various cortical subtype-specific or cortical neuron development gene sets. DEGs enriched in subsetted WT neurons show an enrichment of SCPN-related genes as compared to DEGs enriched in subsetted cKO neurons. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***E.*** Stringent DEGs enriched in various cortical subtype-specific or cortical development gene sets. DEGs enriched in subsetted WT neurons still show an enrichment of SCPN-related genes as compared to DEGs in cKO neurons, but the number of genes is drastically reduced as compared to all DEGs. P-values are from

Supplementary Figure A 2.12 (continued) the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***F.*** Violin plots of DEGs enriched in the 'top adult SCPN' and 'top L2/3' gene sets highlighted previously. Red asterisks denote stringent DEGs.



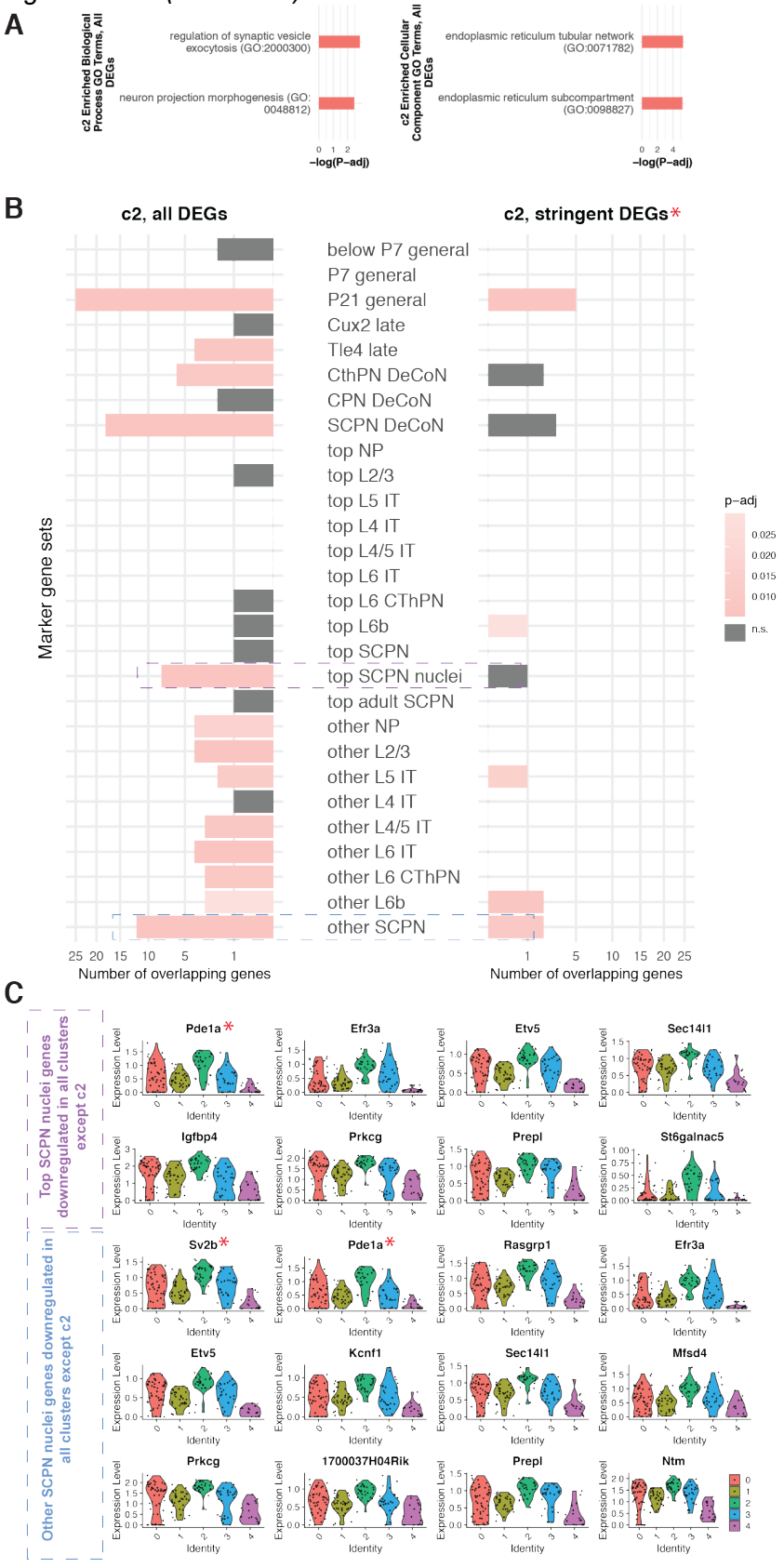
SUPPLEMENTARY FIGURE A2.13. SUBSETTING ONLY NEURONS FROM FEZF2 CKO; AI14 + CRE (CKO) ANIMALS FOR SUBANALYSIS

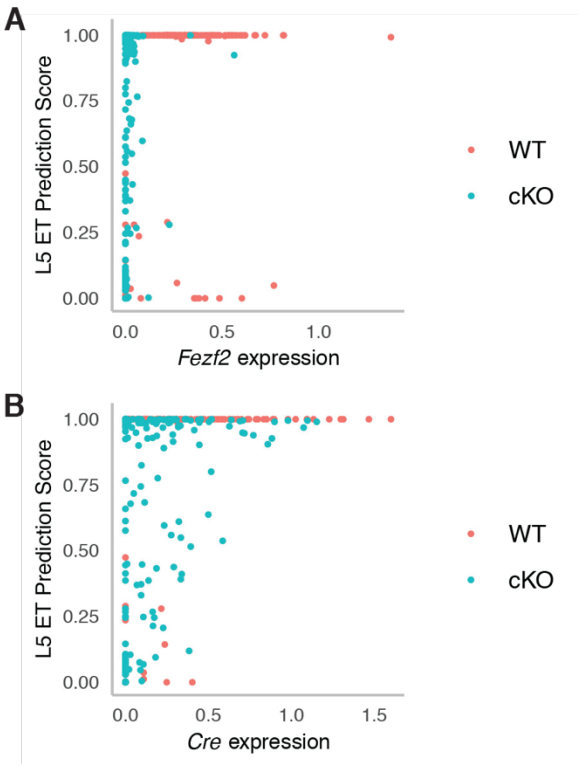
A. Visualization of the subsetted data, showing single cKO neurons and their cluster identification by TSNE plot. ***B.*** Quantification of cell number per cluster. ***C.*** Violin plot of *Fezf2* expression between clusters. ***D.*** Quantification of differentially expressed genes (DEGs) between clusters shows the largest number of significantly upregulated genes comes from cluster c2. All DEGs refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype.

SUPPLEMENTARY FIGURE A2.14. ANALYSIS OF ONLY FEZF2 CKO; AI14 + CRE (CKO) NEURONS SHOWS VARYING LOSS OF SCPN MOLECULAR IDENTITY

A. The only significantly associated GO groups in the DEGs enriched in cluster c2 as defined by biological process and cellular component. P-values are from the Fisher exact test, and adjusted p-values are reported with Benjamini-Hochberg multiple hypothesis correction. ***B.*** Cluster c2 upregulated DEGs (all DEGs and with the stringent cutoff) enriched in various cortical subtype-specific or cortical neuron development gene sets. DEGs upregulated in cluster c2 show enrichment of SCPN-related genes as compared all other clusters, suggesting all other cKO neurons and their clusters lose expression of SCPN-related genes. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. **C.** Violin plots of DEGs enriched in the 'top SCPN nuclei' and 'other SCPN nuclei' gene sets highlighted previously. Red asterisks denote stringent DEGs.

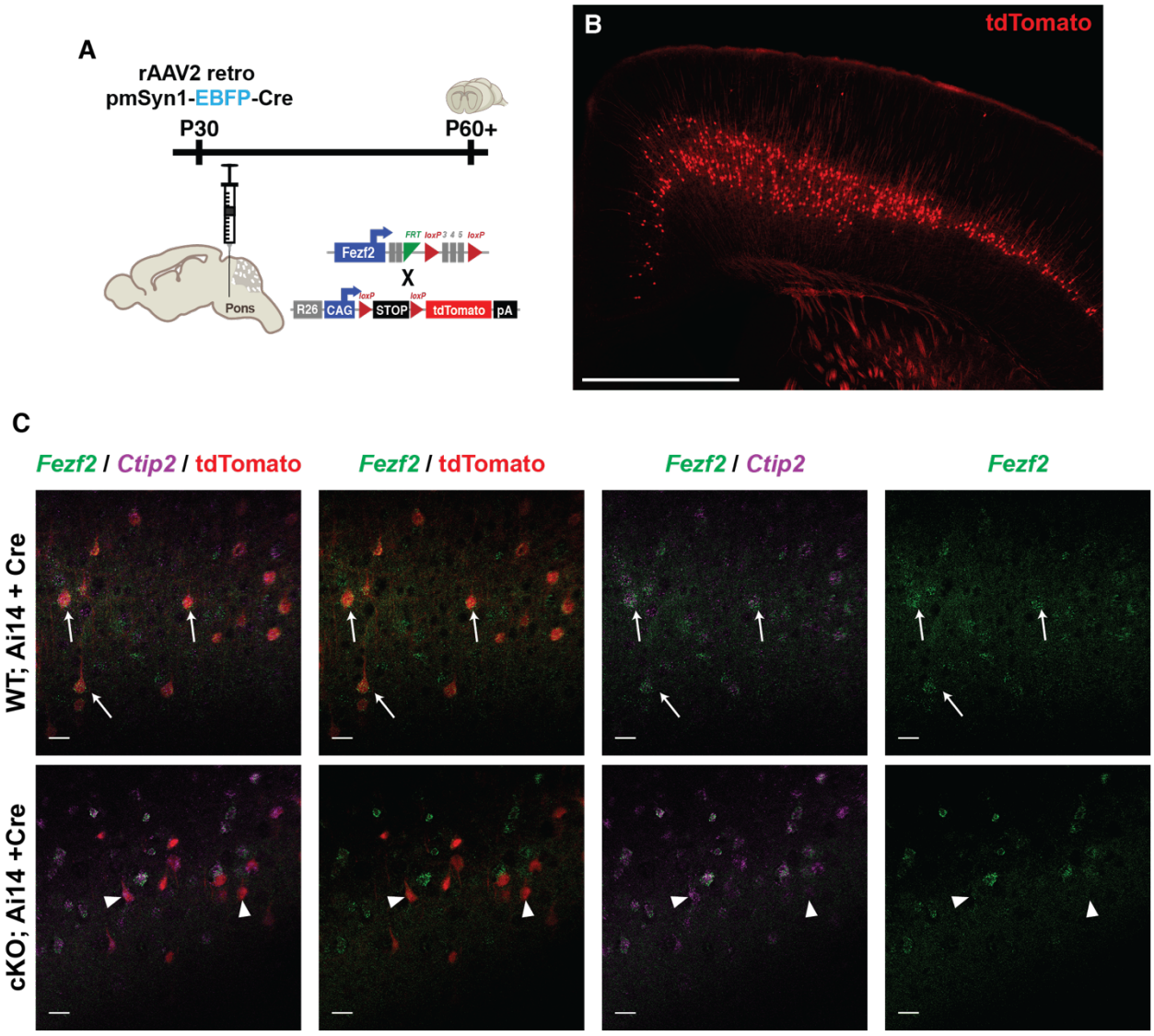
Supplementary Figure A2.14 (continued)





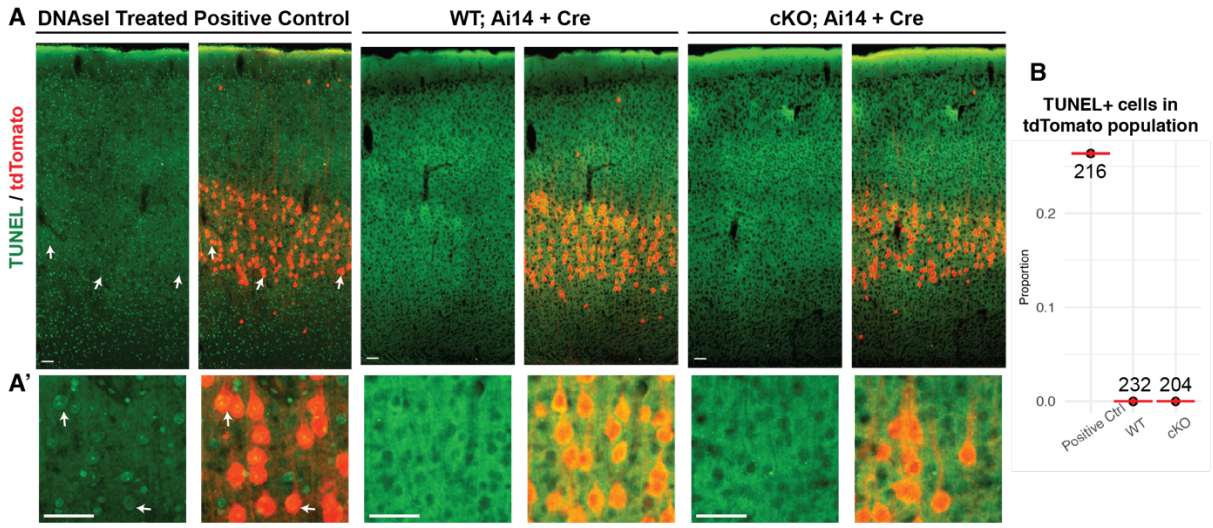
SUPPLEMENTARY FIGURE A2.15. CORRELATION OF L5 ET (EXTRATELENCEPHALIC, OR SCPN) PREDICTION SCORE AND GENE EXPRESSION REVEALS A *FEZF2* EXPRESSION LEVEL INDEPENDENT MECHANISM FOR LOSS OF L5 MOLECULAR IDENTITY

A. Scatter plot of L5 ET predicted scores versus *Fezf2* expression by genotype shows little correlation. ***B.*** Scatter plot of L5 ET predicted scores versus *Cre* expression by genotype shows a slight negative correlation in cKO neurons.



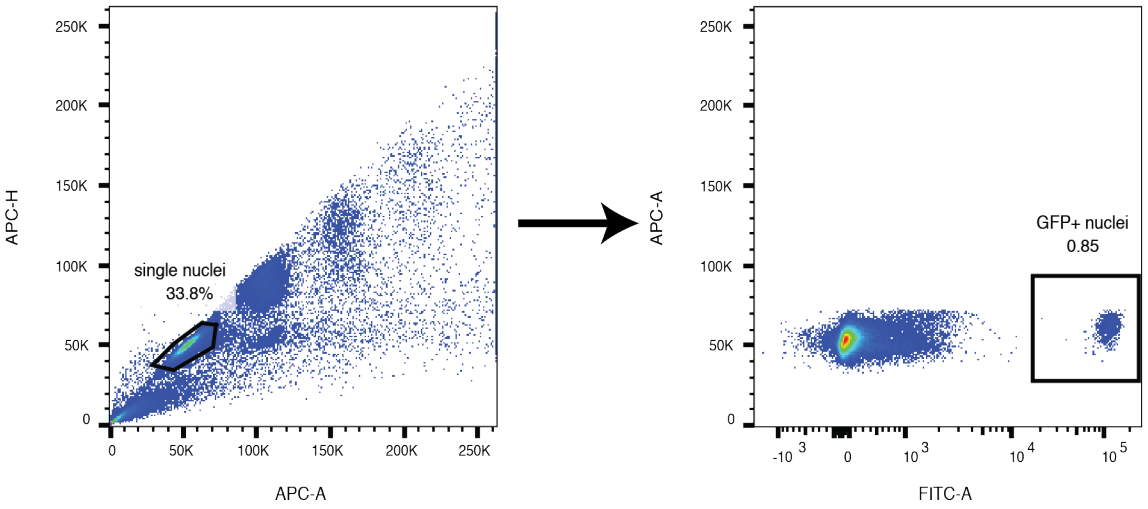
SUPPLEMENTARY FIGURE A2.16. *FEZF2* CAN BE DELETED AT LATE POSTNATAL AGES

A. Experimental design. ***B.*** tdTomato is seen in the cortex after P30 injection of Cre in the pons. Scale bar is 1000 micron. ***C.*** *Fezf2* expression is gone from tdTomato+ cells by RNAscope *in situ* hybridization. Arrow: tdTomato positive SCPN with *Ctip2 (Bcl11b)* and without *Fezf2*; arrowhead: tdTomato positive SCPN with *Fezf2* and *Ctip2 (Bcl11b)*. Scale bars are 50 microns.



SUPPLEMENTARY FIGURE A2.17. LATE *FEZF2* DELETION DOES NOT CAUSE CELL DEATH

A. TUNEL stain of P30 injected animals and sacrificed at 6 WPI, with magnified insets (***A'***). Arrows point to tdTomato, TUNEL double positive cells. Scale bars are 50 microns. ***B.*** Quantification of tdTomato and TUNEL double positive cells. Numbers above bars indicate total tdTomato cells counted per condition. Red diamond indicates mean, red bars indicate standard error of the mean.



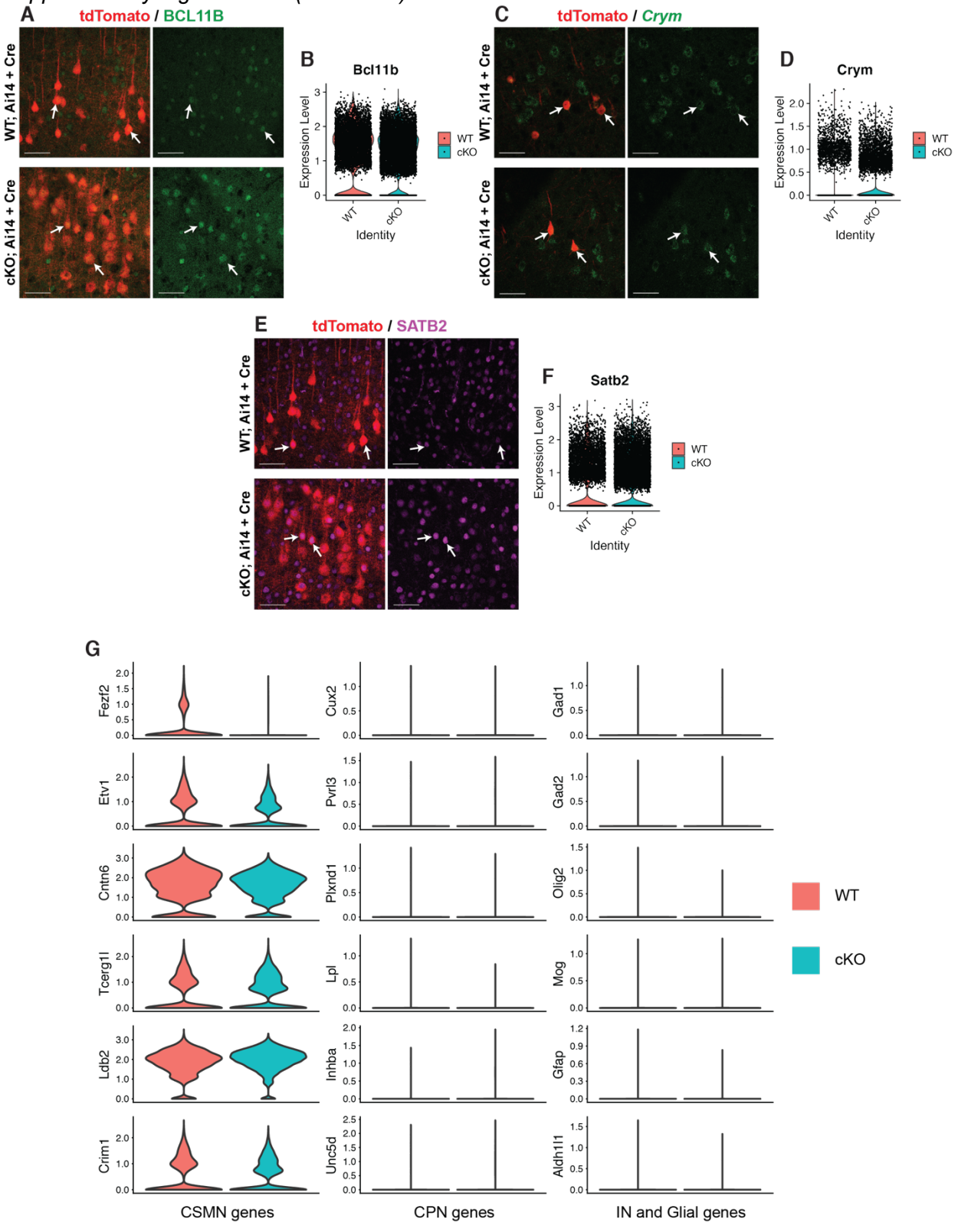
SUPPLEMENTARY FIGURE A2.18. FACS GATING STRATEGY OF FEZF2 WT; SUN1-GFP + CRE AND FEZF2 CKO; SUN1-GFP + CRE ANIMALS

Representative FACS gating strategy. Vybrant DyeCycle Ruby positive singlets and GFP positive nuclei are collected prior to sequencing.

SUPPLEMENTARY FIGURE A2.19. LATE *FEZF2* DELETION DOES NOT THE EXPRESSION OF CLASSICAL SUBTYPE-SPECIFIC MARKERS

A. Immunohistochemistry on Fezf2 WT; Sun1-GFP + Cre (WT) and Fezf2 cKO; Sun1-GFP + Cre (cKO) sections shows no difference in CTIP2 (BCL11B), a SCPN marker. Arrows: tdTomato, BCL11B double positive cells. ***B.*** Violin plot of *Bcl11b* between genotypes ***C.*** *In situ* hybridization of *Crym*, a CSMN marker, with RNAscope on WT and cKO sections shows no difference. Arrows: tdTomato, *Crym* double positive cells. ***D.*** Violin plot of *Crym* between genotypes. ***E.*** Immunohistochemistry on WT and cKO sections shows no difference in SATB2, an early CPN marker. Arrows: tdTomato, SATB2 double positive cells. ***F.*** Violin plot of *Satb2* between genotypes. ***G.*** Violin plots of selected corticospinal motor neuron (CSMN), callosal projection neuron (CPN), interneuron (IN), and glial genes. There is no significant difference between genotypes.

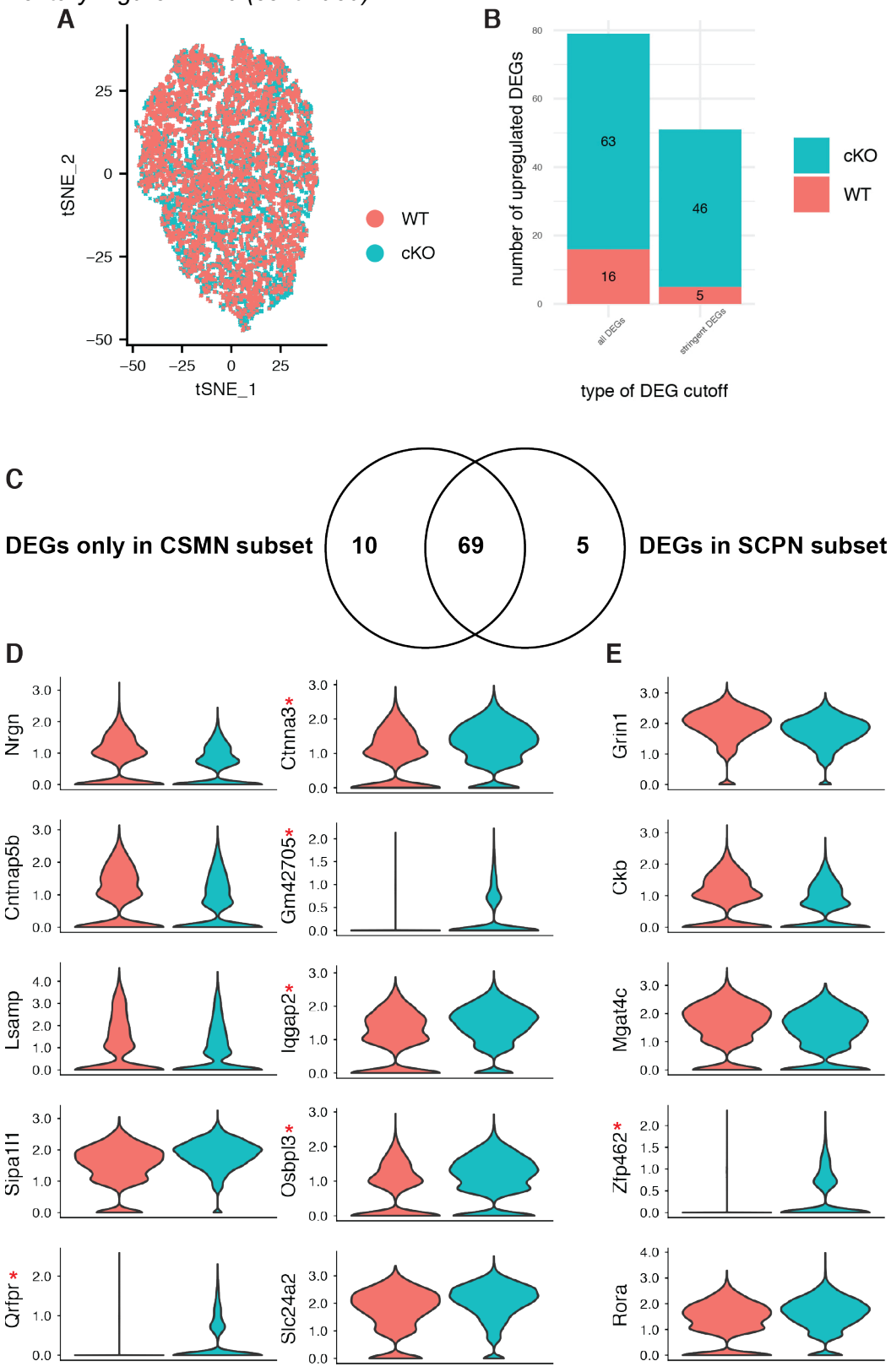
Supplementary Figure A2.19 (continued)



SUPPLEMENTARY FIGURE A2.20. SUBSETTING ONLY PUTATIVE CSMNS FOR SUBANALYSIS SHOWS MANY SHARED GENES AS COMPARED TO ANALYZING ALL SCPNS

A. *Satb2* and *Tshz2* low expressing clusters from Fezf2 cKO; Sun1-GFP (cKO) and Fezf2 WT; Sun1-GFP (WT) dataset were reanalyzed separately as putative CSMNs. Visualization of the subsetted data, showing single nuclei by t-distributed stochastic neighbor embedding plot (TSNE) plot. ***B.*** Quantification of number of differentially expressed genes (DEGs) between genotypes. All DEGs refer to any DEG between genotypes; stringent DEGs refer only to DEGs that have a minimum 10% difference in the number of expressing cells per genotype. ***C.*** Venn diagram showing the number of overlapping of DEGs between the two analyses. When comparing the DEGs between genotypes among all SCPNs versus only among CSMNs, most genes (69) are shared between analyses. ***D.*** Violin plots of DEGs unique to the CSMN only analysis. Red asterisks denote stringent DEGs. ***E.*** Violin plots of DEGs unique to the analysis with all SCPNs. Red asterisks denote stringent DEGs.

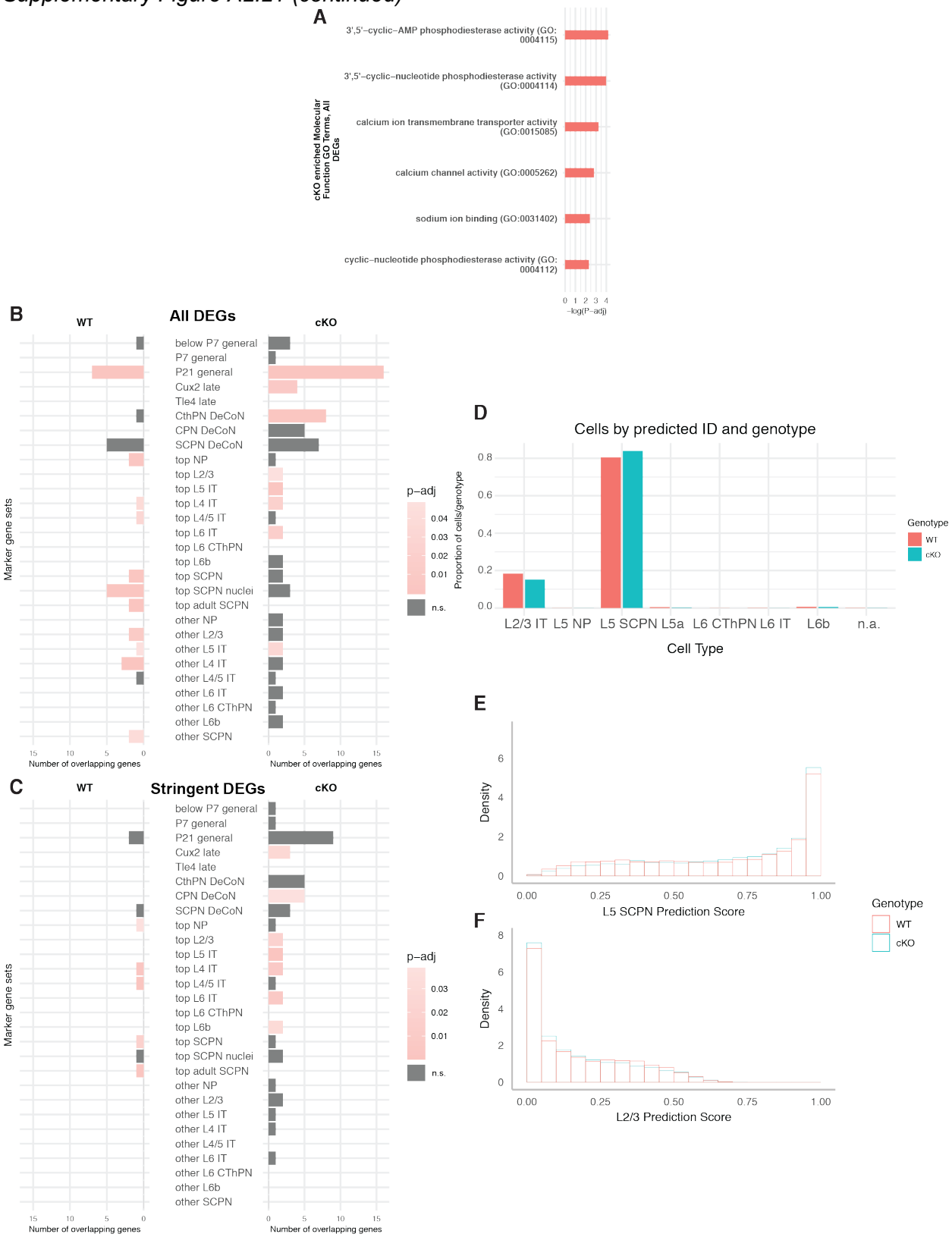
Supplementary Figure A2.20 (continued)



SUPPLEMENTARY FIGURE A2.21. LATE *FEZF2* DELETION IN PUTATIVE CSMNS SHOWS MOLECULAR CHANGES, BUT NOT A CHANGE IN SCPN IDENTITY.

A. The only significantly associated GO groups in the DEGs are enriched in putative CSMNs of Fezf2 cKO; Sun1-GFP neurons, ranked by log adjusted p-value. P-values are from the Fisher exact test, and adjusted p-values are reported with Benjamini-Hochberg multiple hypothesis correction. ***B.*** All DEGs enriched in various cortical subtype-specific or cortical neuron development gene sets. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***C.*** Stringent DEGs enriched in various cortical subtype-specific or cortical development gene sets. P-values are from the hypergeometric test, and adjusted p-values are reported with Bonferroni multiple hypothesis correction. n.s.: not significant. ***D.*** Proportion of 5 WPI neurons and their predicted adult cortical neuron cell type identification. ***E.*** Distribution of L5 SCPN neuron prediction scores by genotype. ***F.*** Distribution of L2/3 IT neuron prediction scores by genotype.

Supplementary Figure A2.21 (continued)



Tables

SUPPLEMENTARY TABLE A2.1. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN LAYER VB FEZF2 HZ AND FEZF2 KO NEURONS AT P1.

Significantly differentially expressed genes between layer Vb Fezf2 HZ and Fezf2 KO neurons at P1. Significant calculated with the Wald test, as implemented in DESeq2 R package. FDR <0.1. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.2. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN TCERG-CREERT2; FEZF2 WT; AI14 AND TCERG-CREERT2; FEZF2 CKO; AI14 ANIMALS

Significantly differentially expressed genes between Tcerg-CreERT2; Fezf2 WT; Ai14 and Tcerg-CreERT2; Fezf2 cKO; Ai14 neurons. Animals received 4-hydroxytamoxifen at P3 and were sacrificed three weeks later. DEGs were calculated with the Wald test as implemented in DESeq2 R package. FDR <0.05. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.3. LISTS OF GENE ONTOLOGY (GO) TERMS ENRICHED IN DEGS UPREGULATED IN TCERG-CREERT2; FEZF2 WT; AI14 NEURONS OR TCERG-CREERT2; FEZF2 CKO; AI14 NEURONS

All associated GO terms found in DEGs upregulated in Tcerg-CreERT2; Fezf2 WT; Ai14 (WT) neurons as compared to Tcerg-CreERT2; Fezf2 cKO; Ai14 (cKO) neurons, as implemented in the EnrichR R package. GO groups are separated into three categories: biological process, cellular component, and molecular function. Overlap: number of DEGs overlapping with genes in GO group; p-value: p-value calculated from the Fisher exact test; adjusted p-value: adjusted p-value, calculated using the Benjamani-Hochberg multiple hypothesis correction; combined score: log of the p-value from Fisher exact test, multiplied by z-score of the deviation from the expected rank; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.4. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN FEZF2 WT; AI14 + CRE AND FEZF2 CKO; AI14 + CRE NEURONS, P1 INJECTION

Significant differentially expressed genes (DEGs) between Fezf2 WT; Ai14 + Cre (WT) and Fezf2 cKO; Ai14 + Cre (cKO) animals. Cre was delivered to SCPNs via the pons at P1, and tdTomato positive neurons were collected three weeks later. DEGs were calculated using the MAST algorithm as implemented in the Seurat R package. Adjusted p-values are reported with Bonferroni correction. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.5. LISTS OF GO TERMS ENRICHED IN ALL DEGS UPREGULATED IN FEZF2 WT; AI14 + CRE NEURONS AS COMPARED TO FEZF2 CKO; AI14 + CRE NEURONS, P1 INJECTION

All associated GO terms found in DEGs upregulated in Fezf2 WT; Ai14 + Cre neurons as compared to Fezf2 cKO; Ai14 + Cre neurons, as implemented in the EnrichR R package. GO groups are separated into three categories: biological process, cellular component, and molecular function. Overlap: number of DEGs overlapping with genes in GO group; p-value: p-value calculated from the Fisher exact test; adjusted p-value: adjusted p-value, calculated using the Benjamani-Hochberg multiple hypothesis correction; combined score: log of the p-value from Fisher exact test, multiplied by z-score of the deviation from the expected rank; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.6. LISTS OF CORTICAL SUBTYPE-SPECIFIC OR CORTICAL NEURON DEVELOPMENT GENE SETS ENRICHED IN DEGS OF FEZF2 WT; AI14 + CRE NEURONS AS COMPARED TO FEZF2 CKO; AI14 + CRE NEURONS, P1 INJECTION

DEGs from P1 Cre injection of Fezf2 WT; Ai14 + Cre neurons as compared to Fezf2 cKO; Ai14 + Cre neurons enriched in various cortical subtype-specific or cortical neuron development gene sets. P-value calculated using the hypergeometric test function in R and adjusted for multiple hypotheses with the Bonferroni correction. Gene set: gene set used (see Appendix 1 for how gene sets were generated); adjusted p-value: Bonferroni adjusted p-value from hypergeometric test; overlap: number of DEGs overlapping with genes in gene set; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.7. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN TOP 50% OF FEZF2 WT; AI14 + CRE AND BOTTOM 50% FEZF2 CKO; AI14 + CRE NEURONS AS RANKED BY *FEZF2* EXPRESSION

Significant DEGs between top 50% of Fezf2 WT; Ai14 +Cre (WT) and bottom 50% of Fezf2 cKO; Ai14 + Cre (cKO) neurons as ranked by *Fezf2* expression. Cre was delivered to SCPNs via the pons at P1, and tdTomato positive neurons were collected three weeks later. DEGs were calculated using the MAST algorithm as implemented in the Seurat R package. Adjusted p-values are reported with Bonferroni correction. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.8. LISTS OF GO TERMS ENRICHED IN ALL DEGS UPREGULATED IN TOP 50% OF FEZF2 WT; AI14 + CRE NEURONS AS COMPARED TO THE BOTTOM 50% OF FEZF2 CKO; AI14 + CRE NEURONS, AS RANKED BY *FEZF2* EXPRESSION

All associated GO terms found in DEGs upregulated in the top 50% of Fezf2 WT; Ai14 + Cre neurons as compared to the bottom 50% Fezf2 cKO; Ai14 + Cre neurons, as ranked by *Fezf2* expression. Associated terms were found using the EnrichR R package. GO groups are separated into three categories: biological process, cellular component, and molecular function. Overlap: number of DEGs overlapping with genes in GO group; p-value: p-value calculated from the Fisher exact test; adjusted p-value: adjusted p-value, calculated using the Benjamani-Hochberg multiple hypothesis correction; combined score: log of the p-value from Fisher exact test, multiplied by z-score of the deviation from the expected rank; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.9. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN FEZF2 CKO; AI14 + CRE NEURONS.

Significant DEGs between Fezf2 cKO; Ai14 + Cre neuronal subclusters. Cre was delivered to SCPNs via the pons at P1, and tdTomato positive neurons were collected three weeks later. DEGs were calculated using the MAST algorithm as implemented in the Seurat R package. Adjusted p-values are reported with Bonferroni correction. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.10. LISTS OF CORTICAL SUBTYPE-SPECIFIC OR CORTICAL NEURON DEVELOPMENT GENE SETS ENRICHED IN DEGS BETWEEN FEZF2 CKO; AI14 + CRE NEURONS

DEGs from P1 Cre injection of Fezf2 cKO; Ai14 + Cre subclusters enriched in various cortical subtype-specific or cortical neuron development gene sets. P-value calculated using the hypergeometric test function in R and adjusted for multiple hypotheses with the Bonferroni correction. Gene set: gene set used (see Appendix 1 for how gene sets were generated); adjusted p-value: Bonferroni adjusted p-value from hypergeometric test; overlap: number of DEGs overlapping with genes in gene set; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.11. LISTS OF GO TERMS ENRICHED IN ALL DEGS UPREGULATED IN CLUSTER C2 OF FEZF2 CKO; AI14 + CRE NEURONS

All associated GO terms found in DEGs upregulated in cluster c2 of Fezf2 cKO; Ai14 + Cre neurons, as implemented in the EnrichR R package. GO groups are separated into two categories: biological process and cellular component. Overlap: number of DEGs overlapping with genes in GO group; p-value: p-value calculated from the Fisher exact test; adjusted p-value: adjusted p-value, calculated using the Benjamani-Hochberg multiple hypothesis correction; combined score: log of the p-value from Fisher exact test, multiplied by z-score of the deviation from the expected rank; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.12. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN FEZF2 WT; SUN1-GFP+ CRE AND FEZF2 CKO; SUN1-GFP+ CRE SCPNS, P30 INJECTION

Significant differentially expressed genes (DEGs) between Fezf2 WT; Sun1-GFP + Cre (WT) and Fezf2 cKO; Sun1-GFP + Cre (cKO) animals. Cre was delivered to SCPNs via the pons at P30, and GFP positive nuclei were collected five weeks later. DEGs were calculated using the MAST algorithm as implemented in the Seurat R package. Adjusted p-values are reported with Bonferroni correction. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.13. LIST OF GO TERMS ENRICHED IN ALL DEGS UPREGULATED IN FEZF2 CKO; SUN1-GFP + CRE AS COMPARED TO FEZF2 WT; SUN1-GFP + CRE NUCLEI, P30 INJECTION

All associated GO terms found in DEGs upregulated in Fezf2 cKO; Sun1-GFP + Cre nuclei, as implemented in the EnrichR R package. Overlap: number of DEGs overlapping with genes in GO group; p-value: p-value calculated from the Fisher exact test; adjusted p-value: adjusted p-value, calculated using the Benjamani-Hochberg multiple hypothesis correction; combined score: log of the p-value from Fisher exact test, multiplied by z-score of the deviation from the expected rank; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.14. LISTS OF CORTICAL SUBTYPE-SPECIFIC OR CORTICAL NEURON DEVELOPMENT GENE SETS ENRICHED IN DEGS OF FEZF2 WT; SUN1-GFP + CRE NUCLEI AS COMPARED TO FEZF2 WT; SUN1-GFP + CRE NUCLEI, P30 INJECTION

DEGs from P30 Cre injection of Fezf2 WT; Sun1-GFP + Cre (WT) and Fezf2 cKO; Sun1-GFP + Cre (cKO) nuclei enriched in various cortical subtype-specific or cortical neuron development gene sets. P-value calculated using the hypergeometric test function in R and adjusted for multiple hypotheses with the Bonferroni correction. Gene set: gene set used (see Appendix 1 for how gene sets were generated); adjusted p-value: Bonferroni adjusted p-value from hypergeometric test; overlap: number of DEGs overlapping with genes in gene set; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.15. LIST OF SIGNIFICANTLY DIFFERENTIALLY EXPRESSED GENES BETWEEN PUTATIVE CSMNS IN FEZF2 WT; SUN1-GFP+ CRE AND FEZF2 CKO; SUN1-GFP+ CRE NUCLEI, P30 INJECTION

Significant differentially expressed genes (DEGs) between putative CSMNs in Fezf2 WT; Sun1-GFP + Cre (WT) and Fezf2 cKO; Sun1-GFP + Cre (cKO) animals. Cre was delivered to SCPNs via the pons at P30, and GFP positive nuclei were collected five weeks later. Only putative CSMNs were used to find DEGs. DEGs were calculated using the MAST algorithm as implemented in the Seurat R package. Adjusted p-values are reported with Bonferroni correction. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.16. LIST OF GO TERMS ENRICHED IN DEGS UPREGULATED IN PUTATIVE CSMNS OF FEZF2 CKO; SUN1-GFP + CRE AS COMPARED TO FEZF2 WT; SUN1-GFP + CRE NUCLEI, P30 INJECTION

All associated GO terms found in DEGs upregulated in Fezf2 cKO; Sun1-GFP + Cre nuclei, as implemented in the EnrichR R package. Only putative CSMNs were used to find DEGs for this analysis. Overlap: number of DEGs overlapping with genes in GO group; p-value: p-value calculated from the Fisher exact test; adjusted p-value: adjusted p-value, calculated using the Benjamani-Hochberg multiple hypothesis correction; combined score: log of the p-value from Fisher exact test, multiplied by z-score of the deviation from the expected rank; genes: overlapping genes. *Electronic attachment.*

SUPPLEMENTARY TABLE A2.17. LISTS OF CORTICAL SUBTYPE-SPECIFIC OR CORTICAL NEURON DEVELOPMENT GENE SETS ENRICHED IN DEGS OF PUTATIVE CSMNS OF FEZF2 WT; SUN1-GFP + CRE NUCLEI AS COMPARED TO FEZF2 WT; SUN1-GFP + CRE NUCLEI, P30 INJECTION

DEGs from P30 Cre injection of Fezf2 WT; Sun1-GFP + Cre (WT) and Fezf2 cKO; Sun1-GFP + Cre (cKO) nuclei enriched in various cortical subtype-specific or cortical neuron development gene sets. Only putative CSMNs were used to find DEGs for this analysis. P-value calculated using the hypergeometric test function in R and adjusted for multiple hypotheses with the Bonferroni correction. Gene set: gene set used (see Appendix 1 for how gene sets were generated); adjusted p-value: Bonferroni adjusted p-value from hypergeometric test; overlap: number of DEGs overlapping with genes in gene set; genes: overlapping genes. *Electronic attachment.*

Appendix 3.

Extracellular Matrix Dependent Maintenance of Cortical Neuronal Identity

While most of the work of my dissertation was on the role of Fezf2 *in maintaining corticospinal motor neuron identity, I initially worked on a separate project, aimed at understanding the role of the extracellular matrix in maintaining neuronal identity. Although related to the work described in Chapter II, it is distinct. I therefore am including it here as an appendix.*

Author contributions: I designed the study with Paola Arlotta. I did all experiments and analysis except: Cux2 and Tle4 RNA-sequencing dataset generation (Wen Yuan and Michael Ziller, manuscript in preparation); DF-ChABC construct generation (Juliana Brown).

Introduction

The neocortex contains a great diversity of neuronal subtypes that wire precisely into local and long-distance circuits to execute higher-order functions. All cortical neurons belong to two broad classes: glutamatergic projection neurons (PNs) that act as the primary drivers of cortical output, and inhibitory, GABAergic, locally-targeting interneurons (INs) that refine and modulate the excitatory output from the PNs (Greig et al., 2013; Harris et al., 2015; Harris and Shepherd, 2015; Lodato et al., 2015). PNs are born during embryonic development from the germinal zones in the dorsal telencephalon; INs, on the other hand, are derived from the ventral telencephalon and migrate a substantial distance tangentially to reach the cortical plate, where they then migrate radially to reach their final laminar location to wire with PNs (Greig et al., 2013; Harris et al., 2015; Harris and Shepherd, 2015; Lodato et al., 2015). To understand the cell-autonomous mechanisms that control the generation and maintenance of individual classes of PNs, defined PN subtypes have been purified and molecularly profiled across development (Arlotta et al., 2005; Molyneaux et al., 2009; Molyneaux et al., 2015; Saunders et al., 2018; Tasic et al., 2016; Tasic et al., 2018). However, much less is known regarding the role of the extracellular environment.

Functions of ECM in the central nervous system (CNS)

The collection of molecules secreted into the extracellular space form heterogeneous aggregates, and is collectively known as the extracellular matrix. The ECM provides structural support and acts as a reservoir of soluble ligands for the surrounding cells to regulate a wide range of function. The ECM and its associated molecules can be thought of as composed of 1. the core molecules (fibronectins, collagens, etc.) and the proteins that share similar biochemical properties, 2. molecules that regulate the core ECM proteins (such as remodeling enzymes), and 3. secreted factors that can reside within the ECM (Hynes, 2009; Naba et al., 2012).

It has long been observed that both glia and neurons contribute to the ECM. Many ECM molecules have been implicated in cortical development, from neuronal migration to the acquisition and maintenance of neuronal circuits in the CNS (Barros et al., 2011). For example, during postnatal maturation, the aggregation of perineuronal nets (PNNs), a condensed and structured form of ECM, around parvalbumin-positive interneurons contributes to the closure of the period of critical period plasticity (Hensch, 2005; Pizzorusso et al., 2002). ECM-related molecules can also regulate synaptic plasticity by modulating receptor composition at the synapse and spine motility (Alexander et al., 2010; Levy et al., 2014); for example, activation of beta-1 integrin can increase phosphorylation of GluN2A and GluN2B to promote LTP in the hippocampus (Bernard-Trifilo et al., 2005). In addition, there is evidence for activity-dependent ECM remodeling. MMP9, a matrix metalloproteinase, has proteolytic activity on many membrane proteins, cytokines, and growth factors, and has been shown to be upregulated upon neuronal stimulation in the hippocampus (Szklarczyk et al., 2002). Curiously, direct lineage reprogramming between cortical neuronal subtypes is able to rewire inhibitory input, suggesting that PNs may directly regulate their microenvironment to influence the behavior of other neurons (Ye et al., 2015). Together, this evidence suggests that neurons contribute substantially to the heterogeneous ECM. However, there is still a lack of direct evidence of the extent of neuronal contribution to their

local ECM. There are references to differential patterns of expression of ECM-related molecules in the cortex, such as the layer-specific expression of various members of integrins (Ralf et al., 2003), but this is not understood at the level of the individual neuronal subtypes.

Reciprocal restriction of the ECM on molecular and structural neuronal identity

Neuron-ECM interactions are reciprocal; while neurons contribute to the establishment of the ECM, the ECM in turn regulates structural plasticity of the neuron as well, both physically and biochemically. Proteolytic digestion of endogenous PNNs after the critical period restores ocular dominance plasticity in adult mice (Hensch, 2005; Pizzorusso et al., 2002). CNS injury paradigms and transplantation studies suggest that glia produce inhibitory ECM molecules that prevent axonal regeneration; degradation of the ECM robustly improves functional recovery and regeneration, suggesting that the ECM may also sequester inhibitory growth cues (Bartus et al., 2014; García-Alías et al., 2009; Silver and Miller, 2004). Furthermore, the ECM can regulate the cell-cycle state of the cell. In cancer biology, misregulated ECM promotes the growth and proliferation of cancerous tissue (Bonnans et al., 2014). These studies raise the question of what other cues the ECM may contain to maintain neuronal properties, such as the cell-intrinsic neuronal identity.

Our lab has extensively studied the identities of the callosal projection neurons (CPNs) that connect the two hemispheres via the corpus callosum, and the corticofugal projection neurons (CFuPNs), which include the subcerebral projection neurons (SCPNs) that project to targets in the brainstem and spinal cord (Arlotta et al., 2005). Fezf2, a master transcription factor for SCPNs (Chen et al., 2005; Chen et al., 2008; Molyneaux et al., 2005; Lodato et al., 2014), is sufficient to instruct direct lineage conversion of early post-mitotic CPNs into CFuPNs; reprogrammed CPNs acquire molecular properties of CFuPNs and can change axonal connectivity from interhemispheric to subcerebral (Rouaux and Arlotta, 2013; Ye et al., 2015) .

Furthermore, the dendritic morphology and intrinsic electrophysiological properties change in early-reprogrammed CPNs toward that of CFuPNs (Ye et al., 2015). All CPNs that receive *Fezf2* early post-mitotically have broad tuft arborization, like endogenous layer V CFuPN (Alexis and Sacha, 2007; Ye et al., 2015). Whole-cell recordings showed that *Fezf2*-overexpression in early post-mitotic CPNs shifts their intrinsic electrophysiological properties in the direction of CFuPNs, as compared to control GFP-expressing CPNs (Ye et al., 2015). However, Fezf2 is unable to reprogram CPNs after P21 (Rouaux and Arlotta, 2013), prompting the question of why this temporal restriction of neuronal reprogramming occurs. Based on current understanding of the role of the ECM in restricting plasticity in the CNS, it is possible that the postnatal maturation of the ECM acts to close the window for reprogramming by either physically or molecularly stabilizing neurons.

Here, I explored for the very first time the contribution of extracellular matrix (ECM) molecules to the maintenance of neuronal identity, using PNs of the cerebral cortex as the model system. I first asked whether different classes of PNs produce distinct ECM molecules that help to define the molecular composition of their local extracellular microenvironment. I then asked whether ECM molecules are needed to maintain PNs through the lifespan of the organism, by testing whether ECM degradation around specific neuronal classes causes a shift in identity and/or increases the capacity of projection neurons to respond to reprogramming signals and acquire identity-defining traits of another neuronal class.

Materials and Methods

Mice

All procedures were in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) of Harvard University and performed in accordance with institutional and federal guidelines.

CD-1 mice. Timed-pregnant mice were obtained from Charles River Laboratories (Wilmington, MA).

Cux2-CreERT2. Mice were described previously in Franco et al., 2012. 4-OHT was administered at embryonic day 17.5 (E17.5) at 1 mg/10 g body weight, and brains were harvested at various postnatal time points.

Tle4-2A-CreERT2. Mice were newly generated in collaboration with Z. Josh Huang's laboratory; it will be described elsewhere. 4-OHT was administered at embryonic day 17.5 (E17.5) at 1 mg/10 g body weight, and brains were harvested at various postnatal time points.

Day of vaginal plus was designated as E0.5, and day of birth was designated postnatal day 0 (P0). All mice were maintained in standard housing conditions on a 12-h light/dark cycle with food and water *ad libitum*.

Immunocytochemistry

Brains were processed and stained using standard methods, as previously described (Arlotta et al., 2005). Primary antibodies and dilutions used were as follows: biotinylated WFA lectin, 1:100 (Vector Lab); mouse anti-chondroitin sulfate Delta Di-4S (2B6) supernatant, 1:20 (AMSBIO); rabbit anti-Cux1 (Santa Cruz). Goat anti-rabbit IgG Alexa Fluor 488, 546 and 647 (Life Technologies A11070, A-11071, A21246), streptavidin 546 (Thermo Fisher Scientific S11225),

goat anti–mouse IgG Alexa Fluor 488, 546, 647 (Life Technologies A11017, A11018, A21237). Secondary antibodies were diluted 1:750. Sections were imaged with a Nikon 90i fluorescence microscope equipped with a Retiga Exi camera (Q-Imaging) and analyzed with Volocity image analysis software v4.0.1 (Improvision).

In Utero *electroporation*

The DF-GFP, DF-Fezf2, and CAG-ERT2CreERT2 constructs have been previously described (Rouaux and Arlotta, 2013). The DF-ChABC construct was generated by subcloning the ChABC construct (Bartus et al., 2014) between the T2A sequence and the inverted *LoxP* sites of the DF-Fezf2 construct. *In utero* electroporation was performed as previously described (Molyneaux et al., 2005). Briefly, 0.8 µl of endonuclease-free purified DNA (2 µg/µl) in sterile PBS mixed with 0.005% Fast Green was injected into the lateral ventricle of embryonic day 14.5 CD-1 mice by visual guidance. Five 35 volt pulses of 50 ms duration at 1 s intervals were delivered outside the uterus in appropriate orientation using 1-cm-diameter platinum electrodes and a CUY21EDIT square-wave electroporator (Nepa Gene). P3 pups received a single dose of 0.5 mg of 4-OHT intraperitoneally.

FACS purification

Electroporated animals were anesthetized, and their sensorimotor cortices were microdissected and enzymatically digested into a single-cell suspension using a modified protocol from the Worthington Papain Dissociation System kit (Worthington Biochemical) (Velasco, et al, 2019). A fluorescence dissecting microscope was used to microdissect GFP positive, electroporated sensorimotor sections. Digested tissues were filtered with a 40 micron strainer before DAPI was added to stain dead cells. GFP positive and DAPI negative cells were sorted on a Beckman Coulter MoFlo XDP Cell sorter with a 100 micron nozzle. Cells were sorted directly into Trizol LS

(Thermo Fisher Scientific) for bulk RNA-sequencing. Cells were spun down and frozen for future downstream RNA extraction and library preparation.

RNA-seq analysis

RNA was recovered from Trizol LS using their published protocol (Thermo Fisher Scientific). 1 ng of RNA was used to generate cDNA libraries with the SMART-Seq v4 Ultra Low Input RNA kit, which then served as input for the standard Nextera XT DNA library preparation kit (Illumina). Libraries from different samples were pooled based on molar concentrations and sequenced on a HiSeq 2500 instrument (Illumina) with 101 base paired-end reads. Reads were clipped and mapped onto the mouse genome (mm9, UCSC) using STAR (Dobin et al., 2013). The resulting matrix of read counts from the STAR alignment were analyzed for differential expression by DESeq2 (Love et al., 2014). Cutoff for non-expressed genes were defined as the lowest five percent of genes expressed in transcripts per million, and genes that did not meet this cutoff in all samples were discarded. Cutoff for differential expression was set at FDR<0.05.

Results

Transcriptome analysis of distinct cortical neuronal subtypes reveals subtype-specific ECM gene expression

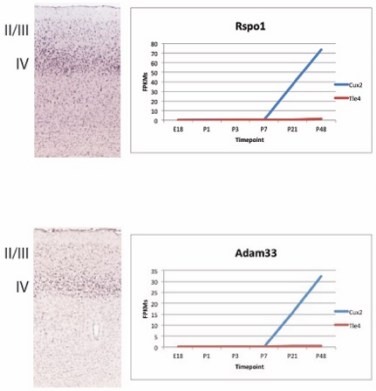
A postdoctoral fellow in the lab had recently generated a RNA-seq dataset from purified upper layer callosal projection neurons (CPNs) from Cux2-CreERT2; Ai14 mice and corticothalamic projection neurons (CThPNs) from Tle4-CreERT2; Ai14 mice. He, in collaboration with a bioinformatician, analyzed these data, and I took their dataset and looked for differentially expressed ECM molecules between the two subtypes. Instead of using GO terms, I used a curated “matrisome” genelist identified through known extracellular domains by the Hynes group

at MIT (Naba et al., 2012). This list includes 1. the core molecules (fibronectins, collagens, etc.) and proteins that share similar biochemical properties, 2. molecules that regulate the core ECM proteins (such as remodeling enzymes), and 3. secreted factors that can reside within the ECM. From this dataset, I started my analysis by comparing equivalent timepoints, with a cutoff of 5 fragments per kilobase per million and 2-fold enrichment. Even at P48, we can still see shared expression of broad families of genes between Cux2+ and Tle4+ neurons, such as the Igfbp superfamily, Semaphorin family, and R-spondin family (Supplementary Figure A4.1). Furthermore, I noticed that axon guidance molecules are continuously differentially expressed throughout postnatal development.

A.

Gene	Cux2 P48 (FPKMs)	Tle4 P48 (FPKMs)
Ptn	458.923	70.2829
Pamr1	180.434	87.1374
S100a10	180.367	43.3913
Igfbp6	92.3709	41.9509
Nell1	87.9474	16.5363
Rspo1	73.5341	1.29738
Coch	60.0976	0.787574
S100a6	40.8663	2.21163
Spock3	39.2044	18.6604
Adam33	32.3168	0.55891
Lgi3	28.2513	12.2785
Igfbp7	25.241	11.5708
Frzb	24.4539	0.800092
Adamtsl2	24.0659	7.53847
Scube1	22.0876	5.78294
Bmp1	21.9484	2.23594
Smoc2	21.0263	0.504699
Megf9	20.8755	8.90031
Bdnf	20.7265	2.44716
Anxa11	20.5988	5.59953
Fgf9	19.7623	7.20946
C1ql3	19.0619	1.2001
Cbln4	15.8862	0.553379
Il12a	14.945	1.5294
Thbs3	14.8328	1.8298
Igfbp5	14.7026	5.88446
Ntng1	13.413	0.772829
Sema3a	13.3183	5.52849
Wnt9a	13.1773	5.42728
Plxnd1	12.4617	0.742397
Ntn5	12.0477	3.25801
Crhbp	11.9593	0.188428
Plxdc2	11.3651	3.69489
Lamc2	10.3881	2.54024
Sema6a	8.8253	4.23765
Wnt4	8.35385	0.663313
Hbegf	8.28644	1.57004
Slit2	7.92502	1.71833
Mfap2	7.81536	2.54717
Inhba	7.4156	0.458958
Sema3c	6.49215	0.341831
Kitl	6.00316	2.4255
Masp1	5.17104	2.31589

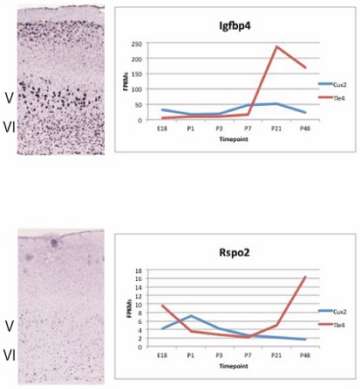
B.



C.

Gene name	Cux2 P48 (FPKMs)	Tle4 P48 (FPKMs)
Serpini1	205.937	480.896
Mmp17	80.2221	193.783
Igfbp4	22.7933	170.579
Htra1	25.4906	97.7404
Sulf2	25.2356	94.7438
Mfge8	10.5569	82.2155
Ctsz	11.2914	73.9638
Spon1	3.58618	68.2841
Sema4a	22.5825	65.9324
Ctgf	1.76002	54.1469
Prss12	2.85056	50.5489
Gpc1	6.5422	36.8628
Bcan	6.58991	30.955
Slit1	5.85468	28.3365
Col6a1	3.66403	27.4617
Plxna2	7.31475	26.4137
Sulf1	0.196593	22.9975
Il17d	9.91854	21.1927
Col12a1	1.31224	18.5321
Nov	8.02042	16.7144
Rspo2	1.69832	16.2999
Wnt7b	3.9025	16.0301
Sema5a	0.626926	15.4809
Igfbp2	2.9058	14.5396
Sema5b	0.656113	13.0168
P4ha2	4.32583	11.8989
Matn2	3.99615	11.499
Fam20a	0.928457	11.3705
Sema3e	1.12654	10.9888
Wisp1	2.96005	10.5438
Timp4	2.00932	8.23577
Vwc2l	0.948486	7.36347
Cxcl12	1.83151	5.96929
Igfbp3	1.45822	5.8741
Gpc5	2.74909	5.71159
Npnt	1.61943	5.39403

D.



SUPPLEMENTARY FIGURE A4.1. DIFFERENTIAL EXPRESSION OF MATRISOME GENES IS MAINTAINED BETWEEN NEURONAL SUBTYPES IN LATE POSTNATAL ANIMALS

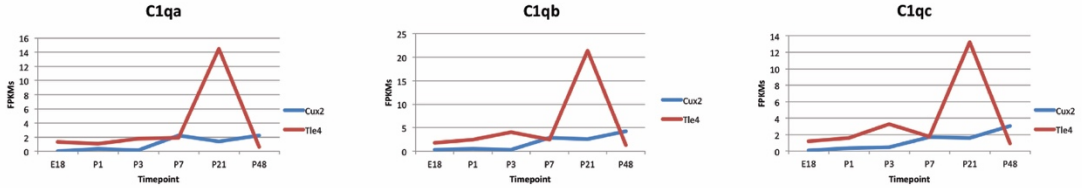
A. Genes upregulated in Cux2+ neurons as compared to Tle4+ neurons at P48. ***B.*** Selected matrisome genes enriched in Cux2+ neurons with their postnatal expression profiles and their layer specificity by *in situ* hybridization (ISH) at P58 from Allen Brain Atlas. ***C.*** Genes upregulated in Tle4+ neurons as compared to Cux2+ neurons at P48. ***D.*** Selected matrisome genes enriched in Tle4+ neurons with their postnatal expression profiles and their layer specificity by ISH at P58 from Allen Brain Atlas.

Intriguingly, I noticed that at P21, Tle4+ CThPN neurons seem to upregulate all three C1q subunit transcripts (Supplementary Figure A4.2A-B). The classical Complement cascade is known to be important for synaptic elimination by microglia in the central nervous system (Schafer et al., 2012; Stevens et al., 2007), and we wondered if it is possible that different classes of neurons use different Complement or Complement-like molecules to differentially regulate synaptic pruning. To this end, I generated a list of domains used in the classical Complement cascade (C1q, anaphylatoxin-like, NTR, Sushi/CCP/SCR, and MACPF domains) and the

A.

Gene name	Cux2 E18	Tle4 E18	Cux2 P1	Tle4 P1	Cux2 P3	Tle4 P3	Cux2 P7	Tle4 P7	Cux2 P21	Tle4 P21	Cux2 P48	Tle4 P48	Shared Complement Domain or Function
Mfge8	3.88448	15.4679	4.02878	17.9678	2.62955	10.9631	2.59298	7.78614	6.20025	85.2893	10.5569	82.2155	in microglia phagocytosis pathway
Sez6l	12.7953	11.6783	19.6002	13.7844	21.6185	11.9845	28.1206	19.5609	30.3761	52.2101	15.7993	45.7286	Sushi/CCP/SCR
Bcan	1.72104	5.50872	2.09532	8.20679	2.56703	9.13107	1.72709	6.5772	7.05757	45.9563	6.58991	30.955	Sushi/CCP/SCR
C1qb	0.340929	1.82398	0.502969	2.49211	0.344508	4.01284	2.91118	2.44363	2.55477	21.3246	4.27975	1.28419	C1q
C1qa	0	1.35202	0.377024	1.03313	0.157733	1.75135	2.26605	1.90927	1.42251	14.4626	2.23376	0.615066	C1q
C1qc	0.0814732	1.22151	0.368643	1.61502	0.459069	3.27013	1.72665	1.73479	1.59883	13.2522	3.04526	0.923202	C1q
Timp4	0.490484	0.434554	0.465842	0.560622	1.03988	0.96566	1.45445	1.19572	2.07663	10.0847	2.00932	8.23577	NTR

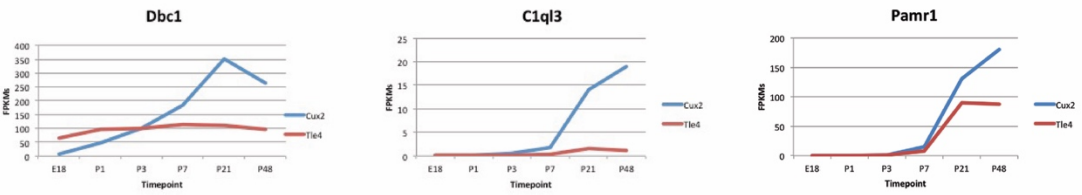
B.



C.

Gene name	Cux2 E18	Tle4 E18	Cux2 P1	Tle4 P1	Cux2 P3	Tle4 P3	Cux2 P7	Tle4 P7	Cux2 P21	Tle4 P21	Cux2 P48	Tle4 P48	Shared Complement Domain or Function
C1qtnf4	60.5004	127.956	81.8505	157.426	109.152	251.027	178.077	270.444	394.036	288.874	325.273	235.775	C1q
Dbc1	7.17212	64.4977	48.1689	98.1473	98.5379	101.142	185.466	114.537	351.646	112.095	265.748	96.3475	MACPF
Pamr1	0.0412673	0.198293	0.277408	0.362779	0.788386	1.11719	15.5007	7.92733	131.041	90.0531	180.434	87.1374	Sushi/CCP/SCR
Susd4	10.934	28.6783	16.4029	31.5638	16.8975	34.2005	21.3798	34.5249	110.412	58.5749	87.681	62.73	Sushi/CCP/SCR
Fam5b	27.9368	19.086	14.7031	24.0084	15.2918	26.4621	38.5633	33.9681	99.7917	45.9754	68.7275	41.1424	MACPF
C1ql3	0.0915113	0.151099	0.14852	0.14619	0.596744	0.123917	1.82817	0.348682	14.0907	1.55892	19.0619	1.2001	C1q
Cbln4	0.20353	0.0971758	0.776864	0.0889647	2.39234	0.953915	3.06092	3.18393	13.5501	0.389677	15.8862	0.553379	C1q
Csmd1	3.2628	0.873068	3.70161	1.38728	5.03413	1.40533	5.35129	2.39378	10.9512	9.59263	10.6213	8.75883	Sushi/CCP/SCR
C1ra	0	0.0384005	0	0.00916312	0.0339225	0.00873337	2.71625	0.093396	9.62147	0.413346	17.1814	1.44109	Sushi/CCP/SCR
Ntn5	0	0.0337933	0	0	0	0.0466566	0.0334261	0.0139525	7.45954	1.55314	12.0477	3.25801	NTR

D.



SUPPLEMENTARY FIGURE A4.2. COMPLEMENT FAMILY AND MOLECULES WITH SHARED DOMAINS ARE DIFFERENTIALLY EXPRESSED BETWEEN NEURONAL SUBTYPES

A. Complement family-related genes with an expression profile in fragments per kilobase of transcript per million (FPKMs) that peak or plateau at P21 and are upregulated in Tle4+ neurons as compared to Cux2+ neurons at P48. ***B.*** Selected Complement family-related genes enriched in Tle4+ neurons with their postnatal expression profiles. ***C.*** Complement family-related genes with an expression profile (in FPKMs) that peak or plateau at P21 and are upregulated in Tle4+ neurons as compared to Cux2+

Supplementary Figure A4.2 (continued) neurons at P48. ***D.*** Selected Complement family-related genes enriched in Cux2+ neurons with their postnatal expression profiles.

molecules that contain these domains to examine their expression over time between neuronal subtypes. In addition, through literature, I added genes known to be important in the microglia-neuron phagocytosis pathway (Brown and Neher, 2014). Several have differential expression, but more importantly, have similar expression profiles that also have a dramatic increase in expression at P21, which suggests a possible functional role of differentially regulating synaptic elimination by microglia between neuronal subtypes. This observation could open up a new line of investigation, in parallel to understanding whether the ECM contributes to the maintenance of class-specific neuronal identity.

I also analyzed the expression of several CPN and CFuPN specific ECM molecules in a single cell qPCR dataset of *Fezf2*-overexpressing CPNs after reprogramming into CFuPNs (Ye et al., 2015). I found that several CFuPN-specific ECM molecules are indeed upregulated (and CPN-specific ECM transcripts downregulated), suggesting that ECM molecules are differentially regulated when the class-specific identity of the PN changes (Supplementary Table A4.1). Together, these data suggest that specific subtypes of PNs do create and modulate their own, unique ECM content.

Gene Name	Marker Type	Cluster 1 Average expression	Cluster 2 Average expression	Fold Change, Cluster 2/Cluster 1	Anova pValue, Cluster 1 vs Cluster 2
Crim1	CFuPN	4.541	7.669	1.690	**0.04787**
Ephb1	CFuPN	2.320	4.538	1.955	**0.003059**
Galntl6	CFuPN	1.093	2.944	2.695	**3.02E-10**
Ntng1	CFuPN	1.230	2.549	2.072	**5.11E-06**
Parm1	CFuPN	1.512	4.050	2.679	**2.18E-06**
Fn1	CPN	1.046	1.548	1.480	**0.004144**
Frem2	CPN	1.000	1.063	1.063	**0.03986**

Mdga1	CPN	20.97	8.610	0.4107	**0.02001**
Ptn	CPN	96.80	66.49	0.6865	**0.004222**
Pvrl3	CPN	16.94	4.698	0.2774	**9.00E-06**

SUPPLEMENTARY TABLE A4.1. SINGLE CELL QPCR ANALYSIS OF SELECTED ECM AND ECM-RELATED GENES

Several ECM and ECM-related genes are differentially expressed upon *Fezf2*-induced reprogramming of CPNs to CFuPNs. Clusters were determined by gene expression using partitioning-based clustering on a panel of marker genes. Cluster 1 contains 54/57 control GFP-overexpressing CPNs and 33/70 *Fezf2*-overexpressing non-reprogrammed CPNs; cluster 2 contains 4 GFP-overexpressing CPNs and 37/70 *Fezf2*-overexpressing induced CFuPNs. Average gene expression is represented in arbitrary units.

RNA-sequencing reveals transcriptional changes upon ECM degradation and Fezf2 *overexpression*

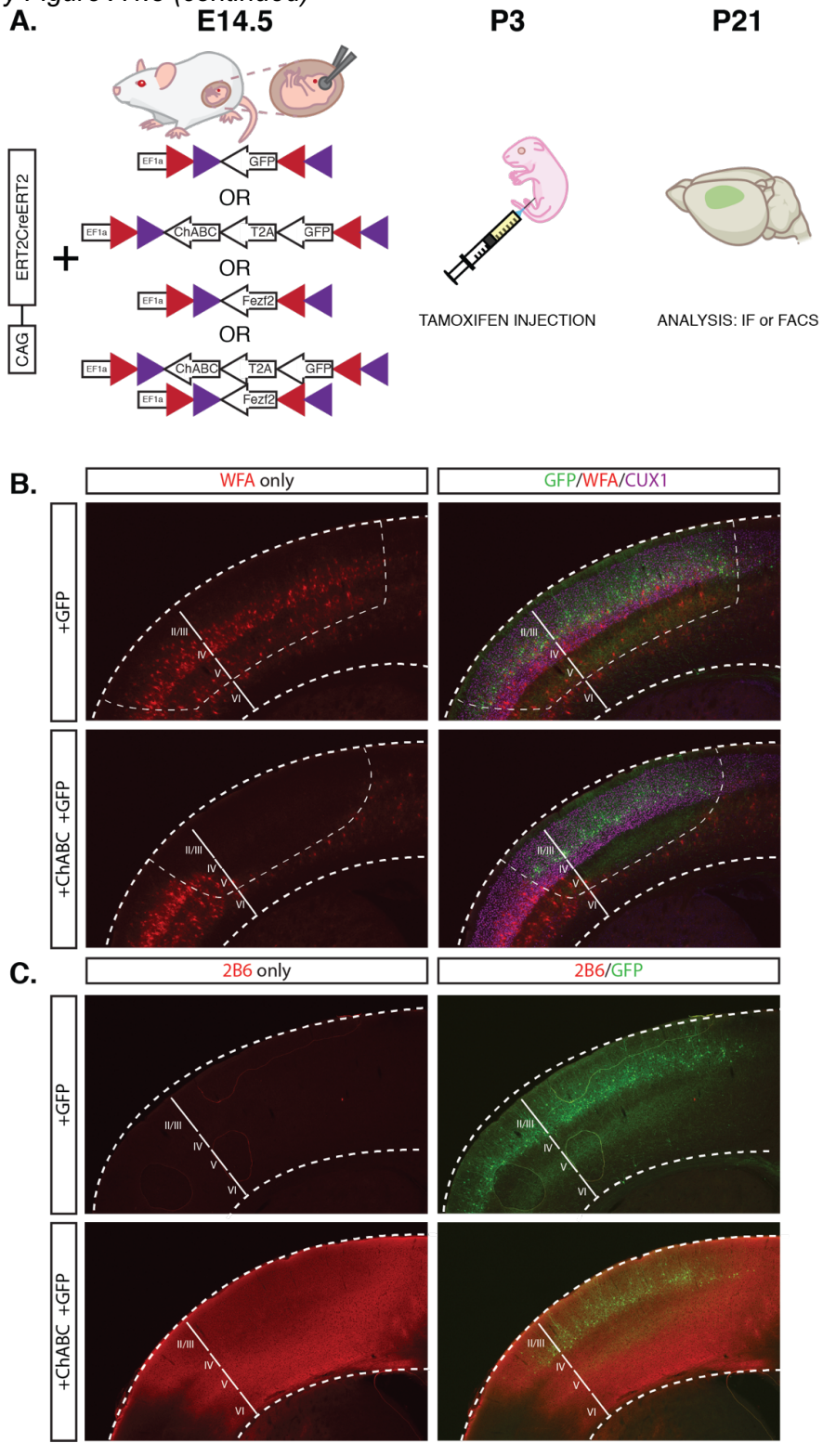
To determine if the ECM is involved in the maintenance of class-specific neuronal identity, I wanted to degrade the ECM surrounding specific populations of neurons and see if their molecular identity changes. First, I established a method to degrade the ECM surrounding CPNs. I used ChABC, a secreted bacterial enzyme that has been proven to be effective at degrading chondroitin sulfate proteoglycans (CSPGs), a component of neural ECM (Bartus et al., 2014; Pizzorusso, et al, 2002). To temporally control the expression ChABC, it was cloned into a Cre-dependent expression plasmid along with a GFP reporter (DF-ChABC). When delivered with CAG-ERT2CreERT2, ChABC can be overexpressed upon 4-hydroxytamoxifen (4OHT) injection (Supplementary Figure A4.3A). ECM degradation was confirmed on fixed sections with wisteria floribunda agglutinin (WFA) staining; WFA stains perineuronal nets, a condensed form of CSPGs, which are degraded upon ChABC secretion (Supplementary Figure A4.3B). Similarly, 2B6, which labels cleaved CSPGs, is seen only with ChABC induction (Supplementary Figure A4.3C).

With the efficacy of ECM degradation established, I characterized the transcriptional changes using bulk RNA-seq. DF-ChABC and CAG-ERT2CreERT2 were delivered by IUE at E14.5 to reach layer II/III CPNs. At P3, I injected 4-OHT to induce ChABC expression.

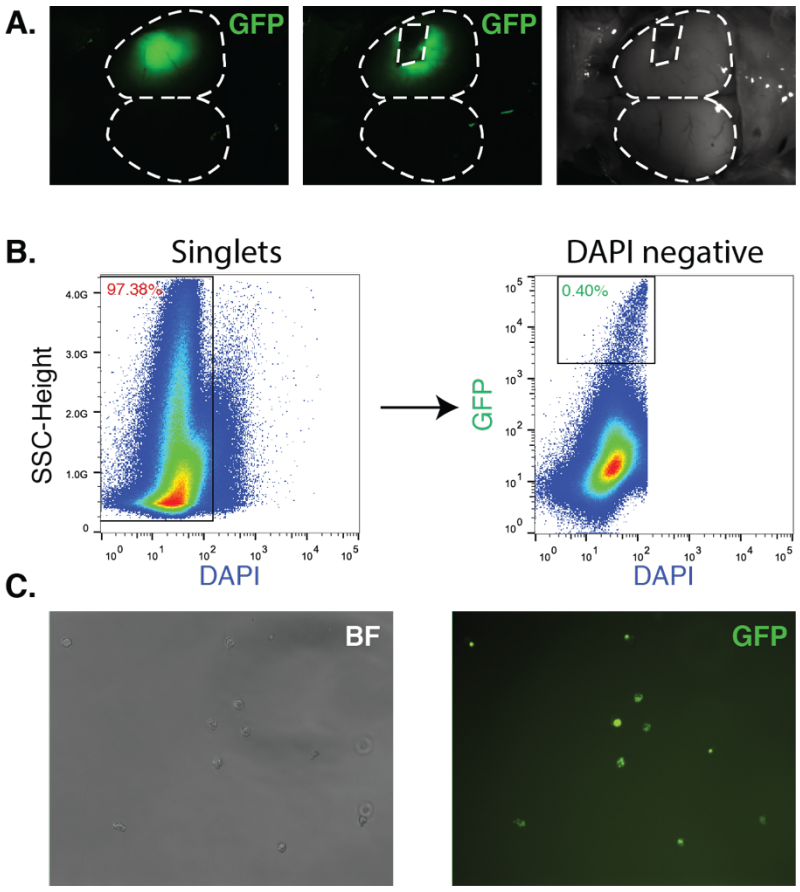
SUPPLEMENTARY FIGURE A4.3. CHABC OVEREXPRESSION PRODUCES LONG-LASTING CPSG DIGESTION

A. Experimental timeline and design. ***B.*** IUE at E14.5 delivers either GFP control or inducible ChABC to upper layer CPNs, as shown by CUX1 staining. As compared to control electroporations, neurons *Supplementary Figure A4.3 (continued)* that overexpress ChABC are surrounded by fewer perineuronal nets, as seen by Wisteria Floribunda Agglutinin (WFA) staining. ***C.*** Neurons that overexpress ChABC are instead surrounded by digested CSPG epitopes, as identified by 2B6 staining.

Supplementary Figure A4.3 (continued)



Electroporated cortices were dissected and dissociated at P21, and electroporated neurons were sorted by FACS and collected for bulk RNA-sequencing (Supplementary Figure A4.3A, Supplementary Figure A4.4). I sequenced 2-4 replicates per condition, and we see that samples that have received Fezf2 largely separate from controls (Supplementary Figure A4.5A). In total, there are 1409 differentially expressed genes across four conditions. Between CPNs that received *ChABC* and *Fezf2* versus *Fezf2* only, only 363 genes are significantly upregulated with Fezf2 only, versus 42 upregulated with *ChABC* and *Fezf2*.



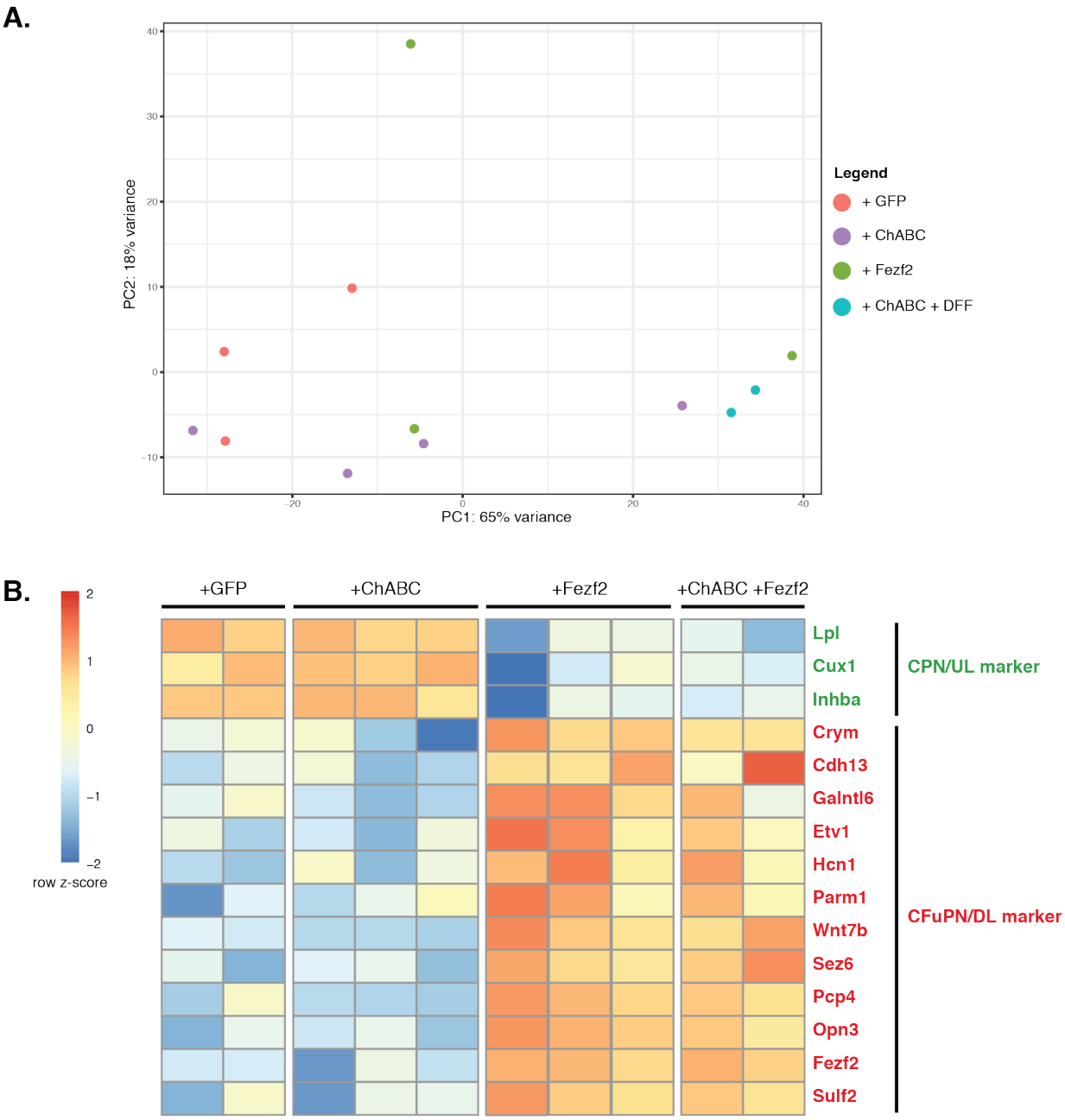
SUPPLEMENTARY FIGURE A4.4. FACS PURIFICATION OF ELECTROPORATED NEURONS ISOLATES A PURE POPULATION OF GFP+ NEURONS

A. GFP+ cortices are dissected from the somatosensory cortex and dissociated into a single cell suspension at P21. ***B.*** Representative FACS plot of sorted, electroporated neurons. Only DAPI negative, GFP+ neurons are collected into Trizol for subsequent RNA extraction, library preparation, and RNA-seq. ***C.*** A small number of neurons are sorted back into media to assess the gating and quality of collected events.

Early post-mitotic *Fezf2* overexpression in CPNs is able to reprogram approximately 42% of CPNs into CFuPNs, while the early postnatal *Fezf2* overexpression can reprogram fewer than 10% (Rouaux and Arlotta, 2013; Ye et al., 2015). We hypothesized the digestion of CSPGs by ChABC may facilitate the early postnatal reprogramming efficiency, and therefore looked at cortical neuronal subtype genes known to change in prior reprogramming studies. Unfortunately, cortical neuronal subtype genes known to change in prior reprogramming studies do not significantly differ between standard early postnatal *Fezf2* overexpression or with the addition of both *Fezf2* and *ChABC* (Rouaux and Arlotta, 2013; Ye et al., 2015; Supplementary Figure A4.5B), suggesting that at least CSPGs do not restrict the nuclear plasticity of cortical neurons.

SUPPLEMENTARY FIGURE A4.5. BULK RNA-SEQUENCING REVEALS TRANSCRIPTIONAL CHANGES UPON *FEZF2* OVEREXPRESSION

A. Principal component analysis (PCA) of the transcriptional data indicate the samples separate along PC1. Graph generated using DeSeq2. ***B.*** Heat map of selected CPN/upper layer neuron (UL) or *Supplementary Figure A4.5 (continued)* CFuPN/deep layer neuron (DL) markers. Using the likelihood ratio test, we find 1409 genes signficantly differentially expressed in at least one condition. When comparing specifically between *+ChABC +Fezf2* versus *+Fezf2* only, we see 363 genes as significantly upregulated with *Fezf2* only, versus 42 upregulated with *ChABC* and *Fezf2*.


A.
PC2: 18% variance
PC1: 65% variance
Legend
+ GFP
+ ChABC
+ Fezf2
+ ChABC + DFF
B.
row z-score
+GFP
+ChABC
+Fezf2
+ChABC +Fezf2
Lpl
Cux1
Inhba
Crym
Cdh13
Galntl6
Etv1
Hcn1
Parm1
Wnt7b
Sez6
Pcp4
Opn3
Fezf2
Sulf2
CPN/UL marker
CFuPN/DL marker

Discussion

The ECM is a dynamic, critical component of the extracellular space implicated in many aspects of neuronal development and function. The work here showed for the first time, how subtypes of cortical PNs differentially contribute to the molecular composition of their local extracellular environment. I also started to explore the question of whether the ECM also controls aspects of maintenance of neuronal class-specific identity.

ECM generation has primarily been thought to be the role of glial and other "supporting" cells. However, recent work has shown that these supporting cells are crucial to proper circuit development in both responding to and secreting essential ECM molecules (Schafer et al., 2012; Stevens et al., 2007). How much neurons themselves contribute to their extracellular environment was largely ignored. We find here that neurons do secrete ECM molecules. Furthermore, there is differential expression across time and subtypes that may correlate with developmental needs. We have only looked at the tip of the iceberg, as we have only looked at transcriptional data. Future work should be dedicated toward the proteomic structure of the ECM around neuronal subtypes, such as their post-translational modifications and localization, which can help elucidate the function of these differentially expressed genes. These data will be pivotal to systematically dissect which pathways PNs use to differentially interact with their environment, thus shedding light on which molecules can be used to redirect cortical assembly in a diseased state.

We find here that CSPGs, a large component of neural ECM, is not necessary to maintain class-specific identity, as CSPG digestion does not further Fezf2's ability to reprogram post-mitotic neurons. We have only examined this using bulk RNA-sequencing, and it is possible that the digestion could marginally improve reprogramming capacity in a few cells. We also not rule out other cell non-autonomous mechanisms for maintaining neuronal identity. As described previously, cortical ECM is incredibly diverse in terms of contribution and composition, and each neuron may contain its own micro niche of extracellular factors and interactions. We speculate

that there may be other hidden signals yet to be discovered in this milieu which may work with cell autonomous mechanisms to maintain neuronal identity. Finally, we also hypothesize that CSPG digestion may affect the local microcircuit. Considering its importance in synaptic connectivity, CSPGs may inhibit the formation of functional microcircuits upon neuronal reprogramming, which would hinder the newly reprogrammed neuron's integration into the new circuit. We did not explore how CSPG digestion affected the local CPN microcircuit, but it would be of interest to then open the possibility of repairing tissue and circuits upon CNS injury.

References

Alcamo, E.A., Chirivella, L., Dautzenberg, M., Dobreva, G., Farinas, I., Grosschedl, R., and McConnell, S.K. (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. Neuron *57*, 364-377.

Alexander, D., Melitta, S., and Peter, S. (2010). The dual role of the extracellular matrix in synaptic plasticity and homeostasis. Nature Reviews Neuroscience *11*, 735-746.

Amamoto, R., Zuccaro, E., Curry, N.C., Khurana, S., Chen, H.H., Cepko, C.L., and Arlotta, P. (2020). FIN-Seq: transcriptional profiling of specific cell types from frozen archived tissue of the human central nervous system. Nucleic Acids Res *48*, e4.

Arlotta, P., Molyneaux, B.J., Chen, J., Inoue, J., Kominami, R., and Macklis, J.D. (2005). Neuronal Subtype-Specific Genes that Control Corticospinal Motor Neuron Development In Vivo. Neuron *45*.

Barros, C.S., Franco, S.J., and Müller, U. (2011). Extracellular matrix: functions in the nervous system. Cold Spring Harbor perspectives in biology *3*.

Bartus, K., James, N.D., Didangelos, A., Bosch, K.D., Verhaagen, J., Yáñez-Muñoz, R.J., Rogers, J.H., Schneider, B.L., Muir, E.M., and Bradbury, E.J. (2014). Large-scale chondroitin sulfate proteoglycan digestion with chondroitinase gene therapy leads to reduced pathology and modulates macrophage phenotype following spinal cord contusion injury. The Journal of neuroscience : the official journal of the Society for Neuroscience *34*, 4822-4836.

Bedogni, F., Hodge, R.D., Elsen, G.E., Nelson, B.R., Daza, R.A., Beyer, R.P., Bammler, T.K., Rubenstein, J.L., and Hevner, R.F. (2010). Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. Proc Natl Acad Sci U S A *107*, 13129-13134.

Bernard-Trifilo, J.A., Kramár, E.A.A., Torp, R., Lin, C.-Y.Y., Pineda, E.A., Lynch, G., and Gall, C.M. (2005). Integrin signaling cascades are operational in adult hippocampal synapses and modulate NMDA receptor physiology. Journal of neurochemistry *93*, 834-849.

Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. Nature reviews Molecular cell biology *15*, 786-801.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525-527.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D.*, et al.* (2008). Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron *57*, 378-392.

Brown, G.C., and Neher, J.J. (2014). Microglial phagocytosis of live neurons. Nature reviews Neuroscience *15*, 209-216.

Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res *44*, D1251-1257.

Chen, B., Schaevitz, L.R., and McConnell, S.K. (2005a). Fezl regulates the differentiation and axon targeting of layer 5 subcortical projection neurons in cerebral cortex. Proceedings of the National Academy of Sciences of the United States of America *102*, 17184-17189.

Chen, B., Wang, S.S., Hattox, A.M., Rayburn, H., Nelson, S.B., and McConnell, S.K. (2008). The Fezf2-Ctip2 genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex. Proc Natl Acad Sci U S A *105*, 11382-11387.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Vaz Meirelles, G., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*.

Chen, J.-G., Rašin, M.-R., Kwan, K.Y., and Šestan, N. (2005b). Zfp312 is required for subcortical axonal projections and dendritic morphology of deep-layer pyramidal neurons of the cerebral cortex. Proceedings of the National Academy of Sciences of the United States of America *102*, 17792-17797.

Clare, A.J., Wicky, H.E., Empson, R.M., and Hughes, S.M. (2017). RNA-Sequencing Analysis Reveals a Regulatory Role for Transcription Factor Fezf2 in the Mature Motor Cortex. Front Mol Neurosci *10*, 283.

De la Rossa, A., Bellone, C., Golding, B., Vitali, I., Moss, J., Toni, N., Lüscher, C., and Jabaudon, D. (2013). In vivo reprogramming of circuit connectivity in postmitotic neocortical neurons. Nature neuroscience *16*, 193-200.

Deneris, E.S., and Hobert, O. (2014). Maintenance of postmitotic neuronal cell identity. Nature neuroscience *17*, 899-907.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
Duggan, A., Ma, C., and Chalfie, M. (1998). Regulation of touch receptor differentiation by the Caenorhabditis elegans mec-3 and unc-86 genes. Development *125*, 4107-4119.

Eckler, M.J., Larkin, K.A., McKenna, W.L., Katzman, S., Guo, C., Roque, R., Visel, A., Rubenstein, J.L., and Chen, B. (2014). Multiple conserved regulatory domains promote Fezf2 expression in the developing cerebral cortex. Neural Development *9*.

Fazel Darbandi, S., Robinson Schwartz, S.E., Qi, Q., Catta-Preta, R., Pai, E.L., Mandell, J.D., Everitt, A., Rubin, A., Krasnoff, R.A., Katzman, S.*, et al.* (2018). Neonatal Tbr1 Dosage Controls Cortical Layer 6 Connectivity. Neuron *100*, 831-845 e837.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M.*, et al.* (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol *16*, 278.

Florio, M., and Huttner, W.B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. Development *141*, 2182-2194.

Foo, L.C., Allen, N.J., Bushong, E.A., Ventura, P.B., Chung, W.S., Zhou, L., Cahoy, J.D., Daneman, R., Zong, H., Ellisman, M.H.*, et al.* (2011). Development of a method for the purification and culture of rodent astrocytes. Neuron *71*, 799-811.

Franco, S.J., Gil-Sanz, C., Martinez-Garay, I., Espinosa, A., Harkins-Perry, S.R., Ramos, C., and Muller, U. (2012). Fate-restricted neural progenitors in the mammalian cerebral cortex. Science *337*, 746-749.

García-Alías, G., Barkhuysen, S., Buckle, M., and Fawcett, J.W. (2009). Chondroitinase ABC treatment opens a window of opportunity for task-specific rehabilitation. Nature neuroscience *12*, 1145-1151.

Gorski, J.A., Talley, T., Qiu, M., Puelles, L., Rubenstein, J.L., and Jones, K.R. (2002). Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. J Neurosci *22*, 6309-6314.

Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H., and Macklis, J.D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci *14*, 755-769.

Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K.*, et al.* (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods *14*, 955-958.

Harb, K., Magrinelli, E., Nicolas, C.S., Lukianets, N., Frangeul, L., Pietri, M., Sun, T., Sandoz, G., Grammont, F., Jabaudon, D.*, et al.* (2016). Area-specific development of distinct projection neuron subclasses is regulated by postnatal epigenetic modifications. Elife *5*, e09531.

Harris, J., Tomassy, G.S., and Arlotta, P. (2015). Building blocks of the cerebral cortex: from development to the dish. Wiley Interdiscip Rev Dev Biol *4*, 529-544.

Harris, K.D., and Shepherd, G.M. (2015). The neocortical circuit: themes and variations. Nature neuroscience *18*, 170-181.

Hensch, T.K. (2005). Critical period plasticity in local cortical circuits. Nature reviews Neuroscience *6*, 877-888.

Hevner, R.F., Shi, L., Justice, N., Hsueh, Y.-P., Sheng, M., Smiga, S., Bulfone, A., Goffinet, A.M., Campagnoni, A.T., and Rubenstein, J.L. (2001). Tbr1 Regulates Differentiation of the Preplate and Layer 6. Neuron *29*, 353-366.

Hirata, T., Suda, Y., Nakao, K., Narimatsu, M., Hirano, T., and Hibi, M. (2004). Zinc finger genefez-like functions in the formation of subplate neurons and thalamocortical axons. Developmental Dynamics *230*, 546-556.

Hobert, O. (2008). Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. Proc Natl Acad Sci U S A *105*, 200067-220071.

Hynes, R.O. (2009). The extracellular matrix: not just pretty fibrils. Science (New York, NY) *326*, 1216-1219.

Kadkhodaei, B., Alvarsson, A., Schintu, N., Ramskold, D., Volakakis, N., Joodmardi, E., Yoshitake, T., Kehr, J., Decressac, M., Bjorklund, A.*, et al.* (2013). Transcription factor Nurr1 maintains fiber integrity and nuclear-encoded mitochondrial gene expression in dopamine neurons. Proceedings of the National Academy of Sciences *110*, 2360-2365.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A.*, et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research *44*.

Levy, A.D., Omar, M.H., and Koleske, A.J. (2014). Extracellular matrix control of dendritic spine and synapse structure and plasticity in adulthood. Frontiers in neuroanatomy *8*, 116.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*.

Lodato, S., and Arlotta, P. (2015). Generating Neuronal Diversity in the Mammalian Cerebral Cortex. Annual review of cell and developmental biology.

Lodato, S., Molyneaux, B.J., Zuccaro, E., Goff, L.A., Chen, H.-H.H., Yuan, W., Meleski, A., Takahashi, E., Mahony, S., Rinn, J.L.*, et al.* (2014). Gene co-regulation by Fezf2 selects neurotransmitter identity and connectivity of corticospinal neurons. Nature neuroscience *17*, 1046-1054.

Lodato, S., Rouaux, C., Quast, K.B., Jantrachotechatchawan, C., Studer, M., Hensch, T.K., and Arlotta, P. (2011). Excitatory projection neuron subtypes control the distribution of local inhibitory interneurons in the cerebral cortex. Neuron *69*, 763-779.

Lois, C., and Alvarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. Science *264*, 1145-1148.

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

Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18-36.

Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg, A., Niino, Y., Egolf, L., Monetti, C.*, et al.* (2015). Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. Neuron *85*, 942-958.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R.*, et al.* (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci *13*, 133-140.

Mangale, V.S., Hirokawa, K.E., Satyaki, P.R., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A.S., Martynoga, B., Paul, J., Mai, M.V.*, et al.* (2008). Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. Science *319*, 304-309.

Mason, A., and Larkman, A. (1990). Correlations Between Morphology and Electrophysiology of Pyramidal Neurons in Slices of Rat Visual Cortex. Journal of Neuroscience *10*, 1415-1428.

Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R.*, et al.* (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron *86*, 1369-1384.

Molyneaux, B.J., Arlotta, P., Fame, R.M., MacDonald, J.L., MacQuarrie, K.L., and Macklis, J.D. (2009). Novel subtype-specific genes identify distinct subpopulations of callosal projection neurons. J Neurosci *29*, 12343-12354.

Molyneaux, B.J., Arlotta, P., Hirata, T., Hibi, M., and Macklis, J.D. (2005). Fezl Is Required for the Birth and Specification of Corticospinal Motor Neurons. Neuron *47*.

Molyneaux, B.J., Goff, L.A., Brettler, A.C., Chen, H.-H.H., Brown, J.R., Hrvatin, S., Rinn, J.L., and Arlotta, P. (2015). DeCoN: Genome-wide Analysis of In Vivo Transcriptional Dynamics during Pyramidal Neuron Fate Selection in Neocortex. Neuron *85*, 275-288.

Muralidharan, B., Khatri, Z., Maheshwari, U., Gupta, R., Roy, B., Pradhan, S.J., Karmodiya, K., Padmanabhan, H., Shetty, A.S., Balaji, C.*, et al.* (2017). LHX2 Interacts with the NuRD Complex and Regulates Cortical Neuron Subtype Determinants Fezf2 and Sox11. J Neurosci *37*, 194-203.

Naba, A., Clauser, K.R., Hoersch, S., Liu, H., Carr, S.A., and Hynes, R.O. (2012). The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Molecular & cellular proteomics : MCP *11*.

Ninkovic, J., Pinto, L., Petricca, S., Lepier, A., Sun, J., Rieger, M.A., Schroeder, T., Cvekl, A., Favor, J., and Gotz, M. (2010). The transcription factor Pax6 regulates survival of dopaminergic olfactory bulb neurons via crystallin alphaA. Neuron *68*, 682-694.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K.*, et al.* (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112-123.

Pizzorusso, T., Medini, P., Berardi, N., Chierzi, S., Fawcett, J.W., and Maffei, L. (2002). Reactivation of ocular dominance plasticity in the adult visual cortex. Science (New York, NY) *298*, 1248-1251.

Pouchelon, G., Gambino, F., Bellone, C., Telley, L., Vitali, I., Lüscher, C., Holtmaat, A., and Jabaudon, D. (2014). Modality-specific thalamocortical inputs instruct the identity of postsynaptic L4 neurons. Nature *511*, 471-474.

Rahe, D.P., and Hobert, O. (2019). Restriction of Cellular Plasticity of Differentiated Cells Mediated by Chromatin Modifiers, Transcription Factors and Protein Kinases. G3: Genes|Genomes|Genetics *9*, 2287-2302.

Ralf, S.S., and Anton, E.S. (2003). Role of Integrins in the Development of the Cerebral Cortex. Cerebral Cortex *13*, 219-224.

Rhee, H.S., Closser, M., Guo, Y., Bashkirova, E.V., Tan, G.C., Gifford, D.K., and Wichterle, H. (2016). Expression of Terminal Effector Genes in Mammalian Neurons Is Maintained by a Dynamic Relay of Transient Enhancers. Neuron *92*, 1252-1265.

Rouaux, C., and Arlotta, P. (2010). Fezf2 directs the differentiation of corticofugal neurons from striatal progenitors in vivo. Nature neuroscience *13*, 1345-1347.

Rouaux, C., and Arlotta, P. (2013). Direct lineage reprogramming of post-mitotic callosal neurons into corticofugal neurons in vivo. Nature cell biology *15*, 214-221.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S.*, et al.* (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell *174*, 1015-1030 e1016.

Schafer, D.P., Lehrman, E.K., Kautzman, A.G., Koyama, R., Mardinly, A.R., Yamasaki, R., Ransohoff, R.M., Greenberg, M.E., Barres, B.A., and Stevens, B. (2012). Microglia Sculpt Postnatal Neural Circuits in an Activity and Complement-Dependent Manner. Neuron *74*, 691-705.

Serrano-Saiz, E., Leyva-Díaz, E., De La Cruz, E., and Hobert, O. (2018). BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice. Current Biology *28*, 2813-2823.e2812.

Shepherd, G.M. (2013). Corticostriatal connectivity and its role in disease. Nat Rev Neurosci *14*, 278-291.

Shetty, A.S., Godbole, G., Maheshwari, U., Padmanabhan, H., Chaudhary, R., Muralidharan, B., Hou, P.S., Monuki, E.S., Kuo, H.C., Rema, V.*, et al.* (2013). Lhx2 regulates a cortex-specific mechanism for barrel formation. Proceedings of the National Academy of Sciences *110*, E4913-E4921.

Shim, S., Kwan, K.Y., Li, M., Lefebvre, V., and Sestan, N. (2012). Cis-regulatory control of corticospinal system development and evolution. Nature *486*, 74-79.

Silver, J., and Miller, J.H. (2004). Regeneration beyond the glial scar. Nature reviews Neuroscience *5*, 146-156.

Stevens, B., Allen, N.J., Vazquez, L.E., Howell, G.R., Christopherson, K.S., Nouri, N., Micheva, K.D., Mehalow, A.K., Huberman, A.D., Stafford, B.*, et al.* (2007). The classical complement cascade mediates CNS synapse elimination. Cell *131*, 1164-1178.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Suter, B.A., O'Connor, T., Iyer, V., Petreanu, L.T., Hooks, B.M., Kiritani, T., Svoboda, K., and Shepherd, G.M. (2010). Ephus: multipurpose data acquisition software for neuroscience experiments. Front Neural Circuits *4*, 100.

Szklarczyk, A., Lapinska, J., Rylski, M., McKay, R.D., and Kaczmarek, L. (2002). Matrix metalloproteinase-9 undergoes expression and activation during dendritic remodeling in adult hippocampus. The Journal of neuroscience : the official journal of the Society for Neuroscience *22*, 920-930.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T.*, et al.* (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci *19*, 335-346.

Tervo, D.G., Hwang, B.Y., Viswanathan, S., Gaj, T., Lavzin, M., Ritola, K.D., Lindo, S., Michael, S., Kuleshova, E., Ojala, D.*, et al.* (2016). A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. Neuron *92*, 372-382.

Velasco, S., Paulsen, B., and Arlotta, P. (2019). Highly reproducible human brain organoids recapitulate cerebral cortex cellular diversity. Protocol Exchange.
Wang, C.F., Hsing, H.W., Zhuang, Z.H., Wen, M.H., Chang, W.J., Briz, C.G., Nieto, M., Shyu, B.C., and Chou, S.J. (2017). Lhx2 Expression in Postmitotic Cortical Neurons Initiates Assembly of the Thalamocortical Somatosensory Circuit. Cell Rep *18*, 849-856.

Wester, J.C., Mahadevan, V., Rhodes, C.T., Calvigioni, D., Venkatesh, S., Maric, D., Hunt, S., Yuan, X., Zhang, Y., Petros, T.J.*, et al.* (2019). Neocortical Projection Neurons Instruct Inhibitory Interneuron Circuit Development in a Lineage-Dependent Manner. Neuron *102*, 960-975 e966.

Wyler, S.C., Spencer, W.C., Green, N.H., Rood, B.D., Crawford, L., Craige, C., Gresch, P., McMahon, D.G., Beck, S.G., and Deneris, E. (2016). Pet-1 Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. The Journal of Neuroscience *36*, 1758-1774.

Yamamoto, M., Wada, N., Kitabatake, Y., Watanabe, D., Anzai, M., Yokoyama, M., Teranishi, Y., and Nakanishi, S. (2003). Reversible suppression of glutamatergic neurotransmission of cerebellar granule cells in vivo by genetically manipulated expression of tetanus neurotoxin light chain. The Journal of neuroscience : the official journal of the Society for Neuroscience *23*, 6759-6767.

Ye, Z., Mostajo-Radji, M.A., Brown, J.R., Rouaux, C., Tomassy, G., Hensch, T.K., and Arlotta, P. (2015). Instructing Perisomatic Inhibition by Direct Lineage Reprogramming of Neocortical Projection Neurons. Neuron *88*, 475-483.

Zembrzycki, A., Perez-Garcia, C.G., Wang, C.F., Chou, S.J., and O'Leary, D.D. (2015). Postmitotic regulation of sensory area patterning in the mammalian neocortex by Lhx2. Proc Natl Acad Sci U S A *112*, 6736-6741.

Zhang, F., Bhattacharya, A., Nelson, J.C., Abe, N., Gordon, P., Lloret-Fernandez, C., Maicas, M., Flames, N., Mann, R.S., Colon-Ramos, D.A.*, et al.* (2014). The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. Development *141*, 422-435.

Zhang, G., Titlow, W.B., Biecker, S.M., Stromberg, A.J., and McClintock, T.S. (2016). Lhx2 Determines Odorant Receptor Expression Frequency in Mature Olfactory Sensory Neurons. eneuro *3*.

Zuccotti, A., Le Magueresse, C., Chen, M., Neitz, A., and Monyer, H. (2014). The transcription factor Fezf2 directs the differentiation of neural stem cells in the subventricular zone toward a cortical phenotype. Proc Natl Acad Sci U S A *111*, 10726-10731.