Title
The Evolution of Cell Types Across the Developing Euarchontogliran Brain

Author
Schmitz, Matthew T

Publication Date
2023

Peer reviewed|Thesis/dissertation

The Evolution of Cell Types Across the Developing Euarchontogliran Brain

by
Matthew Schmitz

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

in

Developmental and Stem Cell Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Daniel E. Wagner
683F6055047E46A...

Daniel E. Wagner
Chair

DocuSigned by:
Alex Pollen
DocuSigned by:
Chun Jimmie Ye
755C0380968F429...

Alex Pollen

Chun Jimmie Ye

Committee Members

For my parents Tom and Debra,
for my siblings Natalie and Evan,
for my wife Anna.
All of whom give everything to the ones they love.

Acknowledgments

I would like to thank many people for their generous support and patience throughout the course of my Ph.D. Alas the winds of fate give me little time before I am swept away to the next challenges, but I shall do my best here.

Firstly, thank you to whichever faculty picked my application from the UCSF Tetrad rejection pile to give me the opportunity in the Developmental and Stem Cell Biology program. I have been incredibly fortunate to have stumbled into two wonderful labs for mentorship during my time at UCSF. To my mentor Jimmie Ye, I'm extremely grateful for all the time you spent in getting me started. Thank you for making science and data a joy to discuss, for always being full of ideas, and for cultivating a rich and fun environment that has been an honor to be a part of, even though my human genetics focused projects did not work out. Thanks to George, Gracie, Elizabeth, and Cody, for being so supportive in starting this journey and always being a joy to run into. Thank you to Mincheol Kim for always being there with kind and statistical words. Thanks to Dr Yang Sun for always staying organized for me, and for always being a joy to chat with, and thank you to everyone else in the Ye lab for being insightful, fun and caring.

To the Pollen Lab, I can hardly begin to express how lucky I have been to spend my days with you. Even when I had no lab work to do, it was worth coming across town just to see you all for a few hours. I fear I can't write all you have meant to me, or the tears will prevent me from meeting the deadline for this work! If you read this or think of me, know you can call on me day or night for help, advice or a chat, for you have been a family for me. And to Alex, thank you firstly for your perseverance in building this group. Thank you for helping me voice thoughts when I could not find them in my mind, for putting people first, for caring, for indulging and for all

the Friday afternoon meetings. Thank you for being an example to me in so many ways, and then for providing the guidance and care to help me begin to grow towards that light.

I'm also so lucky for the friends who were here from the first day of my Ph.D. Evan I'm so lucky you moved out to San Francisco to undertake this adventure with you. You have perfected the art of being both sensitive and fun, you never get enough credit for all that you've accomplished and I love you dearly. To Dr. Nick Sanchez, thank you for being my landing pad and a piece of home, for introducing me to all your friends as I got my feet in the city. I regret that I've never been as good a friend to you as you have been to me. To Sam Patchett, thank you for always being there when I popped my head out from whatever rabbit hole I'd been in for weeks. You've been a rock to me through every month of this journey, and it means more to me than I have shown. To Elina Kostyanovskaya, you're the only person I know who matches my love for jokes in science presentations, and for being a steadfast friend to me. Thank you for being an inspiration in giving yourself to do what is right, taking action and for teaching me some things about conflict aversion that I occasionally remember. To Dr. Bryan Marsh and Eliza Skoler, thank you for being a force for good in my life, and I'm so lucky to have been able to build friendships, undertake adventures and do some deep reflections with you. To Stefan Lundgren and Khurshid Iranpur, thanks for being so much fun, for the adventures and silliness, for having a weird and beautiful dog, for calling on me when you needed it, and for being the most loving of friends.

I also have to thank the mentors who shaped me and set me on this course. From the earliest: To Tony Sikorski, thank you for seeing so much more in me than I saw in myself. To Louise Thompson and Chris Demos thank you for putting up with a biology student who only wanted to

get by enough to dream in my spare time. Dr. Curtis Brandt and Aaron Kolb, for taking a chance on a wildcard undergrad, for teaching me scientific rigor and always being willing to explain the wonders of virology. To Dr. Jamie Thomson, Jen Bolin, thank you for giving me the opportunity to work on inspiring problems. To Dr. Ron Stewart and Dr. Chris Barry, thank you for validating my initiative and for taking my under your wing on exciting science when I was just hired to be a technician. I'm not sure if you know, but when my science career hung in the balance, you gave me the opportunity I needed to choose this academic path.

To Anna and my family, every other word that could ever be written in thanks for the life and love you have given me. Suffice it to say I would know nothing without what you have taught me, and would be nowhere without the life you have given me.

I strive to someday give as much as I have been given.

Contributions

The research that comprises this thesis was conducted under the supervision of Dr. Alex Pollen and Dr. Jimmie Ye at the University of California, San Francisco. This work was supported by the NINDS Ruth L. Kirschstein Predoctoral Fellowship (F31) F31NS124333, as well as DP2MH122400-01 (A.A.P.), U01MH114825 (A.A.P., T.J.N.), R01AI136972 (C.J.Y.).

Chapter 2 describes work published in Nature in 2022 with the following authors Matthew T. Schmitz, Kadellyn Sandoval, Christopher P. Chen, Mohammed A. Mostajo-Radji, William W. Seeley, Tomasz J. Nowakowski, Chun Jimmie Ye, Mercedes F. Paredes, and Alex A. Pollen. I performed all the analysis, generated the data with the help of M.A.M.R, conducted all the experiments except for the late human stains which were performed by K.S.

I wrote this thesis with input from Alex Pollen, Dan Wagner and Jimmie Ye.

Abstract

Matthew Schmitz

The Evolution of Cell Types Across the Developing Euarchontogliran Brain

As the brain develops, a vast diversity of neurons and glia are generated and distributed across the landscape of space and time. While in general this process is highly conserved across mammalian species, interspecies differences cascade into major changes in composition, scaling and ultimately the function of the adult brain. By generating single cell RNA sequencing (scRNAseq) data from across the period of rhesus monkey neurogenesis, we were able to characterize the distinct classes of inhibitory neurons in the cerebrum. With this taxonomy, we were able to identify instances of cell type evolution via generation of a novel primate-specific class of striatal interneurons, as well as differing distributions of olfactory bulb sister cell types between human, monkey and mouse. As cell type differences have their roots in developmental gene expression divergence, we next sought to define the mechanisms of gene expression evolution across the entire brain, which are not well characterized. To explore this, we developed a deep-learning based model of cell type evolution using scRNAseq to identify homologous cell types across species, call gene coexpression modules, and detect differential gene expression. We used this model to parcellate three forms of gene expression divergence and found that these changes were highly modular during both development and adulthood. We also show that the genomic context has a significant effect on whether a gene's expression will change in evolution. This work identifies multiple mechanisms of cell type evolution during

embryonic development and moves toward a formalized model for how gene expression evolves in these cell types.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Mammalian brains differ in remarkable ways. Despite shared developmental programs and predictable relationships between brain region sizes, mammalian brains differ in size by 5 orders of magnitude, in shape from smooth to folded, and in connectivity patterns. Our understanding of the developmental origin of these differences and the extent to which these differences involve the formation of novel cell types or the modification of existing cell types remains limited. However, comparative single cell transcriptomics now enables identification of homologous cell types across species and elucidation of developmental programs. In this thesis, I use single cell transcriptomics to provide a taxonomy of initial classes of inhibitory neurons in the euarchontogliran telencephalon, highlighting contrasting developmental mechanisms contributing to the evolution of novel cell types, including the striatum laureatum neuron, a population I identified lining the primate but not rodent striatum. I extend these comparisons of developmental cell types to a census of initial classes across the developing brain and provide a novel computational approach for determining the architecture of gene regulatory divergence in conserved cell types. Finally, I describe future directions with respect to computational analyses, extending Bayesian variational inference approaches across levels of biological organization, and utilizing a collection of cetacean samples that I have collected for exploring extreme brain adaptations and their potential to inform complex disease.

In the introductory chapter, I outline: the evolution and historical context of mammalian brain evolution in recent geological time; the development and structure of the brain; mechanisms of brain structural evolution, mechanisms of cell type evolution, and their intersection; a historical perspective on the development of the study of cell type diversity in the brain; and lastly, the recent advances that have underpinned the research contained in the subsequent chapters of this work.

The Archetypical Vertebrate Brain

The architecture of vertebrate brains, from lampreys to elephants, follows an evolutionary blueprint that has been conserved across more than 500 million years. Each principal subdivision of the brain is found in essentially all vertebrate species. Within clades, specifically in mammals which are the focus of this introduction, we observe that the allometric relationships between different brain regions also remain predictably stable across related species.

Mammal-like reptiles evolved more than 200 million years ago(Sulej et al., 2020), following a split between the sauropsids (birds and reptiles, including dinosaurs) and synapsids (mammals and extinct mammal-like reptilian lineages). This pivotal divergence in the evolutionary tree resulted in characteristic adaptations that can be seen in the structure and function of the respective descendants' brains.

The original reptilian-like mammalian ancestors were likely very small in size, with very small brains to match. However, a higher brain to body size ratio was coincident with the emergence

of mammalian endothermy, and appeared early in the synapsid lineage.(Northcutt, 2002). Along with the expanded brain as a whole, the pallium of the telencephalon expanded relative to the rest of the brain, and the expansion of the pallium of the forebrain is a hallmark of the mammalian brain, just as expansion of the subpallial telencephalon is for the avian brain. (Northcutt, 2002) During this period, our small ancestors also likely transitioned into a nocturnal niche as they coexisted with dinosaurs, leading to what has been called a "nocturnal bottleneck", which caused irreversible changes in our senses (loss of blue light-sensitive photoreceptors, for instance), and greatly influenced the evolutionary trajectory of our other sensory capabilities.(Musser and Arendt, 2017)

Following the extinction of the dinosaurs at the end of the Cretaceous Period, during which nearly all tetrapods larger than 25kg went extinct(Muench et al., 2000), the small number of surviving mammalian ancestors underwent an explosive radiation resulting in the thousands of mammalian species extant today.(dos Reis et al., 2012) The extant mammalian species fall into three broad groups: monotremes, marsupials, and placentals. Placental mammals are further divided into a group branching into afrotheria (elephants, hyraxes, manatees) and xenarthra (anteaters, armadillos), and a group branching into laurasiatheria (carnivores, ungulates, bats) and euarchontoglires (rodents, primates), each of which had diverged well before cretaceous extinction.(Foley et al., 2023) Despite the fact that brain tissue does not fossilize and early mammals had a cartilaginous skull cap that makes endocasts unreliable or impossible, it is clear

that early mammals had a brain with a form quite similar to today's mammals.(Glenn Northcutt and Kaas, 1995) This deeply-conserved principal adult brain structures are equivalent to the original "vertebrate bauplan", and consists of the telencephalic divisions: olfactory bulb, pallium, subpallium, preoptic area; the diencephalic divisions: retina, pituitary, hypothalamus, ventral thalamus, dorsal thalamus, epithalamus, prectectum, tuberculum; the mesencephalic divisions: the tectum and tegmentum; and the rhombencephalic divisions: cerebellum, isthmus and medulla oblongata.

The Development and Evolution of the Brain

The neuromeric model of brain development proposed by Luis Puelles and John Rubenstein in 1993, following their observations of spatiotemporally restricted expression of transcription factors, posits that the neural tube undergoes progressive longitudinal segmentation by transcription factors driven by patterning factors, with secondary gradients generating intra-neuromeric axes (Puelles and Rubenstein, 1993). The general thrust of this model has since been born out by countless gene expression studies(Carlson, 2014; Flames et al., 2007; Nieuwenhuys and Puelles, 2015). Here I shall proceed anteriorly to posteriorly, mostly by neuromere, to summarize brain development and key evolutionary adaptations of regions along the mammalian or primate lineages. The forebrain is the anterior-most vesicle of the neural tube, also called the prosencephalon. Early in embryogenesis, around the gastrulation stage, the primitive ectoderm gives rise to the neural plate under the influence of signals such as Noggin, Chordin, and Follistatin, which antagonize BMP4 signaling to establish the initial neural

epithelium, which folds to create the neural tube (the lateral plate becomes the dorsal, while the medial plate becomes ventral).(Carlson, 2014) The developing prosencephalon subsequently divides into two main components by embryonic day 10 (E10) in mice (6 weeks post conception in humans): the telencephalon and the diencephalon.(Carlson, 2014) This process is guided by signaling gradients formed by molecules like SHH from the underlying notochord, BMPs and WNT from the dorsal ectoderm, and FGFs from the anterior neural ridge.

Development and Evolution of the Telencephalon

In the telencephalon, the dorsal portion transforms into the cerebral cortex, and the ventral part forms the subpallium, which later gives rise to structures like the basal ganglia. This process, which happens between E11.5 to E18.5 in mice or 6 to roughly 24 weeks post conception in humans, is orchestrated in part by transcription factors like EMX2 in the dorsal telencephalon (pallium) and DLX1/2/5/6 genes in the subpallium.(Bishop et al., 2002; Ma et al., 2013; Marín et al., 2000) A key feature of mammals relative to all other tetrapods is the relative expansion of telencephalic pallium which form striking cerebral cortical hemispheres in mammals. The cytoarchitecture of the cortex is a significant mammalian innovation and its novel stereotyped 5-6 layers, earn it the moniker "neocortex".(Striedter, 2004)

To examine the evolutionary ultrastructure of the cerebral cortices, mammalian ancestor lacked a corpus callosum, and had a smaller cortex with fewer cortical areas.(Krubitzer, 1998) The

corpus callosum, a large bundle of axons crossing the dorsal midline of the cortical hemispheres, is specific to placental mammals, and is likely driven by transient cells in the callosal sling which express SEMA3C, attracting axons across the midline to form the corpus callosum rather than anterior or hippocampal commissure.(Niquille et al., 2009) The first ancestors of extant mammals have been predicted by parsimony to have a gyrified cerebral cortex(Lewitus et al., 2014), largely due to the appearance of gyrencephaly in all three major mammalian lineages due to the gyrification of the echidna brain.("Comparative Mammalian Brain Collections," n.d.) However given that gyrencephaly is strongly predicted by brain size rather than clade and no brain below 5g is gyrencephalic("Comparative Mammalian Brain Collections," n.d.; Striedter, 2004), this approach is likely incorrect as stem mammals were too small to have brains of sufficient size, each lineage has undergone independent expansion, and exactly when gyrification appeared likely depends on the brain size of true mammalian ancestors on each mammalian lineage.(Striedter and Northcutt, 2019)

As a whole, the neocortex has expanded disproportionately along the primate lineage, and followed by the striatum, is the most expanded area of the human brain.(Finlay and Darlington, 1995; Stephan and Andy, 1969) As primary and secondary somatosensory areas do not expand as quickly as cortex size, this means that the expanding neocortex provides substrate for the development of functionally new cortical areas(Buckner and Krienen, 2013; Kaas, 2002). Multimodal association areas are generally located between the primary sensory cortical areas.

In line with this property, the expansion of sensory areas from the most basal mammals, monotremes (Krubitzer et al., 1995) relative to primate (Kaas, 2002; Striedter, 2004) shows a clear increase in the definition and number of secondary and tertiary areas. Another driving mechanism besides sheer increased territory for the formation of new cortical areas is differential regional signaling as the cortex expands. Retinoic acid, which is produced by the meninges, astrocytes, and potentially DA neuron axons, has also been shown to have an effect on neuron morphogenesis as RA induces CBLN2, which in turn increases spinogenesis in the prefrontal cortex(doi:10.1038/s41586-021-03952-y, Shibata et al., doi:10.1038/s41586-021-03953-x), and humans have increased retinoic acid in developing PFC and portions of the temporal lobe.

Owing to a superlinear scaling law governing the ratio of white to gray matter, while the neocortex as a whole is the most enlarged structure of the human brain, it is actually the white matter that is the most expanded element in the human brain(Zhang and Sejnowski, 2000) While the cortical gray matter is composed of densely packed cell bodies of neurons and glia, the white matter is made up of glia, myelin, rare neurons and axonal projections running from neurons in one cortical area to those in another. The white matter thus facilitates integration, connectivity and coherence in the function of large brains. Although neurons are sparse in WM, in aggregate, across the large volume, there are thought to be more than 5 million WM neurons in an infant macaque(Kostovic and Rakic, 1980). In particular, there is an increased density of

superficial WM neurons in the WM of the frontal lobe, while deeper WM neurons have an equal distribution across cortical areas.(García-Marín et al., 2010; Meyer et al., 1992) Ischemic injury can trigger inhibitory neurogenesis in adult mice from regions generally associated with INs of the OB. (Li et al., 2010) Preliminary studies have suggested that postnatally born INs generally destined for the OB may stray from the rostral migratory stream (RMS) to migrate into mouse cortical WM after birth (doi:10.1038/ncomms14219, Frazer et al.; doi:10.1371/journal.pone.0025194). In humans, postnatally born WM-INs appear to display migratory dysregulation in schizophrenia(Fung et al., 2011), and WM-IN density is variably increased in areas associated with schizophrenia, reviewed in (Duchatel et al., 2019). The limited characterization of WM-IN molecular identity has been a major hindrance to these studies. In contrast to gray matter, for which extensive cell type taxonomies have been completed (Bakken et al., 2018; Hodge et al., 2018; Sousa et al., 2017; Tasic et al., 2017), the neuronal types of the WM have been overlooked and have mainly been identified by neuropeptides that do not unambiguously identify cell types.(Suárez-Solá et al., 2009) Thus, despite the expansion of WM-IN number in the primate lineage, especially humans (Sedmak and Judaš, 2019) and their association with schizophrenia, we lack solid understanding of the types, proportion and origin of WM-INs, limiting our understanding of the cellular and developmental basis for underlying pathologies, as I will discuss in chapter 2.

The hippocampus is an ancient non-neocortical portion of the pallium, marked by expression of PROX1, with a homolog identified in all jawed vertebrates, though the mammalian hippocampus

contains a dentate gyrus not found in non-mammalian species.(Briscoe and Ragsdale, 2019; Hevner, 2016) It famously plays a major role in episodic memory consolidation and spatial memory, and is affected in the early stages of Alzheimer's disease.(Hippius and Neundörfer, 2003)

The subpallium on the other hand consists of the septum, striatum, pallidum, and preoptic area. The striatum is the second most proportionally enlarged region of the primate brain, and receives massive dopaminergic innervation from the midbrain.(Stephan and Andy, 1969) The connectivity of this region appears to be significantly divergent as well, with mammals specifically showing a patch-matrix organization within the striatum. Importantly, there also exists a primate-specific cortico-basal ganglia-thalamo-cortical connectivity loop(Utter and Basso, 2008) which appears to be affected in many human neurological diseases.(Maia et al., 2008; Maia and Frank, 2011; Silkis, 2001)

Development and Evolution of the Diencephalon

The diencephalon, which differentiates further into the retina, thalamus, epithalamus and hypothalamus, develops slightly more quickly than the telencephalon. Here, the zona limitans intrathalamica (ZLI), a crucial organizing center that forms by E10.5 in mice(Bulfone et al., 1993), secretes SHH, FGFs, and WNTs to demarcate the boundary between the prethalamus and thalamus and guides thalamus and hypothalamus development.(Scholpp and Lumsden, 2010) The development of both the hypothalamus and thalamus are highly anatomically

complex, and the treating the diencephalon or thalamus and hypothalamus as developmental units is an oversimplification, as in reality it is composed of at least five "fundamental morphological units".(García-Cabezas et al., 2023; Nieuwenhuys and Puelles, 2015) Nonetheless, as the diencephalon will not play a major role in later chapters, I shall simplify here. The thalamus is largely composed of glutamatergic neurons which are born and differentiate *in situ*, while GABAergic neurons of the thalamus are mostly derived from the ZLI, anterior pretectal nucleus (GATA3+) and ventral thalamus (GATA3 or ARX+), though may also have telencephalon origins.(Jager et al., 2021; Krienen et al., n.d.; Virolainen et al., 2012)

The thalamus plays a role in routing signals to and from cortical and subcortical structures, and is much larger relative to the rest of the diencephalon in mammals and birds than other tetrapods. In addition, the primate thalamus is more parcellated than in other species, with more reciprocal connection with the cortex especially in the dorsal thalamus, likely facilitating the increased role of the neocortex in most brain functions.(García-Cabezas et al., 2023; Krubitzer, 1998)

The hypothalamus is initially patterned by high levels of NODAL and SHH from the prechordal axial mesoderm(Xie and Dorsky, 2017), with an anterior-posterior WNT gradient set up by WNT antagonists like DKK secreted from the anterior opposing caudal WNT sources.(Newman et al., 2018) Further subdivisions are set up by BMP and WNT caudal, NOGGIN ventral and SHH in the rostral hypothalamus eventually leading to a diverse array of hypothalamic nuclei.(Xie and

Dorsky, 2017) The hypothalamus has many populations of cells with direct links to physiology(Herber et al., 2019) and behavior through direct function and the hypothalamus' role in governing many hormones via the pituitary(Bendesky et al., 2017; López-Gutiérrez et al., 2022), however these are too complex to be covered in detail here.

Development and Evolution of the Mesencephalon

The development of the mesencephalon, or midbrain, is likewise defined by organizers initiating expression of region defining transcription factors, including a repeat appearance by FGF8. The midbrain receives a number of key signals: SHH from the floor plate, FGF8 from the midbrain-hindbrain-bounding isthmic organizer and LMX1B driven WNT1 from the caudal mesencephalon. This leads to transcription factors defining distinct regions, with corepressive GBX2 initially defining hindbrain and OTX2 defining not-hindbrain, while EN1/PAX2 defines midbrain and SIX3 defines forebrain.(Lagutin et al., 2003) Following this, PAX3/7 expression drives expansion and budding of the dorsal midbrain to form the tectum (Carlson, 2014; Nakamura et al., 2008). The ventral midbrain will eventually differentiate into the dopaminergic neurons that innervate the forebrain and other glutamatergic and GABAergic neurons neurons in the red and other midbrain nuclei(Tiklová et al., 2019). The trends governing the unequal scaling of brain regions mean that the striatum and neocortical cell numbers have exploded relative to the number of midbrain neurons in the human brain, vastly increasing the number of synapses each midbrain dopaminergic projection neuron must form. This has been suggested

as an evolutionary force underpinning the human specificity of Parkinson's disease.(Diederich et al., 2019)

The dorsal midbrain (tectum) consists of glutamatergic and some GABAergic neurons that play a role in routing visual (superior colliculus) and auditory (inferior colliculus) sensory information.(Liu et al., 2023) The tectum is an evolutionary ancient structure, and is generally responsible for the reception of primary visual and auditory signals and routing them to regions of the forebrain and hindbrain for further processing. The superior colliculus (optic tectum) is notably less innervated directly by retinal ganglion cells in primates relative to other species, as in primates these cells predominantly project to the LGN of the thalamus.(Perry and Cowey, 1984)

Development and Evolution of the Rhombencephalon

While there is widespread debate about when each of the various subdivisions discussed above evolved, it is clear that the rhombencephalon, or hindbrain, is the oldest(Holland, 2015). The initial patterning of the hindbrain is driven by the reciprocal gradients of FGF8 from the isthmic organizer retinoic acid from the spinal cord-adjacent somitic mesoderm, with retinoic breakdown enzyme CYP26 genes helping to define hindbrain rhombomere segments and induce HOX gene expression.(Bedois et al., 2021; Hernandez et al., 2007; Nakamura et al., 2008) Each rhombomere gives rise to specific structures within the hindbrain. The more anterior subdivision

of the hindbrain, the metencephalon consists of the cerebellum and pons. Rhombomere 1 generates much of the cerebellum with ATOH1 expression at the rhombic lip driving generation of the abundant cerebellar granule cells.(Aldinger et al., 2021) while the others predominantly contribute to the myelencephalon. Interestingly the pontine nuclei, though positioned amidst rhombomere 3 and 4 derive from the distant caudal rhombic lip expressing PHOX2B, spanning rhombomere 6-8(Kratochwil et al., 2017).

The cerebellum accounts for more than half the neurons in the human brain, containing a relatively constant ratio of 3.6 cerebellar neurons per neocortical neuron across mammals(Herculano-Houzel, 2010). In addition, the cerebellum is clearly highly evolved, with the complex dentate nucleus (and its projection to the diencephalon) as well as the greatly expanded cerebellar lateral lobes appearing to be mammalian innovations.(Green and Wingate, 2014; Smaers et al., 2018) The cerebellum receives strong input from the neocortex via the pontine nuclei (especially the prefrontal cortex in humans), which is in itself another significant mammalian innovation, and has implications for the cerebellum's collaboration with the neocortex in mammal's exceptional cognition.(Ramnani et al., 2006; Smaers et al., 2018) A final piece of evidence further supports the signature of increased cortical control over motor function in primates. Similar to the pontine nuclei, the red nucleus of the caudal midbrain is divided into a magnocellular and parvocellular component.(Olivares-Moreno et al., 2021) The magnocellular component generally receives strong cortical projections and feeds information forward to the spinal cord, but in primates is undifferentiated to the point that it has been considered vestigial,

though has been shown to integrate motor information and even compensate in cases of lost corticospinal projection tracts.(Hicks and Onodera, 2012; Jones and Adkins, 2015) On the other hand, the parvocellular component of the red nucleus passes cortical input from the motor cortex to the cerebellum via the ipsilateral lower olive, where the loop is completed by projections from the dentate nucleus.(Olivares-Moreno et al., 2021) Thus the metencephalon associated-circuitry is clearly divergent in both mammals and primates, and points to innovations in the integration of direct cortical involvement in motor function, though not a complete cortical invasion.

The hindbrain also gives rise to various brainstem nuclei which are associated with various cranial nerves collectively forming the myelencephalon. The assignment of these specific structures and functions to each rhombomere is maintained by Eph-ephrin signaling, which ensures the segregation of these nuclei and eventually the correct axon bundling into the cranial nerves(Terriente and Pujades, 2015).These nuclei are diverse and represent the brain's adrenergic, noradrenergic (locus coeruleus and other medullary nuclei) and serotonergic (Raphe nuclei) centers, which appear to be relatively well conserved, though the locus coeruleus appears affected by human-enriched Alzheimer's and Parkinson's diseases.(Sharma et al., 2010) The cochlear nuclei, crucial in integrating binaural auditory information, are derived from the rhombic lip or rhombomere 2-5, with the help of the olivary nuclei (r5-derived).(Lipovsek and Wingate, n.d.; Ryugo and Parks, 2003) Interestingly the cochlear nuclei are highly divergent across tetrapod evolution, with mammals have an additional cochlear

nucleus, and also incorporating sister neurons of cerebellar granule cells dorsal cochlear nucleus have cerebellar granule cells, less in human than other mammals.(Ryugo and Parks, 2003)

General Principles of Brain Evolution

Homology

When comparing brains, homology must be considered at multiple levels, at least those suggested by Striedter and Northcutt: structures, cell types, and genes.(Striedter and Northcutt, 2019) It is often the case that a cell type is specified that does not express all of the same genes as a homolog in another species, though the cell types maintain the same function in a circuit. Likewise, homologous structures need not contain exactly the same cell types, for instance the primate striatum contains an extra type of inhibitory neuron(Krienen et al., 2020), but the striatum of a mouse and that of a primate are clear homologs as they are composed mostly of the same cell types, governed by the same genes, develop from the same neuromere precursors, and largely fulfill the same functions. As such, it often requires a nuanced analysis to make a statement about what in the brain can be said to be homologous. Finally gene-level homology provides a simpler example, as the DNA sequence is the most fundamental representation of an organism, with segments sharing identity between species considered orthologous and segments sharing identity within species considered paralogous. Many segments of the genome can readily be duplicated evolutionary and detected (with the use of long read sequencing), and thus genes are duplicated as well.(Jiang et al., 2007; Mohajeri et al.,

2016) Some of the duplicated genes undergo neofunctionalization to become new genes which are orthologous by descent, but not in function. Without biochemical data for much of the proteome however, functional orthology is often left to a best bioinformatic guess.(Fiddes et al., 2018; Kirilenko et al., 2023) As the evolution of regions and cell types is in many ways a higher dimensional problem, and there is less consensus on how evolution is to be understood, the rest of this section will focus mostly on the evolution of brain structure and cell types.

Allometry

By far the best studied aspect of brain evolution, the allometry of homologous region sizes relative to the size of the whole brain has long been a key focus in comparative neuroanatomy. In analyses by Finlay and Darlington and others, up to 96% of the total variation in the size of brain regions across mammals (olfactory bulb, cortex,hippocampus, striatum, septum, diencephalon, midbrain,cerebellum, and medulla can be predicted from absolute brain size (Finlay and Darlington, 1995; Jerison, 1989), pointing to a large contribution of concerted effects. Thus, it seems clear that concerted changes dwarf mosaic changes on the fine scale of closely related, but different sized species.(Finlay and Darlington, 1995) On the other hand, as allometric scaling tends to shift across clades, for example the expansion of the telencephalon in mammals, or the higher exponent in primate neocortical scaling, it's clear that mosaic evolution plays a key role in the grand trends of evolution.

An intuitive principle that tasks that require more information processing, require larger dedicated brain regions proposed as the "principle of proper mass" by Jersion in 1973.(Jerison, 2012) Interestingly along with this principle, Jerison also introduced the encephalization quotient (EQ), a brain to body mass scaling ratio which is adjusted by the expected brain size for its clade and places humans atop the list of animals with the highest EQ. This debate demands that we find ways to specific or general intelligence across numerous animals to make a fair comparison in order to answer the embedded question "Are larger brains better?".

Assessing intelligence across animals is extremely difficult as it must take into account each animal's perceptual biases, motivations and sensitivities to human intervention (Macphail, 1985). This is complicated further as it has been well known since before Brodmann's work on cortical lesions that different regions of the brain participate in specialized computational tasks.(Brodmann, 1909; Economo and Koskinas, 1925) It follows that across evolution, intelligence comes in many forms and is perhaps best thought of as a coordinated ensemble of many different processing modules. As such, accurate ratings of intelligence across species or even individuals within a species remains a loose and biased affair, and it must be noted that this impulse has justified the most discriminatory and destructive treatment of humans by each other, time and again, reflecting more the bias of the theorist rather than any estimation of reality. As such, I speak of specialization of intelligence in only the broadest terms, and that we

expect that tasks that are computationally intensive and crucial for survival will require increased or improved computational substrate.

It has been observed that the more functionally relevant a region is to crucial and computationally intensive tasks, the larger, late-developing, more laminated, and better connected it tends to be.(DEACON, 1990; Finlay and Darlington, 1995; Striedter, 2004). While the expansion of a region leading to a scaling in computational potential is obvious, the cause of lamination is slightly less obvious, but it is likely due to the multiplicative scaling of the number of distinct targets or processing "channels" as the number of layers of a structure increases.(Striedter, 2004)

It must be pointed out that there is some skepticism that larger brains are more intelligent, and also that expansion may represent an evolutionary "spandrel" given that brain and body size are strongly linked (though intelligence could still be evolved as a byproduct of body expansion in such cases).(Gould et al., 1997; Manger et al., 2021) It has even been argued that the main function of the enlarged cetacean brain is to generate heat in cold water(Manger et al., 2021). Firstly, there is ample evidence in lineages of especially enlarged brains, namely primates, cetaceans and elephants, that these species are capable of cognitive tasks beyond that of other animals (Foerder et al., 2011; Hart et al., 2008; Marino, n.d.; Marino et al., 2007) A second line of reasoning that "bigger must be better" is that brain tissue is metabolically expensive and so has a high evolutionary cost. On the one hand, evidence suggests there is no relationship

between brain size and metabolic rate(McNab and Eisenberg, 1989). On the other hand, it has been proposed humans and other large brained mammals may also have decreased the size and energy expenditure on our gut and musculature, which may be an adaptation to preserve energy for use in the brain, and other species have likely come to similar equilibria by making unseen sacrifices and gaining strong benefits from larger brains.(Aiello and Wheeler, 1995) As for the argument that a brain simply generates heat, I add that the mammalian innovation of brown adipose tissue is a much simpler and less expensive structure to develop to generate heat more efficiently. Thirdly the idea that absolute brain size does not matter to intelligence is ludicrous, while incoming sensory information and to some degree motor control does scale with body size (sublinearly it must be added (Glenn Northcutt and Kaas, 1995), it is difficult to see how placing a larger brain into a smaller body, thus increasing its EQ would make a brain more intelligent respectively. Indeed, as domestic dogs display more intraspecies variation in body size (and thus overall brain size) than any other mammalian species, studies of *canis familiaris* are able to show that larger brained dogs, differing in scale only, show better short term memory and executive function than do smaller conspecifics. (Horschler et al., 2019)

The principle of increasing functional specialization driving enlargement and definition is also true at the regional level. The auditory information conveying inferior colliculus is larger than the superior colliculus in echolocating dolphins, while the inverse is true in vision-dependent primates.(Oelschläger et al., 2010) Independently of dolphins, this is also true in echolocating

bats (Baron et al., 1996). The primate and predator reliance on vision has also likely had significant influence on other brain structures as well. In the thalamus, the lateral geniculate nucleus is not strictly laminated in small brained mammals like lagomorphs but is in larger brained mammals likely independently along three separate lineages (Kaas, 2002; Kahn and Krubitzer, 2002) Likewise, as covered above, the magnocellular compartment of the red nucleus is diminished in primates, likely owing to invasion of the cortico-spinal projections filling part of this region's role.

The expansion and parcellation of the neocortex, covered above (and everywhere else), is another prime example showing both expansion in many intelligent species and increased lamination in mammals. (Cheung et al., 2010; Glenn Northcutt and Kaas, 1995; Pollen et al., 2009) In addition, in the most fascinating cases of reorganized sensory modalities, for instance bat and cetacean echolocation, remain poorly characterized, though it appears that their visuo-auditory cortices are both reorganized and expanded.(Berns Gregory S. et al., 2015; Macias et al., 2020) More recently a number of studies have further accentuated the complexity of this topic, showing that the functional regionalization of the cortex may change across the lifespans of individuals, adding an interesting possibility that the expanded human lifespan also adds a dimension of dynamic specialization in the cortex (Englund and Krubitzer, 2022).

In conclusion, I must acknowledge that characterizing allometry is possible because we are largely able to take the conservation of the "vertebrate brain bauplan" for granted(Nieuwenhuys and Puelles, 2015).

There are many more cases of questionable regional homology across larger time differences, for instance the functional homology of the three mammalian cochlear nuclei compared to the two in sauropsids (Ryugo and Parks, 2003), or the homology of the mammalian neocortex to the reptile or amphibian structures.(Tosches et al., 2018) Within mammals, the primate thalamus displays an increased number of subdivisions. In homologizing these many demarcations, it is necessary to trace connectivity as well as cell type development and identity. Through these considerations, it would appear in many cases, elements of the function of the original whole have been divided amongst its evolutionary subdivisions.(García-Cabezas et al., 2023) The determination of homology of neocortical regions presents similar problems. As discussed above, cortical expansion drives parcellation of association areas, generally between the simply homologous primary sensory areas. As such, it is often unclear whether this represents a continuum between existing areas, addition of new areas, or nonlinear subdivision of old areas.

Evolution of cell types

Understanding evolution in cell types is doubly difficult relative to brain structures, firstly because cell types are far more numerous than the number of regions in the brain, and cells inhabit a semi-continuous state space often conceptualized as Waddington's landscape(Waddington, 1957). Secondly, it is challenging because cell types may differ in many

modalities in both a quantitative and qualitative fashion and it is the cell types differences that drive regional composition and connectivity differences. Thus a cell type can be defined by its function (circuits in the case of neurons, as I will mostly consider), by its transcriptomic state, or as defined by Detlev Arendt et al, a group of cells which changes together in evolution.(Arendt et al., 2016) This final definition is best expressed in deep evolutionary homologies and is perhaps too retrospective to be considered in my focus within euarchontoglires.

As inspiration for discussion of how cell types can be born or change in evolution, I'll build off the 4 basic examples detailed by Fenna Krienen in her exploration of primate interneurons in the adult brain.(Krienen et al., 2020) The first two examples: that conserved types may change in abundance or be allocated across locations or contexts are linked, as the latter may be the cause of the former. Examples of altered abundance include increased ratio of inhibitory neurons to excitatory neurons in the cortex and the increased abundance of upper layer neurons in the human cortex. The third example of cell type evolution is that cells may change the genetic program of a conserved type. As this category is extremely commonplace across species (though will be discussed in more detail in chapter 3), there are thousands of genes with changed expression in many cell types.(Bakken et al., 2021; Hodge et al., 2018) As I have made much ado of the allometry in the brain, it must also be said that the large brain environment also can affect cell types. For instance the giant Betz cells in the motor cortex, which are larger in humans than non-human primates or mice, likely because they must project

farther to make its corticospinal connections.(Bakken et al., 2021) Notably, the fraction of extratelencephalic layer 5 neurons, including corticospinal neurons, has also decreased in expanded primate brains and may scale with target region size rather than with disproportionately expanded cortex.(Hodge et al., 2018) I will also discuss this in the context of the "Reduce and Reuse" hypothesis in Chapter 2. Krienen's final example is the invention of a novel cell type. The case of the TAC3+ striatal interneuron is especially interesting (see chapter 2). As with genes, the existence of paralogs limits the degree to which an evolutionary change will disrupt the function of established characters, and also makes the origin of a novel cell type or gene much more clear (it is not always easy to tell if a cell type has a heavily modified genetic program or is novel).

To expand on this point, it is also important to take a cell type's function into account. While cell types may be homologs by developmental origin, they may be modified in ways that alter or entirely change or ablate a cell's functional role. As such, functional cell type evolution can be qualitative or quantitative. In the case of neurons, different receptivity, different target cells, or neurotransmitter switches represent changes that can alter a neuron's fundamental role in a circuit. This is a qualitative change, and there is a case to be made that a functionally novel cell type has been born by the modification of a single gene, despite clear homology between a type and its homologs in other species. I expect changes of this manner to be relatively rare in evolution, unless the cell type was duplicated and a sister cell type maintains the original role of

this neuron in a circuit while the paralogous cell type gains a new function, as is seen in the MGE_CRABP1/TAC3 and MGE_CRABP1/MAF striatal interneurons (Chapter 2). Given the continuous nature of Waddington's landscape, the production of identical (or near identical) neuron initial classes followed by separation of subpopulations of these cells into different signaling contexts is sufficient to create two or more distinct mature sister cell types, possibly including the striatum laureatum neuron seen in Chapter 2. By the same token, theoretically it is possible that a homologous cell type could arise from a different lineage in one species relative to another, though no specific example of this is evident. Ultimately, understanding the evolution of cell types will require both data on the function of cell types in the brain, as well as the developmental origin and molecular profile of the diversity of cell types in many dimensions.

A short history of the study of cellular diversity in the brain

From Cells and Genes to Brains

Our understanding of the diversity of cells has unfurled from the initial visual observation that life is, in fact, cellular. In the early 1820s microscope technology was greatly improved, allowing the description of cortical cellular layers, delineation of brain nuclei and tracing of axon fibers, however the understanding of the cellular nature of the brain had not yet taken shape. Among the pioneering neuroscientists at this time, Ukrainian Vladimir Betz undertook foundational work in comparative cytoarchitecture, publishing his discovery of the giant pyramidal neurons now known as Betz cells in 1874 and a comparative atlas of the brain in 1879 (Betz, Vladimir, 1874; Kushchayev et al., 2012) Building upon the work of Betz to describe the cytoarchitecture of the

brain, perhaps the most famous example of examination of the comparative cellular diversity of the brain comes in the late 1800s with the staining and drawings of individual neurons by Camillo Golgi and Santiago Ramon y Cajal. Golgi's use of silver nitrate and potassium chromate to stain random neurons black by the microcrystallization of silver chromate proved crucial in both Santiago Ramon y Cajal's description of differing cellular composition in different species and brain areas, as well as discernment of synapses inspiring the "neuron doctrine"–that the nervous system is built of discrete neurons–and the two shared a Nobel Prize in 1906.(Ramón y Cajal, 1981)

Through the 20th century the development and anatomy of the brain was mapped in great detail. The cellular basis of many neurological diseases became clear, with the cellular degeneration driving Alzheimer's disease and Parkinson's disease were described before 1920(Hippius and Neundörfer, 2003; Lees, 2007). At the same time, acetylcholine was the first neurotransmitter molecule to be identified and isolated in work done by Otto Loewi in 1914, and the first half of the 20th century proved extremely productive in revealing the electrophysiological function of the brain.(Valenstein, 2002) Immunohistochemistry, developed in the 1930s, using antibodies raised to target specific protein epitopes in order to to stain cellular proteins made it clear that there was great proteomic diversity underlying the morphological forms of the cell types in the brain. To follow this, as the central dogma that DNA is transcribed to produce RNA which is translated into proteins was elucidated, it became clear

that genes were expressed differently across cell types, and that cells were a key link in translating genotypes to phenotypes. As individual disease-related genes were identified, it became clear that scaling understanding of more complex phenomena would require the whole collection of genes in the genome and corresponding transcriptome (Velculescu et al., 1997) Despite the many advances in neuroimaging, single cell electrophysiological recording, mouse transgenics, neuronal birthdating(Rakic, 1974), it remained unclear how the cell types in the brain arose and just how diverse the types were.

The initial pushes to understand the composition of the brain in the years following the publication of the complete human genome in 2001(Lander et al., 2001) were marked by a massive broadening of the lens with which the transcriptome brain could be examined. It was also around this time that the stem cells giving rise to many of the neurons in the cortex were identified to the radial glia, and many in situ hybridization studies revealed the transcription factor logic of brain progenitor regionalization(Bulfone et al., 1993; Gould et al., 1999; Noctor et al., 2004, 2001; Wonders and Anderson, 2006)

With these new and largely accurate models of how brain development and neurogenesis unfolded, it was possible to begin to make sense of atlases of gene expression as initial efforts used microarrays and RNAseq to identify regional gene signatures in the developing brain(Kang et al., 2011; Miller et al., 2014). A key shortcoming of these approaches was evident: that bulk 'omics profiling did not reveal the cellular composition of the brain that was known to underlie

the data. Gene correlation networks driven by cell type composition provided some degree of cell type signal, and yielded cell-type specific differences across species in cortical GABAergic neurons. (Oldham et al., 2008; Raju et al., 2018) However, without the ability to assign transcriptomes to single cell types, it was unclear just how many cell types there were in the developing adult brain.

The Single Cell 'Omics Revolution

The first transcriptome of a single cell was achieved in 2009 using cells from the 4-cell stage of mouse embryos.(Tang et al., 2009) Despite this significant accomplished, sequencing each of the trillions cells in a body, would take roughly until the heat death of the universe to complete if done at this scale. Among the first technologies to achieve scalable capture of individual cells was the SMART-seq, in which cDNA libraries could be generated from single cells (Ramsköld et al., 2012). Often this approach was paired with fluorescence activated cell sorting to achieve one cell per well in a more scalable way.

Scalable single cell RNA-sequencing proved to be decisive in surpassing the problem of decomposing cell type signals in RNA-seq data, and ultimately made it possible at last to survey the diversity of cells at a molecular level in an unbiased fashion.The industrialization of these techniques, for instance by Fluidigm Corporation, was key to early studies to determine novel markers of rare cell populations in heterogeneous tissue, like outer radial glia(Pollen et al., 2015). Scalability was further improved with the invention of droplet based single cell

sequencing, in which cells in droplets of water are partitioned in oil, forming an emulsion that allows the sequencing of hundreds of thousands of cells at a time.(Zilionis et al., 2017)

This key innovation has opened the door to countless variants, multiple data modalities and the generation of massive datasets including millions of cells. It is this droplet-based RNA-sequencing, specifically the platform commercialized by 10X Genomics, that I have used to undertake the meta-cell type-atlases that are a major contribution of the discoveries and resources made in the subsequent chapters.

References

Aiello LC, Wheeler P. 1995. The Expensive-Tissue Hypothesis: The Brain and the Digestive System in Human and Primate Evolution. *Curr Anthropol* **36**:199–221. doi:10.1086/204350

Aldinger KA, Thomson Z, Phelps IG, Haldipur P, Deng M, Timms AE, Hirano M, Santpere G, Roco C, Rosenberg AB, Lorente-Galdos B, Gulden FO, O'Day D, Overman LM, Lisgo SN, Alexandre P, Sestan N, Doherty D, Dobyns WB, Seelig G, Glass IA, Millen KJ. 2021. Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci* **24**:1163–1175. doi:10.1038/s41593-021-00872-y

Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, Wagner GP. 2016. The origin and evolution of cell types. *Nat Rev Genet* **17**:744–757. doi:10.1038/nrg.2016.127

Bakken TE, Hodge RD, Miller JM, Yao Z, Nguyen TN, Aevermann B, Barkan E, Bertagnolli D, Casper T, Dee N, Garren E, Goldy J, Gray LT, Kroll M, Lasken RS, Lathia K, Parry S,

Rimorin C, Scheuermann RH, Schork NJ, Shehata SI, Tieu M, Phillips JW, Bernard A, Smith KA, Zeng H, Lein ES, Tasic B. 2018. Equivalent high-resolution identification of neuronal cell types with single-nucleus and single-cell RNA-sequencing. doi:10.1101/239749

Bakken TE, Jorstad NL, Hu Q, Lake BB, Tian W, Kalmbach BE, Crow M, Hodge RD, Krienen FM, Sorensen SA, Eggermont J, Yao Z, Aevermann BD, Aldridge AI, Bartlett A, Bertagnolli D, Casper T, Castanon RG, Crichton K, Daigle TL, Dalley R, Dee N, Dembrow N, Diep D, Ding S-L, Dong W, Fang R, Fischer S, Goldman M, Goldy J, Graybuck LT, Herb BR, Hou X, Kancherla J, Kroll M, Lathia K, van Lew B, Li YE, Liu CS, Liu H, Lucero JD, Mahurkar A, McMillen D, Miller JA, Moussa M, Nery JR, Nicovich PR, Niu S-Y, Orvis J, Osteen JK, Owen S, Palmer CR, Pham T, Plongthongkum N, Poirion O, Reed NM, Rimorin C, Rivkin A, Romanow WJ, Sedeño-Cortés AE, Siletti K, Somasundaram S, Sulc J, Tieu M, Torkelson A, Tung H, Wang X, Xie F, Yanny AM, Zhang R, Ament SA, Behrens MM, Bravo HC, Chun J, Dobin A, Gillis J, Hertzano R, Hof PR, Höllt T, Horwitz GD, Keene CD, Kharchenko PV, Ko AL, Lelieveldt BP, Luo C, Mukamel EA, Pinto-Duarte A, Preissl S, Regev A, Ren B, Scheuermann RH, Smith K, Spain WJ, White OR, Koch C, Hawrylycz M, Tasic B, Macosko EZ, McCarroll SA, Ting JT, Zeng H, Zhang K, Feng G, Ecker JR, Linnarsson S, Lein ES. 2021. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**:111–119. doi:10.1038/s41586-021-03465-8

Baron G, Stephan H, Frahm HD. 1996. Comparative neurobiology in Chiroptera. Basel ; Boston: Birkhäuser Verlag.

Bedois AMH, Parker HJ, Krumlauf R. 2021. Retinoic Acid Signaling in Vertebrate Hindbrain Segmentation: Evolution and Diversification. *Diversity* **13**:398. doi:10.3390/d13080398

Bendesky A, Kwon Y-M, Lassance J-M, Lewarch CL, Yao S, Peterson BK, He MX, Dulac C, Hoekstra HE. 2017. The genetic basis of parental care evolution in monogamous mice. *Nature* **544**:434–439. doi:10.1038/nature22074

Berns Gregory S., Cook Peter F., Foxley Sean, Jbabdi Saad, Miller Karla L., Marino Lori. 2015. Diffusion tensor imaging of dolphin brains reveals direct auditory pathway to temporal lobe. *Proc R Soc B Biol Sci* **282**:20151203. doi:10.1098/rspb.2015.1203

Betz, Vladimir. 1874. Anatomischer Nachweis zweier Gehirncentra, Centralblatt für die medizinischen Wissenschaften. p. pg. 578-580.

Bishop KM, Rubenstein JLR, O'Leary DDM. 2002. Distinct Actions of Emx1, Emx2, andPax6 in Regulating the Specification of Areas in the Developing Neocortex. *J Neurosci* **22**:7627–7638. doi:10.1523/JNEUROSCI.22-17-07627.2002

Briscoe SD, Ragsdale CW. 2019. Evolution of the Chordate Telencephalon. *Curr Biol* **29**:R647–R662. doi:10.1016/j.cub.2019.05.026

Brodmann K. 1909. Vergleichende Lokalisationslehre der Großhirnrinde : in ihren Prinzipien dargestellt auf Grund des Zellenbaues.

Buckner RL, Krienen FM. 2013. The evolution of distributed association networks in the human brain. *Trends Cogn Sci* **17**:648–665. doi:10.1016/j.tics.2013.09.017

Bulfone A, Puelles L, Porteus M, Frohman M, Martin G, Rubenstein J. 1993. Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt- 3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. *J Neurosci* **13**:3155–3172. doi:10.1523/JNEUROSCI.13-07-03155.1993

Carlson BM. 2014. Early Patterning and Development of the Central Nervous SystemReference Module in Biomedical Sciences. Elsevier. p. B9780128012383054000. doi:10.1016/B978-0-12-801238-3.05440-4

Cheung AFP, Kondo S, Abdel-Mannan O, Chodroff RA, Sirey TM, Bluy LE, Webber N, DeProto J, Karlen SJ, Krubitzer L, Stolp HB, Saunders NR, Molnár Z. 2010. The Subventricular Zone Is the Developmental Milestone of a 6-Layered Neocortex: Comparisons in Metatherian and Eutherian Mammals. *Cereb Cortex* **20**:1071–1081. doi:10.1093/cercor/bhp168

Comparative Mammalian Brain Collections. n.d. https://brainmuseum.org/

DEACON TW. 1990. Rethinking Mammalian Brain Evolution1. *Am Zool* **30**:629–705. doi:10.1093/icb/30.3.629

Diederich NJ, Surmeier DJ, Uchihara T, Grillner S, Goetz CG. 2019. Parkinson's disease: Is it a consequence of human brain evolution? *Mov Disord* **34**:453–459. doi:https://doi.org/10.1002/mds.27628

dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B Biol Sci* **279**:3491–3500. doi:10.1098/rspb.2012.0683

Duchatel RJ, Shannon Weickert C, Tooney PA. 2019. White matter neuron biology and neuropathology in schizophrenia. *Npj Schizophr* **5**:1–9. doi:10.1038/s41537-019-0078-8

Economo C von, Koskinas GN. 1925. Die Cytoarchitektonik der Hirnrinde des erwachsenen Menschen. Berlin: J. Springer.

Englund M, Krubitzer L. 2022. Phenotypic Alterations in Cortical Organization and Connectivity across Different Time Scales. *Brain Behav Evol* **97**:108–120. doi:10.1159/000522131

Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, Haussler D, Stanke M, Paten B. 2018. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res*. doi:10.1101/gr.233460.117

Finlay BL, Darlington RB. 1995. Linked regularities in the development and evolution of mammalian brains. *Science* **268**:1578–1584. doi:10.1126/science.7777856

Flames N, Pla R, Gelman DM, Rubenstein JLR, Puelles L, Marín O. 2007. Delineation of Multiple Subpallial Progenitor Domains by the Combinatorial Expression of Transcriptional Codes. *J Neurosci* **27**:9682–9695. doi:10.1523/JNEUROSCI.2750-07.2007

Foerder P, Galloway M, Barthel T, Iii DEM, Reiss D. 2011. Insightful Problem Solving in an Asian Elephant. *PLOS ONE* **6**:e23251. doi:10.1371/journal.pone.0023251

Foley NM, Mason VC, Harris AJ, Bredemeyer KR, Damas J, Lewin HA, Eizirik E, Gatesy J, Karlsson EK, Lindblad-Toh K, Zoonomia Consortium, Springer MS, Murphy WJ. 2023. A genomic timescale for placental mammal evolution. *Science* **380**:eabl8189. doi:10.1126/science.abl8189

Fung SJ, Joshi D, Allen KM, Sivagnanasundaram S, Rothmond DA, Saunders R, Noble PL, Webster MJ, Weickert CS. 2011. Developmental Patterns of Doublecortin Expression and White Matter Neuron Density in the Postnatal Primate Prefrontal Cortex and Schizophrenia. *PLOS ONE* **6**:e25194. doi:10.1371/journal.pone.0025194

García-Cabezas MÁ, Pérez-Santos I, Cavada C. 2023. Mapping the primate thalamus: historical perspective and modern approaches for defining nuclei. *Brain Struct Funct* **228**:1125–1151. doi:10.1007/s00429-022-02598-4

García-Marín V, Blazquez-Llorca L, Rodriguez JR, Gonzalez-Soriano J, DeFelipe J. 2010. Differential distribution of neurons in the gyral white matter of the human cerebral cortex. *J Comp Neurol* **518**:4740–4759. doi:10.1002/cne.22485

Glenn Northcutt R, Kaas JH. 1995. The emergence and evolution of mammalian neocortex. *Trends Neurosci* **18**:373–379. doi:10.1016/0166-2236(95)93932-N

Gould E, Reeves AJ, Graziano MSA, Gross CG. 1999. Neurogenesis in the Neocortex of Adult Primates. *Science* **286**:548–552. doi:10.1126/science.286.5439.548

Gould SJ, Lewontin RC, Maynard Smith J, Holliday R. 1997. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* **205**:581–598. doi:10.1098/rspb.1979.0086

Green MJ, Wingate RJ. 2014. Developmental origins of diversity in cerebellar output nuclei. *Neural Develop* **9**:1. doi:10.1186/1749-8104-9-1

Hart BL, Hart LA, Pinter-Wollman N. 2008. Large brains and cognition: Where do elephants fit in? *Neurosci Biobehav Rev* **32**:86–98. doi:10.1016/j.neubiorev.2007.05.012

Herber CB, Krause WC, Wang L, Bayrer JR, Li A, Schmitz M, Fields A, Ford B, Zhang Z, Reid MS, Nomura DK, Nissenson RA, Correa SM, Ingraham HA. 2019. Estrogen signaling in arcuate Kiss1 neurons suppresses a sex-dependent female circuit promoting dense strong bones. *Nat Commun* **10**:163. doi:10.1038/s41467-018-08046-4

Herculano-Houzel S. 2010. Coordinated Scaling of Cortical and Cerebellar Numbers of Neurons. *Front Neuroanat* **4**:12. doi:10.3389/fnana.2010.00012

Hernandez RE, Putzke AP, Myers JP, Margaretha L, Moens CB. 2007. Cyp26 Enzymes Generate the Retinoic Acid Response Pattern Necessary for Hindbrain Development. *Dev Camb Engl* **134**:177–187. doi:10.1242/dev.02706

Hevner RF. 2016. Evolution of the Mammalian Dentate Gyrus. *J Comp Neurol* **524**:578–594. doi:10.1002/cne.23851

Hicks TP, Onodera S. 2012. The mammalian red nucleus and its role in motor systems, including the emergence of bipedalism and language. *Prog Neurobiol* **96**:165–175. doi:10.1016/j.pneurobio.2011.12.002

Hippius H, Neundörfer G. 2003. The discovery of Alzheimer's disease. *Dialogues Clin Neurosci* **5**:101–108. doi:10.31887/DCNS.2003.5.1/hhippius

Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, Close JL, Long B, Penn O, Yao Z, Eggermont J, Hollt T, Levi BP, Shehata SI, Aevermann B, Beller A, Bertagnolli D, Brouner K, Casper T, Cobbs C, Dalley R, Dee N, Ding S-L, Ellenbogen RG, Fong O, Garren E, Goldy J, Gwinn RP, Hirschstein D, Keene CD, Keshk M, Ko AL, Lathia K, Mahfouz A, Maltzer Z, McGraw M, Nguyen TN, Nyhus J, Ojemann JG, Oldre A, Parry S, Reynolds S, Rimorin C, Shapovalova NV, Somasundaram S, Szafer A, Thomsen ER, Tieu M, Scheuermann RH, Yuste R, Sunkin SM, Lelieveldt B, Feng D, Ng L, Bernard A, Hawrylycz M, Phillips J, Tasic B, Zeng H, Jones AR, Koch C, Lein ES. 2018. Conserved cell types with divergent features between human and mouse cortex. doi:10.1101/384826

Holland LZ. 2015. The origin and evolution of chordate nervous systems. *Philos Trans R Soc B Biol Sci* **370**:20150048. doi:10.1098/rstb.2015.0048

Horschler DJ, Hare B, Call J, Kaminski J, Miklósi Á, MacLean EL. 2019. Absolute brain size predicts dog breed differences in executive function. *Anim Cogn* **22**:187–198. doi:10.1007/s10071-018-01234-1

Jager P, Moore G, Calpin P, Durmishi X, Salgarella I, Menage L, Kita Y, Wang Y, Kim DW, Blackshaw S, Schultz SR, Brickley S, Shimogori T, Delogu A. 2021. Dual midbrain and forebrain origins of thalamic inhibitory interneurons. *eLife* **10**:e59272. doi:10.7554/eLife.59272

Jerison H. 2012. Evolution of The Brain and Intelligence. Elsevier.

Jerison HJ. 1989. Brain Size In: Kimura D, editor. Speech and Language, Readings from The. Boston, MA: Birkhäuser. pp. 4–6. doi:10.1007/978-1-4899-6774-9_2

Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, Eichler EE. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat Genet* **39**:1361–1368. doi:10.1038/ng.2007.9

Jones TA, Adkins DL. 2015. Motor System Reorganization After Stroke: Stimulating and Training Toward Perfection. *Physiology* **30**:358–370. doi:10.1152/physiol.00014.2015

Kaas J. 2002. Convergences in the Modular and Areal Organization of the Forebrain of Mammals: Implications for the Reconstruction of Forebrain Evolution. *Brain Behav Evol* **59**:262–272. doi:10.1159/000063563

Kahn DM, Krubitzer L. 2002. Massive cross-modal cortical plasticity and the emergence of a new cortical area in developmentally blind mammals. *Proc Natl Acad Sci* **99**:11429–11434. doi:10.1073/pnas.162342799

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AMM, Pletikos M, Meyer KA, Sedmak G, Guennel T, Shin Y, Johnson MB, Krsnik Z, Mayer S, Fertuzinhos S, Umlauf S, Lisgo SN, Vortmeyer A, Weinberger DR, Mane S, Hyde TM, Huttner A, Reimers M, Kleinman JE, Sestan N. 2011. Spatio-temporal transcriptome of the human brain. *Nature* **478**:483–489. doi:10.1038/nature10523

Kirilenko BM, Munegowda C, Osipova E, Jebb D, Sharma V, Blumer M, Morales AE, Ahmed A-W, Kontopoulos D-G, Hilgers L, Lindblad-Toh K, Karlsson EK, Consortium‡ Z, Hiller M. 2023. Integrating gene annotation with orthology inference at scale. *Science*. doi:10.1126/science.abn3107

Kostovic I, Rakic P. 1980. Cytology and time of origin of interstitial neurons in the white matter in infant and adult human and monkey telencephalon. *J Neurocytol* **9**:219–242. doi:10.1007/BF01205159

Kratochwil CF, Maheshwari U, Rijli FM. 2017. The Long Journey of Pontine Nuclei Neurons: From Rhombic Lip to Cortico-Ponto-Cerebellar Circuitry. *Front Neural Circuits* **11**. doi:10.3389/fncir.2017.00033

Krienen FM, Goldman M, Zhang Q, C. H. del Rosario R, Florio M, Machold R, Saunders A, Levandowski K, Zaniewski H, Schuman B, Wu C, Lutservitz A, Mullally CD, Reed N, Bien E, Bortolin L, Fernandez-Otero M, Lin JD, Wysoker A, Nemesh J, Kulp D, Burns M, Tkachev V, Smith R, Walsh CA, Dimidschstein J, Rudy B, S. Kean L, Berretta S, Fishell G, Feng G, McCarroll SA. 2020. Innovations present in the primate interneuron repertoire. *Nature* 1–8. doi:10.1038/s41586-020-2781-z

Krienen FM, Levandowski KM, Zaniewski H, Schroeder ME, Goldman M, Lutservitz A, Zhang Q, Li KX, Beja-Glasser VF, Sharma J, Shin TW, Mauermann A, Nemesh J, Kashin S, Vergara J, Chelini G, Dimidschstein J, Berretta S, Boyden E, McCarroll SA, Feng G. n.d. A marmoset brain cell census reveals persistent influence of developmental origin on neurons.

Krubitzer L. 1998. What can monotremes tell us about brain evolution? *Philos Trans R Soc Lond B Biol Sci* **353**:1127–1146. doi:10.1098/rstb.1998.0271

Krubitzer L, Manger P, Pettigrew J, Calford M. 1995. Organization of somatosensory cortex in monotremes: In search of the prototypical plan. *J Comp Neurol* **351**:261–306. doi:10.1002/cne.903510206

Kushchayev SV, Moskalenko VF, Wiener PC, Tsymbaliuk VI, Cherkasov VG, Dzyavulska IV, Kovalchuk OI, Sonntag VKH, Spetzler RF, Preul MC. 2012. The discovery of the pyramidal neurons: Vladimir Betz and a new era of neuroscience. *Brain* **135**:285–300. doi:10.1093/brain/awr276

Lagutin OV, Zhu CC, Kobayashi D, Topczewski J, Shimamura K, Puelles L, Russell HRC, McKinnon PJ, Solnica-Krezel L, Oliver G. 2003. Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev* **17**:368–379. doi:10.1101/gad.1059403

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond Christina, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin

S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen H-C, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang S-P, Yeh R-F, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond Christopher, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A, Morgan MJ, International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research C for GR, The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, Department of Genome Analysis I of MB, GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, Multimegabase Sequencing Center TI for SB, Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, Cold Spring Harbor Laboratory LAHGC, GBF—German Research Centre for Biotechnology:, *Genome Analysis Group (listed in alphabetical order also includes individuals listed under other headings):, Scientific management: National

Human Genome Research Institute UNI of H, Stanford Human Genome Center:, University of Washington Genome Center:, Department of Molecular Biology KUS of M, University of Texas Southwestern Medical Center at Dallas:, Office of Science UD of E, The Wellcome Trust: 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921. doi:10.1038/35057062

Lees AJ. 2007. Unresolved issues relating to the Shaking Palsy on the celebration of James Parkinson's 250th birthday. *Mov Disord* **22**:S327–S334. doi:10.1002/mds.21684

Lewitus E, Kelava I, Kalinka AT, Tomancak P, Huttner WB. 2014. An Adaptive Threshold in Mammalian Neocortical Evolution. *PLOS Biol* **12**:e1002000. doi:10.1371/journal.pbio.1002000

Li Y, Yu SP, Mohamad O, Genetta T, Wei L. 2010. Sublethal Transient Global Ischemia Stimulates Migration of Neuroblasts and Neurogenesis in Mice. *Transl Stroke Res* **1**:184–196. doi:10.1007/s12975-010-0016-6

Lipovsek M, Wingate RJ. n.d. Conserved and divergent development of brainstem vestibular and auditory nuclei. *eLife* **7**:e40232. doi:10.7554/eLife.40232

Liu Y, Savier EL, DePiero VJ, Chen C, Schwalbe DC, Abraham-Fan R-J, Chen H, Campbell JN, Cang J. 2023. Mapping visual functions onto molecular cell types in the mouse superior colliculus. *Neuron* **111**:1876-1886.e5. doi:10.1016/j.neuron.2023.03.036

López-Gutiérrez MF, Mejía-Chávez S, Alcauter S, Portillo W. 2022. The neural circuits of monogamous behavior. *Front Neural Circuits* **16**:978344. doi:10.3389/fncir.2022.978344

Ma T, Wang C, Wang L, Zhou X, Tian M, Zhang Q, Zhang Y, Li J, Liu Z, Cai Y, Liu F, You Y, Chen C, Campbell K, Song H, Ma L, Rubenstein JL, Yang Z. 2013. Subcortical origins of human and monkey neocortical interneurons. *Nat Neurosci* **16**:1588–1597. doi:10.1038/nn.3536

Macias S, Bakshi K, Garcia-Rosales F, Hechavarria JC, Smotherman M. 2020. Temporal coding of echo spectral shape in the bat auditory cortex. *PLoS Biol* **18**:e3000831. doi:10.1371/journal.pbio.3000831

Macphail EM. 1985. Vertebrate intelligence: the null hypothesis. *Philos Trans R Soc Lond B Biol Sci* **308**:37–51. doi:10.1098/rstb.1985.0008

Maia TV, Cooney RE, Peterson BS. 2008. The neural bases of obsessive–compulsive disorder in children and adults. *Dev Psychopathol* **20**:1251–1283. doi:10.1017/S0954579408000606

Maia TV, Frank MJ. 2011. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* **14**:154–162. doi:10.1038/nn.2723

Manger PR, Patzke N, Spocter MA, Bhagwandin A, Karlsson KÆ, Bertelsen MF, Alagaili AN, Bennett NC, Mohammed OB, Herculano-Houzel S, Hof PR, Fuxe K. 2021. Amplification of potential thermogenetic mechanisms in cetacean brains compared to artiodactyl brains. *Sci Rep* **11**:5486. doi:10.1038/s41598-021-84762-0

Marín O, Anderson SA, Rubenstein JLR. 2000. Origin and Molecular Specification of Striatal Interneurons. *J Neurosci* **20**:6063–6076. doi:10.1523/JNEUROSCI.20-16-06063.2000

Marino L. n.d. Cetacean Brain Evolution: Multiplication Generates Complexity 16.

Marino L, Connor RC, Fordyce RE, Herman LM, Hof PR, Lefebvre L, Lusseau D, McCowan B, Nimchinsky EA, Pack AA, Rendell L, Reidenberg JS, Reiss D, Uhen MD, Van der Gucht E, Whitehead H. 2007. Cetaceans Have Complex Brains for Complex Cognition. *PLoS Biol* **5**. doi:10.1371/journal.pbio.0050139

McNab BK, Eisenberg JF. 1989. Brain Size and Its Relation to the Rate of Metabolism in Mammals. *Am Nat* **133**:157–167.

Meyer G, Wahle P, Castaneyra-Perdomo A, Ferres-Torres R. 1992. Morphology of neurons in the white matter of the adult human neocortex. *Exp Brain Res* **88**:204–212. doi:10.1007/BF02259143

Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Aiona K, Arnold JM, Bennet C, Bertagnolli D, Brouner K, Butler S, Caldejon S, Carey A, Cuhaciyan C, Dalley RA, Dee N, Dolbeare TA, Facer BAC, Feng D, Fliss TP, Gee G, Goldy J, Gourley L, Gregor BW, Gu G, Howard RE, Jochim JM, Kuan CL, Lau C, Lee C-K, Lee F, Lemon TA, Lesnar P, McMurray B, Mastan N, Mosqueda NF, Naluai-Cecchini T, Ngo N-K, Nyhus J, Oldre A, Olson E, Parente J, Parker PD, Parry SE, Player AS, Pletikos M, Reding M, Royall JJ, Roll K, Sandman D, Sarreal M, Shapouri S, Shapovalova NV, Shen EH, Sjoquist N, Slaughterbeck CR, Smith M, Sodt AJ, Williams D, Zöllei L, Fischl B, Gerstein MB, Geschwind DH, Glass IA, Hawrylycz MJ, Hevner RF, Huang H, Jones AR, Knowles JA, Levitt P, Phillips JW, Sestan N, Wohnoutka P, Dang C, Bernard A, Hohmann JG, Lein ES. 2014. Transcriptional Landscape of the Prenatal Human Brain. *Nature* **508**:199–206. doi:10.1038/nature13185

Mohajeri K, Cantsilieris S, Huddleston J, Nelson BJ, Coe BP, Campbell CD, Baker C, Harshman L, Munson KM, Kronenberg ZN, Kremitzki M, Raja A, Catacchio CR, Graves TA, Wilson RK, Ventura M, Eichler EE. 2016. Interchromosomal core duplicons drive both evolutionary instability and disease susceptibility of the Chromosome 8p23.1 region. *Genome Res* **26**:1453–1467. doi:10.1101/gr.211284.116

Muench D, Muench M, Gilders MA. 2000. Primal forces. Portland, OR: Graphic Arts Center Pub.

Musser JM, Arendt D. 2017. Loss and gain of cone types in vertebrate ciliary photoreceptor evolution. *Dev Biol*, Development and evolution of sensory cells and organs **431**:26–35. doi:10.1016/j.ydbio.2017.08.038

Nakamura H, Sato T, Suzuki-Hirano A. 2008. Isthmus organizer for mesencephalon and metencephalon. *Dev Growth Differ* **50**:S113–S118. doi:10.1111/j.1440-169X.2008.00995.x

Newman EA, Wu D, Taketo MM, Zhang J, Blackshaw S. 2018. Canonical Wnt signaling regulates patterning, differentiation and nucleogenesis in mouse hypothalamus and prethalamus. *Dev Biol* **442**:236–248. doi:10.1016/j.ydbio.2018.07.021

Nieuwenhuys R, Puelles L. 2015. Towards a New Neuromorphology. Springer.

Niquille M, Garel S, Mann F, Hornung J-P, Otsmane B, Chevalley S, Parras C, Guillemot F, Gaspar P, Yanagawa Y, Lebrand C. 2009. Transient Neuronal Populations Are Required to Guide Callosal Axons: A Role for Semaphorin 3C. *PLoS Biol* **7**:e1000230. doi:10.1371/journal.pbio.1000230

Noctor SC, Flint AC, Weissman TA, Dammerman RS, Kriegstein AR. 2001. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**:714. doi:10.1038/35055553

Noctor SC, Martínez-Cerdeño V, Ivic L, Kriegstein AR. 2004. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci* **7**:136. doi:10.1038/nn1172

Northcutt RG. 2002. Understanding Vertebrate Brain Evolution. *Integr Comp Biol* **42**:743–756. doi:10.1093/icb/42.4.743

Oelschläger HHA, Ridgway SH, Knauth M. 2010. Cetacean Brain Evolution: Dwarf Sperm Whale (Kogia sima) and Common Dolphin (Delphinus delphis) – An Investigation with High-Resolution 3D MRI. *Brain Behav Evol* **75**:33–62. doi:10.1159/000293601

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. 2008. Functional organization of the transcriptome in human brain. *Nat Neurosci* **11**:1271–1282. doi:10.1038/nn.2207

Olivares-Moreno R, Rodriguez-Moreno P, Lopez-Virgen V, Macías M, Altamira-Camacho M, Rojas-Piloni G. 2021. Corticospinal vs Rubrospinal Revisited: An Evolutionary Perspective for Sensorimotor Integration. *Front Neurosci* **15**. doi:https://doi.org/10.3389

Perry VH, Cowey A. 1984. Retinal ganglion cells that project to the superior colliculus and pretectum in the macaque monkey. *Neuroscience* **12**:1125–1137. doi:10.1016/0306-4522(84)90007-1

Pollen AA, Cheung AFP, Molnár Z. 2009. Evolution and Embryological Development of the Cortex in Amniotes In: Binder MD, Hirokawa N, Windhorst U, editors. Encyclopedia of Neuroscience. Berlin, Heidelberg: Springer. pp. 1165–1172. doi:10.1007/978-3-540-29678-2_3111

Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, Lim DA, Leyrat AA, West JA, Kriegstein AR. 2015. Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell* **163**:55–67. doi:10.1016/j.cell.2015.09.004

Puelles L, Rubenstein JLR. 1993. Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization. *Trends Neurosci* **16**:472–479. doi:10.1016/0166-2236(93)90080-6

Raju CS, Spatazza J, Stanco A, Larimer P, Sorrells SF, Kelley KW, Nicholas CR, Paredes MF, Lui JH, Hasenstaub AR, Kriegstein AR, Alvarez-Buylla A, Rubenstein JL, Oldham MC. 2018. Secretagogin is Expressed by Developing Neocortical GABAergic Neurons in

Humans but not Mice and Increases Neurite Arbor Size and Complexity. *Cereb Cortex* **28**:1946–1958. doi:10.1093/cercor/bhx101

Rakic P. 1974. Neurons in Rhesus Monkey Visual Cortex: Systematic Relation between Time of Origin and Eventual Disposition. *Science* **183**:425–427. doi:10.1126/science.183.4123.425

Ramnani N, Behrens TEJ, Johansen-Berg H, Richter MC, Pinsk MA, Andersson JLR, Rudebeck P, Ciccarelli O, Richter W, Thompson AJ, Gross CG, Robson MD, Kastner S, Matthews PM. 2006. The Evolution of Prefrontal Inputs to the Cortico-pontine System: Diffusion Imaging Evidence from Macaque Monkeys and Humans. *Cereb Cortex* **16**:811–818. doi:10.1093/cercor/bhj024

Ramón y Cajal S. 1981. Recuerdos de mi vida: historia de mi labor científica, Alianza universidad. Madrid: Alianza Editorial.

Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R. 2012. Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**:777–782. doi:10.1038/nbt.2282

Ryugo DK, Parks TN. 2003. Primary innervation of the avian and mammalian cochlear nucleus. *Brain Res Bull*, Functional Anatomy of Ear Connections **60**:435–456. doi:10.1016/S0361-9230(03)00049-2

Scholpp S, Lumsden A. 2010. Building a bridal chamber: development of the thalamus. *Trends Neurosci* **33**:373–380. doi:10.1016/j.tins.2010.05.003

Sedmak G, Judaš M. 2019. The total number of white matter interstitial neurons in the human brain. *J Anat* **235**:626–636. doi:10.1111/joa.13018

Sharma Y, Xu T, Graf WM, Fobbs A, Sherwood CC, Hof PR, Allman JM, Manaye KF. 2010. Comparative Anatomy of the Locus Coeruleus in Humans and Non-Human Primates. *J Comp Neurol* **518**:963. doi:10.1002/cne.22249

Shibata M, Pattabiraman K, Lorente-Galdos B, Andrijevic D, Kim S-K, Kaur N, Muchnik SK, Xing X, Santpere G, Sousa AMM, Sestan N. 2021a. Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature* **598**:483–488. doi:10.1038/s41586-021-03953-x

Shibata M, Pattabiraman K, Muchnik SK, Kaur N, Morozov YM, Cheng X, Waxman SG, Sestan N. 2021b. Hominini-specific regulation of CBLN2 increases prefrontal spinogenesis. *Nature* **598**:489–494. doi:10.1038/s41586-021-03952-y

Silkis I. 2001. The cortico-basal ganglia-thalamocortical circuit with synaptic plasticity. II. Mechanism of synergistic modulation of thalamic activity via the direct and indirect pathways through the basal ganglia. *Biosystems* **59**:7–14. doi:10.1016/S0303-2647(00)00135-0

Smaers JB, Turner AH, Gómez-Robles A, Sherwood CC. 2018. A cerebellar substrate for cognition evolved multiple times independently in mammals. *eLife* **7**:e35696. doi:10.7554/eLife.35696

Sousa AMM, Zhu Y, Raghanti MA, Kitchen RR, Onorati M, Tebbenkamp ATN, Stutz B, Meyer KA, Li M, Kawasawa YI, Liu F, Perez RG, Mele M, Carvalho T, Skarica M, Gulden FO, Pletikos M, Shibata A, Stephenson AR, Edler MK, Ely JJ, Elsworth JD, Horvath TL, Hof PR, Hyde TM, Kleinman JE, Weinberger DR, Reimers M, Lifton RP, Mane SM, Noonan JP, State MW, Lein ES, Knowles JA, Marques-Bonet T, Sherwood CC, Gerstein MB, Sestan N. 2017. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**:1027–1032. doi:10.1126/science.aan3456

Stephan H, Andy OJ. 1969. QUANTITATIVE COMPARATIVE NEUROANATOMY OF PRIMATES: AN ATTEMPT AT A PHYLOGENETIC INTERPRETATION*. *Ann N Y Acad Sci* **167**:370–387. doi:10.1111/j.1749-6632.1969.tb20457.x

Striedter GF. 2004. Principles of Brain Evolution. Oxford, New York: Oxford University Press.

Striedter GF, Northcutt RG. 2019. Brains Through Time: A Natural History of Vertebrates, 1st ed. Oxford University Press. doi:10.1093/oso/9780195125689.001.0001

Suárez-Solá ML, González-Delgado FJ, Pueyo-Morlans M, Medina-Bolívar OC, Hernández-Acosta NC, González-Gómez M, Meyer G. 2009. Neurons in the White Matter of the Adult Human Neocortex. *Front Neuroanat* **3**. doi:10.3389/neuro.05.007.2009

Sulej T, Krzesiński G, Tałanda M, Wolniewicz AS, Błażejowski B, Bonde N, Gutowski P, Sienkiewicz M, Niedźwiedzki G. 2020. The earliest-known mammaliaform fossil from Greenland sheds light on origin of mammals. *Proc Natl Acad Sci* **117**:26861–26867. doi:10.1073/pnas.2012437117

Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**:377–382. doi:10.1038/nmeth.1315

Tasic B, Yao Z, Smith KA, Graybuck L, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller JA, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirchstein D, Pendergraft J, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. 2017. Shared and distinct transcriptomic cell types across neocortical areas. doi:10.1101/229542

Terriente J, Pujades C. 2015. Cell segregation in the vertebrate hindbrain: a matter of boundaries. *Cell Mol Life Sci* **72**:3721–3730. doi:10.1007/s00018-015-1953-8

Tiklová K, Björklund ÅK, Lahti L, Fiorenzano A, Nolbrant S, Gillberg L, Volakakis N, Yokota C, Hilscher MM, Hauling T, Holmström F, Joodmardi E, Nilsson M, Parmar M, Perlmann T. 2019. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nat Commun* **10**:581. doi:10.1038/s41467-019-08453-1

Tosches MA, Yamawaki TM, Naumann RK, Jacobi AA, Tushev G, Laurent G. 2018. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**:881–888. doi:10.1126/science.aar4237

Utter AA, Basso MA. 2008. The basal ganglia: An overview of circuits and function. *Neurosci Biobehav Rev*, Special section: Neurobiology of Deep Brain Stimulation: Innovations in Treatment and Basal Ganglia Function **32**:333–342. doi:10.1016/j.neubiorev.2006.11.003

Valenstein ES. 2002. The Discovery of Chemical Neurotransmitters. *Brain Cogn* **49**:73–95. doi:10.1006/brcg.2001.1487

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Hieter P, Vogelstein B, Kinzler KW. 1997. Characterization of the Yeast Transcriptome. *Cell* **88**:243–251. doi:10.1016/S0092-8674(00)81845-0

Virolainen S-M, Achim K, Peltopuro P, Salminen M, Partanen J. 2012. Transcriptional regulatory mechanisms underlying the GABAergic neuron fate in different diencephalic prosomeres. *Development* **139**:3795–3805. doi:10.1242/dev.075192

Waddington CH. 1957. The Strategy of the Genes. Routledge.

Wonders CP, Anderson SA. 2006. The origin and specification of cortical interneurons. *Nat Rev Neurosci* **7**:687–696. doi:10.1038/nrn1954

Xie Y, Dorsky RI. 2017. Development of the hypothalamus: conservation, modification and innovation. *Dev Camb Engl* **144**:1588–1599. doi:10.1242/dev.139055

Zhang K, Sejnowski TJ. 2000. A universal scaling law between gray matter and white matter of cerebral cortex. *Proc Natl Acad Sci* **97**:5621–5626. doi:10.1073/pnas.090504197

Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, Mazutis L. 2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nat Protoc* **12**:44–73. doi:10.1038/nprot.2016.154

Chapter 2: The Birth, Development and Evolution of Inhibitory Neurons in the Primate Cerebrum

Abstract

Neuroanatomists have long speculated that expanded primate brains contain an increased morphological diversity of inhibitory neurons (DeFelipe et al., 2013) and recent studies have identified primate-specific neuronal populations at the molecular level(Krienen et al., 2020). However, we know little about the developmental mechanisms that specify evolutionarily novel cell types in the brain. Here, we reconstructed gene expression trajectories specifying inhibitory neurons spanning the neurogenic period in macaque and mouse by analyzing the transcriptomes of 250,181 cells. We find that the initial classes of inhibitory neurons generated prenatally are largely conserved among mammals. Nonetheless, we identify two contrasting developmental mechanisms for specifying evolutionarily novel cell types during prenatal development. First, we show that recently-identified primate-specific TAC3 striatal interneurons are specified by a unique transcriptional program in progenitors followed by induction of a distinct suite of neuropeptides and neurotransmitter receptors in newborn neurons. Second, we find that multiple classes of transcriptionally-conserved olfactory-bulb bound precursors are redirected to expanded primate white matter and striatum, including a novel peristriatal class, *striatum laureatum* neurons, that resemble dopaminergic periglomerular cells of the olfactory bulb. We propose an evolutionary model in which conserved initial classes of neurons supplying the smaller primate olfactory bulb are reused in the enlarged striatum and cortex. Together, our results provide a unified developmental taxonomy of initial classes of mammalian inhibitory neurons and reveal multiple developmental mechanisms for neural cell type evolution.

Introduction

Functional neural circuits depend on diverse inhibitory neuron types for excitation-inhibition balance, rhythmic firing, and information processing(Kepecs and Fishell, 2014). In the cerebrum, inhibitory neurons mainly emerge prenatally from progenitors lining the lateral ventricles of the ventral telencephalon. These inhibitory neuron progenitors are generally partitioned by anatomical structures, the lateral, medial and caudal ganglionic eminences (LGE, MGE and CGE, respectively) and ventromedial forebrain (VMF) regions including the septum, preoptic area, anterior entopeduncular area, and preoptic hypothalamus(Lim et al., 2018; Ma et al., 2013; Marín et al., 2000). Spatial and temporal patterning of progenitors both within and between these regions influence the initial specification of inhibitory neuron classes. Following differentiation from progenitors, tangential, lateral, caudal, and rostral migratory streams from these subcortical structures supply inhibitory neurons to the cortex, basal ganglia, and olfactory bulb, and local signals further influence the maturation and post-mitotic refinement of initial neuron classes.(Fuentealba et al., 2015; Stenman et al., 2003; Wichterle et al., 2001, 1999)

The expansion of the primate cerebrum could influence the diversity of inhibitory neuron cell types and their migratory pathways by altering the regional destinations and requirements for inhibitory neuron populations. Indeed, primate cortex contains an increased proportion of inhibitory neurons (although this may be due to a murine decrease, rather than primate increase), primate striatum contains a novel population of TAC3-expressing interneurons, and

primate white matter contains millions of diverse neurons. In addition, the distribution of inhibitory neurons expressing tyrosine hydroxylase (TH), the rate-limiting enzyme of catecholamine synthesis has changed in primate cortex and striatum.(Benavides-Piccione and DeFelipe, 2007; DeFelipe, 2011; Krienen et al., 2020; Sousa et al., 2017). Although many inhibitory neuron migratory pathways are conserved among mammals (Hansen et al., 2013; Ma et al., 2013), recent studies suggest additional migratory pathways distinguish primate and rodent brains, including a ventromedial migratory stream supplying TH+ neurons to medial prefrontal cortex, and a dorsomedial migratory pathway called the arc, containing immature neurons in the early neonatal human brain, after cortical neurogenesis has concluded.(Paredes et al., 2016a; Sanai et al., 2011) Despite the increased complexity of primate inhibitory neuron development, we lack a taxonomy of initial primate cell types and their distribution along migratory streams that could illuminate developmental mechanisms underlying the origin of primate inhibitory neuron diversity.

Atlas of initial classes of primate INs

To examine the diversity of primate inhibitory neurons (INs) during development, we dissected progenitor zones in the ventral telencephalon and migratory destinations in the cortex and basal nuclei of prenatal rhesus macaque brains. We focused on the lateral, medial and caudal ganglionic eminences (LGE, MGE and CGE, respectively) and ventromedial forebrain (VMF) regions including the septum, preoptic area, and preoptic hypothalamus, where distinct configurations of transcription factors specify initial IN classes(Flames et al., 2007; Hansen et

al., 2013; Lim et al., 2018; Ma et al., 2013). We also sampled migratory destinations in cortex and basal nuclei where local signals further influence the maturation and post-mitotic refinement of neuron classes(Fishell and Kepecs, 2020). In total, we collected 71 samples across 9 specimens, spanning the onset of cortical neurogenesis, post-conception day (PCD) 40, to the conclusion, PCD100(Rakic, 1974) (**Figure 2.1, Figure 2.6**). We performed single cell RNA sequencing using the 10X Chromium Controller, and incorporated a recent analysis of PCD110(Zhu et al., 2018) as well as mouse studies from PCD13 - PCD21(Loo et al., 2019; Mayer et al., 2018), and adult olfactory bulb (OB)(Tepe et al., 2018). We applied stringent quality control, batch correction, dimensionality reduction, Leiden clustering, and RNA velocity trajectory analysis to identify transcriptionally similar classes of progenitors and post-mitotic INs among 109,112 macaque and 141,065 mouse cells, which were identified by expression of *DLX* and *GAD* genes (**Methods**).

Macaque and mouse IN progenitors mainly clustered by cell cycle phase rather than spatial origin (**Figure 2.1**, **Figure 2.6-8**). Similarly, the most immature newborn neurons clustered by both class and differentiation stage, according to RNA velocity latent time (**Figure 2.1, Figure 2.6-8**). From Leiden clusters, we delineated 11 discrete initial classes of macaque post-mitotic neurons, which resolve to 17 initial classes in mouse **(Figure 2.1)**. We found that canonical marker genes for established progenitor territories exhibited significant correlations among progenitors, suggesting core transcriptional regulatory programs are present at or prior to the

last cell division that predict initial classes identity of post-mitotic neurons. For example, signatures reflecting spatial origin and subtypes in the CGE (*NR2F1* and *NR2F2*), LGE (*MEIS2*, *EBF1* and *ISL1*), MGE (*MAF, LHX6,* and *CRABP1*) and VMF (*ZIC1* and *ZIC2*) are visible in G1 progenitors, and signatures related to four distinct LGE initial classes further emerge in dividing cells as correlations of the pan-LGE marker *MEIS2* to *PAX6*, *FOXP2*, *ISL1* or *PENK*, respectively (**Figure 2.6**). These observations support a model where we refer to differentiating post-mitotic neuron clusters, largely defined by germinal zone regional transcription factors, as "initial classes" of newborn neurons. This terminology reflects that a small number of discrete transcriptional classes are initially produced following a neuron's final cell division, and that each of these immature transcriptional states is later partitioned into one or many different mature classes by "nurture" and "circumstance" as neurons migrate and integrate into circuitry (**Figure 2.1)**(Fishell and Kepecs, 2020). Although many studies have demonstrated that these transcriptionally-defined classes result from shared lineage relationships(Wagner and Klein, 2020), we refrain from referring to classes as "lineages" in the absence of direct lineage tracing data.

Conserved and Divergent Initial Classes

To construct a taxonomy of IN development, we sought to identify evolutionarily conserved cell classes and link these to candidate adult populations(Miller et al., 2020). We found most well-known initial class markers to be conserved between species **(Figure 2.2, Figure 2.9)**. Gene expression signatures for each progenitor and post-mitotic initial class in macaque correlated

strongly with at least one comparable class in mouse (**Figure 2.9**). Most classes have 1-to-1 relationships, however subclasses like LGE_FOXP1/ISL1/NPY1R and LGE_MEIS2/PAX6/SCGN as well as a number of VMF classes are apparent in mouse data but undersampled in macaque. Because cell type correlation methods depend on clustering resolution in each species, we further examined homology at the level of individual cells. Mutual nearest neighbor analysis revealed that all telencephalic initial classes present in mouse were also present in macaque (**Figure 2.2**). To infer putative fates of initial classes in the absence of lineage tracing, we compiled the most complete adult brain data available from mouse (**Figure 2.10**). We then computed terminal class absorption probabilities for prenatal neurons using nearest neighbor relationships and RNA velocity with equal weight in cellrank's Markov chain model(Lange et al., 2020). Our predicted mapping of post-mitotic differentiation and partitioning of each initial class using transcriptional similarities recapitulates known lineage relationships and makes a number of unexpected predictions that support unresolved linkages in the literature, summarized in **Figure 2.2**, such as an NKX2-1+ MGE-derived LAMP5+ cortical chandelier population(Paul et al., 2017; Tasic et al., 2017; Valero et al., 2021) and a shared origin of amygdala intercalated cells and striatal eccentric spiny neurons(Kuerbitz et al., 2018; Märtin et al., 2019; Saunders et al., 2018). The widespread distribution and diversity of derivatives from some initial classes, such as MGE_LHX6/NPY, also highlights shared genetic programs underlying the initial specification of populations that later diversify according to regional destinations, where terminal classes are commonly subdivided into many transcription-

and morpho-types. While our results suggest that initial classes are largely uniform, minor axes of variation may already exist within classes that could trigger downstream cascades that bias terminal fate partitioning, but which cannot be identified here without knowing a cell's fate *a priori*.

Comparative studies of adult primate, rodent, and ferret telencephalon recently revealed a primate-specific population of striatal interneurons that express the neuropeptide TAC3(Krienen et al., 2020), but the developmental origin of this evolutionarily novel population remained unclear. These striatal interneurons proved a striking exception to the 1-to-1 conservation of initial classes between primates and rodents (**Figure 2.2**). Instead, mouse has a single ancestral class of MGE_CRABP1/MAF neurons that shows strong homology to both the macaque MGE_CRABP1/MAF and MGE_CRABP1/TAC3 clusters (**Figure 2.9**). We further examined the gene networks that define this primate-specific population (**Figure 2.3, Figure 2.11**). Using RNAscope, we quantified co-expression of dividing cell marker *MKI67* and initial class markers *CRABP1,TAC3*, *MAF*, and *LHX8* across the rostrocaudal expanse of the MGE and striatum at PCD65. Our results reveal a bias of MGE_CRABP1/MAF neurons rostrally and MGE_CRABP1/TAC3 neurons caudally in the MGE progenitor zone (**Figure 2.3**). In addition, we detected a low fraction of cells co-expressing *CRABP1*, *TAC3* and *MKI67*, displaced from the ventricle, suggesting that subventricular zone (SVZ) progenitors upregulate the program for this novel initial class at or prior to their final cell division (**Figure 2.12)**. In the striatum, both

classes showed uniform distributions, also confirmed by RNAscope for *STXBP6, ANGPT2,* and *RBP4* in two additional individuals (**Figure 2.12, Figure 2.13**). *LHX8* was restricted to a subset of *CRABP1+/TAC3+* cells outside of the MGE (**Figure 2.3**)*,* highlighting early postmitotic specification of a *TAC3/LHX8* subclass observed in adult marmoset (Krienen et al., 2020). Interestingly mouse cholinergic and pallidal neurons (VMF_CRABP1/LHX8), though clearly distinct from primate MGE_CRABP1/TAC3 neurons, also express *Zic1* and *Lhx8* (**Figure 2.3**), hinting that a combination of transcriptional programs used by neighboring initial classes may define the novel TAC3 population. Differential expression and regulon analysis revealed that the earliest molecular programs that distinguish TAC3 interneurons involve distinct neuropeptides, acetylcholine receptors, and immediate early gene networks (**Figure 2.11**) suggesting TAC3 neurons may receive signals from nearby cholinergic neurons. Notably, the primate-specific TAC3 population emerges as a distinct class as cells become postmitotic by PCD65, far earlier in development than the conserved PTHLH+, PVALB+ and TH+ terminal fates that ultimately arise from the related MGE_CRABP1/MAF class(Krienen et al., 2020; Muñoz-Manchado et al., 2018). Lastly, we found that MGE_CRABP1 classes emerge *in vitro* as rare populations in human pluripotent stem cell-derived telencephalon organoids (**Figure 2.11**).

Re-use of OB neurons in primate cerebrum

We next analyzed the initial classes of neurons detected within and likely derived from the LGE(Ma et al., 2013). Two classes, LGE_FOXP2/TSHZ1 and LGE_MEIS2/PAX6, showed unexpected enrichment in the cortical frontal lobe in addition to ventral telencephalon (**Figure**

2.4, Figure 2.14). LGE_MEIS2/PAX6 neurons express *ETV1*, *SP8*, *MEIS2*, *SALL3*, *TSHZ1* and *PAX6* during differentiation, which all mark and are required for proper production of OB granule cells and dopaminergic TH+ periglomerular cells (PGCs)(Agoston et al., 2014; Cave et al., 2010). Indeed the transcriptomes of this class showed strong correlations to mouse adult-born granule cells (OB-GC_MEIS2/PAX6, **Figure 2.9**). Similarly, trajectory analysis linked the mouse LGE_FOXP2/TSHZ1 class to OB-PGC_FOXP2/CALB1 periglomerular cells of the OB, connecting each LGE class to distinct olfactory populations (**Figure 2.2**).

We performed immunofluorescence microscopy to visualize the spatial distribution of the LGE_FOXP2/TSHZ1 and LGE_MEIS2/PAX6 classes, using combinations of MEIS2 together with FOXP2/FOXP4 and SCGN/SP8/PAX6, respectively. Both populations appeared to emanate from the dorsal LGE (dLGE), but showed complementary distributions. LGE_FOXP2/TSHZ1 cells immunoreactive for FOXP2, FOXP4 and SCGN are mainly found in the dorsolateral dLGE (DL-dLGE**, Figure 2.15, Figure 2.16**), but not in the anterior dLGE (A-dLGE) or rostral migratory stream (RMS). Instead, cells of this class migrate directly into the striatum, via the lateral migratory stream (LMS) to the outer OB, and ventro-medially to cortical superficial white matter (**Figure 2.15-17**). Consistent with trajectory analysis, markers of dLGE origin are downregulated (*ETV1*, SCGN, *SP8*), while *FOXP2*, *CASZ1, OPRM1* and projection neuron markers are upregulated as cells differentiate and migrate into the striatum (**Figure 2.14, Figure 2.15**). Expression of *TSHZ1, LYPD1, PCDH8* and *CASZ1,* the absence of canonical

medium spiny projection neuron markers *NPY* and *FOXP1,* and RNA velocity analysis suggest that the LGE_FOXP2/TSHZ1 initial class also explains the previously unknown developmental origin of a recently described striatal projection neuron in adult mice, the "eccentric spiny projection neurons" (eSPN)(Saunders et al., 2018) and amygdala intercalated cells (ITCs)(**Figure 2.2, Figure 2.14**). This linkage is consistent with reports that cells in mouse dLGE initially express SP8 and maintain TSHZ1 expression as they migrate via the LMS to become amygdala ITCs(Kuerbitz et al., 2018, n.d.). This developmental perspective suggests these cells are not "eccentric" deviations from canonical spiny projection neuron development, but instead the LGE_FOXP2/TSHZ1 class converges on a similar striatal and amygdaloid projection neuron transcription profile, despite a distinct origin.

In contrast to the LGE_FOXP2/TSHZ1 class, we observe MEIS2+/PAX6+/SP8+/SCGN+ cells representing the LGE_MEIS2/PAX6 class continuously from the anterior end of the dLGE along the RMS to the OB granule cell layer (**Figure 2.4, Figure 2.16, Figure 2.17**). Notably, we observed dense parenchymal chains of these cells radiating from the dLGE at PCD80 (n=3 hemispheres) (**Figure 2.4, Figure 2.17**). At PCD120, we found large numbers of LGE_MEIS2/PAX6 precursors that express SCGN extending dorsomedially and caudally in the Arc migratory stream(Paredes et al., 2016a) in addition to the RMS (**Figure 2.18**). These cells were densest in chains running along the whole striatum in the primary tier of the Arc with fewer cells radially(Paredes et al., 2016a). Surprisingly, we also observed a robust stream diverted

from the Arc stretching from the A-dLGE into the anterior cingulate cortex (**Figure 2.4**). This stream which we refer to as the Arc-ACC appeared to be bounded by TH+ fibers in superficial white matter **(Figure 2.4).** Cells from the Arc and Arc-ACC were common in dorsomedial cortex deep white matter, but rarely found lateral or ventral of the striatum, while many CGE-derived MEIS2-/SP8+/NR2F2+ neurons were observed throughout white matter (**Figure 2.18**), highlighting regional heterogeneity in the composition of white matter INs.

We confirmed that LGE_MEIS2/PAX6 neurons from A-dLGE also contribute to the RMS, the Arc, and Arc-ACC in perinatal human (**Figure 2.4**) and postnatal macaque (**Figure 2.18**). We further found that these neurons persist postnatally in the deep white matter of the cingulate cortices and the superior corona radiata (**Figure 2.18**). In contrast, in postnatal day 2 (P2) mouse, we only identified rare instances of LGE_MEIS2/PAX6 cells in deep white matter (**Figure 2.19**), consistent with recent reports that sparse MEIS2+/HTR3A+ neurons integrate into cortical circuitry perinatally(Frazer et al., 2017). Instead, the vast majority of these cells appear in the anterior SVZ and RMS in mouse **(Figure 2.19)**. Overall we find that neurons derived from dLGE are more widely distributed than previously recognized in primates, representing a major source of neurons in the primate Arc migratory streams and persisting in the deep white matter.

Our analysis revealed a third presumed dLGE-derived class in and around the striatum, insula and claustrum, derivative of the LGE_MEIS2/PAX6 initial class, termed *striatum laureatum*

neurons (SLNs or Str-SLN_TH/SCGN) for the wreath they form around the striatum. At both PCD120 and 7 months postnatal, SLNs are immunoreactive for PAX6, MEIS2, SP8, TH, and SCGN, but not FOXP2, NKX2-1, or NR2F2 a combination also characteristic of TH+ periglomerular cells (OB-PGC_TH/SCGN) of the OB (**Figure 2.5, Figure 2.18, Figure 2.20**). This distribution matches observations of TH+ cells circumscribing the primate striatum and their reported absence in rodents and illuminates their molecular identity and origin(Dubach et al., 1987). Indeed, we did not identify any MEIS2/PAX6/SCGN/TH+ cells along the mouse striatum border or the claustrum (**Figure 2.19**). Instead, in mouse, these cells were restricted to the olfactory bulb, tract or tubercle, matching the macaque olfactory peduncle domain(**Figure 2.5**). We find that SLNs form a reticule at the white matter boundaries of the caudate and putamen of macaques and humans and persist throughout life (**Figure 2.5**).

Discussion

By identifying transcriptional regulatory programs distinguishing the earliest specification of initial classes, our study provides a resource for identifying conserved molecular mechanisms that specify cell type diversity, for rational in vitro derivation of these populations from pluripotent stem cells, and for interpretation of the cellular substrates of genetic disorders of neural development. TAC3-expressing striatal interneurons represent an exceptional case in which an evolutionarily novel initial class of neurons emerges in differentiating progenitors. A limited number of gene networks distinguish the TAC3 initial class from the related MGE_CRABP1/MAF class consistent with a recent model of cell type evolution. Under this

model an ancestral cell type is partitioned into distinct subtypes by changes in transcription factor expression that enable genomic individuation of sister cell classes that still share many regulatory complexes and developmental trajectories(Arendt et al., 2016). However, the conservation of nearly all other initial classes of INs between macaque and mouse suggests that evolutionary diversification of primate INs arises mainly by radiation of conserved initial classes of newborn neurons and may be shaped by the expanded diversity of primate regional destinations(Fishell and Kepecs, 2020).

The neurons of the dLGE appear to be particularly affected by primate brain reorganization. In both macaque and mouse, the LGE_MEIS2/PAX6 class is among the latest-born INs and migrates to olfactory structures and deep white matter. However, the absolute migration distance of late born A-dLGE neurons to OB is over two orders of magnitude longer in newborn macaque compared to mouse, increasing further as the brain expands after birth.(Paredes et al., 2016b, 2016a) Similarly, the volume of white matter is over three and five orders of magnitude larger in macaque and human than mouse, respectively(Zhang and Sejnowski, 2000), while the relative size of the primate OB is dramatically smaller (**Figure 2.5**)(Stephan and Andy, 1969). Thus in mouse, the birthplace is only several cell lengths from any point in the adjacent deep white matter, but in macaque these homologous cells traverse histologically-distinct dorsal migratory streams, apparently reusing the chain migration strategy. OB granule cells derived from this class contribute to adult plasticity(Lledo et al., 2006), and myelination is

delayed for up to two decades in human frontal lobe white matter (Miller et al., 2012), potentially linking these cells to white matter plasticity. Notably, abnormal accumulations of frontal lobe white matter neurons have been reproducibly associated with schizophrenia and autism(Duchatel et al., 2019). With their prolonged migration to far-flung and ever-changing destinations, the A-dLGE neurons that we identify here may be particularly vulnerable to environmental influences, and the markers we identified will be useful for assessing the molecular heterogeneity of disease-associated populations.

Finally, we identified striatum laureatum neurons, another likely OB sister type redistributed to peristriatal regions that share a molecular resemblance to dopaminergic OB TH+ PGCs. Future studies can examine whether this primate striatal population partly explains the human-specific increase in TH-expressing striatal neurons(Diederich et al., 2019; Sousa et al., 2017), and whether these neurons produce dopamine themselves or play an auxiliary role to compensate for increased demands on midbrain dopaminergic neurons.(Betarbet et al., 1997; Björklund and Dunnett, 2007) Francis Crick speculated that in the claustrum, hitherto undiscovered sparse INs resembling intraglomerular OB cells with dendrodendritic synapses could contribute to binding information(Crick and Koch, 2005), and molecular access to this rare population will enable future circuit-level studies. Together, our results highlight contrasting models for diversification of primate INs by specification of an entirely novel initial class and by redistribution of conserved

initial classes that supply OB and peduncle into primate white matter migratory streams and peristriatal locations.

Figures

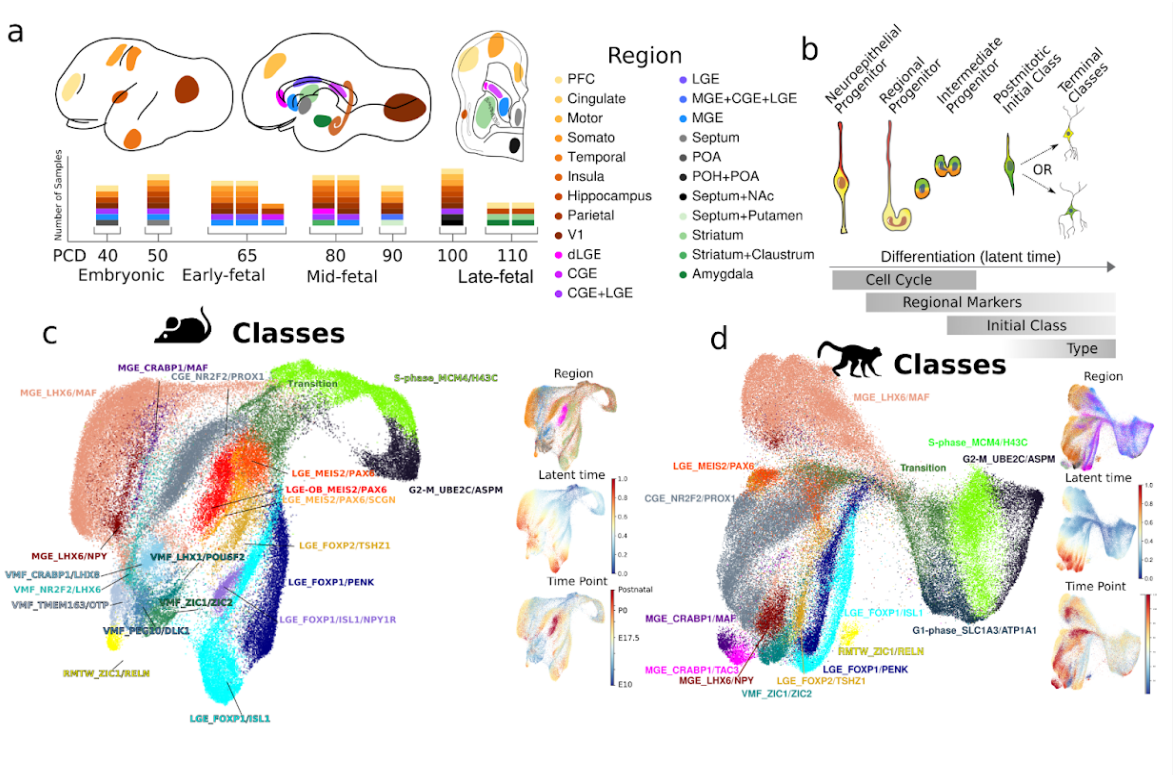

Figure 2.1: Transcriptional diversity of IN precursors in developing macaque and mouse telencephalon.

a. Regions dissected for scRNA-seq, labeled on PCD80 macaque lateral, medial and frontal coronal section traces. Stacked boxes represent samples from an individual; PCD110 samples from(Zhu et al., 2018). **b.** Model of inhibitory neurogenesis. **c-d.** UMAP projections colored by progenitor state and initial class for mouse (**c**) and macaque (**d**). Insets show dissection region, from **a**, scVelo dynamical RNA velocity shared latent time, and age.

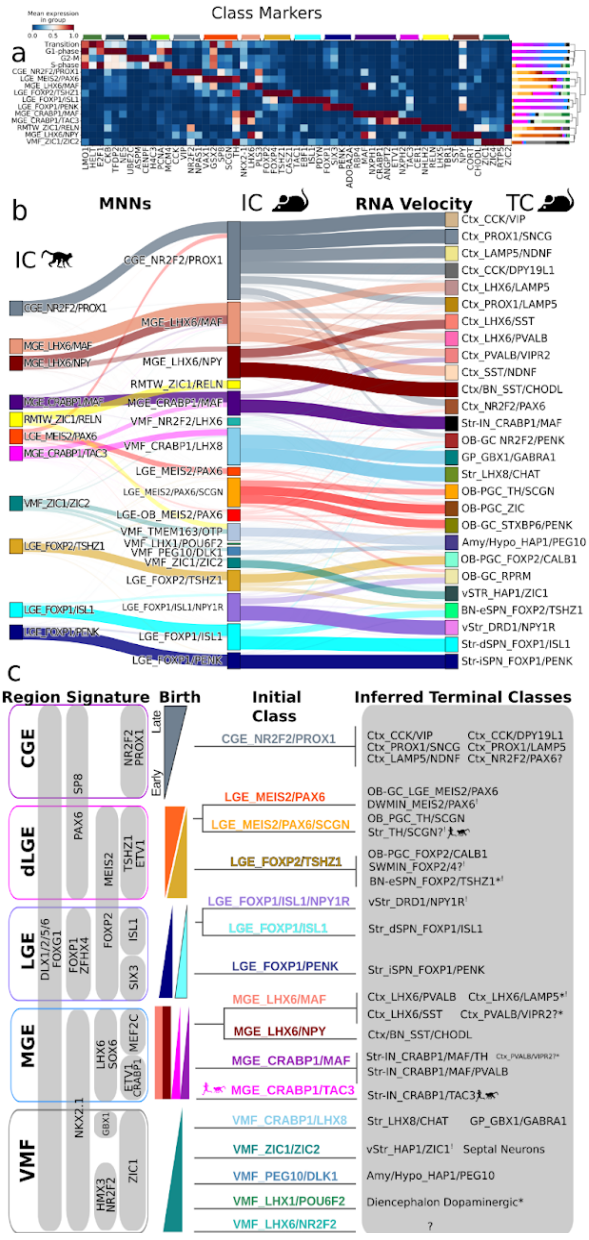
Class Markers
a
b
MNNs
IC
RNA Velocity
TC
CGE_NR2F2/PROX1
MGE_LHX6/MAF
MGE_LHX6/NPY
RMTW_ZIC1/RELN
MGE_CRABP1/MAF
VMF_NR2F2/LHX6
VMF_CRABP1/LHX8
LGE_MEIS2/PAX6
LGE_MEIS2/PAX6/SCGN
LGE-OB_MEIS2/PAX6
VMF_TMEM163/OTP
VMF_LHX1/POU6F2
VMF_PEG10/DLK1
VMF_ZIC1/ZIC2
LGE_FOXP2/TSHZ1
LGE_FOXP1/ISL1
LGE_FOXP1/ISL1/NPY1R
LGE_FOXP1/PENK
Ctx_CCK/VIP
Ctx_PROX1/SNCG
Ctx_LAMP5/NDNF
Ctx_CCK/DPY19L1
Ctx_LHX6/LAMP5
Ctx_PROX1/LAMP5
Ctx_LHX6/SST
Ctx_LHX6/PVALB
Ctx_PVALB/VIPR2
Ctx_SST/NDNF
Ctx/BN_SST/CHODL
Ctx_NR2F2/PAX6
Str-IN_CRABP1/MAF
OB-GC NR2F2/PENK
GP_GBX1/GABRA1
Str_LHX8/CHAT
OB-PGC_TH/SCGN
OB-PGC_ZIC
OB-GC_STXBP6/PENK
Amy/Hypo_HAP1/PEG10
OB-PGC_FOXP2/CALB1
OB-GC_RPRM
vSTR_HAP1/ZIC1
BN-eSPN_FOXP2/TSHZ1
vStr_DRD1/NPY1R
Str-dSPN_FOXP1/ISL1
Str-iSPN_FOXP1/PENK
c
Region Signature
Birth
Initial Class
Inferred Terminal Classes
CGE
dLGE
LGE
MGE
VMF
Late
Early
CGE_NR2F2/PROX1
Ctx_CCK/VIP Ctx_CCK/DPY19L1
Ctx_PROX1/SNCG Ctx_PROX1/LAMP5
Ctx_LAMP5/NDNF Ctx_NR2F2/PAX6?
LGE_MEIS2/PAX6
LGE_MEIS2/PAX6/SCGN
OB-GC_LGE_MEIS2/PAX6
DWMIN_MEIS2/PAX6†
OB_PGC_TH/SCGN
Str_TH/SCGN?†
LGE_FOXP2/TSHZ1
OB-PGC_FOXP2/CALB1
SWMIN_FOXP2/4?†
BN-eSPN_FOXP2/TSHZ1*†
LGE_FOXP1/ISL1/NPY1R
vStr_DRD1/NPY1R†
LGE_FOXP1/ISL1
Str_dSPN_FOXP1/ISL1
LGE_FOXP1/PENK
Str_iSPN_FOXP1/PENK
MGE_LHX6/MAF
Ctx_LHX6/PVALB Ctx_LHX6/LAMP5*†
Ctx_LHX6/SST Ctx_PVALB/VIPR2?*
MGE_LHX6/NPY
Ctx/BN_SST/CHODL
MGE_CRABP1/MAF
Str-IN_CRABP1/MAF/TH
Str-IN_CRABP1/MAF/PVALB
MGE_CRABP1/TAC3
Str-IN_CRABP1/TAC3
VMF_CRABP1/LHX8
Str_LHX8/CHAT GP_GBX1/GABRA1
VMF_ZIC1/ZIC2
vStr_HAP1/ZIC1† Septal Neurons
VMF_PEG10/DLK1
Amy/Hypo_HAP1/PEG10
VMF_LHX1/POU6F2
Diencephalon Dopaminergic*
VMF_LHX6/NR2F2
?

Figure 2.2: Unified taxonomy of Euarchontogliran telencephalic IN specification.

a. Heatmap of macaque initial class markers, scaled by column. Dendrogram represents complete linkage of Pearson correlation distance of mean expression values. Stacked bar plots show regional distribution of each class. **b.** Sankey diagram where thickness of lines between left and middle column represents the number of mutual nearest neighbor cells shared between each class, and between middle and right column reflects initial class identity of 100 cells with the highest (cellrank) probabilities to be absorbed to each terminal class. **c.** Summarized taxonomy of initial and terminal classes observed in macaque and mouse. Forked lines represent subclasses which become apparent post-mitotically. Initial classes of INs are organized by presumptive birthplace based on the expression of regional marker genes and putative birthdates, presented in the manner of Lim et al.(Lim et al., 2018) Inferred terminal fates are based upon our gene expression and histology analysis and the literature, as denoted, and discussed in detail in **Table 2.1**. S/DWMIN: Superficial/Deep White Matter Inhibitory Neuron, BN: Basal Nuclei. RMTW_ZIC1/RELN and VMF_TMEM163/OTP are not included, as they are excitatory cortical and hypothalamic classes, respectively.

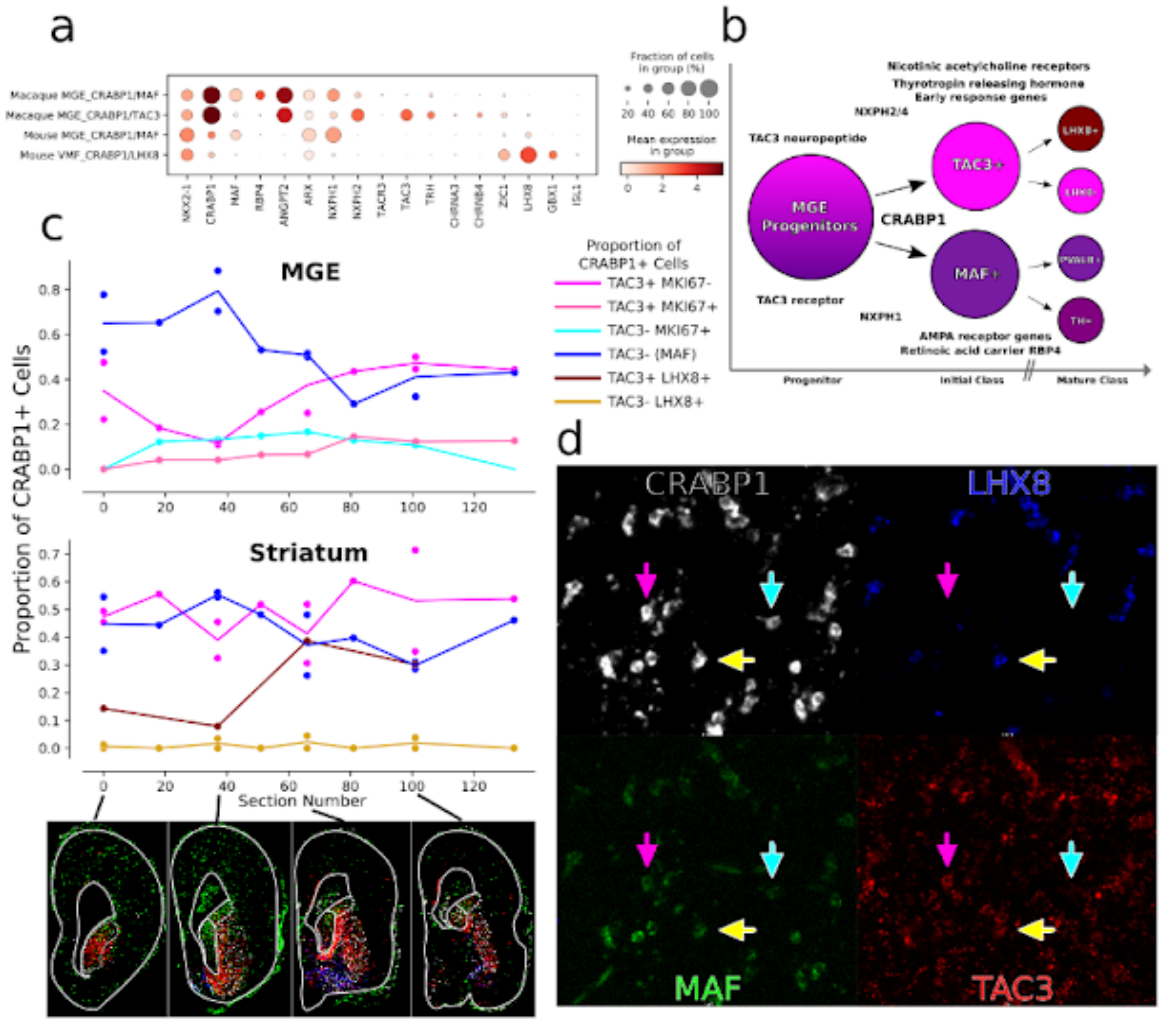
a
Macaque MGE_CRABP1/MAF
Macaque MGE_CRABP1/TAC3
Mouse MGE_CRABP1/MAF
Mouse VMF_CRABP1/LHX8
NKX2-1
CRABP1
MAF
RBP4
ABGPT2
ARX
NXPH1
NXPH2
TACR3
TAC3
TRH
CHRNA3
CHRNB4
ZIC1
LHX8
GBX1
ISL1
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2 4
b
Nicotinic acetylcholine receptors
Thyrotropin releasing hormone
Early response genes
NXPH2/4
TAC3 neuropeptido
MGE Progenitors
TAC3 receptor
CRABP1
TAC3+
MAF+
NXPH1
AMPA receptor genes
Retinoic acid carrier RBP4
Progenitor
Initial Class
Mature Class
c
MGE
Striatum
Proportion of CRABP1+ Cells
Section Number
Proportion of CRABP1+ Cells
TAC3+ MKI67-
TAC3+ MKI67+
TAC3- MKI67+
TAC3- (MAF)
TAC3+ LHX8+
TAC3- LHX8+
d
CRABP1
LHX8
MAF
TAC3

Figure 2.3: Emergence of primate-specific MGE_CRABP1/TAC3 striatal interneurons.

a. Dotplot of striatal interneuron marker genes. **b.** Schematic summarizing properties distinguishing newborn MGE CRABP1+/TAC3+ or /MAF+ neurons. **c.** Line plot showing rostro-caudal distribution of classes of CRABP1+ cells. Each point is the sum of all cells in at least 5 random fields of view in each section/region. Whole section scans of representative sections from which cells were counted (full size in Extended Data). Solid outline GE region represents MGE, while dotted outlines LGE. One individual was used, with 4 pairs of tandem sections, interspersed with 4 single sections. **d.** Representative image of MGE_CRABP1/MAF (blue arrow), MGE_CRABP1/TAC3 (pink arrow), MGE_CRABP1/TAC3/LHX8 (yellow arrows), and VMF_CRABP1/LHX8 interneurons (LHX8+/TAC3- cells) in the putamen from section 66 (**Extended Data Figure 8c**).

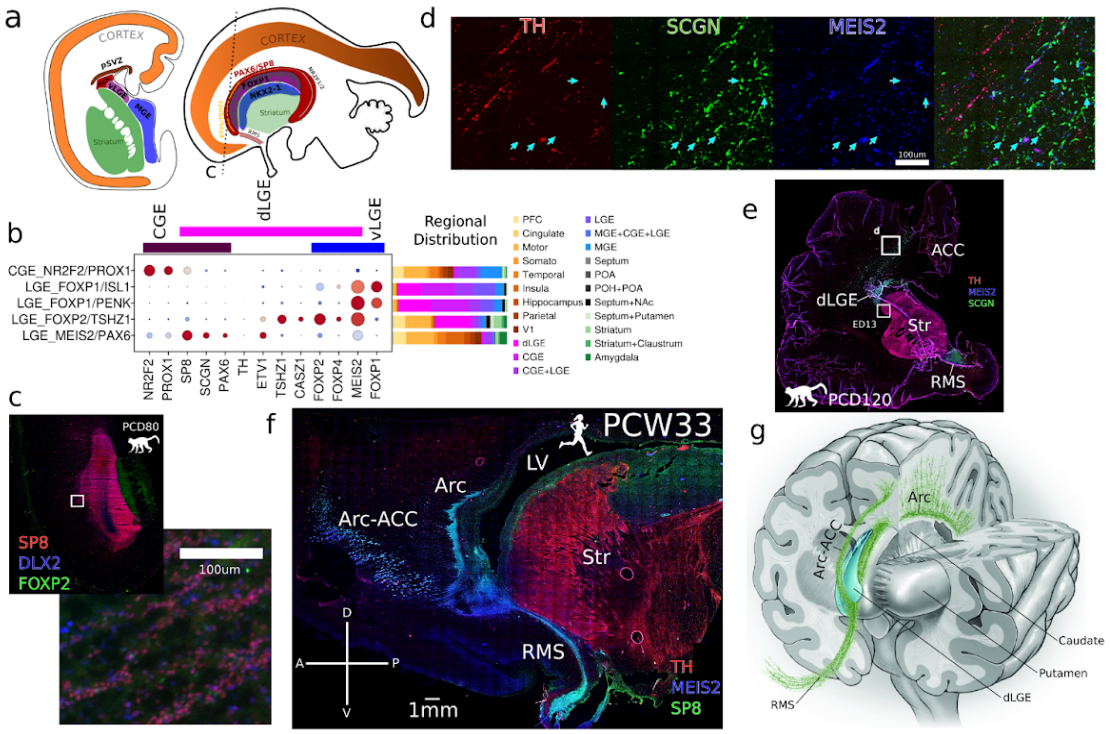

Figure 2.4: Redistribution of LGE_MEIS2/PAX6 granule cells.

a. Approximate ganglionic eminence transcription factor territories in PCD80 macaque brain, with estimated section planes. **b.** Dotplot of markers of LGE and CGE-derived classes shows overlap of transcription factor domains. Expression values are scaled 0-1 per gene. Dot size represents percent of cells that express each gene. Stacked bar plots show regional distribution of each class. **c.** Coronal-axial section of dLGE and PFC with inset highlighting DLX2+/SP8+/FOXP2- parenchymal chains. **d.** Arc-ACC SCGN+/MEIS2+ cells shown migrating within the boundary of dense TH+ axons. **e.** Whole PCD120 coronal section from **d** showing Arc-ACC and RMS wrapping around the striatum from the A-dLGE. **f.** Human PCW33 low magnification image showing large streams of MEIS2+/SP8+ neurons originating from the A-dLGE contributing to the RMS and the Arc. **g.** Schematic of macaque brain. Multiple streams extend from anterior pole of dLGE, or SVZ at later stages, to the RMS leading to the OB, the Arc extending dorsomedially, and a subsidiary stream, the Arc-ACC, extending to the anterior cingulate cortex. LV: lateral ventricle.

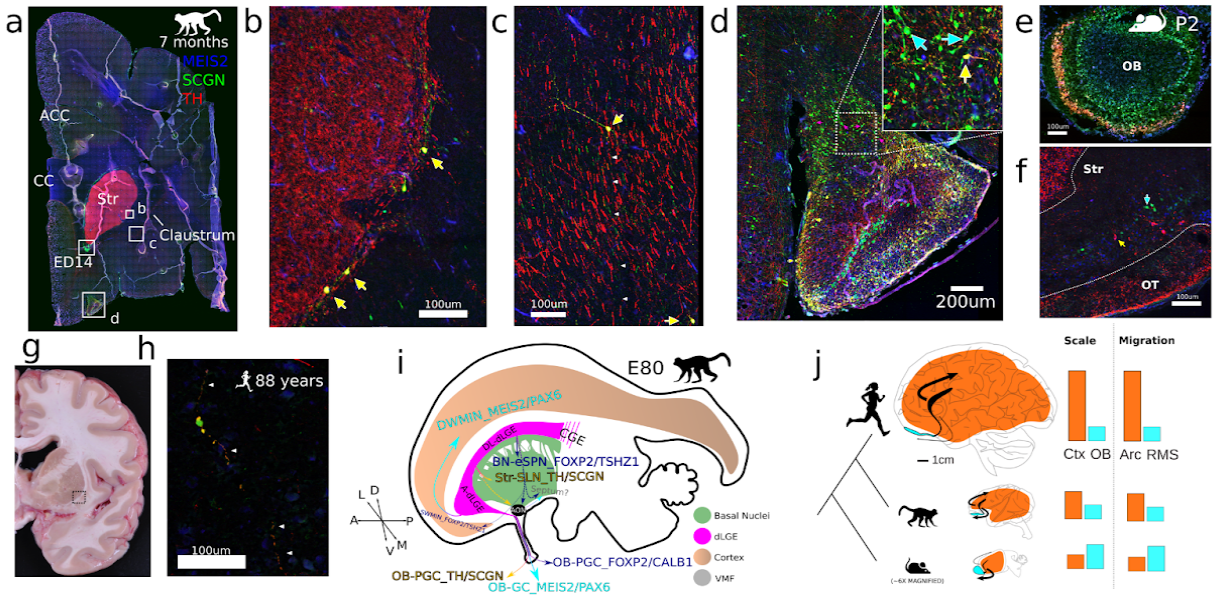

Figure 2.5: **TH+ *striatum laureatum* neurons and ancestral olfactory populations.**

All stains in this figure are MEIS2/SCGN/TH **a.** 7 month postnatal macaque coronal section including remnant RMS. **b.** TH+ SLNs at the border of the striatum. **c.** MEIS2+/SCGN+/TH+ peristriatal SLNs (yellow arrows). **d.** SLNs in claustrum with long, straight processes, one TH+ process (orange arrowhead), one TH- process (white arrowheads), among dense TH+ midbrain-cortical fiber synapses. **d.** Anterior olfactory nucleus at the olfactory peduncle**,** with MEIS2+/SCGN+/TH- cells (magenta arrows) including SCGN+/TH- fibers entering ventral cortex (inset) and triple positive cells (yellow arrows). **e.** Coronal section of mouse P2 OB showing SLN-analogous TH+/SCGN+ cells. **f.** Mouse P2 coronal section showing olfactory tubercle and striatum, with only MEIS2+/TH+ or MEIS2+/SCGN+ cells. **g.** Photograph of 88 year old human brain coronal slab with box outlining approximate section block. **h.** TH+/SCGN+ human SLN, with double positive processes highlighted with white arrowheads. Only two cells were observed across this section. **i.** Schematic summarizing the unequal scaling of cortical and olfactory structures. Relative values in bar plots are arbitrary. **j.** Schematic summarizing migration and cell types likely generated by primate dLGE subdomains.

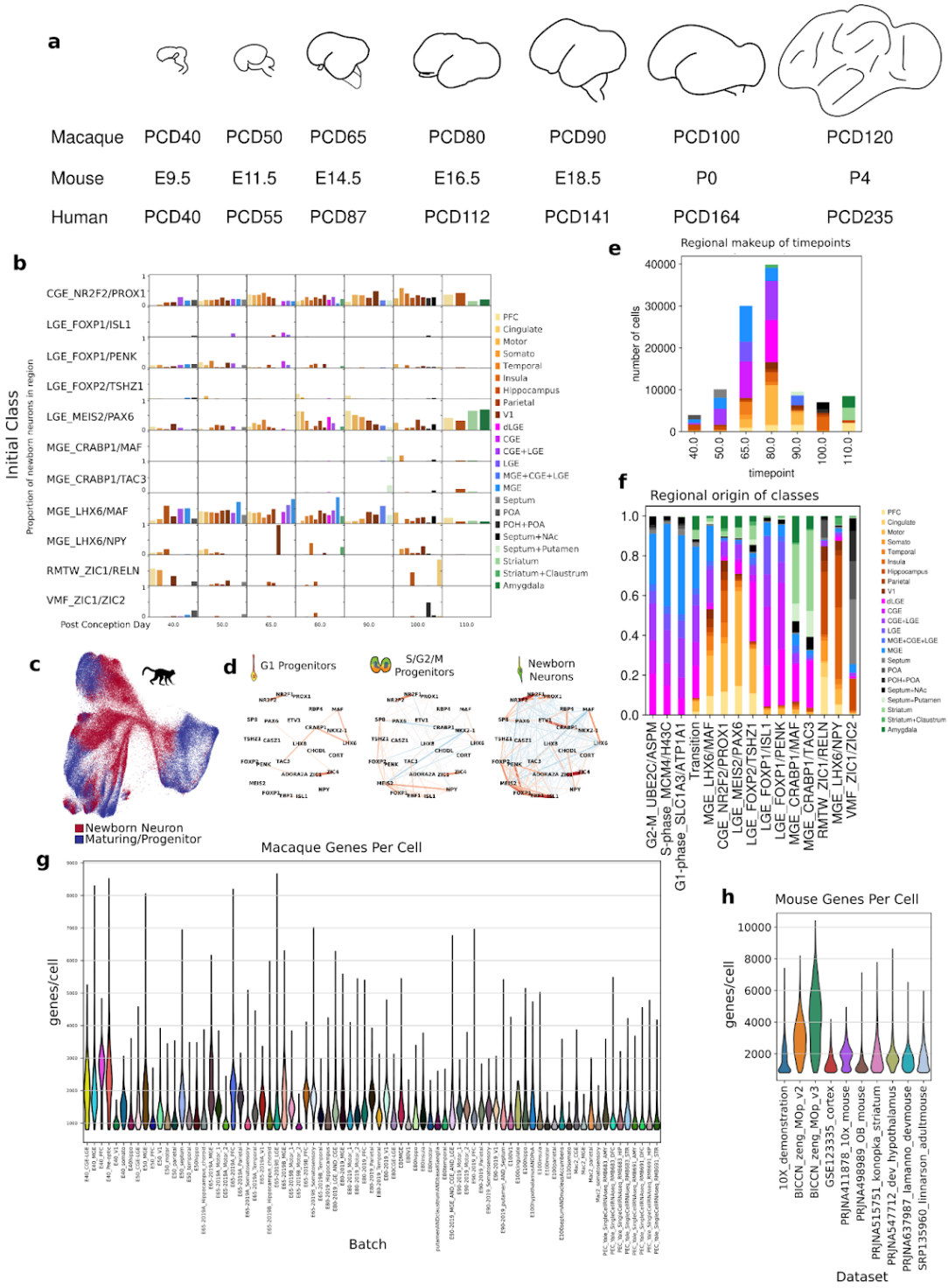
a
Macaque PCD40 PCD50 PCD65 PCD80 PCD90 PCD100 PCD120
Mouse E9.5 E11.5 E14.5 E16.5 E18.5 P0 P4
Human PCD40 PCD55 PCD87 PCD112 PCD141 PCD164 PCD235
b
Initial Class
Proportion of newborn neurons in region
CGE_NR2F2/PROX1
LGE_FOXP1/ISL1
LGE_FOXP1/PENK
LGE_FOXP2/TSHZ1
LGE_MEIS2/PAX6
MGE_CRABP1/MAF
MGE_CRABP1/TAC3
MGE_LHX6/MAF
MGE_LHX6/NPY
RMTW_ZIC1/RELN
VMF_ZIC1/ZIC2
Post Conception Day
PFC
Cingulate
Motor
Somato
Temporal
Insula
Hippocampus
Parietal
V1
dLGE
CGE
CGE+LGE
LGE
MGE+CGE+LGE
MGE
Septum
POA
POH+POA
Septum+NAc
Septum+Putamen
Striatum
Striatum+Claustrum
Amygdala
c
Newborn Neuron
Maturing/Progenitor
d
G1 Progenitors
S/G2/M Progenitors
Newborn Neurons
e
Regional makeup of timepoints
number of cells
timepoint
f
Regional origin of classes
G2-M_UBE2C/ASPM
S-phase_MCM4/H43C
G1-phase_SLC1A3/ATP1A1
Transition
MGE_LHX6/MAF
CGE_NR2F2/PROX1
LGE_MEIS2/PAX6
LGE_FOXP2/TSHZ1
LGE_FOXP1/ISL1
LGE_FOXP1/PENK
MGE_CRABP1/MAF
MGE_CRABP1/TAC3
RMTW_ZIC1/RELN
MGE_LHX6/NPY
VMF_ZIC1/ZIC2
g
Macaque Genes Per Cell
genes/cell
Batch
h
Mouse Genes Per Cell
genes/cell
10X_demonstration
BICCN_zeng_MOp_v2
BICCN_zeng_MOp_v3
GSE123335_cortex
PRJNA411878_10x_mouse
PRJNA498989_OB_mouse
PRJNA515751_konopka_striatum
PRJNA547712_dev_hypothalamus
PRJNA637987_lamanno_devmouse
SRP135960_linnarson_adultmouse
Dataset

Figure 2.6: Birthdates of initial classes of INs in macaque.

a. Drawings of the lateral view of developing macaque brains across stages surveyed in this study and estimated comparable human and mouse stages based on the translating time model (Clancy et al., 2001) of cortical neurogenesis. **b.** Spatiotemporal distribution of newborn neurons, as determined by RNA velocity latent time, from each class as a proxy for birthdate. Bars represent the proportion of cells from each class in each region, at each timepoint (columns sum to 1). Cajal-Retzius neurons, MGE-derived cortical interneurons, and LGE-derived projection neurons first appeared early in development, starting at PCD40, followed by the later appearance of immature CGE-derived cortical interneurons and LGE_MEIS2/PAX6 neurons, consistent with broad patterns of temporal ordering in mouse.(Bielle et al., 2005; Miyoshi et al., 2010; Stenman et al., 2003) **c.** UMAP showing which cells are newborn in red (latent time < 0.5 quantile of latent time for each class) or maturing (>0.5 quantile) in blue. Clearly cycling progenitors are not included. **d.** Kamada-Kawai graph visualization of Pearson correlations between a gene pair's expression highlight the emergence of initial class gene co-expression patterns during macaque neuronal differentiation. Edges shown are Holm-Šídák corrected q value <0.05 calculated by bootstrap, with thickness and color representing correlation. **e.** Stacked bar chart showing number of macaque cells collected at each timepoint, colored by region from which the cells are derived. Note undersampling of VMF structures between PCD50 and PCD100. **f.** Normalized stacked bar chart showing regions from which each macaque initial class is derived across the whole dataset. **g.** Violin plot showing the distribution of genes detected per cell for each macaque batch. **h.** Violin plot showing the distribution genes per cell detected for each mouse dataset.

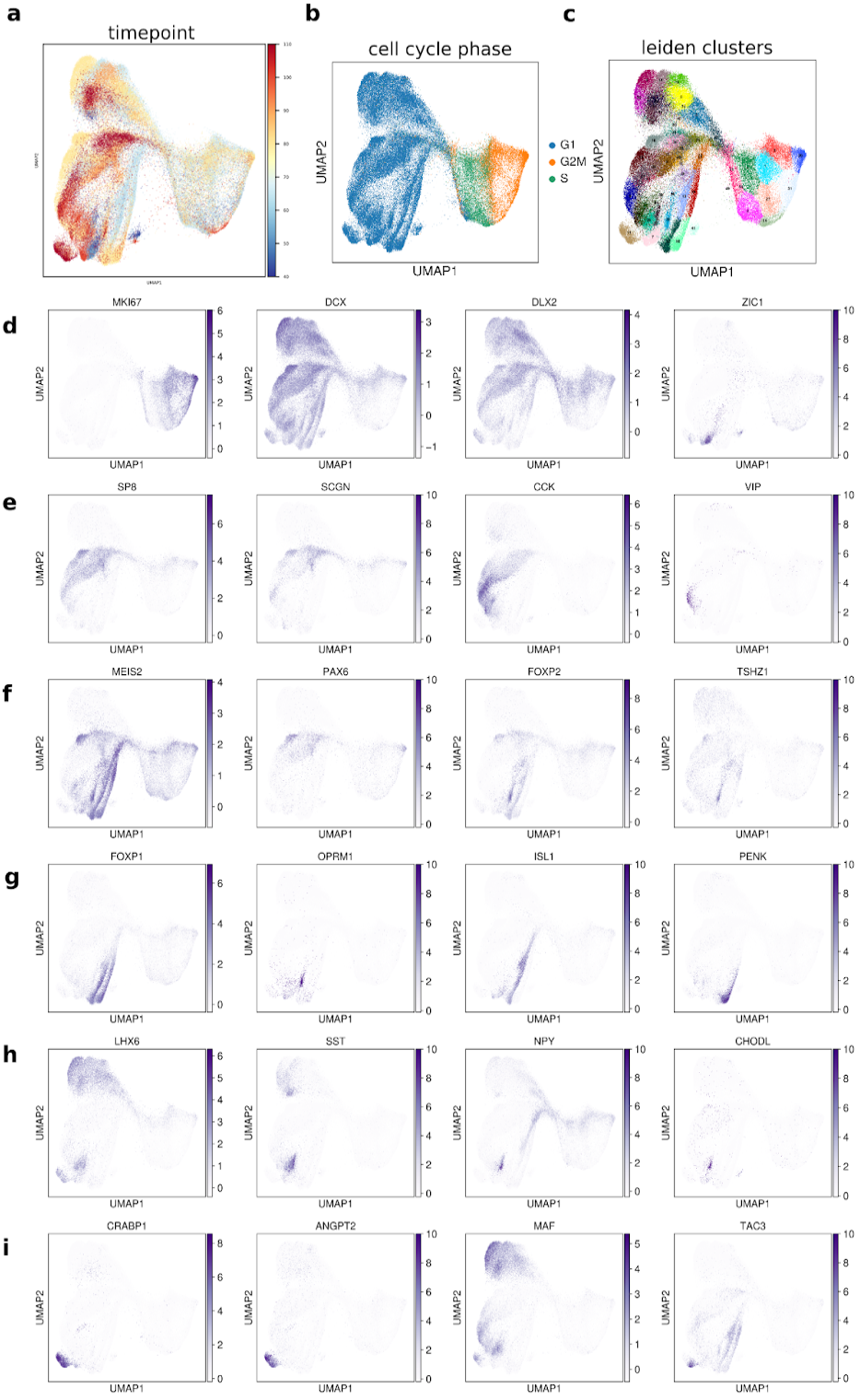
a
timepoint
b
cell cycle phase
G1
G2M
S
UMAP2
UMAP1
c
leiden clusters
UMAP2
UMAP1
d
MKI67
DCX
DLX2
ZIC1
e
SP8
SCGN
CCK
VIP
f
MEIS2
PAX6
FOXP2
TSHZ1
g
FOXP1
OPRM1
ISL1
PENK
h
LHX6
SST
NPY
CHODL
i
CRABP1
ANGPT2
MAF
TAC3

Figure 2.7: Macaque single cell RNAseq gene expression landscape.

a. Macaque scRNAseq UMAP colored by post conception day from which cells are sampled. **b.** UMAP projection colored by cell-cycle phase as classified by scanpy score_genes_cell_cycle function. **c.** UMAP projection colored by Leiden clusters. Although cell intrinsic differences within initial classes may predict their further subclass partitioning, the fine-grained Leiden clusters did not yield groups appearing to match terminal classes and mainly varied by neuronal differentiation trajectories. **d-i.** Scaled and normalized expression of **d,** dividing and newborn neuron marker genes **e,** CGE-derived neuron markers **f,** dLGE-derived neuron markers **g,** LGE-derived projection neuron markers **h,** MGE-derived cortical neuron markers **i,** MGE-derived striatal interneuron markers. An interactive browser for exploring the transcriptional features of inhibitory neuron development is available (https://dev-inhibitory-neurons.cells.ucsc.edu/).(Speir et al., 2021)

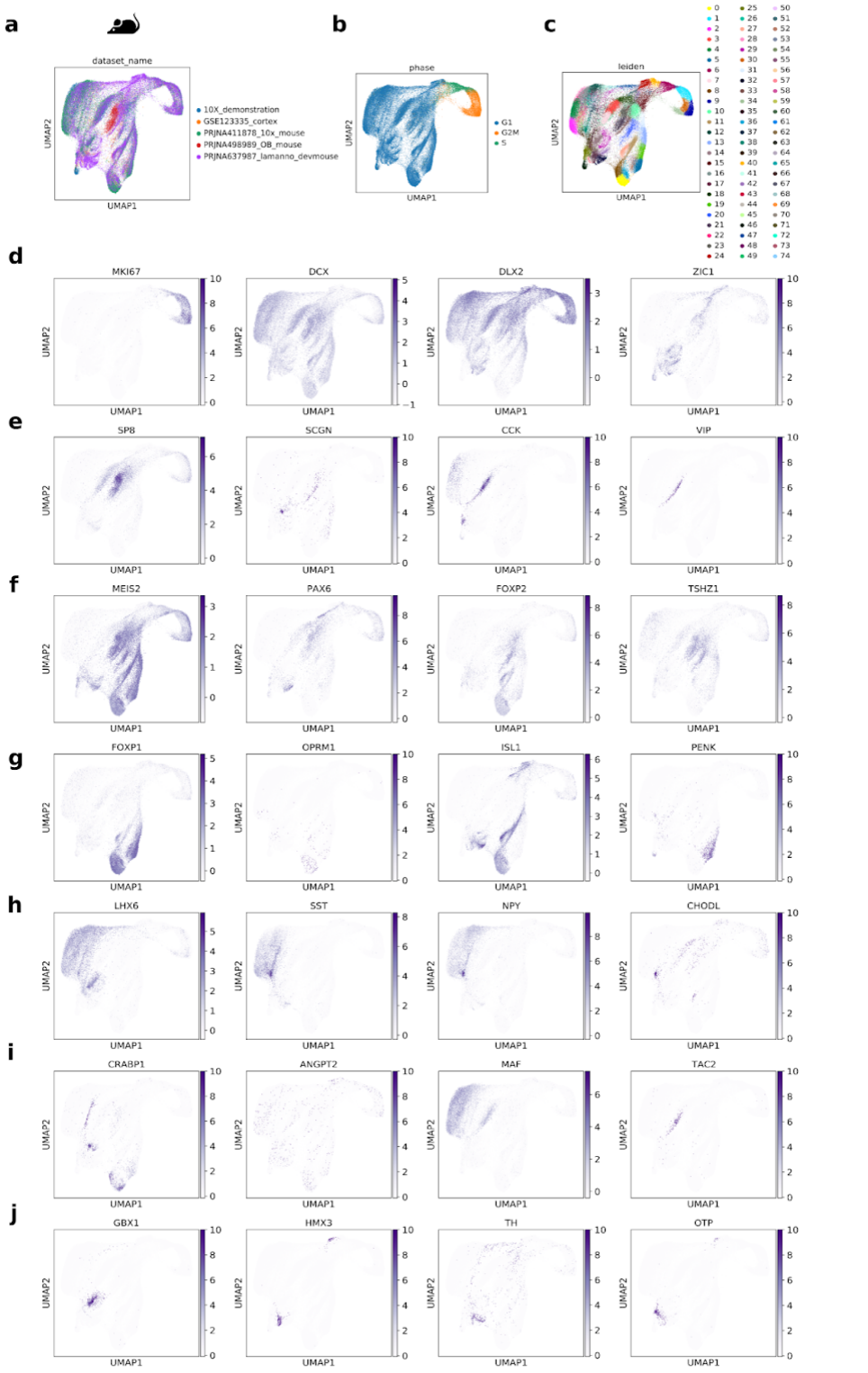
a
dataset_name
10X_demonstration
GSE123335_cortex
PRJNA411878_10x_mouse
PRJNA498989_OB_mouse
PRJNA637987_lamanno_devmouse
b
phase
G1
G2M
S
c
leiden
d
MKI67
DCX
DLX2
ZIC1
e
SP8
SCGN
CCK
VIP
f
MEIS2
PAX6
FOXP2
TSHZ1
g
FOXP1
OPRM1
ISL1
PENK
h
LHX6
SST
NPY
CHODL
i
CRABP1
ANGPT2
MAF
TAC2
j
GBX1
HMX3
TH
OTP
UMAP1
UMAP2

Figure 2.8: Mouse single cell RNAseq gene expression landscape.

a. Mouse scRNAseq UMAP colored by public dataset from which the cells are derived. **b.** UMAP projection colored by cell-cycle phase as classified by scanpy score_genes_cell_cycle function. **c.** UMAP projection colored by Leiden clusters. **d-i** Scaled and normalized expression of **d,** dividing and newborn neuron marker genes **e,** CGE-derived neuron markers **f,** dLGE-derived neuron markers **g,** LGE-derived projection neuron markers **h,** MGE-derived cortical neuron markers **i,** MGE-derived striatal interneuron markers **i,** VMF-derived markers.

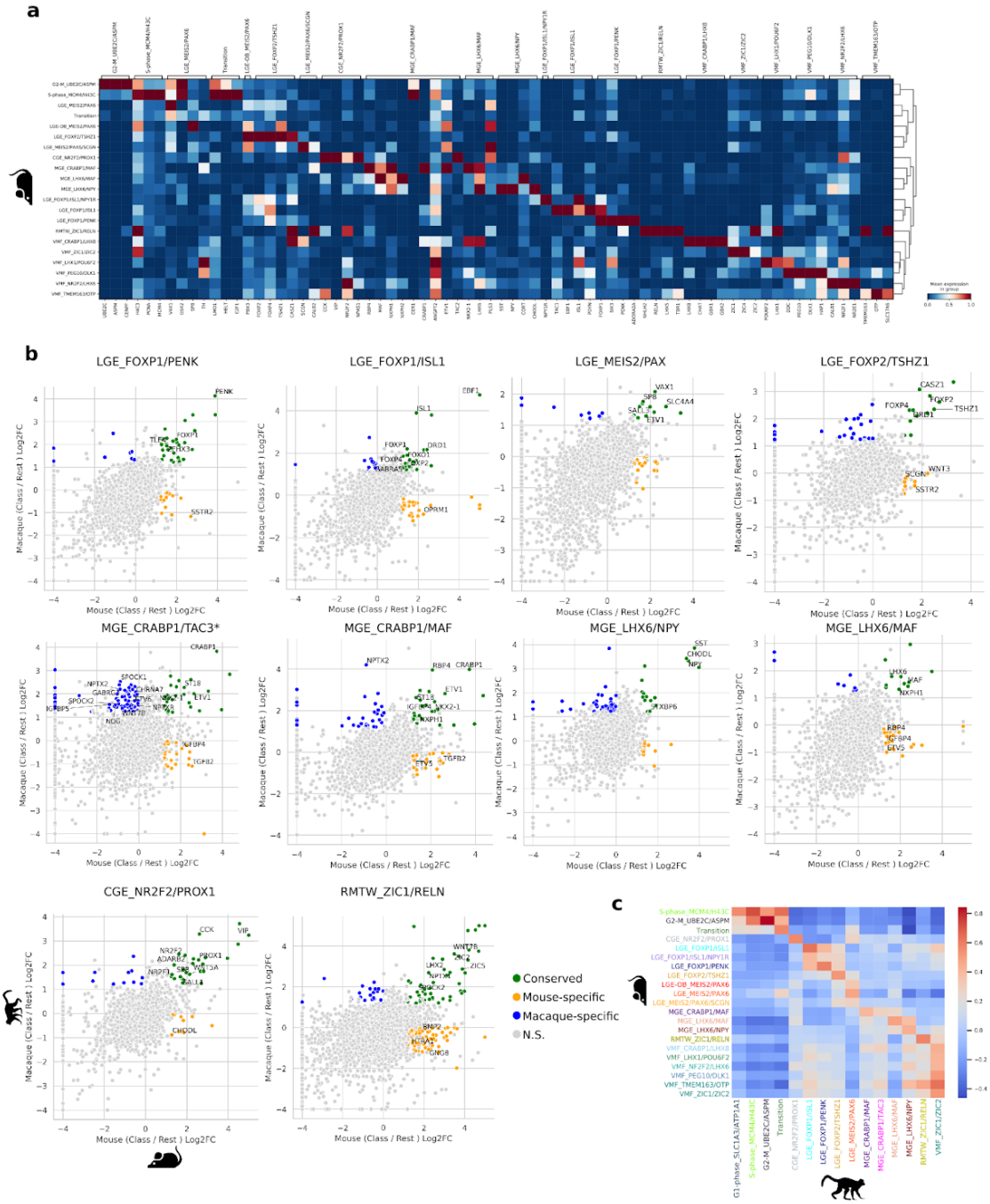
a
b
LGE_FOXP1/PENK
LGE_FOXP1/ISL1
LGE_MEIS2/PAX
LGE_FOXP2/TSHZ1
MGE_CRABP1/TAC3*
MGE_CRABP1/MAF
MGE_LHX6/NPY
MGE_LHX6/MAF
CGE_NR2F2/PROX1
RMTW_ZIC1/RELN
Mouse (Class / Rest) Log2FC
Macaque (Class / Rest) Log2FC
Conserved
Mouse-specific
Macaque-specific
N.S.
c

Figure 2.9: Markers of mouse and macaque initial classes.

a. Heatmap of mouse initial class marker genes selected from the top markers of each, scaled by column. **b.** Scatterplots of log 2 fold changes of each initial class vs the rest of the dataset for mouse vs macaque, with selected gene families of interest labeled. Conserved represents genes significantly upregulated in class vs rest in both species, Mouse-specific in mouse but not macaque, etc. N.S.= Not Significant in either species. Significance defined as |log2fcl>1.2 and adjusted p value < .01. MGE_CRABP1/TAC3* is the comparison of macaque MGE_CRABP1/TAC3 vs mouse MGE_CRABP1/MAF, as this is the ancestral class comparison. Note that more genes show specific correlations to the macaque MGE_CRABP1/TAC3 class versus the macaque MGE_CRABP1/MAF class in the comparison to the single mouse MGE_CRABP1/MAF class. **c.** Pairwise Pearson correlations of mean gene expression in classes across species.

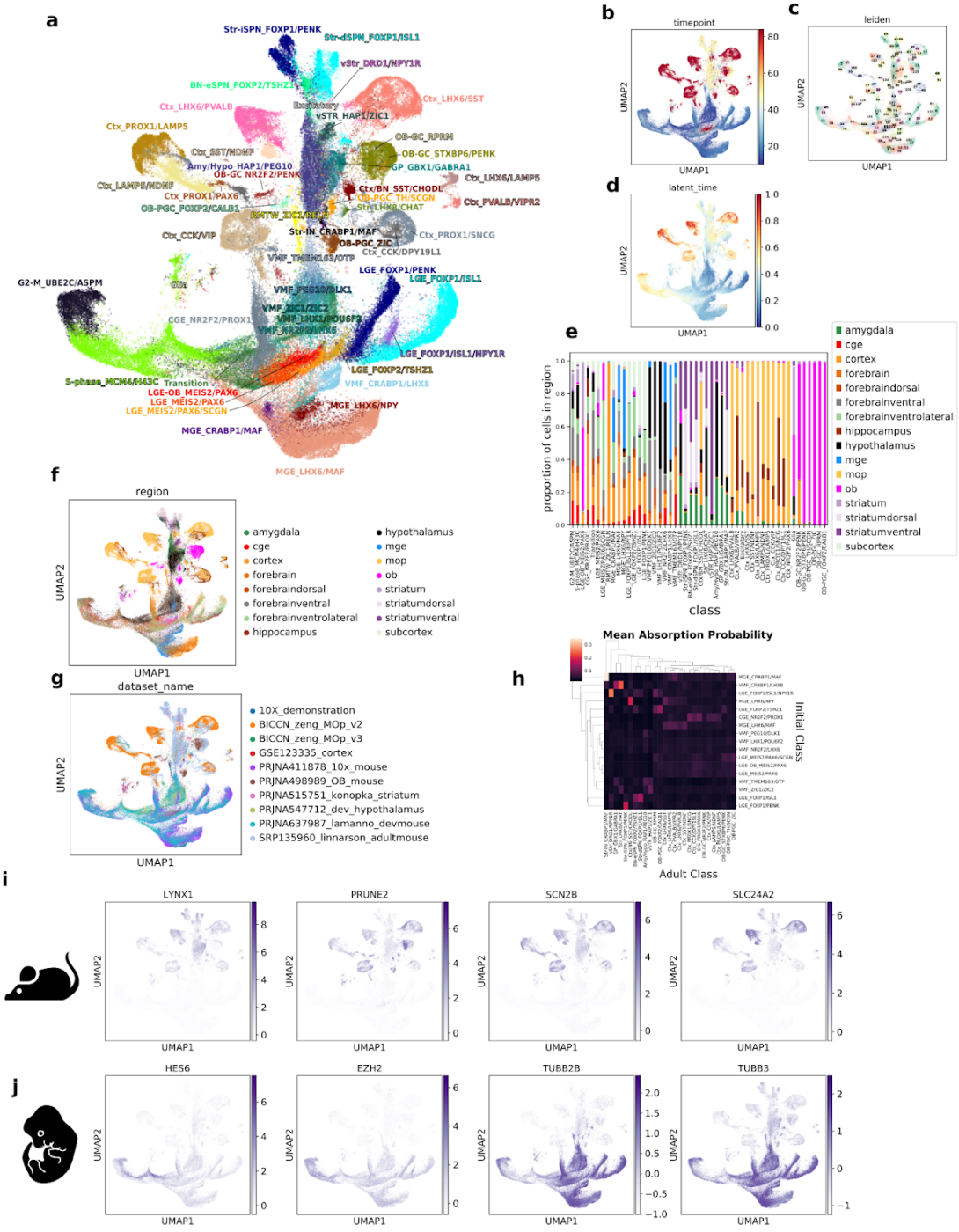
a
Str-iSPN_FOXP1/PENK
Str-dSPN_FOXP1/ISL1
vStr_DRD1/NPY1R
BN-eSPN_FOXP2/TSHZ1
Ctx_LHX6/SST
Ctx_LHX6/PVALB
Ependymy
vSTR_HAP1/ZIC1
Ctx_PROX1/LAMP5
OB-GC_RPRM
Ctx_SST/MDHF
OB-GC_STXBP6/PENK
Amy/Hypo_HAP1/PEG10
OB-GC_NR2F2/PENK
GP_GBX1/GABRA1
Ctx_LAMP5/MDHF
Ctx_LHX6/LAMP5
Ctx_PROX1/PAX6
Ctx/BN_SST/CHODL
OB-PGC_TH/SCGN
Ctx_PVALB/VIPR2
OB-PGC_FOXP2/CALB1
Str_LHX8/CHAT
vSTW_ZIC1/ISL1
Ctx_CCK/VIP
Str-IN_CRABP1/MAF
Ctx_PROX1/SNCG
OB-PGC_ZIC
Ctx_CCK/DPY19L1
VMF_TMEM163/OTP
LGE_FOXP1/PENK
LGE_FOXP1/ISL1
G2-M_UBE2C/ASPM
Glia
VMF_PEG10/DLK1
VMF_ZIC1/ZIC2
CGE_NR2F2/PROX1
VMF_LHX1/POU6F2
VMF_NR2F2/LHX6
LGE_FOXP1/ISL1/NPY1R
LGE_FOXP2/TSHZ1
S-phase_MCM6/H43C
Transition
VMF_CRABP1/LHX8
LGE-OB_MEIS2/PAX6
LGE_MEIS2/PAX6
LGE_MEIS2/PAX6/SCGN
MGE_LHX6/NPY
MGE_CRABP1/MAF
MGE_LHX6/MAF
b
timepoint
UMAP2
UMAP1
c
leiden
UMAP2
UMAP1
d
latent_time
UMAP2
UMAP1
e
proportion of cells in region
class
amygdala
cge
cortex
forebrain
forebraindorsal
forebrainventral
forebrainventrolateral
hippocampus
hypothalamus
mge
mop
ob
striatum
striatumdorsal
striatumventral
subcortex
f
region
UMAP2
UMAP1
amygdala
cge
cortex
forebrain
forebraindorsal
forebrainventral
forebrainventrolateral
hippocampus
hypothalamus
mge
mop
ob
striatum
striatumdorsal
striatumventral
subcortex
g
dataset_name
UMAP2
UMAP1
10X_demonstration
BICCN_zeng_MOp_v2
BICCN_zeng_MOp_v3
GSE123335_cortex
PRJNA411878_10x_mouse
PRJNA498989_OB_mouse
PRJNA515751_konopka_striatum
PRJNA547712_dev_hypothalamus
PRJNA637987_lamanno_devmouse
SRP135960_linnarson_adultmouse
h
Mean Absorption Probability
Initial Class
Adult Class
i
LYNX1
PRUNE2
SCN2B
SLC24A2
UMAP2
UMAP1
j
HES6
EZH2
TUBB2B
TUBB3
UMAP2
UMAP1

Figure 2.10: Inhibitory neurons of the developing and adult mouse forebrain.

a. UMAP projection of developing and adult mouse single cell RNAseq data, with initial and terminal classes labeled. **b.** UMAP of all mouse data, labeled by the post conception timepoint from which the cells are derived. **c.** UMAP of all mouse data, labeled by Leiden clusters used to determine terminal classes. **d.** UMAP of all mouse data, labeled by the scVelo dynamical shared latent time of each cell. **e.** Normalized stacked barplot, showing proportion of total cells of each class from each region. **f.** UMAP of all mouse data, labeled by the region from which the cells are derived. **g.** UMAP of all mouse data, labeled by the public dataset from which cells are derived. **h.** Heatmap representing the mean absorption probabilities of cells in each initial class to each terminal class. **i.** Selected genes differentially expressed in all terminal classes over all initial classes. **j.** Selected genes differentially expressed in initial classes over terminal classes.

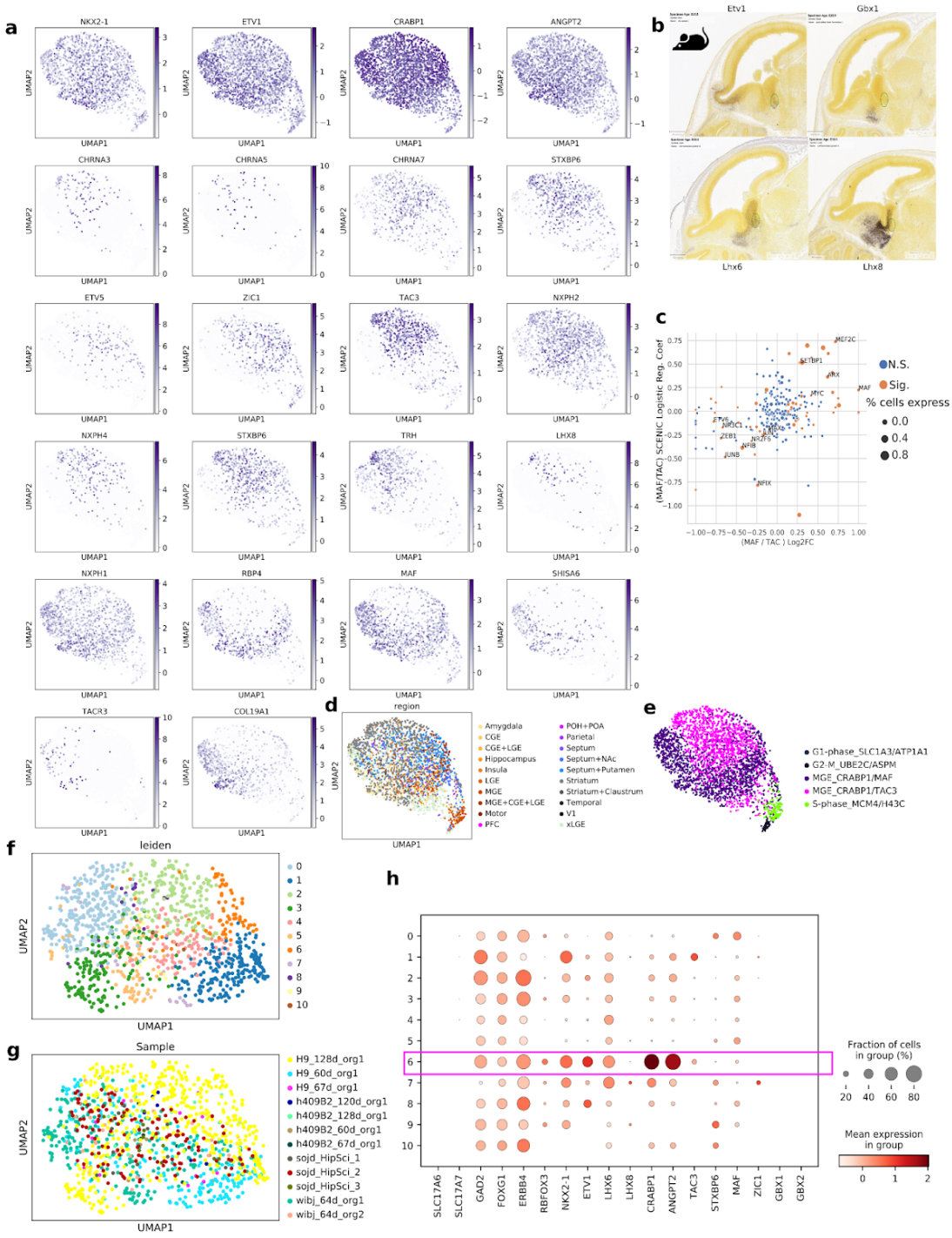
a
NKX2-1
ETV1
CRABP1
ANGPT2
CHRNA3
CHRNA5
CHRNA7
STXBP6
ETV5
ZIC1
TAC3
NXPH2
NXPH4
STXBP6
TRH
LHX8
NXPH1
RBP4
MAF
SHISA6
TACR3
COL19A1
UMAP1
UMAP2
b
Etv1
Gbx1
Lhx6
Lhx8
c
(MAF/TAC) SCENIC Logistic Reg. Coef
(MAF / TAC) Log2FC
N.S.
Sig.
% cells express
0.0
0.4
0.8
d
region
Amygdala
CGE
CGE+LGE
Hippocampus
Insula
LGE
MGE
MGE+CGE+LGE
Motor
PFC
POH+POA
Parietal
Septum
Septum+NAc
Septum+Putamen
Striatum
Striatum+Claustrum
Temporal
V1
xLGE
e
G1-phase_SLC1A3/ATP1A1
G2-M_UBE2C/ASPM
MGE_CRABP1/MAF
MGE_CRABP1/TAC3
S-phase_MCM4/H43C
f
leiden
g
Sample
H9_128d_org1
H9_60d_org1
H9_67d_org1
h409B2_120d_org1
h409B2_128d_org1
h409B2_60d_org1
h409B2_67d_org1
sojd_HipSci_1
sojd_HipSci_2
sojd_HipSci_3
wibj_64d_org1
wibj_64d_org2
h
SLC17A6
SLC17A7
GAD2
FOXG1
ERBB4
RBFOX3
NKX2-1
ETV1
LHX6
LHX8
CRABP1
ANGPT2
TAC3
STXBP6
MAF
ZIC1
GBX1
GBX2
Fraction of cells in group (%)
Mean expression in group

Figure 2.11: Expression of CRABP1+/TAC3+ and MAF+ striatal interneuron markers in developing macaque.

a. UMAP projection of NKX2-1/CRABP1+/ETV1+ cells only colored by scaled and normalized expression of TAC3 or MAF class markers. **b.** Allen Institute E15.5 mouse brain in situ hybridization showing expression of CRABP1+ neuron-related regional transcription factors. Green circle denotes boundary between *Lhx8+* MGE and rostroventral MGE/septum known to produce cholinergic neurons and the *Etv1+* MGE thought to produce CRABP1+ striatal interneurons, indicating partitioning of *Etv1* and *Lhx8* domains in mouse MGE. **c.** SCENIC module scores (Y-axis) vs log2 fold change of hub transcription factor predicted to regulate the module (X-axis). Significance represents multiple testing corrected q-value < 0.1 for both diffxpy differential expression in macaque and also q-value <0.1 SCENIC logistic regression coefficient q-value calculated by shuffling class labels. Size represents the proportion of all CRABP1+ cells which also express the gene. **d.** UMAP projection showing the region from which macaque cells are derived. **e.** UMAP projection showing classes in cells expressing 2 or more of (CRABP1, ETV1, ANGPT2). **f.** Subclustering of rare NKX2-1+ cells from organoid dataset(Kanton et al., 2019), labeled by Leiden subclusters. **g.** NKX2-1+ cells from organoid dataset(Kanton et al., 2019), labeled by the experimental conditions of the differentiation. **h.** Dotplot of expression of MGE_CRABP1 related markers in Leiden subclusters showing cluster 6 likely contains MGE_CRABP1/MAF and MGE_CRABP1/TAC3 cells.

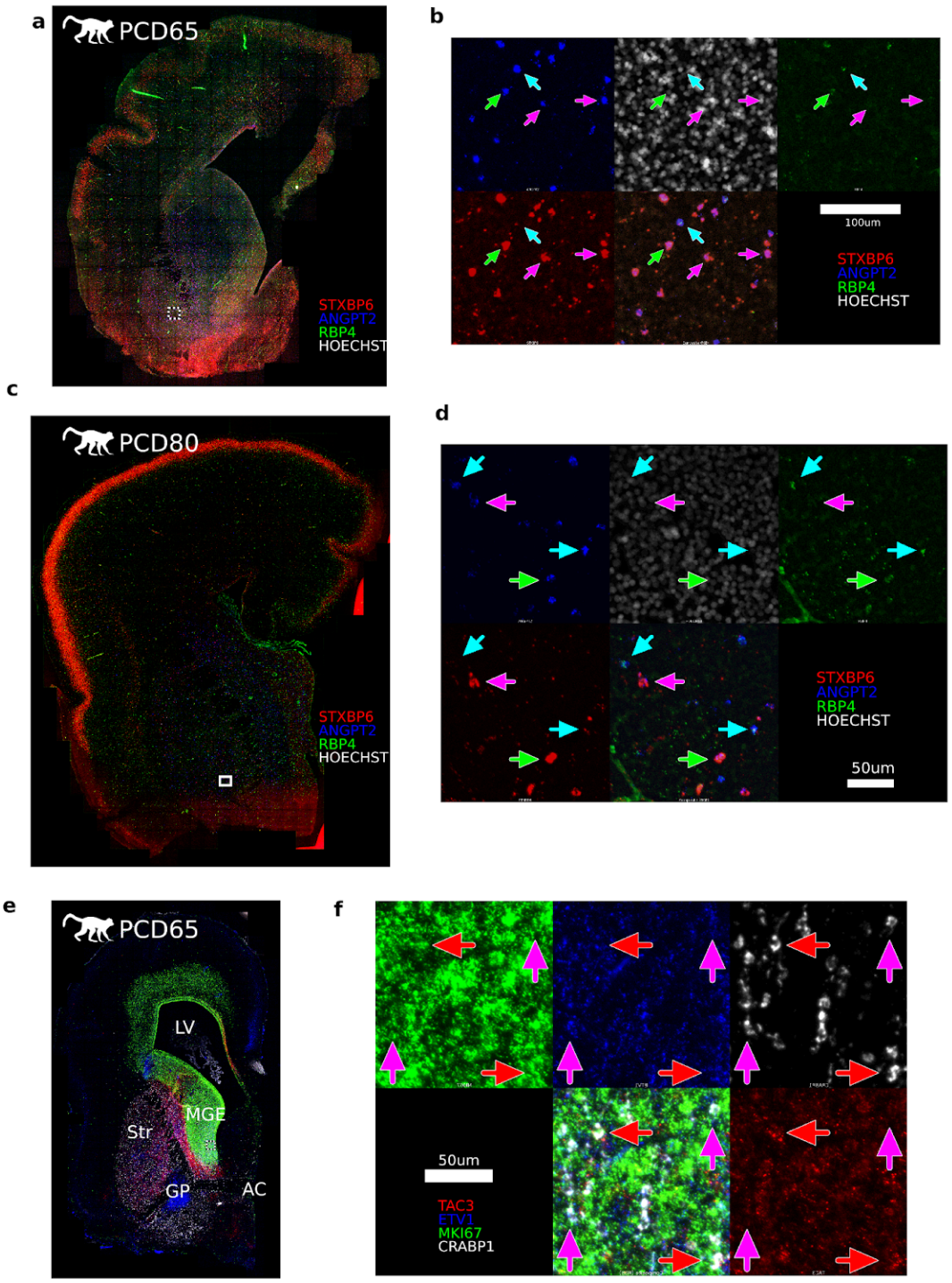
a
PCD65
STXBP6
ANGPT2
RBP4
HOECHST
b
100um
STXBP6
ANGPT2
RBP4
HOECHST
c
PCD80
STXBP6
ANGPT2
RBP4
HOECHST
d
STXBP6
ANGPT2
RBP4
HOECHST
50um
e
PCD65
LV
MGE
Str
GP
AC
f
50um
TAC3
ETV1
MKI67
CRABP1

Figure 2.12: Spatial distribution of CRABP1+/TAC3+ and MAF+ striatal interneuron markers expression.

a. PCD65 macaque brain section, from different individual than main text quantifications showing RNA expression of alternative markers of both CRABP1 classes (*ANGPT2*), the MGE_CRABP1/MAF class (*RBP4*) and the MGE_CRABP1/TAC3 class (*STXBP6*). Note that *RBP4* expression is rare at PCD65, but much more common at PCD80 (see **c-d**). **b.** Montage from **a** showing MGE_CRABP1/MAF (cyan arrows), MGE_CRABP1/TAC3 (magenta arrows) and STXBP6+/RBP4+ cells (green arrows). **c.** PCD80 macaque brain section showing RNA expression of alternative markers of both CRABP1 classes (*ANGPT2*), the MGE_CRABP1/MAF class (*RBP4*) and MGE_CRABP1/TAC3 (*STXBP6*). **d.** Montage from **c** showing MGE_CRABP1/MAF (cyan arrows), MGE_CRABP1/TAC3 (magenta arrows) and STXBP6+/RBP4+ cells (green arrows). **e.** PCD65 macaque brain section showing RNA expression of *CRABP1*, marking both CRABP1 classes, *ETV1,* marking both CRABP1 classes, the dLGE, and the GP, *TAC3* marking the MGE_CRABP1/TAC3 class and MKI67 marking dividing cells. Labeled regions are abbreviated LV: Lateral Ventricle, MGE: Medial Ganglionic Eminence, Str: Striatum, GP: Globus Pallidus, AC: Anterior Commissure **f.** Montage from **e** showing MGE_CRABP1/TAC3 (magenta arrows) and TAC3+/MKI67+ cells (red arrows).

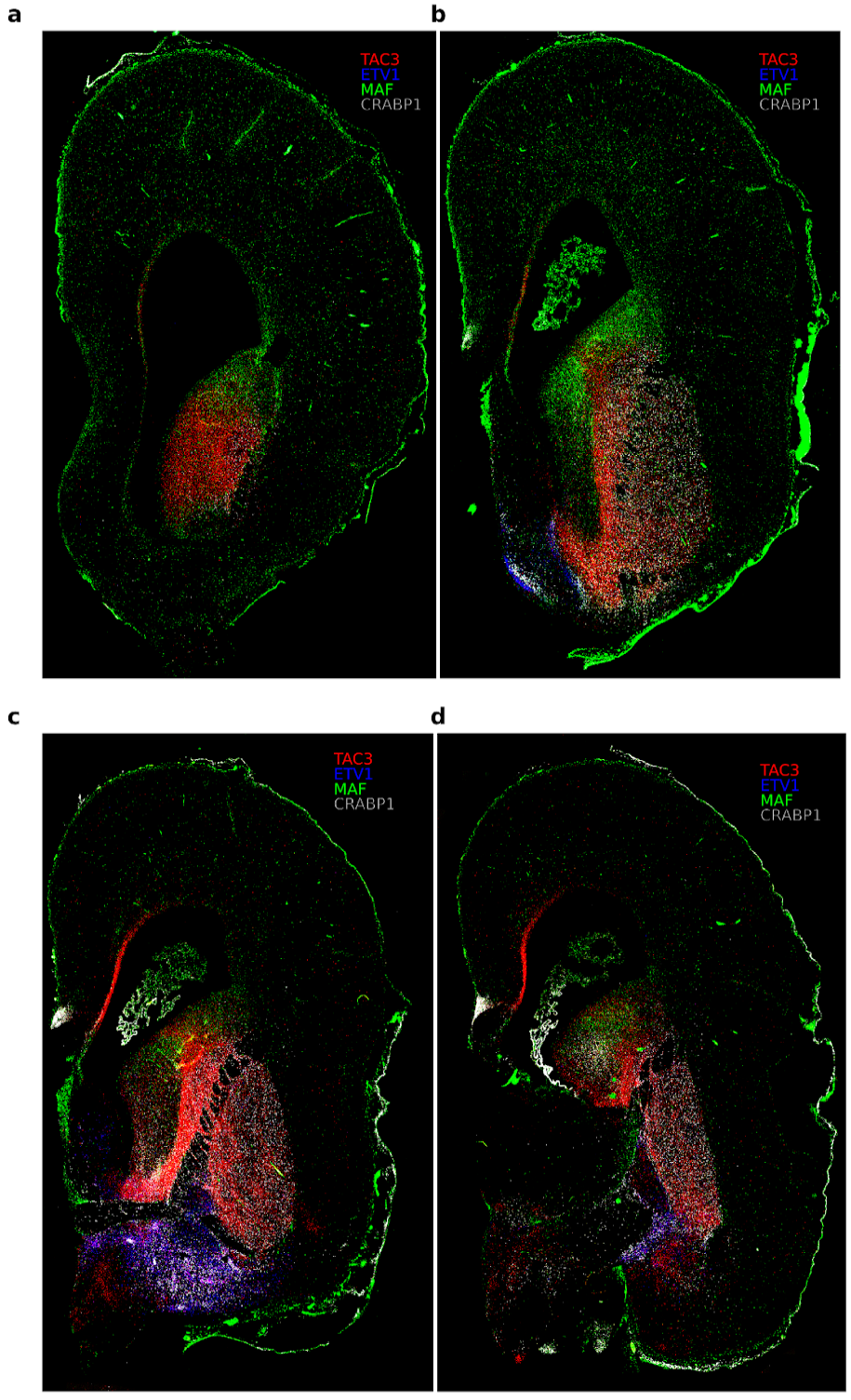
a
b
c
d
TAC3
ETV1
MAF
CRABP1

Figure 2.13: CRABP1+/TAC3+ and MAF+ striatal interneuron markers expression.

Full size tile scanned representative images from Figure 3c. **a-d.** Sections 0, 37, 66, and 101 four color in situ hybridization for CRABP1, LHX8, MAF, TAC3.

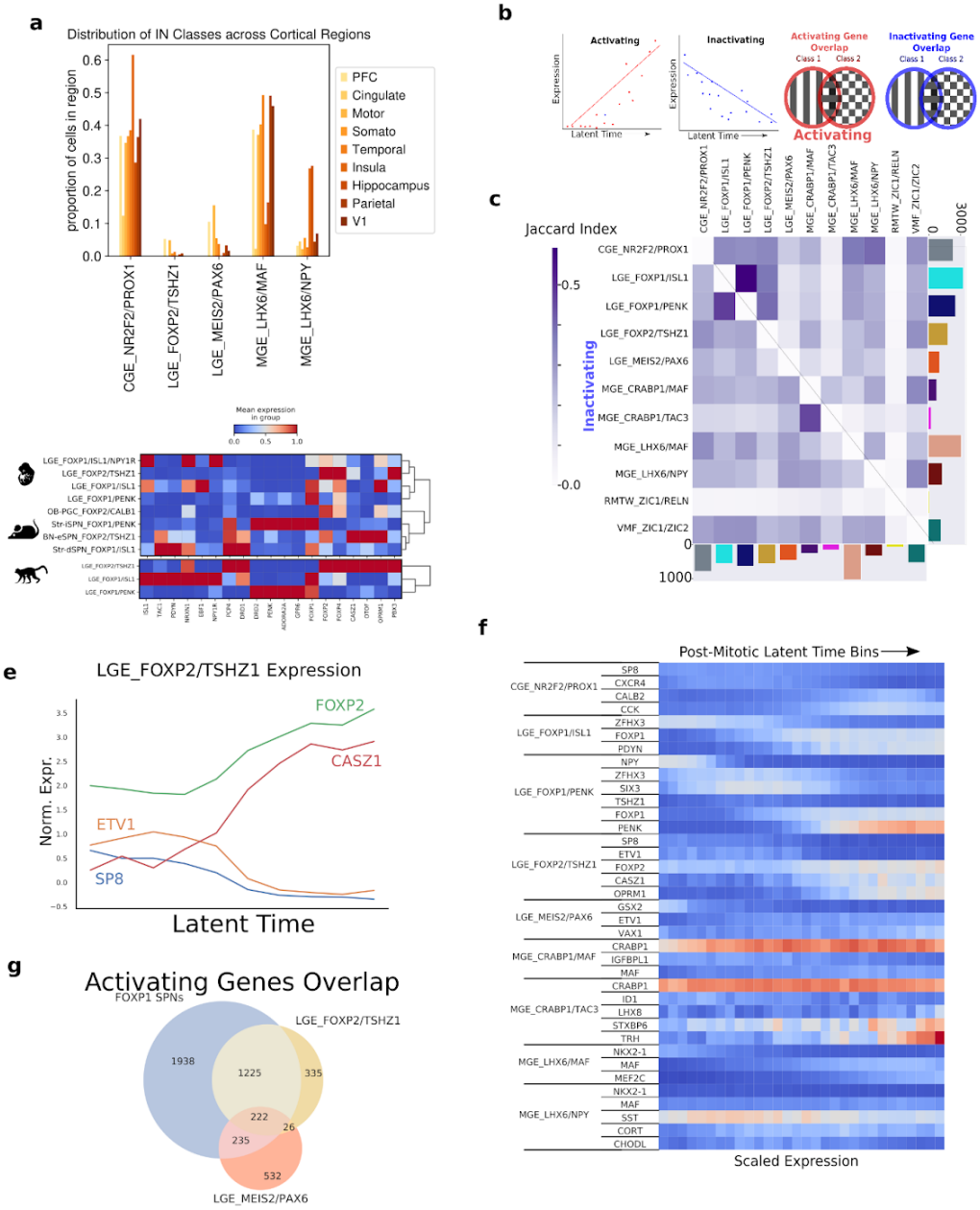
a
Distribution of IN Classes across Cortical Regions
proportion of cells in region
PFC
Cingulate
Motor
Somato
Temporal
Insula
Hippocampus
Parietal
V1
CGE_NR2F2/PROX1
LGE_FOXP2/TSHZ1
LGE_MEIS2/PAX6
MGE_LHX6/MAF
MGE_LHX6/NPY
b
Activating
Inactivating
Activating Gene Overlap
Inactivating Gene Overlap
Expression
Latent Time
Class 1
Class 2
Activating
c
Jaccard Index
Inactivating
d
Mean expression in group
LGE_FOXP1/ISL1/NPY1R
LGE_FOXP2/TSHZ1
LGE_FOXP1/ISL1
LGE_FOXP1/PENK
OB-PGC_FOXP2/CALB1
Str-iSPN_FOXP1/PENK
BN-eSPN_FOXP2/TSHZ1
Str-dSPN_FOXP1/ISL1
e
LGE_FOXP2/TSHZ1 Expression
FOXP2
CASZ1
ETV1
SP8
Norm. Expr.
Latent Time
f
Post-Mitotic Latent Time Bins
Scaled Expression
g
Activating Genes Overlap
FOXP1 SPNs
LGE_FOXP2/TSHZ1
LGE_MEIS2/PAX6
1938
1225
335
222
26
235
532

Figure 2.14: Spatial, temporal, and molecular distinctions among initial LGE-derived neurons.

a. Bar plot of the proportion of cells from each initial class across cortical regions for CGE-, LGE- and MGE-derived classes highlights frontal lobe enrichment of LGE classes during sampling. **b.** Schematic for **c**, showing cartoon scatter plots for the expression of two genes in individual cells regressed against latent time (left) to describe genes that activate or inactivate during neuronal differentiation, followed by Venn diagrams showing the overlap of dynamically-regulated genes between pairs of initial classes used to calculate Jaccard indices. **c.** Heatmap of Jaccard indices of significantly overlapping lists of dynamic genes between cell classes of Holm-Šídák corrected linear regression q value < 0.05 on scaled and normalized gene expression to address the extent of shared versus cell type-specific trajectories during post-mitotic differentiation of initial classes. Bar graphs on axes represent the total number of significant genes activating/inactivating. Direct and indirect medium spiny neurons (LGE_FOXP1/ISL1 and LGE_FOXP1/PENK) show strong overlap of both activating and inactivating genes, despite early partitioning as distinct initial classes, while presumed eccentric spiny neurons from LGE_FOXP2/TSHZ1 show strong overlap of activating genes (see also **g**) but not inactivating genes. Conversely, MGE_CRABP1/MAF and MGE_CRABP1/TAC3 classes show strong overlap of inactivating genes among a smaller overall set of dynamically-regulated genes, but little overlap in activating genes. **d.** Heatmap of selected marker expression of LGE-derived striatal initial and terminal classes, scaled by gene. **e.** Gene expression of dynamic dLGE marker genes in macaque LGE_FOXP2/TSHZ1 cells across shared latent time, ordered by latent time value and divided into 10 equally sized bins to provide stable mean expression values. **f.** Gene expression of dynamic marker genes in macaque cells across shared latent time, grouped by initial class, ordered by latent time value and divided into 30 equally sized bins to provide stable mean expression values. **g.** Venn diagram of intersections of significantly activating gene sets in LGE initial classes. In contrast to LGE_MEIS2/PAX6, the LGE_FOXP2/TSHZ1 class activates a large set of shared SPN genes during neuronal differentiation as inferred by latent time trajectories.

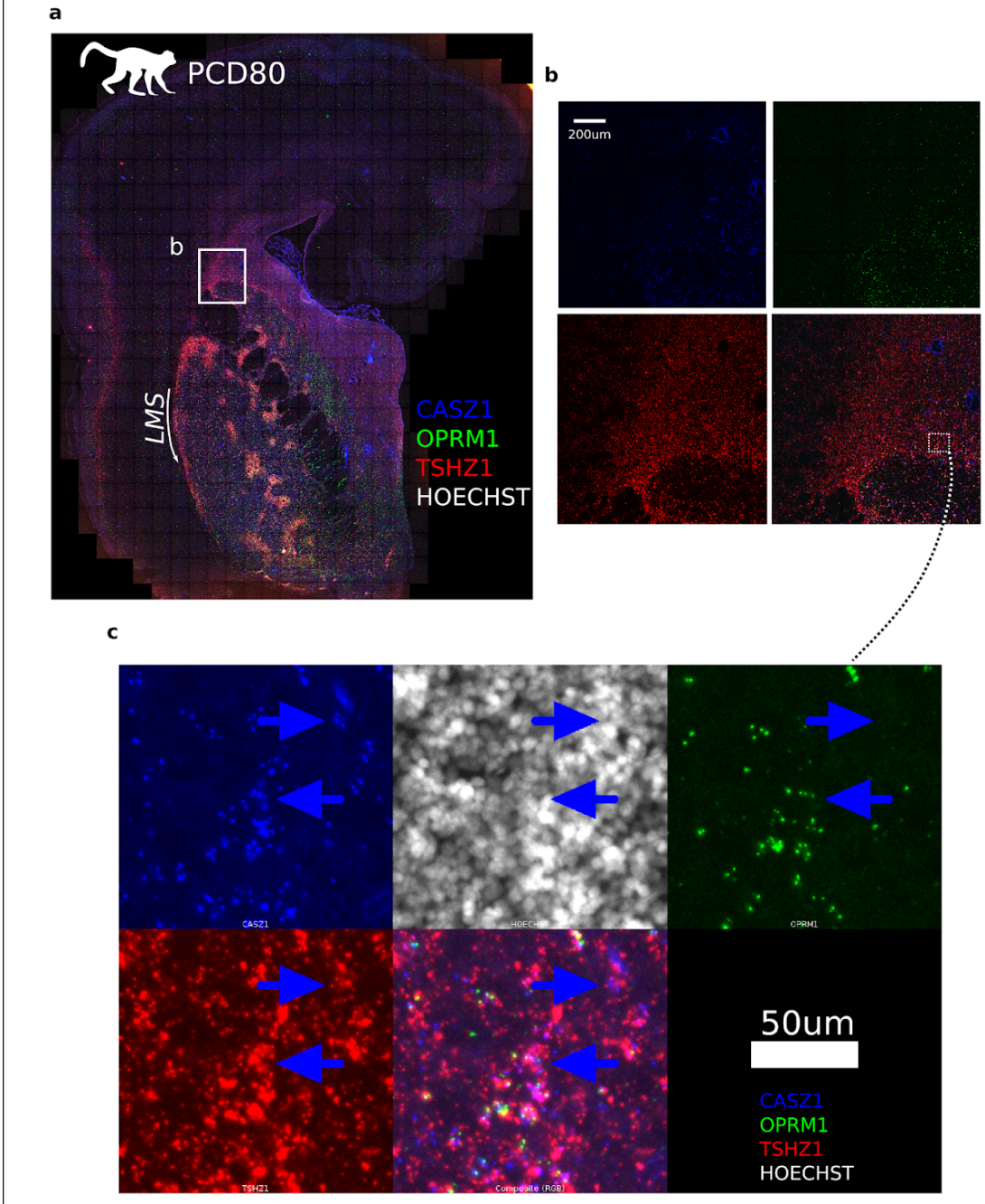
a
PCD80
b
LMS
CASZ1
OPRM1
TSHZ1
HOECHST
b
200um
c
50um
CASZ1
OPRM1
TSHZ1
HOECHST

Figure 2.15: Emergence of eSPN from LGE_FOXP2/TSHZ1 in the dLGE.

a. PCD80 macaque coronal section showing RNA expression of eSPN markers. The Lateral Migratory Stream (LMS) is noted as is shown in Kuerbitz et al.(Kuerbitz et al., 2018) **b.** Montage of magnified dLGE at the striatum-GE boundary. **c.** Montage from the box in **b** showing TSHZ1/CASZ1/OPRM1- (top arrow) and TSHZ1/CASZ1/OPRM1+ (bottom arrow) Str-eSPN_FOXP2/TSHZ1 cells (blue arrows) within the GE.

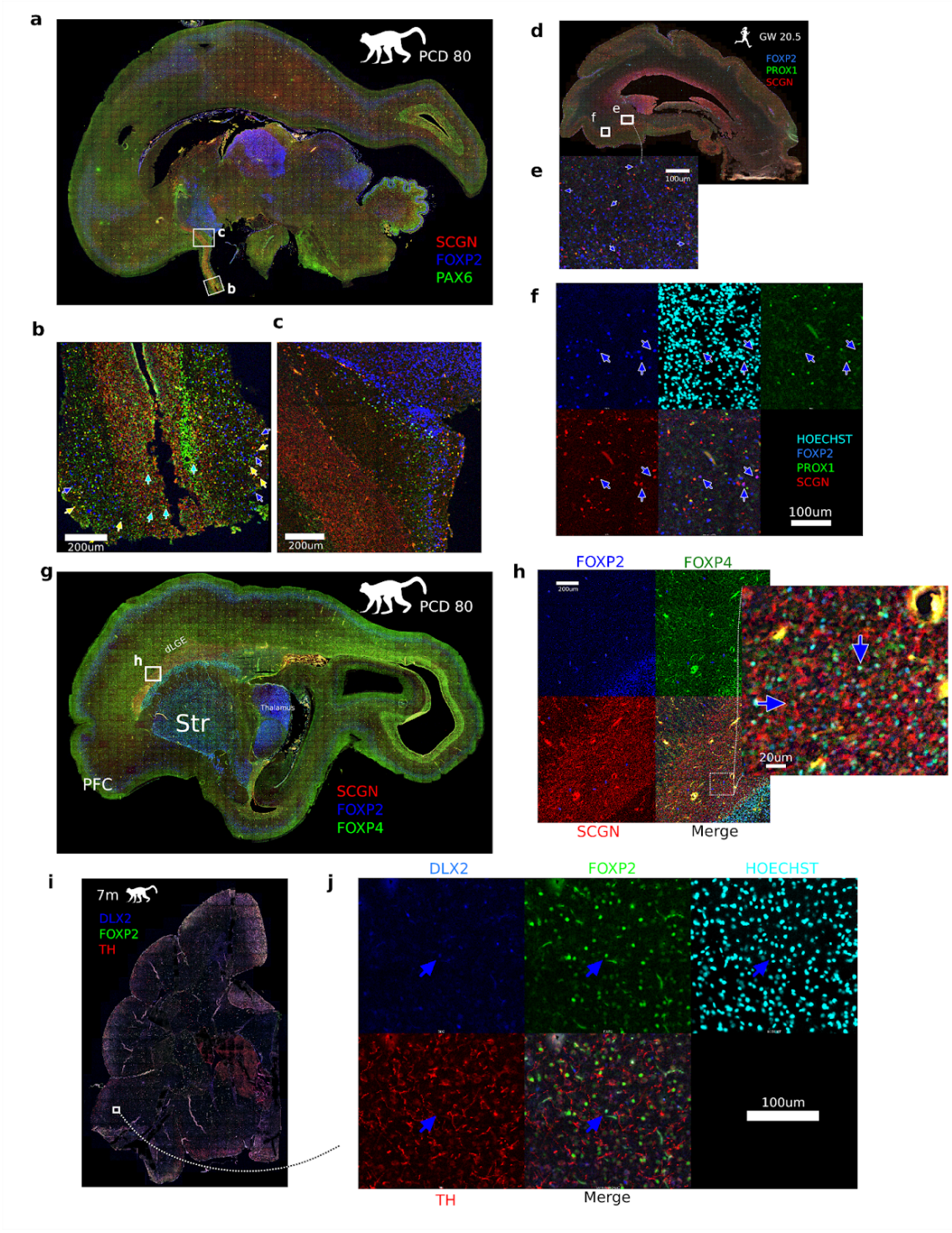
a
PCD 80
SCGN
FOXP2
PAX6
b
c
200um
200um
d
GW 20.5
FOXP2
PROX1
SCGN
e
f
100um
f
HOECHST
FOXP2
PROX1
SCGN
100um
g
PCD 80
dLGE
h
Str
Thalamus
PFC
SCGN
FOXP2
FOXP4
h
FOXP2
FOXP4
200um
20um
SCGN
Merge
i
7m
DLX2
FOXP2
TH
j
DLX2
FOXP2
HOECHST
100um
TH
Merge

Figure 2.16: Distribution of dLGE-derived LGE_FOXP2/TSHZ1 precursors in the superficial white matter.

a. Medial sagittal section of PCD80 macaque brain. The SCGN+ RMS originating at the anterior pole of the dLGE is seen extending from the olfactory ventricle to the OB. **b.** FOXP2+/PAX6+ cells from lateral migratory streams converge with RMS and enter periglomerular layers of OB (see also **Figure 2.6**). Note that FOXP2+ OB-PGC_FOXP2/CALB1 cells are largely absent from the RMS but are found ventral of the nucleus accumbens (NAc), anterior olfactory nucleus (AON) and in outer olfactory tract sheath. **c.** SCGN+/PAX6+ granule cells (OB-GC_MEIS2/PAX6) (cyan arrows), TH+ PGCs (OB-PGC_TH/SCGN) (yellow arrows) and FOXP2+ PGCs (OB-PGC_FOXP2/CALB1) (blue arrows) in OB. **d-f.** Human gestation week 20.5 sagittal cortex section shows newborn FOXP2+/SCGN+/PROX1- neurons (blue arrows) migrating into the ventral cortex superficial white matter. SCGN+ expression decreases as cells mature. **g.** Lateral sagittal section of PCD80 macaque brain. **h.** Immunofluorescence FOXP2+/FOXP4+/SCGN+ dLGE-derived projection class neurons (blue arrows) are seen in large numbers in the dLGE portion dorsal of the caudate, and in adjacent cortical white matter and striatum. **i.** 7 month old macaque coronal section. **j.** Montage from box in **i** with rare DLX2/FOXP2+ superficial white matter IN (SWMIN) marked with a blue arrow.

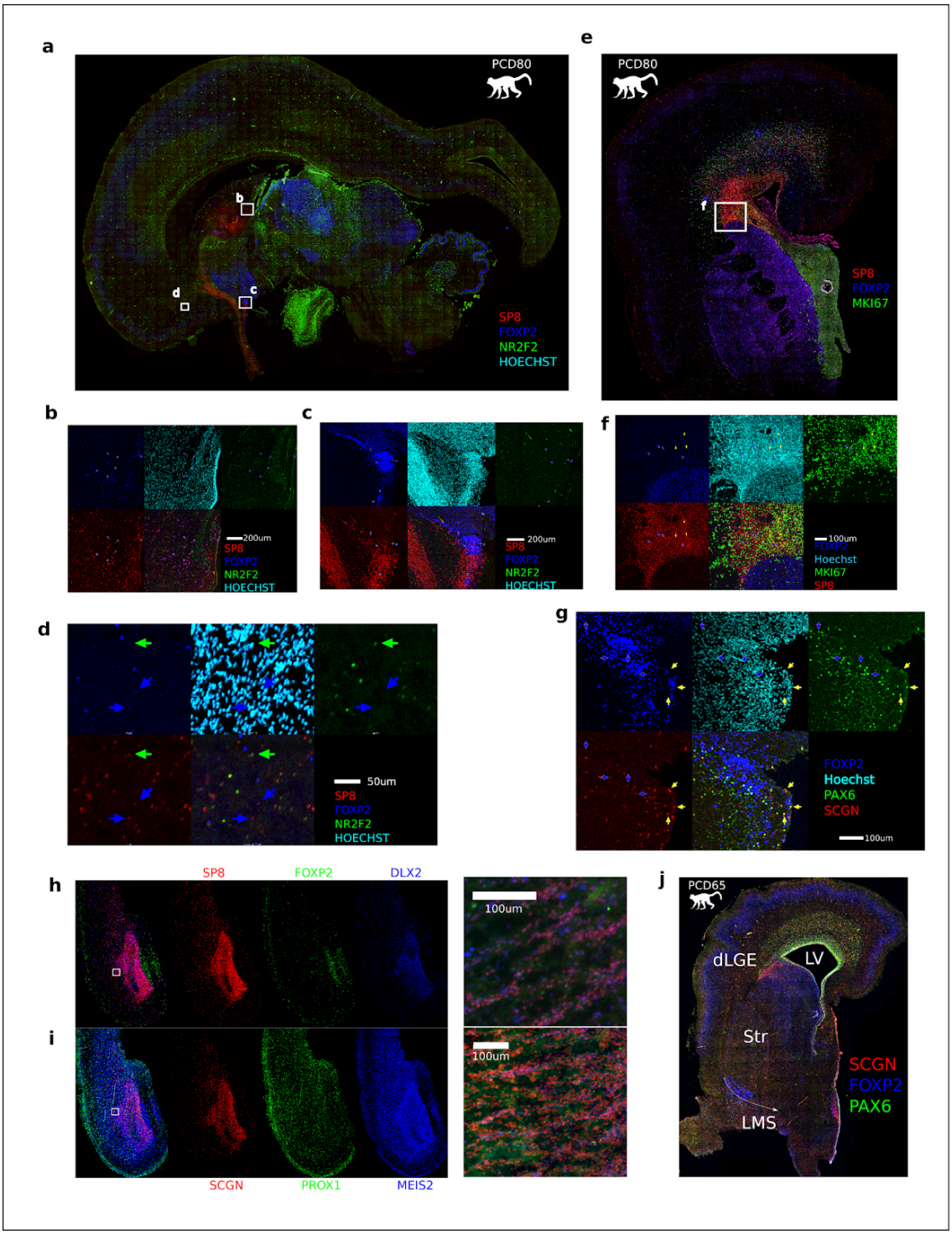
a
PCD80
b
c
d
SP8
FOXP2
NR2F2
HOECHST
e
PCD80
f
SP8
FOXP2
MKI67
b
200um
SP8
FOXP2
NR2F2
HOECHST
c
200um
SP8
FOXP2
NR2F2
HOECHST
f
100um
FOXP2
Hoechst
MKI67
SP8
d
50um
SP8
FOXP2
NR2F2
HOECHST
g
FOXP2
Hoechst
PAX6
SCGN
100um
h
SP8
FOXP2
DLX2
100um
i
100um
SCGN
PROX1
MEIS2
j
PCD65
dLGE
LV
Str
LMS
SCGN
FOXP2
PAX6

Figure 2.17: dLGE migration streams.

a-d. PCD 80 macaque brain sagittal section. Blue arrows represent SP8+/FOXP2+ cells in **(b)** septum and fornix (it is unclear whether these cells are born here, or arrive via an RMS dorsal extension or via a cortex-indusium griseum-fornix route, possibly seen in (Alzu'bi and Clowry, 2020)) **(c)** anterior olfactory nucleus (these cells appear to be near the point where the LMS is converging with the RMS, suggesting a lateral source of FOXP2+ PGCs) and **(d)** vmPFC. **e-f.** Coronal section of macaque PCD80 brain shows large numbers of SP8+/FOXP2+ cells and SP8+/MKI67+ cells in DL-dLGE. **g.** Sagittal section of anterior olfactory nucleus. Newborn dLGE-derived PAX6+/SCGN+/FOXP2- labeled with yellow arrows. **h.** Oblique coronal-axial section of dLGE and PFC with insets highlighting DLX2+/SP8+/FOXP2- parenchymal chains. **i.** Oblique coronal-horizontal section of dLGE and PFC with insets highlighting MEIS2+/SCGN+/PROX1- parenchymal chains. **j.** Coronal section of PCD65 macaque brain with lateral ventricle (LV), striatum (Str), dLGE and approximate lateral migratory stream (LMS) labeled.

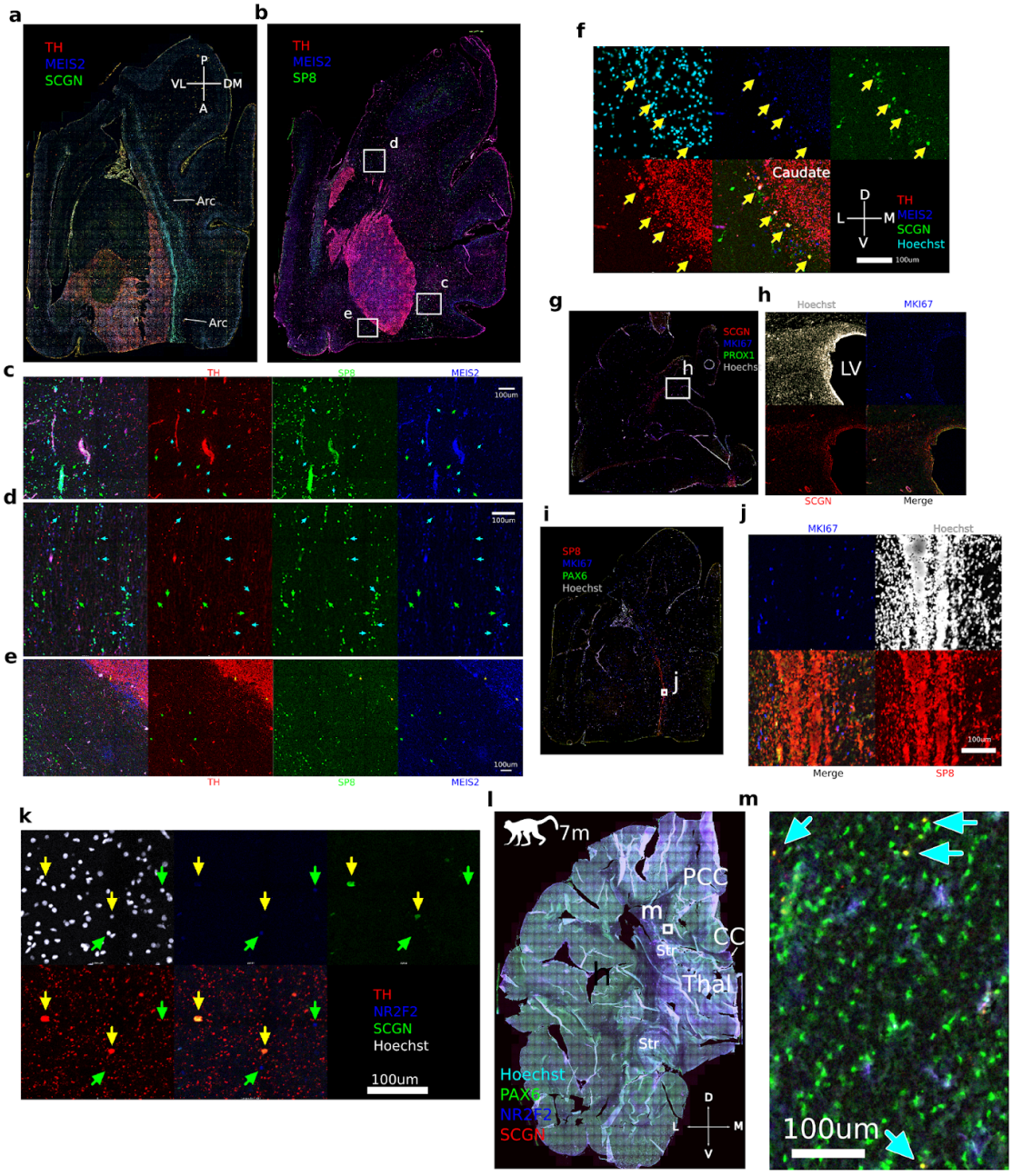
a
TH
MEIS2
SCGN
P
VL
DM
A
Arc
Arc
b
TH
MEIS2
SP8
d
c
e
f
Caudate
D
L
M
V
TH
MEIS2
SCGN
Hoechst
100um
g
SCGN
MKI67
PROX1
Hoechs
h
h
Hoechst
MKI67
LV
SCGN
Merge
c
TH
SP8
MEIS2
100um
d
100um
e
100um
TH
SP8
MEIS2
i
SP8
MKI67
PAX6
Hoechst
j
j
MKI67
Hoechst
100um
Merge
SP8
k
TH
NR2F2
SCGN
Hoechst
100um
l
7m
PCC
m
CC
Str
Thal
Str
Hoechst
PAX6
NR2F2
SCGN
D
L
M
V
m
100um

Figure 2.18: A-dLGE cells in the Arc.

a. Low magnification stitching of oblique horizontal section showing large stream of MEIS2+/SCGN+ chains in the Arc along the dorsomedial edge of the TH+ striatum. **b.** Further dorsal oblique horizontal section showing dorsal and lateral migratory streams. **c.** Enlargement of Arc-ACC. Cyan arrows denote MEIS2+/SCGN+ dLGE cells. Note that the stream is continually bounded by the increased density of TH+ fibers. **d.** Chains caudal of the striatum are of mixed classes. **e.** White matter neurons lateral of the striatum are nearly all MEIS2-negative (green arrows). Also note peristriatal SP8+/MEIS2+/TH+ *striatum laureatum* neurons (SLNs) at lateral border of striatum (yellow arrows). **f.** Coronal section of PCD120 macaque striatum showing an array of TH+ peristriatal SLNs(yellow arrows) at the edge of the caudate nucleus, from **Figure 2.3**, at the same location as 7 month postnatal, but not yet having developed SCGN+/TH+ processes tangential to the external capsule. **g-h.** Coronal section of PCD120. Sparse MKI67+ cells at ventricle, with SCGN+ cells away from the ventricle being MKI67-. **i,j.** Chains in Arc do not appear to be MKI67+. **k.** PCD120 SCGN/TH+ peristriatal SLNs (yellow arrows) are NR2F2 negative (NR2F2+ DWMINs labeled with green arrows). Note that the LGE_MEIS2/PAX6 and LGE_FOXP2/TSHZ1 classes very sparsely expressed the transcript encoding tyrosine hydroxylase (TH), a rate limiting enzyme in dopamine production, and were the only cortical IN classes to do so at developmental stages (see **Figure 2.4**). **l.** Coronal section of macaque cortex at 7 months. **m.** PAX6+/SCGN+/NR2F2- deep white matter neurons in postnatal macaque cingulate cortex (cyan arrows)**.** These DWMINs were found in the cingulate white matter and corona radiata, though not in the corpus callosum itself and rarely near the deep layers of the cortical plate and external capsule white matter

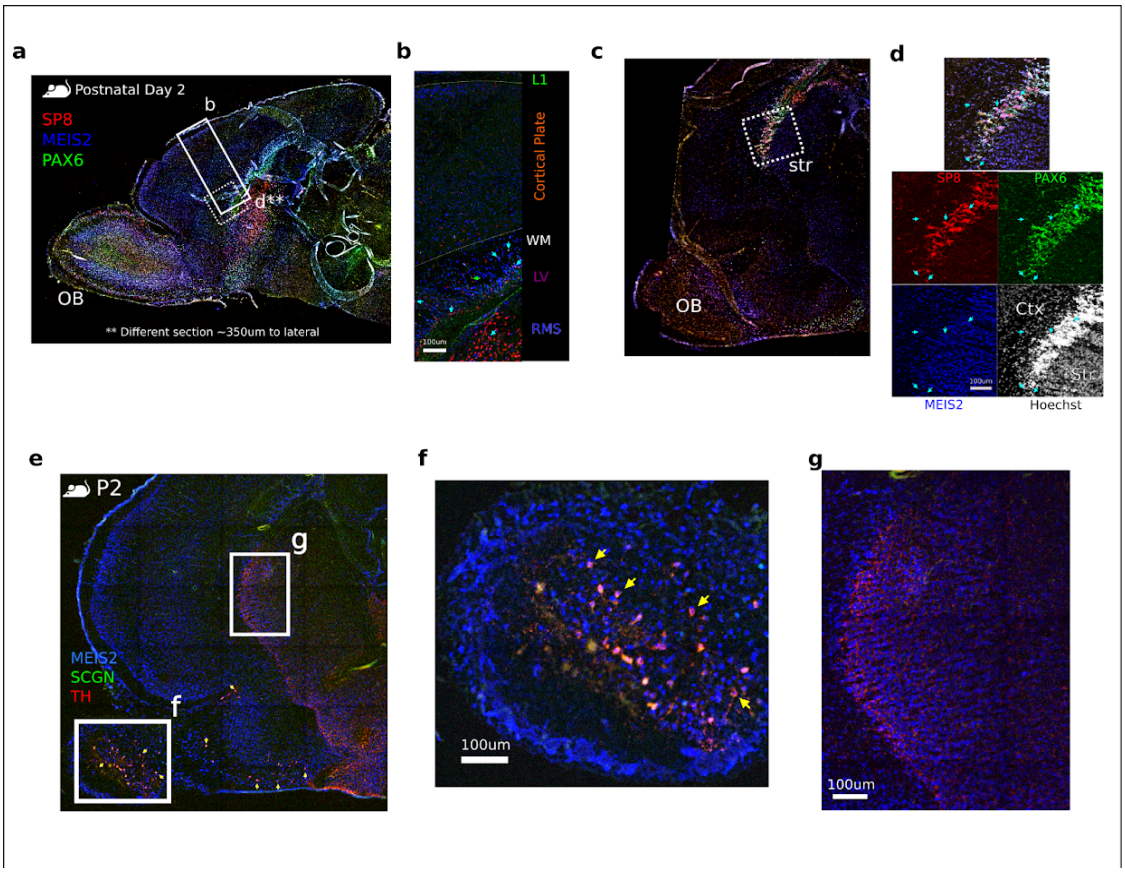

Figure 2.19: Distribution of LGE_MEIS2/PAX6-derived cells in postnatal mouse.

a. Mouse sagittal section showing **b** and approximate **c** magnification locations. **b.** LGE_MEIS2/PAX6 cells in deep white matter (cyan arrows) **c.** Lateral sagittal section **d.** Panel showing SP8+/MEIS2+/PAX6+ cells in remainder of dLGE, likely homologous to dLGE chains in Arc. **e.** Lateral sagittal section of mouse postnatal day 2. **f.** MEIS2+/SCGN+/TH+ periglomerular cells in lateral OB. **g.** Striatum shown with dense TH+ projection fibers and synapses, but no MEIS2+/SCGN+/TH+ cell bodies.

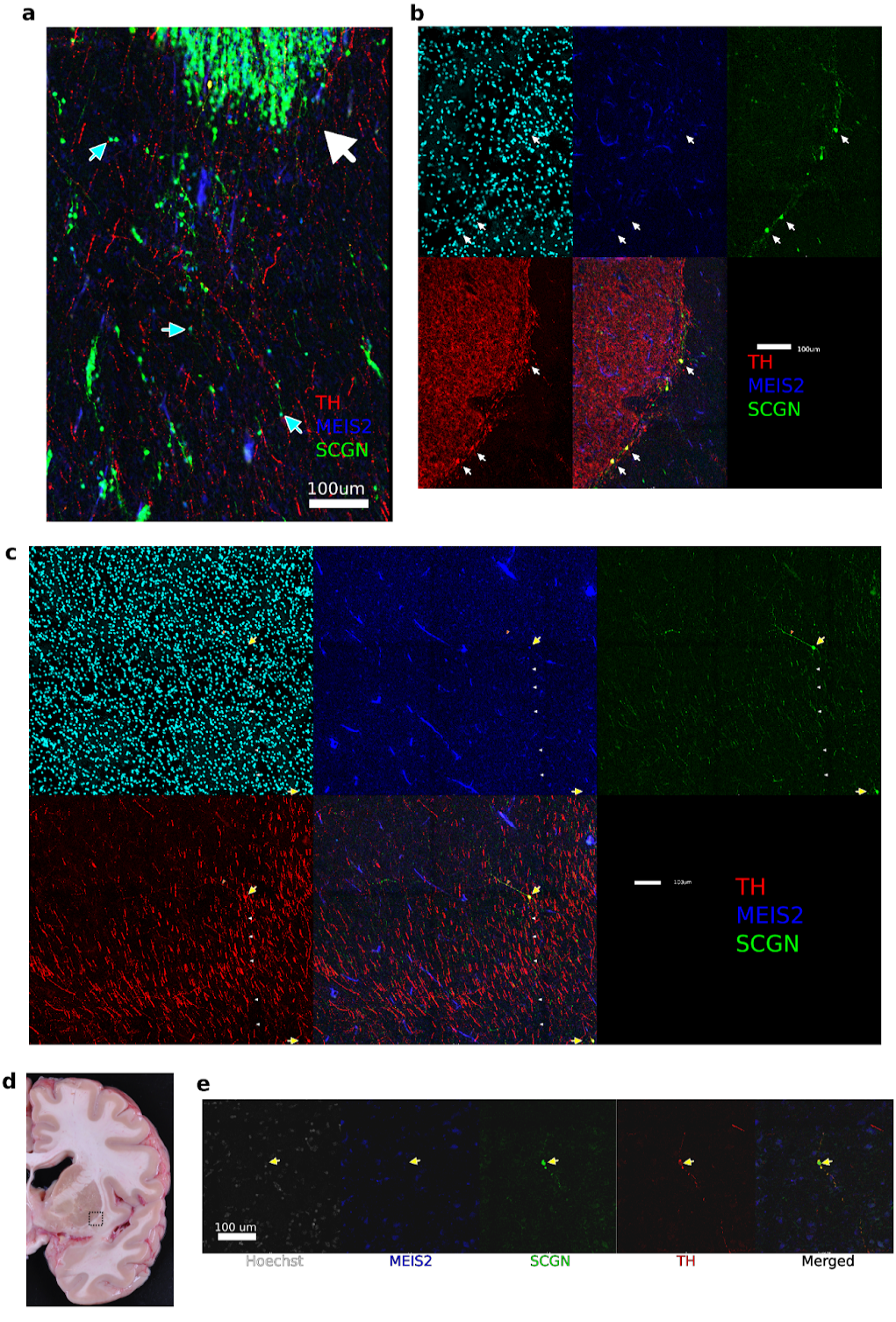
a
TH
MEIS2
SCGN
100um
b
100um
TH
MEIS2
SCGN
c
100um
TH
MEIS2
SCGN
d
e
100 um
Hoechst
MEIS2
SCGN
TH
Merged

Figure 2.20: Full montages of Figure 2.5 peristriatal *striatum laureatum* neurons.

a. Remnant immature neurons of the RMS from 7m macaque (large white arrow) with migrating morphology MEIS2+/SCGN+ cells (teal arrows). **b.** Montage of edge of striatum with arrows pointing to MEIS2+/SCGN+/TH+ SLNs. **c.** Montage of claustrum with yellow arrows pointing to MEIS2+/SCGN+/TH+ peristriatal SLNs. The TH+/SCGN+ process is labeled with orange arrowheads while TH-/SCGN+ process is labeled with white arrowheads.

Methods

Samples:

Macaque cortical tissue was generously provided by the UC Davis Primate Center from 9 specimens at PCD40, PCD50, PCD65 (N=3), PCD80 (N=2), PCD90, and PCD100. All animal procedures conformed to the requirements of the Animal Welfare Act and protocols were approved prior to implementation by the Institutional Animal Care and Use Committee (IACUC) at the University of California, Davis. PCD40 represents embryonic Carnegie stage 20 and marks the approximate beginning of neurogenesis of both excitatory and INs, while PCD100 is the approximate end of excitatory neurogenesis in the cortex(Clancy et al., 2001). Macaque data from PCD110 (N=2) was further incorporated from Zhu et al., 2018(Zhu et al., 2018). Public mouse datasets including E13.5, E14.5 ganglionic eminences (which was enriched for DLX6+ cells)(Mayer et al., 2018), 3 samples from the 10X Genomics 1.3 million cell E18 mouse cortex example dataset (GSE93421 samples 1,3,4), E14 and neonatal cortex and subcortex(Loo et al., 2019), whole brain developmental and adult structures(La Manno et al., 2020; Zeisel et al., 2018), P9 striatum(Anderson et al., 2020) as well as adult OB(Tepe et al., 2018) were also included. In total, we analyzed single cell transcriptomes from 109,112 developing macaque cells, 76,828 developing mouse cells, and 141,065 total mouse cells. De-identified tissue samples were collected with previous patient consent in strict observance of the legal and institutional ethical regulations in accordance with the Declaration of Helsinki. Protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee

and the Committee on Human Research (institutional review board) at the University of California, San Francisco.

Single cell RNA sequencing tissue processing:

For PCD40 - PCD100 macaque, dissections were performed in PBS under a stereo dissection microscope (Olympus SZ61). A number of regions were difficult to distinguish at earlier timepoints as key anatomical landmarks are still forming, and so presumptive regions were dissected (e.g. motor vs somatosensory cortex prior to appearance of the central sulcus or the anterior end of the MGE and LGE). For single cell dissociation, samples were cut into small pieces, and incubated with a pre-warmed solution of Papain (Worthington Biochemical Corporation) prepared according to manufacturer's instructions for 10 min at 37 C. After 30 – 60 min incubation, samples were gently triturated with glass pipet tips and PCD100 macaque samples were further spun through an ovomucoid gradient to remove debris. Cells were then pelleted at 300xg and resuspended in PBS supplemented with 0.1% bovine serum albumin(Sigma). Samples for MULTIseq were prepared in strip tubes and maintained at 4° C for the labeling protocol, as in(McGinnis et al., 2019). Single cell RNA-seq was completed using the 10X Genomics Chromium controller and the version 2 or 3 3-prime RNA capture kits. Most samples were loaded at approximately 10,000 cells per well while up to 25,000 cells were loaded per lane for multiplexed samples. Transcriptome library preparation was completed using the associated 10X Genomics RNA library preparation kit. Multiseq barcode library

preparation was completed as in McGinnis et al.(McGinnis et al., 2019). Following library preparation, libraries were sequenced on Illumina HiSeq and NovaSeq platforms.

Alignments and gene models:

Fastq files were generated from Illumina BCL files using bcl2fastq2. Genes were quantified using kallisto (release 0.46)(Bray et al., 2016) and the RheMac10 genome assembly, newly annotated using the comparative annotation toolkit(Fiddes et al., 2018), and the *Mus musculus* ENSEMBL release 100 transcript annotations. A custom kallisto reference for each species was created for the quantification of exons and introns together, where introns were defined as the complement of exonic and intergenic space. The kallisto index used kmers of length 31. Public data was downloaded as raw fastq files or BAM files (which were converted back to fastq files) and all data was processed from raw reads using the same kallisto pipeline to minimize annotation or alignment artifacts.

Quality control:

Kallisto-Bus output matrix files (including both introns and exons together) were input to Cellbender (release 0.2.0, https://github.com/broadinstitute/CellBender), which was used to remove likely ambient RNA only. Only droplets with greater than 0.99 probability of being cells (not ambient RNA), calculated by the Cellbender model, were included in further analysis.

Droplets with fewer than 800 genes detected, or greater than 40% ribosomal or 15% mitochondrial reads were filtered from the dataset. Doublets were then detected and removed from the dataset using scrublet (release 0.2.2, using threshold parameter 0.5).

Clustering and determining homologous cell types:

Much of the analysis pipeline is based on scanpy infrastructure and anndata data structures(Wolf et al., 2018). Counts in cells were normalized by read depth, log transformed and then scaled for each gene across all cells. Principal component analysis was then performed using the top 12,000 most variable genes (using the original Seurat variable genes selection method, implemented in the scanpy package), with the 100 most variance-encompassing principal components being used for the following steps. Batch correction was limited to the requirement that highly variable genes be variable in more than one sequencing sample and by application of batch-balanced k-nearest neighbors (BBKNN)(Polański et al., 2020), using Euclidean distance of principal components to find 3 neighbors per batch in the developing data, and 12 neighbors per dataset in the developing and adult merged mouse data. Using BBKNN-derived k-nearest neighbors graphs, Leiden clustering was then applied to cluster based upon the KNN graph with scanpy's resolution parameter set to 10 (or 7 in the developing mouse dataset). Glia, along with excitatory progenitor and neuron clusters were removed from the dataset in non-ganglionic eminence batches if they had below mean expression value for two or more GAD1/2 and DLX1/2/5/6 genes, with Cajal-Retzius cells

(RMTW_ZIC1/RELN) meeting this threshold and serving as a useful out-group (these cells were called RMTW-derived based on the ZIC1 and RELN expression, though they are known to have multiple origins(Bielle et al., 2005)). Following removal of non-INs, scaling, PCA and the following steps were repeated with this final IN dataset.

High-resolution Leiden clusters partitioned continuous differentiation trajectories of post-mitotic initial classes into subclusters based on maturation stage. These high resolution clusters were then merged to initial classes manually, using hierarchical clustering of cluster gene expression averages and distinctness of individual Leiden cluster markers as a guide. The nomenclature for merged clusters incorporates the presumptive spatial origin of initial classes and specific marker genes. Spatial origin for each class was inferred based on the expression of canonical marker genes for RMTW, MGE, LGE, CGE, and VMF (e.g., *LHX5, NKX2.1, MEIS2, NR2F2, ZIC1*) and supported by immunostaining and by the enrichment of these genes in cells from region-specific dissections. For merged species analysis, genes were normalized and scaled within species, then merged for downstream analysis using BBKNN (with 25 neighbors across and within species, with the mutual nearest neighbors used for the Sankey plot comparison of developing macaque and developing mouse). Following clustering, mean expression in each class was calculated for each gene which was among the original 12,000 most variable 1-to-1 orthologs from each dataset that were variable in both species (6,227 genes). These classes were then compared across species by Pearson correlation of their gene expression vectors.

Trajectory analysis of activating and inactivating macaque genes:

We applied scVelo's dynamical model (release 0.2.3) (Bergen et al., 2020) to derive a shared latent time based on RNA velocity using spliced and unspliced counts from kallisto. Next, we used the related cellrank (release 1.3.1) package(Lange et al., 2020) to derive absorption probabilities of immature cells in the "Transition" cluster to likely initial classes. This step was necessary due to the effect that, like children, newborn neurons are more similar to each other than to their mature state. By using adjacency along the paths of differentiation it is possible to infer which mature state is likely to absorb a given immature cell. We then called newborn neurons as cells below the 0.5 quantile of latent time for that class. Recent studies indicate that these transcriptionally immature neurons correspond to newborn neurons as labeled by classical nucleoside-based methods(Habib et al., 2016). To call genes activating or inactivating along trajectories, we used linear regression implemented in scipy upon latent time values (x) vs genes' expression values (y). This yielded linear regression coefficients and two-tailed p-values for each gene, which were multiple hypothesis testing corrected using Holm-Sidak method implemented in the statsmodels (release 0.12.2) package to derive q-values. Comparison of gene sets was carried out by calculating the Jaccard indices of set intersections, defined as the number of intersecting elements between two sets divided by the number of elements in the union of the two sets.

Linking developmental and adult data:

Similar to the reassignment of the macaque "Transition" cells, we also used cellrank-derived absorption probabilities, with equally weighted KNN and RNA-velocity kernels, to estimate precursor states of adult cells. Because absorption probabilities are biased by cell numbers in terminal states, and the goal this time was not to assign each developmental cell to a terminal state, we subsample a maximum of 1000 cells (the rarest class was was 707 cells from MGE_CRABP1/MAF) and reported the class identity of the 100 developing cells with the highest probability to be absorbed into each terminal class, we provide an estimate of which developing class is the likely origin of the terminal classes, which is reflected in the weights of the edges in the **Figure 2.2** Sankey diagram. We also calculated the mean absorption probability for cells in each initial class to each terminal state to alleviate compositional effects, presented as a heatmap. Note that the RMTW_ZIC1/RELN and VMF_TMEM163/OTP are not included, as they are excitatory cortical and hypothalamic classes, respectively.

Immunohistochemistry tissue processing and imaging:

Mouse, macaque, and human tissue for histology was fixed in 4% PFA in PBS overnight at 4 C with constant agitation. PFA was then replaced with fresh PBS (pH=7.4) and cryopreserved by 24-48 hour incubation in 30% sucrose diluted in PBS (pH=7.4), and then embedded in a mixture of OCT (Tissue-Tek, VWR) and 30% sucrose. Tissue was then frozen at 80° C and

was cryosectioned at 16 - 20 um. For RNAscope RNA *in situ* hybridization, fixed cryosections were stained according to the Advanced Cell Diagnostics RNAscope Multiplex Fluorescent Reagent Kit V2 Assay (ACD, 323120) protocol. For immunostaining, antigen retrieval was performed by placing tissue slides in a 95° C citrate buffer, then allowing them to cool at room temperature. Antibodies were diluted in blocking buffer (0.1% Triton X-100, 5% donkey serum, 0.2% gelatin in PBS). Primary antibodies were incubated overnight at room temperature under bright light to photobleach autofluorescence in a light box(Sun et al., 2017). Primary antibodies (and dilutions) used were recorded in online methods.

Alexa dye-conjugated donkey secondary antibodies were incubated in the dark at room temperature for 1 hour. All tiled scans were acquired using the Evos M7000 microscope. All images were stitched using a custom python script and ImageJ's max correlation Grid/Collection stitching (release 1.2) and processed using ImageJ (release 1.53c) Rolling Ball background subtraction and manual brightness/contrast adjustment within an ImageJ macro. Image quantification of CRABP1+ cells was carried out with a custom ImageJ macro, with CRABP1+ area automatically thresholded using Maximum entropy. Positivity for other genes was classified manually for every cell in at least 5 random areas in the striatum or MGE, and was defined as >1 puncta within CRABP1+ area not clearly belonging to another cell.

Acknowledgments

The authors thank Aparna Bhaduri, Arturo Alvarez-Buylla, John Rubenstein, Corey Harwell, Miguel Turrero Garcia, Bill Seeley, Alissa Li, Stephanie Gaus, Shaohui Wang, Beatriz Alvarado, Alice Tarantal, Andrew Fox, Madeline Andrews, Sara Nolbrant, Jenelle Wallace, Sofie Salama, Ian Fiddes, and members of the Pollen and Ye labs for helpful discussion, reagents, and samples, and Kenneth Probst for artwork. This study was supported by NIH awards F31NS124333 (M.T.S), DP2MH122400-01 (A.A.P), U01MH114825 (A.A.P, T.JN), R01AI136972 (C.J.Y), DP2NS122550-01 (M.F.P.), Roberta and Oscar Gregory Endowment (M.F.P.), the Schmidt Futures Foundation (A.A.P, T.J.N), the Chan Zuckerberg Biohub (C.J.Y, A.A.P.,T.J.N), and the Shurl and Kay Curci Foundation (A.A.P.).

Contributions

Methodology, M.T.S, T.J.N, M.F.P, A.A.P.; Investigation, M.T.S. K.S., C.P.C., M.R., T.J.N., and A.A.P.; Resources, M.F.P, T.J.N, A.A.P; Histology, M.T.S., C.P.C., K.S.; scRNAseq Processing, M.T.S., M.M.R., T.J.N., A.A.P.; Software, M.T.S; Formal Analysis, M.T.S; Writing, M.T.S, A.A.P. with input from all authors; Funding Acquisition, T.J.N, M.F.P, C.J.Y, A.A.P.; Conceptualization, M.T.S, M.F.P. A.A.P.; Supervision, T.J.N, C.J.Y, M.F.P, A.A.P.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Materials and correspondence should be addressed to alex.pollen@ucsf.edu

Data Availability

The sequencing data have been deposited in GEO under the accession GSE169122, and the

data is browsable at https://dev-inhibitory-neurons.cells.ucsc.edu/

Tables

Table 2.1 Dictionary of initial and terminal classes

Qualitative definitions of classes explored in the atlas with extended explanations for inferences about initial-terminal class relationships.

class name	class category	description	aliases or subclasses	derived from
CGE_NR2F2/PROX1	initial	Newborn neurons from the caudal ganglionic eminence		
LGE-OB_MEIS2/PAX6	initial	Newborn neurons of the LGE_MEIS2/PAX6 class which are found in the adult olfactory bulb	GC-2 (doi: 10.1016/j.celrep.2018.11.034)	
LGE_FOXP1/ISL1	initial	Newborn neurons of the inner LGE which are FOXP1/ISL1 positive		
LGE_FOXP1/ISL1/NPY1R	initial	A ventral subclass of LGE_FOXP1/ISL1, which is EBF1 negative and NPY1R positive, and is apparently distinct early in differentiation		
LGE_FOXP1/PENK	initial	Newborn neurons of the inner LGE which are FOXP1/PENK positive		
LGE_FOXP2/TSHZ1	initial	Newborn neurons from the external/dorsal LGE which are FOXP2/TSHZ1 positive and FOXP1/NPY negative. We observe in immunos and transcriptome trajectories they are likely SP8 and SCGN positive until they become strongly FOXP2 positive, at which point those genes become undetectable.		
LGE_MEIS2/PAX6	initial	Newborn neurons of the external/dorsal LGE which are MEIS2/PAX6/TSHZ1/SCGN(macaque only)/ETV1/FOXP2(but barely detectable with immunofluorescence) positive and FOXP1/NR2F2 negative		
LGE_MEIS2/PAX6/SCGN	initial	A subclass of LGE_MEIS2/PAX6 which is also positive for CALB2, ZIC1 and SCGN in the mouse, and is apparently distinct early in differention.		
MGE_CRABP1/MAF	initial	Newborn neurons of the MGE which are CRABP1/MAF/NKX2-1/ETV1/LHX6 positive and are seen migrating through the MGE and striatum by RNAscope.		
MGE_CRABP1/TAC3	initial	A sister initial class to MGE_CRABP1/MAF. Appears by RNAscope to become LHX8+/- only after reaching striatal destinations. Adult terminal classes covered by		

class name	class category	description	aliases or subclasses	derived from
		Krienen et al. (doi: 10.1038/s41586-020-2781-z)		
MGE_LHX6/MAF	initial	Newborn neurons from the MGE which are positive for NKX-1/LHX6/MAF/MAFB/ERBB4/SST		
MGE_LHX6/NPY	initial	Newborn neurons from the MGE which may be postmitotic derivatives from MGE_LHX6/MAF, but appear to distribute to the basal nuclei as well as the cortex. Expresses LHX6/MAF/NPY/CORT/CHODL. See Ctx/BN_SST/CHODL.		
RMTW_ZIC1/RELN	initial	Newborn Cajal-Retzius neurons found in cortical layer 1 and derived from the Rostro-medial telencephalic wall (RMTW). Highly expresses RELN. Appear to be glutamatergic, but express low levels of GAD genes. Reported to mostly disappear before maturity, and no corresponding terminal class is observed in adult.	Cajal-Retzius Cells, Horizontal Cells of Cajal	
VMF_CRABP1/LHX8	initial	These cells may arise from the MGE-like GBX1+ portion of the septum.		
VMF_LHX1/POU6F2	initial	POH or POA-derived neurons expressing PAX6/NR2F2/ISL1/SP8/MEIS2/LHX1/TSHZ2, reported to become TH+/SCGN+/- hypothalamic dopaminergic cells	may contribute to dopaminergic groups A11-15 (doi:10.1016/j.tins.2007.03.006)	
VMF_NR2F2/LHX6	initial	Neurons with a VMF expression profile that are NR2F2/LHX6/CALB1 positive. Found in mouse MGE, CGE and cortex dissections with unknown derivative adult profile. This is a possible source of POA derived cortical interneurons (doi: 10.7554/eLife.32017 , doi: 10.1523/JNEUROSCI.4068-11.2011)		
VMF_PEG10/DLK1	initial	POH or POA-derived newborn neurons that express PEG10/DLK1/HMX3/HAP1/TSHZ2/NR2F2. These cells are reported to cross the telencephalon-diencephalon boundary and to populate the amygdala. (doi: 10.1038/nn.2556)		
VMF_TMEM163/OTP	initial	Hypothalamus-derived newborn neurons that express VMF_TMEM163/OTP. These cells appear to be excitatory, their top marker is SLC17A6. Thus they are not included in our taxonomy, nor does their derivative terminal class(es) appear to make thresholds to be included in the dataset.		
VMF_ZIC1/ZIC2	initial	Septum-derived newborn neurons expressing ZIC1/2/3/4. Adult septum is undersampled in both macaque and mouse, and so derivative terminal class		

class name	class category	description	aliases or subclasses	derived from
		may be missing from dataset.		
Amy/Hypo_HAP1/PEG10	terminal	These cells were found in both hypothalamus and amygdala mouse samples. They appear to be the neurons which are reported to cross the telencephalon-diencephalon boundary and to populate the amygdala (doi: 10.1038/nn.2556). Interestingly their expression suggests they use both GABA and Glutamate neurotransmitters.		VMF_PEG10/DLK1 or VMF_TMEM163/OTP
BN-eSPN_FOXP2/TSHZ1	terminal	This terminal class appears to consist of two related classes which coclustered in our leiden clustering. In the striatum cells express CASZ1/OTOF and are known as eccentric spiny neurons (eSPNs). The other clear subclass is derived from amygdala samples, where these cells are known as intercalated cells (ITCs). In both locations they express FOXP2/4 but do not express FOXP1. We believe the preponderance of the evidence points to LGE_FOXP2/TSHZ1 being the precursor of BN-eSPN_FOXP2/TSHZ1. This is supported by the RNA velocity absorption analysis, the more perfect matching of markers like OPRM1, CASZ1 and lack of FOXP1 in macaque and mouse, expression of CASZ1 in the dorsal/external LGE where SCGN/FOXP2 are also seen (histology), trajectory expression showing that immature cells of FOXP2/TSHZ1 expressing SCGN, SP8 and ETV1. LGE_FOXP1/ISL1 is also a possible source but we believe the partial transcriptomic correspondence between LGE_FOXP1/ISL1 and BN-eSPN_FOXP2/TSHZ1 is spuriously driven by activation of shared projection markers and DRD1 as these neurons mature.	eccentric spiny neurons, intercalated cells of the amygdala, D1H(doi:10.1016/j.neuron.2019.11.004), PCDH8+ SPNs(doi:10.1016/j.celrep.2016.06.059)	LGE_FOXP2/TSHZ1
Ctx/BN_SST/CHODL	terminal	This terminal class is distributed to the cortex and also the basal nuclei. In the cortex these neurons are well known for high expression of SST and CHODL. They also likely correspond to hippocampal SST/NPY/NOS1+ ivy and neurogliaform (NGF) cells (doi:10.1523/JNEUROSCI.5123-09.2010).	cortical long-projecting SST+ neurons, hippocampal ivy and NGF cells, striatal plateau low-threshold-spiking neurons	MGE_LHX6/NPY
Ctx_CCK/DPY19L1	terminal	Rare class of cortical interneurons		CGE_NR2F2/PROX1
Ctx_CCK/VIP	terminal	Common CCK+ VIP+/- class of CGE-derived cortical interneurons	CCK basket cells, VIP bipolar cells	CGE_NR2F2/PROX1
Ctx_LAMP5/	terminal	Class of cortical/hippocampal formation interneurons	Adarb2/Ndnf-HPF cells	CGE_NR2F2

class name	class category	description	aliases or subclasses	derived from
NDNF			in Allen taxonomy.	/PROX1
Ctx_LHX6/LAMP5	terminal	LHX6/LAMP5+ class of cortical chandelier interneurons. This class is reported to be derived from NKX2-1+ cells using Cre-Lox lineage tracing(doi:10.1016/j.cell.2017.08.032, Paul et al). We find that the most parsimonious explanation is that these are MGE derived interneurons which bear transcriptomic similarity to CGE-derived LAMP5 cells due to very high expression of LAMP5.	CHC2 Cortical Chandelier Cells (doi:10.1038/s41586-018-0654-5)	MGE_LHX6/MAF
Ctx_LHX6/PVALB	terminal	Common PVALB+ MGE-derived cortical interneurons	PVALB basket cells, fast spiking interneurons	MGE_LHX6/MAF
Ctx_LHX6/SST	terminal	Common SST+ MGE-derived cortical interneurons	Martinotti SST+ interneurons	MGE_LHX6/MAF
Ctx_NR2F2/PAX6	terminal	Rare class of cortical interneurons. Terminal class doesn't express MEIS2/FOXP2/STXBP6/ETV1 or any other markers which would be expected from LGE derived cells but expresses CGE/VMF marker NR2F2, and RNA velocity suggests CGE_NR2F2/PROX1 is more likely as origin.	Adarb2/Pax6 cells in Allen taxonomy.	CGE_NR2F2/PROX1 or VMF classes or (unlikely) LGE_MEIS2/PAX6
Ctx_PROX1/LAMP5	terminal	Common LAMP5+ class of CGE-derived cortical interneurons	Lamp5 interneurons	CGE_NR2F2/PROX1
Ctx_PROX1/SNCG	terminal	Common SNCG+ class of CGE-derived cortical interneurons	Sncg interneurons	CGE_NR2F2/PROX1
Ctx_PVALB/VIPR2	terminal	PVALB/VIPR2+ class of cortical chandelier interneurons. RNA velocity analysis is unclear as to which MGE class is the more likely precursor initial class. Cortical PVALB cells are assumed to descend from MGE_LHX6/MAF class, while striatal PVALB cells descend from MGE_CRABP1/MAF, (doi: 10.1016/j.celrep.2018.07.053), but the Ctx_PVALB/VIPR2 class is distinct from the Ctx_LHX6/PVALB class and shows transcriptional similarities to both candidate initial classes, for instance sharing PTHLH and continuing adult NKX2-1 expression with MGE_CRABP1/MAF.	CHC1 Cortical Chandelier Cells (doi:10.1038/s41586-018-0654-5)	MGE_CRABP1/MAF or MGE_LHX6/MAF
Ctx_SST/NDNF	terminal	SST/NDNF+ class of cortical interneurons		MGE_LHX6/MAF
GP_GBX1/GABRA1	terminal	GBX1/GABRA1/PVALB/ZIC1 neurons which appear to make up a large percentage of the neurons in the globus pallidus	fast-spiking “prototypical” globus pallidus neurons (doi:	VMF_CRABP1/LHX8

class name	class category	description	aliases or subclasses	derived from
			10.1016/j.cell.2018.07.028)	
OB-GC NR2F2/PENK	terminal	A class of granule cells of the olfactory bulb	GC-6 (doi: 10.1016/j.celrep.2018.11.034)	LGE_MEIS2/PAX6 (LGE-OB_MEIS2/PAX6)
OB-GC_RPRM	terminal	A rare class of granule cells of the olfactory bulb	GC-3 (doi: 10.1016/j.celrep.2018.11.034)	LGE_MEIS2/PAX6 or LGE_FOXP2/TSHZ1
OB-GC_STXBP6/PENK	terminal	The most prominent class of granule cells of the olfactory bulb	GC-5 (doi: 10.1016/j.celrep.2018.11.034)	LGE_MEIS2/PAX6 (LGE-OB_MEIS2/PAX6)
OB-PGC_FOXP2/CALB1	terminal	FOXP2/CALB1 positive periglomerular cells (PGC) of the olfactory bulb	PGC-3 (doi: 10.1016/j.celrep.2018.11.034), calbindin+ PGC	LGE_FOXP2/TSHZ1
OB-PGC_TH/SCGN	terminal	MEIS2/PAX6/ETV1/TH/SCGN positive periglomerular cells of the olfactory bulb and accessory olfactory bulb. We believe that TH/SCGN+ PSCs are derived from this group rather than the hypothalamic VMF_LHX1/POU6F2 as PSCs are negative for NR2F2, in addition to their location far removed from the telencephalon-diencephalon boundary.	PGC-2 (doi: 10.1016/j.celrep.2018.11.034), dopaminergic group A16 (doi: 10.1016/S0165-0173(00)00034-5 / doi.org/10.1016/j.tins.2007.03.006 short axon dopaminergic cells, THLI striatum cells	LGE_MEIS2/PAX6/SCGN or LGE_MEIS2/PAX6
OB-PGC_ZIC	terminal	CALB2/SP8/ETV1/ZIC1/2/4 positive periglomerular cells of the olfactory bulb	PGC-1 (doi: 10.1016/j.celrep.2018.11.034)	LGE_MEIS2/PAX6 or septum
Str-IN_CRABP1/MAF	terminal	Striatal GABAergic interneurons CRABP1/ETV1/NKX2-1/KIT/MAF+ in adult rodent and primates.	TH+ (MGE-derived) striatal interneurons, PVALB+ (MGE-derived) striatal interneurons, PTHLH+ (MGE-derived) striatal interneurons	MGE_CRABP1/MAF
Str-IN_CRABP1/TAC3	terminal	Striatal GABAergic interneurons CRABP1/ETV1/NKX2-1/MAF/TAC3/STXBP6+ in adult primates but not rodents, nor laurasiatherians.	TAC3+ (MGE-derived) striatal interneurons	MGE_CRABP1/TAC3

class name	class category	description	aliases or subclasses	derived from
Str-dSPN_FOXP1/ISL1	terminal	Direct (striatonigral projecting) spiny projection neurons (DRD1+) of the basal nuclei. Subclasses include orthogonal axes of variation which appear to be overlaid later in development such as dorsoventral and patch-matrix gene expression identities, perhaps due to striosome cells appearing to be born before matrix cells in mouse (doi:10.1016/j.neuron.2018.06.021.)	dSPN, direct medium spiny neuron	LGE_FOXP1/ISL1
Str-iSPN_FOXP1/PENK	terminal	Indirect (striatopallidal projecting) spiny projection neurons (DRD2+) of the basal nuclei. Subclasses include orthogonal axes of variation which appear to be overlaid later in development such as dorsoventral and patch-matrix gene expression identities.	iSPN, indirect medium spiny neuron	LGE_FOXP1/PENK
Str_LHX8/CHAT	terminal	Cholinergic cells of the basal nuclei and septum which express ZIC1/CHAT/LHX8 and choline transporters	cholinergic striatal interneurons, septohippocampal cholinergic pathway cells	VMF_CRABP1/LHX8
DWMIN_NR2F2/SP8	terminal	The terminal class identity and transcriptome is unknown. We observed many cells matching these markers in the deep white matter at late developmental and mature ages in the macaque, but can't speculate about them beyond noting their existence in large numbers. Presumed to be CGE derived, but could also be POA/POH derived given the small number of markers observed in immunostaining		CGE_NR2F2/PROX1 (Presumed), VMF also possible
DWMIN_MEIS2/PAX6	terminal	The terminal class full transcriptome is unknown. Deep white matter inhibitory neurons (DWMIN) derived from the arc and arc-ACC migratory streams. They appear to express MEIS2/PAX6/SP8/SCGN and low/no FOXP2, TH or NR2F2. While it was unclear whether mature cortical DWMIN_MEIS2/PAX6 cells in the adult mouse clustered with olfactory bulb or were hidden amongst another larger terminal class, Figure 4 and ED provides evidence that neurons of class migrate into and remain in the deep white matter. protein.	Possible component of Allen cortical Meis2	LGE_MEIS2/PAX6
SWMIN_FOXP2/FOXP4	terminal	The terminal class full transcriptome is unknown. Superficial white matter inhibitory neurons (SWMIN), which appear to derive from a medial migratory stream from the LMS which continues into VMPFC. Their existence is largely based upon LGE_FOXP2/TSHZ1 cells found in cortical developing macaque samples, and FOXP2/SP8 or FOXP2/DLX2 stains in macaque superficial white matter. These cells appeared very rare	Likely component of Allen cortical Meis2	LGE_FOXP2/TSHZ1

class name	class category	description	aliases or subclasses	derived from
		in histology. Macaque transcriptome and histology suggests these cells still express some SP8 and SCGN while migrating. These cells did not cluster distinctly in mouse adult data.		
Str-SLN_TH/SCGN	terminal	The terminal class full transcriptome is unknown. These *striatum laureatum* neurons wreaTHe the boundaries of the striatum (especially externally), including the claustrum. These cells share many markers with OB-PGC_TH/SCGN, however this requires further transcriptomic characterization. These cells express MEIS2/PAX6/SP8/TH/SCGN but do not appear to express FOXP2 (SPN), NKX2-1 (MGE-derived striatal interneuron) or NR2F2 (CGE or POA/POH-derived). It appears that most studies of TH+ cells of the striatum don't distinguish between Str-SLN_TH/SCGN and Str-IN_CRABP1/MAF/TH. Informally called wreaTH cells.		LGE_MEIS2/PAX6/SCGN or LGE_MEIS2/PAX6 (presumed)
G1-phase_SLC1A3/ATP1A1	technical	Cells in G1-phase of cell cycle		
S-phase_MCM4/H43C	technical	Cells in S-phase of cell cycle		
G2-M_UBE2C/ASPM	technical	Cells in G2/M phase of cell cycle		
Transition	technical	Cells of multiple classes which aren't well distinguished by leiden clustering due to cell cycle and early maturation variation. Excluded from most analyses to avoid cross contamination of initial classes		
Glia	technical	Diverse glial cells which pass filters to be part of the dataset but are not analyzed here.		
Excitatory	technical	Excitatory cells which pass filters to be part of the dataset. Expresses SLC17A6/7 and tiny amount of GAD or DLX genes, so are not analyzed		

References

Agoston Z, Heine P, Brill MS, Grebbin BM, Hau A-C, Kallenborn-Gerhardt W, Schramm J, Götz M, Schulte D. 2014. Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development* **141**:28–38. doi:10.1242/dev.097295

Alzu'bi A, Clowry GJ. 2020. Multiple Origins of Secretagogin Expressing Cortical GABAergic Neuron Precursors in the Early Human Fetal Telencephalon. *Front Neuroanat* **14**. doi:10.3389/fnana.2020.00061

Anderson AG, Kulkarni A, Harper M, Konopka G. 2020. Single-Cell Analysis of Foxp1-Driven Mechanisms Essential for Striatal Development. *Cell Rep* **30**:3051-3066.e7. doi:10.1016/j.celrep.2020.02.030

Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, Wagner GP. 2016. The origin and evolution of cell types. *Nat Rev Genet* **17**:744–757. doi:10.1038/nrg.2016.127

Benavides-Piccione R, DeFelipe J. 2007. Distribution of neurons expressing tyrosine hydroxylase in the human cerebral cortex. *J Anat* **211**:212–222. doi:10.1111/j.1469-7580.2007.00760.x

Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 1–7. doi:10.1038/s41587-020-0591-3

Betarbet R, Turner R, Chockkan V, DeLong MR, Allers KA, Walters J, Levey AI, Greenamyre JT. 1997. Dopaminergic Neurons Intrinsic to the Primate Striatum. *J Neurosci* **17**:6761–6768. doi:10.1523/JNEUROSCI.17-17-06761.1997

Bielle F, Griveau A, Narboux-Nême N, Vigneau S, Sigrist M, Arber S, Wassef M, Pierani A. 2005. Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. *Nat Neurosci* **8**:1002–1012. doi:10.1038/nn1511

Björklund A, Dunnett SB. 2007. Dopamine neuron systems in the brain: an update. *Trends Neurosci*, Fifty years of dopamine research **30**:194–202. doi:10.1016/j.tins.2007.03.006

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**:525–527. doi:10.1038/nbt.3519

Cave JW, Akiba Y, Banerjee K, Bhosle S, Berlin R, Baker H. 2010. Differential Regulation of Dopaminergic Gene Expression by Er81. *J Neurosci* **30**:4717–4724. doi:10.1523/JNEUROSCI.0419-10.2010

Clancy B, Darlington RB, Finlay BL. 2001. Translating developmental time across mammalian species. *Neuroscience* **105**:7–17. doi:10.1016/S0306-4522(01)00171-3

Crick FC, Koch C. 2005. What is the function of the claustrum? *Philos Trans R Soc B Biol Sci* **360**:1271–1279. doi:10.1098/rstb.2005.1661

DeFelipe J. 2011. The Evolution of the Brain, the Human Nature of Cortical Circuits, and Intellectual Creativity. *Front Neuroanat* **5**. doi:10.3389/fnana.2011.00029

DeFelipe J, López-Cruz PL, Benavides-Piccione R, Bielza C, Larrañaga P, Anderson S, Burkhalter A, Cauli B, Fairén A, Feldmeyer D, Fishell G, Fitzpatrick D, Freund TF, González-Burgos G, Hestrin S, Hill S, Hof PR, Huang J, Jones EG, Kawaguchi Y, Kisvárday Z, Kubota Y, Lewis DA, Marín O, Markram H, McBain CJ, Meyer HS, Monyer H, Nelson SB, Rockland K, Rossier J, Rubenstein JLR, Rudy B, Scanziani M, Shepherd GM, Sherwood CC, Staiger JF, Tamás G, Thomson A, Wang Y, Yuste R, Ascoli GA. 2013. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci* **14**:202–216. doi:10.1038/nrn3444

Diederich NJ, Surmeier DJ, Uchihara T, Grillner S, Goetz CG. 2019. Parkinson's disease: Is it a consequence of human brain evolution? *Mov Disord* **34**:453–459. doi:https://doi.org/10.1002/mds.27628

Dubach M, Schmidt R, Kunkel D, Bowden DM, Martin R, German DC. 1987. Primate neostriatal neurons containing tyrosine hydroxylase: Immunohistochemical evidence. *Neurosci Lett* **75**:205–210. doi:10.1016/0304-3940(87)90298-9

Duchatel RJ, Shannon Weickert C, Tooney PA. 2019. White matter neuron biology and neuropathology in schizophrenia. *Npj Schizophr* **5**:1–9. doi:10.1038/s41537-019-0078-8

Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, Haussler D, Stanke M, Paten B. 2018. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res*. doi:10.1101/gr.233460.117

Fishell G, Kepecs A. 2020. Interneuron Types as Attractors and Controllers. *Annu Rev Neurosci* **43**:annurev-neuro-070918-050421. doi:10.1146/annurev-neuro-070918-050421

Flames N, Pla R, Gelman DM, Rubenstein JLR, Puelles L, Marín O. 2007. Delineation of Multiple Subpallial Progenitor Domains by the Combinatorial Expression of Transcriptional Codes. *J Neurosci* **27**:9682–9695. doi:10.1523/JNEUROSCI.2750-07.2007

Frazer S, Prados J, Niquille M, Cadilhac C, Markopoulos F, Gomez L, Tomasello U, Telley L, Holtmaat A, Jabaudon D, Dayer A. 2017. Transcriptomic and anatomic parcellation of 5-HT 3A R expressing cortical interneuron subtypes revealed by single-cell RNA sequencing. *Nat Commun* **8**:14219. doi:10.1038/ncomms14219

Fuentealba LC, Rompani SB, Parraguez JI, Obernier K, Romero R, Cepko CL, Alvarez-Buylla A. 2015. Embryonic Origin of Postnatal Neural Stem Cells. *Cell* **161**:1644–1655. doi:10.1016/j.cell.2015.05.041

Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, Trombetta JJ, Hession C, Zhang F, Regev A. 2016. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**:925–928. doi:10.1126/science.aad7038

Hansen DV, Lui JH, Flandin P, Yoshikawa K, Rubenstein JL, Alvarez-Buylla A, Kriegstein AR. 2013. Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat Neurosci* **16**:1576–1587. doi:10.1038/nn.3541

Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchís-Calleja F, Guijarro P, Sidow L, Fleck JS, Han D, Qian Z, Heide M, Huttner WB, Khaitovich P, Pääbo S, Treutlein B, Camp JG. 2019. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**:418–422. doi:10.1038/s41586-019-1654-9

Kepecs A, Fishell G. 2014. Interneuron cell types are fit to function. *Nature* **505**:318–326. doi:10.1038/nature12983

Krienen FM, Goldman M, Zhang Q, C. H. del Rosario R, Florio M, Machold R, Saunders A, Levandowski K, Zaniewski H, Schuman B, Wu C, Lutservitz A, Mullally CD, Reed N, Bien E, Bortolin L, Fernandez-Otero M, Lin JD, Wysoker A, Nemesh J, Kulp D, Burns M, Tkachev V, Smith R, Walsh CA, Dimidschstein J, Rudy B, S. Kean L, Berretta S, Fishell G, Feng G, McCarroll SA. 2020. Innovations present in the primate interneuron repertoire. *Nature* 1–8. doi:10.1038/s41586-020-2781-z

Kuerbitz J, Arnett M, Ehrman S, Williams MT, Vorhees CV, Fisher SE, Garratt AN, Muglia LJ, Waclaw RR, Campbell K. 2018. Loss of Intercalated Cells (ITCs) in the Mouse

Amygdala of Tshz1 Mutants Correlates with Fear, Depression, and Social Interaction Phenotypes. *J Neurosci* **38**:1160–1177. doi:10.1523/JNEUROSCI.1412-17.2017

Kuerbitz J, Madhavan M, Ehrman LA, Kohli V, Waclaw RR, Campbell K. n.d. Temporally Distinct Roles for the Zinc Finger Transcription Factor Sp8 in the Generation and Migration of Dorsal Lateral Ganglionic Eminence (dLGE)-Derived Neuronal Subtypes in the Mouse. *Cereb Cortex*. doi:10.1093/cercor/bhaa323

La Manno G, Siletti K, Furlan A, Gyllborg D, Vinsland E, Langseth CM, Khven I, Johnsson A, Nilsson M, Lönnerberg P, Linnarsson S. 2020. Molecular architecture of the developing mouse brain (preprint). Developmental Biology. doi:10.1101/2020.07.02.184051

Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, Lickert H, Ansari M, Schniering J, Schiller HB, Pe'er D, Theis FJ. 2020. CellRank for directed single-cell fate mapping. *bioRxiv* 2020.10.19.345983. doi:10.1101/2020.10.19.345983

Lim L, Mi D, Llorca A, Marín O. 2018. Development and Functional Diversification of Cortical Interneurons. *Neuron* **100**:294–313. doi:10.1016/j.neuron.2018.10.009

Lledo P-M, Alonso M, Grubb MS. 2006. Adult neurogenesis and functional plasticity in neuronal circuits. *Nat Rev Neurosci* **7**:179–193. doi:10.1038/nrn1867

Loo L, Simon JM, Xing L, McCoy ES, Niehaus JK, Guo J, Anton ES, Zylka MJ. 2019. Single-cell transcriptomic analysis of mouse neocortical development. *Nat Commun* **10**:134. doi:10.1038/s41467-018-08079-9

Ma T, Wang C, Wang L, Zhou X, Tian M, Zhang Q, Zhang Y, Li J, Liu Z, Cai Y, Liu F, You Y, Chen C, Campbell K, Song H, Ma L, Rubenstein JL, Yang Z. 2013. Subcortical origins of human and monkey neocortical interneurons. *Nat Neurosci* **16**:1588–1597. doi:10.1038/nn.3536

Marín O, Anderson SA, Rubenstein JLR. 2000. Origin and Molecular Specification of Striatal Interneurons. *J Neurosci* **20**:6063–6076. doi:10.1523/JNEUROSCI.20-16-06063.2000

Märtin A, Calvigioni D, Tzortzi O, Fuzik J, Wärnberg E, Meletis K. 2019. A Spatiomolecular Map of the Striatum. *Cell Rep* **29**:4320-4333.e5. doi:10.1016/j.celrep.2019.11.096

Mayer C, Hafemeister C, Bandler RC, Machold R, Batista Brito R, Jaglin X, Allaway K, Butler A, Fishell G, Satija R. 2018. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**:457–462. doi:10.1038/nature25999

McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, Hu JL, Murrow LM, Weissman JS, Werb Z, Chow ED, Gartner ZJ. 2019. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* **16**:619–626. doi:10.1038/s41592-019-0433-8

Miller DJ, Duka T, Stimpson CD, Schapiro SJ, Baze WB, McArthur MJ, Fobbs AJ, Sousa AMM, Šestan N, Wildman DE, Lipovich L, Kuzawa CW, Hof PR, Sherwood CC. 2012. Prolonged myelination in human neocortical evolution. *Proc Natl Acad Sci U S A* **109**:16480–16485. doi:10.1073/pnas.1117943109

Miller JA, Gouwens NW, Tasic B, Collman F, van Velthoven CT, Bakken TE, Hawrylycz MJ, Zeng H, Lein ES, Bernard A. 2020. Common cell type nomenclature for the mammalian brain. *eLife* **9**:e59928. doi:10.7554/eLife.59928

Miyoshi G, Hjerling-Leffler J, Karayannis T, Sousa VH, Butt SJB, Battiste J, Johnson JE, Machold RP, Fishell G. 2010. Genetic Fate Mapping Reveals That the Caudal Ganglionic Eminence Produces a Large and Diverse Population of Superficial Cortical Interneurons. *J Neurosci* **30**:1582–1594. doi:10.1523/JNEUROSCI.4515-09.2010

Muñoz-Manchado AB, Bengtsson Gonzales C, Zeisel A, Munguba H, Bekkouche B, Skene NG, Lönnerberg P, Ryge J, Harris KD, Linnarsson S, Hjerling-Leffler J. 2018. Diversity of

Interneurons in the Dorsal Striatum Revealed by Single-Cell RNA Sequencing and PatchSeq. *Cell Rep* **24**:2179-2190.e7. doi:10.1016/j.celrep.2018.07.053

Paredes MF, James D, Gil-Perotin S, Kim H, Cotter JA, Ng C, Sandoval K, Rowitch DH, Xu D, McQuillen PS, Garcia-Verdugo J-M, Huang EJ, Alvarez-Buylla A. 2016a. Extensive migration of young neurons into the infant human frontal lobe. *Science* **354**. doi:10.1126/science.aaf7073

Paredes MF, Sorrells SF, Garcia-Verdugo JM, Alvarez-Buylla A. 2016b. Brain size and limits to adult neurogenesis. *J Comp Neurol* **524**:646–664. doi:10.1002/cne.23896

Paul A, Crow M, Raudales R, He M, Gillis J, Huang ZJ. 2017. Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell* **171**:522-539.e20. doi:10.1016/j.cell.2017.08.032

Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park J-E. 2020. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**:964–965. doi:10.1093/bioinformatics/btz625

Rakic P. 1974. Neurons in Rhesus Monkey Visual Cortex: Systematic Relation between Time of Origin and Eventual Disposition. *Science* **183**:425–427. doi:10.1126/science.183.4123.425

Sanai N, Nguyen T, Ihrie RA, Mirzadeh Z, Tsai H-H, Wong M, Gupta N, Berger MS, Huang E, Garcia-Verdugo J-M, Rowitch DH, Alvarez-Buylla A. 2011. Corridors of migrating neurons in the human brain and their decline during infancy. *Nature* **478**:382–386. doi:10.1038/nature10487

Saunders A, Macosko E, Wysoker A, Goldman M, Krienen F, de Rivera H, Bien E, Baum M, Wang S, Goeva A, Nemesh J, Kamitaki N, Brumbaugh S, Kulp D, McCarroll SA. 2018.

Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**:1015-1030.e16. doi:10.1016/j.cell.2018.07.028

Sousa AMM, Zhu Y, Raghanti MA, Kitchen RR, Onorati M, Tebbenkamp ATN, Stutz B, Meyer KA, Li M, Kawasawa YI, Liu F, Perez RG, Mele M, Carvalho T, Skarica M, Gulden FO, Pletikos M, Shibata A, Stephenson AR, Edler MK, Ely JJ, Elsworth JD, Horvath TL, Hof PR, Hyde TM, Kleinman JE, Weinberger DR, Reimers M, Lifton RP, Mane SM, Noonan JP, State MW, Lein ES, Knowles JA, Marques-Bonet T, Sherwood CC, Gerstein MB, Sestan N. 2017. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**:1027–1032. doi:10.1126/science.aan3456

Speir ML, Bhaduri A, Markov NS, Moreno P, Nowakowski TJ, Papatheodorou I, Pollen AA, Raney BJ, Seninge L, Kent WJ, Haeussler M. 2021. UCSC Cell Browser: visualize your single-cell data. *Bioinformatics*. doi:10.1093/bioinformatics/btab503

Stenman J, Toresson H, Campbell K. 2003. Identification of Two Distinct Progenitor Populations in the Lateral Ganglionic Eminence: Implications for Striatal and Olfactory Bulb Neurogenesis. *J Neurosci* **23**:167–174. doi:10.1523/JNEUROSCI.23-01-00167.2003

Stephan H, Andy OJ. 1969. QUANTITATIVE COMPARATIVE NEUROANATOMY OF PRIMATES: AN ATTEMPT AT A PHYLOGENETIC INTERPRETATION*. *Ann N Y Acad Sci* **167**:370–387. doi:10.1111/j.1749-6632.1969.tb20457.x

Sun Y, Ip P, Chakrabartty A. 2017. Simple Elimination of Background Fluorescence in Formalin-Fixed Human Brain Tissue for Immunofluorescence Microscopy. *J Vis Exp JoVE*. doi:10.3791/56188

Tasic B, Yao Z, Smith KA, Graybuck L, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller JA, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S,

Shapovalova NV, Hirchstein D, Pendergraft J, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. 2017. Shared and distinct transcriptomic cell types across neocortical areas. doi:10.1101/229542

Tepe B, Hill MC, Pekarek BT, Hunt PJ, Martin TJ, Martin JF, Arenkiel BR. 2018. Single-Cell RNA-Seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons. *Cell Rep* **25**:2689-2703.e3. doi:10.1016/j.celrep.2018.11.034

Valero M, Viney TJ, Machold R, Mederos S, Zutshi I, Schuman B, Senzai Y, Rudy B, Buzsáki G. 2021. Sleep down state-active ID2/Nkx2.1 interneurons in the neocortex. *Nat Neurosci* **24**:401–411. doi:10.1038/s41593-021-00797-6

Wagner DE, Klein AM. 2020. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat Rev Genet* 1–18. doi:10.1038/s41576-020-0223-2

Wichterle H, Garcia-Verdugo JM, Herrera DG, Alvarez-Buylla A. 1999. Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. *Nat Neurosci* **2**:461–466. doi:10.1038/8131

Wichterle H, Turnbull DH, Nery S, Fishell G, Alvarez-Buylla A. 2001. In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. *Development* **128**:3759–3771.

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**:15. doi:10.1186/s13059-017-1382-0

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G, Codeluppi S, Furlan A, Lee K, Skene N, Harris KD, Hjerling-

Leffler J, Arenas E, Ernfors P, Marklund U, Linnarsson S. 2018. Molecular Architecture of the Mouse Nervous System. *Cell* **174**:999-1014.e22. doi:10.1016/j.cell.2018.06.021

Zhang K, Sejnowski TJ. 2000. A universal scaling law between gray matter and white matter of cerebral cortex. *Proc Natl Acad Sci* **97**:5621–5626. doi:10.1073/pnas.090504197

Zhu Y, Sousa AMM, Gao T, Skarica M, Li M, Santpere G, Esteller-Cucala P, Juan D, Ferrández-Peral L, Gulden FO, Yang M, Miller DJ, Marques-Bonet T, Imamura Kawasawa Y, Zhao H, Sestan N. 2018. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**:eaat8077. doi:10.1126/science.aat8077

Chapter 3: Conclusions and Future Directions

Introduction

This thesis provides a review of the development, function and evolution of the brain in chapter 1, while chapter 2 builds a taxonomy of the inhibitory neurons in the euarchontogliran forebrain, highlighting two specific examples of cell type evolution in the developing brain. In the first example, I found that the TAC3+ striatal interneuron, which had been identified as a primate-specific type of neuron in the adult brain, is in fact distinct from its initial postmitotic state. Of the 16 initial classes of inhibitory neuron that I identified in the forebrain, this was the only initial class to be absent in primate or mouse, and even expanding my search to the entire brain I have yet to identify another novel initial class, which suggests that evolution at this developmental juncture may be relatively rare. In the second example, I found differences in the distribution of dorsal LGE-derived inhibitory neurons in the primate brain, which almost exclusively migrate to the olfactory bulb in mice. It turned out this was driven by large chains of these cells migrating into the primate cortex, and I proposed that the allometric reduction of the olfactory bulb in the primate brain may be a driving force in the evolution of the distribution of this cell type. I was able to follow these migratory streams in late development, which led us to identify the *striatum laureatum* neurons, a sister cell type of olfactory bulb dopaminergic neurons which appeared specific to the primate brain. Through this project I have identified two new examples of cell type evolution in the primate brain which adds to a growing number of differences identified in the brains of humans and other mammals.

This work has raised a number of future directions that are being pursued by my heirs and collaborators. Firstly, the discovery that TAC3+ striatal interneurons are developmentally specified raises questions of how this novel cell type is specified. Therefore the Pollen lab is delving into this. One aspect requires using lineage tracing to reveal whether there is a novel progenitor population which gives rise specifically to the TAC3+ neurons, or if it is the ancestral progenitor that gains the potency to produce a new initial class. It is also an open question how striatal interneurons are specified in the first place, and whether there is a difference in signaling that drives the difference in the two striatal interneuron sister cell types. Luckily this can be tested *in vitro* using human and non-human primate induced pluripotent stem cells.

Secondly, my findings about the large numbers of dLGE-derived chains in the arc migratory stream open myriad questions. It's completely unknown what happens to these cells between development and adulthood in large brains, and a full characterization of the neurons of the white matter is an important step to understanding the cortex. It would be interesting to attempt an olfactory bulb-ectomy in the mouse and observe whether the distribution of neurons, as has been performed by Arturo Alvarez-Buylla's group without particular attention to the cell type distribution.(Kirschenbaum et al., 1999) Alas, with the technical challenges of performing these surgeries being beyond my expertise, I am pursuing this question by focusing on a natural experiment, namely the loss of the olfactory bulb in toothed whales, for which I have collected brains and thus will be a part of my work in the next stages of my career. Lastly, Mercedes

Paredes' lab has also observed similar different compositions in the different migratory divisions of the arc, and is hard at work in answering the many questions of how and why these cells migrate so late in the development of large-brained mammals. Further phylogenetic comparisons will also be necessary to determine whether the arc migratory stream convergently evolved in larger brain mammals (Catarrhini, Cetartiodactyla) or is repeatedly lost in species with reduced brain size (marmoset, mouse).

Challenges in development and evolution

Despite the great power of single-cell genomics to observe cell types in high resolution, in identifying and comparing the initial classes of inhibitory neurons in monkey and mouse, I encountered a number of shortcomings in the tools for doing cross-species analysis. First and foremost, while single-cell RNA sequencing (scRNA-seq) has provided unprecedented resolution of cell heterogeneity within species, the lack of robust bioinformatic tools and comprehensive databases to accurately map homologous cell types across species has proven to be a significant limitation. For this reason, I determined that the only way to accurately compare the cell types was to cluster and label homologous cell types by hand across species, an arduous process which relies on the judgment of the individual scientist, and thus is not scalable or unbiased.

Secondly, the evolutionary distance between species introduces an additional layer of complexity into transcript models of homologous genes that makes cross species transcript orthology a challenge. While recent bioinformatics tools leveraging long read isoform

sequencing, graph genomes and a wealth of genome assembly advances, calling transcript homology remains imperfect and thus an artificial source of variation between species. (Fiddes et al., 2018; Kirilenko et al., 2023) Once genes themselves are quantified, it is also not trivial to statistically determine which genes are more highly expressed between species. Most differential expression methods rely on a negative binomial generalized linear model (Love et al., 2014; Robinson et al., 2010), and the assumptions that overall gene expression histograms can be quantile normalized by a simple scalar, that most genes are not differentially expressed, and that equal numbers of genes are up and down regulated. Though specific methods exist to properly normalize sequencing data across species (Zhou et al., 2019), these were designed for bulk RNA-sequencing.

Philosophical models of cell type evolution have been developing for more than 100 years. Meanwhile the single cell genomics revolution has also changed much about how we understand the definition and diversity of cell types. As such, the field has not yet invented mathematical models that describe cell types evolution in a principled way while capturing the rich complexity and scale of data that this technology offers. In this final chapter, I describe progress towards the building of generative models that attempt to capture the landscape of development while modeling evolutionary transcriptional divergence

Current Solutions

As the effort to map the the developing brain and to build an ontology of cell types(Miller et al., 2020; Regev et al., 2017) across time spans many independent projects around the world, it is

necessary to draw upon data from many studies and to compare them together. While this problem is a worthy challenge in the adult brain, during development it is especially difficult. Cell types are dynamic during development. This means that there is no simple way to draw a fence around a group of cells and assume that they are discrete. This is especially important in calculating differential expression across species, as differences in composition across species due to sampling will cause many differentiation-related genes to be differentially expressed if this variation is not accounted for. Likewise, the diversity of progenitor cells is often masked by cell cycle gene signatures, which may consist of only a few regional genes which drive descendant cells down a specific differentiation cascade.

Generally studying cross-species differences has been treated as a "data harmonization" problem. Generally, data harmonization is the integration of two or more transcriptomics datasets into a single dataset for downstream analysis. This term is loosely used instead of "batch effect correction" to include the use case where input datasets can be generated using different technology, and include samples with a varying composition of cell types. A broad variety of methods have been formulated to tackle this essential problem, initially designed for microarrays bulk RNA sequencing, and include methods like ComBat(Johnson et al., 2007), which uses generalized linear models and utilize empirical Bayes shrinkage to prevent over-correction. In the few years since the single-cell revolution, a plethora of dataset harmonization or batch correction methods have been developed. In most cases, the algorithms used for

clustering cell types within a species are not directly applicable to cross-species comparisons, and that methods that attempt to assign homology across species are only loosely built on any biological model. These include KNN-graph based methods that find mutual nearest neighbors across datasets and merge them such that mutual nearest neighbors are at the same points of a shared latent space.(Haghverdi et al., 2018) This also includes methods which assume anchors across datasets and fits a dimension reduction upon the data,(Andreatta and Carmona, 2020; Korsunsky et al., 2019) or that apply nonlinear transformations to merge samples by condition with neural networks(Johansen and Quon, 2019).

A model of cell type evolution

Alas, while many methods have been invented to attempt to deal with the large dataset batch effects that exist in current single cell genomics methods, very few are specifically designed for the evolution problem, and to my knowledge, none attempt an explicit model of cell type evolution. Herein I describe the class of probabilistic generative models I have formulated to model cell types in evolution, in effect solving the cross-species "harmonization" problem by accounting for putative forms of gene expression divergence.

This model combines ideas of a matrix decomposition dimension reduction method that generates count data (Townes et al., 2019), and models various putative sources of variation as modifications to this matrix decomposition. It assumes that each cell of cell type κ exists at a point on an ancestral latent space ζ. This latent space represents the relative abundance of gene coexpression networks, where the ancestral association of genes to networks (latent space components) is a matrix V, and the abundance of transcripts in the ancestral cell type is

represented by $\zeta \times V^T$. In order to model the gene expression of the extant species, I also propose a system for decomposing the differential expression (DE). Whereas standard DE modeling would assume the frequentist standpoint that each gene in each cell type is independent, I propose that DE can be broken into at least three different forms of gene expression divergence. This is possible because it has been widely shown that gene expression is organized into modules, with many genes changing together across cell types.(Harris et al., 2021; Lee et al., 2020; Suresh et al., 2022) As such, I argue that if many genes change together within organisms, and also change together in a cell type across species, this change should be called *differential by module* (δ_{DM}). If a gene gains or loses covariance with a group of genes which otherwise covary, this change should be called *differential by coexpression* (δ_{DC}). Finally, if a gene follows the classical frequentist assumption, that it is best described as a single gene changing in a single cell type, it should be called *differential by identity* (δ_{DI}). (**Figure 3.1**) Expressed in this way, this model is formulated such that it should return the canonical DE as the sum of δ_{DM}, δ_{DC}, and δ_{DI}. Importantly, I note that each of these categories may implicate different causal mechanisms, for instance δ_{DM}, with many genes changing together, suggests that a regulator of this module has gained or lost potency, hence implicating a *trans* change. On the other hand δ_{DC}suggests that the regulators of a module have gained or lost potency over a single gene, which suggests that this gene expression divergence has been driven by a *cis* change.

This model is learned using variational inference and is built upon the SCVI framework.(Lopez et al., 2018) As such, the model is learned as a Bayesian variational autoencoder, where the input and output is the raw counts, and the generative model fills the role of decoder.(**Figure**

3.1) This is in contrast to traditional variational autoencoders, for rather than a latent space which is nonlinearly decoded by a multi-layer neural network, I decode the structured latent space linearly. As such, it is best described as a gaussian mixture model in the latent space, which is decoded linearly by the batch and species effects. In the encoder, the variational distribution of the parameters is learned by the encoder f:

$$\hat{\zeta}_{\mu}, \hat{\zeta}_{\sigma} = f\left(Y\right)$$

$$\hat{\zeta} \sim Normal\left(\hat{\zeta}_{\mu}, \hat{\zeta}_{\sigma}\right)$$

The generative model then works as follows:

$$\kappa \sim Categorical\left(g(\hat{\zeta})\right)$$

$$\zeta \sim Normal\left(\zeta_{\kappa\mu}, \zeta_{\kappa\sigma}\right)$$

$$\ell \sim LogNormal\left(\ell_{\mu}, \ell_{\sigma}^{2}\right)$$

$$\beta, \delta, V \sim Laplace\left(0, 100\right)$$

$$\hat{\mu}_{\kappa sb} = softmax\left(\left(\zeta_{anc} + \delta_{DM} + \beta_{b\kappa}\right) \times \left(V + \delta_{DC}\right)^{T} + \delta + \delta_{DI} + \beta_{b}\right)$$

$$\mu_{\kappa sb} = log\left(\ell\hat{\mu}_{\kappa sb}\right)$$

$$Y \sim NegativeBinomial\left(\theta, sigmoid\left(\mu_{\kappa sb} - log\left(\theta\right)\right)\right)$$

Where Y is the reconstructed UMI counts for genes in cells, $\hat{\zeta}$ is the variational estimation of the latent space, $\zeta_{\kappa\mu}$ and $\zeta_{\kappa\sigma}$ are the parameters of each the gaussians that generate a given cell type, ℓ is the scale factor (sum of counts), κ is the discrete cluster a cell belongs to, $f()$ represents the encoder deep neural network, $g()$ represents the classifier deep neural network, δ are species (s) effects, V is the weight matrix, β are batch (b) effects, and θ is the n total

failures parameter of the negative binomial distribution. Note that $\mu_{\kappa sb}$ represents the logits of the negative binomial p parameter, and represents the expected count value given κ, s, b, and ℓ .

I assume that counts are generated by a non-zero-inflated negative binomial distribution, as modeling zero inflation adds additional modeling complexity and it is unclear whether or not it provides any benefit.(Choi et al., 2020; Jiang et al., 2022; Townes et al., 2019) The model above is learned via variational inference to find variational distributions for the parameters, using the pyro probabilistic programming language as a platform as in SCVI.(Bingham et al., 2019; Lopez et al., 2018) As such, the variational distributions $q\left(\zeta, \ell | Y\right)$ (the encoder) and $q\left(\kappa | \zeta\right)$ (the classifier) are learned by deep neural networks, stacking batchnorm and leaky ReLU layers. Inference is performed using the Adam optimizer, and as such, approaches a minimum by stochastic gradient descent, optimizing for the evidence lower bound of the Bayesian model. (**Figure 3.1**)

Towards the future of evolutionary neuroscience

In considering where both I and the field should go from here, I am often reminded of the words of University of Wisconsin biochemistry professor Michael M. Cox: "In life, structure confers function". The structure of the DNA underlies chromatin, the structure of chromatin underlies the transcriptome, the structure of transcripts translate into how they become the proteome, the structure of proteins defines their function, and the structure of all of these within cells determine how cells will function and work together in the tissues that make up an organism.

While it has been only a decade since the ability to read out the transcriptome, the number of ways by which we can observe biology with single cell resolution is quickly proliferating. New technologies, like spatial transcriptomics and single cell multi-omics, open new avenues by which large scale omics studies can be used to generate new discoveries. While the most biologically interpretable data modality remains transcriptomics, these new modalities form bridges across levels of abstraction in biology.

Given a change in the transcriptome, it cannot be considered understood until its ultimate cause has been identified. As such we must go back in the causal chain to identify the gene regulatory element, or coding mutation in a regulator that could have caused it. For these purposes it is very helpful to use chromatin accessibility in the cell type of interest to identify potential regulatory elements, as the genome is billions of bases and so it is useful to shrink the haystack when one goes searching for a needle. Indeed, it has been shown that changes in noncoding DNA regulatory elements is likely a key mechanism by which cell type evolution takes place, as many of the rapidly evolving regions of the human genome have been shown to act as enhancers.(Capra et al., 2013; Pollard et al., 2006) Likewise, *trans* factor binding assays like "ChIP-seq" or "Cut and Tag", can inform our distribution on likely causes of a transcriptomic change. As a final point, this tracing back the causal chain from an alteration in an adult brain also often means "going back in time" to look at development. The cell types of an adult brain are quite stable over time, and so most evolutionary changes to structure or function of cell

types must be manifest during development, and functional genomics approaches like CRISPR screens promise to prove or disprove causality of specific base pair changes that propagate developmentally to the adult.(Tabula Muris Consortium, 2020; Tasic et al., 2018)

On the other hand, in determining how changes at the molecular level translate into how the brain works, it is necessary to go forward in the causal chain. If a gene is differentially expressed, first we must consider whether it actually represents a differential in its function, either in translated protein level or otherwise. Given a high probability of this, we must then consider what its effect on the properties of the cell are likely to be, and to validate these. Then, at last, we can consider what the function of this cell is in its structure and circuit in the brain, and how this may alter the cognition and behavior of an animal. (**Figure 3.2**)

One can never accurately predict where the next revolution in science will come from, but my personal bet is that it will come from the use of deep learning-powered models that can integrate knowledge from across levels of biological extraction. The Bayesian approach is a well-established way to integrate data from previous experiments and modalities given a model of how they relate to each other. Though these models are limited by human creativity in their formulation and the difficulty of learning the models, it is possible that new iterations of Large Language Models will be helpful in translating the preponderance of the scientific literature into both model structures and priors on the distributions in those models. On top of this, the last few years have seen the expansion of unbiased observation at single cell resolution from

transcriptomics to the genome, DNA methylation, the epigenome, protein-DNA footprinting, the epitranscriptome, low-dimensional proteome and electrophysiology. As these methods become established and datasets are generated around the world, there will be immense opportunity to see new connections in the data. The confluence of innovative computational models, advancements in single-cell genomics techniques, and the proliferation of datasets across a diversity of species heralds an exciting epoch in our quest to understand brain evolution. By bridging computational prowess with cutting-edge biological experimentation, will undoubtedly illuminate the profound mysteries of how diverse brains are built and shape the next chapter of discovery in neuroscience.

Figures

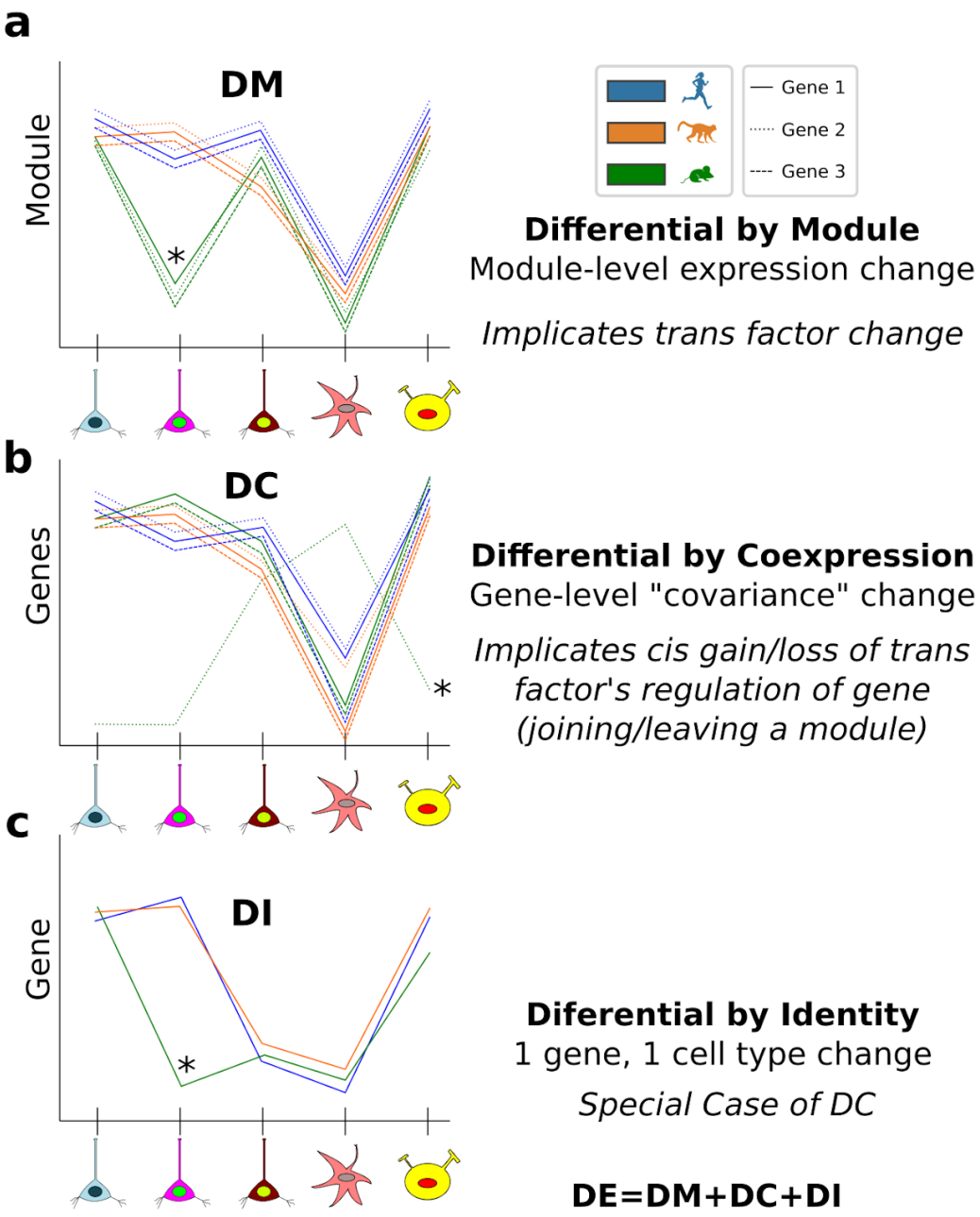
a
DM
Module
*
Gene 1
Gene 2
Gene 3
Differential by Module
Module-level expression change
Implicates trans factor change
b
DC
Genes
*
Differential by Coexpression
Gene-level "covariance" change
Implicates cis gain/loss of trans factor's regulation of gene (joining/leaving a module)
c
DI
Gene
*
Diferential by Identity
1 gene, 1 cell type change
Special Case of DC
DE=DM+DC+DI

Figure 3.1: Classes of gene expression divergence

a. Schematic of differential by module genes, with the group of green genes differentially expressed in one cell type. **b.** Schematic of differential by coexpression gene, with one gene from the group of green genes differentially coexpressed across cell types. **c.** Schematic of differential by identity gene, with a single gene differentially expressed in a single cell type.

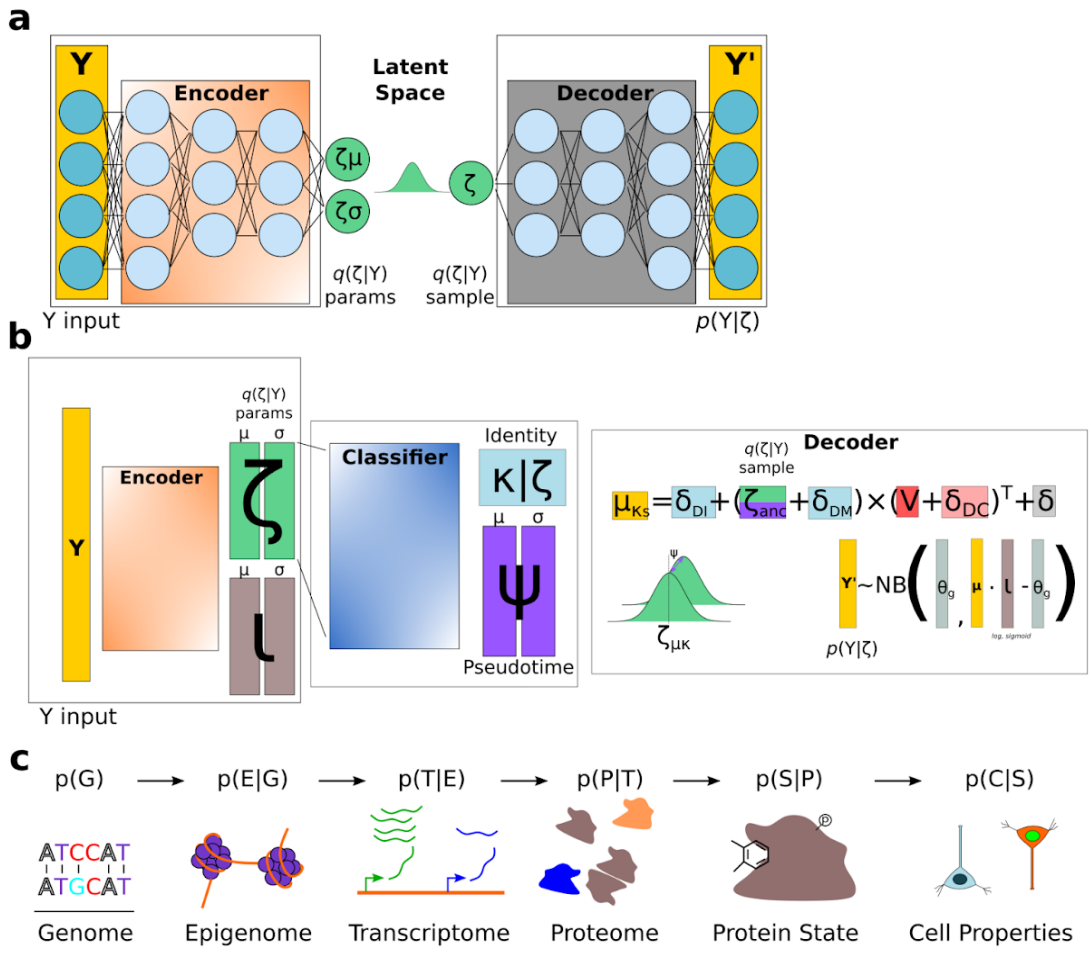

Figure 3.2: Deep learning models of gene expression and cellular phenotypes

a. Schematic of a traditional variational autoencoder. **b.** Schematic of the proposed model, a modified variational autoencoder that includes a discrete cell type classification by which the latent space is generated. NB represents the negative binomial model by which counts are assumed to have been generated. **c.** Schematic of the proposed causal chain by which information from each level of abstraction can be used to predict the distribution of features of the one above.

References

Andreatta M, Carmona SJ. 2020. STACAS: Sub-Type Anchor Correction for Alignment in Seurat to integrate single-cell RNA-seq data. *Bioinformatics* **37**:882–884. doi:10.1093/bioinformatics/btaa755

Bingham E, Chen JP, Jankowiak M, Obermeyer F, Pradhan N, Karaletsos T, Singh R, Szerlip P, Horsfall P, Goodman ND. 2019. Pyro: Deep Universal Probabilistic Programming. *J Mach Learn Res* **20**:1–6.

Capra JA, Erwin GD, McKinsey G, Rubenstein JLR, Pollard KS. 2013. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci* **368**:20130025. doi:10.1098/rstb.2013.0025

Choi K, Chen Y, Skelly DA, Churchill GA. 2020. Bayesian model selection reveals biological origins of zero inflation in single-cell transcriptomics. *Genome Biol* **21**:183. doi:10.1186/s13059-020-02103-2

Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, Haussler D, Stanke M, Paten B. 2018. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res*. doi:10.1101/gr.233460.117

Haghverdi L, Lun ATL, Morgan MD, Marioni JC. 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36**:421–427. doi:10.1038/nbt.4091

Harris BD, Crow M, Fischer S, Gillis J. 2021. Single-cell co-expression analysis reveals that transcriptional modules are shared across cell types in the brain. *Cell Syst* **12**:748-756.e3. doi:10.1016/j.cels.2021.04.010

Jiang R, Sun T, Song D, Li JJ. 2022. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biol* **23**:31. doi:10.1186/s13059-022-02601-5

Johansen N, Quon G. 2019. scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. *Genome Biol* **20**:166. doi:10.1186/s13059-019-1766-4

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**:118–127. doi:10.1093/biostatistics/kxj037

Kirilenko BM, Munegowda C, Osipova E, Jebb D, Sharma V, Blumer M, Morales AE, Ahmed A-W, Kontopoulos D-G, Hilgers L, Lindblad-Toh K, Karlsson EK, Consortium‡ Z, Hiller M. 2023. Integrating gene annotation with orthology inference at scale. *Science*. doi:10.1126/science.abn3107

Kirschenbaum B, Doetsch F, Lois C, Alvarez-Buylla A. 1999. Adult Subventricular Zone Neuronal Precursors Continue to Proliferate and Migrate in the Absence of the Olfactory Bulb. *J Neurosci* **19**:2171–2180. doi:10.1523/JNEUROSCI.19-06-02171.1999

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P, Raychaudhuri S. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**:1289–1296. doi:10.1038/s41592-019-0619-0

Lee J, Shah M, Ballouz S, Crow M, Gillis J. 2020. CoCoCoNet: conserved and comparative co-expression across a diverse set of species. *Nucleic Acids Res* **48**:W566–W571. doi:10.1093/nar/gkaa348

Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. 2018. Deep generative modeling for single-cell transcriptomics. *Nat Methods* **15**:1053–1058. doi:10.1038/s41592-018-0229-2

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8

Miller JA, Gouwens NW, Tasic B, Collman F, van Velthoven CT, Bakken TE, Hawrylycz MJ, Zeng H, Lein ES, Bernard A. 2020. Common cell type nomenclature for the mammalian brain. *eLife* **9**:e59928. doi:10.7554/eLife.59928

Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, Rosenbloom KR, Kent J, Haussler D. 2006. Forces Shaping the Fastest Evolving Regions in the Human Genome. *PLOS Genet* **2**:e168. doi:10.1371/journal.pgen.0020168

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundeberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, Oudenaarden A van, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N, Participants HCAM. 2017. Science Forum: The Human Cell Atlas. *eLife* **6**:e27041. doi:10.7554/eLife.27041

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. doi:10.1093/bioinformatics/btp616

Suresh H, Crow M, Jorstad N, Hodge R, Lein E, Dobin A, Bakken T, Gillis J. 2022. Conserved coexpression at single cell resolution across primate brains (preprint). Systems Biology. doi:10.1101/2022.09.20.508736

Tabula Muris Consortium. 2020. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**:590–595. doi:10.1038/s41586-020-2496-1

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller J, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirschstein D, Pendergraft J, Sullivan HA, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**:72–78. doi:10.1038/s41586-018-0654-5

Townes FW, Hicks SC, Aryee MJ, Irizarry RA. 2019. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol* **20**:295. doi:10.1186/s13059-019-1861-6

Zhou Y, Zhu J, Tong T, Wang J, Lin B, Zhang J. 2019. A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinformatics* **20**:1–10. doi:10.1186/s12859-019-2745-1

Publishing Agreement

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

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

Author Signature

8/30/2023

Date