Genetic Fate Specification of Inhibitory Neurons

May Cheng-Ching Ho

München 2023

Genetic Fate Specification of Inhibitory Neurons

May Cheng-Ching Ho

Dissertation

an der Biologie

der Ludwig–Maximilians–Universität

München

vorgelegt von

May Cheng-Ching Ho

aus Surrey, Canada

München, den 20.06.2023

Erstgutachter: Prof. Dr. Rüdiger Klein

Zweitgutachter: Prof. Dr. Wolfgang Enard

Tag der Abgabe: 20.06.2023

Tag der mündlichen Prüfung: 07.02.2024

EIDESSTATTLICHE RKLÄRUNG
Ich versichere hier mit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist.
Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungs kommission vorgelegt.
Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 20.04.23

May Cheng-Ching Ho

Our wisdom, we prefer to think,
is all of our own gathering, while,
if the truth be told, it is, most of it,
the last coin of a legacy that dwindles with time.

EVELYN WAUGH, BRIDESHEAD REVISTED

Table of Contents

List of Figures

List of Tables

Abstract

Understanding how an enormous diversity of neuronal cell types is generated has been a major objective of neurobiology. This task is particularly challenging in the case of inhibitory neurons because they migrate long distances during development. It is believed that a combination of intrinsic factors and external signals influence progenitor cells to differentiate into distinct types of inhibitory cells, such as interneurons or long-range projection neurons.

To tackle this issue, one approach involves examining the clonal relationships between inhibitory cell types in the brain. In this thesis, I established a single-cell RNA sequencing compatible, lineage-tracing method, TrackerSeq, that enables both the identity of a neuron and its developmental history to be retrieved simultaneously to analyze developmental relationships of inhibitory cell types in the mouse brain. TrackerSeq achieves this by tagging progenitors with inheritable DNA barcodes followed by transcriptome sequencing at a later time point to analyze developmental relationships of inhibitory cell types in the mouse brain. Using TrackerSeq, I found different inhibitory cell types occupying different regions in the brain shared inherited the same lineage barcodes, suggesting that mitotic progenitors can give rise to different cell types.

Subsequently, I explored whether specific transcription factors expressed in inhibitory neurons, such as Meis2 and Lhx6, play a crucial role in determining the fate of inhibitory cell types. Single-cell sequencing compatible perturbation methods, like tCROP-seq, have emerged as an effective way to interrogate the impact of these factors on the outcome of neuronal fates. In a typical tCROP-seq protocol, sgRNAs are delivered to cycling progenitors via in utero electroporation to introduce fameshift mutations in genes of interest, followed by sequencing of neurons at a later timepoint.

By analyzing the tCROP-seq data obtained from perturbing Meis2, I observed that interneuron genes were upregulated in projection neuron cell types, leading to an increased proportion

of interneurons. Interestingly, perturbing Lhx6 had the opposite effect. These findings suggest that when Meis2 is perturbed, progenitor cells originally destined to become projection neurons may instead differentiate into interneurons. To confirm this possibility, I employed TrackerSeq barcodes to tag non-perturbed and Meis2-perturbed cells. The analysis revealed that Meis2-perturbed mitotic cells shared more clones with interneurons than projection neurons, providing further evidence that Meis2 perturbation promotes the preferential differentiation of progenitor cells into interneurons.

My findings reveal that specification of inhibitory subtypes already takes place at the progenitor stage and require the expression of select transcription factors. Gaining a better understanding of how genetic programs such as lineage and transcription factor expression influence subtype specification can improve our modeling of neurodevelopmental disorders.

Abbreviations

GABA γ-amninobutyric acid

PV parvalbumin

MSN medium spiny neuron

VZ ventricular zone

SVZ subventricular zone

GE ganglionic eminence

POA preoptic area

MGE medial ganglionic eminence

CGE caudal ganglionic eminence

LGE lateral ganglionic eminence

GEs ganglioninc eminence

GFP green fluorescent protein

scRNA-seq single-cell RNA sequencing

scLT single-cell lineage tracing

RNA-seq RNA-sequencing

TF transcription factor

DEA differential expression analysis

DEG differentially expressed gene

NGS next-generation sequencing

LB lineage barcodes

IUE *in utero* electroporation

NHEJ non-homologous end joining

ITC intercalated cell

PN projection neuron

IN interneuron

1 Introduction

1.1 Developmental origins of GABAergic neurons

The adult mammalian forebrain has long fascinated scientists with its ability to form conscious thoughts, emotions, and store memories. Such tasks are due to billions of neurons that form thousands of connections with one another, including both glutamatergic excitatory and GABAergic inhibitory cells. Excitatory neurons comprise the largest proportion of cortical cells and are responsible for transmitting information across different regions of the brain through long-range connections. Although inhibitory neurons only represent around 10-20% of the brain, they play an important role in preventing or inhibiting the firing of other neurons. Most GABAergic neurons use the GABAergic neurotransmitter GABA, which binds to synaptic receptors, to dampen nerve cell activity. They are proportionally more abundant in human brain than other species and exhibit different molecular, morphological, and physiological properties.

Ramon y Cajal was one of the first neuroscientists to explore the diversity of GABAergic neurons. By combining microscopy with an innovative staining technique, he documented the intricate morphology of different cortical GABAergic interneurons, which he called "butterflies of the soul". Since Cajal's Nobel-prize-winning work in 1906, our understanding of how this incredible diversity is generated has advanced considerably.

GABAergic neurons can be divided into two main types: interneurons or projection neurons. Interneurons generally project locally and are characterized by their diverse morphology, connectivity pattern, expression of molecular markers, and electrophysiological properties (Kepecs and Fishell 2014). For example, parvalbumin (PV) interneurons are a subset of fast-spiking interneurons. Long-range GABAergic projection neurons, as their name would suggest, project to other regions of the brain with distinct function (Caputi et al. 2013). The

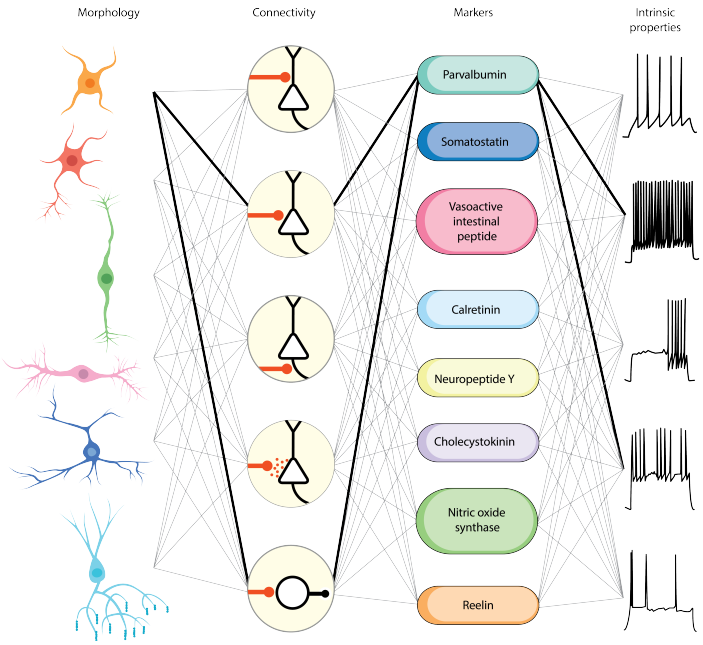

Figure 1: Multiple dimensions of interneuron diversity. Interneuron cell types are usually defined using a combination of criteria based on morphology, connectivity pattern, synaptic properties, marker expression and intrinsic firing properties. The highlighted connections define fast-spiking cortical basket cells. Taken from Kepecs and Fishell 2014.

most abundant GABAergic projection neurons, called medium spiny neuron (MSN)s, are located in the striatum, where they make up 95% neurons in that region.

After the GABAergic cell types are born, they migrate long distances tangentially to settle in various regions such as the cortex, striatum, hippocampus, amygdala or olfactory bulb (Bandler, Mayer, and Fishell 2017). Typically, GABAergic interneurons are found mainly in the cortex, while long-range GABAergic neurons are located in subcortical regions.

1.1.1 GABAergic neurons are born from ganligoninc eminences

Inhibitory interneurons are born separately from their excitatory counterparts during development and do not have a shared origin (Sultan and Shi 2018). Radial glia residing in the ventricular zone (VZ) of the embryonic subpallium generate all GABAergic neurons in the telencephalon until the subventricular zone (SVZ) later takes over the VZ as the main site of

cell proliferation (Garcia and Harwell 2017). Postmitotic cells derived from these proliferative regions move basally into the mantle zone, where they migrate to their final location.

In mice, GABAergic neurons are generated from E11 to E17 in different parts of the ventral telencephalon, in regions termed ganglionic eminence (GE)s and preoptic area (POA). The GE is a transient structure during embryonic development. It first appears as protrusions in the lateral ventricles at E11 in the mouse; as embryonic development comes to an end, the morphological boundaries within the GE disappear until it is no longer visible in the postnatal brain. The GEs can be divided into three subregions: medial, caudal, and lateral GE (MGE, CGE, and LGE, respectively).

A great diversity of cell types is generated from subpallial progenitors that reside in the GE. They produce progenitors that eventually differentiate into non-overlapping neuronal subtypes. Around 60% cortical interneurons are generated from the medial ganglionic eminence (MGE) and express PV or somatostatin (Xu et al. 2004; Butt et al. 2005; Miyoshi et al. 2007). The lateral ganglionic eminence (LGE) is located near the developing neocortex and gives rise to the interneurons of the olfactory bulb, as well as the MSNs of the striatum. The LGE and MGE fuse into caudal ganglionic eminence (CGE), which is towards the caudal end of the telencephalon. Producing a smaller proportion of interneurons, CGE produces cortical interneurons that express the vasoactive intestinal peptide or the Reelin glycoprotein (Bandler, Mayer, and Fishell 2017).

1.1.2 Inhibitory neuron specification by transcription factors

Here, we will dive into the cell-autonomous role of transcription factor (TF)s in inhibitory neuron specification. Loss-of-function studies have identified several early-expressing TFs within inhibitory neurons and their progenitors that are necessary for informing cell fate. Early-expressing TFs tend to be broadly expressed in other ventral telencephalic lineages or are required for early regional patterning and identity of GEs in general (Miyoshi, Machold, and

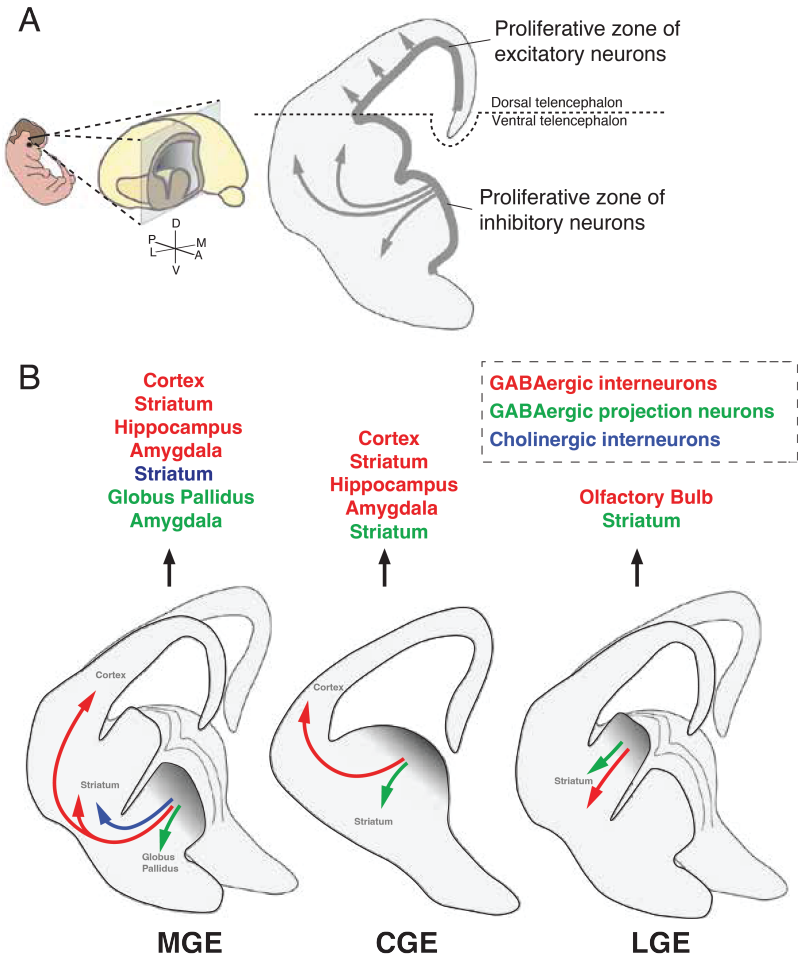

Figure 2: Ganglionic eminences produce GABAergic interneurons and projection neurons. A) Image showing the boundary between the dorsal and ventral telencephalon. The ventral telencephalon produces GABAergic neurons that migrate long distances to different regions in the brain. The dorsal telencephalon generates excitatory cells that migrate short distances radially into the cortex. **B)** Diagram illustrating the various cell types that each ganglionic eminence produces and the brain structures they occupy, which includes GABAergic interneurons (red), GABAergic projection neurons (green), and Cholinergic interneurons (dark blue). Taken from Bandler, Mayer, and Fishell 2017.

Fishell 2013). For example, sonic hedgehog and fibroblast growth factor signaling induce transcriptional pathways are crucial for the proper specification of the MGE, LGE, and CGE (Guillemot and Zimmer 2011; Hébert and Fishell 2008).

Once ganglioninc eminence (GEs) are formed, the birth of more specific types of inhibitory

neurons requires combinations (specific codes) of dynamic TFs expressed at mitotic and postmitotic time points (Lodato and Arlotta 2015). Commitment to the inhibitory lineage is initiated by transcription factors *Dlx1/2*, which then go on to activate downstream TFs (Petryniak et al. 2007). Later, the homeobox1 gene Nk2 homeobox1 (*Nkx2-1*) is expressed in the MGE/POA, where it is critical for the establishment and maintenance of progenitors in the VZ and SVZ. NKX2-1 and its downstream TF, LHX6, are necessary for the development of MGE-derived interneurons, since their absences cause MGE-derived interneurons to fate switch into more CGE and LGE-like derived inhibitory subtypes.

Expressed primarily in the LGE/CGE neuroepithelium, the homoebox factor GSX2 appears to be another TF involved in interneuron specification. It controls the expression of other pro-neural genes such as *Dlx2*, *Olig2*, and *Ascl2*, and its loss leads to a selective reduction in calretinin-expressing interneurons in the cortex (Waclaw et al. 2009). Other TFs involved in CGE-derived interneuron production include NR2f1, NR2f2, PROX1, and ADARB2 (Flames et al. 2007; Miyoshi et al. 2015). The LGE ventricular zone is characterized by the expression of Gsx2, Er81, and low levels of Pax6. MEIS2, a member of the TALE family of TFs containing the homeodomain, has DLX1/2 driven expression in LGE. Its deletion leads to an improper differentiation of MSNs, suggesting that it is important for the specification of LGE-derived GABAergic neurons.

Upon being postmitotic, progenitors diverge and differentiate into transcriptionally distinct precursor cell states that represent populations fated to give rise to interneurons or projection neurons (Mayer et al. 2018). Branch 1 (precursor state 1) expresses a known regulator of interneuron development (*Arx, Maf*; branch 2 (precursor state 2) expresses known projection neuron marker genes (*Isl1, Ebf1*); branch 3 contains the transcription factor Lhx8, a marker gene for MGE cholinergic populations. The expression of some of these precursor-specific genes is driven by the TFs discussed earlier. NKX2-1 induces the fate of cholernergic neurons through induction of *Lhx8*, and LHX6 drives the expression of *Arx* to promote the maturation of cortical

interneurons (Sussel et al. 1999; Vogt et al. 2014). This genetic relationship between these GE-specific TFs and these precursor-specific genes indicates that these early emerging branches of transcriptomic identity are enforced and maintained by TFs.

However, translating these region-specific patterns of gene expression into a combinatorial transcription factor code for cell fate specification has been challenging, especially for finer cell types (Garcia and Harwell 2017). Many of the TFs studied are expressed in gradients or are expressed dynamically. Therefore, it is necessary to find a scalable way to investigate which TFs promote and repress certain GABAergic fates and at which time points.

1.1.3 Intrinsic versus extrinsic determination of GABAergic neuron subtypes

It is not clear to what extent GABAergic neuron subtypes are prespecified at the precursor level and how crucial extracellular cues are during post-mitotic maturation for cell fate determination (Corbin and Butt 2011). The importance of genetic specification of inhibitory neurons, as discussed in the last section, has led researchers to present the hypothesis of progenitor specification. According to this model, cell-type-specific genetic programs are already established at birth (intrinsic), where neurons follow a specific differentiation program to develop into a particular subtype. Supporting the progenitor specification hypothesis, PV+ chandelier cells were observed to originate from a spatially restricted pool of progenitors born relatively late in embryogenesis (Miyoshi, Machold, and Fishell 2013).

In contrast, the progressive specification hypothesis posits that the identity of the cortical interneuron subtype is acquired later through interaction with environmental cues, such as activity-mediated calcium signaling (extrinsic). In fact, some studies indicate that intrinsic genetic mechanisms may not be the sole determinant of GABAergic neuron subtype identity. Many subtype-specific features of cortical interneurons—from morphology, settling position,

synapse sensitivity, and connectivity—rely on activity imprinting upon interneruons as they mature (Wamsley and Fishell 2017). Furthermore, recent lineage-tracing experiments show that clonally related interneurons born from the same progenitor are dispersed over all cortical areas and within the basal telencephalon, sharing no spatial relationship. The ones that are in proximity of each other and form isolated clusters may be influenced by environmental cues (Mayer et al. 2015; Harwell et al. 2015). It has been shown that electrical activity influences the migration patterns of postmitotic interneurons (Bortone and Polleux 2009; Garcıa, Karayannis, and Fishell 2011).

These postmitotic events do not diminish the importance of intrinsic properties for broad subtype determination (Wamsley and Fishell 2017; Pensold 2017). Rather, the mechanism of differentiation for GABAergic neurons is more likely to lie somewhere between both models, where intrinsic genetic programs are instructive for broad subtype determination and extrinsic signals are important for interneuron circuit integration and positioning after migration (Wamsley and Fishell 2017; Pensold 2017). Consistent with this idea, after migrating to the cortex, postmitotic interneurons express certain maturation genes and proteins ((such as potassium–chloride cotransporter 2 (Kcc2; also known as Slc12a5)) and the transcription factors Satb1 and Mef2c), which are modulated by calcium signaling resulting from neuron depolarization (Denaxa et al. 2012; Bortone and Polleux 2009; H. Ma et al. 2012).

Intrinsic genetic cascades are crucial in establishing the broad identity of the inhibitory neuron subtype before migration. However, it appears that intrinsic genetic signaling can be influenced by post-migration activity-mediated regulation to specify the connectivity and positioning of inhibitory neurons. Thus, both intrinsic and extrinsic mechanisms are instructive for the final identity of inhibitory neurons.

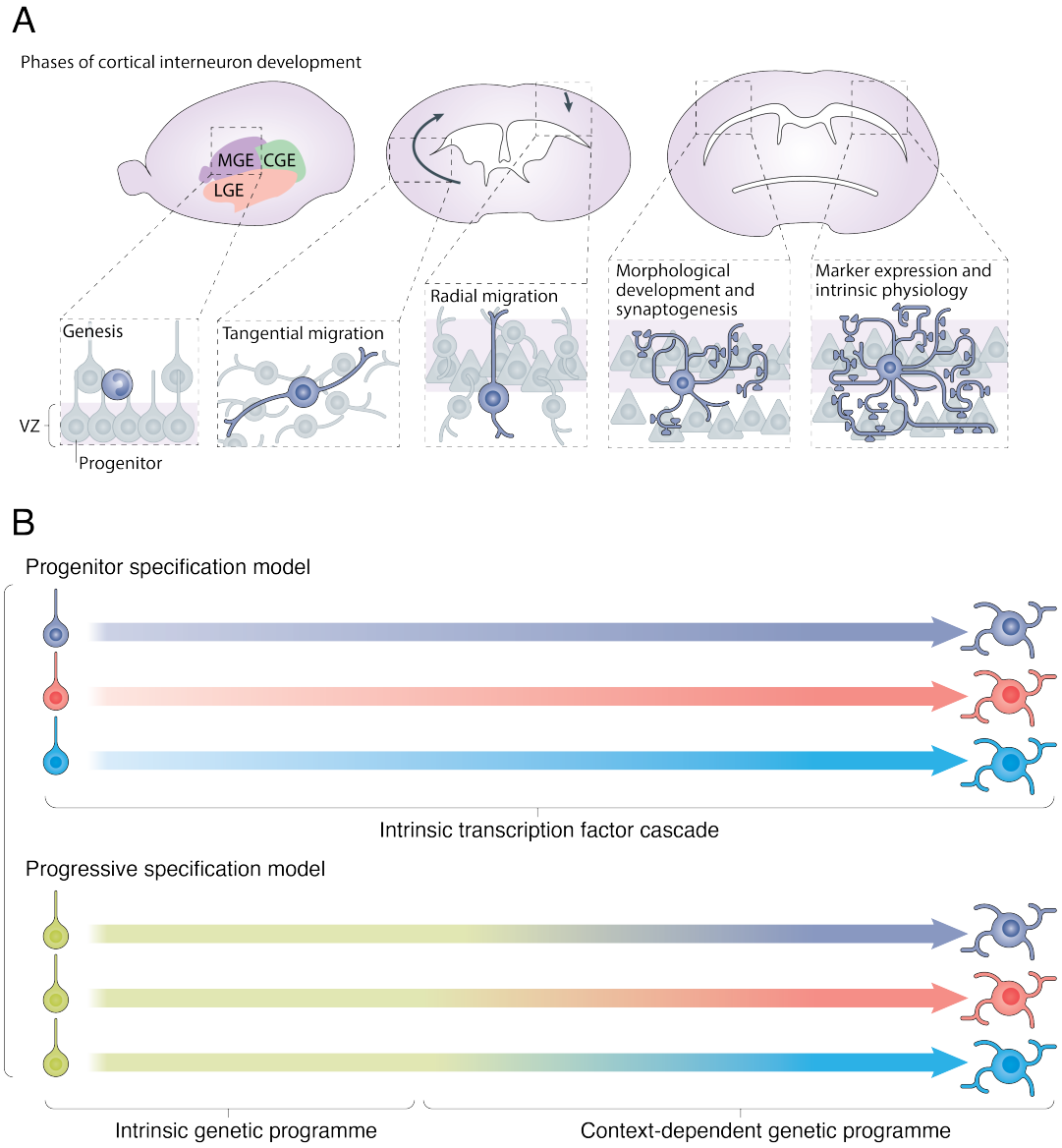

Figure 3: Models of GABAergic neuron specification. A) During embryogenesis, cortical interneurons are generated from the medial ganglionic eminence (MGE) and the caudal ganglionic eminence (CGE). They then undergo a long tangential migration to the cortex, followed by radial migration into the developing cortical layers. During postnatal development, they reach a settling position within a laminar layer and establish their distinct morphology and synaptic contacts. The expression of particular neuronal markers and physiological attributes are acquired in parallel. **B)** The progenitor specification model hypothesizes that an GABAergic neuron's subtype is genetically determined at the progenitor stage during neurogenesis. The progressive specification model speculates that an GABAergic neuron's fate is first defined with intrinsic genetic programmes, followed by context-specific changes that are later induced. Taken and modified from Wamsley and Fishell 2017.

1.2 Lineage tracing in the murine CNS

The gold standard for linking cell states across periods of time is lineage tracing, where the lineage of differentiated cells is examined by tracing it back to the cell it originates from. John Sulston's pioneering work in C. elegans is one such classic example. He elegantly demonstrated how cell lineage for a nematode's developing nervous system is unchanging and progenitors further down the lineage are fate restricted to generating only certain types of neurons (White 1986). In other words, every nematode undergoes the same program of cell division and there was correspondence between lineage ancestry and cell type. However, how cell lineage governs neuronal differentiation in vertebrates is less clear.

Certain neuronal progenitors produce progeny that are more readily influenced by environmental cues rather than intrinsic factors like cell lineage. Nevertheless, lineage has been found to play a role in the developing cortex of mice. Consecutive rounds of asymmetric cell division produce lineage-related sister excitatory neurons that migrate short distances toward the pia. After migration, this results in spatially organized vertical clusters of excitatory sibling neurons that form functional columnar microcircuits in the neocortex (Noctor et al. 2001; Li et al. 2012). Unlike their excitatory counterparts, GABAergic neurons—despite also having embryonic origins—are entirely derived from the ventral telencephalon or subpallium and migrate over large distances to integrate into the developing cortex, hippocampus or other subcortical forebrain structures (Marın and Rubenstein 2001). For this reason, it is difficult to track the complex and wide dispersion of interneurons throughout the brain with conventional techniques like time-lapse imaging.

1.2.1 Mouse engineering for lineage tracing

The earliest studies of lineage tracing of the mammalian forebrain relied on mouse genetics. Mouse genetic engineering labels specific cell populations by driving the expression of marker

genes in a cell. Specificity labelling can be achieved by using certain promoters alone in transgenics or through gene recombination facilitated by sequence-specific recombinase. In the latter case, the system comprises a Cre recombinase that recognizes the *loxP* sequence. To create this mouse line, mouse are engineered to express the recombinase under a cell-type or region specific promoter. These mice are then crossed with a reporter line, in which *loxp* transcriptional/translational stop cassette precedes a marker gene, like green fluorescent protein (GFP) or β-gal. The marker gene is typically inserted into an ubiquitously expressed locus, such as *Rosa26*. Upon Cre-induced recombination, the stop sequence is excised, and the marker gene is expressed.

Using this technique, Xu et al. (2004) could comprehensively fate-mapped Nkx2-1-lineage cells in the mouse telencephalon. In line with previous studies, they found that Nkx2-1 expressing progenitors generate subpopulations of interneurons in the striatum and cerebral cortex, as well as projection neurons of the globus pallidus. Furthermore, a minor population of putative olfactory bulb interneurons was identified.

1.2.2 Viral infection for lineage tracing

Two studies attempted to address whether clonally related interneurons are also exhibiting spatial organization (i.e., confined to discrete anatomical units) using a combination of mouse genetics and retrovirus-based fluorescent labeling. Brown et al. (2011) and Ciceri et al. (2013) found that MGE-derived clones form nonrandom, spatially isolated clusters in cortical columns or laminae. In detail, Brown et al. (2011) postulated that putative clones aligned into horizontal and radial columns similarly to their excitatory counterparts. Rather than columns, Ciceri et al. (2013) described interneuron clones forming clusters in laminae. However, they assumed that dispersed interneurons labeled with the same fluorophere were assumed ab initio to be derived from independent clones.

To overcome these shortcomings, Mayer et al. used a replication-defective retroviral libary

encoding GFP and a highly diverse set of DNA barcodes (Golden, Fields-Berry, and Cepko 1995), comprised of approximately 100, 000 random 24-bp oligonucleotide tags. The barcodes were recovered from the mature progeny of infected progenitor cells with Sanger sequencing, enabling them to specifically determine the lineal relationship between clones regardless of their geometric distribution within the brain. In contrast with the aforementioned studies, they found that clonally related interneurons could disperse broadly across both functional areas and structural boundaries, suggesting that intrinsic genetic programs are not predictive of an interneuron's ultimate circuit specificity or location in the brain.

However, it remains possible that lineage has an influence on the phenotypic identity of clonally related siblings. One hypothesis is that each progenitor domain of the GE contains progenitor cells dedicated to producing specific GABAergic neuron subtypes. Sister cells might share certain traits and form transcriptomically similar cell types across brain regions such as the cortex, hippocampus, and striatum. Such a finding would indicate that lineage predetermines the formation of distinct neuronal types. Alternatively, subtype differentiation might be independent of lineage and instead depend on factors such as environmental cues.

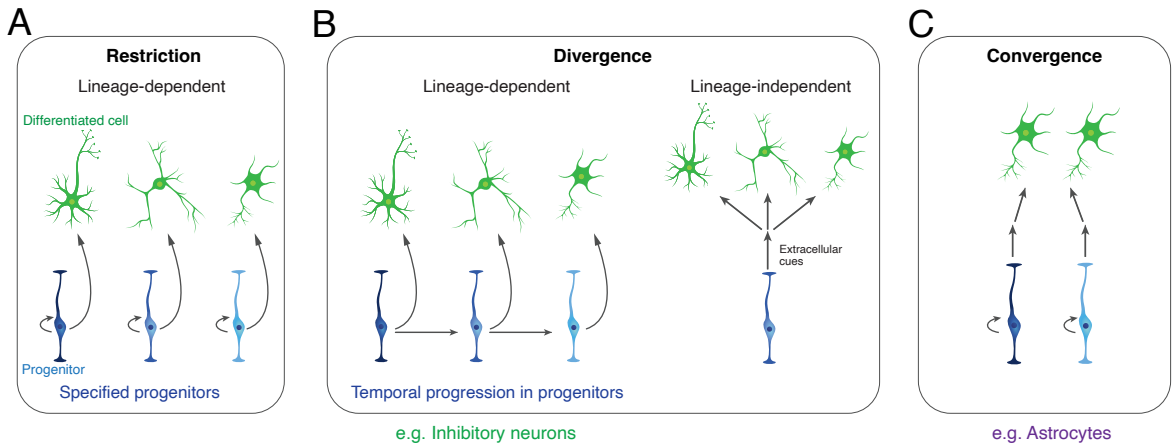

Figure 4: Schematic depicting different scenarios of how lineage could influence cell type diversity. Different cell types could either arise from fate-restricted progenitor cells **A)** or be born temporally from the same pool of progenitor cells (B, left). These lineage-dependent processes suggest that cell-intrinsic mechanisms, or local cues at the mitotic progenitor level, determine the fate of newborn cells. Alternatively, lineage-independent mechanisms such as extracellular induction, activity-dependent processes, or stochastic regulation (B, right) could drive the differentiation into different subtypes. Lineage-dependent and independent mechanisms are not mutually exclusive. Both scenarios in **B)** are examples of lineage divergence, where different cell types could arise from a common progenitor. **C)** Convergence is the process by which similar cell states arise from different lineages. Taken from Bandler et al. 2022.

1.3 Single cell RNA-sequencing

A long-standing question in developmental biology is how much neuronal differentiation is driven by intrinsic genetic programs at the progenitor stage as opposed to extrinsic factors, such as patterned brain activity, during and after migration. In order to disentangle this question, neuronal cell types must first be characterized and defined. Although neurons could be characterized by their morphology, connectivity and patterns of activity, their molecular descriptions were limited to individual genes validated by immunohistochemistry.

The emergence of high-throughput single-cell RNA sequencing (scRNA-seq) methods has revolutionized how cell states are defined and, consequently, how development is studied. Compared to other methodologies such as bulk RNA-sequencing (RNA-seq), microarrays, and in situ

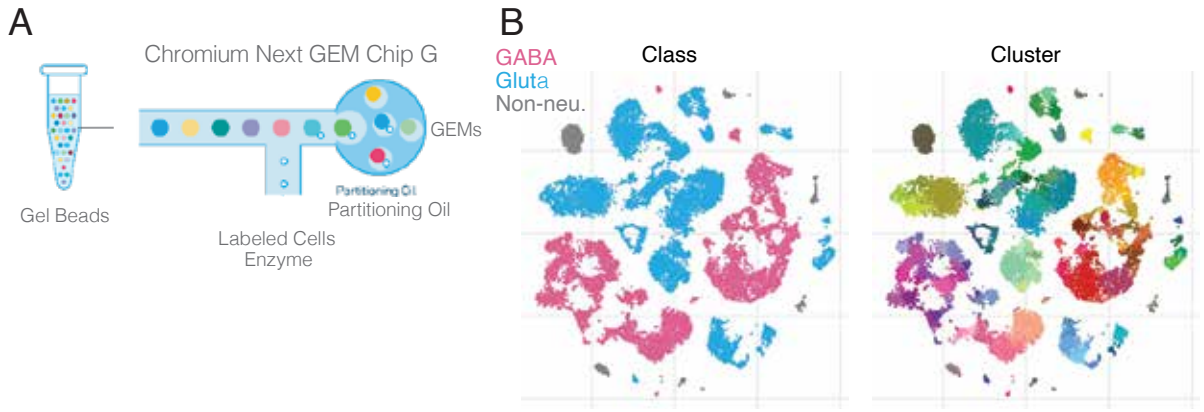

Figure 5: Cell type classification using scRNA-seq A) Diagram showing microfluidic-based high-throughput single cell RNA-sequencing. Single cells are co-encapsulated in oil droplets with beaded barcodes. Cells are lysed inside the droplet and the beaded barcodes attached to the transcriptome of each cell, enabling each gene to be traced back to the cell it came from. **B)** Molecular classification of neurons collected from the mouse neocortex using deep, single-cell RNA sequencing (scRNA-seq). 23, 822 cells are grouped into 133 transcriptomic cell types, of which 61 are GABAergic. Figure adapted from Tasic et al. 2018.

hybridization, which have been limited to querying population averages or a limited number of genes in a supervised manner, scRNA-seq enables monitoring of global expression of thousands of genes within individual cells (Bandler 2019).

The transcriptomes retrieved from these individual cells contain a wealth of information (for example, cell-specific molecular signatures, cell cycle phase, and metabolic state). Many studies have used single cell transcriptomes as an initial framework and anchor to define cell types (Zheng, 22), and one study even identifying up to 60 types of GABAergic neurons in the cortex alone (Tasic et al. 2018).

Retrieving the transcriptome of these cells in different states provides a powerful way to map differentiation dynamics. The densely sampled cells can then be used to construct a manifold of cell states or visualized as a 'landscape', a term inspired by Waddington's illustration depicting cell fate decisions as a ball rolling down a landscape of hills and valleys (Waddington 2014). In the context of cell fate commitment, an uncommitted cell is positioned at the beginning of the landscape and traverses a series of valleys until it goes from a pluripotent state to a committed

one. Similarly, trajectory-building algorithms use transcriptomic information to order cells in a continuum to study changes in average gene expression across the trajectories and for inferring tree-like structures that organize cell clusters into a putative hierarchy.

While state manifolds offer population-level views of differentiation, they do not reveal how clonal lineages of a progenitor population explore these states. It does not account for cell division or death rates, cell state reversibility, difference between clones that can alter the dynamics predicted from snapshot measurements (Weinreb et al. 2018). State manifolds have branch points that may be hypothetical; sister cells from a single division event may progress along both branches of a manifold or continue on a single branch. Unlike trajectory trees, each branch point in a lineage tree represents a division event. Thus, although state manifolds could trace how a single cell might traverse a series of molecular states, it may mislead readers' understanding of fate relationships (Wagner and Klein 2020).

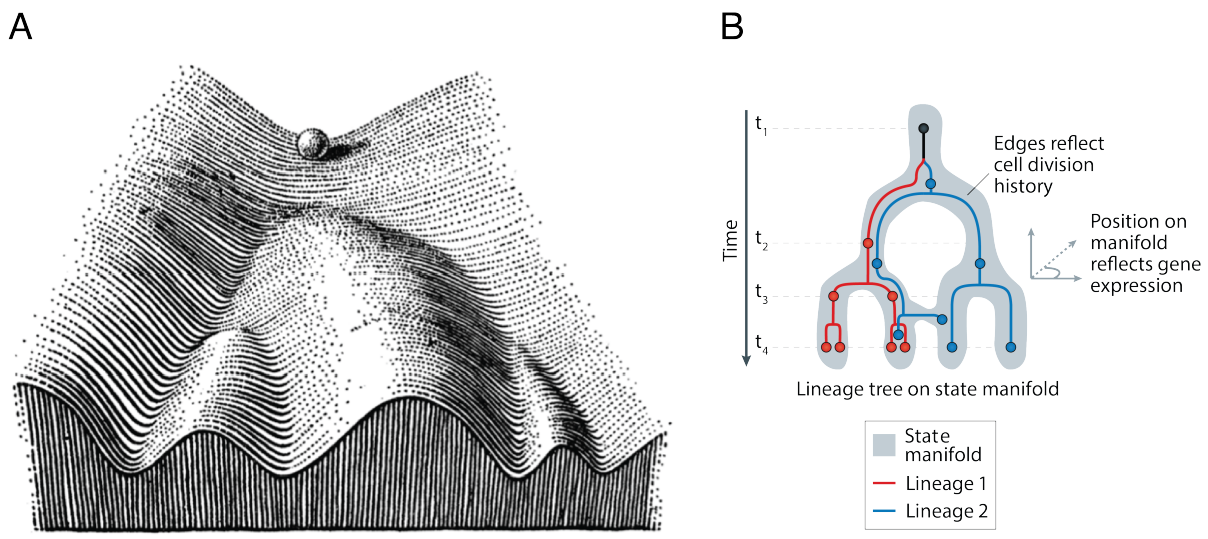

Figure 6: Visualizing cell state transitions using state manifolds. A) Conrad Waddington's epigenetic landscape (EL). The EL model depicts a differentiating cell in the embryo as a ball rolling down a landscape, where the valleys of the landscape represent the developmental choices faced by the cell. **B)** Cell state landscapes vs cell lineage trees. Trajectory relationships are indirectly inferred from gene expression similarities, whereas lineage relationships reflect measured mitotic histories. (**A** Taken from Waddington 2014; **B** modified from Wagner and Klein 2020.

1.3.1 Single-cell lineage tracing

Due to the relatively low resolution determined from low-dimensional measurements and low throughput, many of these early lineage tracing methods are able to identify progeny arising from individual cells but could not meet the stricter definition of lineage tracing. To do so requires that ancestor-progeny relationships are resolved to assemble lineage tree (VanHorn and Morris 2021). By contrast, single-cell genomic technologies support a more objective assessment of cell identity, enabling the capture of many thousands of gene expression measurements while maintaining the cellular resolution required for accurate lineage reconstruction (Kester and Oudenaarden 2018).

single-cell lineage tracing (scLT) is achieved by prospectively introducing a heritable DNA barcodes—-referred to as lineage barcode (LB)s from here on—into cells, then determining clonal relationships and constructing cell lineage retrospectively from sequencing data (Kester and Oudenaarden 2018). The first scLT techniques built on the principals of Cepko's original clonal analysis in the 1980s (Turner and Cepko 1987), where LBs were incorporated into cells via retroviral transgene integration or transposable elements and discriminated from each other via sequencing (VanHorn and Morris 2021). The barcode typically lies within the 3' UTR of a transgene containing a constitutive promoter driving its expression, allowing for parallel capture of barcodes with the transcriptome (VanHorn and Morris 2021).

Another method of scLT relies on CRISPR-Cas9 directed genome editing (Jinek et al. 2012). CRISPR-Cas9 and gRNA are introduced into cells, resulting in scarring of the target sequencing in a given window. McKenna et al. (2016) genetically engineered GESTALT (genome editing of synthetic target arrays for lineage tracing) zebrafish models, to have CRISPR-Cas9 editable casettes within their genomes, and later rendered scLT compatible. Initial experiments performed with GESTALT zebrafish focused on labeled cells from different adult tissues, where after lineage reconstruction, most cells in the adult fish were found to arise from few embry-

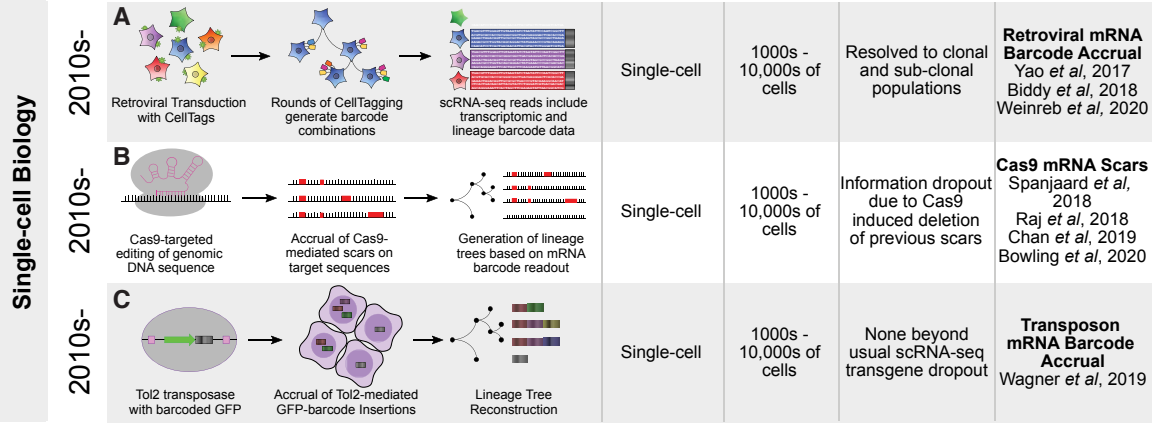

Figure 7: Single-cell lineage tracing techniques. Graphical depictions of **A)** Viral barcoding approaches for clonal and lineage analysis, **B)** Cas9 scar accrual method, and **C)** transposon-mediated barcode accrual. Modified from VanHorn and Morris 2021.

onic progenitors (McKenna et al. 2016). When they looked at neural progenitors, they found that the neural progenitors spread out to many spatial areas, suggesting that they were more migratory than expected. Furthermore, the progeny of these progenitor populations encompassed a high diversity of cell types, indicating that these progenitor populations still maintained high potency (Raj et al. 2018).

However, GESTALT suffers from a number of limitations. Firstly, the cassettes have a large number of Cas9-induced deletions that erase lineage records. Moreover, Cas9 editing saturation (i.e. identical edits introduced into independent cells) and narrow editing periods result in false-positive lineage relationships. These shortcomings make this technique unsuitable for lineage tracing in the mammalian brain, where increased LB diversity is necessary to label substantial cell populations over lengthy periods. To address this gap, a mouse line was created that uses a "homing" guide RNA (hgRNA) to direct CRISPR-Cas9 to its own DNA locus, creating an evolving genetic LB (Kalhor et al. 2018). But it does not have single-cell resolution. A variation of this method is the CARLIN (CRISPR array repair lineage tracing) mouse line (Bowling et al. 2020). It couples inudcible Cas9 expression with a stably integrated Cas9 target allele that contains 10 locations for indel accruals. Using this system, Bowling et al. (2020) observed that

hematopoietic stem cells (HSC) clones preferentially distribute in long bones, indicating that HSCs's expansion potential is influenced by the niche in which they reside. Furthermore, they found that most of replenished blood cells derive from a small progenitor population (Bowling et al. 2020).

Although powerful, the lineage tracing potential of CRISPR-Cas9 strategies is hampered by dropouts due to deletions or false positive relationships resulting from identical edits. By contrast, barcode accumulation achieves a higher diversity of unique heritable sequences through its combinatorial power (VanHorn and Morris 2021). Lineage trees produced by lentivirus-based methods are rudimentary due to the limited number of times cells can be transduced. Wagner et al. bypassed these challenges by opting for a transposon-based approach of labeling cells. TracerSeq. TracerSeq inserts barcode GFP reporters into the genome via the Tol2 transposase, enabling transcription and capture by scRNA-seq to obtain lineage data. Each cell inherits a unique label signature as the barcode accumulates over time, enabling lineage reconstruction.

TracerSeq yielded surprising insights into zebrafish development. Intriguingly, clonally related cells could yield transcriptomically disparate cell types (divergence), and clonally distant clones differentiate into similar cell types (convergence). Sulston's lineage tracing experiments in C. *elegans* also arrived at the same conclusion, that similar neuron types could arise from. Taken together, these examples of lineage tracing demonstrate how single-cell compatible, transposon-based lineage tracing methods can provide new insights into fate commitment.

1.4 Gene editing in the mammalian forebrain

The most direct way to study gene function is to reduce or completely ablate gene expression. Studies traditionally turned to conditional gene knockout mice generated using the Cre / loxP system to investigate the functions of genes in the mammalian forebrain. This technology is powerful as it enables the user to not only selectively knockout a gene in certain populations

but also at select time points.

However, the generation of mutant animals using conventional gene targeting technology is laborious, time consuming, and, more importantly, requires embryonic stem cells, which are not available for all mammalian species (Nishiyama 2019). Rapid advances in genome-editing tools based on engineered nucleases are changing genetic engineering (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013). Cas9 is a type of CRISPR nuclease from bacterial adpative systems that can be used to target virtually any genomic region and is proving to be the most popular approach for genome editing of mammalian cells.

Engineered nucleases such as Cas9 can induce double strand breaks (DSBs) at specific genomic loci, which are then repaired by two types of DNA repair pathways: non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ-mediated DSBs result in the formation of unpredictable patterns of insertions and deleterious mutations (indels) at the break site (Cox, Platt, and Zhang 2015). Thus, the NHEJ mechanism can be used to knock out genes. The CRISPR-Cas9 method of gene editing works as follows: Cas9 is directed to the target site using a short guide RNA (sgRNA). The sgRNA consists of a CRISPR RNA (crRNA), a 17-20 nt sequence complementary to the target DNA, and a transactivating crRNA (tracrRNA) that facilitates Cas9 recruitment to the target site.

Genome editing of cells in the brain requires efficient delivery of the genome editing machinery to cells of interest. Viral vectors, including adeno-associated virus (AAV), lentivirus, adenovirus, and retrovirus, have been used to deliver CRISPR-Cas9-mediated gene editing constructs to cells. Another advantage of viral delivery systems (epsecially AAVs) is that their capsids can be modified to target neuronal subsets of interest (ref Haggerty). AAVs allow persistent long-term expression of transgenes with low immunogenecity but are limited by their transgene capacity (4.7-5kb) (Wu, Yang, and Colosi 2010). However, an advantage of the AAV system is that its capsid can be modified to target neuronal subsets of interest.

in utero electroporation (IUE) is an efficient method to deliver transgenes to brain neuronal

progenitors. It allows the efficient introduction of multiple constructs into the same cells without size limitations. Three plasmids, each encoding a different fluorescent protein, were shown to be co-electroporated in 99% of transcfection neurons by IUE (Nishiyama 2019). Taking advantage of IUE's features enables efficient delivery of CRISPR-Cas9-mediated gene knockouts.

The first studies to apply this technology in the mammalian forebrain was reported in 2014, where two separate groups reported that CRIPSR-Cas9-mediated NHEJ could be used to generate gene knockout in postmitotic neurons (Incontro et al. 2015; Straub et al. 2014). They both targeted *Grin1*, the gene encoding the GluN1 subunit of the N-methyl-D-aspartate-type glutamate receptor (NMDAR), using different gRNAs through IUE. Interestingly, they could not detect NMDAR-mediated current in most of the neurons tested, suggesting disruption of both genomic alleles encoding GluN1. This phenotype was successfully rescued by co-expressing cDNA encoding GluN1, confirming the specificity of CRIPSR-Cas9-mediated gene knockout. The high efficiency of gene knockout in these studies may be attributed to the long-term persistent expression of Cas9 and sgRNA in nondividing neurons.

1.4.1 Single-cell CRISPR perturbations

Previously, biologists had lacked the means to investigate the effects of perturbations on multiple cell types. The role of TFs in specifying inhibitory neuron cell types was typically explored using knockout mice, which is time consuming to generate. Neither classical genetic studies, recent bulk RNA-seq, nor classical genetic studies capture the entire genomic mechanisms through which these TFs induce fate decisions.

The invention of single-cell sequencing provides a rich transcriptomic readout for biological phenotypes that cannot be easily measured by a single marker gene. When single-cell sequencing is coupled with CRISPR screens, it is possible to simultaneously match the gRNAs that induce the perturbation with the corresponding transcriptome in single cells. Methods such as Perturb-seq, CRISPR-seq, CROP-seq and Mosaic-seq can be grouped under the um-

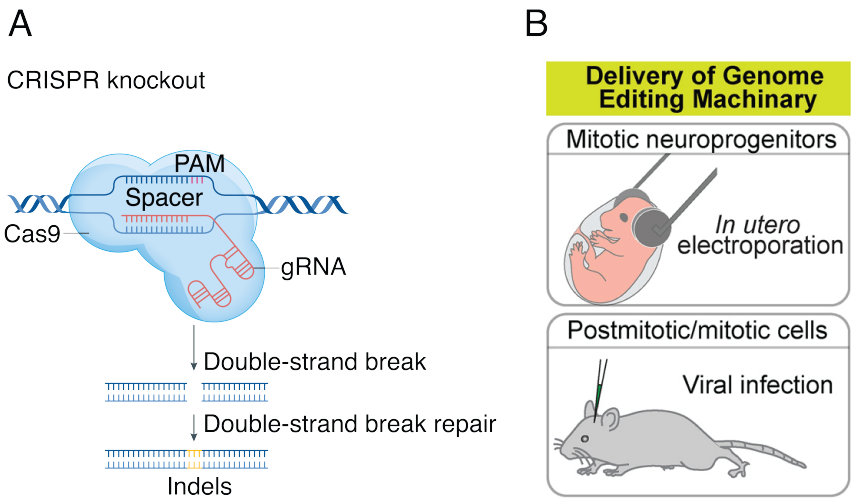

Figure 8: CRISPR-Cas9 genome editing in mammalian cells. A) Gene knockout by CRISPR-Cas9 gene editing. Cas9 nuclease is directed, by a guide RNA (gRNA), to target site to introduce double-strand breaks; subsequent DNA repair results in frameshift mutations or compromised gene function. **B)** Genome editing machinery can be delivered into neuronal progenitors through IUE or postmitotic cells using viral infection. Taken from Bock et al. 2022 and Nishiyama 2019.

brella term scCRIPSR-seq (Bock et al. 2022). Coupling CRISPR screens with a transcriptome readout permits the determination of the type and state of perturbed cells, and allows quantification of induced changes in gene expression, gene regulatory networks, signaling pathway activity, and other properties that can be retrieved from single-cell RNA sequencing.

CRISPR perturbations at single-cell resolution could be the key to revealing more information on the transcriptional codes that regulate the development of inhibitory neurons. One notable advantage of scCRIPSR-seq methods is that the cell-intrinsic effects *in vivo* of the perturbation can be observed due to the minimal disturbance of the tissue environment. Furthermore, a high percentage of gene knockouts in mice are developmentally lethal, and those that survive into adulthood often have morphological or physiological abnormalities that make it difficult to determine the function of a gene in a normal developmental context.

Arlotta et al. were the first to take advantage of this feature, where they developed in vivo Perturb-Seq to functionally evaluate 35 autism spectrum disorders (ASDs) in the mouse brain

(Jin et al. 2020). They identified cell type-specific gene modules in both neuronal and glial cell classes that are affected in different perturbation datasets. Similarly, Fleck et al. (2022) used pooled CRISPR genetic perturbation followed by single-cell transcriptome readout to assess transcription factor regulation of cell fate and state regulation in organoids. They found that certain factors regulate the abundance of cell fates, whereas others affect the states of neuronal cells after differentiation.

Altogether, scCRIPSR-seq methods allow for the unbiased exploration of gene function and systematic delineation of gene regulartory networks (GRNs) (Replogle et al. 2020). It can serve as a scalable tool to develop large gene panels and reveal cell-intrinsinc functions at single-cell resolution in the mammalian brain.

2 Thesis Objectives

A longstanding question in developmental neuroscience is how different types of neuronal cell arise from undifferentiated progenitor cells. To what extent is the identity of neurons determined by intrinsic rather than extrinsic factors is still unclear, particularly for inhibitory neurons. One way of approaching this question is by studying cell lineage, the embryonic genealogy of neurons. In invertebrate systems, the cell lineage of neurons has been well mapped (White 1986), but the clonal relationships between progenitors and their descendants are less clear in vertebrates.

Lineage has been found to play a role in the mammalian neocortex, where asymmetric cell division produces clonally related excitatory neurons in organized columns. Whether lineage dictates how inhibitory neurons develop and integrate into circuits remains an open question. For example, which inhibitory neuron subtypes can be generated from the same progenitor and what logic governs this process, if any? The long distances that inhibitory neurons travel, as well as their wide dispersion, make them difficult to track with conventional lineage-tracing methods such as time-lapse imaging. Thus, the nature of how lineage influences inhibitory neuron specification cannot be delineated unless an appropriate lineage-tracing method is developed. Another genetic mechanism that could drive inhibitory neuron development is the transcription factors expressed in GE during early development. Several transcription factors and their cofactors act spatiotemporally to specify GABAergic neurons (Caputi et al. 2013; Mayer et al. 2018; Leung et al. 2022). The mechanism by which these factors operate and interact with each other is not well understood.

Approaches combining scRNA-seq with DNA barcoding or CRISPR perturbations are appropriate for addressing these questions. scRNA-seq enables a robust and very detailed classification of distinct cell types, while the tagging of progenitors and their progeny with DNA

barcodes allows for the tracing of clonally regulated inhibitory neurons even if they disperse widely throughout the brain. Coupling scRNA-seq with CRISPR perturbation would add another perspective on fate specification by enabling the exploration of transcription factors' role in fate specification.

My thesis can be divided into two parts:

2.1 Establishing TrackerSeq

I developed TrackerSeq, a transposon-based lineage tracing method that uses the piggyBac transposon to successively integrate multiple barcoded GFP reporters into the genome of electroporated mouse cells. It is compatible with the 10X Chromium System, a commercial scRNA-seq platform. The TrackerSeq library has a high diversity of lineage barcodes that can label progenitors *in vivo* and can capture partial clones. Through a collaborative effort, we used TrackerSeq to label dividing progenitors and found that in all ganglionic eminences, newborn GABAergic neurons diverge into different precursor states.

2.2 Fate specification of inhibitory interneurons and projection neurons

In a team effort, I examined the effect that Meis2 perturbation has on the fate specification of inhibitory neurons. To characterize the genomic transcriptional assembly orchestrated by Meis2, we combine IUE CRISPR-based perturbation of Meis2, scRNA-seq, and synthetic oligonucleotide-based lineage tracing in the GE. We found that Meis2 influences the composition of GABAergic neurons by promoting projection neuron fates and suppressing interneuron fates.

3 Experimental Procedures

3.1 Animals

All mouse colonies were maintained according to protocols approved by the Bavarian government at the Max Planck Institute of Neurobiology or the IACUC at the NYU Grossman School of Medicine. Swiss Webster and C57BL/6 wild-type females were used, and embryos were staged in days post-coitus, with E0.5 defined as 12:00 of the day a vaginal plug was detected after overnight mating. Timed pregnant mice were anesthetized with isoflurane (5% induction, 2. 5 % during surgery) and treated with Metamizol (WDT) analgesic. For IUE of the TrackerSeq library, E12.5 embryos were injected unilaterally with 700 nl of DNA plasmid solution made of 0.5 /μl pEF1a-pBase (piggyBac-transposase; a gift from R. Platt) and the TrackerSeq library 0.5 /μl, diluted in endo-free TE buffer and 0.002% Fast Green FCF (Sigma), into the lateral ventricle via a microsyringe pump. Embryos were then electroporated by holding each head between platinum-plated tweezer electrodes (5 mm in diameter, BTX, #45-0489) across the uterine wall, while 5 electric pulses (35 V, 50 ms at 1 Hz) were delivered with a square-wave electroporator (BTX, ECM 830)52. Pregnant dams were kept in single cages and pups were kept with their mothers, in the institutional animal facility under standard 12:12 h light - dark cycles, at a room temperature of 72° F ± 2° F and a humidity of 30–70%.

3.2 Sample Collection

For embryonic lineage tracing, we collected electroporated brains from mouse embryos at E16.5 in Leibowitz medium with 5% FBS. Papain dissociation system was carried out according to the recommended protocol (Wortington, #LK003150), and to isolate positive cells, flow cytometry

was performed using a BD FACSAria III Cell Sorter (BD FACSDiva Software, version 8.0.2) with a 100-um nozzle. For all FACS experiments, non-eGFP-expressing brain tissue was used as a negative control to exclude background fluorescence.

3.3 TrackerSeq library production

TrackerSeq is a *piggyBac* transposon-based34 library, developed to be compatible with the 10x single-cell transcriptomic platform. It records the in vivo lineage history of single cells through the integration of multiple oligonucleotide sequences into the mouse genome. Each of these individual lineage barcodes is a 37-bp long synthetic nucleotide that consists of short random nucleotides bridged by fixed nucleotides. This design results in a library with a theoretical complexity of approximately 4.3 million lineage barcodes (16^8) with each barcode differing from another by at least 5 bp.

3.3.1 Library Cloning

To construct the library, the *piggyBac* donor plasmid (Addgene #40973) was altered to include a number of modifications. The plasmid was digested first with MscI and then the Read2 partial primer was cloned into the 3' UTR of the eGFP to enable retrieval by the 10x platform. The vector is then digested with BstXI so that the sucrose gene can be cloned into the vector. The sucrose gene is used for counter-selection, so empty plasmids that do not incorporate a lineage barcode during the cloning process are removed. After digestion with BstXI to remove the sucrose gene, the plasmid was run on a gel and column purified (Zymo Research, #D4008). The barcode oligo mix was cloned downstream of the Read2 partial primer sequence in the purified donor plasmid through multiple reactions of the Gibson Assembly, as previously described (Gibson et al. 2009). Gibson assembly reactions (NEB, #E2611S) were then pooled and desalted with 0.025 MCE membrane (Millipore, VSWP02500) for 40 min and finally concentrated using

a SpeedVac. 3 μl of the purified assembly is incubated with 50 μl of NEB10--competent Escherichia coli cells (NEB, C3019H) for 30 min at 4 °C, then electroporated at 2.0 kV, 200 , 25 (Bio-Rad, Gene Pulser Xcell Electroporation Systems). Electroporated E. coli was incubated for 90 min shaking at 37 ° C and then seeded in prewarmed sucrose / ampicilin plates.

3.3.2 Vector linearization

To construct the library, the *piggyBac* donor plasmid (Addgene #40973) was altered to include several modifications. The plasmid was digested first with MscI and then the Read2 partial primer was cloned into the 3' UTR of the eGFP to enable retrieval by the 10x platform. The vector is then digested with BstXI so that the sucrose gene can be cloned into the vector (pCAG-SacB).

To begin library cloning, digest 1 ug of the pCAG-SacB vector with BstXI to remove the sucrose gene, and run 4x of these reactions in parallel: assemble 50μl reaction with 5μl Universal Buffer, 1DNA, and molecular grade H_2O. Add 0.2μl of BstXI, then gently mix by pipetting. Place the reaction on a thermoblock for 30 minutes at 50 °C. Run the entire restriction digest on a 1% gel together with a sample of non-digested vector. Excise the 6 kb fragment from the gel, followed by purification of the selected bands with column purification (Zymo Research, #D4008). Finally, measure the concentration on the NanoDrop sepctrophotometer.

3.3.3 Gibson Assembly

The ssDNA oligos to be cloned into the vector were ordered from Integrated DNA Technologies with the following specifications listed in Table 1:

Prepare the ssDNA oligo in 1X NEBuffer 2 to a final concentration of 0.2 (store at -20 °C. Assemble a 10μl reaction mix with 5 μl of 0.2 ssDNA oligo, 50 ng of linearized vector and molecular grade H_2O. Add 10μl of NEBuilder HiFi DNA Assembly Master Mix to the reaction

Table 1: ssDNA oligo specifications

ssDNA oligo	
Product	PAGE Ultramer DNA Oligo
Purification	PAGE Purification
Sequence	5'-GACGTGTGCTCTTCCGATCTCTGANNCT GNNACTNNGACNNTGANNCTGNNACTNN GACNNGACTCTGGCTCACAAATACCAC TG-3'
Length	85

mix, and incubate the assembly reaction for 1 hour at 50 °C in a thermocycler. The resulting stock concentration of the linearized vector should range from 30-60 ng/μl depending on the efficiency of the gel extraction. 6x of these reactions were run in parallel.

3.3.4 Clean-up and sample concentration

The resulting vector from the assembly must be isolated from the salts in the assembly mix so that it does not arc when electroporated into *E.coli*. To desalt the sample, the bottom of a 10 cm Petri dish was filled with milliq-H_2O and the membrane filter disc was placed onto the milliq-H_2O so that it is floating. 120 μl of the sample (6 x 20 μl) were deposited in the center of the membrane. A membrane was placed in the Petri dish to prevent evaporation and the sample was dialyzed for at least 40 minutes. Most of the desalted sample was recovered.

A speed-vac was used to concentrate the sample. The speed-vac was set at 50 °C and operated for 10 minutes initially to ensure that there was sample leftover. The volume was reduced to 20 μl.

3.3.5 Bacteria transformation and plating

3 μl of the purified assembly is incubated with 50 μl of NEB10--competent Escherichia coli cells (NEB, #C3019H) for 30 min at 4 °C, then electroporated at 2.0 kV, 200 , 25 , with a time constant between 3.1 and 3.3 milliseconds (Bio-Rad, Gene Pulser Xcell Electroporation Systems). 950

μl of the pre-warmed recovery medium was immediately added to the cuvette after electroporation, pipetted twice, and then the cells were transferred to a pre-warmed 2 ml Eppendorf. Electroporated E. coli were incubated for 90 min shaking at 37 ° C and 220 rpm and then plated in prewarmed sucrose / ampicillin plates. 190 μl of LB media with ampicillin (LB+Amp) was added to 10 μl of the incubated bacteria before plating on 10 cm pre-warmed sucrose / ampicillin plates. 10 plates of the 200 μl mix (190 μl LB+Amp + 10 μl bacteria) were plated in total. The plates were collected 12 hours later.

3.3.6 Bacteria growth and maxi-prep

Plate colonies were collected by adding 25 ml of LB + amp medium, then scraped with a colony picker. 25 ml of LB +amp medium was added to scrape the remaining colonies. This was repeated for all plates. The colonies of all plates were pooled in an Erlenmeyer flask so that the total volume was 250 ml. LB+Amp medium was added until the total volume reached 4-5L. The initial optical density (OD) of the medium was measured before being incubated at 37 °C at 160 rpm and then monitored hourly until the OD was <0.5. The glycerol stock (1 ml 50% glycerol and 1 ml of bacteria) and the rest were pelleted. The pelleted *E.coli* was maxi-prepped and the resulting library is stored at -20 °C.

3.3.7 Sanger sequencing of colonies

20 colonies were collected with colony pickers and cultured in individual tubes containing 2.5 ml of LB. The QIAprep® Spin Miniprep Kit was used to harvest the donor plasmid from *E. coli*. The Read2 primer was added to each extracted DNA and sent to Eurofins for sequencing..

3.4 TrackerSeq barcode library amplification and cleanup

Following the end of cDNA amplification, step 2 of the 10x protocol, the standard NEB protocol for Q5® Hot Start High-Fidelity 2X Master Mix was used to amplify the library.

The 50 μl l reaction was prepared according to the formula listed in Table 2.

Table 2: NEB protocol for Q5® Hot Start High-Fidelity 2X Master Mix

_PCR reaction mix	
Q5 High-fidelity 2X Master Mix	25 μl
i7 primer 10 μM	2.5 μl
i5 primer 10 μM	2.5 μl
cDNA 1-15ng/μl	10 μl
H_2O	10 μl (adjust depending on amount of cDNA)

The primers and PCR program used to amplify the library are listed in Table 3.

Table 3: NEB protocol for Q5® Hot Start High-Fidelity 2X Master Mix

PCR Program
98 °C 30 sec
⌜ 98 °C 10 sec
18x 63 °C 20 sec
⌞ 72 °C 10 sec
72 °C 2 min
4 °C Hold

Dual-sided SPRI selection was used to recover the 330 bp fragments amplified by the primers. SPRIselect reagent was first resuspended by vortexing then 25 μl (0.5X) of it was added to 50 μl of PCR reaction. The mixture was pipetted 15 times and placed on the high magnet configuration until the solution was clear. 75 μl of the supernatant was transferred to a new PCR tube. The SPRIselect reagent was vortexed again to resuspend it then 60 μl of it was added to the sample. The mixture is pipetted 15 times again and then incubated at room temperature for 5 minutes. The mixture was placed on the high configuration of the magnet again until

Table 4: i7 and i5 primers used to amplify library. Bolded portions are the unique indexes.

i7 primer (5'-3')	**i5 primer (5'-3')**
CAAGCAGAAGACGGCATACGAGAT**GTC AGAAG**GTGACTGGAGTTCAGACGTGT	AATGATACGGCGACCACCGAGATCTACA C**TAGATCGC**ACACTCTTTCCCTACA CGACGCTCTTCCGATCT
CAAGCAGAAGACGGCATACGAGAT**TGG CAAGT**GTGACTGGAGTTCAGACGTGT	AATGATACGGCGACCACCGAGATCTACA C**CTCTCTAT**ACACTCTTTCCCTACAC GACGCTCTTCCGATCT
CAAGCAGAAGACGGCATACGAGAT**GGT TCCTT**GTGACTGGAGTTCAGACGTGT	AATGATACGGCGACCACCGAGATCTACA C**TATCCTCT**ACACTCTTTCCCTACACG ACGCTCTTCCGATCT
CAAGCAGAAGACGGCATACGAGAT**CAC TGGTT**GTGACTGGAGTTCAGACGTGT	AATGATACGGCGACCACCGAGATCTACA C**AGAGTAGA**ACACTCTTTCCCTACA CGACGCTCTTCCGATCT
CAAGCAGAAGACGGCATACGAGAT**CAC CTGTA**GTGACTGGAGTTCAGACGTGT	AATGATACGGCGACCACCGAGATCTA**C ACCAGGACTA**ACACTCTTTCCCTACA CGACGCTCTTCCGATCT
CAAGCAGAAGACGGCATACGAGAT**GAC CACTA**GTGACTGGAGTTCAGACGTGT	AATGATACGGCGACCACCGAGATCTACA C**CGCTACTA**ACACTCTTTCCCTACACG ACGCTCTTCCGATCT

the solution cleared. 135 μl of the supernatant was carefully removed and discarded to avoid distrubing the bead pellet. With the tube still on the magnet, 200 μl of freshly prepared 80% EtOH was added to the pellet for 30 seconds and the EtOH was removed. This step is repeated. The tube was briefly centrifuged and then returned to the magnet in a low configuration. Any residual EtOH was removed with a pipette. The tube was removed from the magnet and 22 μl of EB buffer was added. The mixture was pipetted 15 times and then incubated for 2 minutes at room temperature. The tube is placed in low magnet configuration until the solution clears. 22 μl is transferred to a new tube. This was the barcode library. The success of the library prep was confirmed with the Agilent bioanalyzer.

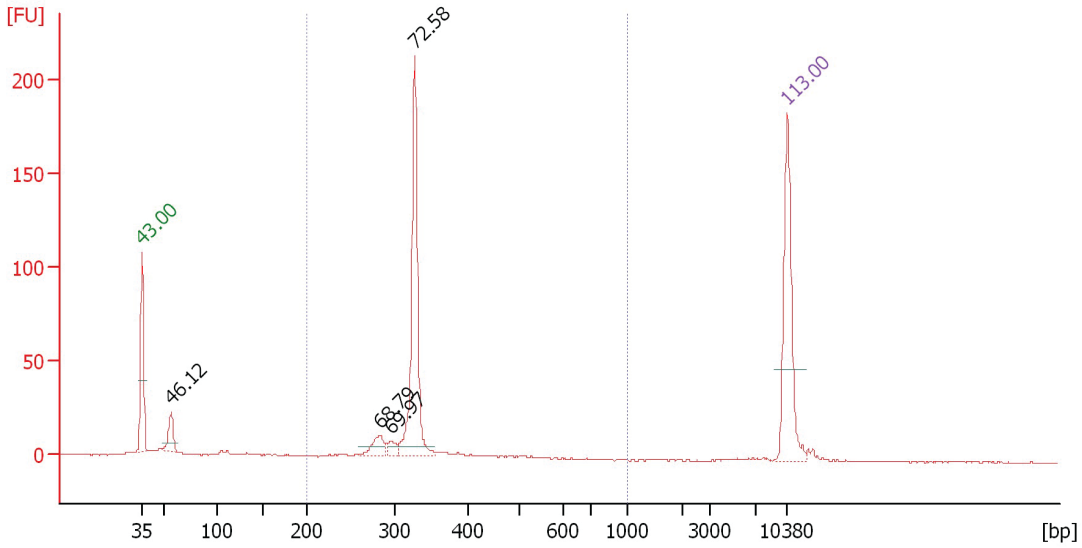

Figure 9: Trace from BioAnalyzer. A peak around 330 bp indicates successful amplification of the library.

3.5 Preparation of RNA-seq and TrackerSeq libraries

For experiments utilizing the 10x Genomics platform, the following reagents were used: Chromium Single Cell 3 Library & Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3' Chip Kit v2 (PN-120236) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3 Reagents Kits V2 User Guide; Chromium Single Cell 3' Library & Gel Bead Kit v3 (PN-1000075), Chromium Single Cell 3 Chip Kit V3 (PN-1000073) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3 Reagents Kits V3 User Guide; Chromium Single Cell 3 Library Gel Bead Kit v3.1 (PN-1000268), Chromium Single Cell 3' Chip Kit V3.1 (PN-1000127) and Dual Index Kit TT Set A (PN-1000215) were used according to the manufacturer's instructions in the Chromium Single Cell 3 Reagents Kits V3.1 User Guide (Dual Index).

The lineage barcode library retrieved from RNA was amplified with a standard NEB protocol for Q5 Hot Start High-Fidelity 2X Master Mix (#M094S) in a 50-μl reaction, using 10 μl of cDNA as template. Specifically, each PCR contained: 25 μl Q5 High-fidelity 2X Master Mix, 2.5 μl 10

μM P7_indexed reverse primer, 2.5 μl 10 μM i5_indexed forward primer, 10 μl molecular grade H_2O, 10 μl cDNA.

3.6 Sequencing and read mapping

Transcriptome and barcode libraries were sequenced either on an Illumina NextSeq 500 at the Next Generation Sequencing Facility of the Max Planck Institute of Biochemistry, at the Genomics Core Facility at the Helmholtz Center in Munich, or on a NovaSeq at the Broad Institute. For a detailed report on each dataset, see Supplementary Data 1. Sequencing reads in FASTQ files were aligned to a reference transcriptome (mm10-2.1.0) and collapsed into UMI counts using the 10x Genomics Cell Ranger software (version 3.0.2 or 5.0.1).

3.7 Processing of TrackerSeq barcode reads

3.7.1 Processing of TracerSeq reads for diversity estimation

The diversity of TrackerSeq barcode libraries were assessed by RNA-seq to check whether any barcode is overrepresented. Unique reads of the lineage barcodes were extracted from the R2 FASTQ files using Bartender (Zhao et al. 2017). Extracted barcodes within 3 bp of each other are collapsed into one cluster, where each cluster is considered to be a unique barcode. 1000 barcodes were sampled randomly to assess the hamming distance of the lineage barcode library. Hamming distance was assessed using the DNABarcodes package (Buschmann and Bystrykh 2013).

3.7.2 Pre-processing of TrackerSeq barcodes

Using BBduk (Brian 2014), reads in the R2 FASTQ files were pre-processed so that the sequences to the left and right of the lineage barcodes (BC) were trimmed. Lineage barcodes shorter than

37 bp were discarded. Whitelists for cell barcodes were generated using two different methods: Cell barcodes (Cell) were extracted from the corresponding Seurat object of the dataset to generate a cell barcode whitelist, or the whitelist was also generated from the R1 sequencing file using UMI-tools. The extracted cell barcodes and UMIs were added to the read names of the lineage barcode FASTQ files. The resulting FASTQ files were processed to output a sparse matrix in csv format, where rows were cells identified by individual cell barcodes and columns were lineage barcodes. Using code modified from Wagner et al. 2018, only Cell–UMI–BC triples supported by at least 10 reads and Cell–BC pairs with at least 6 UMI were considered for further analyses. CloneIDs were assigned to cell barcodes by clustering the matrix using Jaccard similarity and average linkage, as demonstrated by Wagner and colleagues. The resulting dendrogram was cut at a height of 0.999 to obtain the clonal groupings.

3.8 Cell filtering, data normalization batch correction and clustering of datasets

3.8.1 Processing the MUC28072 dataset

The Seurat pipeline (version 3.1.4) was used for cluster identification in scRNA-seq datasets. Embryonic transcriptome datasets (MUC28072) were read into R (version 3.6.0) as a count matrix. Each dataset was filtered with cut-offs for: maximum or minimum gene expression, maximum nCount_RNA and the percentage of total reads that aligned to the mitochondrial genome. In addition, embryonic datasets were filtered with DoubletFinder version 2.0.3 (McGinnis, Murrow, and Gartner 2019).

We used regularized negative binomial regression59 to normalize UMI count data for all embryonic datasets. Cells with UMI counts for Neurod2 > 2 and Neurod6 > 2, which are markers of excitatory neurons, were removed. The TrackerSeq dataset was clustered using Seurat stan-

dard procedures and clusters expressing marker genes for excitatory neurons were removed. We created an 'integrated' data assay including all embryonic datasets for downstream analysis as described by Stuart and colleagues60. Clusters of cells were identified by a shared nearest neighbour modularity optimization-based clustering algorithm. Uniform manifold approximation and projection (UMAP) dimensional reduction (https://github.com/lmcinnes/umap) was applied to the integrated data assay for visualization.

3.8.2 Processing embryonic tCROP-seq datasets.

Embryonic E16 tCROP-seq datasets, including those that contained TrackerSeq barcodes, were processed together for cell filtering, data normalization and cluster annotation following the standard Seurat workflow (4.0.6, Hafemeister and Satija 2019). Data was read into R as a count matrix. Each dataset was preprocessed according to a set of criterions: minimum and maximum genes expressed, maximum nCount_RNA, and mitochondrial mapping percentage. CRISPR-perturbed cells were identified using a CSV file output by CellRanger that contained the cell barcodes and the sgRNA detected in that cell. We removed excitatory clusters by removing those that have UMI counts for Neurod2>2 and Neurod6>2, which are markers of excitatory neurons. To create an 'integrated' data assay, we combined the embryonic tCROP-seq dataset with *wt* GE datasets that were collected at E13.5 and E15.5 as described by Stuart et al. 2019. Briefly, after each dataset is normalized using SCTransoform(), anchors are identified using FindIntegrationAnchors() then the anchors are used to integrate the embryonic tCROP-seq and *wt* datasets with IntegrateData().

To group cells into clusters, we first constructed a shared-nearest neighbour using the FindNeighbors() algorithm, then input the graph into an SLM algorithm that is implemented through the FindClusters() function in Seurat (dimensions = 30, res = 0.8). We obtained cluster-specific marker genes by performing differential expression analysis (DEA) using FindAllMarkers(), comparing cells of each cluster to cells from all other clusters. Genes were considered differen-

tially expressed if they met the fold change, minimum expression and adjusted P value cut-offs as dictated by the Wilcoxon rank sum test implemented via Seurat. Clusters were assigned to cell types based on marker gene expression from literature, primarily http://mousebrain.org/development/ (La Manno et al. 2021).

3.8.3 Processing postnatal tCROP-seq datasets.

Each post-natal P7 dataset was preprocessed according to a set of criterion: minimum and maximum genes expressed, maximum nCount_RNA, and mitochondrial mapping percentage. We normalized and regressed out technical effects introduced by nFeature_RNA, nCount_RNA, and mitochondrial mapping percentage using SCTransform().

We used Harmony (v1.0, Korsunsky et al. 2019) within the Seurat workflow using default settings (theta = 2, lambda = 1, sigma = 0.1) to integrate different CRISPR datasets. We used the first 30 Harmony embeddings for UMAP (https://github.com/lmcinnes/umap) visualizations and clustering analysis.

To group cells into clusters, we first constructed a shared-nearest neighbour graph from Harmony embeddings using the FindNeighbors() algorithm, then input the graph into the FindClusters() function in Seurat (dimensions = 30, res = 0.8). We obtained cluster-specific marker genes by performing differential expression analysis using FindAllMarkers().

3.9 scRNA-seq analysis of tCROP-seq datasets

3.9.1 Comparing cell type composition between perturbations

We compared the perturbation effect on cell type composition using the method described by (Jin et al. 2020). More specifically, we used a Poisson regression model to test the relationship between perturbations and cell number in each cluster, correcting for batch and total number

of cells. The formula is as follows:

Num ~ offset(logTot) + Batch + Pert

3.9.2 Differential expression analysis

We used Libra package to perform DEA (Squair et al. 2021). We ran the run_DE functions on Seurat objects directly with the following parameters (de_family = pseudobulk, de_family = pseudobulk, de_method = edgeR, de_type = LRT). We obtained DEGs of PNs or INs by using run_DE function on cells grouped into classes (mitotic, projection neurons, and interneurons). To obtain differentially expressed gene (DEG)s of individual subclusters, we used the run_DE function on individual clusters. We filtered for statistically significant genes (FDR-adjusted p-value threshold = 0.05). Genes were considered differentially expressed if avg_logFC < -1.0 or avg_logFC > 1.0.

3.9.3 Hotspot analysis of gene coexpression

Hotspot(v0.91) is a tool for identifying co-expressing gene modules in a single-cell dataset (DeTomaso and Yosef 2021). It computes gene modules by evaluating the pairwise correlation of genes with high local autocorrelation, then clusters the results into a gene-gene affinity matrix. *Gad2*-expressing inhibitory population in the P7 dataset was first subset out from the rest to identify inhibitory specific modules in the embryonic dataset. We ran the depth-adjusted negative binomial model on the entire count matrix and Harmony (v1.0) corrected principal components (see batch correction of tCROP-seq datasets). We computed a k-nearest-neighbors (KNN) graph with 30 neighbours, 9154 non-varying genes were subsequently detected and removed. Autocorrelations between each gene were calculated, and the top 500 significant (FDR < 0.05) genes were used to evaluate pairwise gene associations (local correlations). After pairwise local correlations are calculated, we grouped genes into modules. Modules were created

through agglomerative clustering, where the minimum number of genes per module was set to 30. 8 modules were identified, and 103 genes were not assigned to a module. Summary per-cell module scores are calculated using the calculate_module_scores() function as described by (DeTomaso and Yosef 2021).

3.9.4 Testing Hotpot module gene sets

As described by Jin et al. (2020), linear regression was used to test the relationship between perturbation and Hotspot module gene scores. The batch and number of genes were corrected for using the lm function from the stats package (version 3.6.2), with the following formula:

Gene Score ~ perturbation + batch + nGene

3.9.5 GO Term analysis of differentially expressed genes and module genes

GO Term analysis was done using the package enrichR (Kuleshov et al. 2016). The DEGs and module genes of each module were queried against the following databases: GO_Molecular_Function_2018, GO_Cellular_Component_2018, and GO_Biological_Process_2018. Only GO Terms that were significant (p-value adjusted < 0.05) were kept.

3.10 Preparation of RNA-seq and TrackerSeq libraries

For experiments utilizing the 10x Genomics platform, the following reagents were used: Chromium Single Cell 3 Library & Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3' Chip Kit v2 (PN-120236) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3 Reagents Kits V2 User Guide; Chromium Single Cell 3' Library & Gel Bead Kit v3 (PN-1000075), Chromium Single Cell 3 Chip Kit V3

(PN-1000073) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3 Reagents Kits V3 User Guide; Chromium Single Cell 3 Library Gel Bead Kit v3.1 (PN-1000268), Chromium Single Cell 3' Chip Kit V3.1 (PN-1000127) and Dual Index Kit TT Set A (PN-1000215) were used according to the manufacturer's instructions in the Chromium Single Cell 3 Reagents Kits V3.1 User Guide (Dual Index).

The lineage barcode library retrieved from RNA was amplified with a standard NEB protocol for Q5 Hot Start High-Fidelity 2X Master Mix (#M094S) in a 50-μl reaction, using 10 μl of cDNA as template. Specifically, each PCR contained: 25 μl Q5 High-fidelity 2X Master Mix, 2.5 μl 10 μM P7_indexed reverse primer, 2.5 μl 10 μM i5_indexed forward primer, 10 μl molecular grade H_2O, 10 μl cDNA.

3.11 Sequencing and read mapping

Transcriptome and barcode libraries were sequenced on an Illumina NextSeq 500 at the Next Generation Sequencing Facility of the Max Planck Institute of Biochemistry, at the Genomics Core Facility at the Helmholtz Center in Munich, or on a NovaSeq at the Broad Institute. For a detailed report on each dataset, see Supplementary Data 1. Sequencing reads in FASTQ files were aligned to a reference transcriptome (mm10-2.1.0) and collapsed into UMI counts using the 10x Genomics Cell Ranger software (version 3.0.2 or 5.0.1).

3.12 Processing of TrackerSeq barcode reads

3.12.1 Processing of TracerSeq reads for diversity estimation

The diversity of TrackerSeq barcode libraries was assessed by RNA-seq to check whether any barcode is overrepresented. Unique reads of the lineage barcodes were extracted from the R2 FASTQ files using Bartender (Zhao et al. 2017). Extracted barcodes within 3 bp of each other

are collapsed into a cluster, where each cluster is considered to be a unique barcode. 1000 barcodes were randomly sampled to assess the Hamming distance of the lineage barcode library. Hamming distance was assessed using the DNABarcodes package (Buschmann and Bystrykh 2013).

3.12.2 Pre-processing of TrackerSeq barcodes

Using BBduk (Brian 2014), reads in the R2 FASTQ files were preprocessed so that the sequences to the left and right of the LBs trimmed. LBs barcodes shorter than 37 bp were discarded. The whitelists for cell barcodes were generated using two different methods. Cell barcodes (cellbc) were extracted from the corresponding Seurat object of the dataset to generate a cell barcode whitelist, or the whitelist was also generated from the R1 sequencing file using UMI tools. Cell barcodes and extracted UMIs were added to the read names of the lineage barcode FASTQ files. The resulting FASTQ files were processed to output a sparse matrix in csv format, where rows were cells identified by individual cell barcodes and columns were lineage barcodes. Using code modified from Wagner et al. (2018), only Cell-UMI-BC triplets supported by at least 10 reads and Cell–BC pairs with at least 6 UMI were considered for further analyses. CloneIDs were assigned to cell barcodes by clustering the matrix using Jaccard similarity and average linkage, as demonstrated by Wagner et al. (2018). The resulting dendrogram was cut at a height of 0.999 to obtain the clonal groups.

3.13 Cell filtering, data normalization batch correction and clustering of datasets

3.13.1 Processing the MUC28072 dataset

The Seurat pipeline (version 3.1.4) was used for cluster identification in scRNA-seq datasets. Embryonic transcriptome datasets (MUC28072) were read into R (version 3.6.0) as a count matrix. Each dataset was filtered with cut-off points for: maximum or minimum gene expression, maximum nCount_RNA and the percentage of total reads aligned to the mitochondrial genome. Furthermore, embryonic datasets were filtered with DoubletFinder version 2.0.3 (McGinnis, Murrow, and Gartner 2019).

We used regularized negative binomial regression59 to normalize the UMI count data for all embryonic datasets. Cells with UMI counts for Neurod2 > 2 and Neurod6 > 2, which are markers of excitatory neurons, were removed. The TrackerSeq dataset was clustered using Seurat standard procedures and clusters expressing marker genes for excitatory neurons were removed. We created an 'integrated' data assay that includes all embryonic datasets for downstream analysis, as described by Stuart et al. (2019). Clusters of cells were identified by a shared nearest-neighbor modularity optimization-based clustering algorithm. Uniform manifold approximation and projection (UMAP) dimensional reduction (https://github.com/lmcinnes/umap) was applied to the integrated data assay for visualization.

3.13.2 Processing embryonic tCROP-seq datasets.

Embryonic E16 tCROP-seq datasets, including those containing TrackerSeq barcodes, were processed together for cell filtering, data normalization, and cluster annotation following the standard Seurat workflow (4.0.6, Hafemeister and Satija 2019). Data were read into R as a count

matrix. Each data set was pre-processed according to a set of criteria: minimum and maximum genes expressed, maximum nCount_RNA and percentage of mitochondrial mapping. CRISPR-perturbed cells were identified using a CellRanger CSV file output that contained the cell barcodes and the sgRNA detected in that cell. We excluded excitatory clusters by removing those that have UMI counts for Neurod2>2 and Neurod6>2, which are markers of excitatory neurons. To create an 'integrated' data assay, we combined the embryonic tCROP-seq dataset with *wt* GE datasets that were collected at E13.5 and E15.5 as described by Stuart et al. (2019). Briefly, after each dataset is normalized using SCTransform(), anchors are identified using FindIntegrationAnchors(), then the anchors are used to integrate the embryonic tCROP-seq and *wt* data sets with IntegrateData().

To group cells into clusters, we first construct a shared nearest neighbor using the FindNeighbors() algorithm, then input the graph into an SLM algorithm that is implemented through the FindClusters() function in Seurat (dimensions = 30, res = 0.8). We obtained cluster-specific marker genes by performing DEA using FindAllMarkers(), comparing cells of each cluster to cells from all other clusters. Genes were considered differentially expressed if they met the fold change, minimum expression, and adjusted P-value cut-offs as dictated by the Wilcoxon rank sum test implemented via Seurat. Clusters were assigned to cell types according to the expression of marker genes in the literature, mainly at http://mousebrain.org/development/ (La Manno et al. 2021).

3.13.3 Processing postnatal tCROP-seq datasets.

Each post-natal P7 dataset was pre-processed according to a set of criterion: minimum and maximum genes expressed, maximum nCount_RNA and the mitochondrial mapping percentage. We normalized and regressed the technical effects introduced by nFeature_RNA, nCount_RNA, and mitochondrial mapping percentage using SCTransform().

We used Harmony (v1.0, Korsunsky et al. 2019) within the Seurat workflow using default

settings (theta = 2, lambda = 1, sigma = 0.1) to integrate different CRISPR datasets. We used the first 30 Harmony embeddings for UMAP (https://github.com/lmcinnes/umap) visualizations and clustering analysis.

To group cells into clusters, we first constructed a shared-nearest neighbor graph from Harmony embeddings using the FindNeighbors() algorithm, then input the graph into the FindClusters() function in Seurat (dimensions = 30, res = 0.8). We obtained cluster-specific marker genes by performing a differential expression analysis using FindAllMarkers().

3.14 scRNA-seq analysis of tCROP-seq datasets

3.14.1 Comparing cell type composition between perturbations

We compared the effect of perturbation on cell type composition using the method described by Jin et al. (2020). More specifically, we used a Poisson regression model to test the relationship between perturbations and cell number in each cluster, correcting for batch and total number of cells. The formula is as follows:

Num ~ offset(logTot) + Batch + Pert

3.14.2 Differential expression analysis

We used the Libra package to perform DEA (Squair et al. 2021). We ran the run_DE functions on Seurat objects directly with the following parameters (de_family = pseudobulk, de_family = pseudobulk, de_method = edgeR, de_type = LRT). We obtained DEGs of projection neuron (PN)s or interneuron (IN)s by using run_DE function on cells grouped into classes (mitotic, projection neurons, and interneurons). To obtain DEGs of individual subclusters, we used the run_DE function on individual clusters. We filtered for statistically significant genes (FDR-adjusted p-value threshold = 0.05). Genes were considered differentially expressed if avg_logFC < -1.0 or

avg_logFC > 1.0.

3.14.3 Hotspot analysis of gene coexpression

Hotspot(v0.91) is a tool to identify co-expressing gene modules in a single-cell dataset (DeTomaso and Yosef 2021). It computes gene modules by evaluating the pairwise correlation of genes with high local autocorrelation and then clusters the results into a gene-gene affinity matrix. *Gad2*-expressing inhibitory population in the P7 dataset was first subset out from the rest to identify inhibitory specific modules in the embryonic dataset. We ran the depth-adjusted negative binomial model on the entire count matrix and Harmony (v1.0) corrected principal components (see batch correction of tCROP-seq datasets). We computed a k-nearest-neighbors (KNN) graph with 30 neighbors; subsequently, 9154 non-varying genes were detected and removed. Autocorrelations between each gene were calculated, and the top 500 significant (FDR < 0.05) genes were used to evaluate pairwise gene associations (local correlations). After calculating pairwise local correlations, we grouped genes into modules. Modules were created through agglomerative clustering, where the minimum number of genes per module was set to 30. 8 modules were identified, and 103 genes were not assigned to a module. Summary module scores per cell are calculated using the calculate_module_scores() function as described by DeTomaso and Yosef (2021).

3.14.4 Testing Hotpot module gene sets

As described by Jin et al. (2020), linear regression was used to test the relationship between the perturbation and the gene scores of the hotspot module. The batch and number of genes were corrected for using the lm function from the stats package (version 3.6.2), with the following formula:

Gene Score ~ perturbation + batch + nGene

3.14.5 GO Term analysis of differentially expressed genes and module genes

GO term analysis was performed using the enrichR package (Kuleshov et al. 2016). The DEGs and module genes of each module were queried against the following databases: GO_Molecular_Function_2018, GO_Cellular_Component_2018, and GO_Biological_Process_2018. Only GO Terms that were significant (p-value adjusted < 0.05) were kept.

3.14.6 Link to code

The code for the analyses described in this thesis can be found in this GitHub repository: https://github.com/mayho3/PhD-thesis

4 Results

4.1 Establishing TrackerSeq

Acknowledgement

The results presented in Part 1 "Establishing TrackerSeq" were obtained through collaborative efforts. I co-developed the TrackerSeq protocol with Christian Mayer. Ilaria Vitali performed the staining, IUE and scRNA-seq exerpiments. Elena Dvoretskova designed, assisted with the experiments, and analyzed the data. I provided custom scripts to analyze TrackerSeq data. Trajectory analysis was executed by Christian Mayer.

4.1.1 Gibson cloning is appropriate for complex library construction

To study the role that lineage plays in fate specification, we designed a single-cell compatible, transposon-based barcoding approach, named TrakcerSeq.

We designed a 37-bp long synthetic oligos, referred to henceforth as LB, that had 2-bp random regions bridged with 3-bp long fixed regions. This design resulted in a library with a theoretical complexity of approximately 4.3 million LB lineage barcodes. The piggyBac donor plasmid (Addgene #40973) was modified using Gibson Assembly to include a Read2 partial primer sequence to make it compatible with next-generation sequencing. The sucrose gene was cloned downstream of the Read2 sequencing primer for counter-selection process. Between each Gibson Assembly cloning reaction, we confirmed the integrity of the vector using Sanger sequencing. Finally, we cloned the LB into the piggyBac plasmid backbone using Gibson Assembly. We then plated the reaction on sucrose plates so that e.coli containing plasmids that failed to incorporate the LB (i.e. still contain the sucrose gene rather than LB) cannot grow on

sucrose plates and are not harvested. 20 colonies on the plates were sent for Sanger sequencing to confirm the diversity of the LBs, as well as the successful incorporation of the lineage barcode in the plasmid backbone (Figure 10A). In addition, all the LBs were unique.

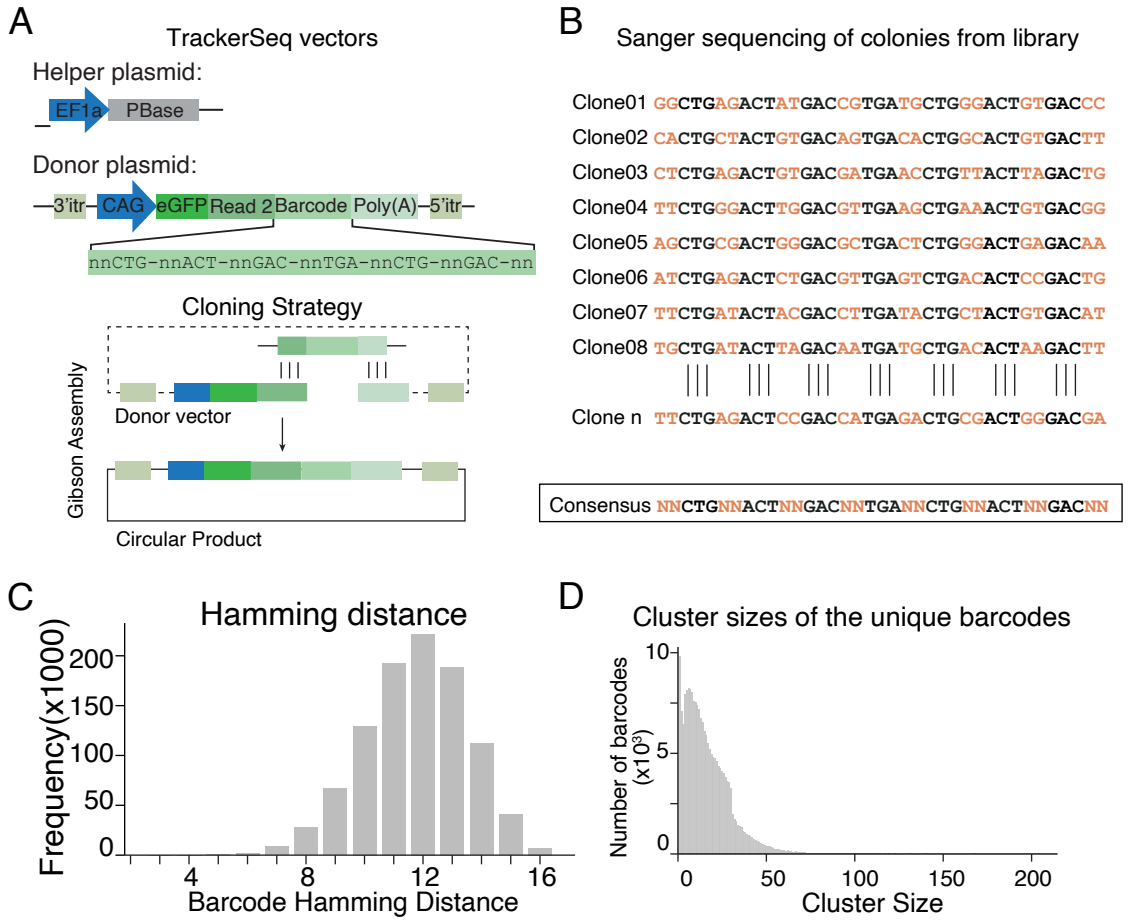

Figure 10: Cloning the TrackerSeq library. A) Vector maps and cloning strategy of TrackerSeq. PBase: piggyBac transposase **B)** Sanger sequencing results of 10 individual E.coli colonies ('clones') depicting the consensus sequence of the TrackerSeq lineage barcodes. **C)** Pairwise hamming distance of 1000 barcodes randomly sampled from the TrackerSeq library. **D)** 3.6 x 10^6 raw sequencing reads were collapsed into 2 x 10^5 clusters, where each cluster is defined as a unique lineage barcode.

4.1.2 The TrackerSeq library is highly diverse

To estimate the diversity of the TrackerSeq library, we first prepared the library for next-generation sequencing (NGS). We took an aliquot of the library, amplified it with primers, performed dual-sided SPRI selection to recover the amplified LBs followed by NGS. We used 3.6 x 10^6 reads to sequence the LB library. These raw sequencing reads were collapsed into approximately 2 x 10^5 clusters, where each cluster is defined as a unique lineage barcode (Figure 10D). We extracted sequencing reads with Bartender and randomly sampled 1000 barcodes to examine the Hamming distance. In this case, the Hamming distance is as the number of bases by which the LBs differ. If the LB library consisted of barcodes that had low Hamming distances, that made the LBs more vulnerable to errors introduced by PCR or sequencing. We evaluated the Hamming distance of the sample barcodes with the stringdistmatrix function from the stringdist package (Loo 2014) and plotted the distribution. The minimum, average, and maximum hamming distance between barcodes in the library were 5, 11, and 16 respectively (Figure 10C).

4.1.3 TrackerSeq can label progenitors *in vivo*

We targeted the TrackerSeq library to ganglionic eminence progenitors at E12.5 and collected electroporated brains at E16.5 to test whether the library can successfully label neuronal cells in vivo (Figure 11A). We examined Hoechst stained coronal section of brains harvested from E16.5 embryos and confirmed that the electroporated cells are expressing GFP (Figure 11B). Once we confirmed that TrackerSeq can label progenitors, we repeated the experiment and FACS-enriched E16.5 electoporated cells and performed scRNA-seq. We embedded the TrackerSeq^{E12} cells with wild-type scNRA-seq datasets that were collected at E13.5 and E15.5 from the medial, caudal and lateral ganglionic eminences to gain a higher resolution of the embryonic cell states. After processing the TrackerSeq barcode reads, we observe that 4, 282 barcodes were

distributed over 2, 370 cells in the total dataset (Figure 11C). Among these labelled cells, 56.0% of them were marked by 2 or more barcode integrations, and 8.4% of them were marked by 5 or more integrations in the total dataset. Within the GABAergic neurons, these numbers were 85.7% and 9.5%, respectively. Hierarchical clustering of TrackerSeq DNA tags organized cells into 256 distinct multicellular clones of GABAergic neurons (Figure 12A).

4.1.4 Emerging embryonic precursor states mapped to postnatal states

We performed pseudotime trajectory analysis on the TrackerSeq dataset using Monocle3, a diffusion pseudotime algorithm that identifies developmental branch by learning the sequencing of gene expression changes. After cell-cycle exit, five different trajectories (or precursor states) of postmitotic GABAergic neurons emerged from a common pool of mitotic progenitors (Figure 13A).

After clustering, we annotated the clusters based on their shared, top marker genes that are expressed ('i_*Six3/Gucy1a3*', 'i_*Ebf1/Isl1*', 'i_*Phlda1/Isl1*, 'i_*Nr2f2*' and 'i_*Nxph1*') (Figure 13B). To examine whether postnatal cell type identity already emerge at this stage, we mapped cells from each embryonic precursor state to GABAergic clusters in postnatal datasets using a correlation-based distance metric (Figure 13C-E). The majority of cells from precursor states mapped to specific postnatal clusters: 83% of cells from the 'i_*Six3/Gucy1a3*' mapped to the postnatal cluster '7a D2 SPNs', and 89% of cells from the 'i_*Ebf1/Isl1*' mapped to the '7b D1 SPNs' cluster (Figure 13E). In concordance with the data, OB interneuron precursors as well as D1 and D2 striatal precursors, sustained the expression of multiple marker genes throughout development.

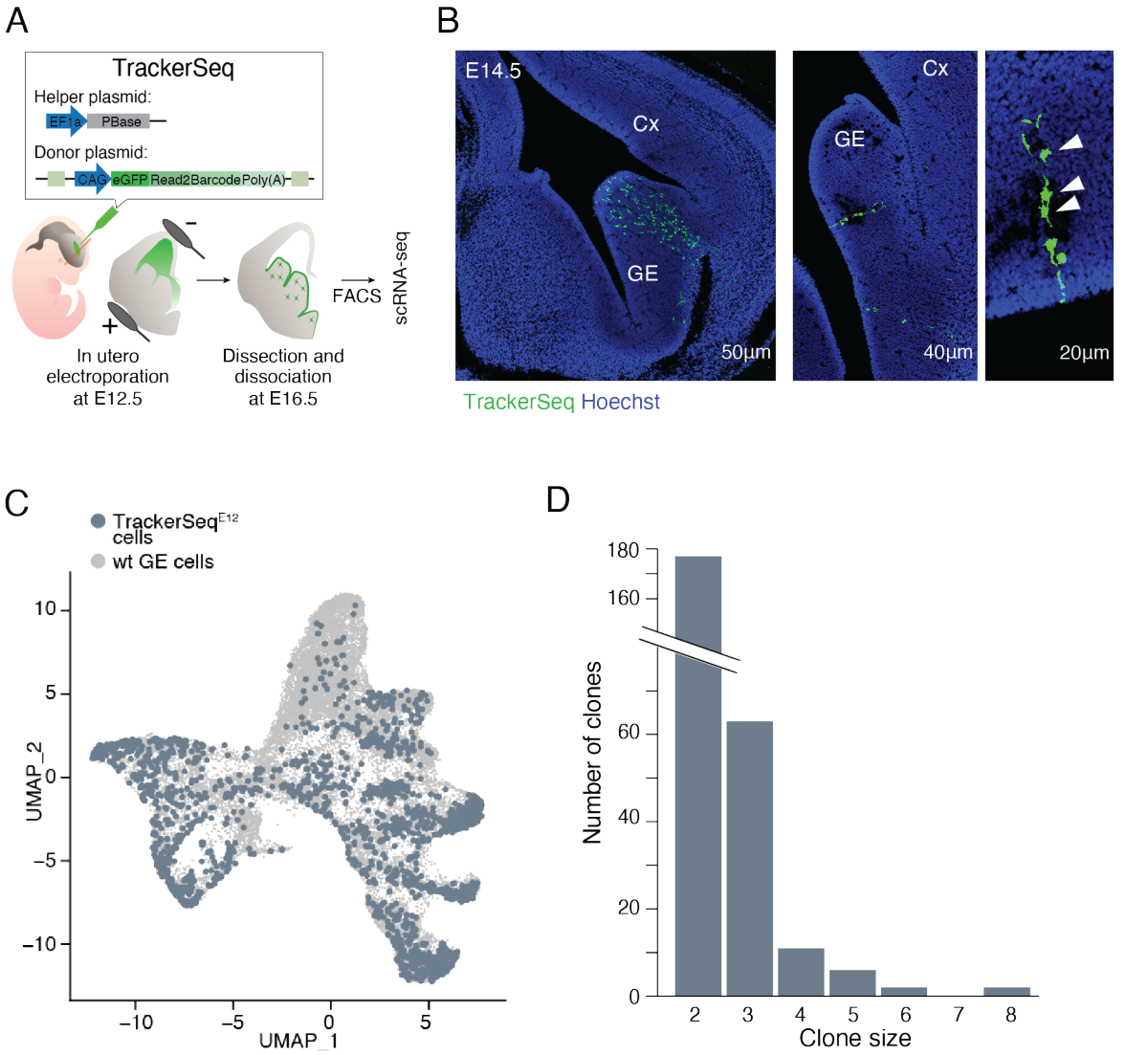

Figure 11: TrackerSeq successfully labels cells in vivo. A) Schematic of the TrackerSeq experimental workflow. PBase, piggyBac transposase. **B)** Images of coronal brain sections electroporated with TrackerSeqE12 and collected at E14.5. Cx, cortex; GE, ganglionic eminence. Magnification on the bottom right panel shows a radial cluster of newborn cells (white arrowheads). **C)** UMAP plot of embryonic scRNA-seq datasets, cells coloured by dataset type (blue, TrackerSeq; grey, wild type). **D)** Histogram showing distribution of clone sizes for TrackerSeq dataset.

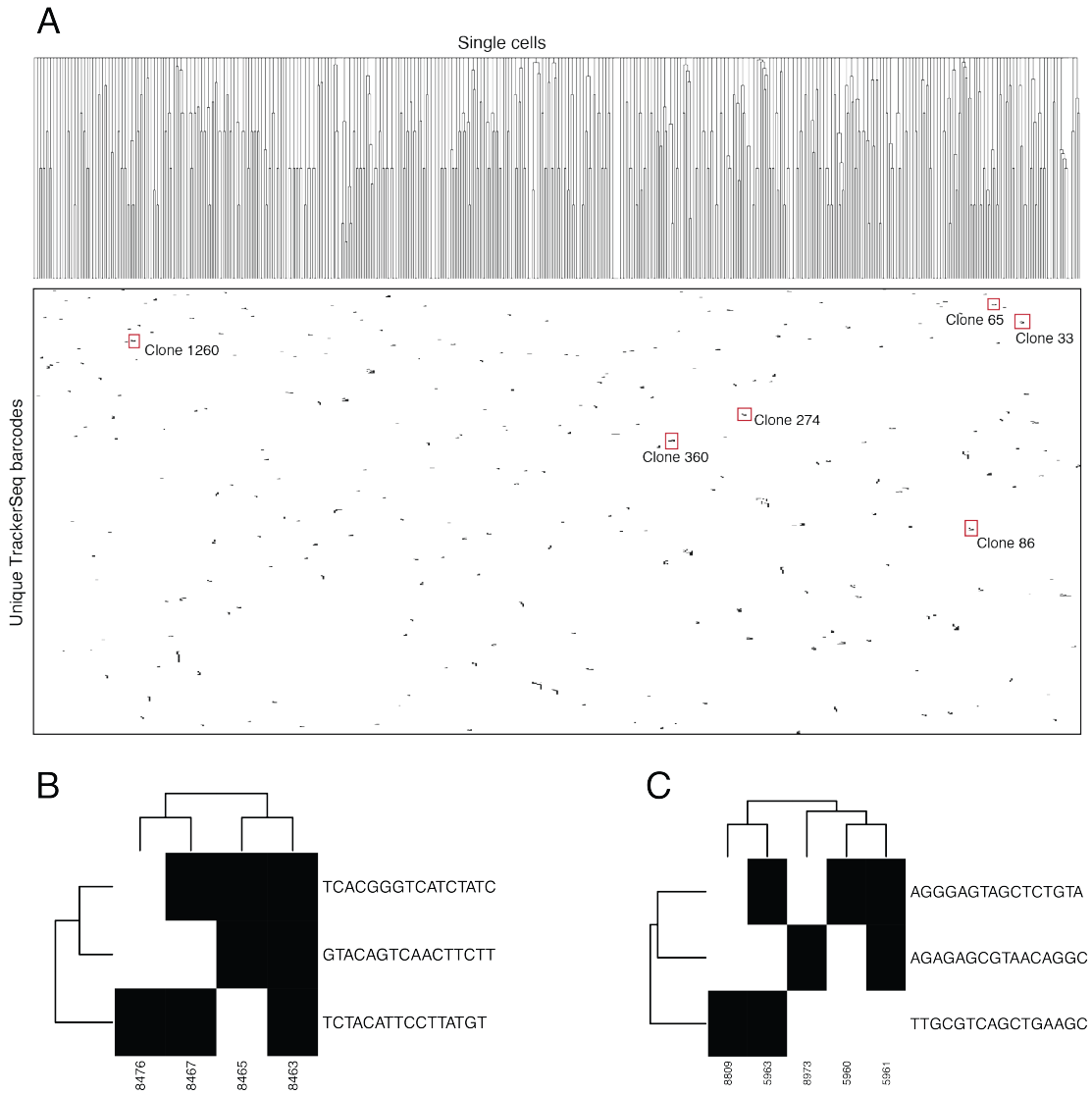

Figure 12: TrackerSeq reveals lineage relationship of multicellular clones. A) Clustered heatmap of TrackerSeqE12 barcodes. Rows are single GABAergic precursor cells for which both transcriptome and >1 TrackerSeq barcodes were retrieved; column represents unique TrackerSeq barcodes. The clonal groupings showed that there were 4,282 barcodes distributed over 2,370 cells in the total dataset, where 56.0% of them were marked by 2 or more barcode integrations, and 8.4% of them were marked by 5 or more integrations in the total dataset. Highlighted barcodes are those represented in Fig4.5(D). **B), C)** Heatmaps of individual clones 1260 and 274, respectively.

4.1.5 Newly born GABAergic sister cells diverge

We then asked whether clonally related cells traverse the same or different trajectories. Intriguingly, although cells derived from 63.6% of the clones entered the same trajectory, 36.4% of

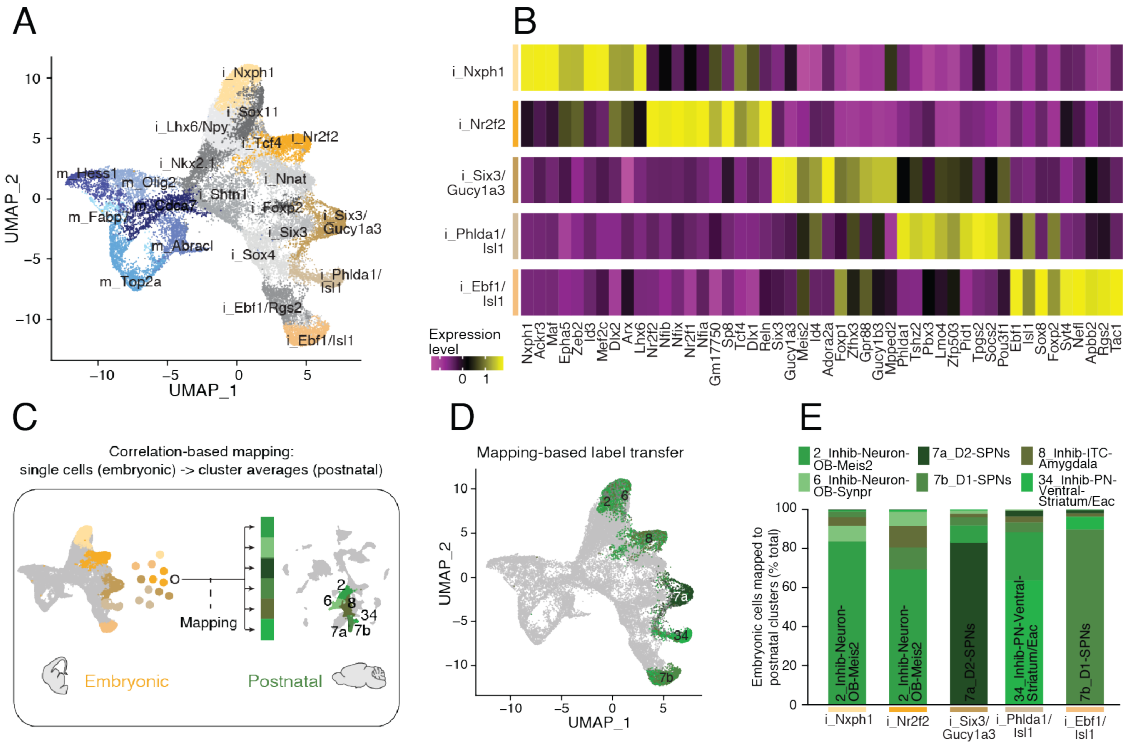

Figure 13: Mapping of embryonic datasets to postnatal GABAergic forebrain neurons. **A)** UMAP plot of integrated embryonic scRNA-seq datasets, coloured by clusters. i, GABAergic; m, mitotic. **B)** Heatmap showing the normalized expression of the top ten marker genes for the five precursor states.**C)** Schematic of the strategy for computationally mapping embryonic precursor state cells to postnatal clusters. **D)** UMAP of the embryonic dataset, with precursor state cells coloured based on the mapping results. **E)** Bar graph quantifying the correlation-based mapping of cells from the five precursor states to selected postnatal ventral GABAergic neuron clusters. The numbers on the bars indicate the dominant mapped postnatal cluster. Inhib., GABAergic; VS, ventral striatum. Figures adapted from Bandler et al. 2022.

the clones had cells that diverged into separate trajectories upon exiting cell cycle (Figure 14B). For example, we discovered sister cells on the '7b D1 SPN' and '7a D2 SPN' trajectories, '7a D2 SPN' and '8 GABAergic ITC-amygdala' trajectories, and the '2 GABAergic neuron OB *Meis2*' and '8 GABAergic ITC-amygdala' trajectories (Figure 14A,D). While the majority of GABAergic clones traverse the same trajectory, the data shows that a subset of progenitor cells in the GE can produce clones that differentiate into various GABAergic cell types at peak neurogenesis. This suggests that clonal divergence of GABAergic sister cells is a lineage-dependent process

that is initiated at the mitotic stage in radial glial cells (Figure 14C).

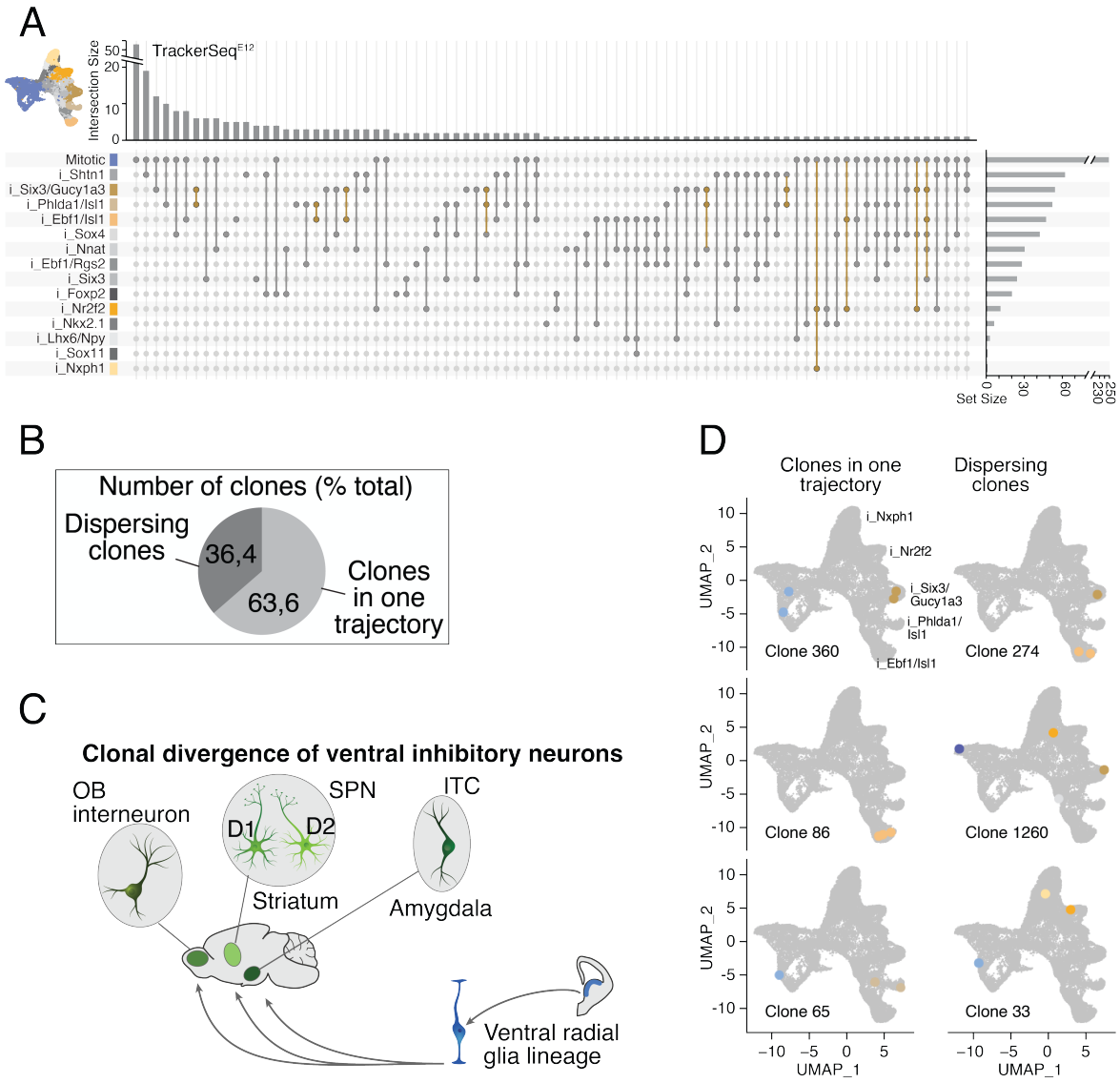

Figure 14: GABAergic sister cells diverge into different precursor states. A) Two hypotheses: Lineage-dependent vs lineage-independent mode of differentiation. **B)** Pie chart represents the percentage of multicellular clones that follow a single trajectory or dispersed across multiple precursor state trajectories. **C)** Schematic of lineage divergence for ventral inhibitory neuron cell types. **D** Examples of clones where sibling cells traverse a single developmental trajectory (left) or different trajectories (right) on the UMAP.

4.2 Fate specification of inhibitory interneurons and projection neurons

Acknowledgement

The results presented in Part II "Fate specification of inhibitory interneurons and projection neurons" were obtained as a group effort. Elena Dvoretskova designed the tCROP-seq workflow. Elena Dvoretskova, Christian Mayer, and Ilaria Vitali conducted the tCROP-seq and TrackerSeq experiments. I analyzed single-cell tCROP-seq and TrackerSeq data. Chao Feng assisted in cloning.

4.2.1 *In vivo* tCROP-seq to assess the function of MEIS2 during GABAergic fate decisions

Our TrackerSeq data showed that diverse types of GABAergic neurons can share a common lineage. The majority of progenitors produce clones that traverse the same developmental landscape, but a sizable subset produce transcriptomically diverse GABAergic cell types. However, the genetic fate specification of GABAergic is influenced by other intrinsic genetic factors, such as the expression of TFs. It is not known which TFs determine the fate of GABAergic neurons or facilitate more fundamental, developmental processes such as maturation and proliferation. Moreover, would the inactivation of certain TFs impede GABAergic differentiation?

How do TFs specify different types of GABAergic neurons? To answer this question, we modified CROP-seq (Datlinger et al. 2017), a pooled CRISPR screen method with single-cell transcriptome readout, to perturb candidate TFs *in vivo*.

MEIS2 was an attractive candidate for perturbation because it has been implicated in the generation of LGE-derived GABAergic PNs (Su et al. 2022). Haploinsufficiency of *Meis2* in

humans results in cardiac and palate abnormalities, developmental delay, and intellectual disability. Moreover, *Meis2* is highly expressed in the LGE SVZ and fairly expressed in the VZ (Su et al. 2022).

To investigate the effects of MEIS2 perturbation on cellular fate decisions in a sparse population of precursors in the GE, we modified CROP-seq (Datlinger et al. 2017), a method for pooled CRISPR screens with single-cell transcriptome readout. Instead of lentiviral vectors to deliver single-guide RNAs (sgRNAs), we used a PiggyBac **t**ransposon-based strategy (**t**CROP-seq) and *in utero* electroporation to efficiently deliver sgRNAs to cycling progenitors in the GE. The transposon system allows genes to be stably integrated into the genomes of electroporated cells and thus to be transmitted to their postmitotic daughter cells (Ding et al. 2005). This increases the pool of perturbed cells and ensures that the perturbation occurs during a period covering the peak of neurogenesis (Bandler et al. 2022). We also added specific capture sequences to the sgRNA vectors that efficiently link sgRNAs to cell barcodes, and enable sequencing of the protospacer from the transcriptome (Replogle et al. 2020). tCROP-seq sgRNA vectors also encode TdTomato to enable the labeling and enrichment of perturbed neurons. The efficiency of sgRNAs to induce frame-shift mutations was validated *in vitro* prior to the tCROP-seq experiments (data not shown).

The tCROP-seq vectors were targeted by *in utero* electroporation at E12.5 to progenitor cells of the GE in a mouse line ubiquitously expressing Cas9 (Platt et al. 2014) (Figure 15A). We electroporated a total of 14 embryos from multiple pregnant females. Of these, 8 received sgRNAs for *Meis2* (gMeis2) and 6 received sgRNAs for *LacZ* (gLacZ), which served as a control. At E16.5, most TdTomato+ cells had migrated away from the ventricular zone and colonized a variety of structures, including the striatum, cerebral cortex, and olfactory bulb (Figure 16A), consistent with the migration patterns of GE-derived inhibitory neurons at this stage (Anderson et al. 2001). Cortices, striata, and olfactory bulbs were then dissected and TdTomato+ cells were enriched by FACS. tCROP-seq allows the retrospective assessment of which sgRNA was

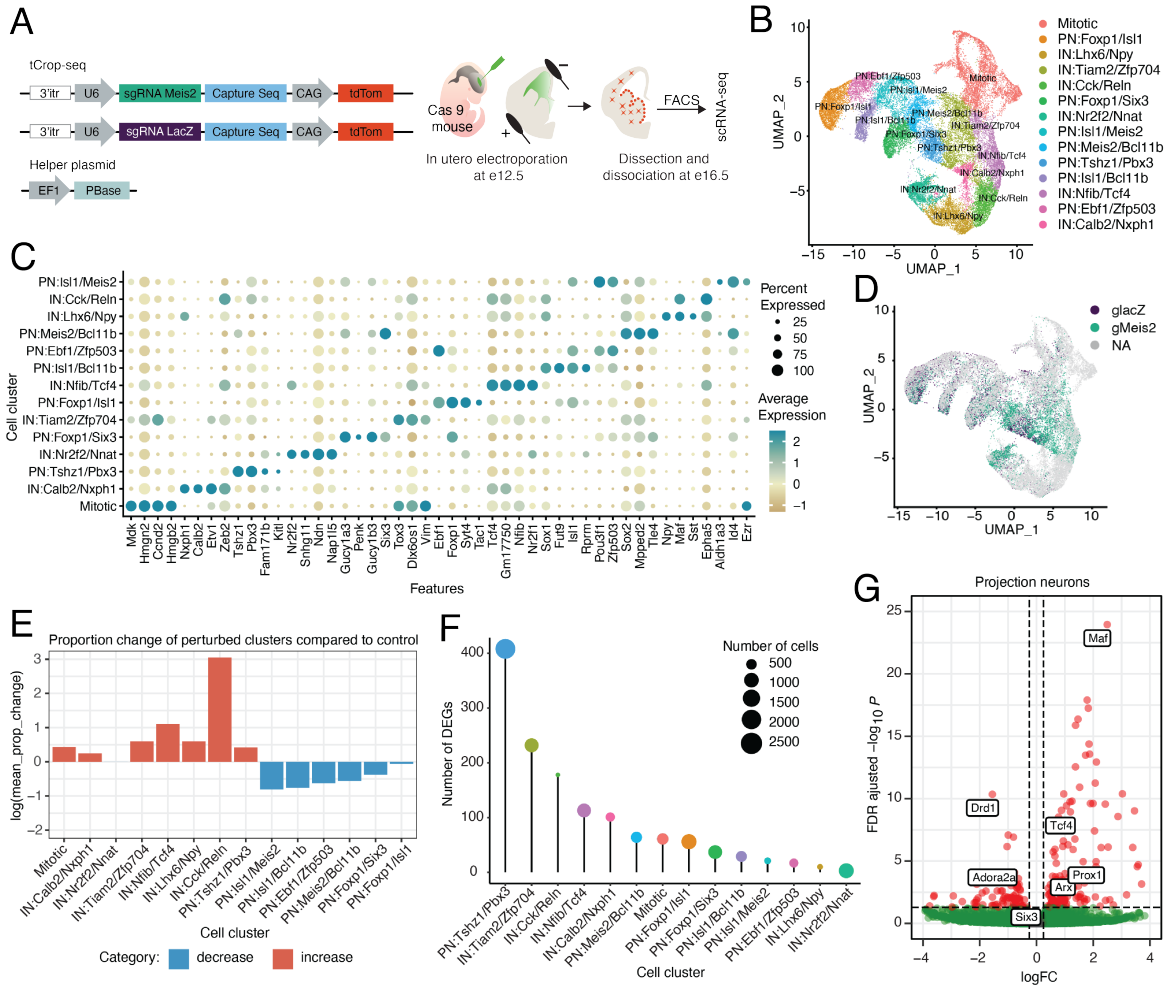

Figure 15: *in vivo* tCROP-seq of *Meis2* in the mouse forebrain. **A)** Vector maps and schematic of the *in vivo* tCROP-seq workflow, in which mutations in individual genes are introduced *in utero* and the effect is determined at a later time point via scRNA-seq. **B)** Uniform Manifold Approximation and Projection (UMAP) plot of inhibitory cells colored by clusters. **C)** Dotplot of the top five marker genes of inhibitory clusters. **D)** UMAP plot of the integrated dataset colored by sgRNA. **E)** Relative increase or decrease in the number of inhibitory cell clusters in gMeis2 compared to gLacZ. **F)** Lollipop plots showing the impact of gMeis2 on inhibitory clusters. **G)** Volcano plot depicting differentially expressed genes in gMeis2 and gLacZ projection neurons.

expressed in which cell. We pooled cells from embryos having received gLacZ or gMeis2, and conducted multiplexed single-cell RNA sequencing to minimize batch effects (Figure 15A) (Jin et al. 2020). We sequenced 6 independent scRNA-seq experiments. Together, this resulted in a

dataset containing 34481 cells passing quality controls and filtering, that could be linked with either gLacZ (11009) or gMeis2 (23472). We projected cells into a shared embedding using Harmony (Korsunsky et al. 2019) and applied a standard Seurat pipeline (Figure 17A).

4.2.2 Single perturbation of MEIS2 alters the proportion of PNs and INs

Louvain clustering grouped glia cells, excitatory neurons, and inhibitory neurons into multiple clusters (Figure 17A). We subset cells from inhibitory clusters (Figure 17B-E) and integrated them with scRNA-seq datasets from wild-type mice (Bandler et al. 2022), to get a higher resolution of inhibitory cell states (Figure15B). We annotated 14 inhibitory clusters based on shared marker gene expression and grouped them into three major classes: mitotic (mitotic), GABAergic PNs (PN:Foxp1/Six3, PN:Foxp1/Isl1, PN:Isl1/Bcl11b, PN:Ebf1/Zfp503, PN:Meis2/Bcl11b, PN:Isl1/Meis2, PN:Tshz1/Pbx3), and GABAergic INs (IN:Calb2/Nxph1, IN:Tiam2/Zfp704, IN:Nfib/Tcf4, IN:Lhx6/Npy, IN:Cck/Reln, IN:Nr2f2/Nnat; Figure 15B-C). Cells expressing gMeis2 contained a reduced proportion of PN cell-types and an increased proportion of IN cell-types, when compared to gLacZ controls (Figure 15D-E, Figure 16B). This suggests that, under normal conditions, MEIS2 promotes the generation of PNs at the expense of INs.

A pseudo-bulk DEA (Squair et al. 2021) of GABAergic neurons comparing gMeis2 and gLacZ showed reduced expression levels of genes known to be involved in PN development and increased expression levels of genes known to be involved in IN development. The impact of gMeis2 on differential gene expression was strongest on immature clusters: PN:Tshz1/Pbx3 and IN:Tiam2/Zfp704 (Figure 15F). In PN clusters, gMeis2+ cells showed decreased expression levels of genes known to be associated with PN identity, such as *Gucy1a3*, *Adora2a*, *Drd1*, *Six3*, and *Zfp503* (Kreitzer and Malenka 2008; Song et al. 2021; Shang et al. 2022; Knowles, Dehorter, and Ellender 2021), compared to gLacZ (Figure 15). Surprisingly, many genes related to IN

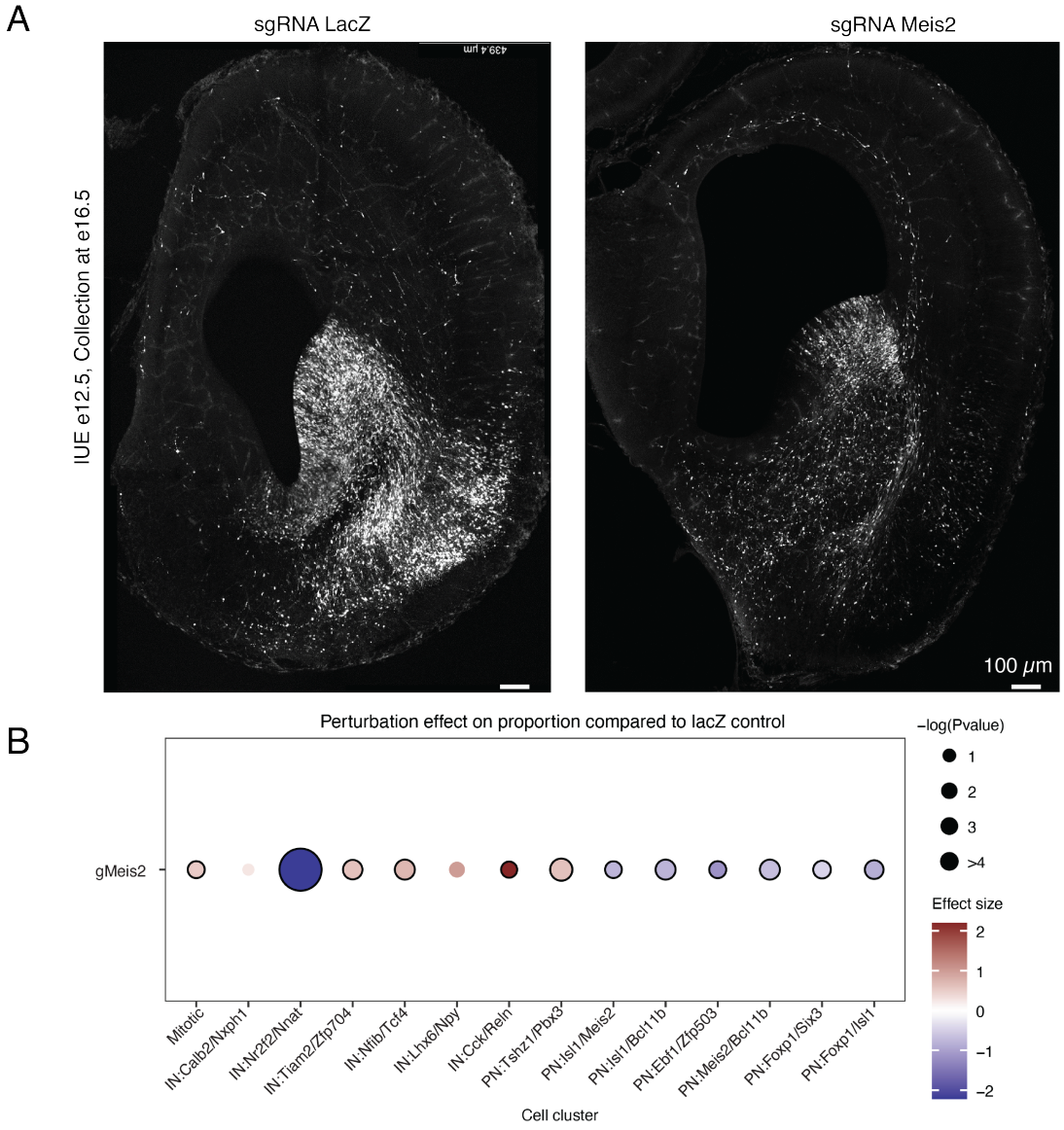

Figure 16: In utero electroporation of cells with gLacZ and gMeis2 vectors carrying a TdTomato reporter at E16.5. A) Distribution of cells in the cortex, striatum, and GE. Scale bar, 0.1 mm. **B)** Dot plot showing the effect of perturbation gMeis2 on the cell type proportion of each cell cluster compared to the control (glacZ). Black outline indicates statistical significance (p-val < 0.05).

development and specification, such as *Maf*, *Tcf4*, *Prox1*, *Arx*, *Sp8*, *Npas1* and *Nxph1* (Lim et al. 2018; Miyoshi et al. 2015; Batista-Brito et al. 2008; Wei et al. 2019) were up-regulated in PN clusters (Figure 15G). Notably, the proportion of mitotic progenitors was increased in gMeis2

compared to gLacZ Genes involved in cell proliferation and differentiation were up-regulated in the mitotic cluster in gMeis2, in particular the gene *Wnt5a*, which is part of the non-canonical Wnt signalling pathway (Megason and McMahon 2002) (Figure 15E, 17F). GO Term analysis of the up and down-regulated DEGs reveal that processes such as neuron development, axon extension, and neuron differentiation are deregulated (Figure 17G). This raises the question of how neurons with a broad PN identity (Louvain clustering grouped them into PNs) acquired CGE/MGE-IN signatures. One possibility would be that, upon the perturbation of gMeis2, progenitors of the LGE-PN lineage fail to establish proper PN identity and switch to a CGE/MGE-IN identity.

4.2.3 Combined *in vivo* lineage tracing and tCROP-seq reveal a shift in clonal compositions of perturbed cells

To test this possibility, we combined tCROP-seq with a barcode lineage tracing method called TrackerSeq (Bandler et al. 2022), that integrates DNA barcodes into the genome of electroporated mitotic progenitors, enabling the tracking of clonal relationships between their postmitotic daughter neurons (Figure 18A). tCROP-seq and TrackerSeq can be used simultaneously because we have implemented a similar transposase strategy for both methods (Figure 18A). If a fate switch occurred in the presence of gMeis2, we would expect to observe a shift in clonal compositions from PNs to INs. We used *in utero* electroporation at E12.5 to introduce the TrackerSeq barcode library and tCROP-seq sgRNAs to cycling progenitors in the GE. We collected TdTomato/EGFP+ cells from 4 independent batches and prepared sequencing libraries for transcriptomes, sgRNAs, and lineage barcodes. The cells with TrackerSeq barcodes were already part of the preceding tCROP-seq analysis and were thus integrated in the same embedding (Figure 18B). Consistent with Bandler et al. (2022), we found clones composed of mitotic cells, PNs, INs, and combinations thereof (Figure 18C-D). The clonal size of multi-cell clones was

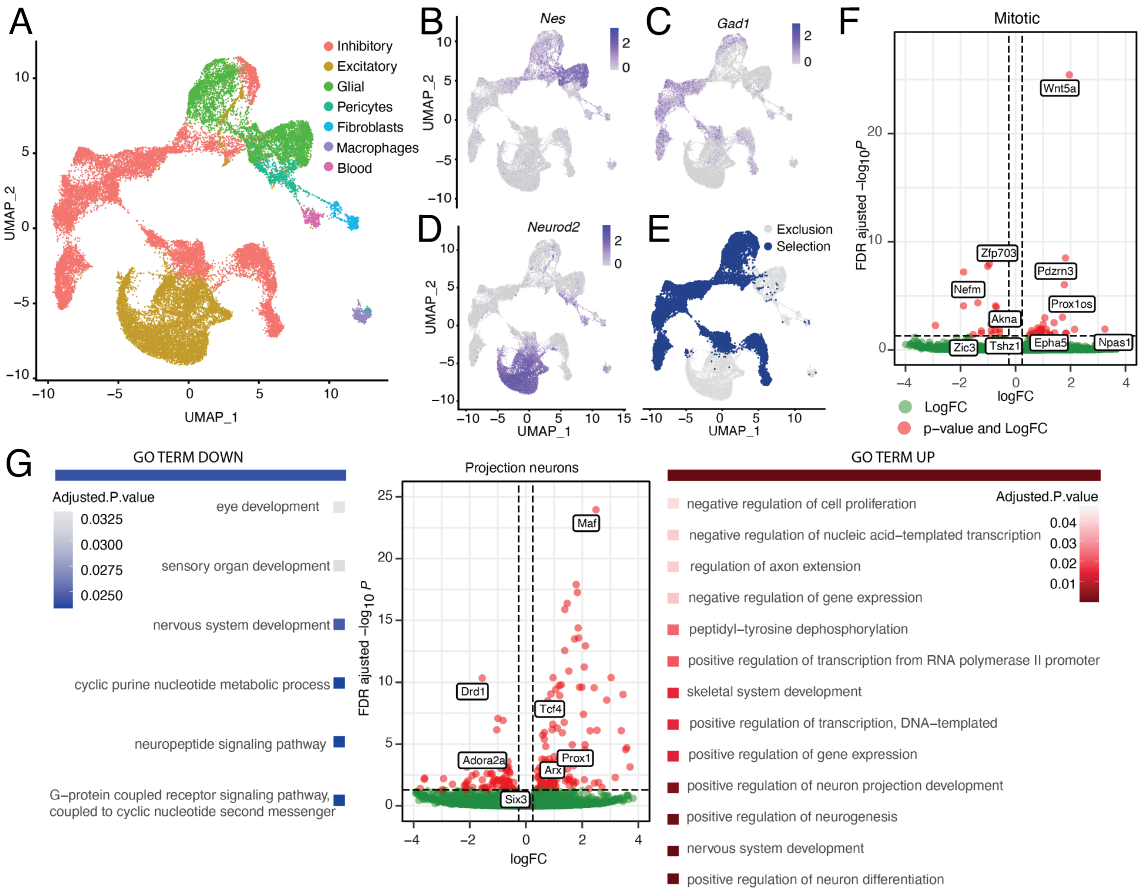

Figure 17: Single cell analysis of cell classes at E16. **A)** UMAP plot of E16 data colored by cell classes. **B)**, **C)**, **D)**, Feature plots of the canonical marker genes *Nes*, *Gad1*, and *Neurod2*. **E)**, UMAP plot depicting the selection of cells for downstream analysis. **F)**, Volcano plot depicting differentially expressed genes in gMeis2 and gLacZ mitotic cells. **G)**, Gene ontology analysis on differentially expressed genes (DEG)s of clusters belonging to the projection neuron class.

unchanged in gMeis2 compared to gLacZ (Figure 18E). The proportion of clones consisting of only mitotic cells was increased in gMeis2 compared to gLacZ, which agrees with a recent report showing that MEIS2 is required for LGE progenitors to leave the cell cycle (Su et al. 2022) (Figure 18F). We found many clones that dispersed across cell states. For example, 225 clones consisted of mitotic cells and PNs (mitotic-PN), and 100 clones consisted of mitotic cells and INs (mitotic-IN; Figure 18F). Strikingly, when we compared clonal patterns of gMeis2 and gLacZ cells, we observed a pronounced shift toward IN-only and mitotic-IN clones. Conversely, the

number of PN-only, and mitotic-PN clones was decreased (Figure 18F). Our results suggest that perturbation of cells with gMeis2 causes a fate switch from PNs to INs.

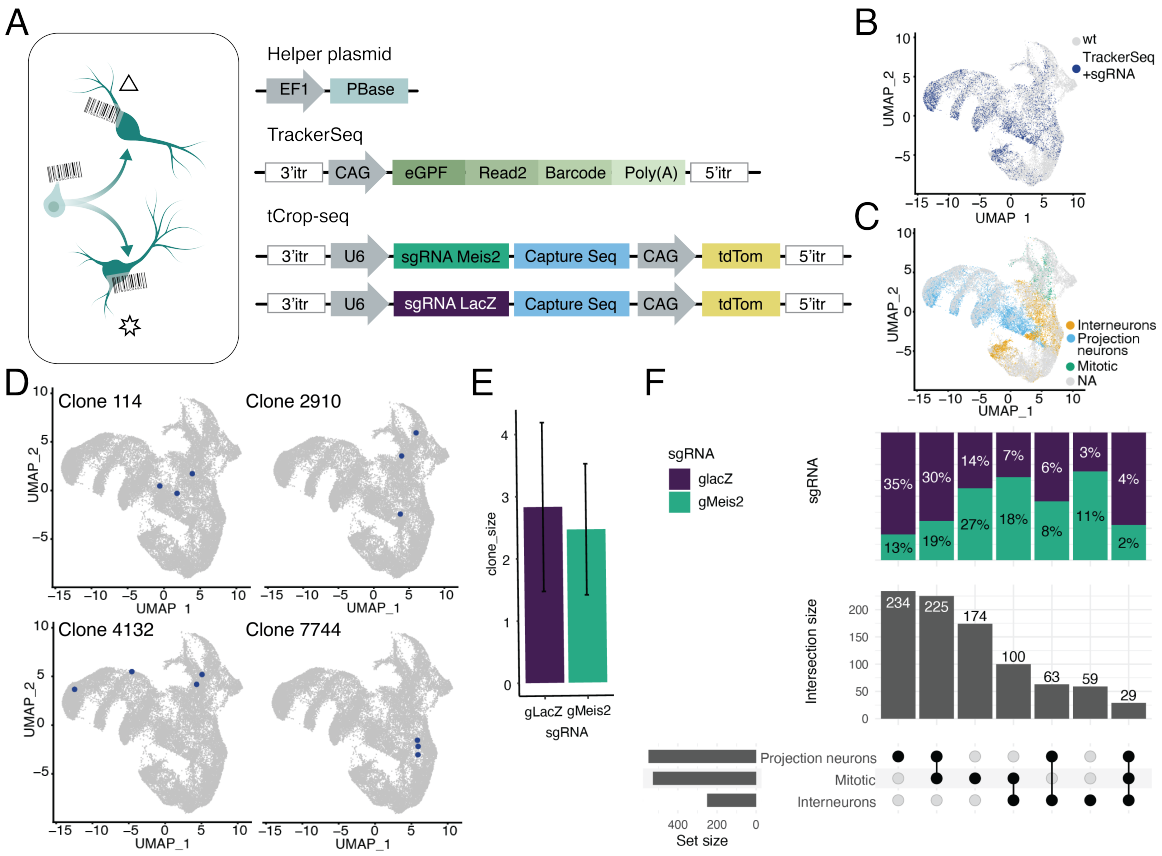

Figure 18: *In vivo* TrackerSeq lineage tracing and tCROP-seq perturbation of *Meis2*. Schematic of the TrackerSeq lineage tracing process, in which clonal boundaries are determined using a diverse library of RNA tags. **B)** UMAP of the integrated dataset, where cells that contained TrackerSeq lineage barcodes are colored. **C)** UMAP of the integrated dataset colored by cell class (mitotic, interneurons, projection neurons). **D)** Examples of clones that are shared between classes, and an example of a clone restricted to one class. **E)**, Bar graph depicting the average clone size of inhibitory clones in the gLacZ and gMeis2 datasets. **F)** UpSet plot showing clonal intersections between cell classes. The bar graph on top displays the proportion of clones belonging to gLacZ or gMeis2, the bar graph in the middle shows the number of observed intersections, and the bar graph on the left indicates the number of cells per cluster.

4.2.4 Meis2 and Lhx6 alter gene modules in PNs and INs

To explore how the depletion of embryonic TFs alters postnatal cell-type composition and identity, we performed pooled tCROP-seq experiments with sgRNAs for *Meis2* (gMeis2), *Lhx6* (gLhx6), *Tcf4* (gTcf4), and LacZ (gLacZ, control). Like LHX6, TCF4 is enriched in INs, but is expressed in all GEs (Kim et al. 2020). A mixture of plasmids encoding one of the sgRNAs, CAS9-EGFP (a gift from Randall Platt) (Platt et al. 2014), and a pB-helper plasmid was targeted to the GE via *in utero* electroporation at E12.5 (Figure 19A-B). At P7 we collected 35 pups, enriched TdTomato/EGFP positive cells with FACS, and performed pooled scRNA-seq. Ten scRNA-seq datasets were combined in silico, clustered, and annotated based on known marker genes (Figure 19C-D, S20A). All three perturbations had a significant effect on the composition of cell types compared to the gLacZ control (Figure 19E-F). Cells expressing gLhx6 showed an increased proportion of medium spiny projection neurons (D1/D2 MSNs), olfactory bulb (OB) precursors, and INs compared to gLacZ. An increase of CGE INs after *Lhx6* deletion has previously been reported (Vogt et al. 2014). Consistent with our embryonic tCROP-seq data, the proportion of INs was also increased in gMeis2. In cells expressing gMeis2, intercalated cell (ITC)s were depleted and the number of olfactory bulb inhibitory neurons and oligodendrocyte progenitor cells (OPCs) was reduced (Figure **??**e-f). The ITC depletion is consistent with the E16.5 tCROP-seq results, where the PN:Tshz1/Pbx3 cluster (likely corresponding to immature ITCs (Kuerbitz et al. 2018) had the most DEGs (Figure 15F). gTcf4 expression had a more modest effect on cell proportions, showing only a slight reduction in inhibitory neurons in the olfactory bulb. Furthermore, both gMeis2 and gTcf4 showed a reduced number of astrocytes. Across all clusters, gLhx6, gMeis2, and gTcf4 positive cells had a total of 90, 58, and 7 DEGs respectively (Figure 19G-H). Many of them were marker genes specifically expressed in IN or PN cell types. gLhx6 perturbed cells were enriched for PN specific genes (*Isl1*, *Foxp1*, *Ebf1*, *Adora2a*, *Drd1*, *Six3*). By contrast, gMeis2 DEGs were enriched for IN-specific genes (*Maf* and

Prox1os) and depleted for PN-specific genes (*Mpped2* and *Pbx3*). Our data support the idea that MEIS2 primarily induces PN fate and LHX6 primarily induces IN fate.

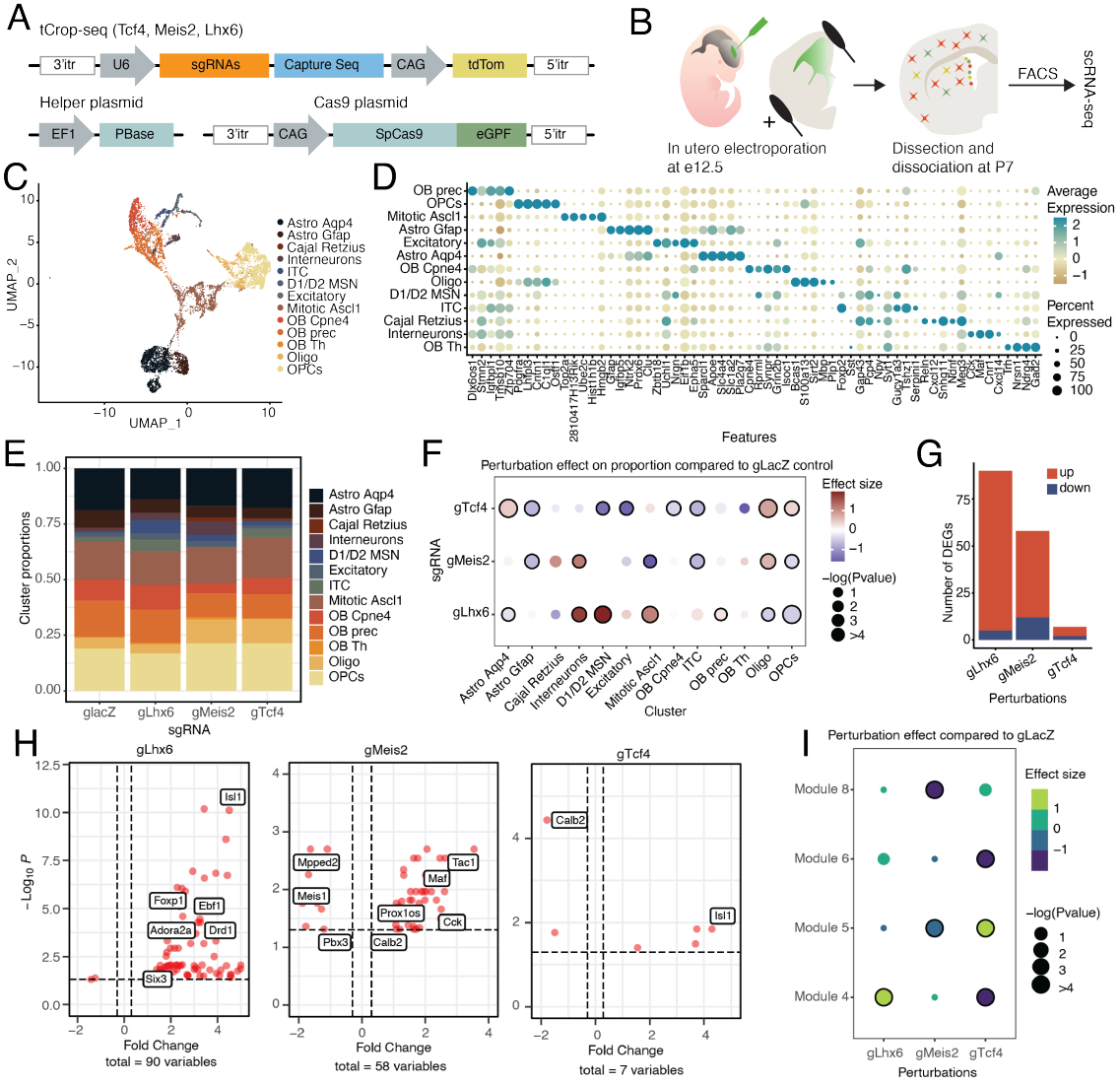

Figure 19: Embryonic disruption of transcription factors alters postnatal cell types. A) Vector maps of tCrop-seq. **B)** Schematic illustrating the workflow of tCROP-seq. **C)** UMAP plot of the P7 data colored by cell type. **D)**, Dot plot showing the top 5 marker genes of each cell type. OB, olfactory bulb cells; OPC, oligodendrocyte progenitor cells; ITC, intercalated cells; MSN, medium spiny neurons; Oligo, oligodendrocyte, Astro, astrocytes. **E)** Cell type compositions for each sgRNA. **F)** Dot plot showing the effect of perturbation on cell type composition for each sgRNA compared to the control (gLacZ). Black outline indicates statistical significance (p-val < 0.05). **G)** Bar plot showing the number of differentially expressed genes detected in each sgRNA. **H)** Differential gene expression analysis of inhibitory neurons in each sgRNA compared to control (LacZ). Volcano plot showing differentially expressed genes that meet the cutoff criteria (FDR < 0.05, avg_logFC > 0.5). **I)** Dot plot showing the effect of perturbation by sgRNAs on the module scores of inhibitory modules.

scRNA-seq data are highly heterogeneous and have numerous zero counts, making it challenging to detect subtler perturbation-based biological changes in single cell datasets. To overcome these limitations, we utilized Hotspot (DeTomaso and Yosef 2021), a tool that identifies co-varying groups of genes (modules). Each cell was assigned a gene module score, with higher scores indicating higher association with that module. We identified 8 Hotspot gene modules (Figure 20B), 4 of which were neuronal (Figure 19I, 20C). Module 5 is represented mostly in olfactory bulb neuroblasts and contains genes enriched for neuronal differentiation. Module 4 is represented in MSN cell types and contained MSN marker genes (*e.g.*, *Foxp1*) and genes involved in retinoic acid receptor signalling (*Rarb*, *Rxrg*). The retinoic acid pathway is involved in the switch between proliferation and differentiation (Berenguer and Duester 2022), which is essential for striatal development (Chatzi, Brade, and Duester 2011). Module 8 was represented in OB precursors and ITC cells. This module contained *Meis2* as well as some of its target genes, such as *Pbx3* and *Etv1*. Module 6 was represented in the OB-Cpne4 population and was characterized by genes involved in calcium response and synapse organization. We fitted a linear regression model that accounted for the batch and number of genes, and extracted the effect sizes to estimate how the module scores in the perturbed cells deviated from gLacZ control cells (Jin et al. 2020). For the three TFs, the perturbations had significant effects across different modules (FDR-corrected $P < 0.05$; Figure 19I). The perturbation of *Lhx6* was positively associated with the expression of module 4, consistent with the change in cell proportion and change in differentially expressed genes. The perturbation of *Meis2* lowered the expression of both modules 8 and 5. The perturbation of *Tcf4* had a significant effect across modules 6, 5, and 4, consistent with previous findings showing that TCF4 is a key facilitator of neurogenesis and neuronal differentiation (Figure 19I) (Mesman, Bakker, and Smidt 2020; Teixeira et al. 2021). Taken together, the tCROP-seq data at P7 indicate a marked influence of MEIS2, LHX6, and TCF4 on PN and IN specification.

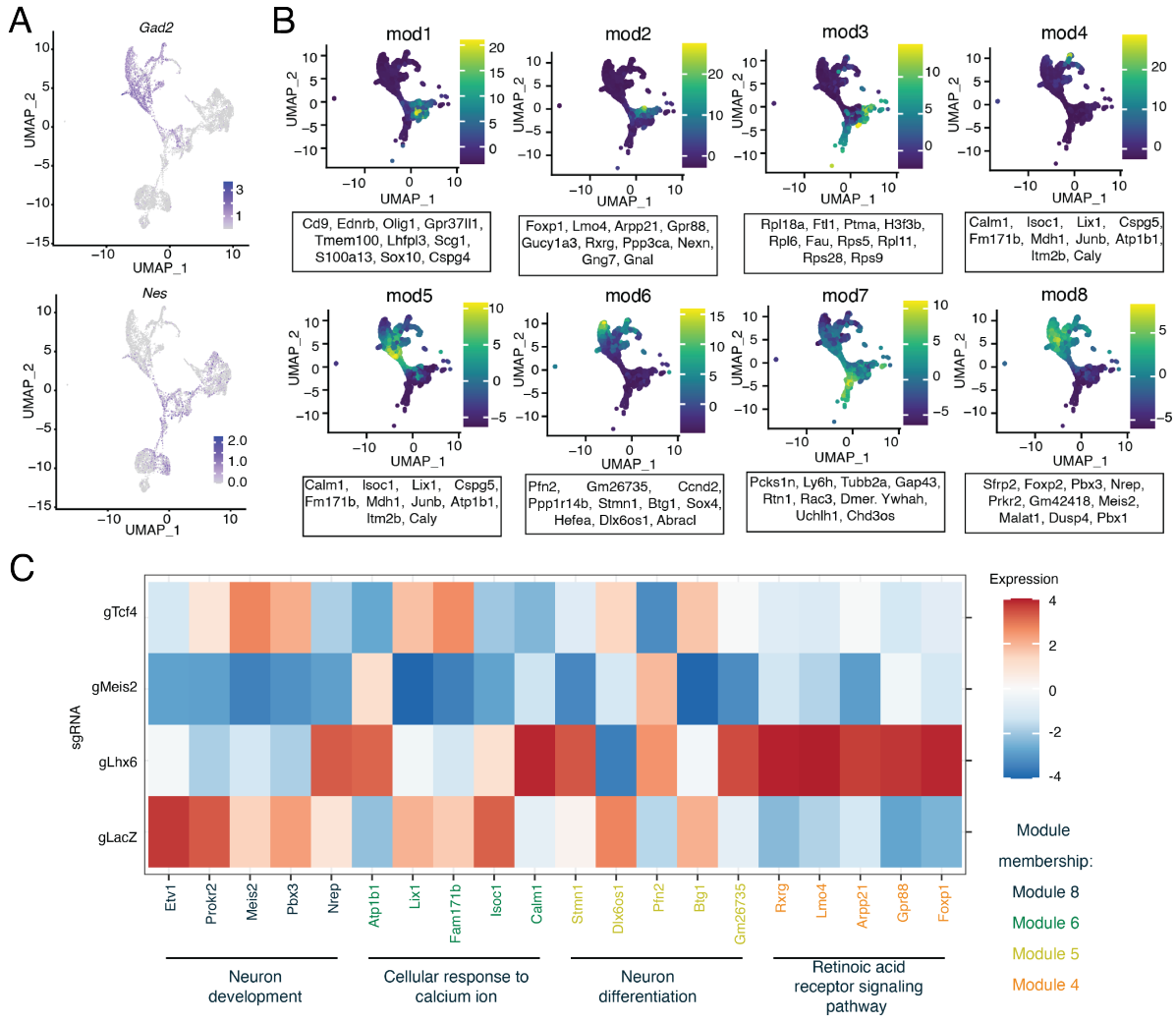

Figure 20: Module analysis of P7 tCROP-seq. A) Feature plots of canonical marker genes *Gad2* and *Nes* at postnatal day 7. The expression is depicted from low (gray) to high (pruple). **B)** Feature plots of gene module expression scores and the correlated genes within each module. **C)** Average expression of top 5 module genes for each sgRNA at postnatal day 7.

Discussion

5.1 Summary of key findings

In this thesis, I explore the different ways in which genetic programs contribute to the diversity of GABAergic cell types through the use of single-cell sequencing compatible techniques. The first way I approached this broad question was by studying cell lineage. It has been hypothesized that transcriptomically similar neurons may arise from the same progenitors. To address this, I developed a method, TrackerSeq, that simultaneously captures both the transcriptomic signatures and lineage histories of individual cells. The TrackerSeq barcode library has high diversity in the 10^5 range and can be electroporated into the embryonic brain to label progenitors *in vivo*. More importantly, TrackerSeq is compatible with the 10x single-cell transcriptomic platform and is capable of identifying multicellular clones. Using this method, we found examples of transcriptomically distinct cell types that arose from the same lineage (divergence) and clones that give rise to transcriptomically similar cell types.

I then approached the idea of genetic specification from another perspective: transcription factor control. Our perturbation of MEIS2 resulted in a marked increase in the proportion of interneurons at the expense of projection neurons. The shift to interneuron gene signatures could also be observed through the bulk analysis of projection neurons, where genes related to IN development and specification were upregulated. We then combined tCROP-seq with TrackerSeq to test whether progenitors of the LGE-PN lineage acquire partial CGE/MGE identity after MEIS2 perturbation, where we observed an increase in the number of IN clones. Finally, perturbing *Meis2* and *Lhx6* altered the composition of GABAergic cell types, gene expression, and modules in the postnatal stage. Consistent with embryonic data, perturbing *Meis2* resulted in an increase in the proportion of interneurons, up-regulation of genes and modules related to

interneurons. The perturbation of *Lhx6* resulted in the opposite effect, leading to an increase in the number of projection neurons, projection neuron-related genes and modules.

In conclusion, both lineage and transcription factor control are intrinsic factors that specify the fate of GABAergic neurons. Additionally, TrackerSeq is an effective technique that can not only identify multicellular clones but also be combined with single-cell perturbation to defect changes in cellular fate.

5.2 Developing a highly-diverse barcode library

Traditional lineage tracing methods that combine mouse genetics with fluorescently tagged retroviruses rely on spatial segregation to infer clonal relationship (J. Ma et al. 2018). This could result in clones that erroneously lump and split together because dispersed inhibitory neurons labeled with the same fluorophore are assumed to be derived from independent clones (J. Ma et al. 2018; Bandler 2019). Due to these limitations, researchers have moved to the use of DNA barcodes to encode clonal information. Unlike fluorescent reporters, which can only query a set number of distinct clones, the complexity of the DNA sequence is exponentially scaled with the length and number of barcodes created. It scales to the extent that it is possible to record each division event (Wagner and Klein 2020). Coupled with high-throughput sequencing, the recorded information can be read out retrospectively. A big challenge I had to overcome was developing a barcode library that was diverse enough to uniquely label hundreds of progenitor cells.

In a collaborative effort, I overcame this problem using various traditional molecular cloning techniques: Gibson Assembly, electroporation of *E.coli.*, and sucrose counterselection. The barcode, with random nucleotides bridged by fixed nucleotides, was designed so that it has a theoretical complexity of 4.3 million lineage barcodes (16^8). Sanger sequencing results showed that each of the 10 colonies we picked had unique barcodes, and the hamming distance analysis

revealed that each barcode differed from each other by at least 5 bp. Finally, RNA-seq results showed that the library made has a diversity of around 2 x 10^5. An IUE experiment typically labels fewer than 1,000 cells per brain. Under these conditions, this batch of TrackerSeq library is still 2 to 3 magnitudes higher and complex enough to label each cell with a unique barcode. However, if the number of cells desired to label exceeds 1,000, overrepresentation of the barcode may pose a problem.

Many DNA barcode library production protocols also use highly competent *E.coli* cells to clone their libraries. The main advantage of using high-efficiency competent cells is that the complexity of the library can be maintained (Kong et al. 2020) and that cells can be grown overnight to obtain greater amounts of the DNA library for IUE. A con of this method is that one may have to repeat the protocol from the very beginning if the batch of competent cells used is of poor quality and a low number of transformants are obtained.

Our protocol for making TrakcerSeq shares similarities with the protocol for making TracerSeq and CellTag, with a number of key differences. Like the TracerSeq protocol, we also opt for Gibson Assembly to clone the library, but rather than amplifying the oligos by PCR to achieve diversity, we ordered the DNA barcodes as oligoes from IDT. Unlike Biddy et al. (2018) and Wagner et al. (2018), we also employed electroporation rather than heat shock of *E.coli* since the transformation efficiency is higher. However, electroporation is more prone to high background than heatshock, and the SacB gene had to be cloned into the vector to mitigate this effect.

The TrackerSeq production protocol is capable of generating a highly diverse library, comparable to that of other methods. Most DNA barcode libraries are limited by two main traits, the theoretical diversity of the library and the transformation efficiency of the method used to clone the library. The theoretical diversity of the library comes down to the number of random base pairs that make up the barcode. For example, TracerSeq's 20-nucleotide sequence yields 10^{12} possible variants. However, the transformation efficiency of *E.coli* ranges from 1 x 10^6 for

chemically competent cells to max 3 x 10^{10} CFU/µg, where each *E.coli* would contain one unique barcode, which is still 2 magnitudes lower. In our own protocol, we were met with the same difficulties. Like them, we had to scale multiple Gibson Assembly and *E.coli* transformation reactions to get as close to theoretical diversity as possible. In the end, regardless of differences in barcode design, it comes down to how competent *E.coli* was at the time of transformation and how many Gibson Assembly reactions and *E.coli* transformations were run in parallel. A direct comparison of our method's library diversity is not possible, due to the difference in barcode length. However, in our hands, no barcodes were shared between embryos that were electroporated with the same batch of TrackerSeq library, eliminating the possibility of non-related cells inheriting identical barcode sequences.

In conclusion, the TrackerSeq production protocol is capable of generating a highly diverse barcode library that can label thousands of cells with unique barcodes. The library production protocol serves as a useful resource for those in the scientific community interested in making a barcode library, whether it be for chemical compound screens, the study of clonal diversity, or genomic screens (Lyons et al. 2017). In particular, the *E.coli* step for library amplification can reduce the need to PCR amplify synthesized DNA barcodes, introducing fewer errors and providing higher quantities of DNA barcodes to use in subsequent steps.

5.3 Lineage tracing in the mammalian forebrain

Investigations of the GABAergic neural clonal relationship were previously limited in scope due to technical difficulties. Before the establishment of scRNA-seq, previous studies explored the dispersion patterns of GABAergic interneuron, but lacked transcriptomic information (Harwell, Mayer). They found that lineage had no influence on the migration patterns of GABAergic neurons.

The TrackerSeq library developed in this thesis enables retrieval of both the transcriptome

and the in vivo lineage history of single cells. Combined with *in utero* electroporation, a technique widely used in neurobiology, cell lineage history is recorded by integrating multiple barcodes.

Compared to other emerging single-cell compatible lineage-tracing techniques, TrackerSeq has a number of advantages. Many other DNA barcoding techniques rely on retrovirus infection of barcoded GFP reporters for lineage tracing, in which GFP expression is used to identify successful integration of the DNA barcode. If there is transcriptional silencing of GFP, this could lead to underestimation of lineally related clones (Swindle and Klug 2002; Mayer et al. 2015), a problem noted in other lineage studies in the forebrain (Cepko et al. 2000; Halliday and Cepko 1992; McCarthy et al. 2001). TrackerSeq circumvents this issue since it relies on transposons rather than retroviruses, to integrate GFP and DNA barcodes into non-specific regions of the genome. Our transposon-based method also has a higher resolution because it enables cells to receive multiple barcodes per cell, whereas retrovirus-based methods allow only one barcode per cell.

Regardless of which method of lineage-tracing is used, it appears that the efficiency of the transcriptome retrieval depends not only on the labeling method, but also on the age at which the cells are harvested. In both lineage tracing and CRIPSR perturbation experiments, we observed that retrieving at E16 compared to P7 resulted in 2-3 times higher rate of cells containing lineage barcode or CRISPR photospacer. This could be due to the extensive number of neurons undergoing apoptosis during development (Hollville, Romero, and Deshmukh 2019).

Transposon-based methods such as TrackerSeq need to be electroporated into cells to enable labeling. This has some advantages and some limitations. It allows us to target barcodes to a spatially defined region, but is less effective for younger embryos, where their young ages and small brains make it more difficult to electroporate. Lentiviral lineage barcoding methods such as STICR can be injected into embryos of earlier stages (Bandler et al. 2022). Intriguingly, introducing barcodes earlier also resulted in much higher number of cells recovered (Bandler

et al. 2022), even for lentiviral lineage barcoding methods like STICR. Of all cells that were injected with the STICR library and recovered in the same way, cells that were injected at E10 had the highest number of recovered cells compared to E12, E12 compared to E14, etc. These results are perhaps unsurprising, given that proliferation dominates earlier on in neurogenesis (Belmonte-Mateos and Pujades 2021). The rapid proliferation of cells in early neurogenesis results in more cells inheriting the barcodes.

Since TrackerSeq is based on another published transposon-method based lineage tracing method, TracerSeq, TrackerSeq is not conceptually novel. But its compatibility with the commercial 10X genomics platform–used by over 4,500 peer-reviewed publications and counting–makes it more accessible than the inDrops compatible TracerSeq (https://www.10xgenomics.com-/resources/publications). Moreover, our method is more clinically relevant since it can be implemented in mammalian systems such as human cell lines and other models organisms such as rats and ferrets.

In summary, TrakcerSeq is a useful method for *in vivo* lineage tracing of GABAergic neurons in the forebrain. Its compatibility with a commercial scRNA-seq platform enables the simultaneous recovery of both the lineage history and the transcriptome of GABAergic neurons. More importantly, TrackerSeq can build on previous research on the role of lineage in the specification of different types of neuronal cells and how they integrate into circuits. However, TrackerSeq, like other DNA barcoding methods, is affected by biological factors such as the timing of library injection and the date of transcriptome retrieval.

5.4 Lineage divergence of GABAergic neurons

The construction of lineage hierarchies reveals valuable information about cell potential, identity, and behavior. Although previous research has found that the dispersion patterns of GABAergic interneurons are independent of lineage (Harwell et al. 2015; Mayer et al. 2015), the extent

to which lineage determines the transcriptomic identity of these neurons has remained obscure.

Since TrackerSeq is able to capture clonal histories, we were able to address this question. Using TrackerSeq, we found transcriptomically distinct cell types that share a clonal relationship (that is, they are lineage divergent). An intriguing example of divergence was GABAergic clones that not only diverged into types of grossly different transcriptomic signatures, but also migrated to distinct brain regions. Most clones (63.6%) were found in one trajectory, but a sizable portion of the clones (36.4%) diverged into differing trajectories immediately after leaving cell cycle. Indeed, precursors of OB interneurons, MSNs in the striatum, and ITC amygdala cells were found to be clonally related. In agreement with Mayer et al. (2018), the clonal divergence we observed indicates that newborn cells born from the same progenitor often diverged into differing trajectories after exit from cell cycle, suggesting that inhibitory neuron specification already takes place at the progenitor stage and not later, when external signals are present.

However, there are still open questions that could not be addressed by the TrackerSeq method. Currently, it is not known whether sequential production of different types of inhibitory cells (temporal progression) follows a set sequence or is influenced by stochastic events that occur during cell cycle exit (Figure 4). The GE can be partitioned into at least a dozen spatial subdomains that are characterized by the combinatorial expression of TF as well as morphogen gradients (Wonders et al. 2008; Flames et al. 2007). It could be that the temporal progression of progenitors depends on the dynamic timing of TF expression and the concentration of morphogens to which they are exposed. How these progenitors interact with TFs and morphogens in different spatial subdomains of GE could explain the enormous diversity of inhibitory types generated in the forebrain. One potential approach to investigate this inquiry could involve integrating lineage barcoding with spatial transcriptomics. This would enable the evaluation of the connection between the distribution of clones and the gradient of morphogens.

In conclusion, TrackerSeq is able to capture partial clones, which is sufficient to detect lineage convergence and divergence. However, it can only deduce, but not prove, lineage restric-

tion because sibling clones that appear to only traverse one trajectory could have sisters cells that are lost during FACS and tissue dissociation. To prove lineage restriction, one would use a lineage tracing method such as zMADM (zebrafish mosaic analysis with double markers) (Xu, Kucenas, and Zong 2022) that permits lineage tracing of cells without tissue dissociation. Although it has lower throughput, zMADM is capable of labeling two daughter cells from a single mother cell with different colors. Furthermore, zebrafish can be directly fixed and stained with antibodies specific for certain neuronal populations and allow *in vivo* imaging, avoiding the problem of cell loss common to scRNA-seq-based methods.

Moving forward, TrackerSeq can also be used to advance our understanding of neurodevelopmental diseases. Somatic mutations are clonally inherited and contribute to numerous neurodevelopmental diseases. The determination of lineage relationships could identify which subsets of cells are most affected by developmental disorders. In fact, single-cell *in vivo* lineage tracing is already being used to explore the clinical phenotypes of tumors (Yang et al. 2022; Simeonov et al. 2021). In the future, we can expect to see the same technique applied in the context of developmental disorders.

5.5 Meis2 promotes projection neuron fate

The degree to which TFs contribute to the tremendous diversity of GABAergic projection neuron and interneuron types is not completely known. Past research indicates that the dynamic expression of certain TFs and their cofactors may be needed to specify GABAergic cell types. For example, ablating *Lhx6* prevented MGE-derived interneurons from migrating tagentially into the cortex. How other GE-expressing TFs, such as *Meis2*, specify the fate of GABAergic neurons is not fully understood. *Meis2* has been implicated in the generation of LGE-derived GABAergic PNs and is a member of the TALE family of homeodomain-containing TFs.

Meis2 is a member of the TALE family of homeodomain-containing TFs and has been im-

plicated in the generation of LGE-derived GABAergic PNs.

Upon embryonic perturbation *Meis2*, we observed an increase in the proportion of INs compared to PNs. We observed ectopic expression of IN-related genes in *Meis2* perturbed PNs. Moreover, we see that its perturbation tended to affect more immature clusters. We identified several miRNA host genes that were downregulated in Meis2-tCROP-seq: *Mirl124-2hg*, *GM27032*(MIR-124a-2), *Arpp21*(miR-128-2), and *GM27032*(miR-124a-3).

Our results are in agreement with what was observed by Su et al. (2022). Like them, we see a strong reduction in MSN cell types. However, our single-cell data allow us to further explore this phenotype at a higher resolution. We see that DRD1 cell types are more reduced; Isl1 expressing PN cell types are more reduced compared to Foxp1 expressing MSN precursors. Furthermore, our DEA data indicate that MEIS2 may inhibit IN fate to promote PN fate by activating repressive TFs such as ISL1, FOXP1/2 and SIX3, through co-repressors such as TLE1/4 or by promoting miRNA expression. Finally, using lineage tracing data, we observe fate switching of GABAergic neurons from projection neuron to interneuron cell types.

Haploinsufficiency of *MEIS2* in humans results in developmental delay and intellectual disability. Our results in mouse could explain these clinical phenotypes. The striatum is responsible for emotions and decision-making functions (Humphries and Prescott 2010; LeDoux 2012). Therefore, the improper specification of striatal MSNs, which make up 95% of the striatum, could be the cause of the clinical phenotype observed in humans.

In summary, our data provide a higher resolution of MEIS2's role as an TF that is necessary to specify the expression of projection neurons, and it may do so by activating corepressors that suppress interneuron fate. Moving forward, we expect that rescue experiments, such as overexpressing FOXP1/2 could clarify this hypothesis. Although we were able to explore MEIS2's function in the GE, MEIS2'S function in the cortex may be different and remains an unexplored topic.

5.6 Similarities and differences between embryonic and postnatal data

Our own results and those published by others indicate that for any scRNA-seq experiment, harvesting the cells at an earlier time point yields more cells. This could lead to some discrepancies between embryonic and postnatal data.

The embryonic and postnatal data from Meis2 show some congruence, as well as discrepancies. The reduction in ITCs in the P7 datasets corresponds to the high number of DEGs observed in the immature ITC cluster. In contrast to the embryonic dataset, the reduction of MSNs in the postnatal data set is not particularly significant. In both the embryonic and postnatal datasets, up-regulation of interneuron-related genes in inhibitory neurons can be observed. Interestingly, an increased proportion of interneurons is observed at both timepoints.

The discrepancies in the cell proportion data of MSNs could be either due to the lower number of cells recovered from the postnatal dataset or due to genetic compensation, a phenomenon whereby the loss of one gene is compensated by another with overlapping function and expression patterns (El-Brolosy and Stainier 2017). It is possible that MEIS1, a paralog of MEIS2, can compensate for the loss of MEIS2 as cells migrate and start to form synaptic connections in the brain at later timepoints. Indeed, *Meis1* and *Meis2* in the lens placode exhibit redundant functions during lens formation.

One way to address this is to examine the loss of MSN at postnatal timepoints in mutant mice where *Meis2* is conditionally KO (Mei2-CKO). The number of MSNs could be compared between control and Meis2-CKO mice at a postnatal timepoint.

Taken together, the embryonic and postnatal data are mostly in agreement. Any observed differences could be due to biological factors or the lower number of MSNs recovered at a postnatal timepoint.

5.7 Lhx6 and Tcf4's role in the specification of GABAergic neurons

Some TFs expressed in the GE and GABAergic cell types during development have been shown to be important for the specification of GABAergic neurons. For example, NKX2-1 is expressed in the MGE and has been demonstrated to be necessary for the specification of MGE-derived cortical interneurons. Therefore, other TF expressed in the GE may also be important for the specification of inhibitory neurons, making them attractive targets for perturbation. Their combinatorial and selective expression in different parts of the GE gives rise to a large variety of GABAergic cell types. The manner in which TFs such as Lhx6 and Tcf4 specify the identities of different types of GABA has not been explored in depth.

The perturbation of *Lhx6* resulted in an increase in MSNs and interneurons. In concordance with the proportion analysis, PN-related genes like *Drd1* and *Foxp1* as well as PN-related modules like Module 4 are upregulated in *Lhx6* perturbed inhibitory neurons. Interestingly, perturbation of *Tcf4* had little effect on cell proportions, but resulted in reduced expression of modules 6 and 4.

Tcf4's modest effect on inhibitory neuron specification is surprising, given that it is linked to a number of severe developmental disorders such as Pitt-Hopkins syndrome and has a broad expression in all cortical and subcortical structures in the developing mouse brain (Jung et al. 2018). Papes et al. (2022) showed that Tcf4 reduced proliferation and impaired the ability to differentiate into neurons in the human cell line. However, mouse models carrying Tcf4 mutations in the clinically relevant heterozygous state exhibit only mild phenotypes, without the severe symptoms observed in patients (Rannals et al. 2016; Badowska et al. 2020). This might explain the mild phenotype we observe in our own P7 t-CROP seq dataset.

LHX6 and MEIS2 may interact antagonistically. Asgarian et al. (2022) explored this hypoth-

esis when they found that PN marker genes, such as *Meis2*, *Pbx3* and *Foxp1*, were up-regulated in *Lhx6* knockout cells collected from the cortex. The same genes were found to be up-regulated in our DEA of *Lhx6* perturbed inhibitory neurons, corroborating their results. A similar study found that conditional KO of NKX2-1, a TF that acts upstream of LHX6, resulted in upregulation of *Meis2* in the SVZ of the MGE (Sandberg et al. 2016). Furthermore, they observed an enrichment of motifs that match the binding sites of MEIS2 (Sandberg et al. 2018). Our findings and others point to *Meis2* and *Lhx6* as key factors that promote opposing inhibitory fates, interneurons vs projection neurons. The absence of *Meis2* gene expression in SVZ and mantel zone could be explained by the possible suppression of Meis2 regulatory elements by LHX6.

In order to definitively prove this competitive relationship between LHX6 and MEIS2, pull down experiments that show interaction between these two TFs could be more conclusive, especially if samples dissected from the MGE show that LHX6 and MEIS2 are interaction partners.

In conclusion, TCF4 may still be crucial for the development of GABAergic neurons, but less so for differentiation. Our results and some literature indicate that it may play a bigger role in earlier neurodevelopmental processes, such as neurogenesis. LHX6 promotes IN fate, and it may do so by either interacting antagnostically with MEIS2 or through the binding of MEIS2 regulatory elements.

Conclusion and Outlook

6.1 Conclusion and Outlook

A wealth of inhibitory cell types are generated in the forebrain during neurogenesis. The intrinsic mechanisms responsible for this phenomenon are still not very well understood due to technological or biological constraints. The long migratory routes undertaken by inhibitory neurons adds another challenge, since that made them difficult to track using traditional methods like time-lapse imaging. This thesis addresses this topic first from the perspective of lineage and TF control of neuronal specification. To overcome this, we developed a transposon-based, lineage barcoding approach that is compatible with the 10x commercial platform.

Building on other DNA-barcoding based methods (Golden, Fields-Berry, and Cepko 1995; Cepko et al. 2000), a highly diverse DNA barcode library was constructed using Gibson Assembly. The library can be targeted into spatially defined regions using *in utero* electroporation. Compared to retrovirus lineage-tracing methods, it has high recovery rate. Using TrackerSeq enables us to retrieve thousands of cells' transcriptome along with their developmental histories. Among these thousands of cells, hundreds of sibling cells could be identified. However, TrackerSeq may be limited in some aspects, retrieving the barcode at later time points may require more replicates.

Previous lineage-tracing methods limited their inquiry to how lineage affected the spatial location of interneurons. TrackerSeq made it possible to investigate how lineage specifies the fate of inhibitory neurons. We find that the majority of clonal cells enter the same trajectory, but a third of the clones enter into different trajectories. Progenitor cells in the GE are capable of producing daughter cells that traverse different developmental trajectories during peak neurogenesis, suggesting that clonal divergence into different GAABergic precursor states is

initiated at the level of mitotic progenitor cells.

We then approached the question of intrinsic specification from the perspective of TF control of GABAergic neuronal differentiation. Perturbing Meis2 in inhibitory neurons lead to a switch to IN fate. GABAergic PN neurons over expressed IN-related genes, and this was also observed in mitotic cells. When we tag *Meis2* perturbed cells with TrackerSeq barcodes, we observed that mitotic cells preferentially differentiated into interneurons as opposed to projection neurons, indicating that perturbing Meis2 lead to fate switching Perturbing Lhx6 lead to an upregulation of PN-related genes and modules, while little changes were observed in glial clusters, indicating that Lhx6 is only important for inhibitory neuron specification.

In conclusion, LHX6 and MEIS2 are TFs that are important for the specification of GABAergic IN and PNs respectively. Preliminary evidence, including the transcriptome data presented in this thesis, indicates that they may play antagonistic roles in the fate specification of inhibitory neurons. This thesis has provided evidence of the effectiveness of TrackerSeq as both a lineage tracing method and a tool that can be integrated with single-cell perturbation techniques to identify changes in cellular fate. Further investigation using TrackerSeq in combination with the hyperpolarization of GABAergic neurons may shed light on the impact of activity-dependent perturbation on the specification of inhibitory neurons.

Bibliography

Anderson, S. A., O. Marın, C. Horn, K. Jennings, and J. L. Rubenstein. 2001. "Distinct cortical migrations from the medial and lateral ganglionic eminences." [in eng]. *Development* (England) 128, no. 3 (February): 353–63.

Asgarian, Zeinab, Marcio Guiomar Oliveira, Agata Stryjewska, Ioannis Maragkos, Anna Noren Rubin, Lorenza Magno, Vassilis Pachnis, et al. 2022. "MTG8 interacts with LHX6 to specify cortical interneuron subtype identity." [in eng]. *Nature communications* (England) 13, no. 1 (September): 5217.

Badowska, D. M., M. M. Brzózka, N. Kannaiyan, C. Thomas, P. Dibaj, A. Chowdhury, H. Steffens, et al. 2020. "Modulation of cognition and neuronal plasticity in gain- and loss-of-function mouse models of the schizophrenia risk gene Tcf4." *Transl Psychiatry.* 10 (1): 343. ISSN: 2158-3188.

Bandler, Rachel C., Christian Mayer, and Gord Fishell. 2017. "Cortical interneuron specification: the juncture of genes, time and geometry." *Curr Opin Neurobiol* 42 (February): 17–24.

Bandler, Rachel C., Ilaria Vitali, Ryan N. Delgado, May C. Ho, Elena Dvoretskova, Josue S. Ibarra Molinas, Paul W. Frazel, et al. 2022. "Single-cell delineation of lineage and genetic identity in the mouse brain." *Nature* 601 (7893): 404–409. ISSN: 1476-4687.

Bandler, Rachel Cohen. 2019. "The Genetic Logic of Interneuron Differentiation." PhD diss., New York University, February.

Batista-Brito, Renata, Jennie Close, Robert Machold, and Gord Fishell. 2008. "The distinct temporal origins of olfactory bulb interneuron subtypes." [in eng]. *J. Neurosci.* (United States) 28, no. 15 (April): 3966–75.

Belmonte-Mateos, Carla, and Cristina Pujades. 2021. "From Cell States to Cell Fates: How Cell Proliferation and Neuronal Differentiation Are Coordinated During Embryonic Development." [in eng]. *Front. Neurosci.* (Switzerland) 15:781160.

Berenguer, Marie, and Gregg Duester. 2022. "Retinoic acid, RARs and early development." [in eng]. *J. Mol. Endocrinol.* (England) 69, no. 4 (November): T59–T67.

Biddy, Brent A., Wenjun Kong, Kenji Kamimoto, Chuner Guo, Sarah E. Waye, Tao Sun, and Samantha A. Morris. 2018. "Single-cell mapping of lineage and identity in direct reprogramming." [in eng]. *Nature* (England) 564, no. 7735 (December): 219–224.

Bock, Christoph, Paul Datlinger, Florence Chardon, Matthew A. Coelho, Matthew B. Dong, Keith A. Lawson, Tian Lu, et al. 2022. "High-content CRISPR screening." *Nat Rev Methods Primers* 2 (1): 8. ISSN: 2662-8449.

Bortone, Dante, and Franck Polleux. 2009. "KCC2 Expression Promotes the Termination of Cortical Interneuron Migration in a Voltage-Sensitive Calcium-Dependent Manner." *Neuron* 62:53–71. ISSN: 0896-6273.

Bowling, Sarah, Duluxan Sritharan, Fernando G. Osorio, Maximilian Nguyen, Priscilla Cheung, Alejo Rodriguez-Fraticelli, Sachin Patel, et al. 2020. "An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells." *Cell* 181, no. 6 (June): 1410–1422.e27.

Brian, Bushnell. 2014. "BBMap."

El-Brolosy, Mohamed A., and Didier Y. R. Stainier. 2017. "Genetic compensation: A phenomenon in search of mechanisms." [in eng]. *PLoS Genet.* (United States) 13, no. 7 (July): e1006780.

Brown, Keith N., She Chen, Zhi Han, Chun-Hui Lu, Xin Tan, Xin-Jun Zhang, Liya Ding, et al. 2011. "Clonal Production and Organization of Inhibitory Interneurons in the Neocortex." *Science* 334, no. 6055 (October): 480.

Buschmann, Tilo, and Leonid V. Bystrykh. 2013. "Levenshtein error-correcting barcodes for multiplexed DNA sequencing." *BMC Bioinformatics* 14 (1): 272. ISSN: 1471-2105.

Butt, Simon J. B., Marc Fuccillo, Susana Nery, Steven Noctor, Arnold Kriegstein, Joshua G. Corbin, and Gord Fishell. 2005. "The temporal and spatial origins of cortical interneurons predict their physiological subtype." *Neuron* 48, no. 4 (November): 591–604.

Caputi, Antonio, Sarah Melzer, Magdalena Michael, and Hannah Monyer. 2013. "The long and short of GABAergic neurons." *Curr. Opin. Neurobiol.* (84 THEOBALDS RD, LONDON WC1X 8RR, ENGLAND) 23, no. 2 (April): 179–186. ISSN: 0959-4388.

Cepko, C. L., E. Ryder, C. Austin, J. Golden, S. Fields-Berry, and J. Lin. 2000. "Lineage analysis with retroviral vectors." [in eng]. *Meth. Enzymol* (United States) 327:118–45.

Chatzi, Christina, Thomas Brade, and Gregg Duester. 2011. "Retinoic acid functions as a key GABAergic differentiation signal in the basal ganglia." [in eng]. *PLoS Biol.* (United States) 9, no. 4 (April): e1000609.

Ciceri, Gabriele, Nathalie Dehorter, Ignasi Sols, Z. Josh Huang, Miguel Maravall, and Oscar Marın. 2013. "Lineage-specific laminar organization of cortical GABAergic interneurons." [in eng]. *Nat. Neurosci.* (United States) 16, no. 9 (September): 1199–210.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex genome engineering using CRISPR/Cas systems." [in eng]. *Science* (United States) 339, no. 6121 (February): 819–23.

Corbin, Joshua G., and Simon J. B. Butt. 2011. "Developmental mechanisms for the generation of telencephalic interneurons." *Dev Neurobiol* 71, no. 8 (August): 710–32.

Cox, David Benjamin Turitz, Randall Jeffrey Platt, and Feng Zhang. 2015. "Therapeutic genome editing: prospects and challenges." *Nat. Med.* 21 (2): 121–131. ISSN: 1546-170X.

Datlinger, Paul, André F. Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C. Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. 2017. "Pooled CRISPR screening with single-cell transcriptome readout." *Nat. Methods* 14, no. 3 (March): 297–301.

Denaxa, Myrto, Melanie Kalaitzidou, Anna Garefalaki, Angeliki Achimastou, Reena Lasrado, Tamara Maes, and Vassilis Pachnis. 2012. "Maturation-Promoting Activity of SATB1 in MGE-Derived Cortical Interneurons." *Cell Rep.* 2:1351–1362. ISSN: 2211-1247.

DeTomaso, David, and Nir Yosef. 2021. "Hotspot identifies informative gene modules across modalities of single-cell genomics." [in eng]. *Cell Syst.* (United States) 12, no. 5 (May): 446–456.e9.

Ding, Sheng, Xiaohui Wu, Gang Li, Min Han, Yuan Zhuang, and Tian Xu. 2005. "Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice." [in eng]. *Cell* (United States) 122, no. 3 (August): 473–83.

Flames, Nuria, Ramón Pla, Diego M. Gelman, John L. R. Rubenstein, Luis Puelles, and Oscar Marın. 2007. "Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes." *J Neurosci* 27, no. 36 (September): 9682–95.

Fleck, Jonas Simon, Sophie Martina Johanna Jansen, Damian Wollny, Fides Zenk, Makiko Seimiya, Akanksha Jain, Ryoko Okamoto, et al. 2022. "Inferring and perturbing cell fate regulomes in human brain organoids." *Nature,* ISSN: 1476-4687.

Garcia, Miguel Turrero, and Corey C. Harwell. 2017. "Radial glia in the ventral telencephalon." *FEBS Lett.* (111 RIVER ST, HOBOKEN 07030-5774, NJ USA) 591, no. 24 (December): 3942–3959. ISSN: 1873-3468.

Garcıa, Natalia V. De Marco, Theofanis Karayannis, and Gord Fishell. 2011. "Neuronal activity is required for the development of specific cortical interneuron subtypes." *Nature* 472:351–355. ISSN: 0028-0836.

Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. 2009. "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nat. Methods* 6 (5): 343–345. ISSN: 1548-7105.

Golden, J. A., S. C. Fields-Berry, and C. L. Cepko. 1995. "Construction and characterization of a highly complex retroviral library for lineage analysis." [in eng]. *Proceedings of the National Academy of Sciences of the United States of America* (United States) 92, no. 12 (June): 5704–8.

Guillemot, François, and Céline Zimmer. 2011. "From Cradle to Grave: The Multiple Roles of Fibroblast Growth Factors in Neural Development." *Neuron* 71:574–588. ISSN: 0896-6273.

Hafemeister, Christoph, and Rahul Satija. 2019. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression." *Genome Biol.* 20. ISSN: 1474-760X.

Halliday, A. L., and C. L. Cepko. 1992. "Generation and migration of cells in the developing striatum." [in eng]. *Neuron* (United States) 9, no. 1 (July): 15–26.

Harwell, Corey C., Luis C. Fuentealba, Adrian Gonzalez-Cerrillo, Phillip R. L. Parker, Caitlyn C. Gertz, Emanuele Mazzola, Miguel Turrero Garcia, Arturo Alvarez-Buylla, Constance L. Cepko, and Arnold R. Kriegstein. 2015. "Wide Dispersion and Diversity of Clonally Related Inhibitory Interneurons." *Neuron* 87:999–1007. ISSN: 0896-6273.

Hébert, Jean M., and Gord Fishell. 2008. "The genetics of early telencephalon patterning: some assembly required." *Nat. Rev. Neurosci.* 9:678–685. ISSN: 1471-003X.

Hollville, Emilie, Selena E. Romero, and Mohanish Deshmukh. 2019. "Apoptotic cell death regulation in neurons." [in eng]. *FEBS J* (England) 286, no. 17 (September): 3276–3298.

Humphries, Mark D., and Tony J. Prescott. 2010. "The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward." [in eng]. *Prog. Neurobiol.* (England) 90, no. 4 (April): 385–417.

Incontro, Salvatore, Cedric S. Asensio, Robert H. Edwards, and Roger A. Nicoll. 2015. "Efficient, Complete Deletion of Synaptic Proteins using CRISPR." *Neuron* 88:617. ISSN: 0896-6273.

Jin, Xin, Sean K. Simmons, Amy Guo, Ashwin S. Shetty, Michelle Ko, Lan Nguyen, Vahbiz Jokhi, et al. 2020. "In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes." [in eng]. *Science* (United States) 370, no. 6520 (November).

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." [in eng]. *Science* (United States) 337, no. 6096 (August): 816–21.

Jung, Matthias, Benjamin M. Häberle, Tristan Tschaikowsky, Marie-Theres Wittmann, Elli-Anna Balta, Vivien-Charlott Stadler, Christiane Zweier, Arnd Dörfler, Christian Johannes Gloeckner, and D. Chichung Lie. 2018. "Analysis of the expression pattern of the schizophrenia-risk and intellectual disability gene TCF4 in the developing and adult brain suggests a role in development and plasticity of cortical and hippocampal neurons." [in eng]. *Mol. Autism* (England) 9:20.

Kalhor, Reza, Kian Kalhor, Leo Mejia, Kathleen Leeper, Amanda Graveline, Prashant Mali, and George M. Church. 2018. "Developmental barcoding of whole mouse via homing CRISPR." *Science* 361, no. 6405 (August).

Kepecs, Adam, and Gordon Fishell. 2014. "Interneuron cell types are fit to function." *Nature* 505, no. 7483 (January): 318–26.

Kester, Lennart, and Alexander van Oudenaarden. 2018. "Single-Cell Transcriptomics Meets Lineage Tracing." [in eng]. *Cell Stem Cell* (United States) 23, no. 2 (August): 166–179.

Kim, Hyojin, Noah C. Berens, Nicole E. Ochandarena, and Benjamin D. Philpot. 2020. "Region and Cell Type Distribution of TCF4 in the Postnatal Mouse Brain." [in eng]. *Frontiers in neuroanatomy* (Switzerland) 14:42.

Knowles, Rhys, Nathalie Dehorter, and Tommas Ellender. 2021. "From Progenitors to Progeny: Shaping Striatal Circuit Development and Function." [in eng]. *J. Neurosci.* (United States) 41, no. 46 (November): 9483–9502.

Kong, Wenjun, Brent A. Biddy, Kenji Kamimoto, Junedh M. Amrute, Emily G. Butka, and Samantha A. Morris. 2020. "CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution." *Nat. Protoc.* 15 (3): 750–772. ISSN: 1750-2799.

Korsunsky, Ilya, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. 2019. "Fast, sensitive and accurate integration of single-cell data with Harmony." *Nat. Methods* 16 (12): 1289–1296. ISSN: 1548-7105.

Kreitzer, Anatol C., and Robert C. Malenka. 2008. "Striatal plasticity and basal ganglia circuit function." [in eng]. *Neuron* (United States) 60, no. 4 (November): 543–54.

Kuerbitz, Jeffrey, Melinda Arnett, Sarah Ehrman, Michael T. Williams, Charles V. Vorhees, Simon E. Fisher, Alistair N. Garratt, Louis J. Muglia, Ronald R. Waclaw, and Kenneth Campbell. 2018. "Loss of Intercalated Cells (ITCs) in the Mouse Amygdala of Tshz1 Mutants Correlates with Fear, Depression, and Social Interaction Phenotypes." [in eng]. *J. Neurosci.* (United States) 38, no. 5 (January): 1160–1177.

Kuleshov, Maxim V., Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, et al. 2016. "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update." [in eng]. *Nucleic Acids Res.* (England) 44, no. W1 (July): W90–7.

La Manno, Gioele, Kimberly Siletti, Alessandro Furlan, Daniel Gyllborg, Elin Vinsland, Alejandro Mossi Albiach, Christoffer Mattsson Langseth, et al. 2021. "Molecular architecture of the developing mouse brain." *Nature* 596 (7870): 92–96. ISSN: 1476-4687.

LeDoux, Joseph. 2012. "Rethinking the Emotional Brain." *Neuron* 73:653–676. ISSN: 0896-6273.

Leung, Ryan F., Ankita M. George, Enola M. Roussel, Maree C. Faux, Jeffrey T. Wigle, and David D. Eisenstat. 2022. "Genetic Regulation of Vertebrate Forebrain Development by Homeobox Genes." [in eng]. *Front. Neurosci.* (Switzerland) 16:843794.

Li, Ye, Hui Lu, Pei-Lin Cheng, Shaoyu Ge, Huatai Xu, Song-Hai Shi, and Yang Dan. 2012. "Clonally related visual cortical neurons show similar stimulus feature selectivity." *Nature* 486, no. 7401 (June): 118–121.

Lim, Lynette, Da Mi, Alfredo Llorca, and Oscar Marın. 2018. "Development and Functional Diversification of Cortical Interneurons." [in eng]. *Neuron* (United States) 100, no. 2 (October): 294–313.

Lodato, Simona, and Paola Arlotta. 2015. "Generating Neuronal Diversity in the Mammalian Cerebral Cortex." *Annu. Rev. Cell Dev. Biol.* 31:699–720. ISSN: 1081-0706.

Loo, Mark P.J. van der. 2014. "The stringdist Package for Approximate String Matching." *The R Journal* 6 (1): 111–122.

Lyons, Eli, Paul Sheridan, Georg Tremmel, Satoru Miyano, and Sumio Sugano. 2017. "Large-scale DNA Barcode Library Generation for Biomolecule Identification in High-throughput Screens." [in eng]. *Sci. Rep.* (England) 7, no. 1 (October): 13899.

Ma, Huan, Samuel Cohen, Boxing Li, and Richard W. Tsien. 2012. "Exploring the dominant role of Cav1 channels in signalling to the nucleus." [in eng]. *Biosci Rep.* (England) 33, no. 1 (December): 97–101.

Ma, Jian, Zhongfu Shen, Yong-Chun Yu, and Song-Hai Shi. 2018. "Neural lineage tracing in the mammalian brain." [in eng]. *Curr. Opin. Neurobiol.* (England) 50 (June): 7–16.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church. 2013. "RNA-Guided Human Genome Engineering via Cas9." *Science* 339:823–826. ISSN: 0036-8075.

Marın, Oscar, and John L. R. Rubenstein. 2001. "A long, remarkable journey: Tangential migration in the telencephalon." *Nat. Rev. Neurosci.* 2:780–790. ISSN: 1471-003X.

Mayer, Christian, Christoph Hafemeister, Rachel C. Bandler, Robert Machold, Renata Batista Brito, Xavier Jaglin, Kathryn Allaway, Andrew Butler, Gord Fishell, and Rahul Satija. 2018. "Developmental diversification of cortical inhibitory interneurons." *Nature* 555 (7697): 457–462.

Mayer, Christian, Xavier H. Jaglin, Lucy V. Cobbs, Rachel C. Bandler, Carmen Streicher, Constance L. Cepko, Simon Hippenmeyer, and Gord Fishell. 2015. "Clonally Related Forebrain Interneurons Disperse Broadly across Both Functional Areas and Structural Boundaries." *Neuron* 87:989–998. ISSN: 0896-6273.

McCarthy, M., D. H. Turnbull, C. A. Walsh, and G. Fishell. 2001. "Telencephalic neural progenitors appear to be restricted to regional and glial fates before the onset of neurogenesis." [in eng]. *J. Neurosci.* (United States) 21, no. 17 (September): 6772–81.

McGinnis, Christopher S., Lyndsay M. Murrow, and Zev J. Gartner. 2019. "DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors." [in eng]. *Cell Syst.* (United States) 8, no. 4 (April): 329–337.e4.

McKenna, Aaron, Gregory M. Findlay, James A. Gagnon, Marshall S. Horwitz, Alexander F. Schier, and Jay Shendure. 2016. "Whole-organism lineage tracing by combinatorial and cumulative genome editing." *Science* 353, no. 6298 (July): aaf7907.

Megason, Sean G., and Andrew P. McMahon. 2002. "A mitogen gradient of dorsal midline Wnts organizes growth in the CNS." [in eng]. *Development* (England) 129, no. 9 (May): 2087–98.

Mesman, Simone, Reinier Bakker, and Marten P. Smidt. 2020. "Tcf4 is required for correct brain development during embryogenesis." [in eng]. *Mol. Cell. Neurosci.* (United States) 106 (July): 103502.

Miyoshi, Goichi, Simon J. B. Butt, Hirohide Takebayashi, and Gord Fishell. 2007. "Physiologically distinct temporal cohorts of cortical interneurons arise from telencephalic Olig2-expressing precursors." *J Neurosci* 27, no. 29 (July): 7786–98.

Miyoshi, Goichi, Robert P. Machold, and Gord Fishell. 2013. "Specification of GABAergic Neocortical Interneurons." In *Cortical Development: Neural Diversity and Neocortical Organization,* edited by Ryoichiro Kageyama and Tetsuo Yamamori, 89–126. Tokyo: Springer Japan. ISBN: 978-4-431-54496-8.

Miyoshi, Goichi, Allison Young, Timothy Petros, Theofanis Karayannis, Melissa McKenzie Chang, Alfonso Lavado, Tomohiko Iwano, et al. 2015. "Prox1Regulates the Subtype-Specific Development of Caudal Ganglionic Eminence-Derived GABAergic Cortical Interneurons." *J Neurosci* 35:12869–12889. ISSN: 0270-6474.

Nishiyama, Jun. 2019. "Genome editing in the mammalian brain using the CRISPR–Cas system." *Neurosci Res* 141:4–12. ISSN: 0168-0102.

Noctor, Stephen C., Alexander C. Flint, Tamily A. Weissman, Ryan S. Dammerman, and Arnold R. Kriegstein. 2001. "Neurons derived from radial glial cells establish radial units in neocortex." *Nature* 409, no. 6821 (February): 714–720.

Papes, Fabio, Antonio P. Camargo, Janaina S. de Souza, Vinicius M. A. Carvalho, Ryan A. Szeto, Erin LaMontagne, José R. Teixeira, et al. 2022. "Transcription Factor 4 loss-of-function is associated with deficits in progenitor proliferation and cortical neuron content." *Nat. Commun.* 13 (1): 2387. ISSN: 2041-1723.

Pensold, Daniel. 2017. "Single cell transcriptomics reveal regulators of progenitor cell fate and postmitotic maturation during brain development." PhD diss., Friedrich-Schiller-Universit{ä}t Jena.

Petryniak, Magdalena A., Gregory B. Potter, David H. Rowitch, and John L. R. Rubenstein. 2007. "Dlx1 and Dlx2 Control Neuronal versus Oligodendroglial Cell Fate Acquisition in the Developing Forebrain." *Neuron* 55:417–433. ISSN: 0896-6273.

Platt, Randall J., Sidi Chen, Yang Zhou, Michael J. Yim, Lukasz Swiech, Hannah R. Kempton, James E. Dahlman, et al. 2014. "CRISPR-Cas9 knockin mice for genome editing and cancer modeling." [in eng]. *Cell* (United States) 159, no. 2 (October): 440–55.

Raj, Bushra, Daniel E. Wagner, Aaron McKenna, Shristi Pandey, Allon M. Klein, Jay Shendure, James A. Gagnon, and Alexander F. Schier. 2018. "Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain." *Nat. Biotechnol.* 36 (5): 442–450. ISSN: 1546-1696.

Rannals, Matthew D., Gregory R. Hamersky, Stephanie Cerceo Page, Morganne N. Campbell, Aaron Briley, Ryan A. Gallo, BaDoi N. Phan, et al. 2016. "Psychiatric Risk Gene Transcription Factor 4 Regulates Intrinsic Excitability of Prefrontal Neurons via Repression of SCN10a and KCNQ1." [in eng]. *Neuron* (United States) 90, no. 1 (April): 43–55.

Replogle, Joseph M., Thomas M. Norman, Albert Xu, Jeffrey A. Hussmann, Jin Chen, J. Zachery Cogan, Elliott J. Meer, et al. 2020. "Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing." *Nat. Biotechnol.* 38 (8): 954–961. ISSN: 1546-1696.

Sandberg, Magnus, Pierre Flandin, Shanni Silberberg, Linda Su-Feher, James D. Price, Jia Sheng Hu, Carol Kim, Axel Visel, Alex S. Nord, and John L. R. Rubenstein. 2016. "Transcriptional Networks Controlled by NKX2-1 in the Development of Forebrain GABAergic Neurons." [in eng]. *Neuron* (United States) 91, no. 6 (September): 1260–1275.

Sandberg, Magnus, Leila Taher, Jianxin Hu, Brian L. Black, Alex S. Nord, and John L. R. Rubenstein. 2018. "Genomic analysis of transcriptional networks directing progression of cell states during MGE development." *Neural Dev.* 13 (1): 21. ISSN: 1749-8104.

Shang, Zicong, Lin Yang, Ziwu Wang, Yu Tian, Yanjing Gao, Zihao Su, Rongliang Guo, et al. 2022. "The transcription factor Zfp503 promotes the D1 MSN identity and represses the D2 MSN identity." [in eng]. *Front. Cell Dev. Biol.* (Switzerland) 10:948331.

Simeonov, Kamen P., China N. Byrns, Megan L. Clark, Robert J. Norgard, Beth Martin, Ben Z. Stanger, Jay Shendure, Aaron McKenna, and Christopher J. Lengner. 2021. "Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states." *Cancer Cell.* 39:1150–1162.e9. ISSN: 1535-6108.

Song, Xiaolei, Haotian Chen, Zicong Shang, Heng Du, Zhenmeiyu Li, Yan Wen, Guoping Liu, et al. 2021. "Homeobox Gene Six3 is Required for the Differentiation of D2-Type Medium Spiny Neurons." [in eng]. *Neurosci Bull.* (Singapore) 37, no. 7 (July): 985–998.

Squair, Jordan W., Matthieu Gautier, Claudia Kathe, Mark A. Anderson, Nicholas D. James, Thomas H. Hutson, Rémi Hudelle, et al. 2021. "Confronting false discoveries in single-cell differential expression." *Nat. Commun.* 12 (1): 5692. ISSN: 2041-1723.

Straub, Christoph, Adam J. Granger, Jessica L. Saulnier, and Bernardo L. Sabatini. 2014. "CRISPR/Cas9-mediated gene knock-down in post-mitotic neurons." [in eng]. *PloS one* (United States) 9 (8): e105584.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. 3rd Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. "Comprehensive Integration of Single-Cell Data." [in eng]. *Cell* (United States) 177, no. 7 (June): 1888–1902.e21.

Su, Zihao, Ziwu Wang, Susan Lindtner, Lin Yang, Zicong Shang, Yu Tian, Rongliang Guo, et al. 2022. "Dlx1/2-dependent expression of Meis2 promotes neuronal fate determination in the mammalian striatum." *Development* 149, no. 4 (February).

Sultan, Khadeejah T., and Song-Hai Shi. 2018. "Generation of diverse cortical inhibitory interneurons." *Wiley Interdiscip Rev Dev Biol* 7, no. 2 (March).

Sussel, Lori, Oscar Marin, Shioko Kimura, and John L. R. Rubenstein. 1999. "Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum." *Development* 126:3359–3370. ISSN: 0950-1991.

Swindle, C. Scott, and Christopher A. Klug. 2002. "Mechanisms that regulate silencing of gene expression from retroviral vectors." [in eng]. *J Hematother Stem Cell Res* (United States) 11, no. 3 (June): 449–56.

Tasic, Bosiljka, Zizhen Yao, Lucas T. Graybuck, Kimberly A. Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, et al. 2018. "Shared and distinct transcriptomic cell types across neocortical areas." [in eng]. *Nature* (England) 563, no. 7729 (November): 72–78.

Teixeira, José R., Ryan A. Szeto, Vinicius M. A. Carvalho, Alysson R. Muotri, and Fabio Papes. 2021. "Transcription factor 4 and its association with psychiatric disorders." [in eng]. *Transl. Psychiatry* (United States) 11, no. 1 (January): 19.

Turner, D. L., and C. L. Cepko. 1987. "A common progenitor for neurons and glia persists in rat retina late in development." [in eng]. *Nature* (England) 328, no. 6126 (July): 131–6.

VanHorn, Sadie, and Samantha A. Morris. 2021. "Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development." *Dev Cell* 56, no. 1 (January): 7–21.

Vogt, Daniel, Robert F. Hunt, Shyamali Mandal, Magnus Sandberg, Shanni N. Silberberg, Takashi Nagasawa, Zhengang Yang, Scott C. Baraban, and John L. R. Rubenstein. 2014. "Lhx6 Directly Regulates Arx and CXCR7 to Determine Cortical Interneuron Fate and Laminar Position." *Neuron* 82:350–364. ISSN: 0896-6273.

Waclaw, Ronald R., Bei Wang, Zhenglei Pei, Lisa A. Ehrman, and Kenneth Campbell. 2009. "Distinct Temporal Requirements for the Homeobox Gene Gsx2 in Specifying Striatal and Olfactory Bulb Neuronal Fates." *Neuron* 63:451–465. ISSN: 0896-6273.

Waddington, Conrad Hal. 2014. *The strategy of the genes.* Routledge.

Wagner, Daniel E., and Allon M. Klein. 2020. "Lineage tracing meets single-cell omics: opportunities and challenges." *Nat. Rev. Genet.* 21, no. 7 (July): 410–427.

Wagner, Daniel E., Caleb Weinreb, Zach M. Collins, James A. Briggs, Sean G. Megason, and Allon M. Klein. 2018. "Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo." [in eng]. *Science* (United States) 360, no. 6392 (June): 981–987.

Wamsley, Brie, and Gord Fishell. 2017. "Genetic and activity-dependent mechanisms underlying interneuron diversity." *Nat. Rev. Neurosci.* 18:299–309. ISSN: 1471-003X.

Wei, Song, Heng Du, Zhenmeiyu Li, Guangxu Tao, Zhejun Xu, Xiaolei Song, Zicong Shang, et al. 2019. "Transcription factors Sp8 and Sp9 regulate the development of caudal ganglionic eminence-derived cortical interneurons." [in eng]. *J Comp Neurol* (United States) 527, no. 17 (December): 2860–2874.

Weinreb, Caleb, Samuel Wolock, Betsabeh K. Tusi, Merav Socolovsky, and Allon M. Klein. 2018. "Fundamental limits on dynamic inference from single-cell snapshots." [in eng]. *Proc. Natl. Acad. Sci. U. S. A.* (United States) 115, no. 10 (March): E2467–E2476.

White, J. G., ed. 1986. *The structure of the nervous system of the nematode Caenorhabditis elegans.* Philosophical transactions of the Royal Society of London. London.

Wonders, Carl P., Lauren Taylor, Jelle Welagen, Ihunanya C. Mbata, Jenny Z. Xiang, and Stewart A. Anderson. 2008. "A spatial bias for the origins of interneuron subgroups within the medial ganglionic eminence." [in eng]. *Dev. Biol.* (United States) 314, no. 1 (February): 127–36.

Wu, Zhijian, Hongyan Yang, and Peter Colosi. 2010. "Effect of Genome Size on AAV Vector Packaging." *Mol. Ther.* 18:80–86. ISSN: 1525-0016.

Xu, Bing, Sarah Kucenas, and Hui Zong. 2022. "zMADM (zebrafish mosaic analysis with double markers) for single-cell gene knockout and dual-lineage tracing." [in eng]. *Proc. Natl. Acad. Sci. U. S. A.* (United States) 119, no. 9 (March).

Xu, Qing, Inma Cobos, Estanislao De La Cruz, John L. Rubenstein, and Stewart A. Anderson. 2004. "Origins of cortical interneuron subtypes." *J Neurosci* 24, no. 11 (March): 2612–22.

Yang, Dian, Matthew G. Jones, Santiago Naranjo, William M. Rideout, Kyung Hoi (Joseph) Min, Raymond Ho, Wei Wu, et al. 2022. "Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution." *Cell* 185:1905–1923.e25. ISSN: 0092-8674.

Zhao, Lu, Zhimin Liu, Sasha F. Levy, and Song Wu. 2017. "Bartender: a fast and accurate clustering algorithm to count barcode reads." *Bioinformatics* 34, no. 5 (October): 739–747. ISSN: 1367-4803.

Acknowledgement

Completing a PhD thesis is a challenging journey, and I am grateful to have had the unwavering support of my friends and family, who have been my pillars of strength throughout this process. To all of those who have contributed to this work, I offer my deepest thanks.

I would like to thank Christian for giving me the opportunity to pursue my graduate studies in his lab. Throughout my PhD, I had the privilege of contributing to the research community through technique development and was allowed to explore my interest in cutting edge biotechnology. Christian taught me to be meticulous and thorough in my approach to solving scientific questions, and always took care to give valuable feedback.

My sincere thanks to the staff of IMPRS for Molecular Life Sciences (IMPRS-LS) graduate school for providing me with the support, resources, and guidance necessary to pursue my academic and professional goals. I am very grateful for the lifelong friends I made through IMPRS-LS. I want to give a special shoutout to my comrade in arms Alexandra, Eduard, Chao, Songwei, Sagar, Taisha, Jakob, and Shuhan. They have been there for me through thick and thin, providing me with the encouragement, motivation, and laughter that has kept me going throughout this journey. From mid-day lunches, spontaneous after work hangouts, weekend adventures, you were always there for me. You were my home away from home, and I look forward to continuing our friendship for many years to come. Thank you for being such an important part of my life.

To my thesis committee members, Rüdiger, Magdalena, and Wolfgang, my deepest gratitude for their guidance, expertise, and patience throughout this process. Their mentorship has been invaluable.

Finally, I would like to express my gratitude to my parents. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

May Cheng-Ching Ho

EDUCATION

LMU Munich/International Max Planck Research School **Sept 2018-Feb 2023**
Ph.D., Neurobiology *Munich, Germany*

- Thesis title: Genetic Fate Specification of Inhibitory Neurons

Investigated lineage and fate commitment of inhibitory neurons through the development of a technique called TrackerSeq, a method compatible with the 10x platform that labels cells with a unique lineage barcode. Using this technique, the team discovered that progenitor cells in the GE can produce daughter cells traversing different developmental trajectories during peak neurogenesis.

University of British Columbia **Sept 2013-May 2018**
B.Sc., Microbiology and Immunology *Vancouver, Canada*

- Grade: 3.75 cGPA

Final year laboratory project, based in the Department of Microbiology & Immunology, where I investigated the sensitivity of *E. coli* to EDTA, a chelation agent used for heavy metal toxicity in a team.

Final year courses taken:
Developmental Neurobiology, Linear Algebra, Environmental Microbiology Laboratory, Microbial Ecological Genomics, Molecular Genetics, Advanced Immunology, Genetics and Cell Biology of Cancer

SCHOLARSHIPS & AWARDS

2019	**Best Poster Award 2nd Prize –** Was voted by peers to have the 2nd best poster at our bi-annual seminar.
2017	**Best Poster** – Awarded best poster by professors at a conference for undergraduate neuroscience research.
2017	**Clayton Person Memorial Scholarship in Biology** – The scholarship is awarded to the student having the highest standing in BIOL 234/335 (Genetics) and is made on the recommendation of the Department of Botany.
2016	**CBR Summer Studentship Award -** $5,000 award value to subsidize research at Aly Karsan

EMPLOYMENT

Single Cell Discoveries **May 2023 – Present**
Account Executive *Utrecht, NL*

- Build relationships with 100+ clients to understand their scientific needs resulting in hundreds of thousands worth of won projects, primarily in new territories.
- Thoroughly prepared for over 100+ technical meetings, including academic and prepared presentations for high-level executives in the biopharma space
- Attended 3 industry conferences and engaged in outbound e-mail prospecting to increase lead generation by 10%.
- As an Account Executive, I oversee territories in the US, Germany, and Nordics
    - Price services and drafting complex proposals in collaboration with the R&D team.
    - Prospecting clients using coordinated outbound cold e-mailing campaigns.
- Experienced with CRM systems such as SalesForce

Max Planck Institute for Biological Intelligence **Sept 2018 – Feb 2023**
PhD Research Fellow *Munich, Germany*

- Engage in hypothesis-driven research in the neurodevelopment field under the guidance of a supervisor.
- Developed a single-cell compatible, lineage-tracing method now used by 10+ researchers in the development field.
- Applied different statistical methods to analyze high-dimensional single-cell sequencing data.

University of Calgary **May 2017 – Dec 2017**
Research Assistant *Calgary, Canada*
- Using lentivirus-mediated CRISPR-Cas9 construct to knockout prion gene.
- Run western blots to validate the absence of the prion protein.

BC Cancer Agency **Jan 2016 – Aug 2016**
Research Assistant *Vancouver, Canada*
- Analyzing microarray data and annotating genes using software programs

PUBLICATIONS

- Dvoretskova, E., **Ho, M.C**., Kittke, V. *et al.* Spatial enhancer activation influences inhibitory neuron identity during mouse embryonic development. *Nat Neurosci* **27**, 862–872 (2024).
- Bandler, R.C., Vitali, I., Delgado, R.N. *et al.* Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).
- Walia, R., **Ho, C.C**., Lee, C. *et al.* Gene-edited murine cell lines for propagation of chronic wasting disease prions. *Sci Rep* **9**, 11151 (2019).

SKILLS & INTERESTS

- **Skills:** Sales processes; negotiations; R&D; solution selling; communication; business development; molecular biology; research; presenting; teamwork; strategy
- **Interests:** Photography; skiing; tennis; museums; cinema; piano