PNAS RESEARCH ARTICLE | DEVELOPMENTAL BIOLOGY OPEN ACCESS

Single cell enhancer activity distinguishes GABAergic and cholinergic lineages in embryonic mouse basal ganglia

Linda Su-Feher[a,b,1], Anna N. Rubin[c,1], Shanni N. Silberberg[c], Rinaldo Catta-Preta[a,b,2], Kenneth J. Lim[a,b,c], Athena R. Ypsilanti[c], Iva Zdilar[a,b], Christopher S. McGinnis[d], Gabriel L. McKinsey[e], Thomas E. Rubino Jr.[a,b], Michael J. Hawrylycz[f], Carol Thompson[f], Zev J. Gartner[d,g,h,i], Luis Puelles[j], Hongkui Zeng[f], John L. R. Rubenstein[c,3], and Alex S. Nord[a,b,3]

Edited by Ulrich Mueller, The Johns Hopkins University, Baltimore, MD; received June 4, 2021; accepted December 8, 2021 by Editorial Board Member Jeremy Nathans

Enhancers integrate transcription factor signaling pathways that drive cell fate specification in the developing brain. We paired enhancer labeling and single-cell RNA-sequencing (scRNA-seq) to delineate and distinguish specification of neuronal lineages in mouse medial, lateral, and caudal ganglionic eminences (MGE, LGE, and CGE) at embryonic day (E)11.5. We show that scRNA-seq clustering using transcription factors improves resolution of regional and developmental populations, and that enhancer activities identify specific and overlapping GE-derived neuronal populations. First, we mapped the activities of seven evolutionarily conserved brain enhancers at single-cell resolution in vivo, finding that the selected enhancers had diverse activities in specific progenitor and neuronal populations across the GEs. We then applied enhancer-based labeling, scRNA-seq, and analysis of in situ hybridization data to distinguish transcriptionally distinct and spatially defined subtypes of MGE-derived GABAergic and cholinergic projection neurons and interneurons. Our results map developmental origins and specification paths underlying neurogenesis in the embryonic basal ganglia and showcase the power of scRNA-seq combined with enhancer-based labeling to resolve the complex paths of neuronal specification underlying mouse brain development.

genetics | neuroscience | development | enhancer | neurogenesis

Significance

During brain development, neurons are generated by spatially and temporally distinct processes that remain to be fully characterized. The ganglionic eminences (GEs) in the embryonic subpallium give rise to GABAergic and cholinergic neuron lineages that form the basal ganglia or migrate to the cerebral cortex. Beyond a limited set of canonical RNA markers, the transcriptional states of GE progenitors and immature neurons cells remain poorly defined. We combine enhancer labeling, single-cell transcriptomics using transcription factor-anchored clustering, and integration with in situ hybridization data to distinguish emerging neuronal populations in embryonic mouse basal ganglia. Our results demonstrate the specificity of enhancer-based labeling at single-cell resolution and reveal developmental origins and specification processes of critical neuronal lineages.

During brain development, transcriptional programs governed by the genomic interplay of transcription factors and *cis*-regulatory enhancer and promoter sequences drive the proliferation and specification of neuronal and glial lineages (1, 2). The embryonic basal ganglia (BG) include spatially distinct proliferative zones of the ganglionic eminences (GEs), which include the medial, lateral, and caudal GEs (MGE, LGE, and CGE) (3). Progenitor cells in the ventricular (VZ) and subventricular (SVZ) domains in the GEs give rise to GABAergic projection neurons and cholinergic neurons (4) that form the ventral pallidum, globus pallidus (5, 6), and striatal structures (3) that make up the mature BG. In addition, the GEs generate interneurons (INs) that populate the cortex, striatum, olfactory bulb, and other brain regions (7–10).

Building on bulk transcriptomics and in situ hybridization studies (ISH), single-cell RNA-sequencing (scRNA-seq) has been applied to embryonic mouse BG, revealing generalized progenitor populations and early born GABAergic lineages, with a focus on cortical interneuron (CIN) specification (11, 12). While CINs are one major output of embryonic BG, single-cell characterization of GABAergic and cholinergic as well as early-born CIN lineages that arise in the BG remains largely unexplored. Resolving the early stages of neurogenesis via scRNA-seq and ISH of the various neuronal classes that originate in the BG has been limited by major barriers: lack of region- and lineage-specific single gene markers, similarity of early transcriptional programs, and regional organization of BG neurogenesis that is missed by unguided scRNA-seq analysis.

Fate mapping represents a powerful tool to resolve the complexity of neurogenesis in vivo (8). Notably, enhancers drive highly specific transcription pattens, including in developing telencephalon (13). We previously demonstrated the utility of enhancer-driven transgenic reporter mouse lines for fate mapping and genetic manipulation of neuronal populations originating in embryonic BG and cortex (14, 15). Fate mapping with these enhancer-driven CreER^{T2}-GFP mice demonstrated that lineages marked by early developmental enhancer activity produce varied mature neuron populations. More broadly, enhancer-based expression constructs are emerging as a powerful tool for cell-type identification, enrichment, and modulation in the brain, and knowledge of sensitivity and specificity of enhancer-driven expression at single-cell resolution is needed.

We applied the strategy of pairing enhancer-based transgenic reporter mouse lines with scRNA-seq and immunohistochemistry (IHC) to define enhancer-labeled

Competing interest statement: J.L.R.R. is cofounder, stockholder, and currently on the scientific board of Neurona, a company studying the potential therapeutic use of interneuron transplantation.

This article is a PNAS Direct Submission. U.M. is a guest editor invited by the Editorial Board.

[1]L.S.-F. and A.N.R. contributed equally to this work.

[2]Present address: Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115.

[3]To whom correspondence may be addressed. Email: john.rubenstein@ucsf.edu or asnord@ucdavis.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2108760119/-/DCSupplemental.

Published April 4, 2022.

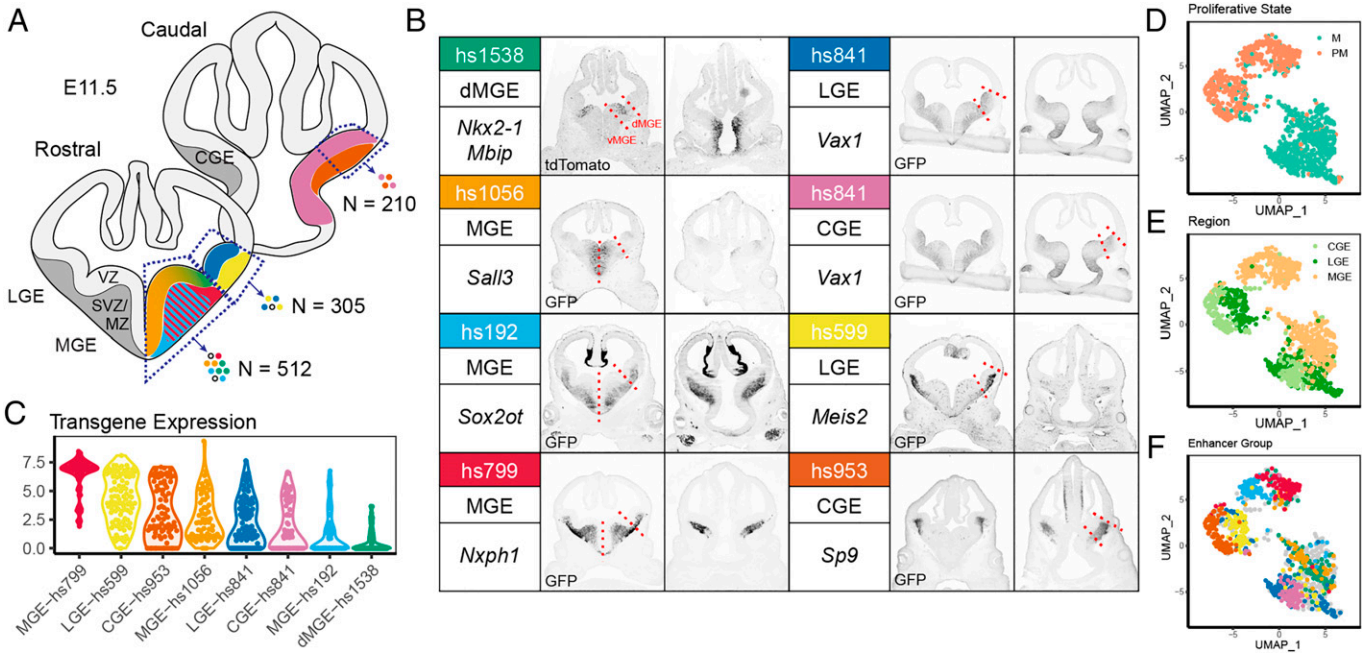

Fig. 1. Profiling enhancer-labeled single cells in E11.5 BG. (*A*) Schematic of dissection, with colors representing the activity of seven transgenic enhancer reporters characterized using C1 scRNA-seq. (*B*) Summary of the enhancers profiled, including dissection region, putative target gene, and representative GFP IHC imaging of enhancer transgenic reporters at E11.5, depicting activity within the ganglionic eminences. Red lines indicate microdissection boundaries. GFP IHC images reprinted with permission (15). (*C*) Violin plot of normalized transgene expression by enhancer. (*D*) Visualization of single cells by UMAP, colored by mitotic state (green: M, mitotic; orange: PM, postmitotic). (*E*) Visualization of single cells by UMAP, colored by region of dissection (light orange: MGE; light green: CGE; dark green: LGE). (*F*) Visualization of single cells by UMAP, colored by enhancer. Colors correspond to header colors in *B*. Enhancer-negative cells are depicted in gray.

BG-derived neuronal lineages at single-cell resolution. We profiled FACS-purified enhancer-positive cells as well as unsorted or enhancer-negative cells to define enhancer activity in the context of embryonic BG populations. We used a transcription factor-anchored approach to derive cell clusters in our scRNA-seq data, which improved developmental relevance of identified clusters. Our study identified proliferative and postmitotic cells that are distinctly labeled by enhancers active in MGE, LGE, and CGE, and revealed specification paths of enhancer-labeled and spatially defined populations of early BG-derived GABAergic and cholinergic lineages.

Results

Comparative Activity of Seven Enhancers in Embryonic Day 11.5 BG via scRNA-seq. To understand developmental origins of BG-arising neuronal populations, we profiled enhancer-labeled cells from embryonic day (E)11.5 MGE, LGE, or CGE for seven representative subpallial enhancer transgenic mouse lines (15) (Fig. 1*A*). The transgenic lines express GFP and CreER^{T2} with enhancer-driven divergent patterns in VZ, SVZ, and mantle (MZ) zones of the GEs, summarized in Fig. 1*B*. The objective of these experiments was threefold: first, to establish the utility and sensitivity to detect enhancer-driven reporter expression via scRNA-seq; second, to understand developmental origins of BG-arising neuronal populations in MGE (enhancers *hs1538*, *hs1056*, *hs799*, and *hs192*), LGE (*hs841* and *hs599*), and CGE (*hs841* and *hs953*) at E11.5; and third, to resolve developmental trajectories labeled by regionally distinct MGE progenitor (*hs1538* and *hs1056*) and neuronal (*hs799* and *hs192)* enhancers.

We performed BG microdissection and preparation of reporter-positive and ungated single cells from each enhancer line, as indicated in Fig. 1 *A* and *B*. Cell suspensions were either enriched for cells with transgene expression through FACS (*SI Appendix*, Fig. S1) or passed directly to the Fluidigm C1 system for capture and amplification of full-length mRNA from single cells. After sequencing and quality control (*SI Appendix*, Fig. S2 *A–L*), 1,027 cells were included for analysis, with ∼594,000 reads and ∼5,140 genes per cell on average.

We used a combination of FACS^{+} gating and transgene (CreER^{T2}-IRES-GFP or tdTomato) RNA expression to distinguished enhancer-labeled cells from cells with no enhancer activity: 315 cells were unsorted or FACS^{-} and transgene negative; 712 cells were FACS^{+} or had detectable reporter transcripts. Enhancer-driven reporter transcripts were detectable at single-cell resolution across all enhancers, with variation in presence and transcript level captured via scRNA-seq (Fig. 1*C*). Five enhancers exhibited strong parallels between reporter protein GFP^{+} gating and transgene transcript detected via scRNA-seq (*SI Appendix*, Fig. S2*M*). The other two lines, dMGE-*hs1538* and MGE-*hs192*, had weaker transgene scRNA-seq sensitivity, with 30 to 40% of FACS reporter-positive cells having detectable transgene transcript. Nonetheless, even for enhancers with weaker transcriptional activity, reporter transgene was reliably detected via scRNA-seq in a substantial fraction of FACS-determined reporter-positive cells, demonstrating the overall utility of this approach and enabling comparison of enhancer-labeled cells. By analyzing both enhancer-positive and unlabeled cells, we tested how specific the enhancer activities were in cells within and across these BG lineages.

Transcription Factor Expression Organizes scRNA-seq Data by Proliferative State and BG Region. Next, we sought to define the cell types present in E11.5 BG via clustering using both enhancer-positive and unsorted cells. Using highly variable genes in scRNA-seq analysis is a common approach for feature selection (16); however, this method did not adequately separate regional and cell state identity in our data (*SI Appendix*, Fig. S3). As an alternative, we used a transcription factor (TF)-curated approach, with the rationale that TFs drive lineage

specification and cell identity. We rooted this analysis using 689 TFs profiled for RNA ISH patterns at E11.5 and E13.5 in the Allen Developing Mouse Brain Atlas (ABA) (17) (*SI Appendix*, Fig. S4 *A* and *B* and Dataset S2). Of these TFs, 455 were expressed in our E11.5 scRNA-seq data, of which 292 (64.2%) had detectable ISH expression in the BG (*SI Appendix*, Fig. S4 *C–E*). We used these 455 TFs to define scRNA-seq cell identity and for visual representation via UMAP plots (Fig. 1 *D–F*). We compared TF-curated analysis to cluster identities using highly variable genes before and after performing regression analysis to reduce the influence of cell-cycle phase (*SI Appendix*, Figs. S3 *A–E*). Excluding non-TFs reduced the contribution of cell-cycle phase and confounding sources of variation (e.g., sequencing batch) while maintaining cell cycle (principal component [PC]1: 15.06% variance explained) and dissection (PC3: 4.48% variance explained) as the main sources of variance (*SI Appendix*, Fig. S3 *F* and *G*). Finally, we assessed clustering using only enhancer-positive cells, which largely reproduced results from the full set of enhancer-positive and unlabeled cells (*SI Appendix*, Fig. S3*H*). In summary, we found that using TFs as features for clustering improved resolution of cell clusters by regional origin and proliferative state.

Having shown that scRNA-seq can reliably identify enhancer-positive cells and that our TF-curated approach enables separation across regional origin and proliferative states, we next modeled transcriptional differences across the full set of enhancer-positive and enhancer-negative cells. First, we modeled transitional cell states via diffusion mapping. Second, we examined differences across cluster-based transcriptional identities. We identified two major diffusion components (DC) corresponding to proliferative state (DC1) and MGE from LGE/CGE origin (DC2) (Fig. 2*A*). DC1 captured the stem cell, proneural, and neurogenic transition, with genes such as *Hes1*, *Ccnd2*, *Gadd45g*, and *Slc34a2* marking cells at various stages of this transition (Fig. 2*B*). Lower values of DC2 were associated with MGE identity, marked by expression of MGE-specific genes, such as *Nkx2-1*, *Lhx6*, and *Lhx8* (18) (Fig. 2*B*). Higher values of DC2 were associated with LGE or CGE identity (Fig. 2 *A* and *B*). DC2 diversity was driven by expression of region-defining TFs, such as *Nkx2-1* in the MGE and *Pax6* in the CGE and LGE (Fig. 2*B*). DC1 and DC2 distinguished cells labeled by different enhancers and indicate that these developmental enhancers are active across maturation states within the GEs.

We next performed clustering using TF-curated scRNA-seq expression, identifying 12 cell clusters that separated by proliferative state and regional or cell-type identity (Fig. 2*C* and Dataset S3). We further used random forest classification to define informative transcripts that discriminate cells labeled by specific enhancers (Dataset S4). Cells labeled by enhancers dMGE-*hs1538*, MGE-*hs1056*, CGE-*hs841*, and LGE-*hs841* primarily grouped into mitotic clusters (cl)-1, cl-2, cl-3, cl-8, and cl-9, further separated by regional identity (MGE vs. non-MGE). Within these regional boundaries, VZ/SVZ-associated enhancers split across multiple clusters (Fig. 2*D*), paralleling diffusion mapping results suggesting mitotic enhancers label multiple proliferative states. Compared to enhancers with progenitor activity, enhancers active in postmitotic cells (MGE-*hs192*, MGE-*hs799*, LGE-*hs599*, and CGE-*hs953*) were biased toward specific cell-type clusters within broader regional identities (Fig. 2*D*). To characterize cell types that were differentially labeled by these enhancers, we performed differential gene-expression analysis using the full transcriptome of 17,015 expressed genes to identify differentially expressed (DE) genes for each TF-defined cluster (Fig. 2*E* and Dataset S3).

Proliferative clusters cl-1, -3, -2, and -9 encompassed VZ and SVZ proliferating cells including those labeled by mitotic-associated enhancers MGE-*hs1538*, MGE-*hs1056*, and CGE- and LGE-*hs841*. Cells in cl-1 expressed neural stem cell markers including higher expression of *Nfia*, *Nfib*, and *Olig2* (Fig. 2 *E* and *F*). cl-3 was associated with higher levels of intermediate progenitor markers such as *Asb4* and *Ascl1*. Proliferative LGE and CGE cells from *hs841* comprised the majority of cl-2 and cl-9 and expressed VZ markers including *Lhx2* and *Fzd8*. Enhancer-negative LGE and LGE-*hs841* mitotic cells additionally form clusters cl-8 and cl-11, which expressed genes such as *Lrp4*, *Has1*, and *Kitl*. cl-10 was composed primarily of enhancer-negative cells from MGE dissections and expressed *Lhx9* and *Tbr1*, indicative of cortical or diencephalic rather than BG identity. From random forest classification, markers that distinguished LGE, CGE, and MGE progenitor-associated enhancers recapitulated region-associated TFs from DC2 (Dataset S4). MGE rostrodorsal (*hs1538*) and caudoventral (*hs1056*) biased enhancers were distinguished by quantitative differences across TFs including *Otx2* and *Id4*, identifying TF expression gradients that distinguished progenitor cells across MGE regional axes.

Compared to proliferative enhancers, postmitotic enhancers active in SVZ/MZ mapped to distinct transcriptional clusters corresponding to emerging neuron types. LGE-*hs599* and CGE-*hs953* are both represented in cl-4, which expressed higher levels of genes, including *Sp8*, *Tiam2*, and *Tshz1* (Fig. 2 *E* and *F*), suggesting cl-4 is more immature than cl-5 and cl-7 and has not yet acquired strong LGE or CGE regional specificity. cl-5, composed predominantly of LGE-*hs599* cells, expressed genes including *Rbfox1*, *Tac1*, and *Zfp503*. cl-7, composed predominantly of CGE-*hs953* and enhancer-negative cells, expressed genes such as *Six3*, *Tshz2*, and *Sema3a*. Genes defining these clusters shared general markers of early GABAergic projection neurons and were consistent with fate mapping of $hs599^{+}$ and $hs953^{+}$ cells to projection neuron populations in the adult forebrain, including striatal medium spiny neurons (15, 19) and $Sp8^{+}$ neurons in the amygdala (15). MGE MZ-associated enhancers *hs192* and *hs799* also exhibited markers of early neuronal fate commitment (Fig. 2*E* and Datasets S3 and S4). cl-0, composed predominantly of $hs799^{+}$ cells, expressed early MGE-derived cortical IN lineage markers, including *Maf*, *Mafb*, *Nxph1*, *Calb1*, and *Sst*. Conversely, cl-6, composed of primarily $hs192^{+}$ cells, expressed a wide range of TFs including *Tle4* and *Zic1, Zic3*, and *Zic4*, suggestive of GABAergic and cholinergic projection neuron commitment (20).

These experiments mapped the cell type-specific activity of seven evolutionarily conserved enhancers across E11.5 GEs. We found regionally separated but otherwise similar mitotic identities among enhancer-labeled progenitors across MGE, LGE, and CGE, and localized regional signatures within MGE via comparing dorsal $hs1538^{+}$ cells with more ventral $hs1056^{+}$ cells. Across GEs, enhancer-labeled neuronal cell types emerged from more general early postmitotic clusters. Postmitotic cells in MGE separated into signatures suggesting GABAergic projection, cholinergic, and IN trajectories that were differentially labeled by *hs192* and *hs799*. This initial survey mapped the activity of these enhancers to distinct cell populations and identified known and novel markers for maturation state and regional identity for progenitor and early born neuronal cell types.

Dissection of Early Born Neuron Types and Positional Identity in MGE via Enhancer-Labeled 3′ scRNA-seq and ISH. To evaluate the impact of deeper sampling and to comprehensively interrogate emerging neuronal types in E11.5 MGE labeled by

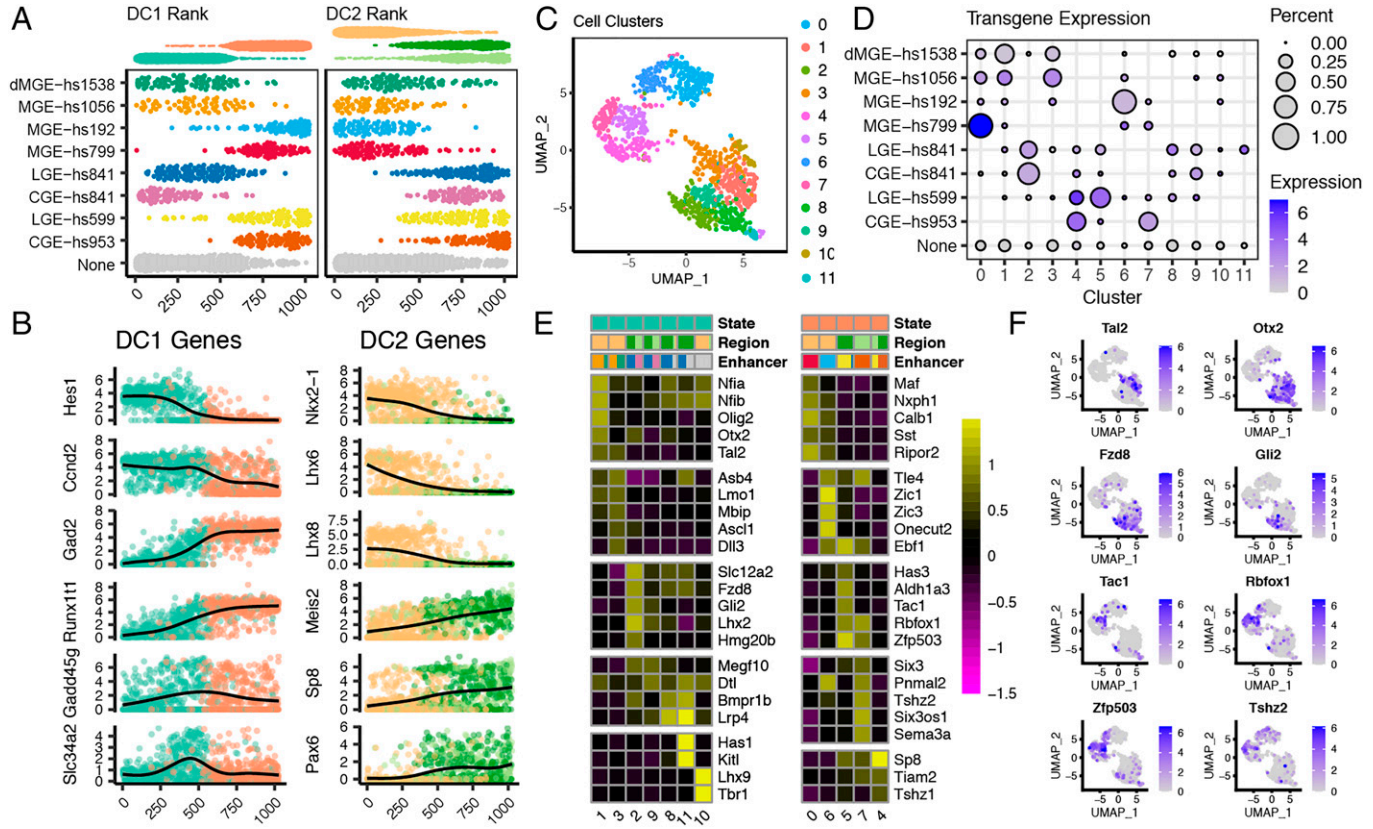

Fig. 2. Enhancers label cells across progenitor state and regional identity gradients. (*A*) Enhancer-labeled cells ranked by DC value (DC)1 (*Left*) or DC2 (*Right*). Density diagrams represent proportion of cells that are mitotic or postmitotic (green and orange) or MGE or LGE/CGE (orange and dark/light green). (*B*) Relative expression of DE genes across DC1 (*Left*) or DC2 (*Right*). Cells on the *x* axis are ordered by DC1 or DC2 rank. Line represents generalized additive model (gam) line. (*C*) UMAP colored by TF-defined cell cluster. (*D*) Dot plot of relative transgene expression and representation of each enhancer across clusters. Dot size represents percent of enhancer corresponding to specific cluster. Color gradient represents mean normalized transgene expression for each enhancer by cluster. (*E*) Representative DE genes by cluster. Color represents mean *z*-score of normalized expression across clusters. Each column color bar represents the proportion representation of proliferative state, region, or enhancer group. (*F*) UMAP plots of representative genes from *E*, colored by relative expression.

hs192 and *hs799*, we performed 3′ scRNA-seq using the 10x Genomics Chromium system. FACS-purified MGE ungated and reporter-positive cells were dissected from embryos across three litters for each transgenic line and prepared for multiplexed scRNA-seq using MULTI-sEq. (21) (Fig. 3*A* and Dataset S1). After quality control (*SI Appendix*, Fig. S5 *A* and *B*), 4,001 single cells were used for downstream analysis. We used the same TF-curated approach to drive cell clustering. Of the 463 TFs detected in the 3′ scRNA-seq MGE data, 421 (90.9%) were represented in both full-length and 3′ scRNA-seq approaches (Fig. 3*B* and *SI Appendix*, Fig. S4 *F*–*I*). We identified 18 cell clusters in the 3′ scRNA-seq data (Fig. 3*C*). Consistent with the full-length scRNA-seq data, using TFs as features reduced separation driven by cell-cycle phase (*SI Appendix*, Fig. S5 *C* and *D*). The primary component of transcriptional variation (PC1: 24.24% variance explained) in TF-curated data were associated with proliferative state (*SI Appendix*, Fig. S5*D*). Finally, we tested clustering using only enhancer-labeled cells (*SI Appendix*, Fig. S5*E*), finding that postmitotic cluster structure remained stable. Mitotic clusters were not captured by the enhancer-positive cells, as expected considering these two enhancers primarily label postmitotic cells. Overall, the MGE-derived 3′ scRNA-seq and full-length scRNA-seq data had similar cluster topology and identified similar cell states, as determined by canonical correlation analysis (*SI Appendix*, Fig. S1).

Cells were defined as enhancer-positive for *hs192* or *hs799* if they had at least one UMI (unique molecular identifier) count for CreER^{T2}-IRES-GFP. Enhancer-positive cells from independent samples clustered together, demonstrating reproducibility across biological replicates and distinct distributions of enhancer-labeled *hs192* vs. *hs799* cells (Fig. 3*E*). As expected, most enhancer-labeled cells mapped to postmitotic clusters. *hs799*$^{+}$ cells were substantially enriched (>50% of cluster composition) in cl-0, -10, and -13, and to a lesser extent in cl-2, -4, -8, and -11 (Fig. 3*F*). In contrast, *hs192*$^{+}$ cells were broadly distributed, with the greatest representation in postmitotic clusters cl-7, -8, -14, -15, and -16, and decreased representation in cl-0, -2, -10, and -13 (Fig. 3*F*). Cells labeled by at least one of the two enhancers were present in all postmitotic clusters alongside ungated enhancer negative cells, indicating that these two enhancers exhibit differential activity across scRNA-seq–defined cell type trajectories, and that together they have activity across major emerging MGE-derived neuronal lineages at E11.5. The proportion of positive cells for each enhancer varied by cell cluster (Fig. 3*F*). These results are consistent with fate mapping data showing unequal numbers of CIN, GABAergic projection, and cholinergic neuronal descendants from cells having activity of *hs192* or *hs799* (15).

We next performed DE and gene ontology (GO) analysis by cluster, expanding to include all 18,088 expressed genes to identify TF and non-TF markers (Fig. 3 *G* and *H* and Datasets S3 and S5). Proliferative clusters (cl-1, -5, -6, and-12), corresponding to cells within the MGE VZ, were enriched in terms including "DNA replication" and "chromosome segregation" (Fig. 3*H*). In comparison, proliferative clusters cl-9 and cl-3, likely corresponding to cells within the SVZ, were enriched for terms such as "exit from mitosis" (Fig. 3*H*). Initiation of

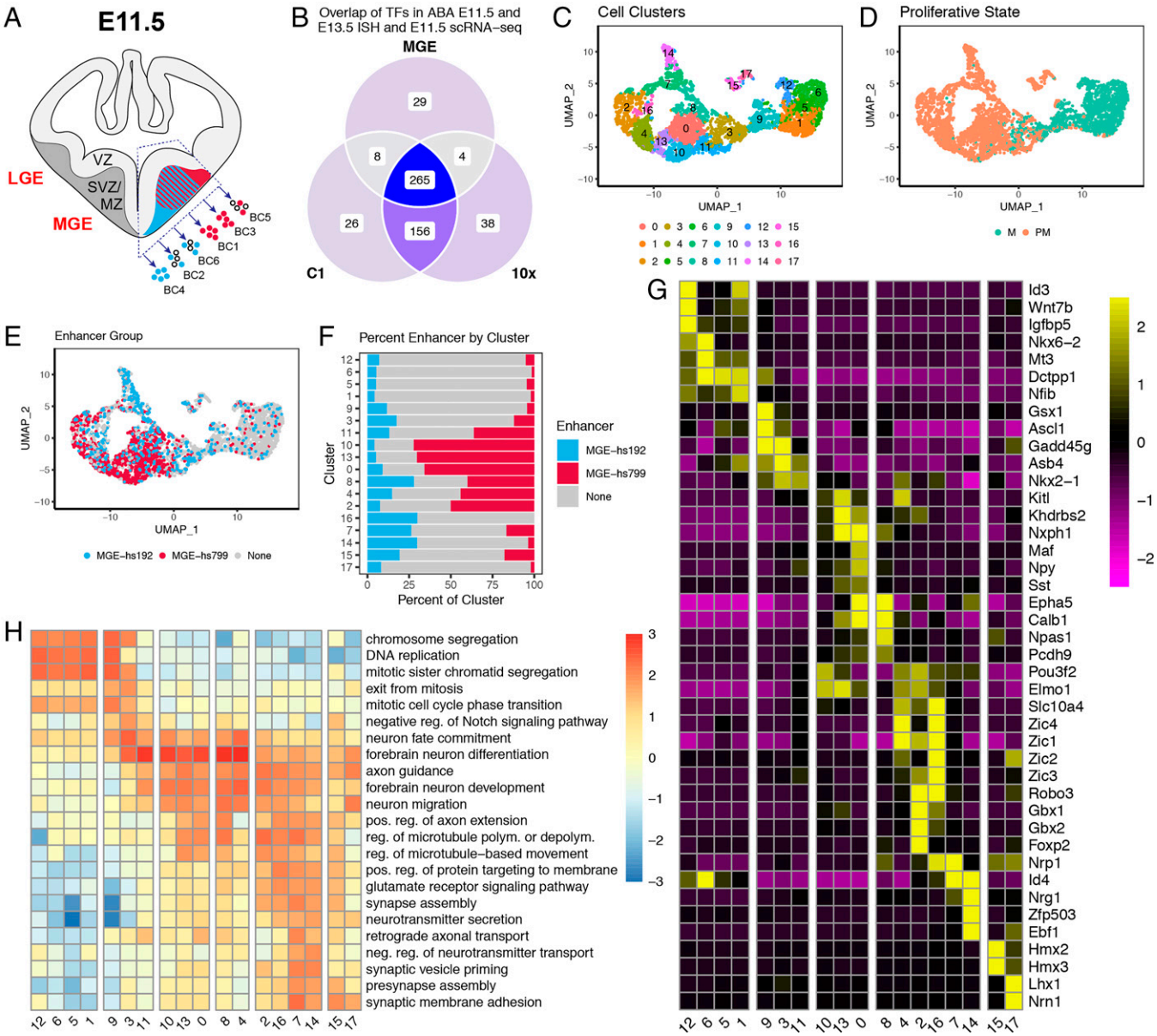

Fig. 3. Regional and proliferative gradients within E11.5 MGE. (*A*) Schematic of dissections used for MULTI-seq. (21), with colors corresponding to enhancer transgenic reporter expression in the MGE (red: *hs799*; blue: *hs192*; open circles: unlabeled). BC: barcode. (*B*) Venn diagram depicting overlap of TFs in 10x and C1 scRNA-seq datasets. (*C*) UMAP colored by TF-anchored cell clusters. (*D*) Visualization of single cells from E11.5 MGE by UMAP (green: M, mitotic; orange: PM, postmitotic). (*E*) UMAP colored by enhancer (red: *hs799*; blue: *hs192*; gray: enhancer-negative). (*F*) Proportion representation of *hs192*, *hs799*, and enhancer-negative by cluster. (*G*) Representative DE genes by cluster. Color represents mean *z*-score of normalized expression across clusters. (*H*) Representative gene ontology biological process terms separating clusters. Color represents $\log_2$(observed/expected).

enhancer activity is first evident for both *hs799* and *hs192* among cells in late SVZ clusters based on transgene expression (Fig. 3 *D* and *E*). Postmitotic clusters subdivided into MGE-derived maturing GABAergic and cholinergic trajectories (cl-0, -2, -4, -7, -8, -10, -11, -13, and -16) differentially labeled by *hs192* and *hs799*, as described in detail below.

Some clusters were made up of postmitotic cells that appeared to originate outside the MGE (cl-14, -15, and -17), and were likely migrating through MGE at E11.5 or captured at dissection boundaries. Cl-14 cells were enriched for *hs192*$^+$ cells and had properties of LGE-derived immature medium spiny neurons, which express *Ebf1*, *Nrg1*, and *Zfp503* (*Nolz1*), but not *Nkx2-1* (Fig. 3*G*). Cl-15 contained both *hs192*$^+$ and *hs799*$^+$ cells and may derive from preoptic area (POA), based on *Hmx2* (*Nkx5-2*) and *Hmx3* (*Nkx5-1*) expression. cl-17 cells were mostly transgene-negative, and may originate from regions adjacent to subpallium (e.g., hypothalamus or prethalamic eminence). These non-MGE clusters are not further discussed.

Projecting scRNA-seq Cell Identities onto Developing Mouse MGE via ISH Data. Using transcriptional identities defined by scRNA-seq, we used ABA ISH data (17) from the 689 TFs to map the anatomical distribution of cell populations in the E11.5 telencephalon. We graded expression of these TFs in VZ, SVZ, and MZ of MGE and LGE by reviewing individual sections from E11.5 and E13.5 (Dataset S2): 332 genes had visually detectable mRNA ISH patterns in the BG, of which 283 were also detected in E11.5 scRNA-seq (*SI Appendix*, Fig. S4*C*). We examined E11.5, E13.5, and E15.5 ISH time points, as ISH expression patterns that were undetectable or ambiguous at E11.5 could resolve spatial or anatomical lineages at E13.5 and E15.5. From this, we identified sentinel genes that

could be used as spatiotemporal markers to presumptively assign cells or clusters to distinct neuroanatomical regions and specific neuronal trajectories on the basis of mRNA expression. Using these sentinel markers and our scRNA-seq data, we characterized emergent specification trajectories and regional distributions of proliferative (Fig. 4) and postmitotic (Fig. 5) populations in MGE.

MGE Progenitors Stratify by VZ to SVZ and Dorsoventral and Rostrocaudal Axes. Analysis of mitotic cells (*Mki67*$^+$) in the developing MGE provided evidence for distinct progenitor stages and regional patterning (Fig. 4 *A*–*C*). We identified four mitotic clusters driven by DE genes associated with progenitor maturation steps, suggesting discrete histogenetic stages: VZ1 (neuroepithelium), VZ2 (radial glial), SVZ1 (secondary progenitor 1), and SVZ2 (secondary progenitor 2).

The most immature VZ stage, the neuroepithelium (VZ1), was represented by cells with highest expression of early mitotic markers such as *Hes1* and *Id4* (22) and correspond to cl-5 and cl-6 (Fig. 4*D* and *SI Appendix*, Fig. S7*A*). More mature VZ2 cells expressed *Wnt7b*, *Id3*, and *Ttyh1*, and were mostly within cl-12 and cl-1 (Fig. 4*E* and *SI Appendix*, S7 *F* and *G*). VZ2 cells had high expression of *Hes5* and *Fabp7*, which mark radial glia (22, 23) (*SI Appendix*, Fig. S7 *E* and *H*). Other genes marked both VZ zones, including *Lhx2*, *Rest*, and *Rgcc* (*SI Appendix*, Fig. S7 *B*–*D*). Genes like *Ascl1*, *Gsx2*, and *Dlx2* began expression in VZ (Fig. 4*I* and *SI Appendix*, Fig. S7 *J* and *K*). SVZ organized into progressively more mature SVZ1 in cl-9 and SVZ2 in cl-3. SVZ1 identity was assigned to cl-9 based on overlapping *Olig2* and *Dlx2* expression (24) (Fig. 4 *H* and *I*). *Gsx1* had perhaps the most specific SVZ1 periventricular expression by ISH and was largely confined to cl-9 (Fig. 4*F*). Cells in this cluster also expressed high levels of known SVZ genes (*Ascl1*, *Gsx2*, and *Hes6*) (25–27) and neurogenic transition markers (*Btg2*) (28) (*SI Appendix*, Fig. S7 *J*–*L*). *Gadd45g*, a marker of intermediate progenitors in the cortex (29), was highest in cl-9. SVZ2 identity was linked to cl-3. This most mature progenitor state was associated with loss of *Olig2* and high *Dlx1*, *Dlx2*, and *Dlx5* (24, 30) (Fig. 4 *H*–*K*). cl-3/SVZ2 markers (e.g., *Prox1*, *Sp9*, and *St18*) were expressed in a distinct layer of cells superficial to SVZ1 markers (Fig. 4*G* and *SI Appendix*, Fig. S7 *M* and *N*). *Insm1* and *Isl1* appeared to be expressed in both SVZ1 and SVZ2 (*SI Appendix*, Fig. S7 *O* and *P*). cl-8 and cl-11 may represent the earliest stage of neuronal commitment as SVZ2 cells exit the cell cycle; these clusters are discussed further below.

Markers indicating dorsal MGE identity (e.g., *Nkx6-2*, *Gli1*, and *Gli2*) were expressed in a subset of cl-12 and cl-6 cells (Fig. 4*L* and *SI Appendix*, Fig. S8 *A* and *B*). *Dach2*, a novel marker in this category, was expressed in the dMGE by ISH (*SI Appendix*, Fig. S8*C*). Genes indicating ventral (v)MGE identity (*Shh*, *Slit2*, and *Tal2*) (31) were expressed in a subset of cl-5 and cl-1 cells (Fig. 4*M* and *SI Appendix*, Fig. S8 *D* and *E*), with highest expression in medial sagittal planes of ISH sections (*SI Appendix*, Fig. S8 *D* and *E*). *Bcan*, *Dach1*, and *Sulf1* may also mark vMGE progenitor identity (*SI Appendix*, Fig. S8 *F*–*H*). The rostral MGE had high *Otx2* expression (31), whereas the caudal MGE was marked by high *Nr2f1* and *Nr2f2* expression (32) (Fig. 4 *N* and *O* and *SI Appendix*, Fig. S8*M*). The POA and preoptic hypothalamus (POH) are contiguous with the caudal mitotic zone in the MGE. cl-12 may represent a mixture of POA2 and POH progenitors, based on expression of *Nkx6-2*, *Dbx1*, and *Pax6* (18) (*SI Appendix*, Fig. S8 *B*, *O*, and *P*). POA1 cells lack these markers and have higher expression of *Etv1* (18); these cells may be intermixed with MGE cells within cl-1. The septum is contiguous with the rostral MGE, and the septal markers *Fgf15* (33), *Zic1*, and *Zic4* (34, 35) were also expressed by rostral MGE progenitors (*SI Appendix*, Fig. S8 *I* and *J*). *Pou3f1* and *Cntnap2* were novel markers of rostral cells (*SI Appendix*, Fig. S8 *K* and *L*), whereas *Ptx3* was a new caudal MGE marker (*SI Appendix*, Fig. S8*N*). We leveraged canonical correlation analysis of MGE cells across both C1 and 10x datasets and ISH data, showing general patterns were concordant across the datasets (*SI Appendix*, Fig. S6). Patterning markers, including *Id4*, *Otx2*, and *Tcf7l2*, distinguished *hs1538* (rostrodorsal-biased) and *hs1056* (caudoventral-biased) cells (Dataset S4), validating the regional identity evident among enhancer-labeled MGE progenitors. For all markers represented in Fig. 4 and *SI Appendix*, Figs. S7 and S8, we show quantitative differences across clusters in a summary violin plot (*SI Appendix*, Fig. S7*Q*).

Emergence of MGE Neuronal Lineages Revealed by Differential Enhancer Labeling. Postmitotic clusters were identified as the precursors of distinct MGE-derived cell lineages: GABAergic INs destined for the cortex (CINs) or other structures, GABAergic projection neurons, and cholinergic neurons (Fig. 5*A*). Early-born neurons emerge in SVZ2 and make up the MZ, which can be further visualized via a topological projection map that preserves spatial relationships across the telencephalon (15, 36) (Fig. 5 *B* and *C*). The organization and cellular outputs of the BG primordia identified herein using this topological map of E11.5 telencephalon correlate with progenitor zone anatomical map defined in Flames et al. (18) (*SI Appendix*, Table S1). Importantly, some of the cell types identified by scRNA-seq appeared to emerge in different spatial subdomains of the MGE based on overlapping marker gene expression in these topological maps.

Previous scRNA-seq studies have captured some of these emergent neuronal identities but largely have not resolved the spatial organization underlying this process (11, 12). *Nkx2-1*, *Lhx6*, and *Lhx8* marked the MGE-derived postmitotic neurons (Fig. 3*G*). To analyze these relationships, we focused on major emergent IN, GABAergic projection neuron, and cholinergic lineages and the differential activity of *hs192* and *hs799* that distinguishes these specific emerging neuronal populations.

cl-0 and a subset of cl-13 were composed primarily of *hs799*$^+$ with only a few *hs192*$^+$ cells present, and expressed immature IN markers. A subset of cl-8 included *hs799*$^+$ and *hs192*$^+$ that expressed *Mafb*. cl-0 showed the most consistent expression of immature CIN markers, including *Calb1*, *Cux2*, *Erbb4*, *Lmo3*, *Maf*, *Npy*, *Sox6*, *Sst*, and *Zeb2* (Figs. 3*G* and 5 *D*–*F* and *SI Appendix*, Fig. S9 *A*–*G*). Earlier lineage markers, such as *Lhx8*, were reduced in these cells. Subsets of cl-13 and cl-8 cells were *Mafb*$^+$, but lacked *Maf* and other CIN markers. The small proportion of *Maf*$^+$ *hs192*$^+$ cells in cl-8 and cl-0 was consistent with our C1 data and in line with fate mapping of a subset of *hs192*-lineage cells to CINs (15). Sentinel genes for cells from cl-0 and cl-13 were expressed in a distinct area in the periventricular MZ of the caudal MGE based on ISH at E11.5. We found that *Calb1* expression is largely located in the caudal Diagonal (Dg) region, as illustrated on the topological map of the telencephalon (Fig. 5*F*). The caudal Dg also showed expression of *Cux2*, *Erbb4*, *Npy*, *Sst*, *Maf*, and *Mafb*; the latter two genes are perhaps the most specific markers of MGE-derived CINs (37, 38) (Fig. 5 *D* and *E* and *SI Appendix*, Fig. S9 *A*–*D*). Contiguous with this region, in more lateral ISH sections, were cells expressing CIN markers, presumably migrating to the

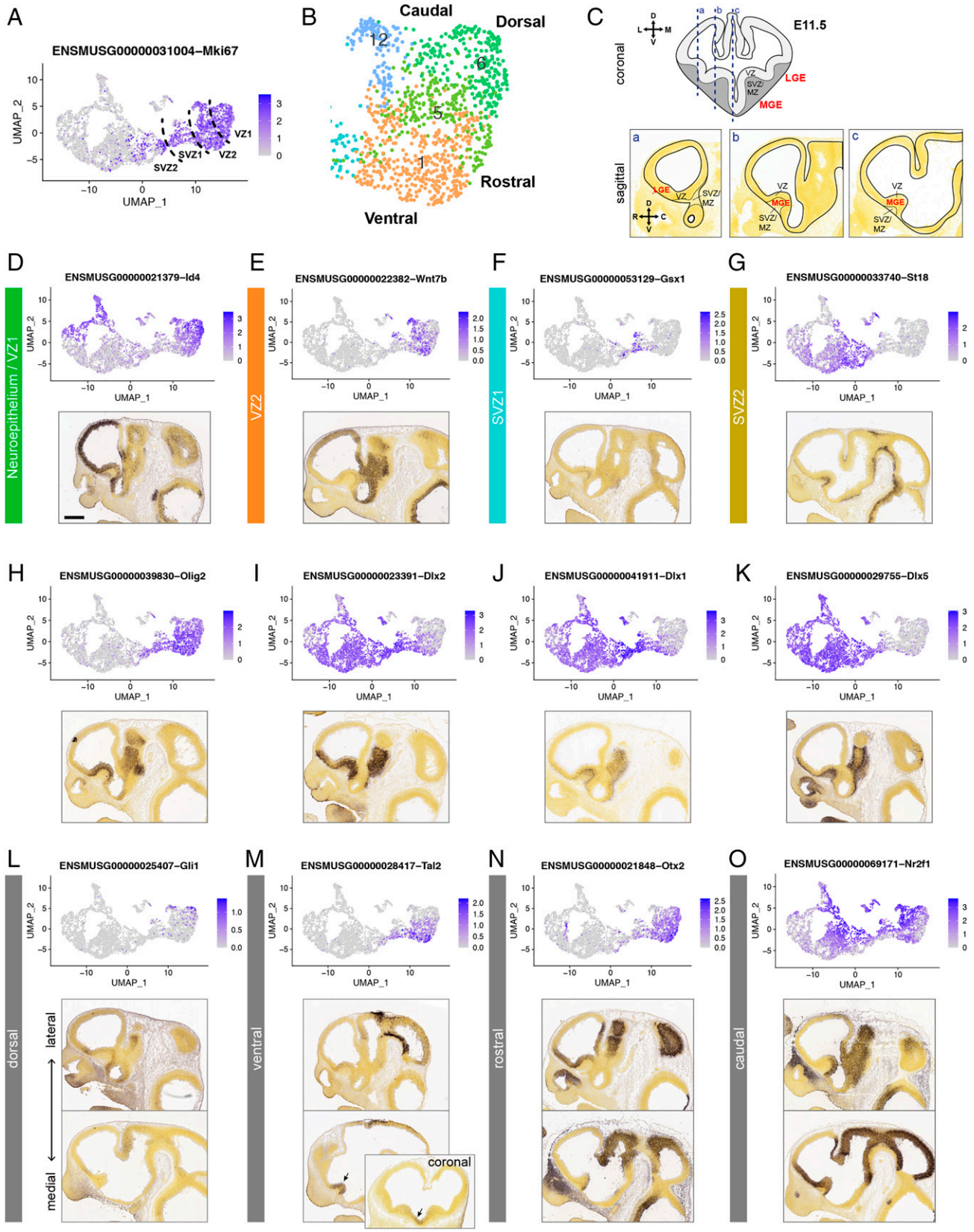

Fig. 4. Separation of mitotic progenitors in E11.5 MGE. (*A*) Visualization of single cells from E11.5 MGE by UMAP, colored by expression of mitotic marker *Mki67*. Dotted lines show proposed boundaries for cells from VZ1, VZ2, SVZ1, and SVZ2. (*B*) Region of UMAP covering the VZ cells, colored by TF-defined cell cluster to show proposed regional identities (also shown in Fig. 3*E*). (*C*, *Upper*) Coronal schematic showing positions (blue dotted lines) of parasagittal sections used to illustrate gene-expression patterns. (*Lower*) Reference images of parasagittal sections from the ABA with relevant subpallial anatomical domains labeled. (*D*–*O*) Gene-expression UMAPs and representative ISH on E11.5 medial parasagittal sections showing expression domains for mitotic cell identity markers. (*D*–*G*) Gene markers for VZ-SVZ developmental stages (*Id4*, *Wnt7b*, *Gsx1*, *St18*). Colored labels correspond to cluster identity. (*H*–*K*) UMAPS and ISH for *Olig2* and *Dlx2/1/5*, genes that help define SVZ1 and SVZ2 identity. (*L*–*O*) Gene markers with known regional expression patterns (*Gli1*, *Tal2*, *Otx2*, *Nr2f1*) in MGE VZ correlate with spatially defined UMAP locations, corresponding to the schematic in *C*. Lateral and medial sections shown. (*M*) *Tal2* expression is highest at the medial level; coronal *Inset* shows this corresponds to ventral MGE (arrow). Additional genes marking these developmental stages and regional identities shown in *SI Appendix*, Figs. S7 and S8. (Scale bar, 500 μm.)

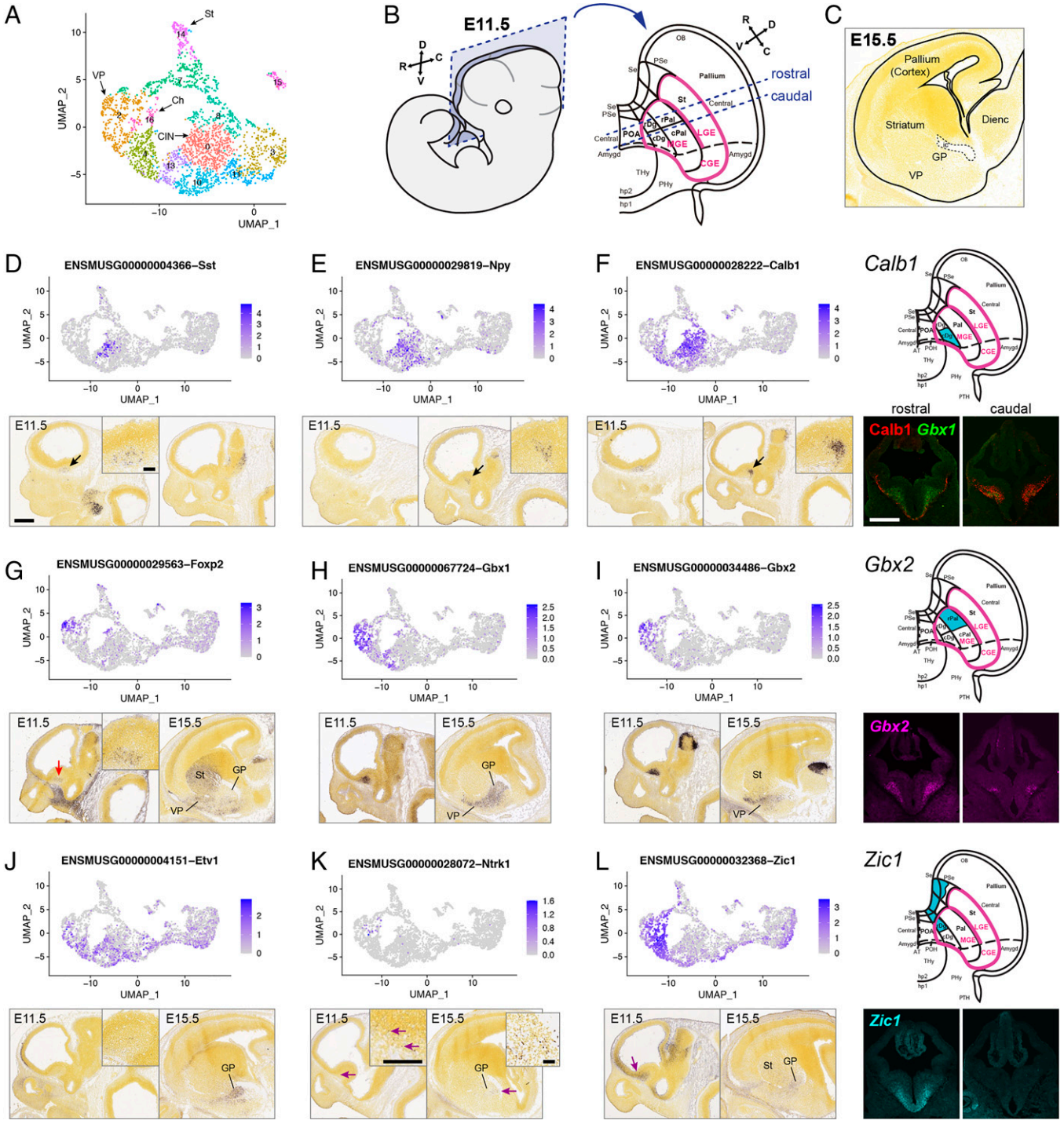

Fig. 5. Separation of early neuronal lineages in E11.5 MGE. (*A*) UMAP colored by TF-defined cell cluster (also shown in Fig. 3*E*) covering postmitotic cells, with proposed identities of cell clusters labeled. Abbreviations: Amygd, amygdala; Ch, cholinergic; GP, globus pallidus; St, striatal medium spiny neurons; VP, ventral pallidum. (*B*) Collapsed two-dimensional topological map of E11.5 MZ, as viewed from the sagittal section plane indicated. Note the rostral and caudal Pal (Pallidum) and Dg. See *SI Appendix*, Table S1 for abbreviations. Blue dotted lines (rostral and caudal) indicate positions of coronal fluorescent images shown in *F*, *I*, and *L*. (*C*) Schematic depicting broad anatomical domains at E15.5. Dienc: diencephalon. (*D*–*L*) Gene-expression UMAPs and representative ISH on E11.5 parasagittal sections from ABA showing markers for distinct postmitotic neuronal lineages. Arrows indicate regions of higher expression. Scale bars: low magnification, 500 μm; high magnification insets, 100 μm.

LGE and CGE on their way to the cortex (e.g., *Sst* and *Calb1* in Fig. 5 *D* and *F*). Additionally, *Adamts5*, *Bend4*, *Dlgap1*, *Kitl*, and *Rai2* may be novel markers of immature CINs (*SI Appendix*, Fig. S9 *H* and *I*). Based on putative maturation states and overlapping transcriptional signatures, immature INs from these three clusters may map back to late progenitor cells in SVZ2 (cl-3) via cl-11 for cl-0, cl-10 to cl-11 for cl-13, and within cl-8. These findings map the neurogenic progression and distinct spatial niches for emerging early MGE-derived INs, including *hs799*$^+$ cells that contribute to SST$^+$ CINs, as well as other populations.

The enhancer-labeled transcriptional signatures and corresponding ISH patterns suggested at least two distinct classes of GABAergic projection neurons differentially labeled by *hs799* and *hs192* and distinguished by *Gbx1*$^+$ or *Zic1*$^+$ expression, respectively, that originate in different parts of the MGE (Fig. 5 *H* and *L*). The first GABAergic projection class was preferentially labeled by *hs799* and expressed *Gbx1*, with more

restricted expression of *Foxp2* and *Gbx2,* and corresponded to *Shh*$^+$ cl-13 cells and to subsets of cl-4 and cl-2 (Fig. 5 *G–I* and *SI Appendix*, Fig. S8*D*). *Gbx1*, *Kitl*, *Lmo3*, *Sox6*, *Th*, *Tle4*, *Tshz2*, *Zeb2*, and *Zic1* (5, 37) are additional markers for cells from cl-2, cl-4, and cl-13 (Figs. 3*G*, 5*L*, and 6*D* and *SI Appendix*, Fig. S9 *E–H*, *J*, and *K*). These cells may map to less mature states in cl-10 and cl-1 (for cl-13 only) or cl-4 and cl-11 (for cl-2). On the topological map of the MGE, *Gbx2* expression was in the rostral portion of the pallidal subdivision (Fig. 5*I*). Anatomically, the E15.5 expression domains of *Foxp2*, *Gbx2*, *Gbx1*, and *Etv1*, determined by ISH, were nested along the pallidum's radial axis. *Foxp2* expression was largely superficial, possibly in the ventral pallidum (39) (Fig. 5*G*). *Gbx1* expression encompassed the ventral pallidum and the entire globus pallidus (Fig. 5*H*). *Gbx2* expression included the ventral pallidum and part of the globus pallidus (20) (Fig. 5*I*). *Etv1* was expressed throughout the globus pallidus but in few cells in the ventral pallidum (5) (Fig. 5*J*). *Foxp2*$^+$ cells become superficial pallidal projection neurons (i.e., ventral pallidum), whereas other cl-2 cells, and some cl-4 and cl-13 cells, may contribute to deeper pallidal structures (i.e., globus pallidus). Thus, we propose these populations that share increased *hs799* activity branch from immature neuron types within cl-2, cl-4, and cl-13.

The second GABAergic lineage expressed *Zic1*, as well as more restricted expression of *Zic3* and *Zic4,* and was enriched for clusters with high *hs192* activity (Fig. 5*L*). ISH MGE expression of *Zic* TFs was restricted to the rostroventral MGE and overlapped little with *Gbx2*; *Zic1* may not overlap with the more caudal CIN markers (Fig. 5 *D–F* and *SI Appendix*, Fig. S9 *A–D*). On the topological map, *Zic1* expression was restricted to the rostral Dg region, and was continuous with *Zic1* expression in the septum (Fig. 5*L*). As noted, *Zic1* was also expressed in the VZ and SVZ of the rostral MGE and septum, suggesting that this progenitor zone may generate the *Zic1*$^+$ postmitotic cells in cl-11 and cl-4. It is unclear from ISH how *Zic1*-associated GABAergic projection neurons are spatially organized within the GP, though the spatial organization appears more diffuse across the GP. These *hs192*-biased *Zic1*$^+$ cells may constitute a distinct cell type within the GP.

Cells in cl-16 were exclusively enhancer *hs192*$^+$ (i.e., no *hs799*$^+$ cells were identified) and expressed definitive cholinergic marker genes. Among these, *Ntrk1* (40) was the most specific to this cluster (Fig. 5*K*). Although *Ntrk1* expression was weak in ISH at E11.5, scattered positive cells were visible in the striatum and GP by E15.5, consistent with the distribution of cholinergic INs (Fig. 5*K*). Cl-16 cells also expressed *Gbx1, Gbx2*, *Isl1*, and *Zic4*, all of which are cholinergic lineage markers (4, 20, 41–43), as well as *Zic1*, *Zic2*, and *Zic5* (Fig. 5 *H*, *I*, and *L* and *SI Appendix*, Fig. S9 *L* and *M*). *Fgf15* appeared to be a novel marker of this population, but its expression was not maintained as cells mature (*SI Appendix*, Fig. S9*N*). A subset of *hs192*$^+$ cl-4 cells also expressed some of these markers (e.g., *Zic2*) and may represent a less mature state or another type of cholinergic neuron. *Fgf8*, *Fgf17*, and *Nkx2-1*/*Zic4* fate mapping provides evidence that these cells arise from the junction of the rostromedioventral MGE with the septum (41, 44), consistent with the expression of *Zic1*.

The final postmitotic group, cl-7, included two subpopulations: the upper part was almost exclusively *hs192*$^+$, while the lower part also contained hs799$^+$ cells. The upper part of cl-7 was *Nkx2-1*$^-$ and strongly expressed *Six3* and *Sp8* (*SI Appendix*, Fig. S9 *O* and *P*), suggesting a CGE-derived CIN or LGE-derived olfactory bulb IN identity (45). The lower part of cl-7 was *Nkx2-1*$^+$ and *Lhx8*$^+$ (and thus likely MGE-derived); these cells also expressed genes shared by cl-2 and cl-14, including *Id4* and *Tle4* (Figs. 4*D* and 6*D*). Earlier states of cl-7 cells may map back to cl-8 then SVZ2 cl-3 cells, with *Nr2f2* expression in these clusters indicating caudal MGE origin (*SI Appendix*, Fig. S8*M*).

Enhancer-labeled scRNA-seq and ABA ISH indicated specific spatial localizations for cell types with divergent transcriptomic identities. To verify this finding, we performed fluorescent IHC and ISH to map protein expression across coronal sections for four sentinel genes representative of these populations (Fig. 5 *F*, *I*, and *L*). We examined Calb1 for early CINs, with predicted expression in caudal Dg; *Gbx1* and *Gbx2* for GABAergic projection neurons in rostral Palladium; and *Zic1* for rostral Dg GABAergic projection and cholinergic populations. These experiments validated that early born neuronal populations indeed exhibit spatial segregation to specific rostrocaudal and dorsoventral zones of the MGE MZ. For all postmitotic markers represented in Fig. 5 and *SI Appendix*, Fig. S9, we show quantitative differences across clusters in a summary violin plot (*SI Appendix*, Fig. S9*Q*). Our experiments captured specification trajectories and associations between spatial and transcriptomic identity of E11.5 MGE-derived neuron populations labeled by *hs192* and *hs799*.

Validation of MGE Enhancer-Labeled Populations. Lastly, we sought to independently validate spatial and transcriptomic differences between proliferative and postmitotic enhancer-labeled cell populations. We used ISH to map the highest expression of the *hs1538*-associated marker *Id4* in the VZ of the dorsal and rostral MGE, where *hs1538* is active (*SI Appendix*, Fig. S10). In postmitotic populations, we tested for overlap between *hs192* and *hs799* GFP$^+$ enhancer-labeled cells and *Tle4* and *Zic1*, two TFs that were enriched in *hs192*$^+$ cells (Fig. 6). *hs192* activity and *Zic1* expression were highest in the rostroventral MGE and paraseptal region, whereas *hs799* GFP expression was in the dorsal MGE (Fig. 6 *A* and *B*; see section series in *SI Appendix*, Fig. S11). For *Tle4,* scRNA-seq and ABA ISH show that it is restricted to specific cell clusters in the MGE. We performed immunofluorescent colabeling of TLE4 with *hs192* GFP and *hs799* GFP; quantification of the proportion of cells labeled by both TLE4 and GFP from each enhancer (Fig. 6*D*) showed significantly more *hs192* GFP$^+$ cells were TLE4$^+$ compared to *hs799* GFP$^+$ cells (Fig. 6 *E* and *F*). Thus, ISH and colabeling experiments validate the results from scRNA-seq–based separation of enhancer-labeled MGE cells.

Discussion

Transcriptional profiling at single-cell resolution has transformed our understanding of the diversity of cell types in the brain. Here, we combined enhancer-based cell labeling, TF-anchored scRNA-seq clustering, and ISH-based spatial annotation to map the neurogenic landscape of embryonic mouse BG. Through this integrated approach, we quantified enhancer activity in specific cells in vivo and generated new insights regarding the specification paths for early GABAergic neurogenesis in the ganglionic eminences.

Enhancer activity is often tightly restricted to specific cell types and developmental stages (46–48), making enhancers potent tools for genetic labeling and manipulation (13–15). However, the specificity of enhancer action in vivo at the single-cell level has not been deeply explored. Our results demonstrate that scRNA-seq can capture reporter transcripts across histogenetic subtypes labeled by individual enhancers. By profiling both enhancer-positive cells and unlabeled cells, our scRNA-seq analysis revealed the onset and offset of enhancer

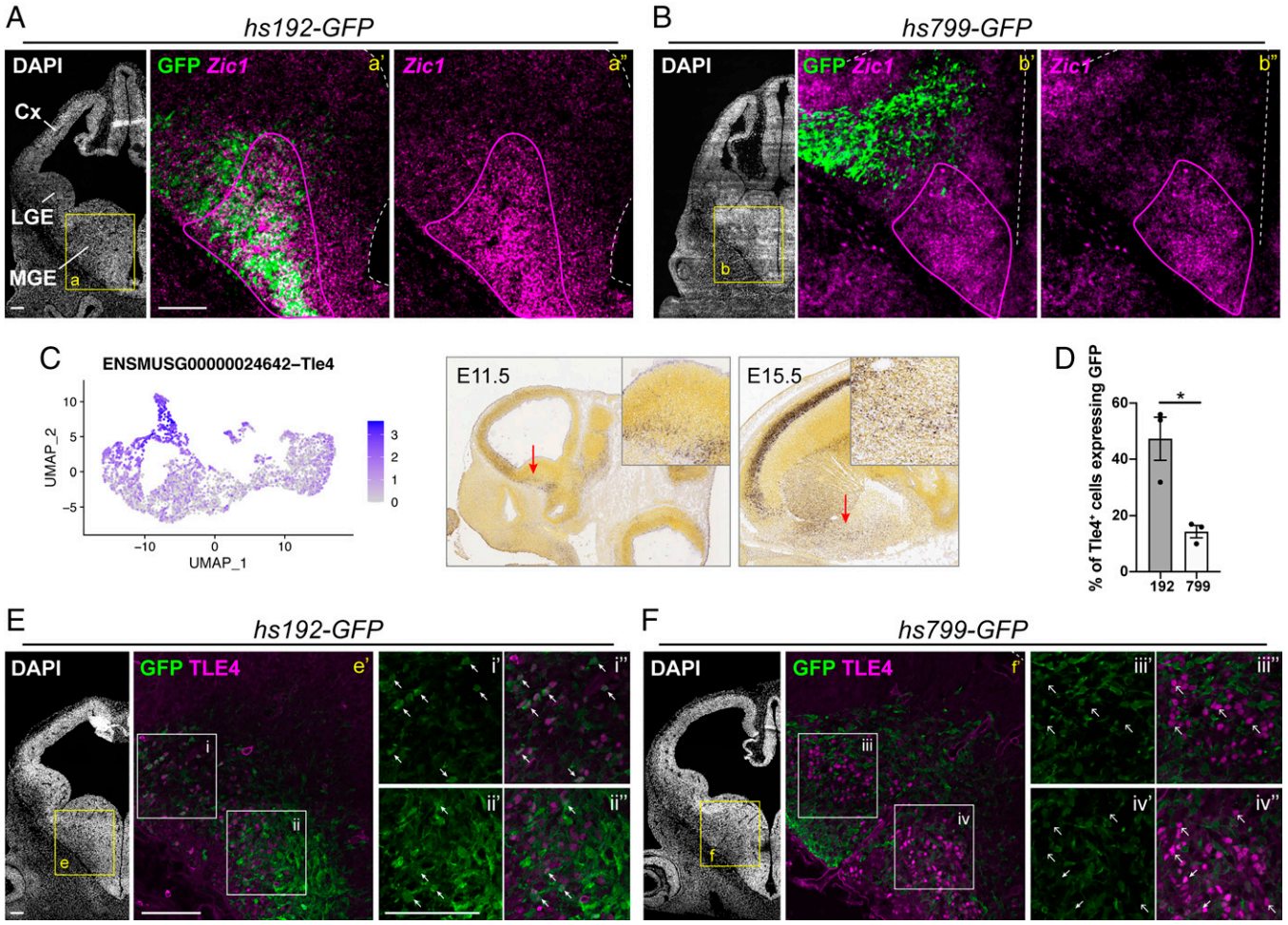

Fig. 6. Spatial differences in the activity of E11.5 MGE enhancers. Zic1 fluorescent ISH (magenta) and GFP IHC (green) on superimposed adjacent coronal sections of E11.5 *hs192* (*A*) and *hs799* (*B*) forebrain. Lower magnification views of DAPI-stained coronal hemisections; yellow boxes (a and b) outline MGE regions shown for adjacent immunostained/ISH panels (a′ and b′). Zic1$^+$ domain demarcated by purple outline. *hs192* GFP overlaps with Zic1 expression in ventral MGE (a′ and a′′). Conversely, *hs799* GFP$^+$ cells in dorsal MGE do not extend into the region of high Zic1 expression (b′ and b′′). See *SI Appendix*, Fig. S11 for additional sections. (*C*) UMAP of *Tle4* expression and representative Tle4 ISH on E11.5 and E15.5 ABA sagittal sections. Red arrows indicate areas of higher expression. Scale bars: low magnification, 500 μm; high magnification insets, 100 μm. (*D*) Quantification of percent Tle4$^+$ cells that are *hs192* GFP$^+$ and *hs799* GFP$^+$ (from *E* and *F*). (*E* and *F*) Double-IHC for Tle4 and GFP on *hs192* (e′) and *hs799* (f′) coronal hemisections. Closed arrows indicate double-positive cells and open arrows GFP$^-$ cells. Lower magnifications show DAPI-stained coronal hemisections. MGE ventricular surface indicated by dotted white lines. Abbreviations: Cx: cortex. Unpaired *t* test, **P* = 0.0145. (Scale bars, 100 μm.)

activity, as well as cell populations where each enhancer was active with fine-scale resolution. Our results highlight the specificity and complexity of developmental enhancer activity during regional and cell type specification, as epitomized here by the activity for *hs192* and *hs799* across emerging MGE neuronal lineages. This study offers a new perspective of single cell-resolved in vivo activity for seven evolutionarily conserved neurodevelopmental enhancers.

Two recent studies (11, 12) used scRNA-seq to follow BG development and CIN lineage specification in the E12.5 to E14.5 embryonic mouse. Our results capture an earlier timepoint when alternative neuronal lineages originate, and we identified transcriptomic differences among progenitor and postmitotic cells across the GEs. We show that using a TF-anchored clustering approach improved resolution of developmentally relevant clusters in our E11.5 data. Furthermore, integrating enhancer labeling and ISH data, we assigned likely identities to four types of MGE progenitors; distinct MGE progenitor regions; and enhancer-labeled subtypes of maturing MGE-derived INs, GABAergic projection neurons, and cholinergic neurons that together mature to form pallidal structures or migrate to become INs in the cortex, amygdala, and striatum.

Progenitor cohorts with overlapping transcriptional states and region-specific signatures were labeled by spatially distinct enhancers with VZ activity (*hs1538*, *hs1056*, *hs841*) Figs. 2 and 3). Similar to studies at older ages (11, 12), we found at E11.5 that scRNA-seq resolves the maturation gradient of BG progenitor populations (Fig. 4*A*). Comparison of *hs1538*$^+$ and *hs1056*$^+$ enabled the discovery of rostrocaudal (*Id4*, *Otx2*, *Tcf7l2*, and *Zic1*), and dorsoventral (*Nr2f1/2* and *Tal2*) axes for MGE progenitors. Interestingly, *Id4* is both associated with neuroepithelial progenitors (49, 50) and biased toward rostrodorsal MGE VZ, indicating potential differences in progenitor state composition across regional domains within E11.5 MGE VZ. Positional information in E11.5 MGE progenitors appears largely encoded by gradients of these and other TFs rather than the expression of specific TF domains. Combinations of TFs then activate region-specific enhancers such as *hs1538*, similar to the classic model of ectoderm patterning in *Drosophila* embryos (51) and as has been reported for the cortical VZ (14).

Early-born CGE, LGE, and MGE GABAergic projection, cholinergic, and inhibitory neuronal developmental trajectories are labeled by differential enhancer activities (*hs953*, *hs599*, *hs192*, and *hs799*, respectively) (Figs. 2, 3, and 5). Compared

to the distinct MGE signature, LGE and CGE cells had more similar transcriptomic identities, but diverged between *hs599* and *hs953* cells in later postmitotic neuron clusters. Comparison of MGE enhancers *hs192* and *hs799* in combination with ISH annotation defined three spatially distinct regions in the MGE MZ that give rise to molecularly distinct cells. CINs (expressing multiple markers) were detected in the caudal Dg region (Fig. 5*F*). *Gbx1/2*$^+$ cells were mapped to the Pallidal region and *Zic1*$^+$ cells to the rostral Dg regions (Fig. 5 *I* and *L*). We are intrigued by the possibility that *Gbx*$^+$ and *Zic*$^+$ cells contribute to distinct types of MGE GABAergic neurons, including within the GP, in addition to the cholinergic neurons already described (20, 41). Previous studies have not identified a specific MGE region giving rise to CINs, nor spatially distinct zones for generation of pallidal neurons and CINs. This contrast to our findings may be because our analysis focused on a younger age (E11.5) than most studies. Consistent with our results, Puelles et al. (36) previously showed that *Sst* expression begins in the Dg region. It is probable that as development proceeds, additional MGE regions also generate CINs (11, 12, 15). Overall, the results from our experiments define emerging GABAergic neuron types and elucidate spatial and transcriptomic relationships in embryonic BG during mammalian brain development.

Beyond BG neurogenesis, our study has broader implications for the application of scRNA-seq to developing tissues. We demonstrate that anchoring scRNA-seq analysis on TFs reduces weight of cell cycle and technical sources of variation, improving the power of histogenetic cell-type classification. This is consistent with the vast body of literature defining TF gradients as master regulators of lineage specification. Our results highlight the value and need for curated approaches in scRNA-seq analysis (e.g., our focus on TF transcripts) and the utility of orthogonal data, in this case ISH and enhancer labeling, to understand complex developmental processes. In summary, this study represents a unique approach, pairing scRNA-seq with enhancer-driven reporter expression in vivo, and offers insight into GABAergic neurogenesis.

Methods

Resource Availability.

Lead contacts. Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding authors, A.S.N. (asnord@ucdavis.edu) and J.L.R.R. (john.rubenstein@ucsf.edu).

Materials availability. The enhancer transgenic mouse lines used in this study have been previously published (15) and deposited to the Mutant Mouse Research and Resource Centers (MMRRC) repository.

Data and code availability. The datasets generated are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO). The analysis codes can be found on the Nord Lab Git Repository (https://github.com/NordNeurogenomicsLab/).

Experimental Model and Subject Details.

Mice. Embryos of either sex from the seven enhancer transgenic mouse lines were used, and all embryos of the correct genotype from a single litter were pooled as a single biological replicate for all sequencing experiments. All animal care, procedures, and experiments were conducted in accordance with the NIH guidelines and approved by the University of California, San Francisco Animal Care Committee's regulations (Protocol AN180174-02).

Methods Details.

scRNA-seq library generation and analysis. E11.5 transgene-positive embryos were identified by screening on a fluorescent microscope. MGEs, LGEs, or CGEs were dissected, pooled, and single-cell suspensions prepared. For gated samples, GFP$^+$ and tdTomato$^+$ cells were isolated by FACS. For ungated samples, DAPI-stained cells were excluded but GFP$^+$ and GFP$^-$ cells were collected. For full-length transcriptome scRNA-seq, cells were loaded onto Fluidigm integrated fluidics chips on the Fluidigm C1 system and cell lysis, reverse transcription and cDNA amplification were performed according to manufacturer protocol. One microliter of diluted cDNA per well was used for library preparation using the Nextera XT DNA Sample Prep Kit. For multiplexed 3′ scRNA-seq library generation (MULTI-seq), dissociated cells for each single-litter pooled MGE sample were labeled with barcoded lipid-modified oligonucleotides, as previously published (21), and processed on the Chromium 10x system. Single-cell cDNA libraries were generated using the Chromium Single Cell 3′ GEM, Library & Gel Bead Kit (v3) according to manufacturer's instructions. After cDNA clean-up, supernatant containing the barcode library fraction was processed as described previously (21) and sequenced.

C1 libraries were sequenced on an Illumina HiSEq. 4000 instrument using a single-end 50-bp protocol. Reads were uniquely aligned to a custom mouse genome including sequences for CreER^{T2}, IRES, EGFP, and tdTomato, duplicates were removed, and gene counts were generated. 10x MULTI-seq libraries were sequenced on an Illumina NovaSeq SP instrument and processed using CellRanger (v3.0.2) (52) using the custom mouse genome. Samples were demultiplexed and negative or multiplet cells were removed using the MULTI-sEq. (21) pipeline. Raw and aligned read quality was assessed for 5′ or 3′ bias, exonic read distribution and GC content distribution. We used the R package Seurat (v3.2.2) (16, 53) for feature selection, clustering, and visualization. To define proliferative state and cell cycle phase, we used the Seurat function *CellCycleScoring* on a published list of cell cycle genes (54) to assign mitotic (M) or postmitotic (PM) state. We removed all cells from clusters with high mtRNA expression and outlier cell clusters.

We extracted counts for 689 transcription factors scored for E11.5 and E13.5 BG and cortex (Dataset S2) using ISH data available in the ABA (17). Genes with expression in >10 cells were used. Principal components analysis (PCA) were used to define clusters. For full transcriptome analyses, the top 3,000 highly variable genes were used. We performed diffusion mapping via the R package *destiny* (55) on the TF-based C1 dataset. For DE analysis, we expanded to all expressed genes and compared clusters via Wilcoxon rank-sum test. Genes with adjusted $P < 0.05$ were considered statistically significant. We assessed DE along DCs using a generalized additive model via the *gam* function in R. We used canonical correlation analysis to combine C1 and 10x data using both TF-curated and full transcriptome datasets. The integrated dataset was used for PCA and the first 15 PCs were used to define 17 clusters. We used the R package topGO (v2.36.0) (56) to perform GO analysis by cluster and enhancer group on DE genes in the 10x dataset. We used random forest (57) to identify TFs important for classifying cells in pairwise comparisons of enhancers in the C1 data, using out-of-box error rates for all forests, mean decrease in accuracy, and mean decrease in node impurity scores for all genes.

ISH transcription factor scoring. Expression levels of TFs in the ABA (17) were estimated in subdomains of the subpallium for two ages: E11.5 and E13.5. The subpallium was divided into LGE, MGE, septum, and POA. The LGE and MGE were further subdivided into laminar subdomains: VZ/SVZ and MZ for E11.5; VZ, SVZ1, SVZ2, and MZ for E13.5. Expression was annotated from sagittal sections in each subdomain.

Histology. E11.5 heads were drop-fixed in 4% paraformaldehyde, cryoprotected in 20% sucrose overnight, embedded in OCT (TissueTek), and frozen on dry ice. The 15-μm cryostat sections were cut onto SuperFrost slides and dried, rinsed in PBS, blocked for 1 h at room temperature in PBST, and incubated with primary antibody overnight. Sections were washed then incubated with Alexa fluor-conjugated secondary antibodies. Finally, sections were washed and coverslipped with Fluorescence Mounting Medium (DAKO #S3023). For FISH, antisense RNA probes have been described previously (26). ISH was performed on 15-μm cryostat sections as described previously (58) up until antibody blocking, with the addition of peroxidase quenching for 20 min after SSC washes. Low-magnification epifluorescent images were taken using a 4× or 10× objective, and confocal images with 20× air and 40× oil objectives. Raw images were preprocessed to adjust brightness/contrast and convert to 8-bit RGB. Confocal images were stitched to create composites. Cell counting was performed on single confocal image planes. Counts were summed from at least three rostrocaudal sections for each brain. Tle4$^+$ cells were counted first with the GFP channel hidden, excluding cells in the VZ (designated by nuclear staining), then scored as positive or negative for overlap with GFP.

Data Availability. The data have been deposited in Github (https://github.com/NordNeurogenomicsLab/; and the NCBI GEO database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE199352). Some study data available. (Requests for resources and reagents should be directed to and will be fulfilled by corresponding authors, A.S.N. and J.L.R.R.). Previously published data were used for this work. [The enhancer transgenic mouse lines used in this study have been previously published (1) and deposited to the MMRRC repository.]

ACKNOWLEDGMENTS. University of California, San Francisco (UCSF) and Davis (UCD) cores supported this work. Sequencing was carried out at the UCSF Center for Advanced Technology and IHG Genomics Core, UCD DNA Core; confocal microscopy was carried out at the UCSF Nikon Imaging Center (1S10OD017993); cell sorting was carried out at the UCSF Helen Diller Family Comprehensive Cancer Center Laboratory for Cell Analysis (P30CA082103) and UCSF Parnassus Flow Core (RRID:SCR_018206, NIH P30 DK063720, 1S10OD021822). L.S.-F. was supported by the UCD Schwall Fellowship in Medical Research and Werner and Jacobsen Fellowship, and National Institute of General Medical Sciences (NIGMS)-NIH T32-GM008799. R.C.-P. was supported by a Science Without Borders Fellowship (Conselho Nacional de Pesquisas, Brazil). This work was supported by the following research grants: National Institute of Mental Health (NIMH) R01 MH081880 and NIMH R37/R01 MH049428 (to J.LR.R.); NIH/NIGMS R35 GM119831 (to A.S.N.); Department of Defense W81XWH-10-1-1023 and W81XWH-13-1-0221 (to Z.J.G.); NIH U01CA199315 and DP2 HD080351-01 (to Z.J.G.); NSF MCB-1330864 (to Z.J.G.); and the UCSF Center for Cellular Construction DBI-1548297 (to Z.J.G.). Z.J.G. is a Chan-Zuckerberg BioHub Investigator. L.P. is supported by Seneca Foundation (5672 Fundación Séneca, Autonomous Community of Murcia) Excellency Research Contract 19904/GERM/15. This work was partially supported by the Allen Institute for Brain Science. We thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.

Author affiliations: [a]Department of Psychiatry and Behavioral Sciences, University of California, Davis, CA 95616; [b]Department of Neurobiology, Physiology, and Behavior, University of California, Davis, CA 95616; [c]Nina Ireland Laboratory of Developmental Neurobiology, Department of Psychiatry, University of California, San Francisco Medical School, San Francisco, CA 94143; [d]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143; [e]Department of Pediatrics, University of California, San Francisco, CA 94143; [f]Allen Institute for Brain Science, Seattle, WA 98109; [g]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA 94158; [h]Chan Zuckerberg BioHub, University of California, San Francisco, CA 94143; [i]Center for Cellular Construction, University of California, San Francisco, CA 94143; and [j]Department of Human Anatomy and Psychobiology, IMIB-Arrixaca Institute, University of Murcia, 30100 Murcia, Spain

Author contributions: L.S.-F., A.N.R., S.N.S., M.J.H., C.T., Z.J.G., H.Z., J.L.R.R., and A.S.N. designed research; L.S.-F., A.N.R., S.N.S., I.Z., and G.L.M. performed research; M.J.H., C.T., and H.Z. contributed to in situ hybridization data generation and access; L.S.-F., A.N.R., R.C.-P., K.J.L., A.R.Y., C.S.M., T.E.R., L.P., J.L.R.R., and A.S.N. analyzed data; and L.S.-F., A.N.R., J.L.R.R., and A.S.N. wrote the paper.

1. L. Beccari, R. Marco-Ferreres, P. Bovolenta, The logic of gene regulatory networks in early vertebrate forebrain patterning. *Mech. Dev.* **130**, 95–111 (2013).
2. A. S. Nord, Learning about mammalian gene regulation from functional enhancer assays in the mouse. *Genomics* **106**, 178–184 (2015).
3. J. L. R. Rubenstein, K. Campbell, "Neurogenesis in the basal ganglia" in *Comprehensive Developmental Neuroscience. Patterning Cell Type Specification in the Developing CNS and PNS*, J. Rubenstein, P. Rakic, Eds. (Academic Press, 2020), pp. 399–403.
4. A. Fragkouli, N. V. van Wijk, R. Lopes, N. Kessaris, V. Pachnis, LIM homeodomain transcription factor-dependent specification of bipotential MGE progenitors into cholinergic and GABAergic striatal interneurons. *Development* **136**, 3841–3851 (2009).
5. P. Flandin, S. Kimura, J. L. R. Rubenstein, The progenitor zone of the ventral medial ganglionic eminence requires Nkx2-1 to generate most of the globus pallidus but few neocortical interneurons. *J. Neurosci.* **30**, 2812–2823 (2010).
6. S. Nóbrega-Pereira *et al.*, Origin and molecular specification of globus pallidus neurons. *J. Neurosci.* **30**, 2824–2834 (2010).
7. S. A. Anderson, D. D. Eisenstat, L. Shi, J. L. R. Rubenstein, Interneuron migration from basal forebrain to neocortex: Dependence on Dlx genes. *Science* **278**, 474–476 (1997).
8. R. Batista-Brito, C. Ward, G. Fishell, "The generation of cortical interneurons" in *Comprehensive Developmental Neuroscience. Patterning Cell Type Specification in the Developing CNS and PNS*, J. Rubenstein, P. Rakic, Eds. (Academic Press, 2020), pp. 461–480.
9. L. Lim, D. Mi, A. Llorca, O. Marín, Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).
10. O. Marín, S. A. Anderson, J. L. R. Rubenstein, Origin and molecular specification of striatal interneurons. *J. Neurosci.* **20**, 6063–6076 (2000).
11. C. Mayer *et al.*, Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
12. D. Mi *et al.*, Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018).
13. A. Visel *et al.*, A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013).
14. K. Pattabiraman *et al.*, Transcriptional regulation of enhancers active in protodomains of the developing cerebral cortex. *Neuron* **82**, 989–1003 (2014).
15. S. N. Silberberg *et al.*, Subpallial enhancer transgenic lines: A data and tool resource to study transcriptional regulation of GABAergic cell fate. *Neuron* **92**, 59–74 (2016).
16. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
17. E. S. Lein *et al.*, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
18. N. Flames *et al.*, Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).
19. M. M. McGregor *et al.*, Functionally distinct connectivity of developmentally targeted striosome neurons. *Cell Rep.* **29**, 1419–1428.e5 (2019).
20. L. Chen, M. Chatterjee, J. Y. H. Li, The mouse homeobox gene Gbx2 is required for the development of cholinergic interneurons in the striatum. *J. Neurosci.* **30**, 14824–14834 (2010).
21. C. S. McGinnis *et al.*, MULTI-seq: Sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
22. R. Kageyama, T. Ohtsuka, T. Kobayashi, Roles of Hes genes in neural development. *Dev. Growth Differ.* **50** (suppl. 1), S97–S103 (2008).
23. L. Feng, M. E. Hatten, N. Heintz, Brain lipid-binding protein (BLBP): A novel signaling system in the developing mammalian CNS. *Neuron* **12**, 895–908 (1994).
24. M. A. Petryniak, G. B. Potter, D. H. Rowitch, J. L. R. Rubenstein, Dlx1 and Dlx2 control neuronal versus oligodendroglial cell fate acquisition in the developing forebrain. *Neuron* **55**, 417–433 (2007).
25. K. Roychoudhury *et al.*, Physical interactions between Gsx2 and Ascl1 balance progenitor expansion versus neurogenesis in the mouse lateral ganglionic eminence. *Development* **147**, dev185348 (2020).
26. J. E. Long *et al.*, Dlx1&2 and Mash1 transcription factors control striatal patterning and differentiation through parallel and overlapping pathways. *J. Comp. Neurol.* **512**, 556–572 (2009).
27. M. H. Porteus *et al.*, DLX-2, MASH-1, and MAP-2 expression and bromodeoxyuridine incorporation define molecularly distinct cell populations in the embryonic mouse forebrain. *J. Neurosci.* **14**, 6370–6383 (1994).
28. W. Haubensak, A. Attardo, W. Denk, W. B. Huttner, Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3196–3201 (2004).
29. S. A. Yuzwa *et al.*, Developmental emergence of adult neural stem cells as revealed by single-cell transcriptional profiling. *Cell Rep.* **21**, 3970–3986 (2017).
30. D. D. Eisenstat *et al.*, DLX-1, DLX-2, and DLX-5 expression define distinct stages of basal forebrain differentiation. *J. Comp. Neurol.* **414**, 217–237 (1999).
31. R. V. Hoch, S. Lindtner, J. D. Price, J. L. R. Rubenstein, OTX2 transcription factor controls regional patterning within the medial ganglionic eminence and regional identity of the septum. *Cell Rep.* **12**, 482–494 (2015).
32. J. S. Hu *et al.*, *Coup-TF1* and *Coup-TF2* control subtype and laminar identity of MGE-derived neocortical interneurons. *Development* **144**, 2837–2851 (2017).
33. U. Borello *et al.*, FGF15 promotes neurogenesis and opposes FGF8 function during neocortical development. *Neural Dev.* **3**, 17 (2008).
34. T. Inoue, M. Ota, M. Ogawa, K. Mikoshiba, J. Aruga, Zic1 and Zic3 regulate medial forebrain development through expansion of neuronal progenitors. *J. Neurosci.* **27**, 5461–5473 (2007).
35. A. N. Rubin *et al.*, The germinal zones of the basal ganglia but not the septum generate GABAergic interneurons for the cortex. *J. Neurosci.* **30**, 12050–12062 (2010).
36. L. Puelles *et al.*, Radial and tangential migration of telencephalic somatostatin neurons originated from the mouse diagonal area. *Brain Struct. Funct.* **221**, 3027–3065 (2016).
37. G. L. McKinsey *et al.*, Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons. *Neuron* **77**, 83–98 (2013).
38. E. L.-L. Pai *et al.*, Mafb and c-Maf have prenatal compensatory and postnatal antagonistic roles in cortical interneuron fate and function. *Cell Rep.* **26**, 1157–1173.e5 (2019).
39. P. Campbell, R. L. Reep, M. L. Stoll, A. G. Ophir, S. M. Phelps, Conservation and diversity of Foxp2 expression in muroid rodents: Functional implications. *J. Comp. Neurol.* **512**, 84–100 (2009).
40. E. Sanchez-Ortiz *et al.*, TrkA gene ablation in basal forebrain results in dysfunction of the cholinergic circuitry. *J. Neurosci.* **32**, 4065–4079 (2012).
41. L. Magno *et al.*, NKX2-1 is required in the embryonic septum for cholinergic system development, learning, and memory. *Cell Rep.* **20**, 1572–1584 (2017).
42. C. H. J. Asbreuk *et al.*, The homeobox genes Lhx7 and Gbx1 are expressed in the basal forebrain cholinergic system. *Neuroscience* **109**, 287–298 (2002).
43. Y. Elshatory, L. Gan, The LIM-homeobox gene Islet-1 is required for the development of restricted forebrain cholinergic neurons. *J. Neurosci.* **28**, 3291–3297 (2008).
44. R. V. Hoch, J. A. Clarke, J. L. R. Rubenstein, Fgf signaling controls the telencephalic distribution of Fgf-expressing progenitors generated in the rostral patterning center. *Neural Dev.* **10**, 8–15 (2015).
45. J. E. Long *et al.*, Dlx-dependent and -independent regulation of olfactory bulb interneuron differentiation. *J. Neurosci.* **27**, 3230–3243 (2007).
46. I. Dunham *et al.*; ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
47. A. S. Nord *et al.*, Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
48. J. P. Reddington *et al.*, Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. *Dev. Cell* **55**, 648–664.e9 (2020).
49. L. Bedford *et al.*, Id4 is required for the correct timing of neural differentiation. *Dev. Biol.* **280**, 386–395 (2005).
50. K. Yun, A. Mantani, S. Garel, J. Rubenstein, M. A. Israel, Id4 regulates neural progenitor proliferation and differentiation in vivo. *Development* **131**, 5441–5448 (2004).
51. M. Levine, A systems view of Drosophila segmentation. *Genome Biol.* **9**, 207 (2008).
52. G. X. Y. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
53. T. Stuart *et al.*, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
54. M. S. Kowalczyk *et al.*, Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
55. P. Angerer *et al.*, destiny: Diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
56. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment analysis for gene ontology (Bioconductor, 2020). https://bioconductor.org/packages/release/bioc/html/topGO.html. Accessed 28 March 2022.
57. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
58. S. Lindtner *et al.*, Genomic resolution of DLX-orchestrated transcriptional circuits driving development of forebrain GABAergic neurons. *Cell Rep.* **28**, 2048–2063.e8 (2019).