ARTICLE

Developmental diversification of cortical inhibitory interneurons

Christian Mayer[1,2,3,4]*, Christoph Hafemeister[2]*, Rachel C. Bandler[1]*, Robert Machold[1], Renata Batista Brito[1,5], Xavier Jaglin[1,3], Kathryn Allaway[1,3], Andrew Butler[2,6], Gord Fishell[1,3,4,7] & Rahul Satija[2,6]

Diverse subsets of cortical interneurons have vital roles in higher-order brain functions. To investigate how this diversity is generated, here we used single-cell RNA sequencing to profile the transcriptomes of mouse cells collected along a developmental time course. Heterogeneity within mitotic progenitors in the ganglionic eminences is driven by a highly conserved maturation trajectory, alongside eminence-specific transcription factor expression that seeds the emergence of later diversity. Upon becoming postmitotic, progenitors diverge and differentiate into transcriptionally distinct states, including an interneuron precursor state. By integrating datasets across developmental time points, we identified shared sources of transcriptomic heterogeneity between adult interneurons and their precursors, and uncovered the embryonic emergence of cardinal interneuron subtypes. Our analysis revealed that the transcription factor Mef2c, which is linked to various neuropsychiatric and neurodevelopmental disorders, delineates early precursors of parvalbumin-expressing neurons, and is essential for their development. These findings shed new light on the molecular diversification of early inhibitory precursors, and identify gene modules that may influence the specification of human interneuron subtypes.

Cortical inhibitory neurons are a diverse population that varies widely in morphology, connectivity and patterns of activity[1]. This group of neurons is developmentally derived from progenitors located in embryonic proliferative zones known as the medial, caudal and lateral ganglionic eminences (MGE, CGE and LGE, respectively)[1]. Although each eminence gives rise to non-overlapping types of interneurons, the genetic programs driving interneuron fate specification and maintenance are not well understood. Diversity is first apparent in the regional expression of a limited number of transcription factors within the ganglionic eminences[2,3]. For example, the transcription factor Nkx2-1 is expressed throughout the entire MGE, but is not expressed in the CGE or LGE[4], whereas the transcription factor Lhx8 is expressed only within a subdomain of the MGE[2]. However, how these early sources of heterogeneity generate the vast diversity of adult interneurons remains unclear, a question that is complicated by the fact that the ganglionic eminences also generate numerous subcortical projection neuron types such as the cholinergic cells of the basal ganglia[5,6].

Here we combine multiple single-cell RNA sequencing (scRNA-seq) approaches with genetic fate-mapping techniques to explore the emergence of cellular heterogeneity during early mouse development. Within mitotic progenitors, we found a highly conserved maturation trajectory, accompanied by eminence-specific transcription factor expression that seeds the emergence of later cell diversity. Alongside the exit from the cell cycle, we reconstructed bifurcations into three distinct precursor states, which were highly correlated across eminences, and included a cortical interneuron ground state. Lastly, guided by the genetic diversity seen in mature populations, we connected the transcriptomic heterogeneity of adult interneurons with their embryonic precursors. Our integrated longitudinal analysis reveals the emergence of interneuron subtype identity during development, and identifies genetic regulators responsible for these fate decisions.

Transcriptional profiling of ganglionic eminence cells

We manually dissected ganglionic eminence cells from wild-type mouse embryos at embryonic day (E)13.5 (for the MGE) or E14.5 (for the CGE and LGE)—time points corresponding to peak neurogenesis in these structures[7,8], which include both dividing mitotic progenitors as well as postmitotic precursor cells (Fig. 1a, Supplementary Table 1). After cell dissociation, we used droplet-based single-cell mRNA sequencing (Drop-seq)[9] to sequence the transcriptomes of 5,622 single cells from the MGE, 7,401 from the CGE and 8,543 from the LGE, from replicate experiments, observing on average 1,626 unique molecular identifiers (UMIs) per cell. We performed latent variable regression to mitigate heterogeneity resulting from cell-cycle state[10,11] (Extended Data Fig. 1)—preventing subsequent analysis from being dominated by mitotic phase-specific gene expression—and filtered out rare contaminating populations of excitatory neurons (*Neurod6*; 2.6% of cells) and endothelial cells (*Igfbp7*; 0.7% of cells) (Fig. 1b, c). The remaining 96.7% of cells were neuronal progenitors and precursors derived from the ganglionic eminences (for example, *Dlx1*; Fig. 1b, c). Within this population, the expression of early, intermediate and late marker genes was strongly associated with the top diffusion map coordinates (DMC; Extended Data Fig. 1). To establish a quantitative temporal account of differentiation programs within each eminence, we fit a principal curve through the DMC, representing an ordered 'maturation trajectory' for single cells based on their expression profiles[12] (Fig. 1d). We obtained very similar trajectories using approaches based on principal component analysis or reverse graph embedding[13] (Extended Data Fig. 1), and observed that the maturation trajectory recapitulated known dynamics associated with neuronal maturation (Fig. 1e) while also segregating ganglionic eminence cells into mitotic and postmitotic phases (Fig. 1f, Extended Data Fig. 1). To independently confirm the association of the maturation trajectory with real time, we used FlashTag technology[14] to fluorescently label cells in the ventricular zone[15] of the

[1]NYU Neuroscience Institute, Langone Medical Center, New York, New York 10016, USA. [2]New York Genome Center, New York, New York 10013, USA. [3]Harvard Medical School, Department of Neurobiology, Boston, Massachusetts 02115, USA. [4]Broad Institute, Stanley Center for Psychiatric Research, Cambridge, Massachusetts 02142, USA. [5]Dominick P Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [6]Center for Genomics and Systems Biology, New York University, New York, New York 10012, USA. [7]Center for Genomics and Systems Biology, New York University, PO Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates.
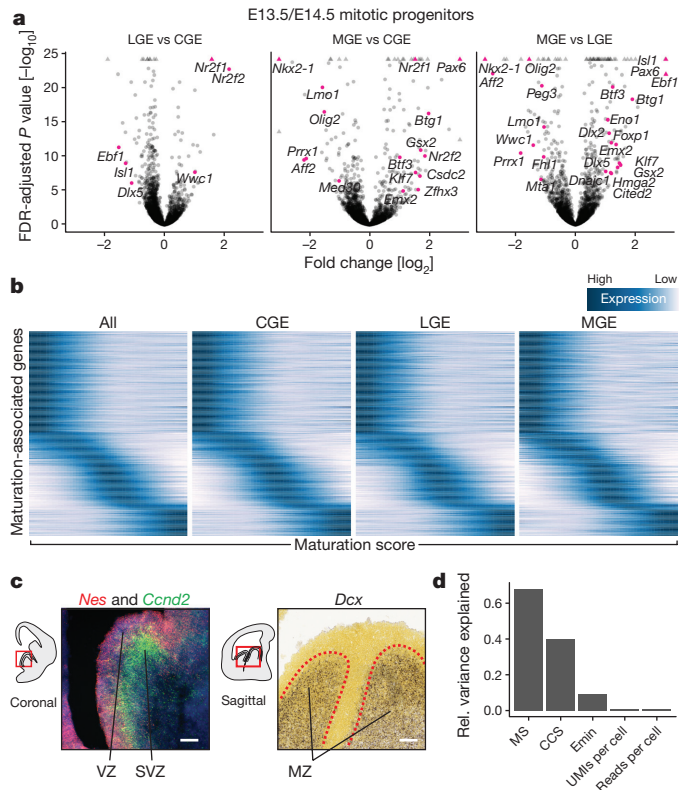*These authors contributed equally to this work.
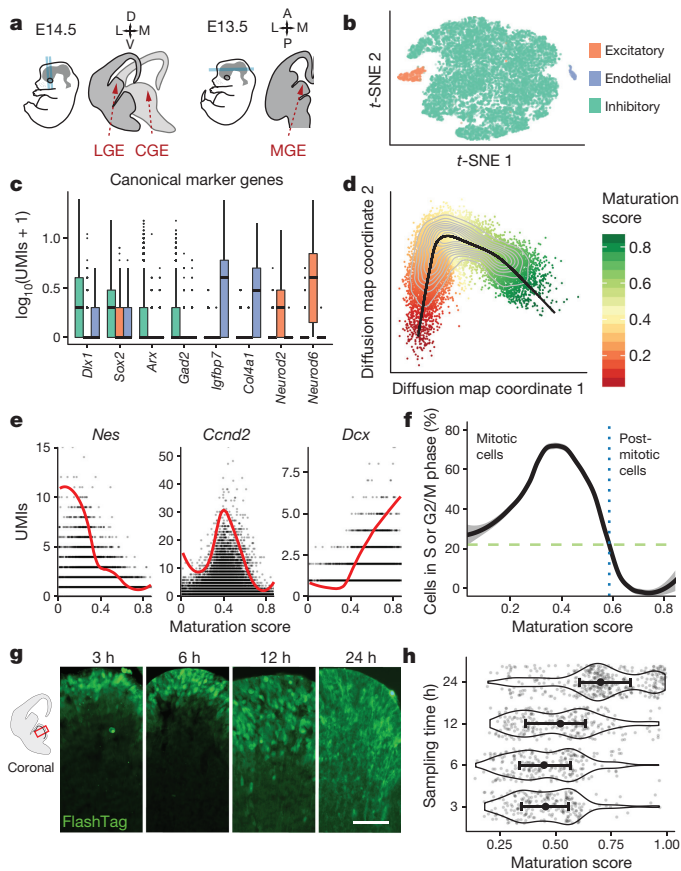
Figure 1 | Transcriptional landscape of single cells in the ganglionic eminences. a, Schematic of experimental workflow. Axes: A, anterior; D, dorsal; L, lateral; M, medial; P, posterior; V, ventral. **b**, Visualization of Drop-seq data from ganglionic eminence cells using *t*-distributed stochastic neighbour embedding. **c**, Canonical marker expression in ganglionic eminence precursors, excitatory neurons and vascular endothelial cells; colours as in **b**. **d**, A principal curve was fitted to the dominant diffusion map coordinates to order cells along a maturation trajectory. **e**, Expression (molecules per cell) of canonical regulators, as a function of the position along the maturation trajectory. The curve reflects local averaging of single-cell expression. Locally averaged values were multiplied by five for visualization on the same scale as the molecule counts. **f**, Percentage of cycling cells as a function of the position along the maturation trajectory; the dotted blue line marks the inferred mitotic-to-postmitotic transition. **g**, Coronal brain sections of the ganglionic eminences, as cells migrate away from the ventricular zone (the apical surface of the ventricular zone is at the top of the images). Images were taken 3, 6, 12 and 24 h after fluorescent labelling with FlashTag technology. Scale bar, 50 μm. **h**, Maturation score distributions of FlashTag labelled cells, separated by time point.

Figure 2 | A common developmental program of gene expression functions in the mitotic progenitors of all three ganglionic eminences. a, Volcano plots depicting differential gene expression across eminences for early mitotic cells (maturation score < 0.3). Transcription factors are annotated. **b**, Gene-expression dynamics in mitotic cells, based on local averaging of single-cell data, plotted along maturation score for 840 developmentally regulated genes that were conserved across eminences. **c**, ISH patterns of early, intermediate and late maturation-trajectory genes in the ganglionic eminences that are highly expressed within anatomical boundaries of the ventricular zone, subventricular zone and mantle zone, respectively. **d**, The variance explained individually by a set of annotated factors, relative to the variance explained by the first principal component. Calculated independently for maturation score (MS), cell cycle score (CCS), eminence of origin (Emin), UMIs per cell and reads per cell.

ganglionic eminences, and performed scRNA-seq on cohorts of 3-, 6-, 12- and 24-h-old neurons as they migrated away from the ventricle (Supplementary Table 1, Fig. 1g). As expected, neurons generated at these sequential time points were distributed progressively along the maturation trajectory timeline (Fig. 1h, Extended Data Fig. 1).

The MGE and CGE are known to produce non-overlapping types of cortical interneurons[16]. To identify regionally expressed transcription factors[2,3,17], we performed a differential expression analysis and found a small number of genes for transcription factors that were enriched in mitotic progenitors within particular eminences (Fig. 2a, Extended Data Fig. 2, Supplementary Table 2), many of which (for example, *Nr2f1*, *Nr2f2*, *Nkx2-1*) have previously been characterized[6]. Next, we identified the sequential patterns of gene expression characterizing the initial stages of cell differentiation. The majority of dynamically expressed genes followed robust and highly reproducible sequential waves of gene expression in all three eminences (Fig. 2b, Extended Data

Fig. 2d, Supplementary Table 3). *In situ* hybridization (ISH) confirmed that these waves describe the sequential expression of stem-cell (for example, *Nes*), proneural (for example, *Ascl1*) and neurogenic genes (for example, *Dcx*), approximately correlating with the spatiotemporal progression from the ventricular zone to the mantle zone (Fig. 2c, Extended Data Fig. 3). Developmental progression and cell cycle were the primary sources of transcriptional variance in these progenitors (Supplementary Methods), with maturation proportionally explaining sixfold more variance compared to eminence-of-origin (Fig. 2d).

To detect the potential fate divergence of cells along the maturation trajectory, we bootstrapped the construction of a minimum spanning tree[18] (Fig. 3a, Supplementary Methods), and summarized the combined result using multidimensional scaling. We first observed evidence of clear fate bifurcations as cells become postmitotic, and precursors from all ganglionic eminences branched into distinct precursor states (Fig. 3b, Supplementary Methods). The sequencing of MGE progenitors at substantially higher depth with plate-based scRNA-seq revealed no transcriptomic evidence of similar bifurcations within mitotic cells (Extended Data Fig. 4a–c; Supplementary Table 1). Moreover, when we performed the unsupervised branching analysis only in mitotic progenitors, we found no evidence for the specification of distinct interneuron fates. Instead, consistent with our previous analysis of the maturation trajectory, heterogeneity was driven primarily by maturation state or cell cycle, which may reflect the existence of mitotic
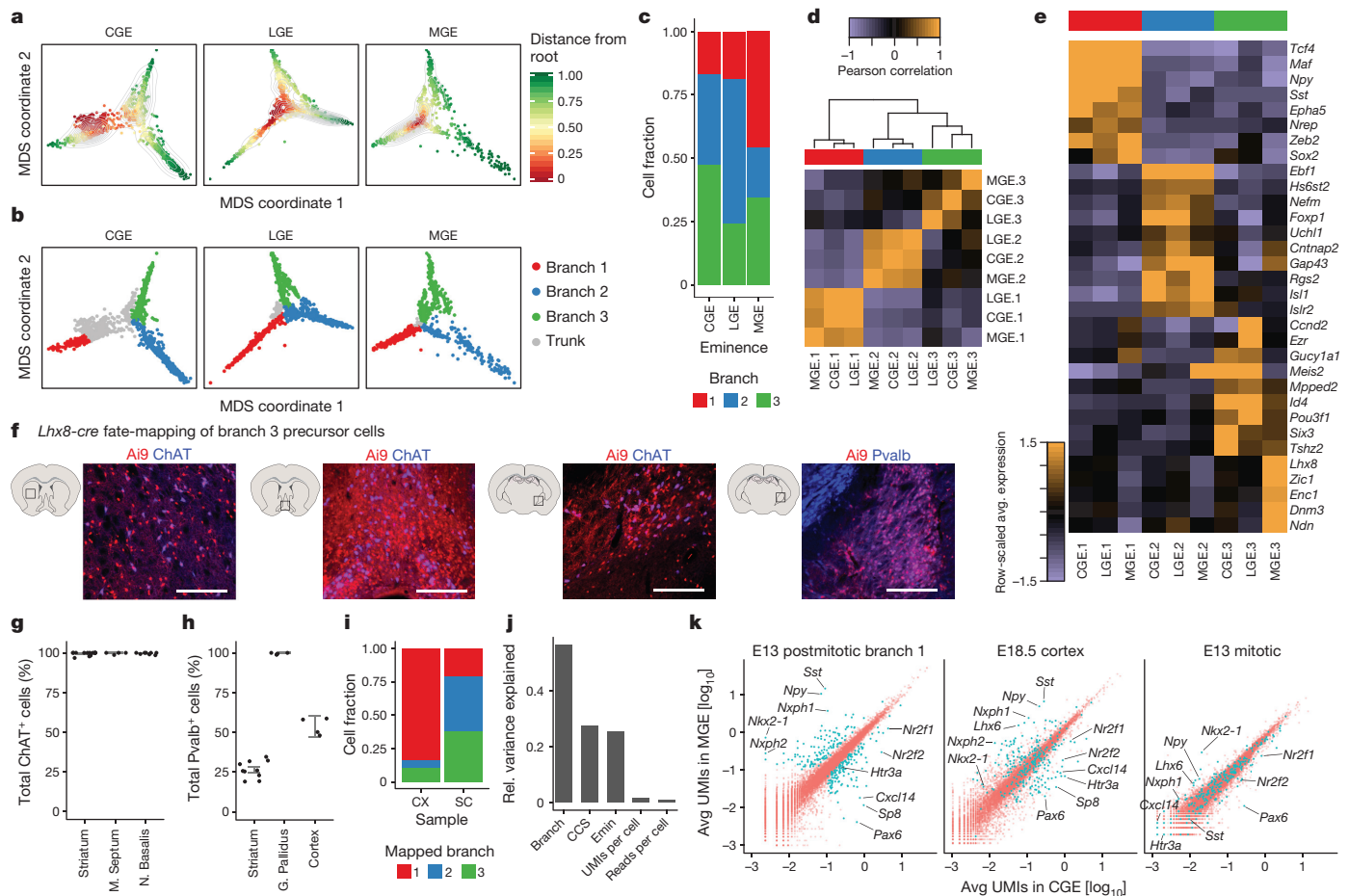
Figure 3 | Postmitotic cells from all eminences pass through distinct precursor states. a, Multidimensional scaling (MDS) based on the average distance along bootstrapped minimum spanning trees. **b**, Minimum spanning tree traversal assigned cells to the trunk and one of three branches. **c**, Quantitative contributions of cells per branch plotted for each ganglionic eminence. **d**, Hierarchical clustering of branch gene expression correlation. Gene expression was averaged for cells from the same ganglionic eminence and branch. **e**, Heat map depicting the top transcriptomic markers for each branch. **f**, Co-localization of Ai9 from *Lhx8-cre;Ai9* mice with, from left to right, choline acetyltransferase (ChAT) in the striatum, medial septum, and nucleus basalis, and Pvalb in the globus pallidus. Scale bars, 300 μm. **g**, Percentage of total ChAT+ cells labelled with tdTomato in *Lhx8-cre;Ai9* mice. $n = 15$ brain sections (striatum), $n = 4$ (medial septum), $n = 8$ (nucleus basalis); 2 mice. **h**, The percentage of total Pvalb+ cells labelled with tdTomato in *Lhx8-cre;Ai9*

mice. $n = 10$ brain sections (striatum), $n = 5$ (globus pallidus), $n = 4$ (cortex), 2 mice. Error bars in **g** and **h** indicate standard deviation across all quantified sections. **i**, Mapping of E18.5 cortical (CX) and subcortical (SC) cells to E13.5 (MGE) or E14.5 (CGE, LGE) branches based on marker gene expression correlations. **j**, Relative variance explained individually by annotated factors for postmitotic cells at E13.5 or E14.5 (branch, CCS, Emin, UMIs per cell and reads per cell) relative to the variance explained by the first principle component. Residual cell cycle variation is due to our conservative cutoff for the mitotic–postmitotic transition. **k**, Differential expression analysis between MGE and CGE postmitotic cells in the interneuron precursor state at E13.5 or E14.5 (left). These genes tend to remain differentially expressed between MGE and CGE-derived populations at E18.5 (middle), which is not the case in E13.5 mitotic progenitors (right); differentially expressed genes are depicted in blue.

progenitors undergoing direct and indirect neurogenesis within the ventricular and subventricular zones[19] (Extended Data Fig. 4d–f). Nonetheless, we cannot fully exclude the possibility of earlier fate-determination in mitotic progenitors.

We assigned cells to branches by traversing the final minimum spanning tree and annotating major splits (Fig. 3b, c). Notably, even though branched trajectories for each eminence were calculated independently, branch gene expression markers were highly correlated across eminences (Fig. 3d, e). This indicates that, although each ganglionic eminence generates different cell populations, upon becoming postmitotic, cells from all eminences pass through conserved precursor states. One group of highly correlated branches (precursor state 1) expressed known regulators of interneuron development (*Arx*, *Maf*; Fig. 3e, Supplementary Table 4), whereas a second group of branches (precursor state 2) expressed known projection neuron marker genes (*Isl1*, *Ebf1*; Fig. 3e, Supplementary Table 4). The third group of branches (precursor state 3) exhibited weaker correlation across eminences, with the transcription factor Lhx8 representing a marker gene for the MGE

branch 3 (Fig. 3e). Genetic fate-mapping using *Lhx8-cre* suggested that neurons within this branch account for the majority, if not all, of the cholinergic projection (nucleus basalis, medial septum) and cholinergic interneuron (striatum) populations, as well as the majority of parvalbumin (Pvalb)-positive projection neurons in the globus pallidus[5,6] (Fig. 3f–h, Extended Data Fig. 4g).

Diversity emerges from a common precursor state

To confirm that cells passing through precursor state 1 give rise to cortical interneurons, we used genetic fate-mapping strategies to enrich for postmitotic cells derived from ganglionic eminences at E18.5 for scRNA-seq (Supplementary Methods, Extended Data Fig. 4). Using a correlation-based distance metric (Supplementary Methods) we found that, as expected[20], more than 80% of *Dlx6a-cre* fate-mapped cortical cells at E18.5 were assigned to precursor state 1, on the basis of their expression of canonical regulators of interneuron development (Fig. 3i, Extended Data Fig. 5). The remaining *Dlx6a-cre* fate-mapped cortical population were assigned to precursor states 2 and 3 (Fig. 3i,
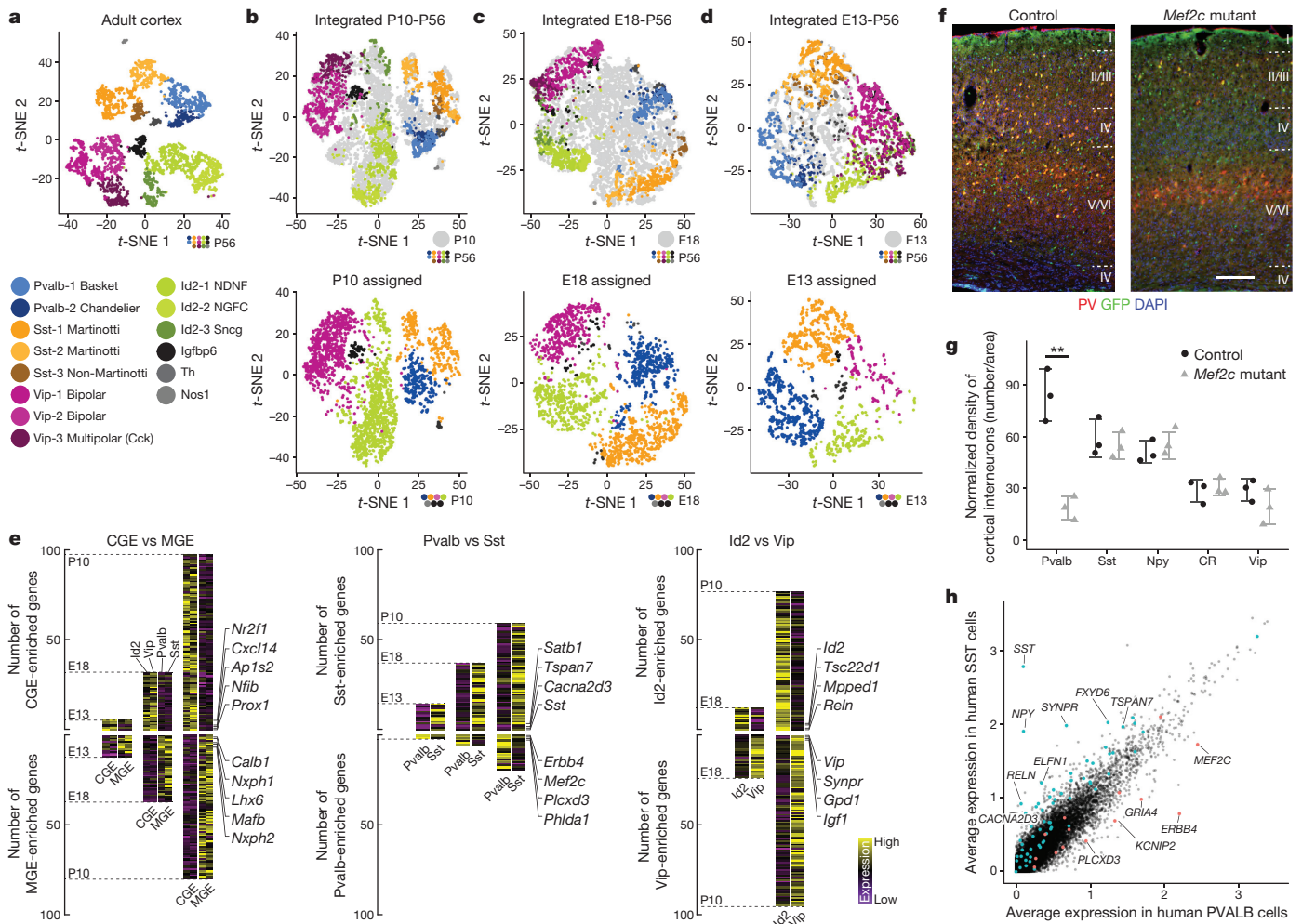
Figure 4 | Integrating developmental scRNA-seq datasets to link embryonic heterogeneity with adult interneuron subtypes. a, Graph-based clustering of interneurons from the adult mouse visual cortex (data from ref. 22 and Allen Cell Types Database, http://celltypes.brain-map.org/download (2015)). Cluster names denote cardinal classes associated with canonical markers (for example Pvalb or Sst), with additional enumerated subdivisions (for example, Vip-1 or Vip-2). **b–d**, Integration of P10 (**b**), E18.5 (**c**), E13.5 (**d**) precursors with P56 cortical interneurons based on shared sources of variation. Top, adult cells coloured by subtype and precursor cells in grey. Bottom, precursor cells coloured by adult cardinal types to which they are assigned; blue, Pvalb; orange, Sst; violet, Vip; green, Id2; shades of gray, Igfbp6, Th or Nos1. **e**, Differentially expressed genes between CGE- and MGE-derived subsets (left); these genes are conserved in both developmental and adult cells. Each conserved gene is placed on the respective heat map when it is first observed to be differentially expressed during development. The same analysis was performed for Pvalb

and Sst subsets (middle), and Vip and Id2 subsets (right). **f**, Conditional deletion of Mef2c in inhibitory neurons using *Dlx6a-cre*;*Mef2c^{loxP/loxP}RCE*. Immunostaining of P20–P22 somatosensory cortex using anti-GFP (green) and anti-Pvalb (red); counterstaining with 4,6-diamidino-2-phenylindole (DAPI) shows cortical layers. Scale bar, 200 μm. **g**, Density quantification of cortical-interneuron subtypes in the P21 somatosensory cortex using antibodies for Pvalb, Sst, Vip, Npy, and calretinin (CR). Error bars reflect s.e.m.; two-tailed unpaired *t*-test, **P < 0.01; n = 3 brains each for cortical knockout and control. Error bars reflect s.e.m.; two-tailed unpaired *t*-test, **P < 0.01; n = 3 brains each for cortical knockout and control. **h**, Scatter plot comparing average expression of GABAergic single nuclei from post-mortem human neurons after segregation into Pvalb and Sst types. Each dot represents the expression of a human gene. Markers of embryonic cardinal types are shown in green or blue dots, with a subset of gene names annotated.

Extended Data Fig. 5), probably including the *Meis2*-expressing CGE-derived GABAergic population that has recently been described[21] (Extended Data Fig. 5). Comparison of the expression profiles of cortical interneuron precursors (precursor state 1) from the MGE and CGE revealed differentially expressed genes, the expression patterns of which are largely maintained in the cortex at later time points (Fig. 3k). Consistently, branching trajectories represented the most important source of variation in these cells, with an increasing contribution attributable to eminence of origin compared to mitotic progenitors (Fig. 3j). Thus, our data reveal how postmitotic pan-eminence transcriptional programs (precursor states) emerge, and in parallel, eminence-specific transcriptional programs escalate.

We next asked when subtype-specific gene expression patterns first appear during interneuron development. In the adult mouse, using a

publicly available dataset[22] (© 2015 Allen Institute for Brain Science, Allen Cell Types Database, available from: http://celltypes.brain-map.org/download) we identified 14 inhibitory interneuron subpopulations that encompass known anatomically and physiologically defined sub-types[23,24] (Fig. 4a, Supplementary Methods, Extended Data Fig. 6). These could be allocated into non-overlapping cardinal types of cortical interneurons (Pvalb, Sst, Vip, Id2, Th, Nos1, Igfbp6). We reasoned that if we could identify heterogeneous gene modules in developing cells that were shared with adult interneurons, we could identify early patterns of specification in precursors. We therefore applied our recently developed tool for the pairwise integration of scRNA-seq datasets[25,26] (Fig. 4b–d), which 'aligns' cell types across datasets based on conserved sources of variation as identified by canonical correlation analysis. This procedure therefore links the heterogeneity observed in

adult cells with heterogeneity in their precursors. On the basis of this alignment, P10 cells exhibited strong evidence of transcriptomic separation beyond cardinal types (Fig. 4b), including clear segregation between Sst Martinotti and non-Martinotti (X94), Vip bipolar and multipolar, and Id2 neurogliaform and non-neurogliaform interneuron subtypes (Fig. 4e, Extended Data Figs 7–9).

Embryonic stages also displayed strong evidence of interneuron specification. On examination of the earliest stages, we observed a separation of Pvalb- and Sst-precursor cells within the E13.5 postmitotic populations (Fig. 4d), and identified transcriptomic markers that were conserved into adulthood (early marker genes for Pvalb neurons: *Mef2c*, *Erbb4*, *Plcxd3*; early marker genes for Sst neurons: *Sst*, *Tspan7*, *Satb1*; Fig. 4e, Extended Data Fig. 7). A minority of E13.5 cells also mapped to Vip and Id2 subsets, but conserved transcriptomic markers did not reach statistical significance until E18.5 (E18.5 markers of Vip neurons: *Vip*, *Synpr*, *Igf1*; E18.5 markers of Id2 neurons: *Reln*, *Mpped1*, *Id2*). By E18.5, all cardinal types of interneurons could be identified, and additional subtypes appeared to be transcriptionally specified as well (Fig. 4e, Extended Data Fig. 9). Notably, segregation into subtypes became evident at different developmental stages. For example, the clear emergence of Sst, Vip and Id2 subtypes was apparent for a subset of cells at E18.5 (Extended Data Fig. 8), but we were unable to clearly subdivide Pvalb neurons by P10, in accordance with their late maturation[27]. The results of our integrated analyses were in agreement with independent unsupervised analysis of each developmental stage (Extended Data Fig. 8). Consistent with our earlier findings (Fig. 2), we did not observe common sources of variation shared between adult interneurons and mitotic progenitors.

In addition to observing the potential specification of embryonic precursors, our list of cardinal type and subtype markers that are conserved from the ganglionic eminences through adulthood suggests a set of genetic regulators that may have important roles in this process. For example, the gene encoding the transcription factor Mef2c was among those that discriminated early Pvalb-precursors from other MGE-derived interneuron types (Fig. 4e). Genome-wide association studies have linked mutations in this gene to Alzheimer's disease, schizophrenia and other neurodevelopmental disorders[28]. Consistent with our predictions, conditional deletion of *Mef2c* in inhibitory neurons led to a specific loss of Pvalb-interneurons by P20 in cortical layers 2–6 (Fig. 4f, g, Extended Data Fig. 10), indicating that *Mef2c* is essential for the generation of this population. Notably, when examining a published single-nucleus RNA-seq dataset of human post mortem tissue[29], we found that a subset of embryonic cardinal type markers from our mouse dataset (including *Mef2c*) was also differentially expressed in adult human interneurons (Fig. 4h). Therefore, the genes we identified as defining embryonic cardinal types are candidates for the regulation of interneuron fate determination and maintenance across species.

Discussion

Our work reveals how subtype-specific heterogeneity progresses from the expression of cardinal genes in progenitors to the emergence of specific subtypes that populate the mature cortex. Postmitotic cells in the ganglionic eminences branch into distinct precursor states, representing populations fated to give rise to interneurons or projection neurons. It seems probable that the superimposition of precursor-state genes and eminence-specific genes act coordinately to bestow the common and unique characteristics within particular GABAergic populations, respectively.

Consequently, precursor genes are likely to direct the developmental cascade and acquisition of general properties that are shared within a given type. This probably ensures, for instance, that interneurons migrate tangentially to the cortex or the hippocampus, whereas projection neurons remain positioned ventrally and form long-range projections. Supplementing these more general programs are the eminence-specific genes that, for example, may direct the axons of

parvalbumin cortical interneurons to form perisomal baskets and the efferents of somatostatin cortical interneurons to reliably target dendrites. These distinct differentiation modules reflect the major cardinal types of cortical interneuron precursors.

The identification of early precursors offers insight into how specific cell types emerge and provides genetic access to immature cortical interneuron subtypes. To broaden the implications of these results, our findings indicate that components of the transcriptional networks underlying interneuron fate specification are conserved between mouse and human, including *Mef2c* and other genes associated with neuropsychiatric disorders. This highlights the power of combining single-cell genomics with analytical tools to identify genes that have important functional roles in the establishment and maintenance of interneuron fates. Our findings mark an initial but important step towards the goal of ultimately linking specific genes to their aetiology in neurodevelopmental and neuropsychiatric disorders.

1. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
2. Flames, N. *et al.* Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).
3. Yun, K., Garel, S., Fischman, S. & Rubenstein, J. L. R. Patterning of the lateral ganglionic eminence by the *Gsh1* and *Gsh2* homeobox genes regulates striatal and olfactory bulb histogenesis and the growth of axons through the basal ganglia. *J. Comp. Neurol.* **461**, 151–165 (2003).
4. Shimamura, K., Hartigan, D. J., Martinez, S., Puelles, L. & Rubenstein, J. L. Longitudinal organization of the anterior neural plate and neural tube. *Development* **121**, 3923–3933 (1995).
5. Nóbrega-Pereira, S. *et al.* Origin and molecular specification of globus pallidus neurons. *J. Neurosci.* **30**, 2824–2834 (2010).
6. Zhao, Y. *et al.* The LIM-homeobox gene *Lhx8* is required for the development of many cholinergic neurons in the mouse forebrain. *Proc. Natl Acad. Sci. USA* **100**, 9005–9010 (2003).
7. Inan, M., Welagen, J. & Anderson, S. A. Spatial and temporal bias in the mitotic origins of somatostatin- and parvalbumin-expressing interneuron subgroups and the chandelier subtype in the medial ganglionic eminence. *Cereb. Cortex* **22**, 820–827 (2012).
8. Miyoshi, G. *et al.* Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *J. Neurosci.* **30**, 1582–1594 (2010).
9. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
10. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
11. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
12. Petropoulos, S. *et al.* Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016); erratum 165, 1012–1026, (2016).
13. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
14. Quah, B. J. C. & Parish, C. R. The use of carboxyfluorescein diacetate succinimidyl ester (CFSE) to monitor lymphocyte proliferation. *J. Vis. Exp.* https://doi.org/10.3791/2259 (2010).
15. Telley, L. *et al.* Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**, 1443–1446 (2016).
16. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
17. Waclaw, R. R., Ehrman, L. A., Pierani, A. & Campbell, K. Developmental origin of the neuronal subtypes that comprise the amygdalar fear circuit in the mouse. *J. Neurosci.* **30**, 6944–6953 (2010).
18. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
19. Petros, T. J., Bultje, R. S., Ross, M. E., Fishell, G. & Anderson, S. A. Apical versus basal neurogenesis directs cortical interneuron subclass fate. *Cell Rep.* **13**, 1090–1095 (2015).
20. Wichterle, H., Turnbull, D. H., Nery, S., Fishell, G. & Alvarez-Buylla, A. *In utero* fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. *Development* **128**, 3759–3771 (2001).

21. Frazer, S. *et al.* Transcriptomic and anatomic parcellation of 5-HT$_{3A}$R expressing cortical interneuron subtypes revealed by single-cell RNA sequencing. *Nat. Commun.* **8,** 14219 (2017).
22. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19,** 335–346 (2016).
23. Tremblay, R., Lee, S. & Rudy, B. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91,** 260–292 (2016).
24. Petilla Interneuron Nomenclature Group. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9,** 557–568 (2008).
25. Butler, A. & Satija, R. Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. Preprint at https://www.biorxiv.org/content/early/2017/07/18/164889 (2017).
26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4096 (2018).
27. Alcántara, S., de Lecea, L., Del Río, J. A., Ferrer, I. & Soriano, E. Transient colocalization of parvalbumin and calbindin D28k in the postnatal cerebral cortex: evidence for a phenotypic shift in developing nonpyramidal neurons. *Eur. J. Neurosci.* **8,** 1329–1339 (1996).
28. Harrington, A. J. *et al.* MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *eLife* **5,** 140 (2016).
29. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14,** 955–958 (2017).

Supplementary Information is available in the online version of the paper.

Author Contributions C.M., C.H., R.C.B., G.F. and R.S. conceived the research. C.M. and R.C.B. led experimental work, assisted by R.M., R.B.B., X.J., K.A. and supervised by G.F. C.H. led computational analysis, assisted by C.M. and A.B., and supervised by R.S. All authors participated in interpretation and writing the manuscript.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Animals. All mouse colonies were maintained in accordance with protocols approved by the Institutional Animal Care and Use Committee at the NYU School of Medicine. Mouse strains used are the following: wild-type Swiss Webster females (Taconic Biosciences), *Dlx6a(Tg)-cre* (ref. 30), *Lhx6(BAC)-GFP* (GENSAT), *Lhx8-cre/cerulean* (Jax stock #023453), *Rosa26^{LSL-tdTomato}* (*Ai9*) (ref. 31), *Rosa26* (*CAG*)-*LSL-eGFP* (*RCE^{loxP}*) (Jax stock #32037) (ref. 32), *Dlx5/6(Tg)-cre/eGFP* (Jax stock #023724; for characterization information, see images at the Allen Institute for Brain Science website (http://connectivity.brain-map.org/transgenic/imageseries/list/1.html?gene_term=Dlx5-CreERT2)). and *Mef2c^{fl/fl}* (ref. 33). Both male and female mice were used for all single-cell RNA sequencing experiments.

Wild-type Drop-seq experiments. Mouse embryos at 13.5 (MGE) and 14.5 (CGE and LGE) days' gestation were isolated from 6–8-week-old wild-type Swiss Webster timed-pregnant dams ordered from Taconic Biosciences. Embryos were staged in days post coitus, with embryonic day (E) 0.5 defined as noon of the day a vaginal plug was detected after overnight mating. The method of euthanasia for pregnant dams was inhaled-isoflurane overdose, and death was confirmed with decapitation. Surgical access to the uterine horns enabled removal of embryos. After removal from the mother, embryos were stored on ice in Leibovitz's L-15 medium and 1% fetal bovine serum. Brains were removed from the embryos and embedded in 1% ultrapure low melting point agarose and sectioned in 50-µm sections with a vibratome (Leica VT1200S). The MGE, CGE or LGE were dissected from each embryo. MGEs were dissected from horizontal brain sections, whereas CGEs and LGEs were dissected from coronal brain sections. Tissue from several embryos was pooled together before dissociation.

Single-cell dissociation. Embryonic brain tissue pooled from several embryos was dissociated into a single-cell suspension using a papain dissociation system (Worthington Biochemical) according to the manufacturer's instructions. Postnatal brain tissue was dissociated with 1 mg ml^{-1} of pronase (Roche, #10 165 921 001) in ice-cold prebubbled artificial cerebrospinal fluid for 25 min.

Fluorescence-activated cell sorting. Fluorescent cells from *Dlx6a-cre;Ai9*, *Dlx6a-cre;RCE^{loxP}*, *Lhx6(BAC)-GFP* and CellTrace injected (FlashTag) brain tissue were sorted on a Sony SY3200 sorter with a 100-µm nozzle. Cells were sorted in bulk for experiments using Drop-seq and the 10x Genomics platform, whereas for experiments using plate based scRNA-seq methods, single cells were sorted into 96-well plates and immediately frozen on dry ice.

Single-cell RNA sequencing and library preparation. Drop-seq was run on single cells according to the Online Dropseq Protocol v.3.1 (December 2015) and the methods published in ref. 9. Drop-seq flow rates (oil: 6,000 µl h^{-1}, cells: 2,000 µl h^{-1}, beads: 2,000 µl h^{-1}) were optimized based on human–mouse species mixing experiments with a 1–2% doublet rate. Libraries were prepared with the Nextera XT DNA Library Preparation Kit according to the manufacturer's instructions. For experiments using the 10x Genomics platform, the 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.

For single cells sorted into 96-well plates, cells were immediately lysed and mRNAs were released when single cells were sorted into wells with 5× Maxima reverse transcription buffer, dNTP mixture, RNase inhibitors (SUPERase In RNase Inhibitor, Thermo Fisher Scientific #AM2696) and water. We reverse-transcribed the mRNAs using Superscript II Reverse Transcriptase (Thermo Fisher Scientific #18064071), and amplified cDNAs for each cell in individual wells using the Smart-seq2 (ref. 34) protocol, with a custom modification in which a 12-base cell barcode was included in the 3′-end reverse transcriptase primer. This enabled us to perform multiplexed pooling before library preparation with the Nextera XT DNA sample prep kit (Illumina), and returned 3′ biased data similar to the Drop-seq protocol. We quantified the cDNA libraries on an Agilent BioAnalyzer and sequenced them on a HiSeq 2500.

FlashTag. Immediately before use, 10 mM CFSE (Life Technologies, #C34554) CellTrace solution was prepared according to the manufacturer's instructions, and 2–3 µl was injected into the lateral ventricle of E12.5 or E13.5 wild-type mouse embryos. Embryos were collected 3, 6, 12 and 24 h post-injection.

Fluorescent *in situ* hybridization. Fluorescent *in situ* hybridization (ISH) for nestin (*Nes*) and cyclin D2 (*Ccnd2*) transcripts (Fig. 2d) was performed as previously described[35]. Antisense cRNA probes were prepared by T7 polymerase *in vitro* transcription of PCR product templates generated using the following primers: Nestin, 5′-AGCAGTGCCTGGAAGTGGAAG-3′ and 5′-GCACATTAATACGACTCACTATAGGGCTGGATCCCCTCAGCTTGG-3′; Cyclin D2, 5′-ACCTCCCGCAGTGTTCCTA-3′ and 5′-AATTAATACGACTCACTATAGGCTGCTCTTGACGGAACTGCT-3′

Immunohistochemistry. Lhx8 *fate-mapping.* To fate-map *Lhx8*-lineage neurons, *Lhx8-cre/cerulean* mice were crossed with *Ai9* mice. Offspring with both alleles were then transcardially perfused with PBS followed by 4% paraformaldehyde (PFA) in PBS at postnatal day 21. The brains were collected, fixed overnight in 4% PFA at 4 °C, embedded in 4% agarose, and sectioned at 50 µm on a Leica VT1200S vibratome before proceeding with immunohistochemistry. *Lhx8-cre/cerulean;Ai9* brain sections were blocked for 1 h in 10% normal donkey serum, 0.3% Triton-X. Sections were incubated at 4 °C overnight in the following primary antibodies: goat anti-ChAT (1:250, Millipore AB144P), goat anti-Pvalb (1:1000, Swant PVG213), rabbit anti-SST (1:3000, Peninsula Labs T-4102), rabbit anti-GABA (1:2000, Sigma A2052), chicken anti-GFP (Aves Labs 1020), rabbit anti-DsRed (Clontech 632496). Following several washes in PBS, sections were incubated at room temperature for 1 h in the following secondary antibodies: Alexa Fluor 488-donkey anti-chicken, Alexa Fluor 594-donkey anti-rabbit, and Alexa Fluor 647-donkey anti-goat and rabbit (all Jackson Immunoresearch). Sections were then washed again several times in PBS, treated with DAPI as a counterstain, and mounted on slides. Sections were imaged using a Zeiss Axioimager A2 and processed in ImageJ.

Mef2c *conditional inactivation*. We conditionally inactivated *Mef2c* by crossing the *Mef2c^{fl/fl}* allele with either a *Dlx6a-cre* or *Dlx5/6-cre* driver line. As the recombination mediated by these two driver lines is indistinguishable, they are used interchangeably in Fig. 4f, g, respectively. This results in the deletion of *Mef2c* in all cortical interneurons during embryogenesis, shortly after they become postmitotic. We took advantage of the *RCE^{loxP}* reporter line, which upon Cre-mediated recombination enables the expression of GFP.

Density of cortical-interneuron subtypes in *Mef2c* conditional mutants. Brains from *Dlx5/6-cre;Mef2c^{loxP/loxP}RCE* mice were fixed by transcardial perfusion with 4% PFA in PBS followed by a 1-h post-fixation period on ice with 4% PFA/PBS solution. Brains were rinsed with PBS and cryoprotected using 15% sucrose/PBS solution for 6 h and 30% sucrose/PBS solution overnight at 4 °C. Tissues were embedded in Tissue-Tek, frozen on dry ice, and cryosectioned at thicknesses of 20 µm. Sections for immunohistochemistry analysis were processed using 1.5% normal goat serum and 0.1% Triton X-100 in all procedures except for washing steps, in which only PBS was used. Sections were blocked for 1 h, followed by incubation with the primary antibodies overnight at 4 °C. Cryostat tissue sections were stained with the primary antibodies rat anti-SST (1:250, Chemicon), mouse anti-Pvalb (1:1,000, Sigma), and rabbit anti-VIP (1:250, ImmunoStar), rabbit anti-Neuropeptide Y (1:500; Incstar), mouse anti-Calretinin (1:1,500; Chemicon). Secondary antibodies conjugated with Alexa fluorescent dyes were applied for 1 h at room temperature to visualize the signals. Nuclear counterstaining was performed with DAPI solution. All analysis was evaluated in the somatosensory cortex. Density quantification for cortical-interneuron subtypes was calculated as number of (peptide marker) + expressing cortical interneurons/area in the P21 control and *Mef2c* cortical knockout somatosensory cortex. To minimize counting bias, we compared sections of equivalent bregma positions, defined according to the Mouse Brain Atlas[36]. The total number of cells expressing the marker were counted for a defined and normalized optical area. Three brains each were used for knockout and control.

Layer distribution of Pvalb-expressing cortical interneurons in *Mef2c* conditional mutants. Tissue from conditional *Mef2c* mutants (*Dlx6a^{cre};Mef2c^{fl/fl}*) and controls (*Dlx6a^{cre};Mef2c^{fl/+}*) were analysed at P20–P21. Adult mice were transcardially perfused with 4% PFA after being anaesthetized by intraperitoneal administration of Sleepaway. Brains that were processed for immunofluorescence on slides were post fixed in 4% PFA in PBS at 4 °C and cryopreserved following the perfusion and brain collection. 16-µm coronal sections were obtained using Cryostat (Leica Biosystems) and collected on super-frost coated slides, then allowed to dry and stored at −20 °C until use. For immunofluorescence, cryosections were thawed and allowed to dry for 5–10 min and rinsed twice in 1× PBS. They were incubated at room temperature in a blocking solution of PBST (PBS, 0.1% Triton X-100) and 10% normal donkey serum (NDS) for 60 min, followed by incubation with primary antibodies in PBST and 1% NDS at 4 °C overnight. Primary antibodies are as follows: anti-GFP chicken polyclonal IgY (1:1,000) (Abcam Ab13970), anti-parvalbumin (Pvalb) goat (Swant PVG 213) and anti-parvalbumin (Pvalb) rabbit (Swant PV25). Samples were then washed three times with PBST and incubated with fluorescence conjugated secondary Alexa antibodies (Life Technologies) in PBST with 1% NDS at room temperature for 60–90 min. Slides were then incubated for 30 s with DAPI, washed three times with PBST and once with PBS. Finally, slides were mounted with Fluoromount G (Southern Biotech) and imaged. To quantify the layer distribution and density of various populations of cortical interneurons, the proportion of interneurons of given subtypes over the total number of fate-mapped interneurons across cortical layers was manually determined in ImageJ. Percentages presented in Fig. 4 were calculated by dividing the number of marker-positive neurons in each layer (for example, layer I, layer II/III, layer IV and layer V/VI) by the total number of reporter-positive neurons. Percentages were compared with

repeated *t*-tests in GraphPad Prism, and means ± s.d. are represented. Three brains each were used for knockout and control, 3–4 sections per brain. See Extended Data Fig. 10a.

Single-cell RNA-seq data processing. The raw Drop-seq data was processed with the standard pipeline (Drop-seq tools version 1.12 from McCarroll laboratory). Reads were aligned to the ENSEMBL release 84 *Mus musculus* genome. 10x Genomics data was processed using the same pipeline as for Drop-seq data, adjusting the barcode locations accordingly.

Data filtration. We selected cells for downstream processing in each Drop-seq run, using the quality control metrics output by the Drop-seq tools package[9], as well as metrics derived from the UMI matrix. 1) We first removed cells with a low number (<700) of unique detected genes. From the remaining cells, we filtered additional outliers. 2) We removed cells for which the overall alignment rate was less than the mean minus three standard deviations. 3) We removed cells for which the total number of reads (after $\log_{10}$ transformation) was not within three standard deviations of the mean. 4) We removed cells for which the total number of unique molecules (UMIs, after $\log_{10}$ transformation) was not within three standard deviations of the mean. 5) We removed cells for which the transcriptomic alignment rate (defined by PCT_USABLE_BASES) was not within three standard deviations of the mean. 6) We removed cells that showed an unusually high or low number of UMIs given their number of reads by fitting a loess curve (span = 0.5, degree = 2) to the number of UMIs with number of reads as predictor (both after $\log_{10}$ transformation). Cells with a residual more than three standard deviations away from the mean were removed. 7) With the same criteria, we removed cells that showed an unusually high or low number of genes given their number of UMIs. Of these filter steps, step 1 removed the majority of cells. Steps 2 to 7 removed only a small number of additional cells from each eminence (2% to 4%), and these cells did not exhibit unique or biologically informative patterns of gene expression.

Data normalization. The raw data per Drop-seq run is a UMI count matrix with genes as rows and cells as columns. The values represent the number of UMIs that were detected. The aim of normalization is to make these numbers comparable between cells by removing the effect of sequencing depth and biological sources of heterogeneity that may confound the signal of interest, in our case cell cycle stage.

A common approach to correct for sequencing depth is to create a new normalized expression matrix x with $x_{i,j} = \log\left(\frac{c_{i,j} \times 10,000}{m_j}\right)$, in which $c_{i,j}$ is the molecule count of gene i in cell j and m_j is the sum of all molecule counts for cell j. This approach assumes that $c_{i,j}$ increases linearly with m_j, which is true only when the set of genes detected in each cell is roughly the same. However, for Drop-seq, in which the number of UMIs is low per cell compared to the number of genes present, the set of genes detected per cell can be quite different. Hence, we normalize the expression of each gene separately by modelling the UMI counts as coming from a generalized linear model with negative binomial distribution, the mean of which can be dependent on technical factors related to sequencing depth. Specifically, for every gene we model the expected value of UMI counts as a function of the total number of reads assigned to that cell, and the number of UMIs per detected gene (sum of UMI divided by number of unique detected genes). To solve the regression problem, we use a generalized linear model (glm function of base R package) with a regularized overdispersion parameter theta. Regularizing theta helps us to avoid overfitting which could occur for genes whose variability is mostly driven by biological processes rather than sampling noise and dropout events. To learn a regularized theta for every gene, we perform the following procedure.

1) For every gene, obtain an empirical theta using the maximum likelihood model (theta.ml function of the MASS R package) and the estimated mean vector that is obtained by a generalized linear model with Poisson error distribution.

2) Fit a line (loess, span = 0.33, degree = 2) through the variance–mean UMI count relationship (both $\log_{10}$ transformed) and predict regularized theta using the fit. The relationship between variance and theta and mean is given by variance = mean + (mean²/theta).

Normalized expression is then defined as the Pearson residual of the regression model, which can be interpreted as the number of standard deviations by which an observed UMI count is higher or lower than its expected value. Unless stated otherwise, we clip expression to the range $[-30, 30]$ to prevent outliers from dominating downstream analyses.

Removal of cell cycle effect. The normalization method described above aims to reduce the effect of technical factors in scRNA-seq data (primarily, depth) from downstream analyses. However, heterogeneity in cell cycle stage, particularly among mitotic cells transitioning between S and G2/M phases, also can drive substantial transcriptomic variation that can mask biological signal. To mitigate this effect, we use a two-step approach: 1) quantify cell cycle stage for each cell using supervised analyses with known stage-specific markers, 2) regress the effect of cell cycle stage using the same negative binomial regression as outlined above. For the first step we use a previously published list of cell cycle dependent genes (43 S phase genes, 54

G2/M phase genes) for an enrichment analysis similar to that proposed in ref. 11. For each cell, we compare the sum of phase-specific gene expression ($\log_{10}$ transformed UMIs) to the distribution of 100 random background genes sets, where the number of background genes is identical to the phase gene set, and the background genes are drawn from the same expression bins. Expression bins are defined by 50 non-overlapping windows of the same range based on $\log_{10}$(mean UMI). The phase-specific enrichment score is the expression *z*-score relative to the mean and standard deviation of the background gene sets. Our final 'cell cycle score' (Extended Data Fig. 1) is the difference between S-phase score and G2/M-phase score.

For a final normalized dataset with cell cycle effect removed, we perform negative binomial regression with technical factors and cell cycle score as predictors. Although the cell cycle activity was regressed out of the data for downstream analysis, we stored the computed cell cycle score before regression, enabling us to remember the mitotic phase of each individual cell. Notably, our regression strategy is tailored to mitigate the effect of transcriptional heterogeneity within mitotic cells in different phases, and should not affect global differences between mitotic and non-mitotic cells that may be biologically relevant.

Dimensionality reduction. Throughout the manuscript we use diffusion maps, a non-linear dimensionality reduction technique[37]. We calculate a cell-to-cell distance matrix using 1 − Pearson correlation and use the diffuse function of the diffusionMap R package with default parameters to obtain the first 50 DMCs. To determine the significant DMCs, we look at the reduction of eigenvalues associated with DMCs. We determine all dimensions with an eigenvalue of at least 4% relative to the sum of the first 50 eigenvalues as significant, and scale all dimensions to have mean 0 and standard deviation of 1.

Initial clustering of all cells. To identify contaminating cell populations and assess overall heterogeneity in the data, we clustered all single cells. We first combined all Drop-seq samples and normalized the data (21,566 cells, 10,791 protein-coding genes detected in at least 3 cells and mean UMI at least 0.005) using regularized negative binomial regression as outlined above (correcting for sequencing depth related factors and cell cycle). We identified 731 highly variable genes; that is, genes for which the *z*-scored standard deviation was at least 1. We used the variable genes to perform dimensionality reduction using diffusion maps as outlined above (with relative eigenvalue cutoff of 2%), which returned 10 significant dimensions. For clustering we used a modularity optimization algorithm that finds community structure in the data with Jaccard similarities (neighbourhood size 9, Euclidean distance in diffusion map coordinates) as edge weights between cells[38]. With the goal of overclustering the data to identify rare populations, the small neighbourhood size resulted in 15 clusters, of which two were clearly separated from the rest and expressed marker genes expected from contaminating cells (*Neurod6* from excitatory neurons, *Igfbp7* from epithelial cells). These cells represent rare cellular contaminants in the original sample (2.6% and 1%), and were excluded from further analysis, leaving 20,788 cells.

Identifying a maturation trajectory. To assign each cell a maturation score that is proportional to the developmental progress, we first performed dimensionality reduction as described above using all genes that were detected in at least 2% of the cells (8,014 genes). This resulted in four significant dimensions. We then fit a principal curve (R package princurve, smoother = 'lowess', $f = 1/3$) through the data. The maturation score of a cell is then the arc-length from the beginning of the curve to the point at which the cell projects onto the curve. The resulting curve is directionless, so we assign the 'beginning' of the curve so that the expression of *Nes* is negatively correlated with maturation. *Nes* is a known ventricular zone marker and therefore should only be highly expressed early in the trajectory. Maturation scores are normalized to the interval [0, 1]. In an independent analysis, we also used Monocle2 to order cells along a pseudo-time. We used Monocle version 2.3.6 with expression response variable set to negative binomial. We estimated size factors and dispersion using the default functions. For ordering cells, we reduced the set of genes based on results of the monocle dispersion Table function, and only considered 718 genes with mean expression ≥ 0.01 and an empirical dispersion at least twice as large as the fitted dispersion. Dimensionality reduction was carried out using the default method (DDRTree).

Defining mitotic and post mitotic populations. We observed a sharp transition point along the maturation trajectory at which cells uniformly transitioned into a postmitotic state, corresponding to the loss of proliferation potential and exit from the cell cycle (Fig. 1f, Extended Data Fig. 1). We therefore subdivided the maturation trajectory into a mitotic and postmitotic phase to facilitate downstream analyses. We defined cells with a high phase-specific enrichment score (score >2, see section 'Removal of cell cycle effect') as being in the S or the G2/M phase. We then fitted a smooth curve (loess, span = 0.33, degree = 2) to number of cells in S, G2/M phases as a function of maturation score. The point where this curve falls below half the global average marks the dividing threshold (Fig. 1f).

Smoothed expression for visualization. Although all statistical analyses (differential expression, branch detection, etc.) were performed on single-cell data (UMI counts or

normalized expression), we created smoothed expression estimates for visualization in Fig. 2b, c, and generated these by first fitting a loess curve (span = 0.5, degree = 2) to the normalized expression of each gene with maturation score as predictor. We then predicted values at 100 points at regular intervals from 0 to 1.

Identifying developmentally regulated genes. To identify genes that are developmentally regulated during the mitotic phase, we used mutual information between expression and maturation score.

We selected all cells that we determined to be mitotic (as defined above, Fig. 1f, Extended Data Fig. 1) but ignored the 1% of cells with the lowest maturation score to be more robust against outliers. We discretized maturation scores by placing each cell into one of 13 equal sized bins, and did the same for the expression of each gene. We then calculated mutual information between each gene and the maturation score. We also calculated a random background mutual information (rbMI) distribution for each gene using shuffled maturation scores. These rbMI values allowed us to z-score the mutual information values by subtracting the mean of rbMI and then dividing by the standard deviation of rbMI. We determined 1,294 genes to be highly developmentally regulated (z-score > 20).

We call 840 of these genes highly conserved, because all eminence-specific expression fits have a Pearson correlation >0.9 with the fitted values obtained using the combined data (Fig. 2b).

Differential expression. We wanted to identify genes that were differentially expressed in the early mitotic cells between the eminences (Fig. 2a, Supplementary Table 3), and also between cells assigned to different branches (Fig. 3e, Supplementary Table 4). As has previously been observed[39], expression values in scRNA-seq are overdispersed, and we model expression values as drawn from a negative binomial distribution. Concordant with our model for data normalization, our test is based on the same negative binomial regression model with regularized overdispersion parameter. For a gene i and its vector of UMI counts c_i and a group indicator variable g, we fit the two models:

$$\text{Model 1}: \log(E(c_i)) = \alpha + \beta_1 r + \beta_2 m + \beta_3 c + \varepsilon$$

$$\text{Model 2}: \log(E(c_i)) = \alpha + \beta_1 r + \beta_2 m + \beta_3 c + \beta_4 g + \varepsilon$$

with the technical factors r, the total number of reads per cell, and m, the average number of molecules per gene per cell, and the biological factor c, the cell cycle score. The overdispersion parameter theta is determined using model 1 and used for both models. The comparison of the two models using a likelihood ratio test determines the P value of model 2 providing a better fit. The log-fold change is directly given by the coefficient of the group indicator variable as $\log_2(e^{\beta_4})$. We called genes differentially expressed if the adjusted P value (false discovery rate, FDR) is smaller than 10^{-4} and the absolute fold change is larger than 1.

Branch analysis. To check for emerging heterogeneity in the cells, we focused on the postmitotic cells and performed a trajectory analysis that allows for branching, that is, one population of cells may give rise to multiple precursors. Minimum spanning trees (MSTs) have been used previously to identify putative branching structures in developing populations[18,40]. However, spurious edges in the MST or similar graph structures, previously referred to as short circuits, can introduce stochasticity into these analyses. To overcome this problem, we applied a bootstrapped approach using an ensemble of graphs, an approach inspired by the Wanderlust algorithm[41], which also constructs an ensemble of graphs to gain robustness to short circuits. We repeatedly constructed MSTs based on subsamples of the data, and combined their results to obtain a new cell-to-cell distance matrix and final tree structure that connects all cells. Branches are determined by traversing the final tree and identifying major splits.

We performed this analysis separately for each eminence, and input were the expression data of cells that we considered postmitotic based on maturation score and cell cycle score (1,992 CGE cells, 1,750 LGE cells, 1,271 MGE cells). For the analysis, we normalized the expression data as described above, regressing out the following factors: number of reads, molecules per gene, and cell cycle score. We considered all genes that were detected in at least 2% of the cells and that were highly variable, that is, had a z-scored standard deviation larger than 1 (997 genes in CGE, 954 in LGE, 1,017 in MGE). We carried out dimensionality reduction as outlined above. We performed multiple runs of constructing an MST using a random set of 66% of all cells each time. The process was repeated until each pair of cells has been sampled at least 30 times (number of bootstraps for CGE, LGE, MGE: 91, 89, 89). We combined the MSTs by averaging the cell-to-cell distances along the tree structures, followed by MDS to two dimensions. In theory, the MDS was not necessary, but in practice we observed a more robust final consensus MST on the MDS coordinates than on the averaged cell-to-cell distances. The consensus MST is given directionality by choosing a cell with a low maturation score as the root. Instead of using the cell with the lowest maturation score, an approach that can

be sensitive to outliers, we chose the cell that maximizes the correlation between maturation score and distance from the root for all cells.

To determine significant branches, we traversed the tree starting from the root. Any cell that has outgoing edges to two or more cells is a potential branch point. However, we considered branches significant only when the number of cells in the branch was at least 8% of all cells in the MST. We marked only cells that are connected to two or more significant branches as branch points. Only terminal branches, that is, tree segments not containing branch points, are considered for further analysis.

Mapping E14.5 fate-mapped 10× scRNA-seq data to branches. *Lhx6-GFP* mice were crossed with wild-type Swiss Webster females. At E14.5, the MGE and CGE were each dissected from transgenic embryos and dissociated into a single-cell suspension. FACS sorting was used to collect GFP-positive cells from the MGE and GFP-negative cells from the CGE. Single-cell libraries were prepared using the 10x Genomics pipeline. A total of 12,513 cells passed our initial filtering (5,998 Lhx6-negative and 6,515 Lhx6-positive cells).

Since Lhx6 is expressed in postmitotic precursors, the Lhx6-positive sample from the MGE dissection should contain only postmitotic MGE precursors. However, the Lhx6-negative sample derived from the CGE will contain both mitotic progenitors and postmitotic precursors, as well as a minority of cells expressing *Lhx6* mRNA, probably owing to errors during FACS sorting or a time delay in GFP translation. To remove mitotic progenitors from the Lhx6-negative CGE dataset, we performed a maturation trajectory analysis as outlined above, and kept only postmitotic cells (2,905 cells, Extended Data Fig. 4h–j). To conservatively remove potential MGE cells from the same dataset, we clustered the data (as described in 'Initial clustering of all cells') and removed all cells belonging to clusters with an Lhx6 detection rate of more than 20%. This step removed 465 cells, leaving 2,440 Lhx6-negative postmitotic cells (Extended Data Fig. 4k).

To determine the branch identity of these cells, we mapped them to the E13.5 Drop-seq branches using a correlation-based approach. We focused on 279 genes that were differentially expressed in one branch compared to the other two. For these genes, we averaged the normalized expression in all branches (three branches in each eminence) to create branch model vectors. We then calculated Pearson correlations between all individual cells that we wanted to map and the model vectors (CGE model vectors for the Lhx6-negative sample, MGE model vectors for the Lhx6-positive sample). We assigned each cell to the branch with the highest correlation, but also calculated empirical P values to determine the significance of the assignment by permuting the single-cell data for a random background. We left the model vectors unchanged, but permuted the single-cell expression data 100 times. For each permutation and each cell we kept track of the largest Pearson correlation to the model vectors, and calculated a P value for the branch assignment by counting what fraction of correlation scores was larger than the one used for the branch assignment. In a final step, we turned all P values into FDRs and mapped only cells with an FDR <0.1 to the branches (Extended Data Fig. 4l–n).

Mapping E18.5 cortex and subcortex cells to branches. To fate-map *Dlx6a*-lineage neurons, *Dlx6a-cre* mice were crossed with either *Ai9* or *RCE^{loxP}* mice. Cortical and subcortical brain regions were dissected and collected from transgenic embryos at E18.5 and postnatal day (P) 10, dissociated into a single-cell suspension, and cells were collected with FACS based on their fluorescence expression. Single-cell libraries were prepared using the 10x Genomics pipeline. To identify the potential branch of origin of E18.5 cortical (8,382 cells) and subcortical (8,237) neurons, we mapped the cells of the E18.5 samples to the branches using the same approach as for the E13.5 cells. We identified 774 differentially expressed genes between the branches in the E13.5 10x data and used the average across cells as branch model vectors (separately for CGE and MGE derived cells). We then used the same procedure as for the E13.5 cells to map to the branches, allowing each E18.5 cell to map to any of the six E13.5 branches (3 CGE branches, 3 MGE branches), and applying the 0.1 FDR cutoff. Detailed results are shown in Extended Data Fig. 5 and Fig. 3i.

Relative variance explained. To quantify the contribution of different factors to the overall heterogeneity in our data, we compared the amount of variance of individual factors to the variance explained by the first principal component. Given an expression matrix (normalized using negative binomial regression as described above, regressing out number of reads and average number of molecules per detected gene), we first selected the most variable genes (z-scored standard deviation >1). To quantify the variance associated with an annotated factor of interest, we first constructed a vector representing the annotation of each cell. For a continuous factor of interest (for example, cell cycle score), we centred the vector and length-normalized it to length 1. We then projected the expression data onto this vector, and calculated the variance of the projected dataset. For a discrete factor (for example, branch), we first turned the vector into a set of indicator variables, and applied principal component analysis to obtain independent continuous vectors. We then projected the dataset onto each of these vectors, calculated the

variance of the projected dataset and took the sum over all vectors. This enabled us to compare the variance explained between different annotated factors (for example, cell cycle and maturation score). For visualization and interpretation in Figs. 2e and 3k, we normalized these values by the variance explained by the first principal component of the dataset.

Smart-seq2 data processing. Reads were aligned using the same pipeline as for the Drop-seq data. We kept only cells with at least 2,000 unique genes detected. We further removed cells with a z-scored alignment rate < -3, or with an absolute z-scored number of reads (after $\log_{10}$ tranformation) > 3. This resulted in 1,099 cells in the Flashtag experiments, and 400 cells in the Dlx6a-negative experiments.

Expression data were normalized to x with $x_{i,j} = \log_{10}\left(\frac{c_{i,j} \times 1,000}{m_j} + 1\right)$ in which $c_{i,j}$ is the read count of gene i in cell j and m_j is the sum of all mapped reads for cell j. Cell cycle score was calculated as described above, and regressed out using a linear model including an intercept term. We used the residuals for all further analyses. For dimensionality reduction we used diffusion maps on all genes that were detected in at least 5% of the cells (approximately 10,000 genes). All further analysis was carried out as for the Drop-seq data.

Integration of developmental and adult datasets. We used genetic fate-mapping strategies in combination with the 10x Genomics Chromium system for scRNA-seq to study cortical interneuron development at embryonic (E13.5, E18.5), postnatal (P10) and adult (P56) stages. Specifically, we used the *Lhx6-GFP* transgenic mouse line to select for postmitotic MGE cells and to precisely discriminate MGE versus CGE precursor cells at E14.5 (Extended Data Fig. 4). For later stages, in which cells have migrated out of the ganglionic eminences, we used a *Dlx6a-cre; RCE^{loxP}* pan-ganglionic-eminence fate-mapping strategy to collect cortical interneurons at E18.5 and P10 (Extended Data Fig. 5) and used the publicly available Allen Brain Institute scRNA-seq dataset[22] (Allen Cell Types Database; http://celltypes.brain-map.org/rnaseq) for the adult time point.

We applied our recently developed integration tool for scRNA-seq datasets[25], to identify shared sources of variation between embryonic (E13.5, E18.5), postnatal (P10) and adult (P56) datasets.

For the P56 dataset, we downloaded FPKM expression values for 8,432 single cells from the mouse visual cortex, sequenced with the Smart-Seq2 protocol, from a publicly available resource at the Allen Brain Atlas (Allen Cell Types Database; http://celltypes.brain-map.org/rnaseq). We selected 3,432 GABAergic cells for downstream analysis based on an initial clustering analysis and a selection of Gad1-positive clusters (Extended Data Fig. 6). For E18.5 and P10 datasets, we performed a similar initial clustering, removing populations of microglia, astrocytes, oligo-dendrocytes and smooth muscle cells that probably represent FACS false-positives and are unlikely to give rise to cortical interneurons (Extended Data Fig. 5). For the E13.5 and E14.5 dataset, we took all cells from our Lhx6-positive and Lhx6-negative datasets that were assigned to the interneuron precursor state (branch 1).

We performed three separate pairwise analyses, aligning E13.5 and P56, E18.5 and P56, and P10 and P56 datasets. In each case, we applied the Seurat alignment procedure as previously described[25]. We first detected variable genes in each dataset independently, using the FindVariableGenes function with default parameters. We used the union of the two variable gene sets used as input to canonical correlation analysis, and aligned the resulting canonical correlation vectors (CCV) across datasets with the AlignSubspace function. In brief, AlignSubspace constructs 'metagenes' representing the average expression of genes exhibiting robust correlations to the CCV in both datasets, and applies nonlinear 'warping' algorithms to align these metagenes between datasets. We performed this analysis for each of the top 15 CCV independently, and used biweight midcorrelation (bicor), a median-based similarity metric implemented in the WGCNA R package[42]. For downstream analysis (t-distributed stochastic neighbour embedding (t-SNE) and subtype mapping), we selected CCV for which at least 30 genes exhibited a minimum bicor of 0.15 in both datasets, applying the same cutoff across all three analyses.

For visualization (Fig. 4b), we constructed a distance matrix from these selected components as input to tSNE with default parameters. We next assigned cells from developmental datasets to adult subtypes, performing the following analysis separately for each of the three pairwise comparisons. For each cell in the developmental dataset, we calculated the $k = 10$ closet neighbours in the P56 dataset, using the selected CCV for the input distance matrix. If at least 9 of these neighbours in the P56 dataset were of the same subtype, the developmental cell was assigned to this subtype. We note that this represents a stringent threshold, which we apply equally across all comparisons. As a secondary check, for each developmental cell, we calculated the nearest $k = 10$ neighbours across all cells in the merged developmental and adult dataset. If none of these 10 neighbours represented cells in the P56 dataset, we also considered the cell to be unassigned. We performed this mapping procedure twice, to assign cells to the four main cardinal types (Fig. 4b–d), and also to assign them to the 14 finer subtypes (Extended Data Fig. 9)

To identify differentially expressed genes that were conserved across development, we used the FindConservedMarkers command in Seurat, which runs differential expression tests separately on both developmental and adult datasets. We required genes to have a 1.25-fold change, with a Bonferroni-corrected P value threshold of 0.05 in both developmental and adult datasets to be considered a conserved marker. Figure 4e and Extended Data Fig. 7 lists these genes, in order of when they first become annotated as differentially expressed.

Lastly, we explored whether our early transcriptomic markers of cardinal type separation in the mouse were also differentially expressed in human adult neurons. We used data from a recently published dataset of 14,963 single human nuclei from post-mortem tissue[28]. We performed standard log-normalization in Seurat, and calculated the average expression level of nuclei that were annotated as belonging to Pvalb, Sst, Id2, and Vip types. We used these values for the scatter plots in Extended Data Fig. 10.

Code availability. Code for preprocessing and analysis of scRNA-seq data are available from the Center for Open Science at https://osf.io/xjmtr. The Seurat package, which was used for the integration of developmental datasets, is open-source and freely available on GitHub (https://github.com/satijalab/seurat) and CRAN (https://cran.r-project.org/web/packages/Seurat/index.html).

Data availability. All source data, including sequencing reads and single-cell expression matrices, are available from the Gene Expression Omnibus (GEO) under accession code GSE104158.

30. Monory, K. *et al.* The endocannabinoid system controls key epileptogenic circuits in the hippocampus. *Neuron* **51,** 455–466 (2006).
31. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13,** 133–140 (2010).
32. Sousa, V. H., Miyoshi, G., Hjerling-Leffler, J., Karayannis, T. & Fishell, G. Characterization of Nkx6-2-derived neocortical interneuron lineages. *Cereb. Cortex* **19** (Suppl 1), i1–i10 (2009).
33. Vong, L. H., Ragusa, M. J. & Schwarz, J. J. Generation of conditional *Mef2c^{loxP/loxP}* mice for temporal- and tissue-specific analyses. *Genesis* **43,** 43–48 (2005).
34. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10,** 1096–1098 (2013).
35. Lee, S., Hjerling-Leffler, J., Zagha, E., Fishell, G. & Rudy, B. The largest group of superficial neocortical GABAergic interneurons expresses ionotropic serotonin receptors. *J. Neurosci.* **30,** 16796–16808 (2010).
36. Paxinos, G. & Franklin, K. B. J. *The Mouse Brain in Stereotaxic Coordinates 2nd edn* (Academic, 2003)
37. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31,** 2989–2998 (2015).
38. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008).
39. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11,** 740–742 (2014).
40. Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* **29,** 886–891 (2011).
41. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157,** 714–725 (2014).
42. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
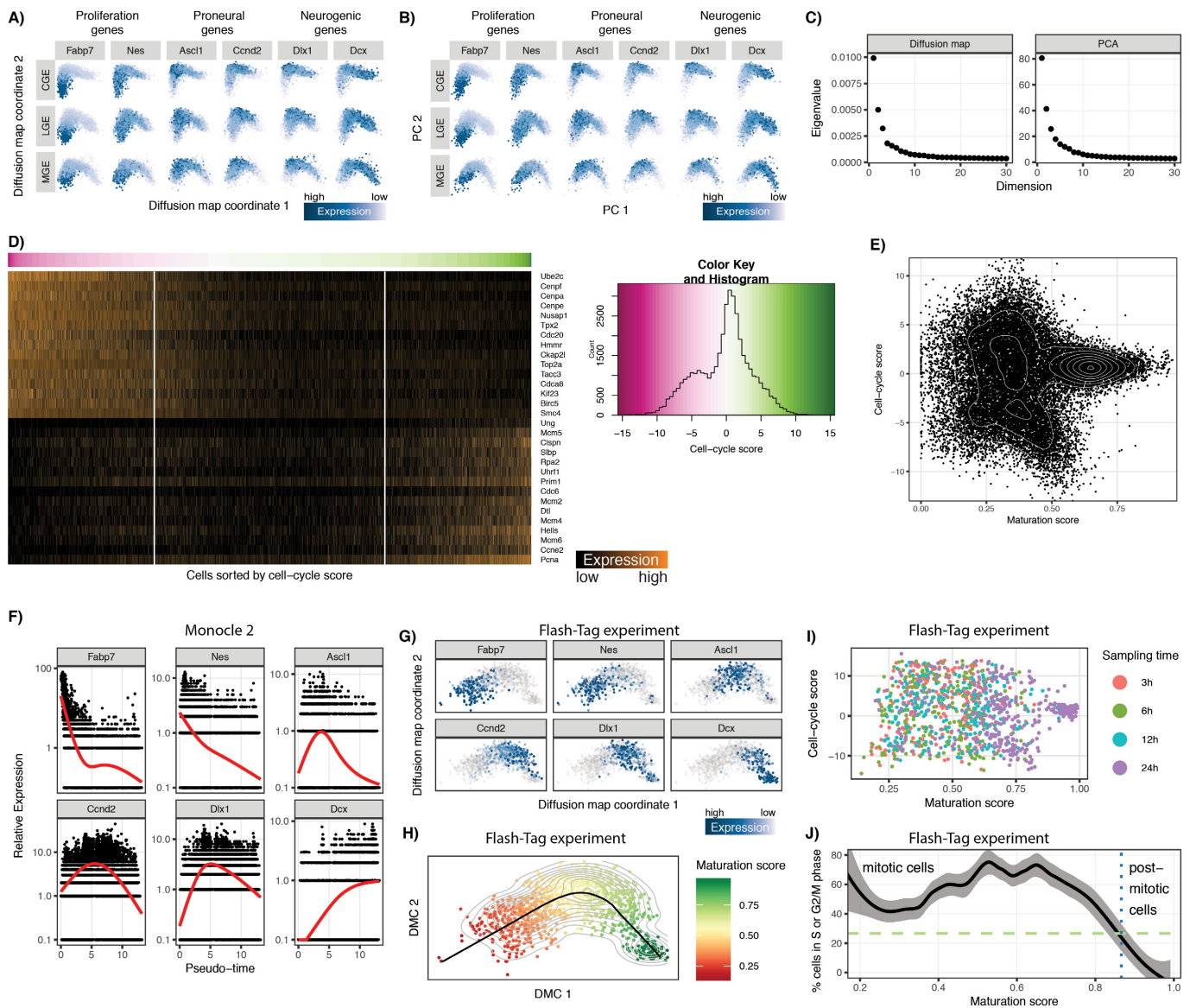
Extended Data Figure 1 | Ordering cells along a maturation trajectory.
a, Diffusion map analysis of eminence datasets suggests a pan-eminence developmental continuum. Each eminence was analysed independently, revealing nearly identical patterns. Cells are coloured according to the expression of canonical regulators. **b**, Using principal component analysis to reconstruct developmental maturation returns nearly identical results to the diffusion map analysis in Fig. 1. Principal component analysis was calculated for all eminences independently, and cells are coloured by their expression of canonical markers. **c**, Eigenvalues for the two dimensionality reduction methods. We observe a substantial eigenvalue drop-off after the initial components, demonstrating that the majority of the variance is captured in the first few dimensions. **d**, Single-cell heat map showing scaled expression levels of top genes that were correlated with cell cycle score. Cells on the *x* axis are sorted by cell cycle score. Negative scores correspond to cells in S phase, positive scores correspond to cells in G2/M phase. **e**, Scatter plot illustrating the relationship between maturation score and cell cycle score for all cells in the dataset. Each dot corresponds

to a single cell. Early progenitors span a wide range of cell cycle states, whereas late cells do not express G2/M or S-phase specific genes and express postmitotic genes. **f**, Expression of canonical marker genes as a function of 'pseudotime', as calculated with Monocle2[13]. Monocle2 pseudotime was strongly correlated with our maturation trajectory (both Pearson and Spearman $R = 0.94$). **g, h**, Diffusion map (**g**) and maturation trajectory (**h**) analysis of 1,099 single cells obtained from FlashTag animals, and sequenced using a custom version of the Smart-seq2 protocol (Supplementary Methods). Cells are coloured by their expression of canonical markers, which exhibit dynamics that are concurrent with the maturation trajectory learned from the Drop-seq data. **i, j**, Relationship between the maturation trajectory and cell cycle scores derived from the FlashTag datasets replicates our observations from Drop-seq. Therefore, our FlashTag maturation trajectory serves as complementary validation of our Drop-seq maturation trajectory, and exhibits strong association with biological time.
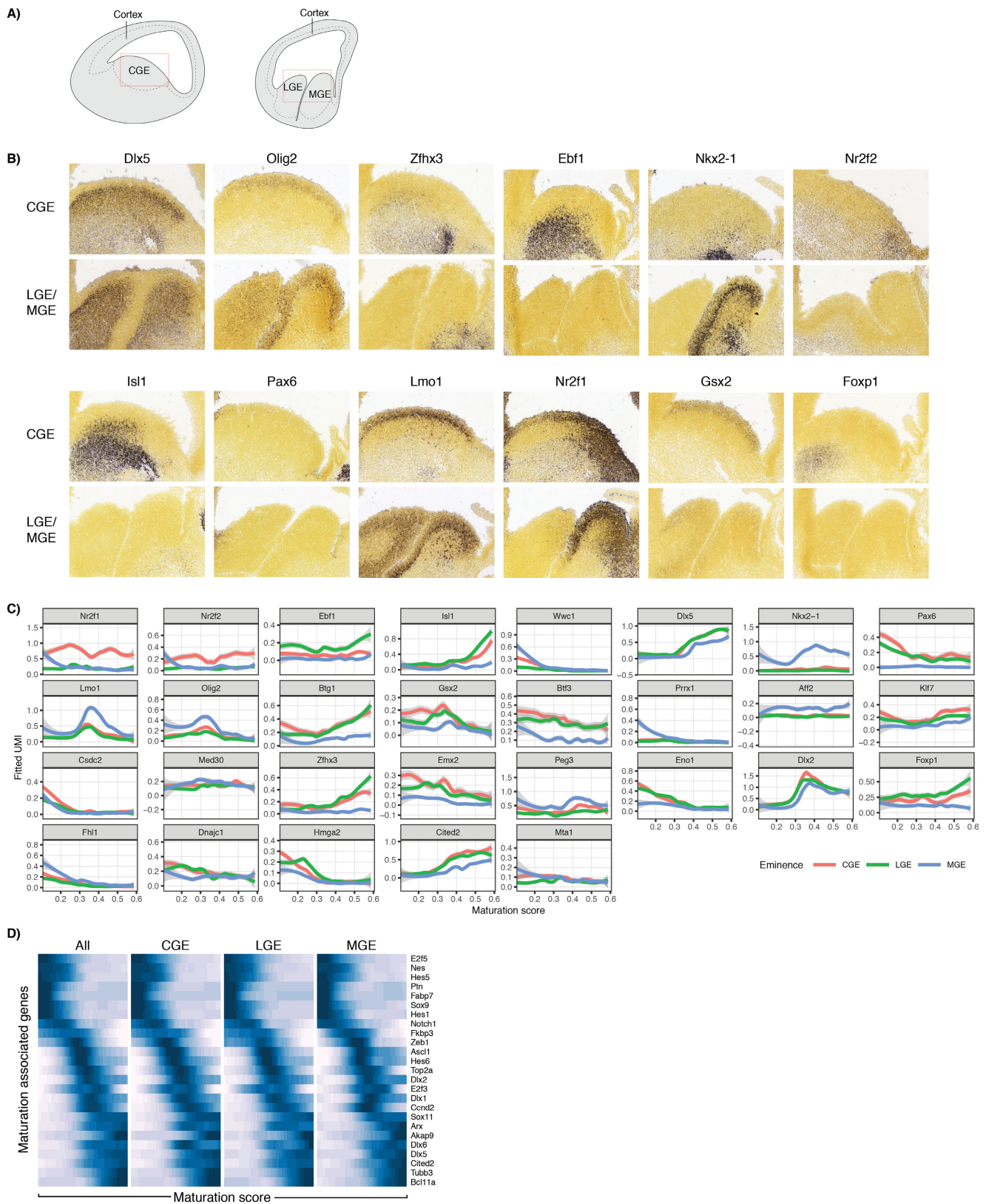
Extended Data Figure 2 | Enrichment of differentially expressed genes in the MGE, CGE and LGE. a, Schematic of embryonic brain sections at E13.5/E14.5. One sagittal section shows the MGE and LGE next to one another (right), whereas the other shows the CGE (left). **b**, ISH images from the Allen Brain Institute Developing Mouse Brain Atlas at E13.5 for genes that our analysis identified as being differentially expressed between the eminences. For each gene, ISH images are shown for the MGE, CGE and LGE. Image credit: Allen Institute. **c**, Temporal dynamics

for differentially expressed genes in early mitotic cells. Curves represent local averaging of single-cell expression, as a function of progression along the maturation trajectory, for each eminence independently. Grey area indicates 95% confidence interval. Genes are selected from the differentially expressed genes in early mitotic cells (Fig. 2a). **d**, Gene expression dynamics in mitotic cells, based on local averaging of single-cell data, plotted along maturation score for selected developmentally regulated genes.
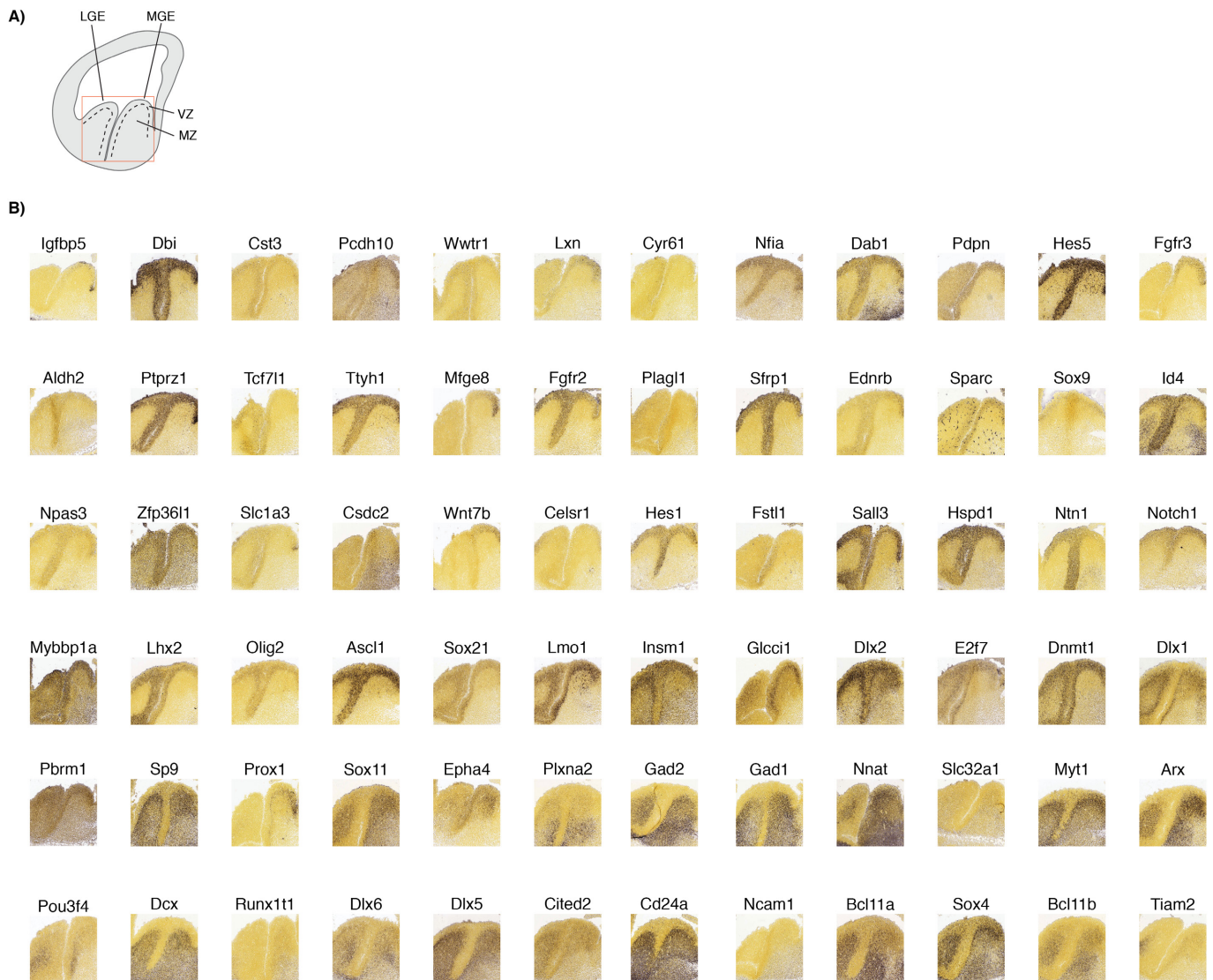
Extended Data Figure 3 | Enrichment of dynamically expressed genes in the ventricular zone, subventricular zone and mantle zone. **a**, Schematic of an embryonic brain section at E13.5/E14.5. The locations of the ventricular zone (VZ) and mantle zone (MZ) are indicated. **b**, Sagittal ISH images from the Allen Brain Institute Developing Mouse Brain Atlas at E13.5. Genes are ordered from lowest to highest maturation score rank. The trend overall shows that genes with peak expression at low maturation score tend to have higher expression in the ventricular zone, and as maturation score rank increases the expression pattern shifts to the subventricular zone and then to the mantle zone. Image credit, Allen Institute.
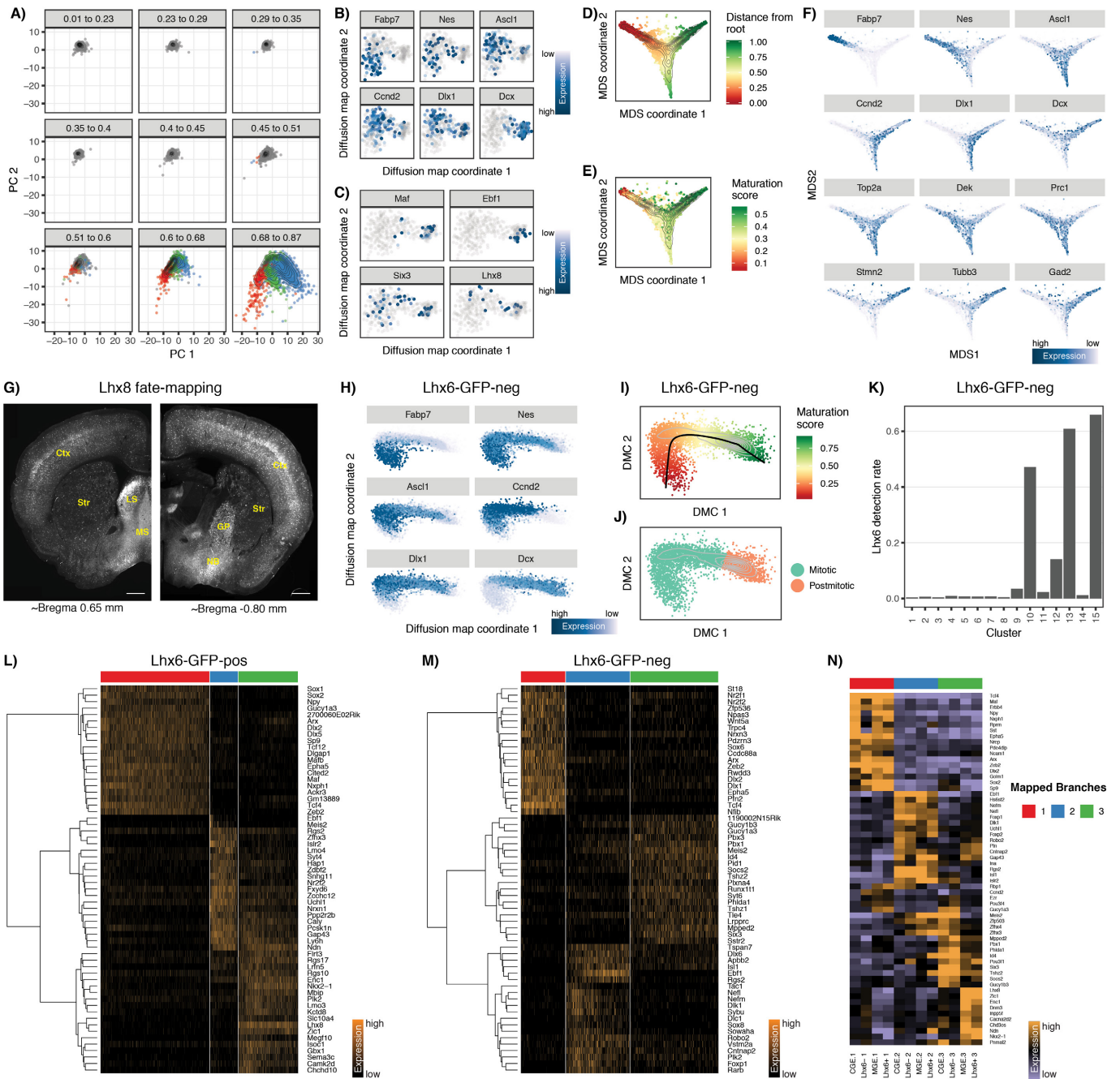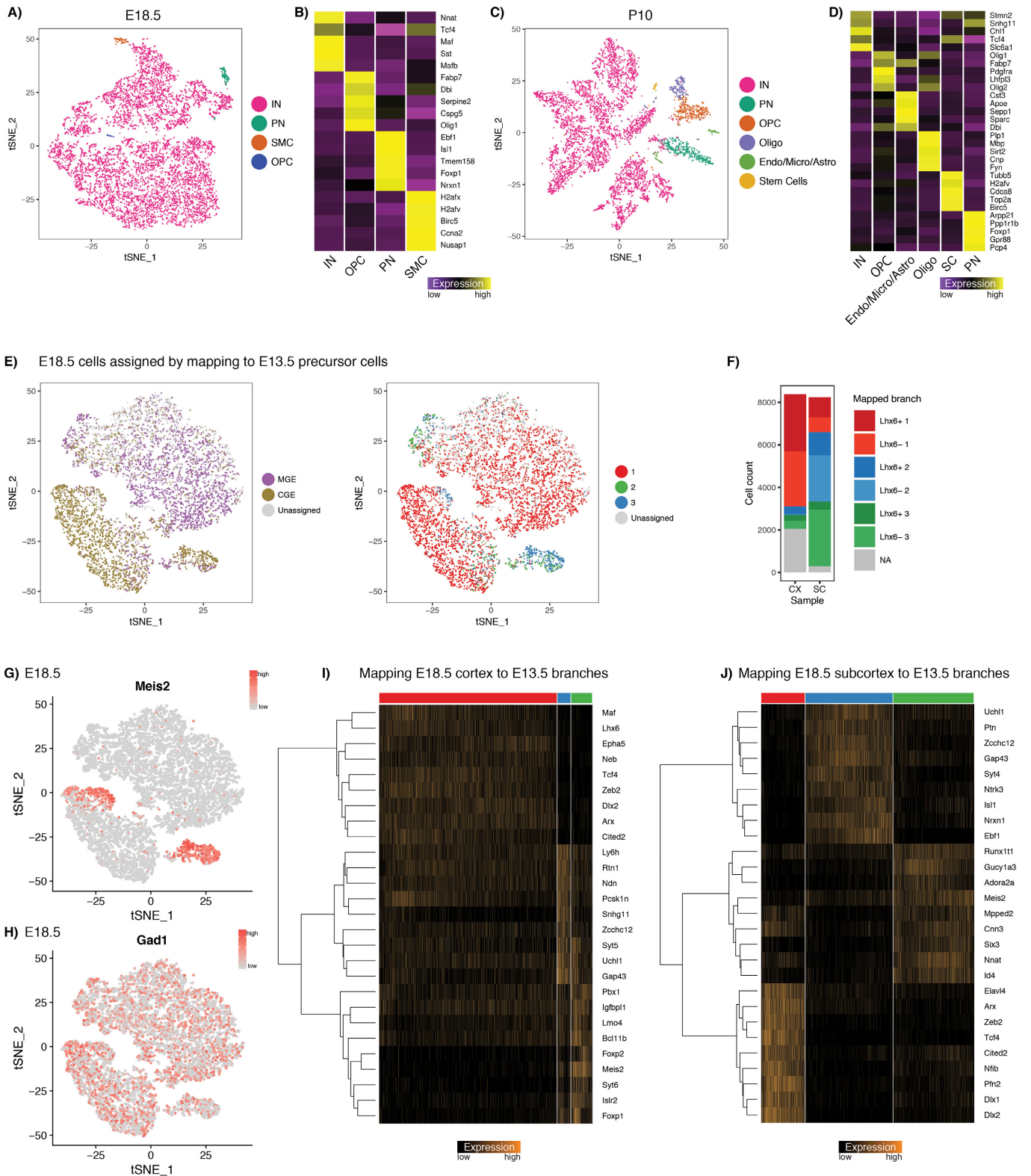
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Fate divergence occurs as cells become postmitotic. a, Supervised analysis: PCA of full dataset, run using only branch-dependent genes. Cells are grouped based on the maturation-trajectory bin: the first five bins represent mitotic progenitors, the last four bins represent postmitotic cells which are coloured by branch ID. Mitotic cells fall within a homogeneous point cloud, with low variance on principal components 1 and 2, showing no evidence of fate bifurcation. **b**, To test whether our inability to detect fate bifurcations earlier in development was due to the lower sequencing depth of Drop-seq, we sequenced 400 *Dlx6a-cre;RCE^{loxP}* negative ganglionic eminence cells (thereby enriching for mitotic progenitors), using a modified Smart-seq2 protocol. Diffusion map analysis of these cells returned only two significant principal components, with no evidence of further structure. These components reflect our previously defined maturation trajectory, with DMC1 separating mitotic cells (left). **c**, Rare mitotic cells expressing canonical branch markers do not segregate on the diffusion plot. **d–f**, Branching analysis on mitotic progenitors. We repeated the branch analysis, previously computed on postmitotic cells (Fig. 3a), on mitotic progenitors from all three ganglionic eminences. Although we did observe computational evidence of branching, this does not represent fate bifurcations as we observed in postmitotic cells. Instead, cells from different branches could largely be separated into 'early', 'intermediate' and 'late' regions of mitotic pseudotime, with one branch being largely defined by the expression of pro-neural cell cycle regulators (for example, Ascl1). As these genes peak at intermediate stages, our branching patterns could reflect either the aberrant assignment of intermediate cells to a new branch, or reflect the potential of multiple modes of cell division (namely,

direct versus indirect neurogenesis) occurring in the ventricular zone and subventricular zone. **g**, Genetic fate-mapping using *Lhx8-cre/cerulean* demonstrates that MGE branch three precursors give rise to the entire breadth of cholinergic projection (globus pallidus and nucleus basalis) and interneuron (striatum) populations. The cumulative longitudinal use of a constitutive Cre driver also results in extensive labelling of cortical interneurons owing to transient expression within this population. Scale bar, 500 μm. Ctx, cortex; Str, striatum; LS, lateral septum; MS, medial septum; NP, nucleus basalis; GP, globus pallidus. **h**, Our Lhx6–GFP-negative dataset contains both mitotic and postmitotic cells from the CGE and diffusion map analysis shows our previously defined maturation trajectory. **i**, **j**, To isolate postmitotic cells, we calculated a maturation trajectory (**i**), and used the cell cycle scores to identify the transition point between mitotic and postmitotic cells (**j**) as with the eminence datasets in Fig. 1. **k**, To avoid the possibility of FACS false-negative MGE cells contaminating our Lhx6–GFP-negative dataset, we clustered the postmitotic cells from this dataset, and filtered out three rare clusters where *Lhx6* mRNA expression was detected in more than 20% of cells (Supplementary Methods). **l**, **m**, We mapped postmitotic cells from the Lhx6–GFP-positive (**l**) and Lhx6–GFP-negative (**m**) datasets to the branches determined from the Drop-seq dataset (Supplementary Methods). Heat maps show scaled single-cell expression markers associated with each branch. **n**, Analogous to Fig. 3e, but also including the Lhx6–GFP-positive and Lhx6–GFP-negative datasets generated using 10x Genomics, as a validation of the original Drop-seq datasets that were performed on wild-type mice.
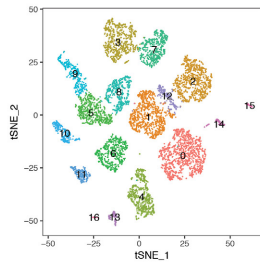
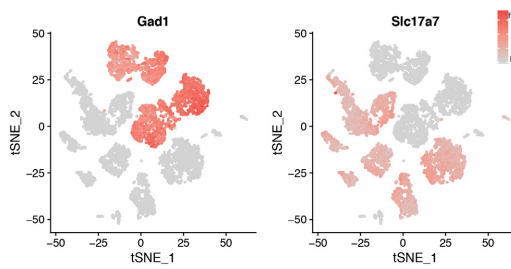**Extended Data Figure 5** | See next page for caption.

Extended Data Figure 5 | Filtering of E18.5 and P10 10x datasets and mapping of E18.5 cortex and subcortex neurons to E13.5/E14.5 branches. a, **c**, *t*-SNE visualization of *Dlx6a-cre;RCE^loxP* positive E18.5 cortical cells (**a**) and *Dlx6a-cre;RCE^loxP* positive P10 cortical cells (**c**). Although the *Dlx6a-cre* should mark only GABAergic eminence-derived cells, we identified rare populations that did not express canonical interneuron (IN) markers, probably representing false positives from FACS. **b**, **d**, Gene expression in these populations (E18.5 cells **b**, P10 cells **d**; heat map shows average expression in group) identifies rare contaminating populations of microglia (micro), astrocytes (astro), oligodendrocyte precursor cells (OPCs) and oligodendrocytes (oligo); smooth muscle cells (SMC), stem cells (SC), projection neurons (PN). For all downstream analyses, we considered only cells in the interneuron cluster. **e**, *t*-SNE visualization of 8,382 *Dlx6a-cre;RCE^loxP* positive E18.5 cortical cells (same dataset as in Extended Data Fig. 5a, but after removing contaminating populations). Each E18.5 cell was mapped to one of six precursor states (branch 1, 2, and 3 for Lhx6–GFP-positive and Lhx6–GFP-negative datasets), using a correlation-based distance metric (Supplementary Methods). This enabled us to assign a putative eminence and branch of origin for each of the E18.5 cortical cells. **f**, As expected, the vast majority of *Dlx6a-cre;RCE^loxP* positive E18.5 cortical cells map to the interneuron precursor state, and are split between MGE and CGE-derived precursors. By contrast, *Dlx6a-cre;RCE^loxP* positive E18.5 cells from the subcortex primarily map to branches 2 and 3, consistent with our interpretation of these branches as precursor states for projection neurons; CX, cortex; SC, subcortex. **g**, **h**, The minority of Dlx6a-positive cortical cells mapping to precursor states 2 and 3 primarily co-express Meis2 (**g**) and Gad1 (**h**), probably representing a CGE-derived GABAergic population. These cells have been recently described as being present in the cortical white matter and probably represent projection neuron precursors[21]. **i**, **j**, Heat maps showing single-cell expression markers for the three different mapped branches of *Dlx6a-cre;RCE^loxP* positive E18.5 cells from the cortex (**i**) and the subcortex (**j**).
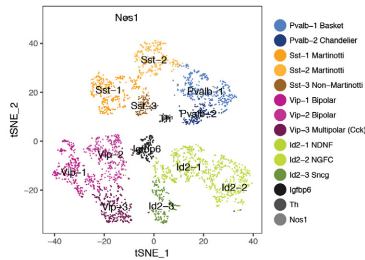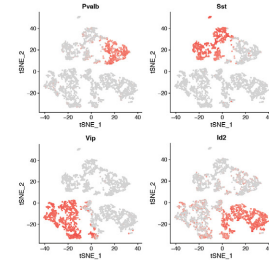
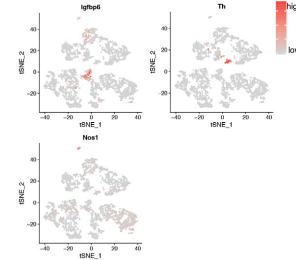**A)** Adult Visual Cortex (Data from the Allen Institute)

B)

Gad1   Slc17a7

C) GABAergic neurons in the adult visual cortex

- Pvalb–1 Basket
- Pvalb–2 Chandelier
- Sst–1 Martinotti
- Sst–2 Martinotti
- Sst–3 Non–Martinotti
- Vip–1 Bipolar
- Vip–2 Bipolar
- Vip–3 Multipolar (Cck)
- Id2–1 NDNF
- Id2–2 NGFC
- Id2–3 Sncg
- Igfbp6
- Th
- Nos1
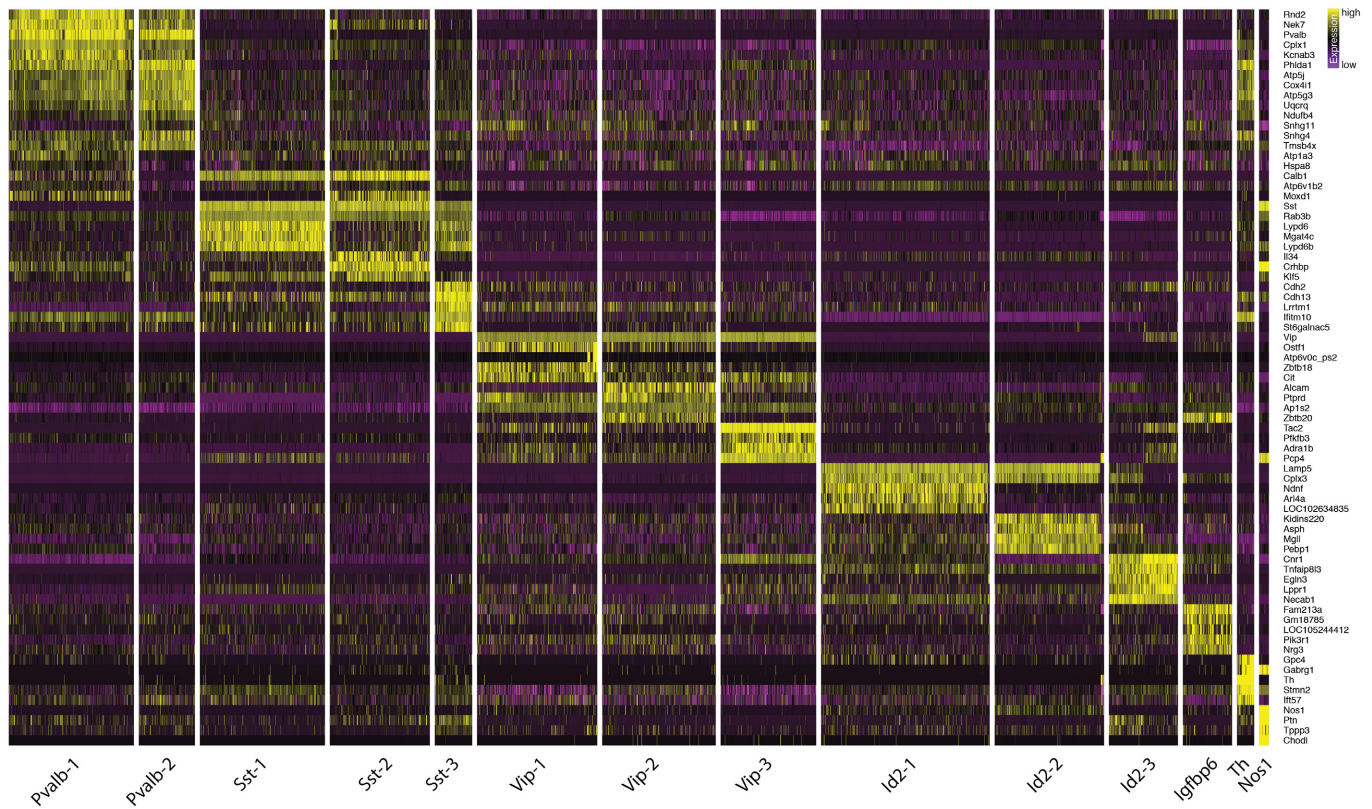
D)   Pvalb   Sst   Vip   Id2

E)   Igfbp6   Th   Nos1

F) GABAergic neurons in the adult visual cortex

Extended Data Figure 6 | Clustering of adult visual cortical neurons into 14 major non-overlapping inhibitory interneuron subtypes. **a**, Initial *t*-SNE visualization and graph-based clustering of 8,329 single cells individually isolated from P56 mouse visual cortex and sequenced with the Smart-Seq2 protocol. Data was downloaded from the publicly available resource hosted by the Allen Brain Atlas[22] (Allen Cell Types Database, http://celltypes.brain-map.org/download (2015)). **b**, Of all cells, 3,432 GABAergic interneurons were easily identified by the expression of *Gad1* (left) and the absence of *Slc17a7* (right), and were selected for downstream analysis. **c**, *t*-SNE visualization and graph-based clustering of the 3,432 GABAergic cells reveals 14 clusters. **d**, **e**, The clusters revealed in **c** could be broadly grouped into cardinal types based on the expression of canonical markers. **f**, Single-cell heat map showing scaled expression values for the best transcriptomic markers in each cluster.
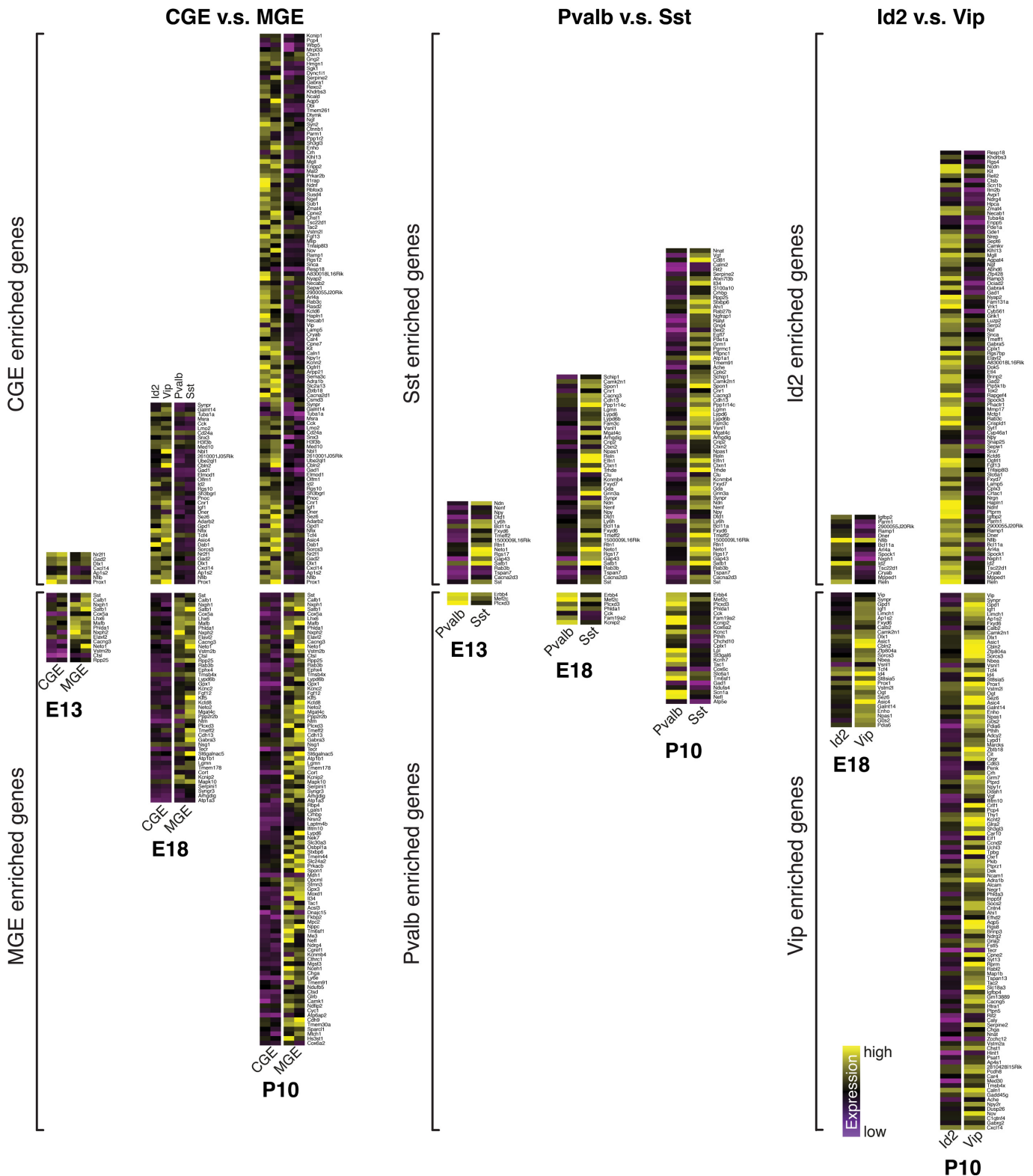
Extended Data Figure 7 | Emergence of transcriptomically defined subtypes across development. Differentially expressed genes between MGE and CGE derived subsets (left), that are conserved in both developmental and P56 cells. Each conserved gene is placed on the heat map when it is first observed to be differentially expressed during development, and the number of conserved differentially expressed genes grows over time. The same analysis is shown for Pvalb and Sst subsets (middle), and for Vip and Id2 subsets (right). This figure is identical to Fig. 4e, but with all gene names displayed.

A) Cells assigned by mapping to P56 neurons

B)
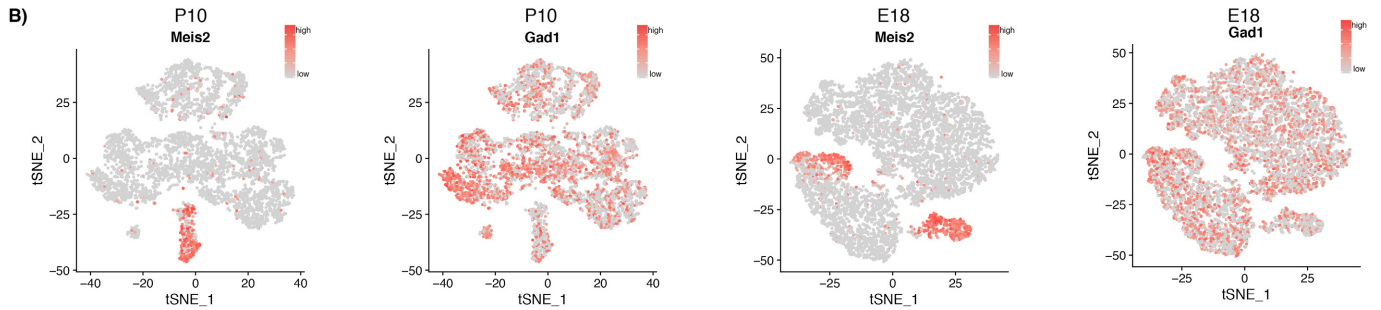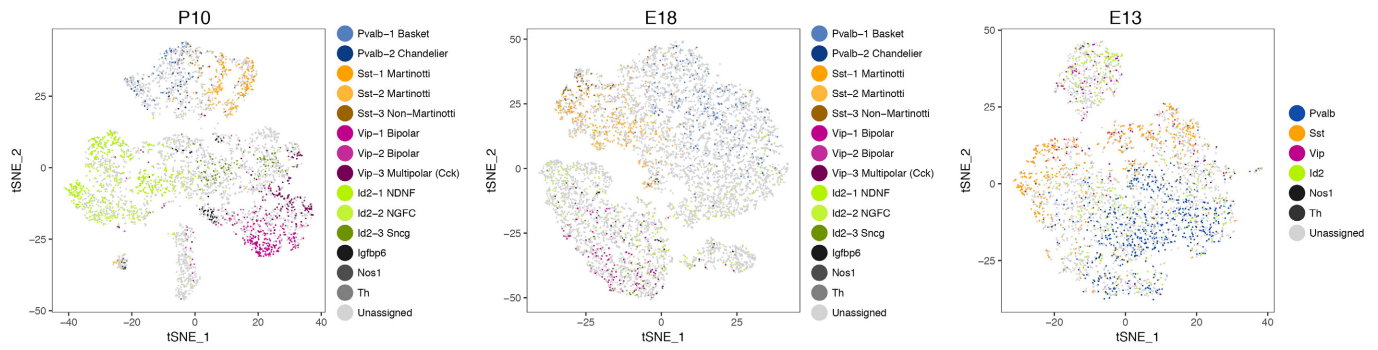
**Extended Data Figure 8 | The integrated analysis agrees with an independent *t*-SNE analysis of each time point. a**, *t*-SNE visualizations of interneuron precursors from E13.5, E18.5 and P10, calculated independently for each time point. Cells are coloured as in Fig. 4b–d, based on their mapping to P56 datasets in integrated analysis. However, since the *t*-SNE was performed separately for each time point, we can assess how the integrated analysis agrees with an independent analysis of each time point. In each case, we can see that the cardinal type separation

that we observe via integrated analysis (Fig. 4b–d) is consistent with an independent analysis of each dataset. Integrated analysis with the P56 dataset results in clearer separation, and enables us to map developmental precursors to adult subtypes. **b**, Expression of *Gad1* and *Meis2* in single-cell datasets. Cells expressing both genes are probably projection neuron precursors that have recently been described in the CGE[21], but whose progeny is not captured in the mouse visual cortex dataset. Therefore, these cells are correctly mapped as unassigned.
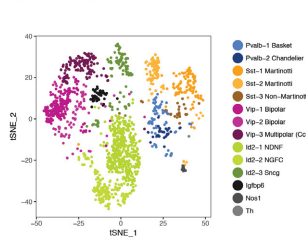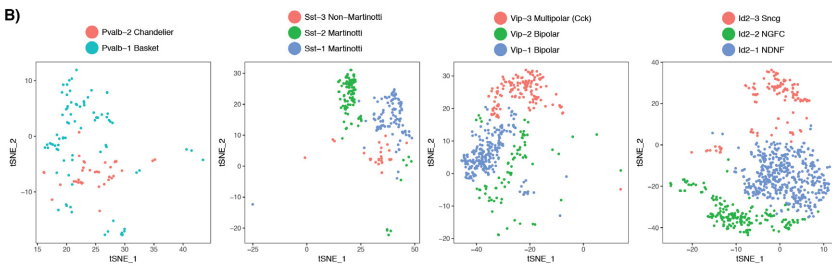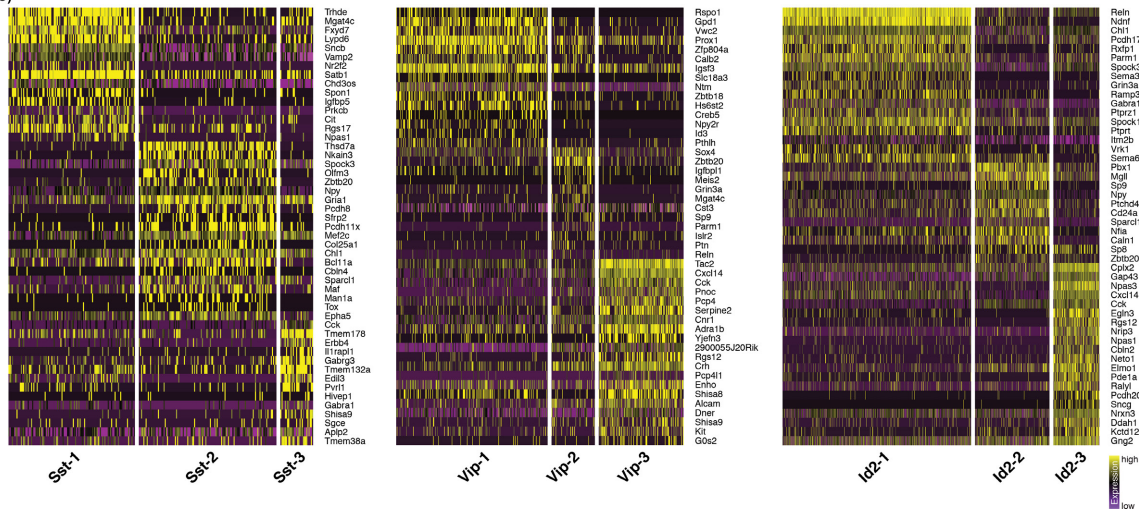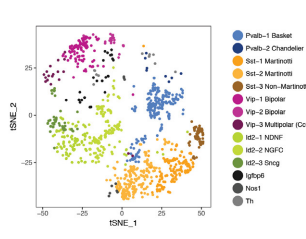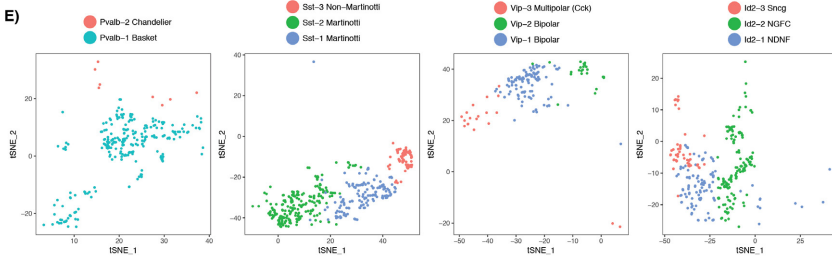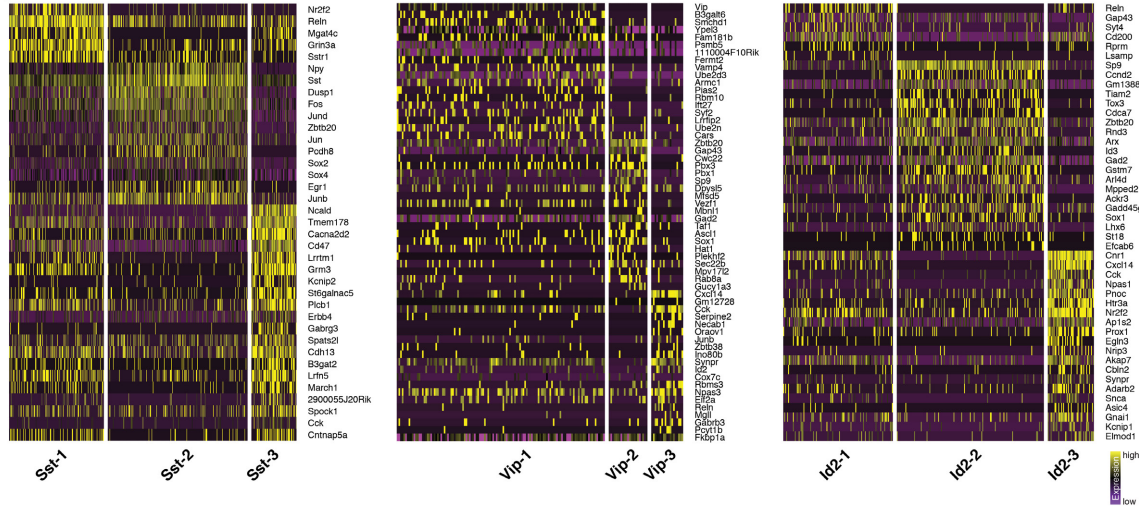
Extended Data Figure 9 | Transcriptional segregation into cortical interneuron subtypes at different developmental stages. a, *t*-SNE visualization of all P10 cells mapping to a P56 subtype (as in right column of Fig. 4c, but cells are coloured by subtype instead of cardinal type). **b**, *t*-SNE visualization as in **a**, but zoomed in on each cardinal type independently. **c**, Single-cell heat maps showing the best transcriptomic markers marking each subtype, for the Sst (left), Vip (middle) and Id2 (right) cardinal types, within P10 cells. We did not observe any statistically significant markers subdividing Pvalb subtypes. **d**, *t*-SNE visualization of all E18.5 cells mapping to a P56 subtype (as in right column of Fig. 4c, but cells are coloured by subtype instead of cardinal type). **e**, *t*-SNE visualization as in **d**, but zoomed in on each cardinal type independently. **f**, Single-cell heat maps showing the best transcriptomic markers marking each subtype, for the Sst (left), Vip (middle) and Id2 (right) cardinal types, within E18.5 cells. We did not observe any statistically significant markers subdividing Pvalb subtypes.

Extended Data Figure 10 | A subset of embryonic markers of cardinal type specification in mouse are conserved in adult human neurons.
a, Quantification of Pvalb-positive cortical interneurons across the different cortical layers of the control and *Mef2c* cortical knockout (*Dlx6a-cre;Mef2c*$^{loxP/loxP}$*RCE*) animals. *Mef2c* cortical knockout results in a reduction in Pvalb density in all cortical layers except for layer 1. Error bars reflect s.e.m.; unpaired *t*-test; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; $n = 4$ brains each for cortical knockout and control, 3–4 sections per brain. **b–d**, Scatter plot comparing average expression of 3,035 GABAergic single nuclei from post-mortem human neurons, after segregation into Pvalb and Sst (**b**), Vip and Id2 (**c**) and MGE and CGE inferred origins (**d**). Each dot represents the expression of a gene in human cells. Markers of transcriptomic cardinal types from our E13.5 and E18.5 datasets (from Fig. 4e) are shown in red or blue dots. Mouse embryonic markers that also differ by 1.5-fold in human have gene names annotated on the plot.

nature research

Corresponding author(s): Gord Fishell, Rahul Satija

☐ Initial submission   ☐ Revised version   ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▸ Experimental design

1. **Sample size**

   Describe how sample size was determined.

   > N/A

2. **Data exclusions**

   Describe any data exclusions.

   > See manuscript page 2, paragraph 3:
   > The focus of our study was to investigate inhibitory interneuron development. Therefore, when we identified small populations of excitatory neurons (Neurod6; 2.6% of cells) and endothelial cells (Igfbp7; 0.7% of cells) (Fig. 1B, C), both were excluded from further analysis.

3. **Replication**

   Describe whether the experimental findings were reliably reproduced.

   > See manuscript page 2, Paragraph 3:
   > After cell dissociation, we utilized Drop-seq to sequence the transcriptomes of 5,622 single cells from the MGE, 7,401 from the CGE, and 8,543 from the LGE, using three independent biological replicates for each eminence.

4. **Randomization**

   Describe how samples/organisms/participants were allocated into experimental groups.

   > N/A

5. **Blinding**

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   > N/A

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Combination of custom R scripts and existing R packages as outlined in the supplemental methods section

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	See Supplemental methods Primary antibodies: goat anti-ChAT (1:250, Millipore AB144P), goat anti-Pvalb (1:1000, Swant PVG213), rabbit anti-SST (1:3000, Peninsula Labs T-4102), rabbit anti-GABA (1:2000, Sigma A2052), chicken anti-GFP (Aves Labs 1020), rabbit anti-DsRed (Clontech 632496),Anti-GFP Chicken Polyclonal IgY (1:1000) (Abcam Ab13970), Anti-Parvalbumin (Pvalb) Goat (Swant PVG 213) and Anti-Parvalbumin (Pvalb) Rabbit (Swant PV25), rat anti-SST (1:250, Chemicon), mouse anti-Pvalb (1:1000, Sigma), and rabbit anti-VIP (1:250, ImmunoStar), rabbit anti-Neuropeptide Y (1:500; Incstar), mouse anti-Calretinin (1:1500; Chemicon). Secondary antibodies: Alexa Fluor 488-donkey anti-chicken, Alexa Fluor 594-donkey 5'-anti-rabbit, and Alexa Fluor 647-donkey anti-goat and rabbit (all Jackson Immunoresearch), secondary Alexa antibodies (Life Technologies).

10. Eukaryotic cell lines

    a. State the source of each eukaryotic cell line used.

        N/A

    b. Describe the method of cell line authentication used.

        N/A

    c. Report whether the cell lines were tested for mycoplasma contamination.

        N/A

    d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

        N/A

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

    See supplementary methods, section "Animal Work"

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

    N/A