# Molecular landscapes of human hippocampal immature neurons across lifespan

https://doi.org/10.1038/s41586-022-04912-w

Received: 25 October 2021

Accepted: 26 May 2022

Published online: 06 July 2022

Check for updates

Yi Zhou[1,19], Yijing Su[1,19], Shiying Li[1,2,19], Benjamin C. Kennedy[3,4], Daniel Y. Zhang[5], Allison M. Bond[1], Yusha Sun[6], Fadi Jacob[1], Lu Lu[1], Peng Hu[7], Angela N. Viaene[8], Ingo Helbig[9,10,11,12], Sudha K. Kessler[9,12], Timothy Lucas[4], Ryan D. Salinas[4], Xiaosong Gu[2], H. Isaac Chen[4,13], Hao Wu[7,13,14], Joel E. Kleinman[15], Thomas M. Hyde[15], David W. Nauen[16], Daniel R. Weinberger[15], Guo-li Ming[1,13,17,18,20] ✉ & Hongjun Song[1,13,14,17,20] ✉

Immature dentate granule cells (imGCs) arising from adult hippocampal neurogenesis contribute to plasticity and unique brain functions in rodents[1,2] and are dysregulated in multiple human neurological disorders[3–5]. Little is known about the molecular characteristics of adult human hippocampal imGCs, and even their existence is under debate[1,6–8]. Here we performed single-nucleus RNA sequencing aided by a validated machine learning-based analytic approach to identify imGCs and quantify their abundance in the human hippocampus at different stages across the lifespan. We identified common molecular hallmarks of human imGCs across the lifespan and observed age-dependent transcriptional dynamics in human imGCs that suggest changes in cellular functionality, niche interactions and disease relevance, that differ from those in mice[9]. We also found a decreased number of imGCs with altered gene expression in Alzheimer's disease. Finally, we demonstrated the capacity for neurogenesis in the adult human hippocampus with the presence of rare dentate granule cell fate-specific proliferating neural progenitors and with cultured surgical specimens. Together, our findings suggest the presence of a substantial number of imGCs in the adult human hippocampus via low-frequency de novo generation and protracted maturation, and our study reveals their molecular properties across the lifespan and in Alzheimer's disease.

During adult hippocampal neurogenesis, activated neural stem cells generate proliferating intermediate neural progenitors (IPCs) and neuroblasts, which in turn give rise to post-mitotic imGCs that mature over time[2] (Fig. 1a). The presence of adult-born dentate granule cells (GCs) in humans was first demonstrated in specimens from patients who previously received nucleotide analogues that dated newborn cells[10], and was independently confirmed and further characterized using the radiocarbon dating approach[11,12]. As accumulating evidence has attributed the function of adult neurogenesis to unique properties of immature neurons that are distinct from mature neurons[2,4,13–15], the immature neuron population is an important target for analysis. Recent contradictory reports have provided immunohistological evidence for[3,5,7,16–19] and against[6,20,21] the existence of immature neurons in the adult human dentate gyrus. These studies have relied largely on immunostaining of doublecortin (DCX), an immature neuron marker that requires intricate histological protocols for postmortem adult human brain specimens[5,22]. These controversies highlight a major gap in our knowledge about immature neurons in the human hippocampus with limited markers and call for new approaches for their identification and analysis. A more precise identification of immature neurons could be

[1]Department of Neuroscience and Mahoney Institute for Neurosciences, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Key Laboratory of Neuroregeneration of Jiangsu and Ministry of Education, Co-innovation Center of Neuroregeneration, Nantong University, Nantong, China. [3]Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [4]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [5]Biochemistry and Molecular Biophysics Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [6]Neuroscience Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [8]Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [9]Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [10]The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hospital of Philadelphia, Philadelphia, PA, USA. [11]Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia, Philadelphia, PA, USA. [12]Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [13]Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA, USA. [14]The Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [15]Lieber Institute for Brain Development, The Solomon H. Snyder Department of Neuroscience, Department of Neurology, and Department of Psychiatry, School of Medicine, Johns Hopkins University, Baltimore, MD, USA. [16]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [17]Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [18]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [19]These authors contributed equally: Yi Zhou, Yijing Su, Shiying Li. [20]These authors jointly supervised this work: Guo-li Ming, Hongjun Song. ✉e-mail: gming@pennmedicine.upenn.edu; shongjun@pennmedicine.upenn.edu

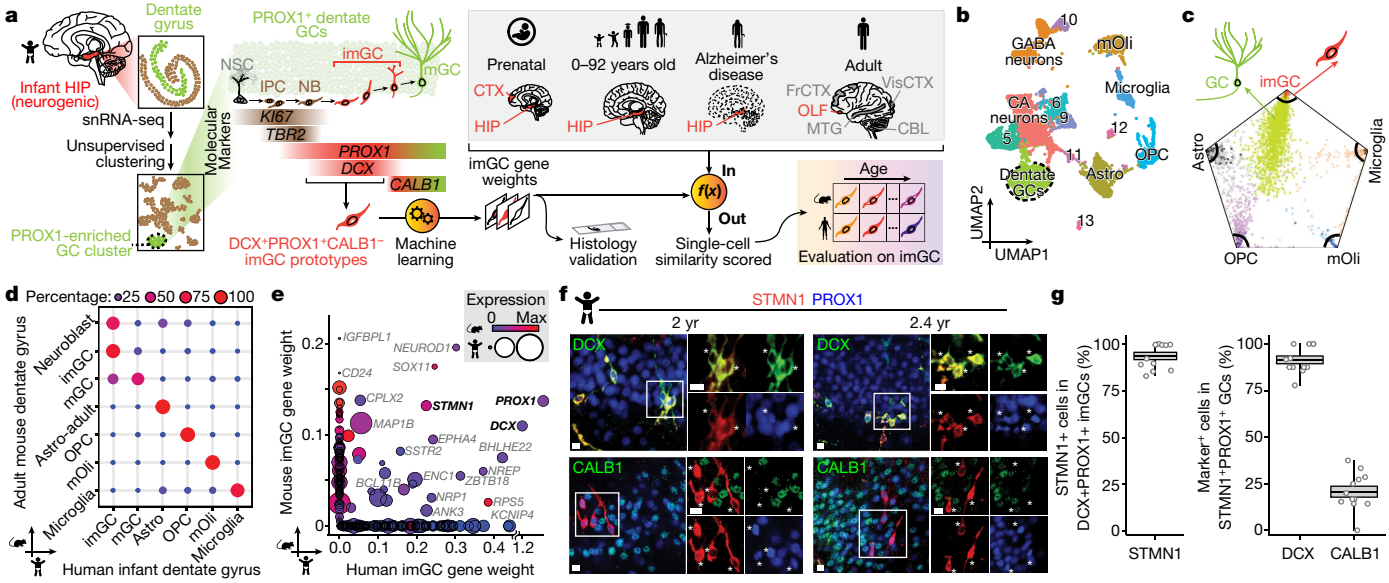



**Fig. 1 | snRNA-seq and immunohistological analyses of imGCs in the human infant hippocampus. a**, Schematic illustration of the experimental design. CBL, cerebellum; CTX, cortex; FrCTX, frontal cortex; HIP, hippocampus; MTG, middle temporal gyrus of cortex; NB, neuroblast; NSC, neural stem cell; OLF, olfactory epithelium; VisCTX, visual cortex. **b**, UMAP visualization of 15,434 nuclei from 4 human infant hippocampal specimens, coloured by cluster. The GC cluster is highlighted with a dashed circle. **c**, Wheel plot visualizing scores of each cell to each prototype by the machine learning model. Dots represent individual cells whose distance to each prototype is proportional to the similarity score of that prototype. Each black line indicates a similarity score of 0.85 to each prototypical cell type. **d**, Transcriptional congruence between the corresponding mouse[9] and human cell types measured by a multi-class random forest classifier[24,32] trained on different human cell types. The confusion matrix plot indicates the percentage of cells of a given mouse cell cluster (row, based on published annotations[9]) assigned to a corresponding human cell type (column, classified by the machine learning model). **e**, Comparison of positive gene weights defining imGCs in humans and mice generated by separate machine learning models. **f**,**g**, Sample confocal immunostaining images (**f**) and quantification (**g**) of STMN1 enrichment in imGCs in the human infant hippocampus. yr, years. Scale bars, 10 µm. Asterisks indicate DCX+ or CALB1− cells among STMN1+PROX1+ cells (**f**). Dots represent data from individual sections; the centre line represents the mean, box edges show s.e.m. and whiskers extend to maximum and minimum values ($n = 4$ subjects) (**g**).

obtained by considering simultaneous expression of multiple genes, ideally the whole transcriptome, at single-cell resolution. To investigate the existence, abundance and molecular properties of neurons with immature neuronal characteristics in the human hippocampus, we performed single-nucleus RNA sequencing (snRNA-seq) aided by a machine learning-based analytical approach to examine human imGCs across the human lifespan (Fig. 1a).

## snRNA-seq of human infant hippocampi

We first performed snRNA-seq analysis of four infant hippocampus specimens (Supplementary Tables 1 and 2), a developmental stage with abundant imGCs[16,20]. Unsupervised clustering identified 14 clusters on the basis of their defining markers (Fig. 1b and Extended Data Fig. 1a,b). The imGCs express immature neuronal marker DCX and pan-GC marker prospero homeobox 1 (PROX1), whereas calbindin (CALB1) is expressed in some imGCs, but is more enriched in mature GCs[5,9,23] (mGCs) (Fig. 1a). Within the GC cluster marked by prominent PROX1 enrichment, DCX+ imGCs were intermingled with other cells and could not be separated by finer partitioning (Extended Data Fig. 1c), in contrast to clear clustering of imGCs in the mouse dentate gyrus single-cell mRNA-sequencing (scRNA-seq) dataset[9] (Extended Data Figs. 2a and 3). Therefore, the conventional unsupervised method is insufficient to identify the imGC population in human snRNA-seq datasets, similar to previous analyses of the adult human hippocampus[24–27].

## Identifying imGCs by machine learning

To identify imGCs, we next explored a supervised machine learning approach[28]. This prototype-based scoring method uses a training set of cell prototypes to extract weighted panels of molecular features de novo, which are then used to quantify the resemblance of each individual cell from query datasets to each prototypical cell type for classification with high fidelity. Applications of supervised models have been shown to be highly effective in differentiating transcriptionally ambiguous cell subtypes in scRNA-seq datasets in multiple systems[24,28–32]. As a validation of model performance, we first tested this approach to identify immature neurons using scRNA-seq datasets from the mouse hippocampus across ages[9] (Supplementary Table 3) for comparison with those from unsupervised clustering, which results in distinct subclusters of neuroblasts, imGCs and mGCs identified via established markers (Extended Data Figs. 2a–c and 3). To mirror our human analysis, we selected high-confidence Dcx+Prox1+Calb1− imGCs from the GC clusters in the P5 mouse hippocampal dataset as a prototype, as well as all major non-neuronal cell-type prototypes (astroglia, oligodendrocyte progenitor cells (OPCs) and microglia) for training (Fig. 1a and Extended Data Fig. 2c; Methods). The trained model was used to score each cell from the query mouse hippocampal datasets across ages independently of the clustering information (Extended Data Fig. 2d–f and Supplementary Table 4). We identified immature neurons using a conservative, empirical cut-off ($P \geq 0.85$) for the similarity score to the mouse imGC prototype and compared our classifications with published clustering annotations[9]. Model-classified immature neurons were largely within the neuroblast and imGC clusters, with some appearing in the immature CA neuron and GABAergic (γ-aminobutyric acid-producing) neuron clusters at P5, yet they resided almost exclusively in the GC lineage in the juvenile and adult mouse hippocampus (Extended Data Fig. 3). Notably, within the GC lineage, imGCs include the majority of cells in the neuroblast and imGC clusters, fewer in the IPC clusters, and almost none in the mGC clusters (Extended Data

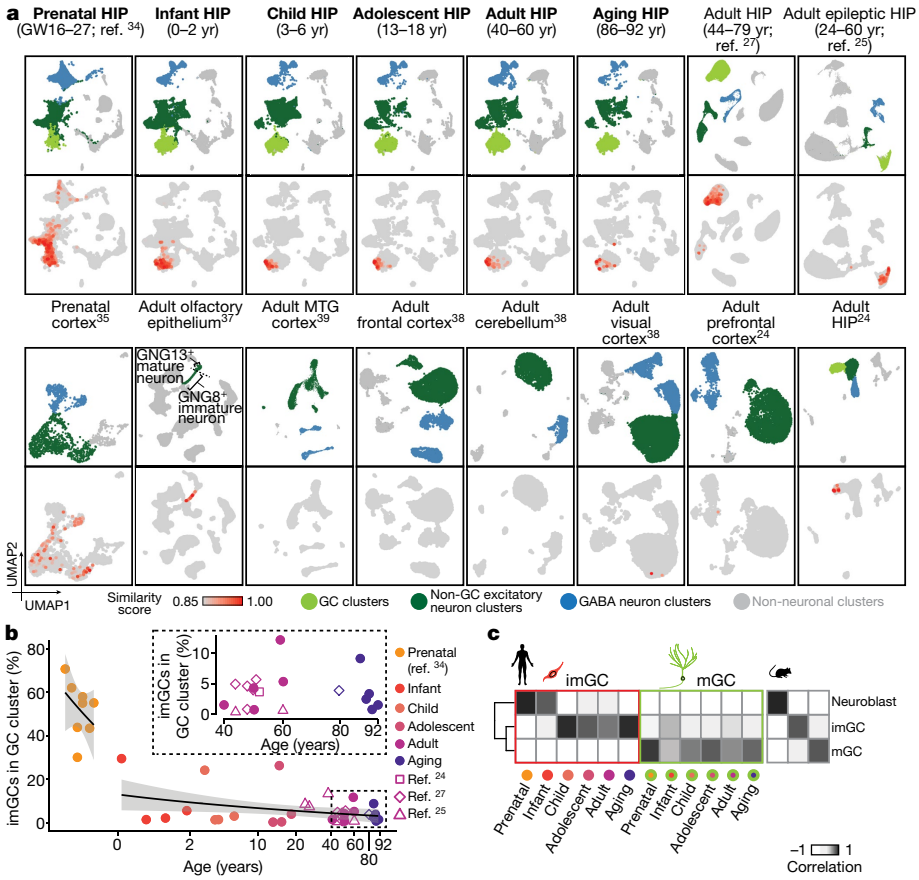



**Fig. 2 | snRNA-seq analysis of human imGCs across ages. a**, UMAP plots showing scRNA-seq and snRNA-seq datasets of human brain specimens coloured by four broad cell classes (top rows) and by similarity score to prototypical imGCs (bottom rows). Datasets in bold were integrated and are shown in aggregate for each age group (with four or five subjects for each age). GW, gestational week. **b**, Quantification of proportions of imGCs (with similarity scores $P \geq 0.85$) among all GCs in each human hippocampal specimen across ages. Prenatal and postnatal data points are fitted separately with generalized linear model fitting (black lines) and 95% confidence intervals (grey shaded areas). Datasets from donors aged 40 to 92 years are highlighted in the inset. **c**, Pearson correlation of gene expression of the corresponding mouse[9] and human imGCs and mGCs.

Fig. 3c,f,i). Therefore, proof-of-principle analysis using well-defined mouse hippocampal datasets demonstrates the efficacy and selectivity of our machine learning-based approach to reliably identify immature neurons with almost no contamination of mature neurons.

We next implemented the same strategy for the human infant hippocampal dataset by training a new scoring model using high-confidence human infant imGC prototype cells (DCX$^+$PROX1$^+$CALB1$^-$ cells from the GC cluster) and prototypes from all major non-neuronal cell types (astrocytes (Astro), OPCs, mature oligodendrocytes (mOli) and microglia) at a 99% accuracy rate (Fig. 1a and Extended Data Fig. 4a). Consistent with their immature nature, the positive gene weights for human imGCs are enriched for gene ontology (GO) terms related to nervous system development, neurogenesis and synaptogenesis, and are closely connected to DCX in the gene network (Extended Data Fig. 4b–d and Supplementary Tables 4 and 5). Immature neurons in the infant human hippocampus were then identified by the trained model using the same conservative similarity score cut-off as for mouse ($P \geq 0.85$) (Fig. 1c). As a validation for cell identities, we compared corresponding cell types at the whole transcriptome level between our model-classified human cell types and published cluster annotations in the adult mouse dataset[9] using an independent random forest classifier[24,32] and found high transcriptomic congruence (Fig. 1d). For example, the identified human imGCs displayed much higher resemblance to mouse neuroblasts and imGCs than to mGCs (Fig. 1d). Human mGCs, defined as cells from the GC cluster with lower similarity scores ($P < 0.85$) to the imGC prototype, displayed high resemblance to mouse mGCs, but not to neuroblasts or imGCs (Fig. 1d). Despite this general conservation, orthologous positive gene weights generated separately by machine learning models for human and mouse imGCs showed substantial species differences (Fig. 1e and Extended Data Fig. 4e).

To confirm the enriched expression of top weighted genes in imGCs, we screened candidates on the basis of antibody availability and focused on STMN1 (Fig. 1e), a tubulin-depolymerizing protein[33], for immunohistology. In independent infant human dentate gyrus specimens (Supplementary Table 1), 93.8% of DCX$^+$PROX1$^+$ imGCs were STMN1$^+$, whereas 91.4% and 20.7% of STMN1$^+$PROX1$^+$ cells were DCX$^+$ and CALB1$^+$, respectively (Fig. 1f,g). Similar results were obtained in the adult mouse dentate gyrus (Extended Data Fig. 4f,g).

## Human imGC abundance across lifespan

We then applied our trained model to assess each cell in query scRNA-seq or snRNA-seq datasets from human brain specimens of various developmental stages and regions (Supplementary Tables 2 and 3). We first examined published datasets of the prenatal human hippocampus[34] and prefrontal cortex[35], both of which contain abundant immature neurons. Indeed, we found a large number of cells with high similarity scores ($P \geq 0.85$), including, most prominently, GCs, some CA neurons and GABAergic neurons in the hippocampus, and some neurons in the cortex (Fig. 2a), suggesting that, as in mice (Extended Data Fig. 3a–c), our approach selects for cells with immature neuronal features but not exclusively for imGCs. We next performed

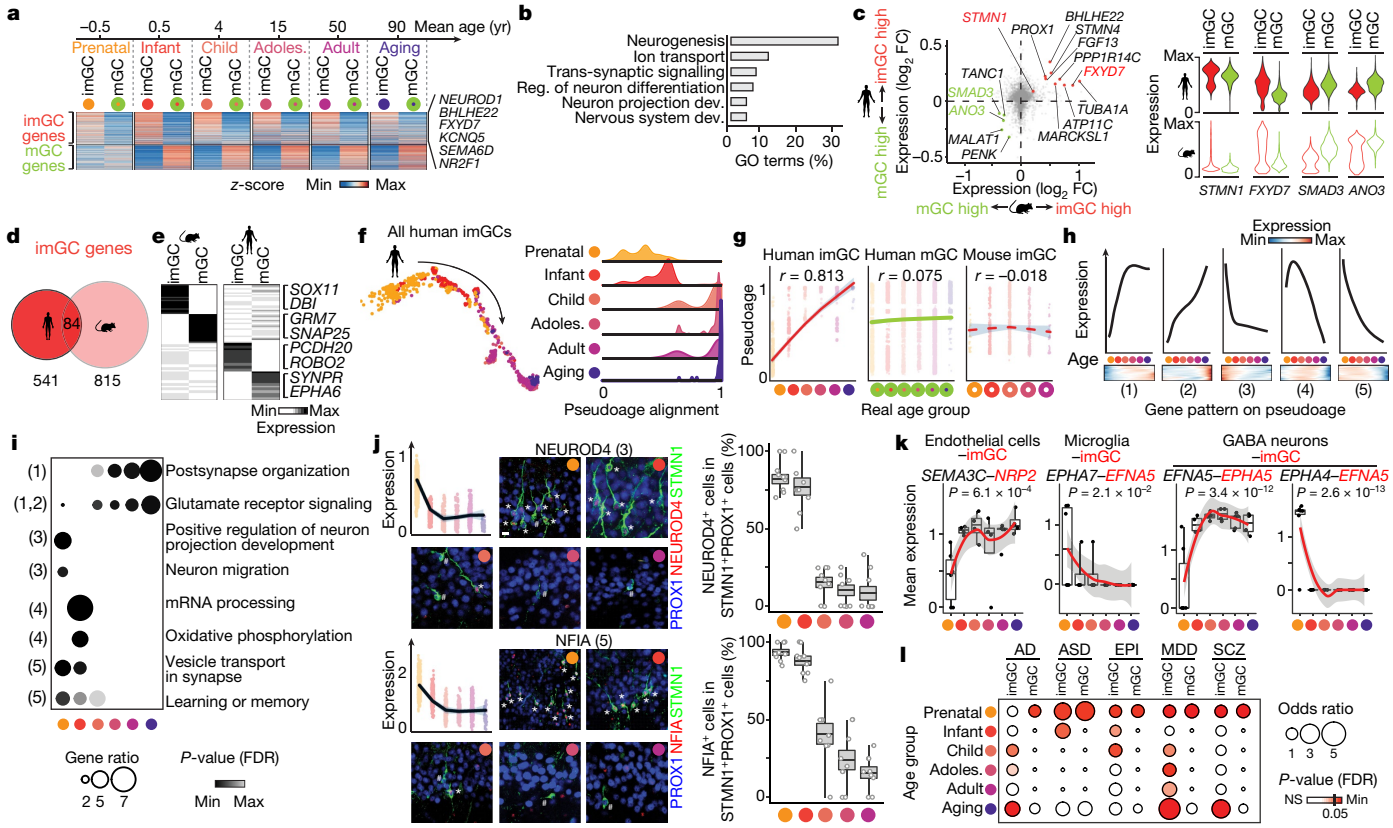



**Fig. 3 | Common and divergent molecular features of imGCs across the lifespan and between humans and mice. a**, Common enriched genes in human imGCs or mGCs across age (two-sided Wilcoxon rank-sum test, false-discovery rate (FDR)-adjusted *P*-value < 0.05). Adoles., adolescent. **b**, Top GO term groups for common human imGC-enriched genes. Dev., development; reg., regulation. **c**, Scatter plot (left) showing $\log_2$ fold change (FC) of expression between imGCs and mGCs and violin plots (right) of exemplary genes. **d**, Venn diagram of imGC-enriched genes in humans and mice. **e**, Unique features of imGC- and mGC-enriched genes in humans and mice. **f**,**g**, Pseudo-age alignment of human imGCs coloured by age group, in scatter (left) or density (right) plots (**f**) and summarized (**g**). Dots representing imGCs in each age group are fitted with loess (lines) with 95% confidence interval (grey shaded areas) with Pearson's *r* for correlations of pseudo- and real-age groups (**g**). **h**,**i**, Distinct patterns of age-dependent gene expression in human imGCs (**h**; likelihood ratio test, Benjamini–Hochberg-adjusted *P*, *q* < 0.01) and representative GO terms (**i**; one-sided Fisher's exact test, FDR *P* < 0.05). **j**, Sample confocal immunostaining images and quantification of two exemplary genes displaying age-dependent expression in human imGCs. Scale bar, 10 µm. Dot plots showing gene expression values as in **g**. Box plots as in Fig. 1g ($n = 3$ subjects per group). Asterisks indicate NEUROD4$^+$ or NFIA$^+$ cells and hashtag symbols indicate NEUROD4$^-$ or NFIA$^-$ cells among PROX1$^+$ and STMN1$^+$ imGCs. **k**, Exemplary ligand–receptor pairs of imGCs interacting with neighbouring cell types (using CellPhoneDB[42]) with age-dependent gene expression changes (two-sided Moran's *I* test, Bonferroni *P* < 0.05; $n = 28$ specimens). Dots represent mean expression of the ligand–receptor pair for the cell-type pair in each specimen with fitting as in **g**. In box plots, the centre line represents the median, box edges show quantiles and whiskers extend to maximum and minimum values. **l**, Enrichment patterns of brain disorder risk gene expression in human imGCs and mGCs across the lifespan (one-sided Fisher's exact test, FDR *P* < 0.05). AD, Alzheimer's disease; ASD, autistic spectrum disorders; EPI, epilepsy; MDD, major depressive disorder; SCZ, schizophrenia.

snRNA-seq on human postnatal hippocampal specimens across ages, with four to five subjects each for child, adolescent, adult, and aging stages (Supplementary Tables 1 and 2). We integrated the published prenatal[34] dataset and all of our postnatal hippocampal datasets using canonical correlation analysis[36] (CCA) (Supplementary Fig. 1) and classified 15 cell clusters into 4 broad classes on uniform manifold approximation and projection (UMAP) plots of 6 age groups (Fig. 2a, in bold). Cells with high similarity ($P \geq 0.85$) to the human imGC prototype were identified in every hippocampal specimen across all ages, most of which are clustered together in UMAP plots, suggesting their transcriptomic proximity (Fig. 2a). The identified immature neurons resided almost exclusively in the GC cluster in postnatal datasets (Fig. 2a and Extended Data Fig. 5). We also applied the same model to several published datasets from various postnatal human brain regions (Supplementary Table 3). We identified immature neurons in three published adult human hippocampus datasets[24,25,27] (Fig. 2a). Adult human olfactory epithelium exhibits continuous neurogenesis; we also identified immature neurons, which matched the published annotations based on unsupervised clustering[37] (Fig. 2a). By contrast, almost no immature neurons were identified from datasets of the adult human frontal cortex, cerebellum, visual cortex[38], middle temporal gyrus of the cortex[39] and prefrontal cortex[24] (Fig. 2a and Extended Data Fig. 5). Together, these results demonstrate the sensitivity and specificity of our approach to identify human immature neurons in various brain scRNA-seq or snRNA-seq datasets.

Following identification of imGCs, quantification of all hippocampal datasets showed that the average percentages of imGCs among all cells in the GC cluster in each age group range from 51.8% in the prenatal stage, 9.4% in infancy, to 3.1–7.5% from 4 years old and beyond (Fig. 2b), which are very similar to results reported on the basis of DCX immunohistology[5,21].

As a validation for the imGC identity in our datasets, Pearson correlation analysis showed that the identified human imGCs resembled mouse neuroblasts and imGCs, but not mGCs[9], whereas human mGCs resembled mouse mGCs, but not neuroblasts or imGCs (Fig. 2c). Immunohistological analysis using independent postmortem human dentate gyrus specimens across ages (Supplementary Table 1) showed that over 70% of DCX$^+$PROX1$^+$ imGCs were STMN1$^+$, whereas only 18–39%

of STMN1$^+$PROX1$^+$ neurons were CALB1$^+$ (Extended Data Fig. 6a–d). We confirmed the neuronal identity of STMN1$^+$PROX1$^+$ cells with additional markers (Extended Data Fig. 6e,f).

## imGC molecular profiles across lifespan

Capturing human imGCs by snRNA-seq across the lifespan enables a systematic analysis of their immature neuronal signature and transcriptomic landscape. To account for batch bias prior to quantitative gene expression comparison, we aligned published prenatal[34] and our postnatal human hippocampal datasets across ages using single-cell variational inference[40] (scVI) in addition to CCA[36] (Supplementary Fig. 2). To identify the common molecular signature of human imGCs irrespective of age, we compared imGCs to their mGC counterparts at different ages and found a preferential enrichment of genes in imGCs of all ages related to nervous system development (for example, *NEUROD1* and *BHLHE22*), ion transport (for example, *FXYD7* and *KCNQ5*) and neuron projection development (for example, *SEMA6D* and *NR2F1*) (Fig. 3a,b, Supplementary Fig. 3 and Supplementary Table 6). Among these genes, only 15.5% overlapped with orthologous genes enriched in mouse imGCs across ages[9] (for example, *FXYD7*), indicating substantial interspecies differences (Fig. 3c–e and Supplementary Table 7).

To deconvolve potential age-dependent molecular changes, we aligned all human imGCs on a pseudo-age trajectory using Monocle[41]. We observed a marked temporal transcriptomic shift correlated with the specimen age (Pearson's $r = 0.813$) (Fig. 3f,g). This correlation was unique to imGCs, but not mGCs, and was observed only in humans but not in mice during the time window examined[9] (Fig. 3g and Extended Data Fig. 7). A gene co-variation kinetics analysis encapsulated five distinct age-dependent patterns, including a continuous upregulation of glutamate receptor signalling pathways and downregulation of neuronal migration- and projection morphogenesis-related genes (Fig. 3h,i, Supplementary Fig. 4 and Supplementary Table 8). We confirmed age-dependent expression of two exemplary genes, *NEUROD4* and *NFIA*, in human imGCs, but consistent expression in mouse imGCs across ages using immunohistology (Fig. 3j and Extended Data Fig. 8). Transcriptomic mapping of different cell types in the hippocampus together enables probing cell–cell interactions on the basis of cognate ligand–receptor expression. An imGC-centric analysis using CellPhoneDB[42] revealed age-dependent interactions between imGCs and their neighbouring cell types in the dentate gyrus (Fig. 3k and Supplementary Fig. 5). To explore the potential contribution of imGCs to different brain disorders, we performed disease-risk gene enrichment analysis for Alzheimer's disease, autistic spectrum disorders, epilepsy, major depressive disorder and schizophrenia, and revealed their selective expression in imGCs at specific ages, many of which coincide with critical periods of the suspected aetiologies of the disorders, such as the aging stage for Alzheimer's disease and early developmental stages for autistic spectrum disorders (Fig. 3l and Supplementary Fig. 6).

## Dysregulated imGCs in Alzheimer's disease

To directly examine how neurological disorders may affect imGCs, we performed snRNA-seq of hippocampal specimens from eight patients with Alzheimer's disease and eight matched controls integrated[36] together with five controls in the original aging group (Fig. 4a and Supplementary Tables 1 and 2). In all these specimens, we identified imGCs almost exclusively in GC clusters, and quantification showed that the percentage of imGCs among all GCs was twofold lower in Alzheimer's disease compared with controls, whereas the percentage of GCs among all cells sequenced per sample was similar (Fig. 4b,c). Our finding is similar to previously reported immunohistological quantification of DCX$^+$ imGCs among NeuN$^+$ GCs regarding the proportion of imGCs and the level of decrease in Alzheimer's disease[5,18].

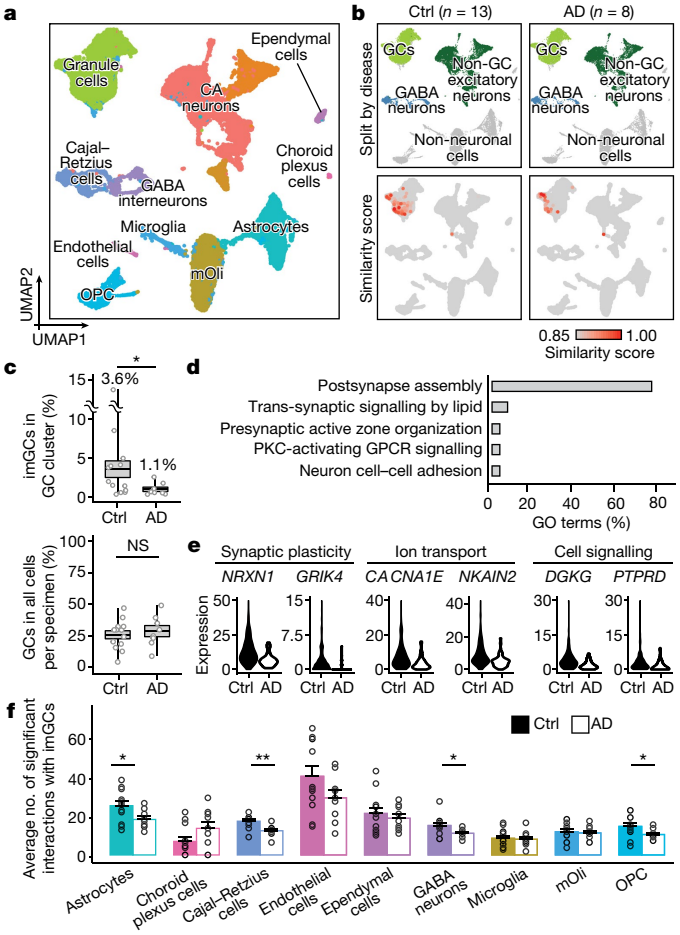



**Fig. 4 | Reduced number and altered gene expression of imGCs in patients with Alzheimer's disease. a**,**b**, UMAP plots of the integrated dataset of patients with Alzheimer's disease and controls (Ctrl) coloured by cluster (**a**) and broad cell class (top row) and similarity score to prototypical imGCs (bottom row) (**b**). **c**, Quantification of proportion of imGCs among GCs (top) and GCs among total cells obtained per specimen (bottom). Each dot represents data from one specimen; the centre line represents the mean, box edges show s.e.m. and whiskers extend to maximum and minimum values ($n = 8$ and 13 individuals for Alzheimer's disease and control, respectively; $^*p = 0.0197$; NS, not significant; one-tailed Mann–Whitney test). **d**,**e**, GO terms (**d**) and examples (**e**) of genes downregulated in imGCs in Alzheimer's disease. **f**, Quantification of the number of significant ligand–receptor pairs of imGCs interacting with neighbouring cell types (using CellPhoneDB[42]). Each dot represents data from one specimen. Data are mean + s.e.m. ($n = 8$ and 13 individuals for Alzheimer's disease and control, respectively; $^*P < 0.05$; $^{**}P < 0.005$; $P$-value of significant pairs from left to right: 0.013, 0.001, 0.017 and 0.012; one-tailed Mann–Whitney test).

Quantitative analysis identified 14 downregulated genes in imGCs in Alzheimer's disease, which are mostly associated with synaptic plasticity and signalling (for example, *NRXN1*) (Fig. 4d,e and Supplementary Table 9). An imGC-centric cell–cell interaction analysis[42] revealed significantly decreased interactions of imGCs with astrocytes, OPCs, GABAergic interneurons and Cajal–Retzius cells in Alzheimer's disease, indicating aberrant niche interactions (Fig. 4f).

## Postnatal human hippocampal neurogenesis

Our transcriptomic analysis could not differentiate whether immature neurons in the adult human brain were born late in life or born earlier and remained in the immature state[8,43]. In an attempt to examine the capacity for neurogenesis, we observed GC fate-specific KI67$^+$PROX1$^+$

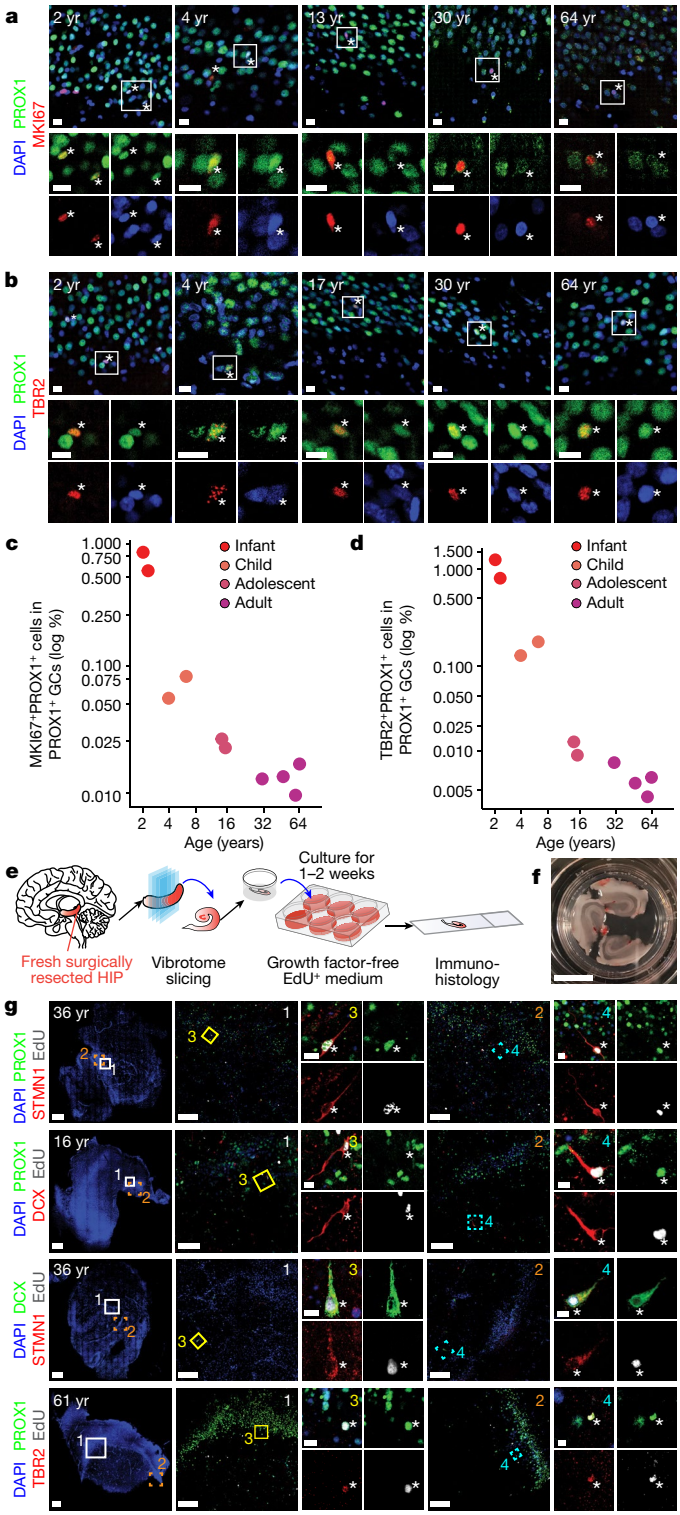



**Fig. 5 | Capacity for neurogenesis in the postnatal human hippocampus across age. a–d**, Sample confocal immunostaining images of MKI67 (**a**) and TBR2 (**b**) and quantification of MKI67$^+$ (**c**) and TBR2$^+$ (**d**) cells among PROX1$^+$ neuronal progenitors in the human dentate gyrus across ages. Scale bars, 10 µm. Asterisks indicate MKI67$^+$ or TBR2$^+$ cells among PROX1$^+$ GCs (**a**,**b**). Each dot represents the sum value of quantification of multiple sections from one specimen ($n$ = 10 specimens) (**c**,**d**). **e–g**, A slice culture system to demonstrate the capacity for neurogenesis in the adult human dentate gyrus. **e**, Schematic illustration of the experimental procedure. **f**, Sample image of a well containing three slices. Scale bar, 1 cm. **g**, Sample confocal staining images of EdU-incorporating newborn imGCs expressing different markers in the postnatal human dentate gyrus. Scale bars, 100 µm (low-magnification images) and 10 µm (expanded insets 3 and 4).

proliferating neural progenitors and TBR2$^+$PROX1$^+$ IPCs[2] in the human hippocampus across ages (Fig. 5a,b). However, the numbers of these precursor cells were very low in the adult human hippocampus (Fig. 5c,d), indicating low frequencies of de novo generation of imGCs.

Because there is almost no practical means to birthdate newborn neurons in humans in vivo, we developed an ex vivo culture method to directly examine the capacity for neurogenesis qualitatively in the postnatal human hippocampus. We culture freshly surgically resected human hippocampi from patients diagnosed with epilepsy in growth factor-free, chemically defined medium[44] in the presence of EdU to label dividing cells (Fig. 5e,f and Extended Data Fig. 9a–c). We observed EdU$^+$PROX1$^+$ newborn GCs after 1–2 weeks in culture in 8 out of 10 specimens from patients ranging in age from 2 to 61 years (Fig. 5g and Supplementary Table 1), and these cells were S100B$^-$ (ref. [19]) and CALB1$^-$ (Extended Data Fig. 9d). More than 80% of EdU$^+$PROX1$^+$ cells were DCX$^+$ or STMN1$^+$, and 88.6% of DCX$^+$EdU$^+$ cells were STMN1$^+$ (Fig. 5g and Extended Data Fig. 9e). We also observed EdU$^+$TBR2$^+$PROX1$^+$ IPCs (Fig. 5g). These results indicate the capacity for the adult human dentate gyrus to generate new GCs and validate the enrichment of STMN1 in imGCs.

## Discussion

Rather than relying on a few pre-selected marker genes[3,5,17–21,45], our study highlights the advantage of the snRNA-seq analysis to precisely define a cell subtype in silico on the basis of its whole transcriptome. Our study reveals dynamic molecular properties of imGCs compared with mGCs in the human hippocampus across the lifespan. As expected, imGCs predominantly express transcripts related to immature neuronal hallmarks, such as development, neurogenesis and plasticity. The significant differences observed in the molecular landscapes between imGCs and mGCs support the notion of unique contributions of human immature neurons to brain functions. We also mined the datasets for novel candidate genes enriched in human immature neurons, such as *STMN1*. Of note, we found differential transcriptional programmes in human imGCs across ages, as well as other properties that diverged from imGCs in mice[9], highlighting interspecies variance. Furthermore, we observed a decreased number of imGCs in Alzheimer's disease and identified altered gene expression and reduced interactions with niche cells. The mechanisms underlying such age- and disease-related changes in human imGCs remain to be determined.

As the similarity score cut-off for human imGCs was determined on the basis of validation in the mouse dataset, we could have under- or over-estimated the numbers of human imGCs in our study. Although we could not determine when these imGCs were born, the presence of GC fate-specific proliferating progenitors in the adult human hippocampus revealed by immunohistology indicates that at least some of imGCs are born in adulthood. Future larger-scale snRNA-seq analysis may capture these rare proliferating neural progenitors for molecular analyses. Our slice culture birth-dating study directly demonstrates the capacity of the adult human hippocampus to generate new neurons. Compared with early development, newborn GCs in adult rodents[46] exhibit immature characteristics for a prolonged period, which lasts longer in aging mice[47,48], and more than ten times longer in adult primates[49,50]. Our results support a model in which new neurons are continuously generated at low frequencies, but exhibit protracted neuronal maturation and are maintained in an immature state for a long period of time, leading to an accumulation of a significant number of neurons with immature neuronal characteristics at any given time in the adult human hippocampus (Extended Data Fig. 10). The function of adult neurogenesis arises primarily from the unique properties of immature neurons[2,4,13–15], rather than proliferating neural progenitors per se.

Our study reveals the molecular landscape of human imGCs across the lifespan and provides resources and methods that will facilitate future investigations into their functions and disease relevance.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04912-w.


1. Gage, F. H. Adult neurogenesis in mammals. *Science* **364**, 827–828 (2019).
2. Ming, G. L. & Song, H. Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron* **70**, 687–702 (2011).
3. Terreros-Roncal, J. et al. Impact of neurodegenerative diseases on human adult hippocampal neurogenesis. *Science* **374**, 1106–1113 (2021).
4. Christian, K. M., Song, H. & Ming, G. L. Functions and dysfunctions of adult hippocampal neurogenesis. *Annu. Rev. Neurosci.* **37**, 243–262 (2014).
5. Moreno-Jimenez, E. P. et al. Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nat. Med.* **25**, 554–560 (2019).
6. Sorrells, S. F. et al. Positive controls in adults and children support that very few, if any, new neurons are born in the adult human hippocampus. *J. Neurosci.* **41**, 2554–2565 (2021).
7. Moreno-Jimenez, E. P., Terreros-Roncal, J., Flor-Garcia, M., Rabano, A. & Llorens-Martin, M. Evidences for adult hippocampal neurogenesis in humans. *J. Neurosci.* **41**, 2541–2553 (2021).
8. Kempermann, G. et al. Human adult neurogenesis: evidence and remaining questions. *Cell Stem Cell* **23**, 25–30 (2018).
9. Hochgerner, H., Zeisel, A., Lonnerberg, P. & Linnarsson, S. Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. *Nat. Neurosci.* **21**, 290–299 (2018).
10. Eriksson, P. S. et al. Neurogenesis in the adult human hippocampus. *Nat. Med.* **4**, 1313–1317 (1998).
11. Spalding, K. L. et al. Dynamics of hippocampal neurogenesis in adult humans. *Cell* **153**, 1219–1227 (2013).
12. Ernst, A. et al. Neurogenesis in the striatum of the adult human brain. *Cell* **156**, 1072–1083 (2014).
13. Schmidt-Hieber, C., Jonas, P. & Bischofberger, J. Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature* **429**, 184–187 (2004).
14. Ge, S., Yang, C. H., Hsu, K. S., Ming, G. L. & Song, H. A critical period for enhanced synaptic plasticity in newly generated neurons of the adult brain. *Neuron* **54**, 559–566 (2007).
15. Marin-Burgin, A., Mongiat, L. A., Pardi, M. B. & Schinder, A. F. Unique processing during a period of high excitation/inhibition balance in adult-born neurons. *Science* **335**, 1238–1242 (2012).
16. Knoth, R. et al. Murine features of neurogenesis in the human hippocampus across the lifespan from 0 to 100 years. *PLoS ONE* **5**, e8809 (2010).
17. Boldrini, M. et al. Human hippocampal neurogenesis persists throughout aging. *Cell Stem Cell* **22**, 589–599.e585 (2018).
18. Tobin, M. K. et al. Human hippocampal neurogenesis persists in aged adults and Alzheimer's disease patients. *Cell Stem Cell* **24**, 974–982.e973 (2019).
19. Ammothumkandy, A. et al. Altered adult neurogenesis and gliogenesis in patients with mesial temporal lobe epilepsy. *Nat. Neurosci.* **25**, 493–503 (2022).
20. Sorrells, S. F. et al. Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* **555**, 377–381 (2018).
21. Cipriani, S. et al. Hippocampal radial glial subtypes and their neurogenic potential in human fetuses and healthy and Alzheimer's disease adults. *Cereb. Cortex* **28**, 2458–2478 (2018).
22. Flor-Garcia, M. et al. Unraveling human adult hippocampal neurogenesis. *Nat. Protoc.* **15**, 668–693 (2020).
23. Kempermann, G., Jessberger, S., Steiner, B. & Kronenberg, G. Milestones of neuronal development in the adult hippocampus. *Trends Neurosci.* **27**, 447–452 (2004).
24. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
25. Ayhan, F. et al. Resolving cellular and molecular diversity along the hippocampal anterior-to-posterior axis in humans. *Neuron* **109**, 2091–2105.e6 (2021).
26. Davila-Velderrain, J. et al. Single-cell anatomical analysis of human hippocampus and entorhinal cortex uncovers early-stage molecular pathology in Alzheimer's disease. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.01.450715 (2021).
27. Franjic, D. et al. Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469.e414 (2022).
28. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e519 (2016).
29. van Galen, P. et al. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. *Cell* **176**, 1265–1281.e1224 (2019).
30. Telley, L. et al. Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522 (2019).
31. Bielecki, P. et al. Skin-resident innate lymphoid cells converge on a pathogenic effector state. *Nature* **592**, 128–132 (2021).
32. Shekhar, K. et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323.e1330 (2016).
33. Gavet, O. et al. The stathmin phosphoprotein family: intracellular localization and effects on the microtubule network. *J. Cell Sci.* **111**, 3333–3346 (1998).
34. Zhong, S. et al. Decoding the development of the human hippocampus. *Nature* **577**, 531–536 (2020).
35. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
36. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
37. Durante, M. A. et al. Single-cell analysis of olfactory neurogenesis and differentiation in adult humans. *Nat. Neurosci.* **23**, 323–326 (2020).
38. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
39. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
40. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
41. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
42. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
43. La Rosa, C., Parolisi, R. & Bonfanti, L. Brain structural plasticity: from adult neurogenesis to immature neurons. *Front. Neurosci.* **14**, 75 (2020).
44. Jacob, F. et al. A patient-derived glioblastoma organoid model and biobank recapitulates inter- and intra-tumoral heterogeneity. *Cell* **180**, 188–204.e122 (2020).
45. Kumar, A. et al. Transcriptomic analysis of the signature of neurogenesis in human hippocampus suggests restricted progenitor cell progression post-childhood. *IBRO Rep.* **9**, 224–232 (2020).
46. Cole, J. D. et al. Adult-born hippocampal neurons undergo extended development and are morphologically distinct from neonatally-born neurons. *J. Neurosci.* **40**, 5740–5756 (2020).
47. Trinchero, M. F., Herrero, M., Monzon-Salinas, M. C. & Schinder, A. F. Experience-dependent structural plasticity of adult-born neurons in the aging hippocampus. *Front. Neurosci.* **13**, 739 (2019).
48. Trinchero, M. F. et al. High plasticity of new granule cells in the aging hippocampus. *Cell Rep.* **21**, 1129–1139 (2017).
49. Kohler, S. J., Williams, N. I., Stanton, G. B., Cameron, J. L. & Greenough, W. T. Maturation time of new granule cells in the dentate gyrus of adult macaque monkeys exceeds six months. *Proc. Natl Acad. Sci. USA* **108**, 10326–10331 (2011).
50. Ngwenya, L. B., Heyworth, N. C., Shwe, Y., Moore, T. L. & Rosene, D. L. Age-related changes in dentate gyrus cell numbers, neurogenesis, and associations with cognitive impairments in the rhesus monkey. *Front. Syst. Neurosci.* **9**, 102 (2015).

# Methods

### Tissue specimens

De-identified human tissue specimens were collected and processed under protocols approved by the Institutional Review Boards of the University of Pennsylvania and the Children's Hospital of Philadelphia. A total of 62 human postmortem hippocampal specimens taken between GW20 and 92 years of age, including 54 specimens from individuals free from neurological disorders and 8 specimens from patients with Alzheimer's disease, were used for snRNA-seq and immunohistological analyses (Supplementary Table 1). Specimens were collected from tissue banks at the Children's Hospital of Philadelphia, the Johns Hopkins University Pathology Archive, the Lieber Institute for Brain Development, and the NIH NeuroBioBank at the following repositories: University of Pittsburgh Brain Tissue Donation Program, the University of Maryland Brain and Tissue Bank, the University of Miami Brain Endowment Bank, the Harvard Brain Tissue Resource Center, the Human Brain and Spinal Fluid Resource Center at the VA West Los Angeles Healthcare Center, and the Mount Sinai School of Medicine. All embryonic tissues were from diagnostic autopsies. As postmortem interval could affect results of snRNA-seq[51] and immunohistology analysis (for example, of DCX[52]), we tried to collect specimens with as short postmortem intervals as possible (listed in Supplementary Table 1). In addition, fresh surgically resected human hippocampal tissue from 10 patients between the ages of 2 and 61 years were used for ex vivo slice culture, collected from the Children's Hospital of Philadelphia and the Hospital of the University of Pennsylvania (Supplementary Table 1). Informed consent for each specimen was obtained by its corresponding institution prior to tissue collection.

For mouse immunohistological analysis, postnatal day (P)14, P60 and 1.4-year-old, wild-type, male and female C57BL/6 mice were used. No obvious sex phenotype was observed in any of the experiments. Animals were housed in a 12-hour dark/ light cycle with food and water ad libitum. Animal procedures were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania.

### Single-nucleus RNA sequencing

We used a modified SPLiT-seq approach for nuclei isolation and snRNA-seq[53,54]. Nuclei isolation from snap-frozen hippocampal tissue was performed as previously described with minor modifications[54,55]. In brief, after a visual inspection to include the dentate gyrus by its distinct anatomical structure, tissue was minced with a razor blade and Dounce (Fisher Scientific, 8853000002) homogenized for 5 to 10 strokes using a chilled tissue grinder in 1 ml of chilled homogenization buffer consist of 1 mM DL-dithiothreitol (DTT, Sigma-Aldrich, D0632), 0.15 mM spermine (Sigma-Aldrich, S4264-1G), 0.5 mM spermidine (Sigma-Aldrich, S0266-1G), EDTA-free protease inhibitor (Roche, 11836170001), 0.3% IGEPAL-630 (Sigma-Aldrich, I8896-50ML), 0.25 M sucrose (Sigma-Aldrich, S5016-500G), 25 mM $MgCl_2$ (ThermoFisher Scientific, AM9530G), and 20 mM tricine-KOH (Sigma-Aldrich, T5816-100G). Homogenates were filtered through a 40-µm cell strainer (Fisher Scientific, 22-363-547), and mixed with 200 µl Myelin Removal Beads II (Miltenyi Biotec, 130-096-733) for a 15-min incubation on ice. The mixture was then transferred on top of a sucrose cushion buffer (0.5 mM $MgCl_2$, 0.5 mM DTT, EDTA-free protease inhibitor, 0.88 M sucrose) at a volume ratio of 1:1 in centrifuge tubes, and centrifuged at 2,800*g* for 10 min in a swinging bucket centrifuge at 4 °C. Nuclei were collected as pellets and resuspended in 0.01% Bovine serum albumin (BSA, Sigma-Aldrich, B6917) in chilled phosphate-buffered saline (PBS, Corning, 21-040-CV). Nuclei were spun down for 3 min at 500*g* at 4 °C before resuspension in 1 ml of chilled PBS-RI (PBS containing 0.05 U $\mu l^{-1}$ RNase Inhibitor (Enzymatics, Y924L)) and filtered through a 40-µm cell strainer. For specimens analysed by the SPLiT-seq method, additional processing steps were applied as follows: 3 ml of chilled 1.33% formaldehyde solution was added to the nuclei suspension for fixation for 10 min. Next, nuclei were permeabilized with 160 µl of 5% Triton X-100 (Sigma-Aldrich, T9284) in chilled PBS for 3 min and spun down at 500*g* for 3 min at 4 °C. Nuclei were then resuspended in 500 ml chilled PBS-RI before 500 µl of chilled 100 mM Tris-HCl (pH 8) was added. Nuclei were spun down again at 500*g* for 3 min at 4 °C and resuspended in 300 µl chilled 0.5× PBS-RI. Finally, nuclei were filtered through a 40-µm strainer again before being counted with a haemocytometer and diluted to one-million nuclei per ml.

The majority of the hippocampal specimens was processed using SPLiT-seq as previously described[53,54]. In brief, single-nuclei mRNA was tagged in three rounds with barcoded primers (Integrated DNA Technologies) for in-cell ligation using the T4 DNA ligase (New England Biolabs, M0202S). Ligation products were purified with Dynabeads MyOne Streptavidin C1 beads (ThermoFisher Scientific, 65001) and resuspended with Kapa HiFi HotStart Master Mix (KAPA Biosystems, KK2600) for a PCR thermocycling. Next, beads were removed from the PCR products, followed by the addition of EvaGreen dye (Biotium, 31000) for a qPCR thermocycling. The PCR products were then purified using KAPA Pure Beads (KAPA Biosystems, KK8000). One hippocampal specimen was processed using droplet-based snRNA-seq technique[56] with modifications[57]. In brief, the single-nucleus suspension and the barcoded beads (ChemGenes, MACOSKO-2011-10) were diluted and co-encapsulated using a microfluidic device (µFluidix, Batch #9508). Droplets were broken and reverse transcription was performed to generate cDNA.

Tagmentation was performed with Nextera XT Library Prep Kits (Illumina, FC-131-1024). The tagmented cDNA libraries were further amplified with 12 enrichment PCR cycles using the Illumina Nextera XT i7 primers and the P5-TSO hybrid primer[56]. After quality control analysis by a Qubit Fluorometer (ThermoFisher Scientific, Q33238) and a Bioanalyzer (Agilent), libraries were sequenced on an Illumina NextSeq 550 instrument using Illumina High Output Kit v2.5 (75-cycle (20024906) for libraries prepared with Drop-seq; 150-cycle (20024907) for libraries prepared with SPLiT-seq). Paired-end sequencing reads were pre-processed using the Drop-seq software (v1.13, http://mccarrolllab.com/dropseq/) with modifications[54,56,57]. In brief, each mRNA read was tagged with a barcode and a unique molecular identifier (UMI), trimmed off sequencing adaptors and poly-A sequences, and aligned to the human reference genome assembly (Genome Reference Consortium hg38, Gencode release v28) using Spliced Transcripts Alignment to a Reference (STAR, v2.5.2a)[58] using default settings. Both exonic and intronic reads mapped to the predicted strands of annotated genes were retrieved for the cell type classification[57]. Uniquely mapped reads were grouped by cell barcodes. To generate a digital expression matrix, a list of UMIs in each gene, within each nucleus, was assembled, and UMIs that differ in just one nucleotide were merged into a single UMI (edit distance = 1). The total number of UMIs was counted and reported as the number of transcripts of that gene for a given nucleus. Raw digital expression matrices were generated for each sequencing run (summarized in Supplementary Table 2).

### Quality control, cell clustering, and dataset integration

Raw count matrices were loaded into the R (v3.6) package Seurat (v3.1.4)[36]. For each specimen, genes expressed in <10 nuclei were discarded. Nuclei with <400 or >5,000 genes were discarded; nuclei with >5% UMIs mapped to mitochondrial genes were discarded. For normalization, UMI counts for all nuclei were scaled by library size (total UMI counts), multiplied by 10,000 and transformed to a log scale. Highly variable genes were identified using the function FindVariableFeatures in Seurat. The top principal components (PCs), determined by the PCElbowPlot function, were selected for dimensionality reduction, clustering and visualization with UMAP or *t*-distributed stochastic neighbour embedding (*t*-SNE). Marker genes for each cluster were identified with a Wilcoxon rank-sum test implemented in the FindAllMarkers function

with the following criteria: adjusted *P*-value < 0.01 (controlled for FDR), log fold change ≥ 0.5, and genes detected in <25% of the cells within its corresponding cluster were excluded. In particular, *PROX1*, the defining marker for excitatory dentate GCs, was used to determine whether a hippocampal specimen contains cells or nuclei from the dentate gyrus. For the hippocampus, only specimens with a distinct *PROX1*-enriched excitatory neuronal cluster containing at least 50 cells or nuclei were included. For non-hippocampal specimens, a distinct *PROX1*-enriched excitatory neuronal cluster precluded further analysis. UMI count matrices from published datasets[9,24,25,27,34,35,37–39] were retrieved from the respective repositories and processed independently using the same criteria (summarized in Supplementary Table 3).

To investigate imGCs across the human lifespan under physiological conditions, hippocampal datasets from our postnatal specimens at the infant (0–2 years old, 4 individuals, 15,434 nuclei), child (3–6 years old, 4 individuals, 24,607 nuclei), adolescent (13–18 years old, 4 individuals, 16,310 nuclei), adult (40–60 years old, 5 individuals, 29,832 nuclei), and aging (86–92 years old, 5 individuals, 16,055 nuclei) stages, and from a published prenatal report[34] (GW16–GW27) were normalized within each age group to remove sequencing variation (implemented in sctransform function in Seurat[59]) prior to integration using a CCA in Seurat[36] (Supplementary Fig. 1 and Supplementary Tables 1 and 2). To investigate imGCs in Alzheimer's disease, hippocampal datasets from our patient specimens (73–88 years old, 8 individuals, 27,508 nuclei) and matched controls (73–88 years old, 8 individuals, 21,955 nuclei) and the 5 control specimens from the aging group (Supplementary Tables 1 and 2) were normalized using the sctransform method[59] and integrated using CCA[36] (Fig. 4a,b). For both integrated datasets, the top 2,000 highly variable genes and the first 30 principal components were used for cell alignment before clustering and UMAP visualization.

### Immature neuron signature extraction and prototype-based cell scoring using machine learning

To provide a precise and holistic characterization of human imGCs, we implemented a supervised learning approach to learn comprehensive gene features from imGCs of unambiguous identities (prototypes), which we then used to quantitatively evaluate the similarity of each cell in query (test) brain snRNA-seq datasets. A multinomial machine learning method using a L2-norm regularized logistic regression model (implemented in the LogisticRegression function in scikit-learn[60] in Python v3.7) was applied with modifications[28]. To validate the sensitivity and specificity of this approach, we first recapitulated our analytic paradigm (Fig. 1a) in scRNA-seq datasets from the mouse hippocampus across ages[9] by selecting prototypes from the early postnatal dataset (P5) for model training, scoring each cell in the juvenile (P12–P35) and the adult (P120–P132) datasets, and benchmarking the classifier performance to the published annotations based on unsupervised clustering and known marker expression[9] (Extended Data Figs. 2a–c and 3). In the context of the mouse dentate gyrus, despite clear separation by unsupervised clustering, we disregarded the finer partitioning of subtypes in the GC lineage to mimic the scenario within human GCs and selected immature neuron prototypes from all the GC clusters in the mouse P5 dataset. Cell selection criteria described below for the human dataset were strictly followed with one difference, which was the lack of mOli as a prototypical cell type, as the mouse P5 dataset does not contain a mOli cluster (Extended Data Fig. 2a).

Given the significant species differences between mice and humans and the potential technical variability in the published datasets, we chose cell-type prototypes from our unsupervised-clustered snRNA-seq dataset of the infant human hippocampi as the training data for the prototype-based scoring model. The prototypes consist of imGCs and all major non-neuronal cell types, including Astro, OPC, mOli and microglia. As *DCX* transcripts are not exclusive to the GC cluster, the imGC prototypes were selected based on consideration of their defining gene expression features, *DCX*$^+$*CALB1*$^-$*PROX1*$^+$, only from the GC cluster, which was included as the only neuronal cell prototype to avoid the potential contamination from non-GC immature neurons in the infant hippocampus. Other cell-type prototypes were selected from their respective clusters by their defining features, including *AQP4*$^+$ cells from the astrocyte cluster, *PDGFRA*$^+$ cells from the OPC cluster, *MOBP*$^+$ cells from the mOli cluster, and *CX3CR1*$^+$ cells from the microglia cluster. To further refine the most representative cell populations as prototypes, two negative selection criteria were applied: (1) cells expressing common markers of the other prototypical cell types were excluded from prototypes. For example, imGC prototypes were expected to exhibit no expression of markers of astrocytes (*SLC1A2* and *AQP4*), OPCs (*PDGFRA*), oligodendrocyte lineage cells (*OLIG2*, *CNP*, *MBP* and *MOBP*) and microglia (*CX3CR1* and *PTPRC*); (2) cells were excluded from all prototypes if they expressed defining markers of other known cell types in the hippocampal dataset, including GABAergic interneurons (*GAD1* and *GAD2*), CA neurons (*SATB2*), ependymal cells (*FOXJ1*), endothelial cells (*FLT1*) and blood cells (*HBA1*).

The cell scoring model was trained on the log-transformed, max-normalized count matrix of the prototype cells with all genes retained, followed by a gene ranking procedure[28,61] to refine for highly variable cell-type specific markers. An optimal regularization parameter of 0.75 for the logistic regression model was chosen by plotting the regularization strength against the classifier accuracy, looking for the most stringent value of regularization with the maximal accuracy rate (~99%). A cross-validation procedure was applied to the training set to estimate the average accuracy of the model (implemented in the LogisticRegressionCV function in scikit-learn) with the following parameters, training set: validation set = 85%:15%, and training set randomly split for 35 iterations (using a stratified *k*-fold cross-validation approach). The resulting trained model uses a list of positively- and negatively-weighted coefficients to rank genes according to their ability to predict each cell category. Importantly, the trained model relies on a combinatorial gene panel rather than a few arbitrarily picked markers to define the transcriptomic profile of imGCs, which strikes a balance among immature, neuronal, and regional (dentate gyrus) features.

A collection of human scRNA-seq and snRNA-seq datasets from various brain regions and developmental stages were individually prepared using Seurat[36] as query (test) datasets (Supplementary Tables 2 and 3). For optimal performance of the machine learning-based classification, we ensured that query datasets have similar sequencing characteristics as our training dataset by performing random down-sampling (implemented in the rbinom function in R) on query expression matrices with significantly higher average number of genes and reads per cell (>2,000 genes per cell and >4,000 reads per cell) to a similar level of depth as our training set (~1,100 genes per cell and ~2,000 reads per cell) prior to quality control and downstream processing on query datasets. This process was repeated ten times to ensure robustness and consistency. The trained model scored the probabilistic similarity (*P*) to each prototype of each individual cell from the log-transformed, max-normalized count matrices of the test datasets without prior knowledge of clustering information. The predicted probability, ranging from 0 to 1, was calculated using the softmax function (implemented in the predict_proba function in scikit-learn)[28]. The $P_{(imGC)} \geq 0.85$ was empirically set as a conservative cut-off of the similarity score to classify a cell as an immature neuron, which was first validated with the mouse dataset. To ensure the specificity of our method, we compared our scoring model predictions on astrocytes, OPCs, mOlis, and microglia using $P \geq 0.85$ as the cut-off of similarity score for cell type classifications to the unsupervised clustering labels in the hippocampal dataset across ages and found that our method classified cells with 94% to 99% specificity. Two types of plots were used for visualization: (1) each cell was plotted on a wheel plot polygon using the polygonalPlot function[28] to show its similarity to each prototype (Fig. 1c and Extended Data Fig. 2f); (2) a similarity score for each cell to imGC prototypes from the test datasets was projected to its corresponding UMAP or *t*-SNE plots (Figs. 2a and 4b and Extended Data Fig. 3b,e,h).

### Comparison of human and mouse cell type classifications

To independently evaluate the transcriptomic similarity of nuclei or cells in the human and mouse dentate gyrus, a previously described multi-class random forest classifier (using the R package 'randomForest' (v4.6.14)[24,32]) was trained on the human infant hippocampal dataset on six cell types, including imGCs and mGCs (determined by the machine learning model), astrocytes, OPCs, mOlis, and microglia. The number of nuclei from each cluster $k$ used as the training set ($N_k$) was determined by $N_k = \min(200, 75\% \text{ of } |\text{nuclei}_k|)$. A common set of highly variable orthologous genes in both human and mouse datasets (identified using homology tables in Ensembl BioMart[62]) was used to train the classifier on $N_k$ for 500 decision trees. The remaining nuclei from each of the six cell types were used as the validation set to estimate the accuracy, resulting in an out-of-bag classification error rate of 11.84%. The classifier was then used to map cells of the corresponding cell types from a scRNA-seq dataset of the adult mouse hippocampus[9] using the published annotations based on unsupervised clustering (Fig. 1d).

In addition, we performed a Pearson correlation analysis (Fig. 2c) comparing human hippocampal imGCs and mGCs across ages determined by the machine learning model to the mouse neuroblasts, imGCs, and mGCs based on the top orthologous highly variable genes. Dataset A in ref. [9] was used as the mouse dataset, where imGCs and mGCs were annotated as 'granule-immature' and 'granule-mature', respectively.

### Deep generative model for batch correction

To eliminate sequencing variation within the human hippocampal datasets, we took precautions prior to quantitative gene expression comparison by correcting the data matrix using scVI[40] (v0.6.8) in Python (v3.7), a neural network-based deep generative modelling method. Consequently, we obtained a shared, batch-corrected latent space among all the human hippocampal datasets across ages with the following parameters: (1) selecting the 20,000 most variable genes (using the subsample_genes function in scVI); (2) training the variational auto-encoder model (VAE) with 90% of cells, holding 10% for validation to monitor overfitting and to measure accuracy (Supplementary Fig. 2a), using 128 hidden layers, and generating 30 latent dimensions (implemented in the UnsupervisedTrainer(model = VAE) function in scVI); and (3) training the model at a learning rate of 1e-3 and 100 epochs (implemented in the 'train' function in scVI). UMAP was applied to the latent dimensions for visualization (Supplementary Fig. 2b). Robust clustering was achieved post-scVI-correction with excellent cluster correspondence to the results from the canonical correlation analysis by Seurat[36], a state-of-the-art cell alignment tool, indicating effective batch correction (Supplementary Fig. 2c). We measured the efficacy of data matrix correction by comparing the expression of the housekeeping 'stably expressed genes'[39,63] across all age groups, benchmarking the efficacy of the SCTransform correction method[59] (implemented in Seurat) (Supplementary Fig. 2d). Prior to the differential gene expression comparisons, mouse datasets[9] across ages were processed separately with scVI for consistency using the same parameters.

### Differentially expressed gene analyses

To investigate differentially expressed genes (DEGs) in imGCs and mGCs across ages under physiological conditions, analysis was performed on the scVI-processed datasets in humans and in mice separately using a two-sided Wilcoxon rank-sum test (implemented in the FindMarkers function in Seurat). For the human DEG analysis, all human imGCs and mGCs across ages were included. Separately for the mouse DEG analysis, all cells in the mouse granule-immature (imGC) and granule-mature (mGC) populations from dataset A in ref. [9] were included. Only DEGs with orthologues in humans (identified using homology tables in Ensembl BioMart[62]) were included for further analyses. To compare gene expression in imGCs between Alzheimer's disease and controls, DEG analysis was performed on the integrated dataset (using the RNA slot of the Seurat object) (Fig. 4a,b) using a two-sided Wilcoxon rank-sum test (implemented in FindMarkers). All imGCs in the Alzheimer's disease and control groups were included. For all analyses, max.cells.per.ident in the FindMarkers function was determined by the cell number of the group with fewer cells for a fair statistical comparison. Genes with an FDR-adjusted $P$-value <0.05 and fold change (log scale, absolute value) >0.1 were considered significantly differentially expressed.

### Pseudotime analysis

The R package Monocle[41] (v2.8) was applied to construct single-cell and single-nuclei pseudo-temporal trajectories of human imGCs, human mGCs, mouse imGCs, and mouse mGCs across ages. The scVI-processed data matrices of the four cell types were individually imported into the Monocle pipeline. The highly variable genes within each cell type across ages, identified heuristically using the vst method (implemented in the FindVariableFeatures function in Seurat), were used to sort cells into a pseudotime order. The DDRTree method was used to reduce dimension (implemented in the 'reduceDimension' function in Monocle). The minimum spanning tree on cells was plotted for visualization (implemented in the plot_cell_trajectory function in Monocle).

Gene expression patterns were grouped after aligning cells on a pseudo-age trajectory (Fig. 3h and Supplementary Fig. 4a). Significant genes were determined using a likelihood ratio test (implemented in the differentialGeneTest function in Monocle), with Benjamini–Hochberg-adjusted $P$-value < 0.01 and $q$-value < 0.01.

### Gene ontology, disease-risk gene enrichment and functional protein association analyses

GO networks of biological processes were built with the ClueGO (v.2.5.5) plug-in[64] in Cytoscape[65] (v3.7.2) with the following settings: 'GO Biological process (January 9, 2020)' was selected; running the default one-sided hypergeometric test, only pathways with FDR-adjusted $P$-value < 0.05 were displayed; the 'GO fusion' option was enabled. Genes identified in the machine learning model or from differential expression analyses were selected as the significantly regulated genes and used as input. For groups with more than 200 significantly regulated genes, a minimum of 7 genes per cluster were used; and for all other groups, a minimum of 3 genes per cluster were used. In addition, the compareCluster function (implemented in the R package clusterProfiler[66]) was applied to obtain representative GO term enrichment patterns (Fig. 3i) with the following parameters, fun = 'enrichGO', ont = 'BP', minGSSize = 3, pAdjustMethod = 'fdr', FDR-adjusted $P$-value <0.05, and $q$-value <0.05.

To map risk genes for brain disorders, we calculated the DEGs of imGCs and mGCs at each age using the 'one_vs_all_degenes' function (implementing the Bayes' method) in scVI[40] using the following parameters, mode = 'vanilla', min_cells = 1, n_samples = 10000. DEGs with natural log Bayes factor ≥1.1 were considered significant. We then analysed the enrichment of the significant DEGs in each category with disease annotations collected from the Phenopedia database[67] (accessed on 25 March 2021) by calculating odds ratios and the enrichment p-values. P-values were determined by a one-tailed Fisher's exact test (implemented in the fisher.test function in R) and corrected by controlling for the FDR for multiple comparisons.

Functional protein association network analysis was performed using the stringApp (v1.5.1) plug-in[68] in Cytoscape[65] with default settings (Extended Data Fig. 4d). First-degree neighbours representing high-confidence connections were calculated with the following parameter: score = 0.35.

### Cell–cell interaction analysis

We applied the CellPhoneDB[42] tool (v2.1.4) with its default settings to infer potential ligand–receptor interactions between imGCs and their neighbouring cell types in the dentate gyrus in each sample, including astrocytes, OPC, mOli, microglia, choroid plexus cells, ependymal cells,

endothelial cells, GABAergic interneurons, and Cajal–Retzius cells. Other excitatory neurons that were spatially separated from imGCs were excluded from the analysis. To investigate imGC niche interactions across ages, the scVI-processed hippocampal datasets (Supplementary Fig. 2) were used. Gene expression levels ≤1.0 were considered negligible and set to 0. To investigate perturbations of Alzheimer's disease on imGC niche interactions, the integrated dataset (using the SCT slot of the Seurat object) (Fig. 4a,b) were used. For both analyses, the mean expression of each ligand–receptor interaction pair for each cell type pair was calculated. A null distribution was generated by a one-sided random permutation of cell type identities over 1,000 times, followed by computation of the mean of each interaction pair for each iteration. The specificity of each interaction pair was determined by comparing the actual mean expression level against the null distribution. Statistically significant ligand–receptor interaction pairs, called using a threshold of *P*-value <0.05, were used to quantify the number of interaction pairs for each cell type pair across ages (Supplementary Fig. 5) or in Alzheimer's disease analysis (Fig. 4f). To determine age-related changes of human imGCs across the lifespan, we further assessed expression patterns of each significant ligand–receptor pair across ages using a Moran's *I* test[69]. A specific cell-type interaction pair with a Bonferroni-adjusted *P*-value <0.05 was considered age-dependent (exemplary gene pairs shown in Fig. 3k).

### Human dentate gyrus ex vivo slice culture analysis

Fresh surgically resected hippocampal tissue was placed in ice-cold sterile cutting solution and taken immediately for vibratome-slicing. The cutting solution, an artificial cerebro-spinal fluid (aCSF) sucrose-based solution, containing 210.3 mM sucrose, 3 mM KCl (Sigma-Aldrich, P9333, CAS 7447-40-7), 1.3 mM $MgCl_2 \cdot 6H_2O$ (Sigma-Aldrich, 442615-M, CAS 7791-18-6), 2 mM $CaCl_2 \cdot 2H_2O$ (Sigma-Aldrich, C3306, CAS 10035-04-8), 26 mM $NaHCO_3$ (Sigma-Aldrich, S5761, CAS 144-55-8), 1.25 mM $NaH_2PO_4$ (Sigma-Aldrich, S3139, CAS 7558-80-7), and 20 mM D-(+)-glucose (Sigma-Aldrich, G6152, CAS 50-99-7), was pre-saturated with carbonated oxygen (95% $O_2$/5% $CO_2$)[70]. Tissue specimens were first visually inspected to ensure inclusion of the dentate gyrus by its distinct anatomical structure. Slicing was performed within a laminar flow biosafety cabinet in continually oxygenated (95% $O_2$/5% $CO_2$) cutting solution, using a Leica VT1200S vibratome at 0.1 mm $s^{-1}$ speed, 1.2 mm vibration amplitude, and with 300 µm thickness interval[54]. In each well, 1.5 ml per well pre-warmed (37 °C) $EdU^+$ BrainPhys medium[44,71], containing BrainPhys Neuronal Medium (StemCell Technologies, 05790), 2% SM1 Neuronal Supplement (StemCell Technologies, 05711), 1% N2 Supplement (ThermoFisher Scientific, 17502048), 1% antibiotic-antimycotic (ThermoFisher Scientific, 15240062), and 1 µM EdU (ThermoFisher Scientific, A10044), was added between a sterile Millicell tissue culture plate well insert (Millipore, PICM03050) and a well of a 6-well plate. Hippocampal slices were transferred and sparsely distributed onto the Millicell well inserts for better access to medium and oxygen during culture. Slices were cultured within a 37 °C, 5% $CO_2$, 90% humidity sterile incubator, with half of the medium in each well replenished with fresh medium every two days. To prepare for whole-mount immunohistological analysis, slices were fixed with 4% paraformaldehyde (wt/vol; in 0.1 M phosphate buffer, pH 7.4) for 4–6 h depending on the tissue size, followed by overnight cryoprotection with 30% sucrose (wt/vol). EdU incorporation was detected using Click-iT EdU Alexa Fluor 647 kit (ThermoFisher Scientific, C10340) prior to primary antibody incubation as previously described[72]. There are limitations for this approach, such as the use of pathological specimens from patients with epilepsy (Supplementary Table 1), and axonal and neuronal injuries from slicing. We characterized the cellular composition, viability, and oxidative stress state[73] of slice cultures and compared them to those of human postmortem specimens by immunohistology (Extended Data Fig. 9a–c). Importantly, to avoid the contribution from FGF-2 on neural progenitor reprogramming[74], we cultured slices in the absence of any exogenous growth factors (EGF or FGF-2) and for a short period of cell culture time to assess the intrinsic capacity for postnatal human neurogenesis.

### Immunostaining and confocal microscopy

Brain tissue sections were pre-treated and immunohistology was performed following a published protocol[22] with modifications for optimal antigen retrieval. In brief, brain tissue blocks were fixed with 4% paraformaldehyde (PFA) at 4 °C for 24–48 h, and cryoprotected with 30% sucrose (wt/vol). Forty-micrometre-thick sections were cut on a frozen sliding microtome (Leica, SM2010R) as previously described[75]. A small proportion of the brain specimens was prepared as formalin-fixed, paraffin-embedded (FFPE) sections. Prior to further pre-treatment, 10 µm FFPE sections were deparaffinized in 4 times xylene (Fisher Scientific, X5-1), 4 times 100% ethanol, and 4 times 95% ethanol, each for 5 min. Tissue sections were incubated with fresh-made 0.5% $NaBH_4$ (Sigma-Aldrich, 213462; in 0.1 N phosphate buffer) for 30 min and washed 4 times with PBS, each for 5 min. The sections then underwent antigen retrieval prior to antibody application by being incubated in 1× target-retrieval solution (DAKO) at 95 °C for 12.5 min, followed by 15 min of cooling to room temperature. Antibodies were diluted in Tris buffered saline (TBS) with 0.1% Triton X-100, 5% (vol/vol) donkey serum (Millipore, S30), and sodium azide (Sigma, S2002, 1:100). Sections were incubated with primary antibodies at 4 °C for five days. The following primary antibodies were applied: ATF4 (CREB-2, rabbit, Abcam, ab28830, 1:250), calbindin (rabbit, Abcam, ab49899, 1:250), calbindin (rabbit, SWANT, D-28k, CB38, 1:500), cleaved caspase 3 (rabbit, Cell Signaling Technology, 9661s, 1:250), doublecortin (rabbit, Cell Signaling Technology, 4604s, 1:500), doublecortin (goat, Santa Cruz Biotechnology, sc8066, 1:250, only works against mouse tissue), IBA1 (rabbit, WAKO, 019-19741, 1:500), Math3 (Neurod4, mouse, Santa Cruz Biotechnology, sc393724, 1:100), MKI67 (mouse, BD Biosciences, 550609, 1:500), NeuN (mouse, Millipore, MAB377X, 1:500), NEUROD1 (mouse, Abcam, ab60704, 1:250), NFIA (mouse, CDI Laboratories, 1.2C6, 1:500), OLIG2 (goat, R&D Systems, AF2418, 1:500), OP18 (Stmn1, mouse, Santa Cruz Biotechnology, sc48362, 1:250, only works against human tissue), PROX1 (rabbit, Abcam, ab101851, ab11941, 1:500), PROX1 (goat, R&D Systems, AF2727, 1:500), S100b (rabbit, Sigma, s2644, 1:500), STMN1 (goat, GeneTex, GTX89411, 1:500, only works against mouse tissue), STMN1 (rabbit, Abcam, ab24445, 1:500), and Tbr2 (Eomes, rabbit, Abcam, ab216870, 1:250). The cyanine (Cy)-conjugated secondary antibodies raised in donkey (Jackson ImmunoResearch; 1:300), including Cy2 anti-goat (705-225-147), mouse (715-225-151), rabbit (711-225-152), Cy3 anti-goat (705-165-147), mouse (715-165-151), rabbit (711-165-152), Cy5 anti-goat (705-175-147), mouse (715-175-151), and rabbit (711-175-152), were incubated at room temperature for 2 h along with DAPI (Roche, 10236276001, 1:1000). After washing with TBS, sections were incubated with 1× TrueBlack (Biotium, 23007; diluted 1:20 in 70% ethanol) for 1 min to block the autofluorescent lipofuscin and blood components. After washing with PBS, stained sections were mounted and imaged as *Z*-stacks on a Zeiss LSM 800 confocal microscope (Carl Zeiss) using a 20× or a 40× objective with Zen 2 software (Carl Zeiss).

### Image processing and data analysis

All confocal images were blindly acquired among different specimens under the same laser power and gain, and analysed as Z-stacked images using Imaris 9.0 software (BitPlane) as previously described[76,77]. The Spots module in Imaris was used to digitize cell-nucleus locations in 3D space and to code cell type classifications according to distinct morphological and molecular markers. A minimum of three randomly chosen areas of equal dimensions within each dentate gyrus tissue were quantitated. The sum of quantifications of these areas per section was considered as one data point. In Fig. 5c,d, owing to the sparsity of marker positive cells, quantification of all sections from one patient specimen were summed as one data point. No statistical methods were used to predetermine sample size.

To quantify total numbers of PROX1$^+$ GCs (Fig. 5c,d) or DAPI$^+$ cells (Extended Data Fig. 9b), semi-automated nuclear staining quantification was performed using Fiji-ImageJ[78]. DAPI and PROX1 channels of confocal image files (.czi) were converted to .tiff format and imported to ImageJ. The DAPI channel was utilized to manually select the sub-granular zone and GC layer as a region of interest (ROI) in each image using the 'polygon' tool. The resulting cropped image was utilized to generate individual ROIs for each nucleus in an image by background subtraction with a rolling ball radius of 50, auto-thresholding with the default algorithm, despeckling, nuclear segmentation using the watershed function, and finally ROI generation via the Analyze Particles function with a minimal size of 5 and circularity 0.2 to 1.0. For quantification of number of PROX1$^+$ GCs (Fig. 5c,d), the corresponding PROX1 channel image for each file was then opened and each nucleus was background subtracted with rolling ball radius of 50. Mean intensities of each nucleus within the previously determined ROIs were measured, and results were inspected in the R software with attention to the overall intensity distributions. Thresholds for assigning marker positivity were determined manually by measuring the mean intensity of nuclei with the minimal signal that would have been determined to be marker positive by traditional manual counting. This process was repeated three times for each image file and results were averaged to ensure consistency and reproducibility.

### Quantification and statistical analysis

The studies were blinded during data collection and quantification. Data in figure panels reflect several independent experiments performed on different days. No data were excluded. An estimate of variation within each group of data is indicated using s.e.m. All data are shown as mean ± s.e.m. All statistical analyses are indicated in the text or figure legends, performed with the R language for statistical computing (v3.6; https://www.r-project.org/).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The snRNA-seq data are available at the Gene Expression Omnibus database under accession numbers GSE185553, GSE185277 and GSE198323. Specimen information and sequencing statistics are described in Supplementary Tables 1 and 2. Sources of the published scRNA-seq or snRNA-seq datasets used in this study are described in Supplementary Table 3. Source data are provided with this paper.

## Code availability

The computational code used in this study is available at GitHub (https://github.com/zhoujoeyyi/humanImmatureNeurons) or upon request.


51. Ferreira, P. G. et al. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nat. Commun.* **9**, 490 (2018).
52. Terstege, D. J., Addo-Osafo, K., Teskey, G. C. & Epp, J. R. New neurons in old brains: a cautionary tale for the analysis of neurogenesis in post-mortem tissue. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.12.468443 (2021).
53. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
54. Qian, X. et al. Sliced human cortical organoids for modeling distinct cortical layer formation. *Cell Stem Cell* **26**, 766–781.e769 (2020).
55. Su, Y. et al. Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat. Neurosci.* **20**, 476–483 (2017).
56. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
57. Hu, P. et al. Dissecting cell-type composition and activity-dependent transcriptional state in mammalian brains by massively parallel single-nucleus RNA-seq. *Mol. Cell.* **68**, 1006–1015 e1007 (2017).
58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
60. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
61. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
62. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
63. Lin, Y. et al. Evaluating stably expressed genes in single cells. *Gigascience* **8**, giz106 (2019).
64. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
65. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
66. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
67. Yu, W., Clyne, M., Khoury, M. J. & Gwinn, M. Phenopedia and genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* **26**, 145–146 (2010).
68. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: network analysis and visualization of proteomics data. *J. Proteome Res.* **18**, 623–632 (2019).
69. Moran, P. A. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
70. Khlghatyan, J. & Saghatelyan, A. Time-lapse imaging of neuroblast migration in acute slices of the adult mouse forebrain. *J. Vis. Exp.* https://doi.org/10.3791/4061 (2012).
71. Bardy, C. et al. Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. *Proc. Natl Acad. Sci. USA* **112**, E2725–E2734 (2015).
72. Berg, D. A. et al. A common embryonic origin of stem cells drives developmental and adult neurogenesis. *Cell* **177**, 654–668.e615 (2019).
73. Lange, P. S. et al. ATF4 is an oxidative stress-inducible, prodeath transcription factor in neurons in vitro and in vivo. *J. Exp. Med.* **205**, 1227–1242 (2008).
74. Palmer, T. D. et al. Progenitor cells from human brain after death. *Nature* **411**, 42–43 (2001).
75. Zhou, Y. et al. Autocrine Mfge8 signaling prevents developmental exhaustion of the adult neural stem cell pool. *Cell Stem Cell* **23**, 444–452.e444 (2018).
76. Sun, G. J. et al. Tangential migration of neuronal precursors of glutamatergic neurons in the adult mammalian brain. *Proc. Natl Acad. Sci. USA* **112**, 9484–9489 (2015).
77. Sun, G. J. et al. Latent tri-lineage potential of adult hippocampal neural stem cells revealed by Nf1 inactivation. *Nat. Neurosci.* **18**, 1722–1724 (2015).
78. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
79. Duque, A., Arellano, J. I. & Rakic, P. An assessment of the existence of adult neurogenesis in humans and value of its rodent models for neuropsychiatric diseases. *Mol. Psychiatry* **27**, 377–382 (2022).
80. Snyder, J. S. Recalibrating the relevance of adult neurogenesis. *Trends Neurosci.* **42**, 164–178 (2019).
81. Marchetto, M. C. et al. Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. *eLife* **8**, e37527 (2019).



**Acknowledgements** We thank all brain donors, patients and their families for the tissue specimens used in this study and members of Ming and Song laboratories for discussion; B. Temsamrit, E. LaNoce and A. Garcia for technical support; A. Angelucci and J. Schnoll for lab coordination; F. H. Gage, G. Kempermann, S. Jessberger and K. M. Christian for comments; O. Spicer, J. Glausier, A. LeFevre, J. Cottrell and J. Hussain for assistance with human tissue acquisition; K. McCanon, Z. Yang, H. Zhu and CDI Laboratories for providing antibody samples; and M. Llorens-Martin for suggestions on immunohistology. Some of the schematic illustrations were modified and/or created with images from https://togotv.dbcls.jp/ or BioRender.com. This work was supported by grants from the National Institutes of Health (K01MH125144 to A.M.B., R35NS097370 to G.-l.M. and R35NS116843 to H.S.), The Lieber Institute for Brain Development (to J.E.K., T.M.H. and D.R.W.), and Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (to G.-l.M.).


**Author contributions** Y.Z. performed bioinformatics and immunohistological analyses and slice culture experiments. Y. Su performed snRNA-seq experiments. S.L. performed snRNA-seq library generation with the help of L.L. D.Y.Z. performed cell-interaction analysis. A.M.B. contributed to the ex vivo slice culture idea. F.J. contributed to the initial slice culture setup. B.C.K., T.L. and H.I.C. provided surgical specimens for slice culture with the coordination of I.H., S.K.K. and R.D.S. A.N.V. and D.W.N. provided some of the human postmortem specimens for immunohistological validation. B.C.K., J.E.K., T.M.H. and D.R.W. provided some of the human specimens for snRNA-seq. Y. Sun, P.H., H.W. and X.G. contributed to additional data collection. Y.Z., Y. Su, G-l.M. and H.S. conceived the project and wrote the manuscript with inputs from all authors.

**Competing interests** The authors declare no competing interests.

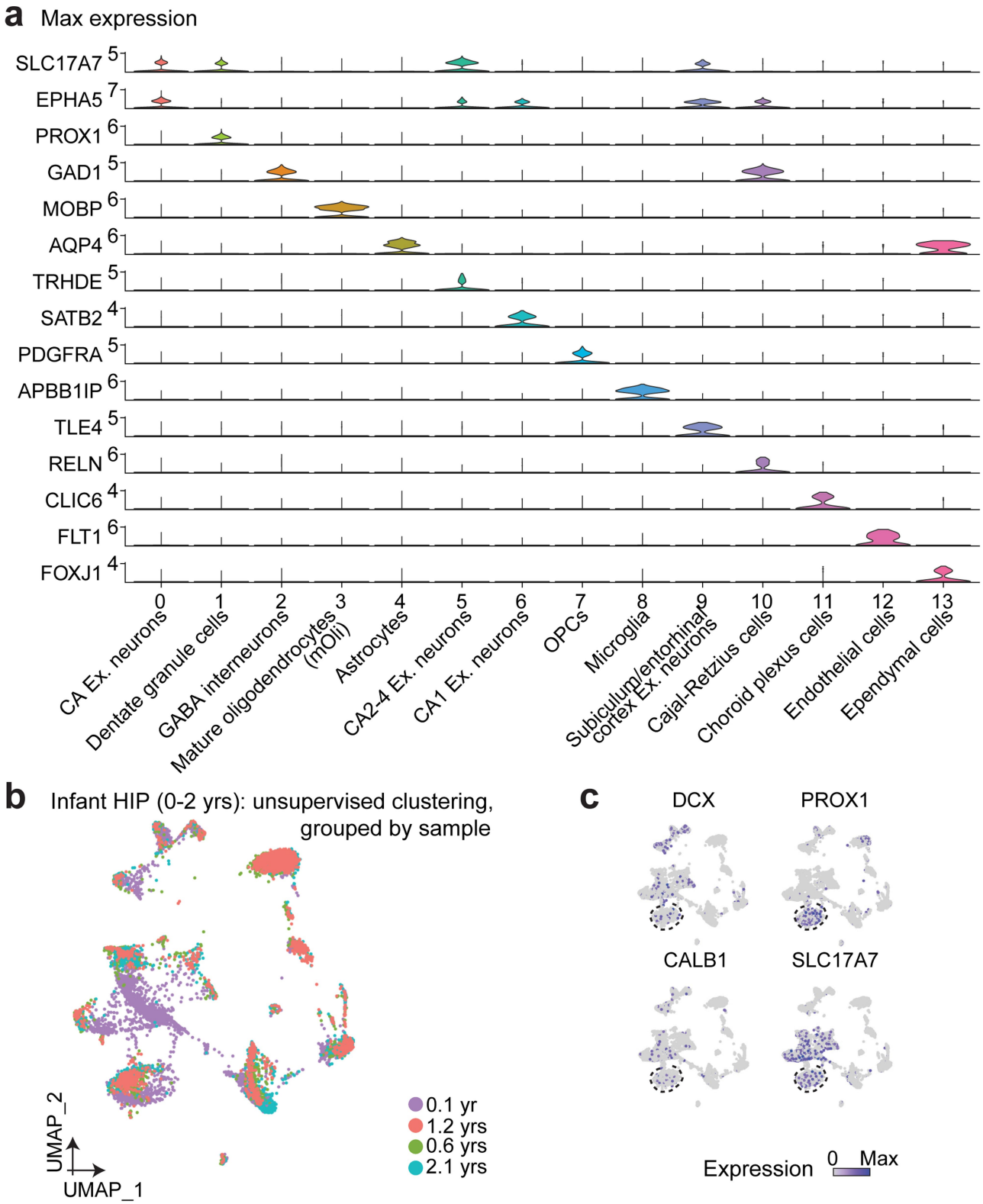
**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04912-w.

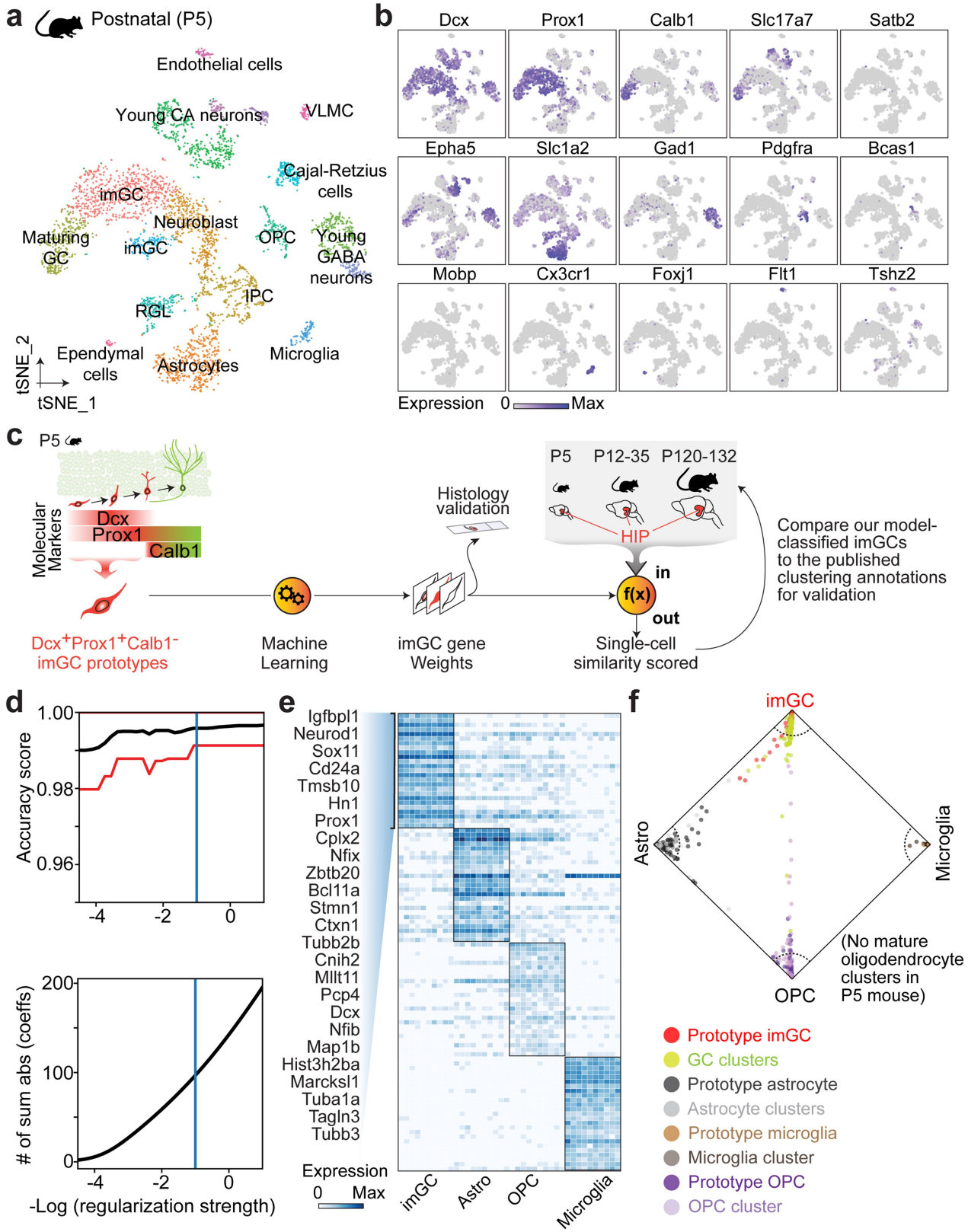

**Correspondence and requests for materials** should be addressed to Guo-li Ming or Hongjun Song.



**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

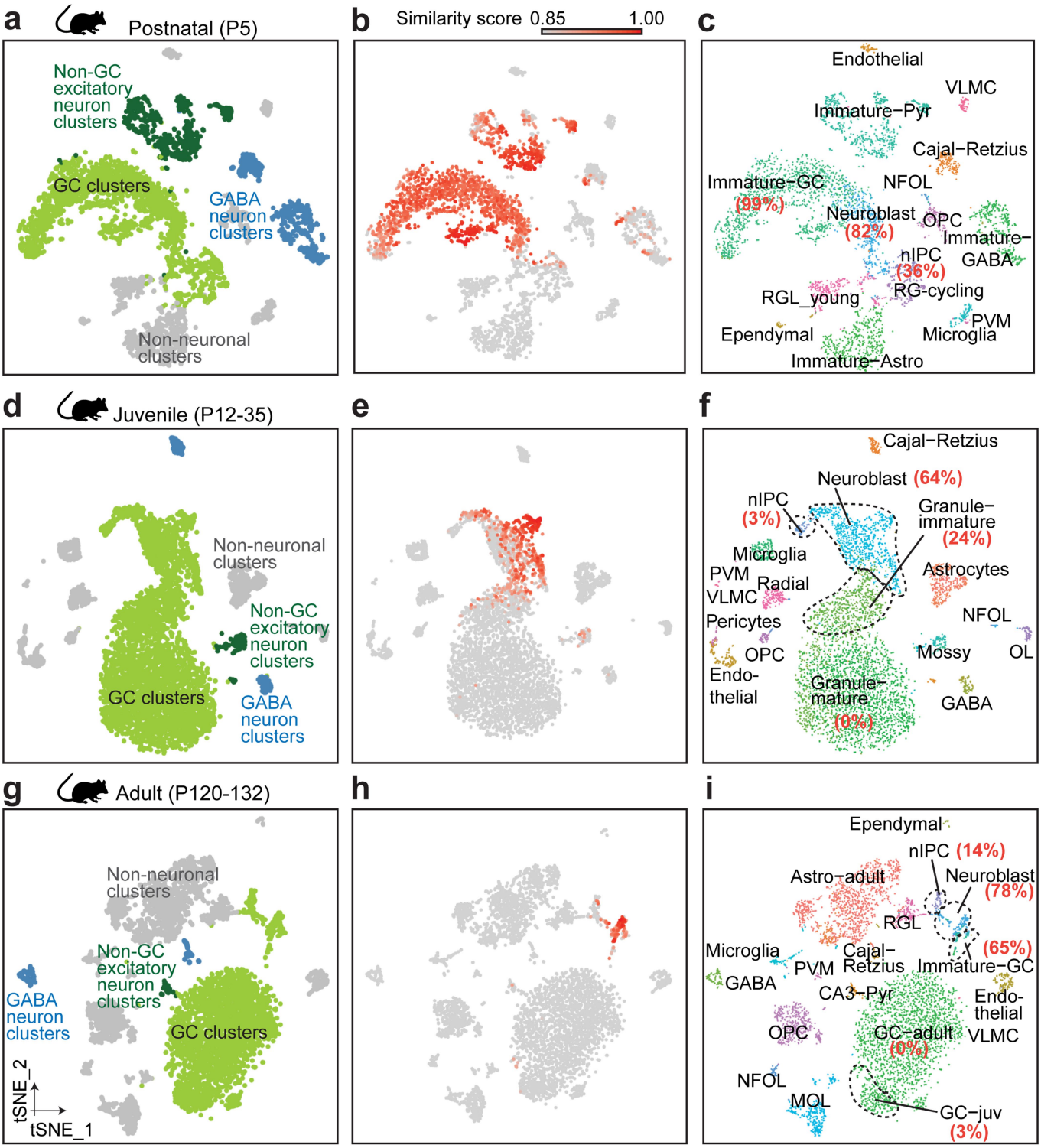



**Extended Data Fig. 1 | Characteristics of the snRNA-seq dataset of the infant human hippocampus. a**, Expression patterns of marker genes used to determine cluster identities. Ex.: excitatory; OPC: oligodendrocyte precursor cells. **b**, Uniform Manifold Approximation and Projection (UMAP) visualization of all cells from the four infant hippocampi (0-2 years) colored by specimen. HIP: hippocampus; yrs: years. **c**, UMAP plots of nuclei from four human infant hippocampal specimens by marker gene expression. The dentate granule cell cluster is highlighted with a dashed line circle.

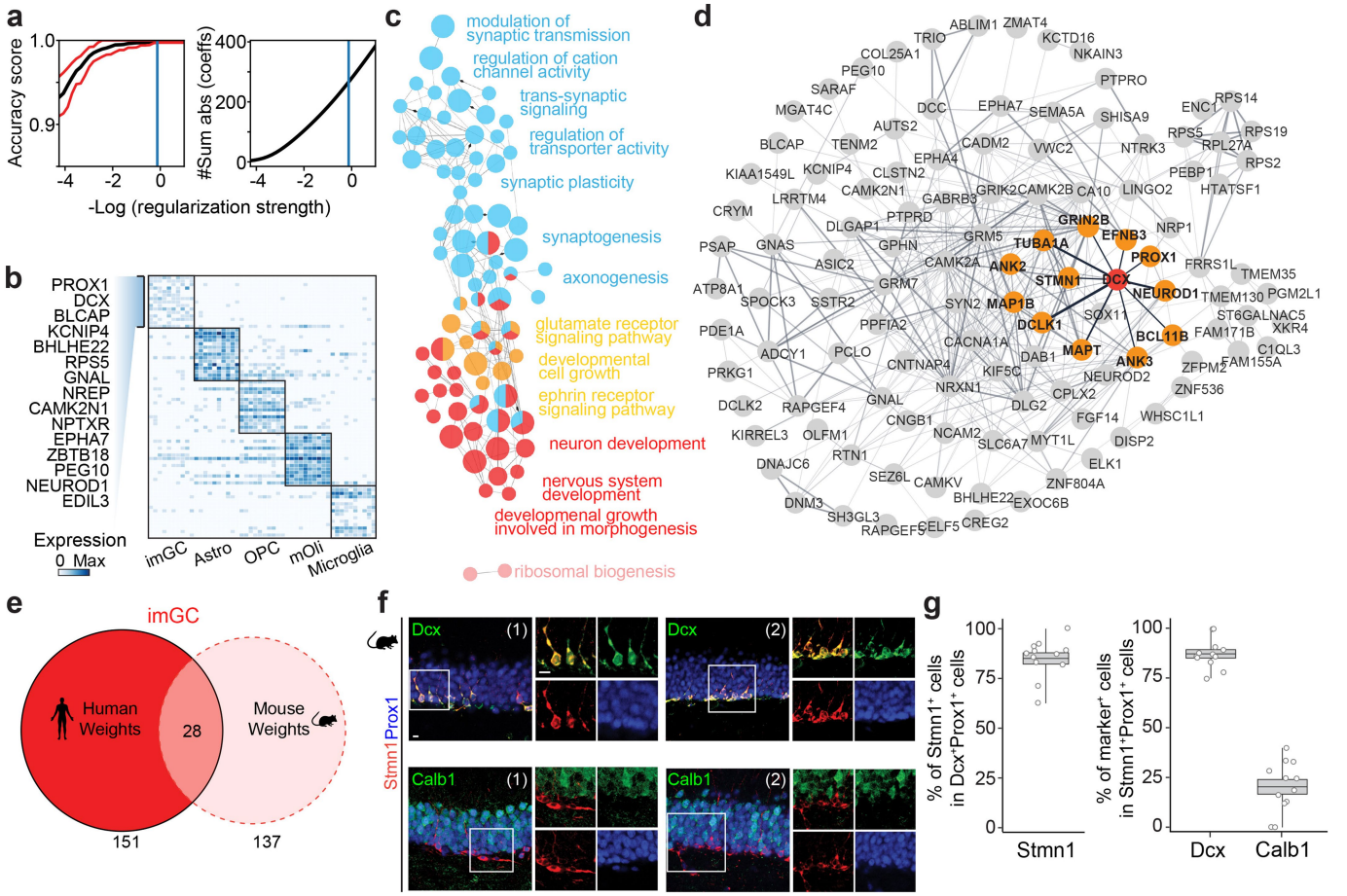



**Extended Data Fig. 2** | See next page for caption.

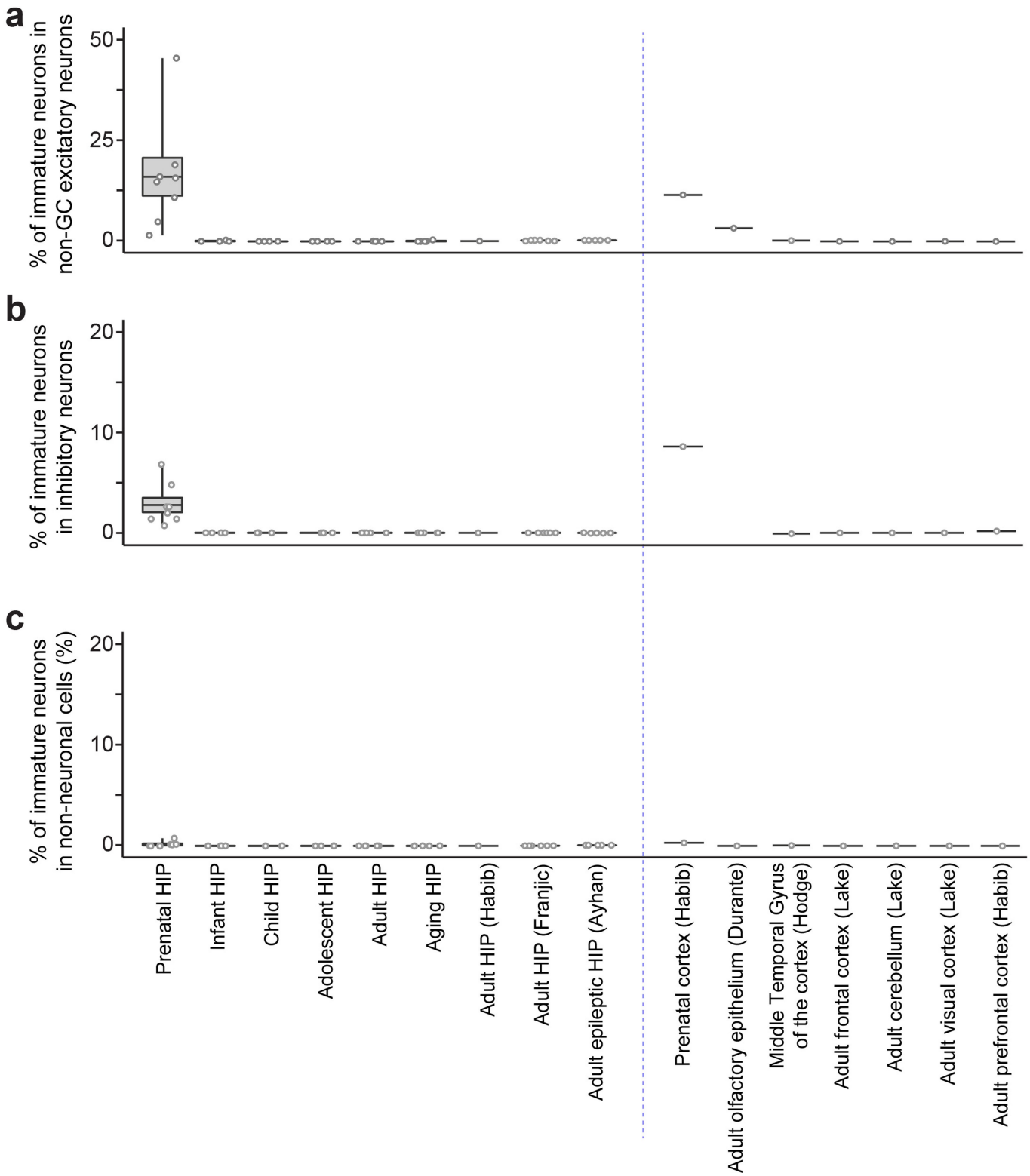
**Extended Data Fig. 2 | Machine learning model trained with the mouse early postnatal hippocampal scRNA-seq dataset. a, b**, Unsupervised clustering and t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of all cells from the mouse postnatal (P5) hippocampus[9] colored by cluster (**a**) and marker gene expression (**b**). imGC: immature dentate granule cell; GC: dentate granule cell; IPC: intermediate progenitor cell; OPC: oligodendrocyte precursor cell. RGL: radial glia-like cell; VLMC: vascular and leptomeningeal cell. **c**, A schematic illustration of the machine learning-aided analysis using the mouse hippocampal scRNA-seq datasets[9], mirroring our analysis pipeline in human studies (Fig. 1a). In brief, $Dcx^{+}Calb1^{-}Prox1^{+}$ imGCs in the P5 mouse dentate gyrus were selected as prototypes to train a scoring model to comprehensively learn their gene features. The trained model containing an aggregate of weighted features ("gene weights") was then used to quantitatively evaluate the similarity of each cell to the imGC prototype in query (test) datasets of the early postnatal (P5; self-scoring), the juvenile (P12-35) and the adult (P120-132) hippocampus[9]. To assess the efficacy of our method, we classified cells with high similarity scores to the imGC prototype as imGCs and compared our model classifications to the published annotations based on unsupervised clustering[9] (Shown in Extended Data Fig. 3). **d**, Measuring performance of the machine learning model. Line plot showing the accuracy score of the machine learning classifier varying with decreasing regularization strength as estimated by cross-validation. Red line shows 95% confidence interval on the estimation of the accuracy score. #Sum abs (coeffs): sum of the absolute value of regression coefficients. **e**, Heatmap showing expression of top-weighted genes in top-scoring cells of each prototype determined by the machine learning model. Genes listed are the top 25 weights defining mouse imGCs. **f**, Wheel plot visualizing the scores of each cell to each prototype. Dots represent individual cells whose distance to each prototype is proportional to the score of that prototype. Red and lime green dots represent the prototypical imGCs and all other GCs, respectively. Dotted line indicates a similarity score of 0.85 to each prototypical cell type. Note that unlike in the human system (Fig. 1c), no mature oligodendrocyte (mOli) cluster was present in the P5 mouse hippocampus.



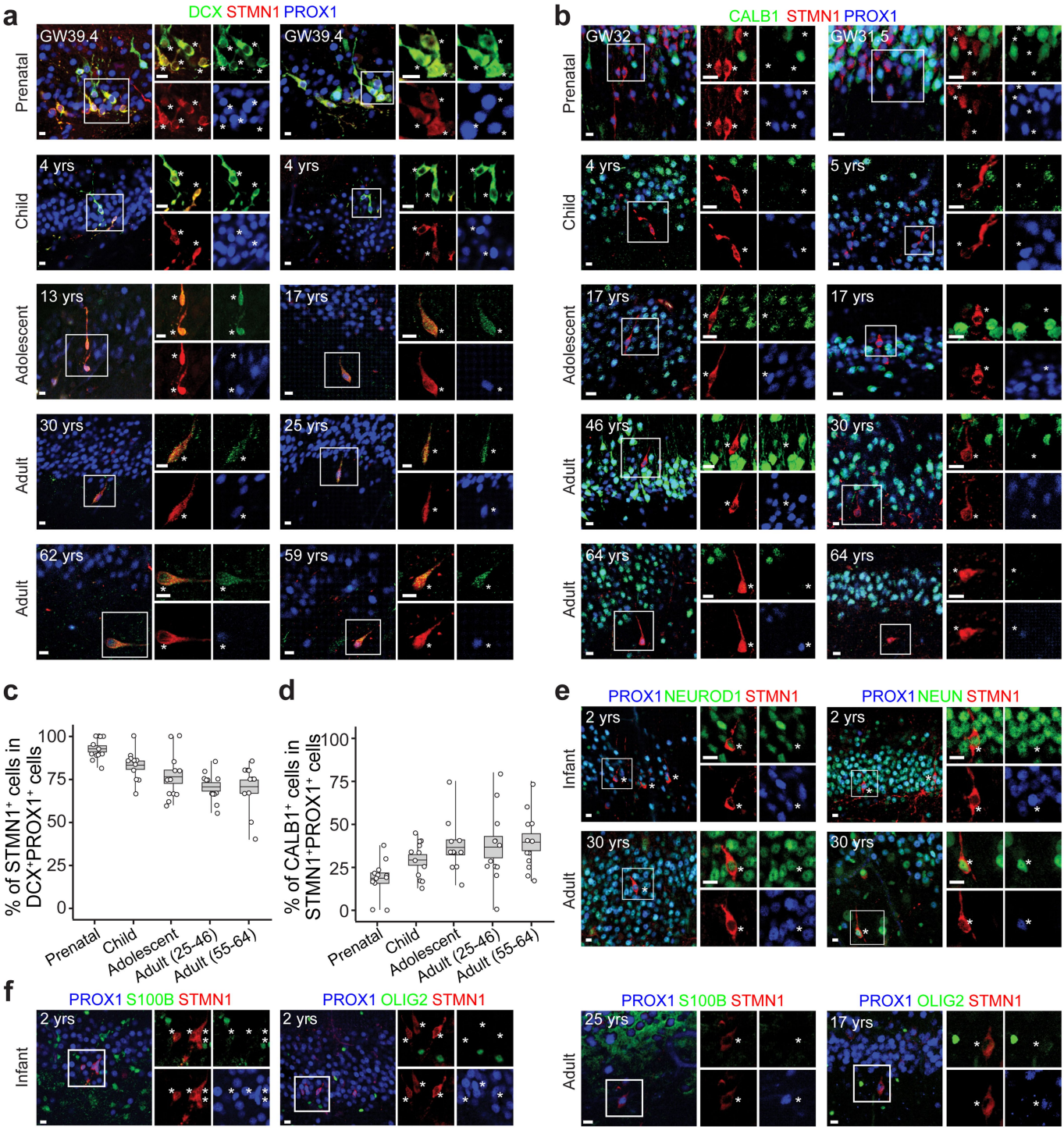

**Extended Data Fig. 3 | Validation of prototype-based scoring of mouse imGCs across ages by the trained machine learning model with published annotations based on unsupervised clustering. a, b, d, e, g, h**, t-SNE visualization of previously published mouse hippocampal datasets[9] at postnatal (**a**), juvenile (**d**), and adult (**g**) stages, colored by four broad cell classes and by similarity score to prototypical imGCs (**b, e, h**). **c, f, i**, Benchmarking cells with high similarity scores ($P \geq 0.85$) with the published annotations[9]. Percentage of cells in the GC lineage clusters (based on published annotations[9]) that are selected as imGCs by our trained machine learning model are indicated in red, bold text.



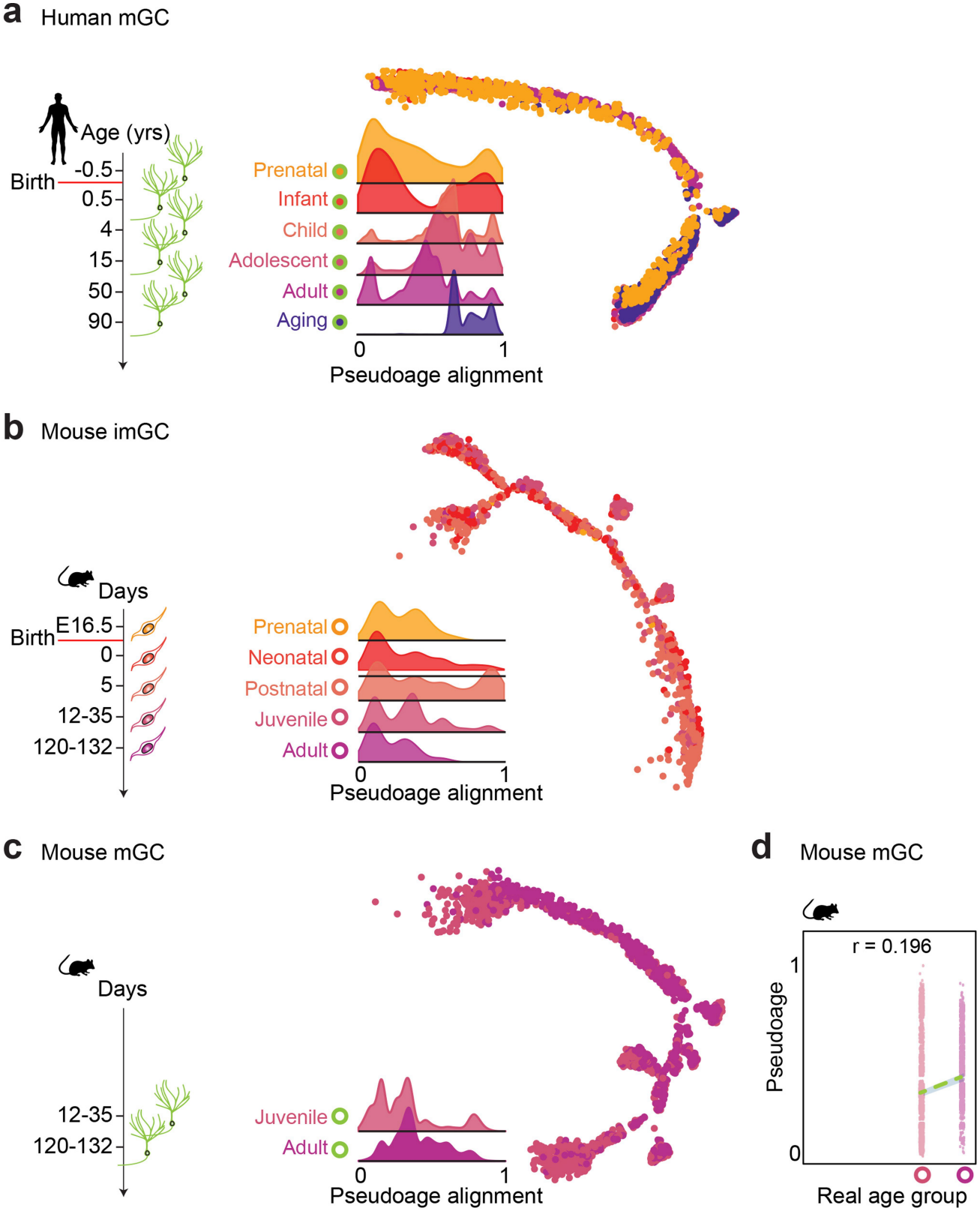

**Extended Data Fig. 4 | Machine learning model performance and feature extraction of gene weights defining human imGCs. a**, Efficacy of the machine-learning approach. Line plot showing the accuracy score of the machine learning model varying with decreasing regularization strength as estimated by cross-validation. Red line shows 95% confidence interval on the estimation of the accuracy score. **b**, Heatmap showing expression of top gene weights in top-scoring cells of each prototype determined by the machine learning model. Genes listed are the top 15 weights defining human imGCs. **c**, Gene ontology (GO) network of biological processes of the positive gene weights defining human imGCs, colored by functionally related ontology group. Only significantly enriched nodes are displayed (one-sided hypergeometric test, false-discovery rate-adjust p value (FDR) < 0.05). The node size represents the term enrichment significance. Examples of the most significant terms per group are shown. See also Supplementary Table 5 for the list of GO terms. **d**, Functional protein-protein association network[68] of the positive gene weights defining human imGCs, highlighting the first-degree neighbors (high-confidence connections) in orange related to DCX. **e**, Overlap of the positive gene weights defining imGCs in humans and in mice that were generated by separate machine learning models. See Supplementary Table 4 for the lists of genes. **f, g**, Immunohistological analysis showing Stmn1 enrichment in immature neurons in the adult mouse dentate gyrus. Shown are sample confocal images (**f**) and quantification (**g**) of Stmn1 expression in imGCs in the adult mouse hippocampus. Individual dots represent value of quantification for different sections (**f**). Scale bars, 10 µm. Box plots similar as in Fig. 1g (n = 4 mice) (**g**).



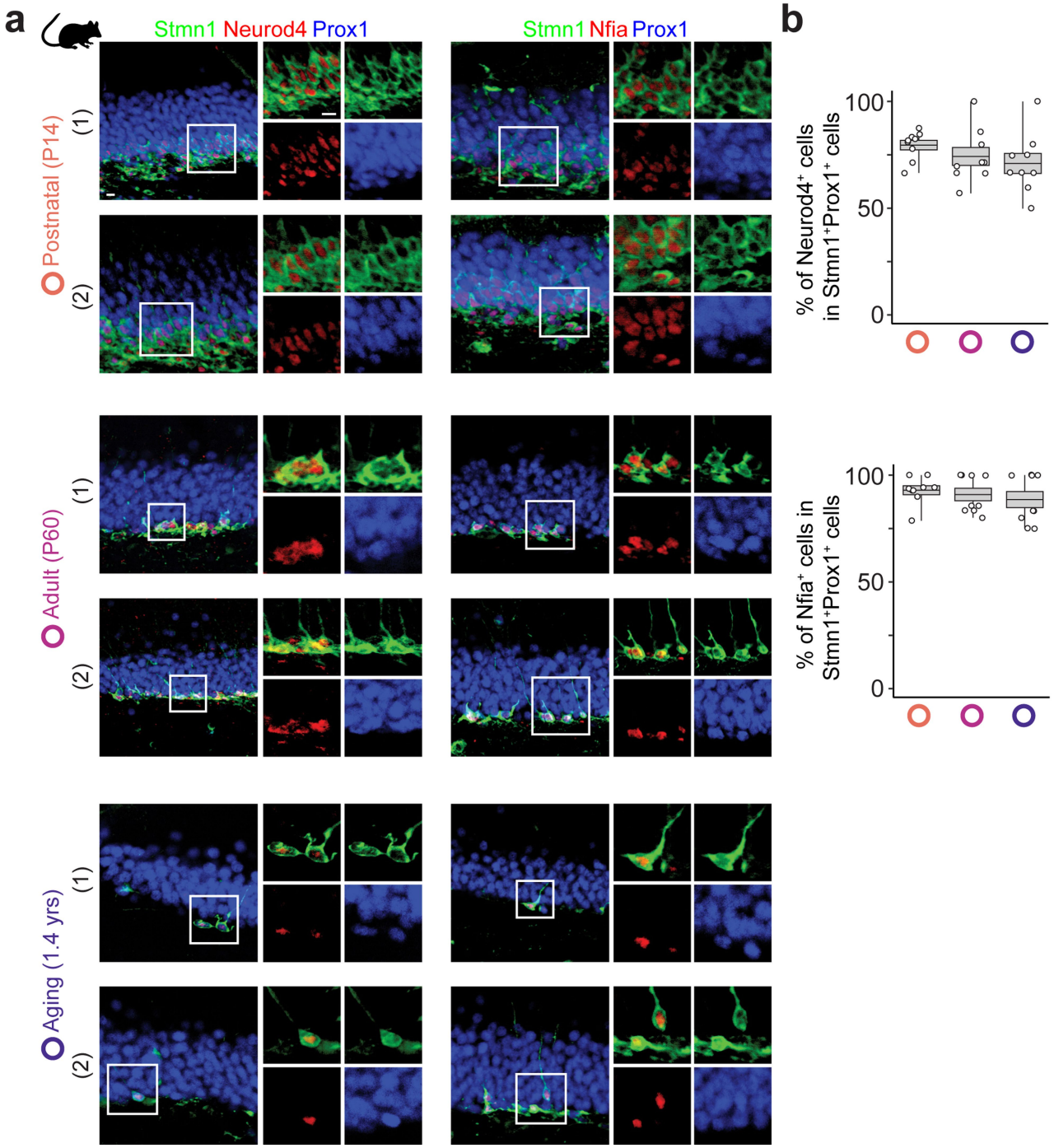

**Extended Data Fig. 5 | Specificity of the machine learning approach for identification of human immature neurons.** The fractions of cells with high similarity scores ($P \geq 0.85$) among non-GC excitatory neuron (**a**), GABA interneuron (**b**), and non-neuronal cell (**c**) clusters in various scRNA-seq or snRNA-seq datasets of the human brains. Box plots represent mean ± s.e.m. with whiskers for max and min. See Supplementary Tables 1, 2, 3 for the specimens used in ours and all published datasets.



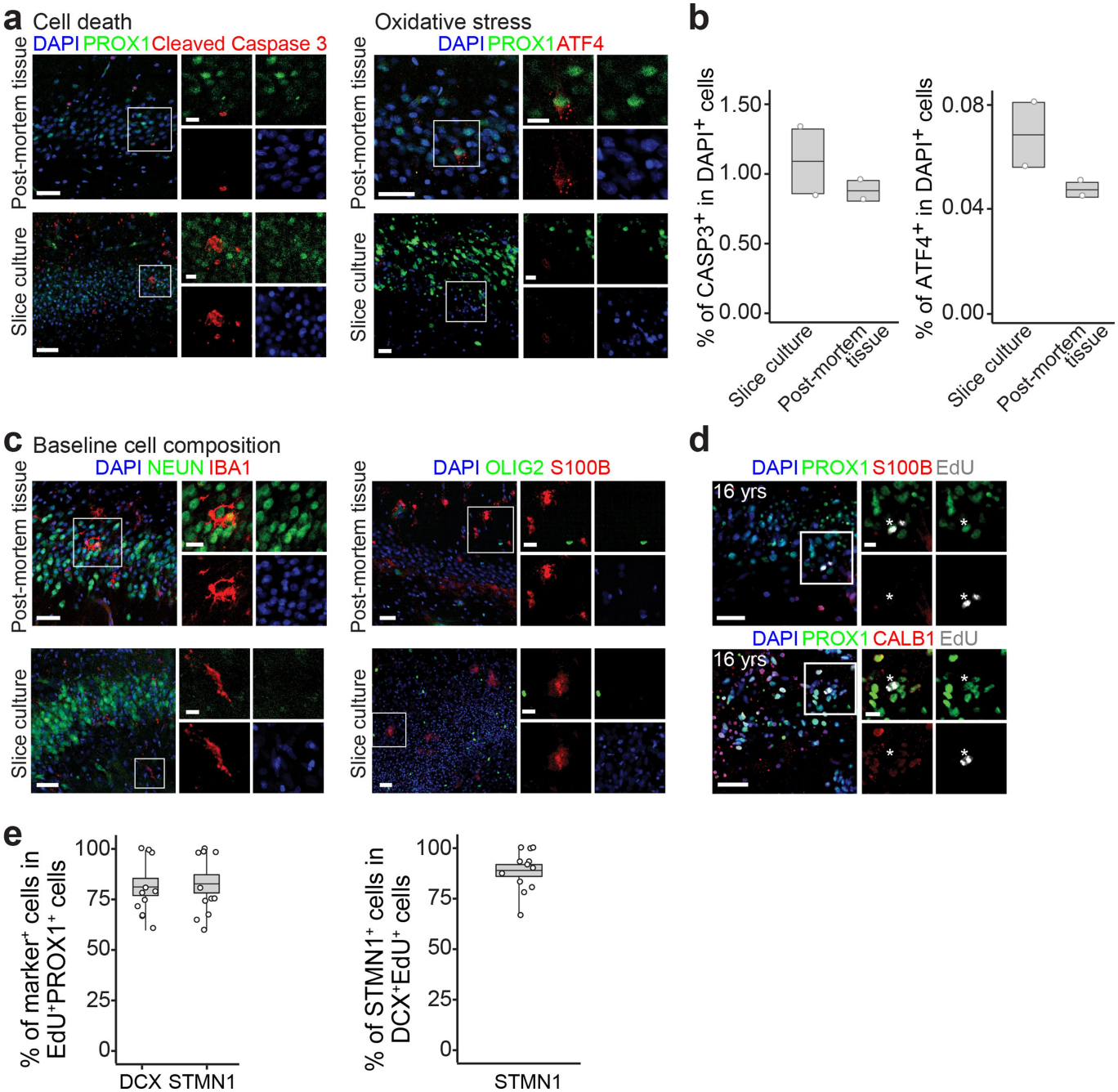

**Extended Data Fig. 6 | Immunohistological analysis of STMN1 enrichment in human imGCs across the lifespan. a-d**, Sample confocal images (**a, b**) and quantifications (**c, d**) of imGCs in the human dentate gyrus across the lifespan. Asterisks indicate DCX$^+$ or CALB1$^-$ among STMN1$^+$PROX1$^+$ GCs (**a, b**). Box plots similar as in Fig. 1g (n = 4 subjects each group) (**c, d**). The immunohistological signal of STMN1 was noticeably more robust than that of DCX in adult specimens. **e, f**, Sample confocal images showing NEUROD1$^+$, NEUN$^+$ (**e**), S100B$^-$, or OLIG2$^-$ (**f**) among STMN1$^+$PROX1$^+$ imGCs in infant or adult human dentate gyrus, confirming their neuronal identity. Asterisks indicate STMN1$^+$PROX1$^+$ imGCs (n = 1 specimen for each immunostaining). All scale bars, 10 µm.



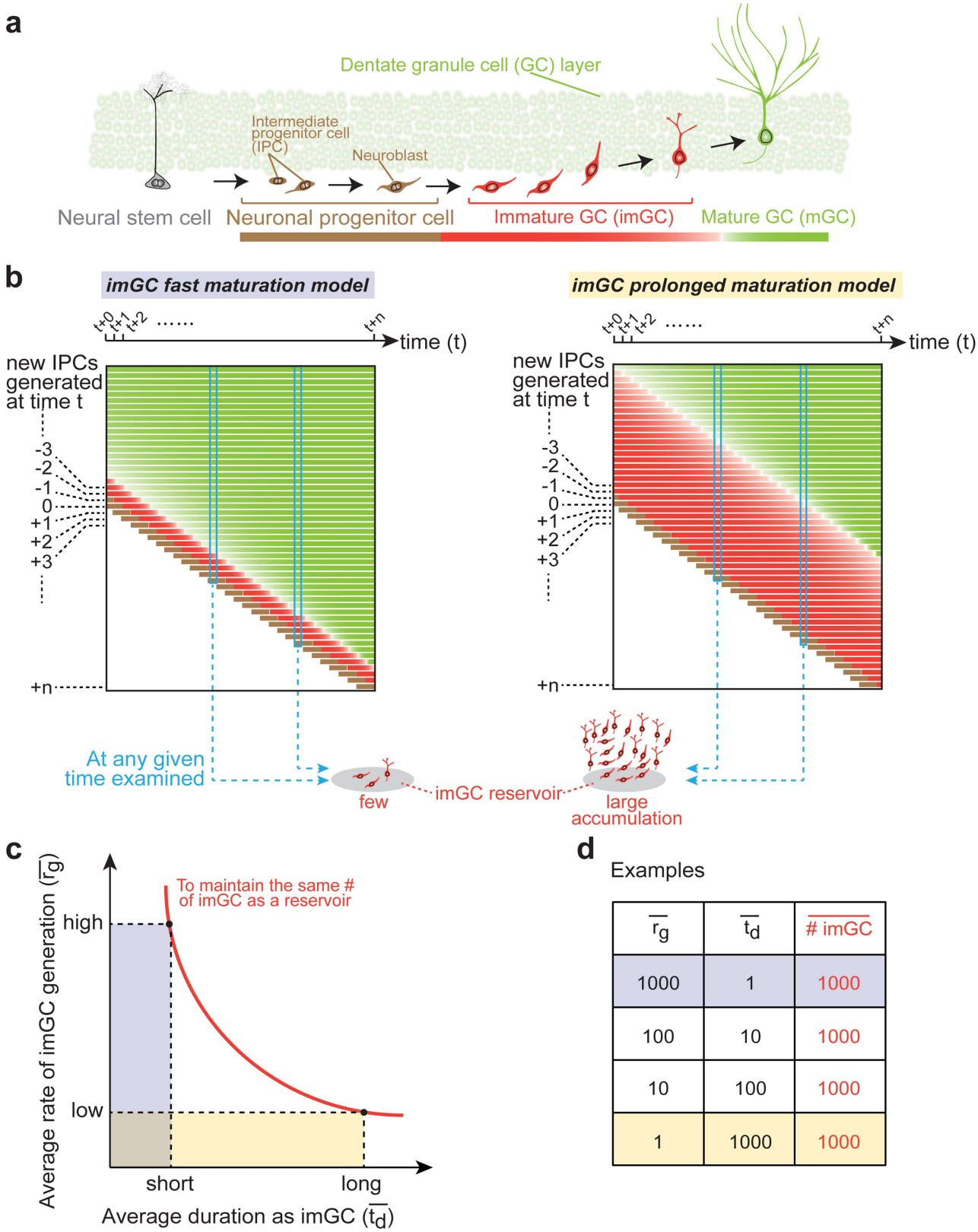

**Extended Data Fig. 7 | Age-dependent transcriptomic dynamics are specific to human imGCs. a-c**, In contrast to human imGCs (Fig. 3f), pseudo-age cell alignment of human mature (**a**), mouse immature (**b**), and mouse mature (**c**) GCs[9] shows very little age-related divergence, visualized as scatter plots. Cells were colored by age group. Distribution of cells within each age group on the pseudo-age trajectory is displayed in the density plots (bottom left). See summary plots in Fig. 3g. **d**, Summary plot comparing pseudo-age alignment (y-axis) of mouse mGCs to real age groups (x-axis), with each mGC of the different age groups plotted as a data point in the background. Data points are fitted with loess fitting (lines) with 95% confidence interval (grey shades). Pearson's r was measured for correlation of pseudo-age and real-age groups. Mouse datasets[9] at prenatal (E16.5), neonatal (P0), or early postnatal (P5) stages do not contain mGC populations.




**Extended Data Fig. 8 | Consistent expression of Neurod4 and Nfia in imGCs of the postnatal mouse hippocampus across ages. a, b**, Sample confocal immunostaining images (**a**) and quantification (**b**) of two exemplary genes that display age-dependent expression patterns in human imGCs (Fig. 3j), but consistent expression in mouse imGCs across ages. Scale bar, 10 µm (**a**). Box plots similar as in Fig. 1g (n = 3 mice per age group) (**b**).




**Extended Data Fig. 9 | Characterization of the slice culture system.** **a, b**, Sample confocal images (**a**) and quantification (**b**) of cell death and oxidative stress level measured in our human hippocampal slice culture in comparison to the post-mortem tissue, using immunohistological analysis of cleaved Caspase 3 and ATF4 (a marker of oxidative stress[73]), respectively. Dots represent value of quantification for individual sections and boxes represent mean ± s.e.m with whiskers for max and min (n = 2 sections) (**b**). **c**, Sample confocal immunostaining images showing baseline cellular composition of slice culture and post-mortem tissue. NEUN$^+$ neurons, IBA1$^+$ microglia, S100B$^+$ astrocytes, and OLIG2$^+$ oligodendrocyte lineage cells were observed. **d**, Sample confocal images showing EdU-incorporated PROX1$^+$ newborn GCs are absent of the astrocyte marker S100B or the more mature neuron marker CALB1 in slice cultures. Asterisks indicate EdU$^+$PROX1$^+$ GCs. For **c, d**, n = 1 section for each immunostaining. Scale bars: 50 µm for main panels and 10 µm for insets. **e**, Quantification of EdU-incorporated newborn imGCs expressing different markers in slice culture of the postnatal human dentate gyrus. Box plots similar as in Fig. 1g (n = 4 subjects).




| $\overline{r_g}$ | $\overline{t_d}$ | # imGC |
|---|---|---|
| 1000 | 1 | 1000 |
| 100 | 10 | 1000 |
| 10 | 100 | 1000 |
| 1 | 1000 | 1000 |

**Extended Data Fig. 10** | See next page for caption.

**Extended Data Fig. 10 | Protracted neuronal maturation leads to accumulation of immature neurons in the presence of low frequency of *de novo* new neuron generation. a**, Process of adult hippocampal neurogenesis[1,2]. Proliferating intermediate progenitor cells (IPCs) and neuroblasts (brown) arising from activated neural stem cells (NSCs, grey) generate new post-mitotic immature dentate granule cells (imGCs, red), which develop over time into mature dentate granule cells (mGCs, lime-green). **b**, An "imGC protracted maturation" model explaining how low-rate, continuous IPC generation can lead to a large number of imGCs as a reservoir, as opposed to a "fast maturation" model. The size of the imGC reservoir in the adult hippocampus depends on a number of factors at the cellular level, such as the rate of stem cell activation and IPC generation, the number of progeny each IPC generates, the percentage of progeny that survives[79], and the duration of imGCs remaining in the immature state, and these parameters may vary tremendously across species and ages[80]. Here we illustrate side-by-side two schematic models showing how changing one factor, the length of imGC maturation duration, alone while keeping all other parameters the same can lead to significant differences in the outcome on the number of imGCs at a given time. For IPCs in a newly generated cohort at a given time t, they go through stereotypical developmental stages to become imGCs and then mGCs (x-axis). At time t+1, a new IPC cohort is generated (y-axis). With all other parameters the same, if the imGCs mature fast, very few imGCs will be observed at any given time (left model). In contrast, if the average length of imGC maturation duration is substantially longer, imGCs in various maturation stages accumulate over time and are present as a large population in any "snapshot" (right model). Prolonged maturation duration of new neurons in the hippocampus has been demonstrated in non-human primates using nucleotide analog tracing analysis to be at least six months[49] and over a year[50]. Furthermore, human induced pluripotent stem cell-derived transplanted neurons display significantly slower maturation compared to those of three non-human primates[81]. **c, d**, An indifference curve qualitatively depicting different combinations of two factors, the average rate of new neuron generation ($\bar{r}_g$) and the average duration of imGC maturation ($\bar{t}_d$), to achieve an equal size of imGC reservoirs (**c**). Hypothetical examples shown in **d**. A significantly longer $\bar{t}_d$ in the adult human hippocampus spares the system from high demand of $\bar{r}_g$ to maintain the same size of imGC reservoir, which is a potential model to explain the seemingly counterintuitive discrepancy between the few IPCs and a large number of imGCs in our results.

nature portfolio

Corresponding author(s): Guo-li Ming, Hongjun Song

Last updated by author(s): Apr 18, 2022

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Sequencing data was obtained from the NextSeq 550 sequencer (Illumina) and was pre-processed using Drop-seq_tools (v1.13, http://mccarrolllab.com/dropseq/, RRID: SCR_018142). Spliced Transcripts Alignment to a Reference (STAR, v2.5.2a, RRID: SCR_015899) was used to align sequencing reads to the human reference genome assembly (Genome Reference Consortium hg38, Gencode release v28). |

Data analysis

#### Bioinformatic analysis:
#### R based:
R (v3.6, RRID: SCR_001905); R Studio (v1.1, RRID: SCR_000432); Seurat (v3.1.4, RRID: SCR_007322); randomForest(v4.6.14, RRID: SCR_015718); biomaRt (v2.46.3, RRID:SCR_019214); Monocle (v2.8, RRID: SCR_016339); sctransform (v0.3); clusterProfiler (v3.18.1, RRID: SCR_016884).
#### Python based:
Python (v3.7, RRID:SCR_008394); Spyder (v4.2.0, RRID:SCR_017585); scikit-learn (v0.24.1, RRID: SCR_002577); backspinpy (v0.2.1); The prototype-based scoring machine learning method and the 'polygonPlot' tool were modified from La Manno et al., Cell 2016 (ref 28; https://github.com/linnarsson-lab/ipynb-lamanno2016); scVI (single-cell Variational Inference, v0.6.8); seaborn (v0.11.1, RRID:SCR_018132); matplotlib (v3.3.2, RRID:SCR_008624); scanpy (v1.6.0, RRID:SCR_018139); pytorch (v1.7.1, RRID:SCR_018536); scipy (1.6.0, RRID:SCR_008058); CellPhoneDB (v2.1.4, RRID: SCR_017054).
#### Others:
Cytoscape (v3.7.2, RRID: SCR_003032); ClueGO (v2.5.5, RRID: SCR_005748); stringApp (v1.5.1).

#### Image analysis:
Imaris (BitPlane, v.9.0, RRID: SCR_007370); Zen 2 (Carl Zeiss, RRID: SCR_013672); Fiji Image J (NIH, v1.53f51, RRID:SCR_003070).

#### Making figures:
Adobe Illustrator CS6 (Adobe, v16.0.3, RRID: SCR_010279).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The snRNA-seq data are available at the GEO database (accession numbers, GSE185553, GSE185277, and GSE198323). Specimen information and sequencing statistics are described in Supplementary Tables 1, 2. Sources of the published scRNA-seq or snRNA-seq datasets used in this study are described in Supplementary Table 3.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

We used a total of 62 human post-mortem specimens and 10 surgical specimens to ensure that there were multiple biological and technical replicates for each experiment. No statistical method was used to predetermine sample size. Sample size was determined by the number of available specimens.

For snRNA-seq analysis of the neurotypical human hippocampus across ages, 4-5 specimens were included in each age group. For snRNA-seq analysis of the human hippocampus in AD, at least 8 specimens were included in each group. Collecting nuclei from multiple donors for sequencing allows us to minimize technical artifacts and ensure that the transcriptomic differences observed between experimental groups are consistent between donors.

For immunohistological analysis of immature neurons in the neurotypical human hippocampus across ages, 4 specimens were included in each age group. For immunohistological analysis of neural progenitor cells in the neurotypical human hippocampus across ages, a total of 10 specimens were included in the experiment. For immunohistological analysis of the wild-type mouse hippocampus, 3-4 mice were included in each age group or per experiment.

Data exclusions

No data were excluded for quantifications of histological experiments.

For snRNA-seq analysis, well-established quality-control filters were applied to each specimen prior to downstream analyses: genes expressed in < 10 nuclei were discarded; Nuclei with < 400 or > 5,000 genes were discarded; nuclei with > 5% UMIs mapped to mitochondrial genes were discarded.

PROX1, the defining marker for excitatory dentate granule cells, was used to determine whether a hippocampal specimen contains cells or nuclei from the dentate gyrus. For each hippocampus specimen, only specimens with a distinct PROX1-enriched neuronal cluster containing at least 50 cells or nuclei were included. For non-hippocampal specimens, a distinct PROX1-enriched neuronal cluster precluded further analysis.

| | |
|---|---|
| Replication | Replication attempts were successful. For snRNA-seq of immature neuron age-dependent features, 4-5 human donors were used per age group. For snRNA-seq of immature neurons in AD, 8 AD patients and 8 sex- and age-matched human donors were sequenced. Data integration methods were applied prior to any quantification or gene expression analysis and we observed that nuclei from different donors were clustered by their cell type but not by donor. Immunohistological experiments were performed on different days using multiple specimens from a total of 24 post-mortem donors and 10 surgical donors. |
| Randomization | For snRNA-seq analysis of the neurotypical human hippocampus across ages, subjects were classified based on their age group: infant (0-2.1 years old), child (3.2-6.7 years old), adolescent (13-18.5 years old), adult (40-60 years old), and aging (86-92 years old). The age numbers were determined by the min and max age of the available specimens we have within each age group.<br><br>For snRNA-seq analysis of the human hippocampus in AD, subjects were classified based on their disease status: AD patients or neurological disease-free "control" donors.<br><br>For immunohistological experiments, subjects were classified based on their age group: prenatal (GW20-GW39.4), infant (0-2.4 years old), child (4-7 years old), adolescent (13-17 years old), younger adults (25-46 years old), and more mature adults (55-64 years old). The age numbers were determined by the min and max age of the available specimens we have within each age group. |
| Blinding | Human specimens were de-identified and assigned with a unique code. Investigators only had access to basic information about donors, such as age, sex, ethnicity, PMI, etc.<br><br>Investigators were not blinded during snRNA-seq analysis as such analysis does not require human quantification, but rather computer-based quantification.<br><br>Investigators were blinded in all immunohistological experiments during tissue processing, image acquisition, and cell quantification. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | Antibodies |
| ☒ | ☐ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology and archaeology |
| ☐ | ☒ | Animals and other organisms |
| ☐ | ☒ | Human research participants |
| ☒ | ☐ | Clinical data |
| ☒ | ☐ | Dual use research of concern |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | ChIP-seq |
| ☒ | ☐ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | The following primary antibodies were used in this study: Atf4 (CREB-2, rabbit polyclonal, Abcam, ab28830, RRID:AB_725570), Calbindin (rabbit polyclonal, Abcam, ab49899, RRID: AB_1267903), Calbindin (rabbit polyclonal, SWANT, D-28k, CB38, RRID: AB_10000340), cleaved Caspase 3 (rabbit polyclonal, Cell Signaling Technology, 9661s, RRID:AB_2341188), Doublecortin (rabbit polyclonal, Cell Signaling, 4604s, RRID: AB_561007), Doublecortin (goat polyclonal, Santa Cruz, sc8066, RRID: AB_2088494), Iba1 (rabbit polyclonal, WAKO, 019-19741, RRID:AB_839504), Math3 (Neurod4, mouse monoclonal (D-10), Santa Cruz, sc393724), Mki67 (mouse monoclonal (B56), BD Biosciences, 550609, RRID:AB_393778), NeuN (mouse monoclonal (clone A60), Millipore, MAB377X, RRID:AB_2149209), Neurod1 (mouse monoclonal (3H8), Abcam, ab60704, RRID:AB_943491), Nfia (mouse monoclonal (1.2C6), CDI Laboratories, RRID: AB_2618885), Olig2 (goat polyclonal, R&D Systems, AF2418, RRID:AB_2157554), Op18 (Stmn1, mouse monoclonal (A-4), Santa Cruz, sc48362, RRID: AB_628297), Prox1 (rabbit polyclonal, Abcam, ab101851, RRID: AB_10712211), Prox1 (rabbit polyclonal, Abcam, ab11941, RRID: AB_562212), Prox1 (goat polyclonal, R&D Systems, AF2727, RRID: AB_2170716), S100b (rabbit polyclonal, Sigma, s2644, RRID:AB_477501), Stmn1 (goat polyclonal, GeneTex, GTX89411, RRID: AB_10726709), Stmn1 (rabbit polyclonal, Abcam, ab24445, RRID: AB_778117), and Tbr2 (EOMES, rabbit monoclonal (EPR21950-241), Abcam, ab216870).<br><br>The following Cyanine (Cy)-conjugated secondary antibodies, all raised in donkey from Jackson ImmunoResearch, were used at 1:300 in this study to detect the binding of primary antibodies: Cy2 donkey anti-goat (705-225-147, RRID: AB_2307341), Cy2 donkey anti-mouse (715-225-151, RRID: AB_2340827), Cy2 donkey anti-rabbit (711-225-152, RRID: AB_2340612), Cy3 donkey anti-goat (705-165-147, RRID: AB_2307351), Cy3 donkey anti-mouse (715-165-151, RRID: AB_2315777), Cy3 donkey anti-rabbit (711-165-152, RRID: AB_2307443), Cy5 donkey anti-goat (705-175-147, RRID: AB_2340415), Cy5 donkey anti-mouse (715-175-151, RRID: AB_2340820), and Cy5 donkey anti-rabbit (711-175-152, RRID: AB_2340607). |
| Validation | Atf4 (rabbit polyclonal, Abcam, ab28830, RRID:AB_725570): validated by manufacturer to detect antigen from human tissue by IHC.<br><br>Calbindin (rabbit polyclonal, Abcam, ab49899, RRID: AB_1267903): validated by manufacturer to detect antigen from human, mouse, and rat tissue by WB and IP.<br><br>Calbindin (rabbit polyclonal, SWANT, D-28k, CB38, RRID: AB_10000340): validated by manufacturer to detect antigen from human, |

monkey, mouse, chicken, rat, and fish tissue by WB and IHC.

cleaved Caspase 3 (rabbit polyclonal, Cell Signaling Technology, 9661s, RRID:AB_2341188): validated by manufacturer to detect antigen from human, monkey, mouse, and rat tissue by WB, IP, IHC, IF, and FC.

Doublecortin (rabbit polyclonal, Cell Signaling, 4604s, RRID: AB_561007): validated by manufacturer to detect antigen from mouse and rat tissue by IF.

Doublecortin (goat polyclonal, Santa Cruz, sc8066, RRID: AB_2088494): validated by manufacturer to detect antigen from mouse, rat, human, and avian tissue by WB, IP, IF, IHC, and ELISA.

Iba1 (rabbit polyclonal, WAKO, 019-19741, RRID:AB_839504): validated by manufacturer to detect antigen from mouse, rat and human tissue by ICC and IHC.

Math3 (Neurod4, mouse monoclonal (D-10), Santa Cruz, sc393724): validated by manufacturer to detect antigen from mouse, rat and human tissue by WB, IP, IF, and ELISA.

Mki67 (mouse monoclonal (B56), BD Biosciences, 550609, RRID:AB_393778): validated by manufacturer to detect antigen from human, monkey, mouse, chicken, rat, and fish tissue by FC and IHC.

NeuN (mouse monoclonal (clone A60), Millipore, MAB377X, RRID:AB_2149209): validated by manufacturer to detect antigen from mouse, rat and human tissue by IHC.

Neurod1 (mouse monoclonal (3H8), Abcam, ab60704, RRID:AB_943491): validated by manufacturer to detect antigen from human tissue by IHC-P, WB, and FC.

Nfia (mouse monoclonal (1.2C6), CDI Laboratories, RRID: AB_2618885): validated by manufacturer to detect antigen from human tissue by WB, IP, and ChIP.

Olig2 (goat polyclonal, R&D Systems, AF2418, RRID:AB_2157554): validated by manufacturer to detect antigen from human tissue by ELISA and WB.

Op18 (Stmn1, mouse monoclonal (A-4), Santa Cruz, sc48362, RRID: AB_628297): validated by manufacturer to detect antigen from mouse, rat and human tissue by WB, IP, IF, IHC(P), and ELISA.

Prox1 (rabbit polyclonal, Abcam, ab101851, RRID: AB_10712211): validated by manufacturer to detect antigen from mouse and human tissue by WB, IP, and IHC.

Prox1 (rabbit polyclonal, Abcam, ab11941, RRID: AB_562212): validated by manufacturer to detect antigen from human tissue by WB.

Prox1 (goat polyclonal, R&D Systems, AF2727, RRID: AB_2170716): validated by manufacturer to detect antigen from human tissue by WB and ELISA.

S100b (rabbit polyclonal, Sigma, s2644, RRID:AB_477501): validated by manufacturer to detect antigen from human, guinea pig, rat, and bovine tissue by IHC/ICC/IF and microarray.

Stmn1 (goat polyclonal, GeneTex, GTX89411, RRID: AB_10726709): validated by manufacturer to detect antigen from mouse and human tissue by WB and ELISA.

Stmn1 (rabbit polyclonal, Abcam, ab24445, RRID: AB_778117): validated by manufacturer to detect antigen from mouse and rat tissue by IHC-FoFr, WB, IHC-P, IHC (PFA fixed), IHC-FrFl, and ICC/IF.

Tbr2 (EOMES, rabbit monoclonal (EPR21950-241), Abcam, ab216870): validated by manufacturer to detect antigen from human, mouse and rat tissue by ICC, IHC, and IP.

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | The wild-type, male and female C57BL/6 mice of different ages were used in this study: Postnatal day 14 (n = 3), Postnatal day 60 (n = 4), and 1.4 years old (n = 3). No obvious sex phenotype was observed in any of the experiments. Animals were housed in a 12-hour light/dark cycle with food and water ad libitum, room temperature: 67-75 °F, humidity: 26-73 %. |
| Wild animals | No wild animals were used in this study. |
| Field-collected samples | No field-collected samples were used in this study. |
| Ethics oversight | Animal procedures were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | For snRNA-seq and histological analyses, a total of 62 human post-mortem hippocampal specimens between the ages of gestational week (GW) 20 to 92 years old, including 54 specimens from subjects free from neurological disorders and 8 specimens from AD patients, were used (Supplementary Table 1). The neurological disease-free donors were separated into six age groups.<br>Prenatal: 2 male, 2 female, median age: GW31.7, median PMI (post-mortem interval): 16.5 hours.<br>Infant: 5 male, 3 female, median age: 1 year, median PMI: 12 hours.<br>Child: 5 male, 3 female, median age: 4.2 years, median PMI: 20.5 hours.<br>Adolescent: 4 male, 4 female, median age: 15 years, median PMI: 18.6 hours.<br>Adult: 10 male, 3 female, median age: 50.2 years, median PMI: 20.5 hours.<br>Aging: 8 male, 5 female, median age: 83 years, median PMI: 12.4 hours.<br><br>For AD patients, they are composed of 4 males and 4 females, with a median age of 80 years, a median PMI of 15.5 hours, and Braak stage ranging from III to VI.<br><br>Tissue samples were visually inspected to include the dentate gyrus by its distinct anatomical structure. For each hippocampus specimen used from snRNA-seq analyses, only specimens with a distinct PROX1-enriched neuronal cluster containing at least 50 cells or nuclei were included.<br><br>In addition, for the slice culture experiment using fresh surgically resected human hippocampal tissue: a total of 10 patients between the ages of 2 to 61 years old were used (5 male, 5 female, median age: 15 years). Tissue samples were visually inspected to include the dentate gyrus by its distinct anatomical structure. |
| Recruitment | De-identified patient information was described in Supplementary Table 1. Informed consent for each specimen was obtained by its corresponding institution prior to tissue collection. All embryonic tissues were from diagnostic autopsies.<br><br>For snRNA-seq and histological analyses: specimens were collected from tissue banks at the Children's Hospital of Philadelphia, the Johns Hopkins University Pathology Archive, the Lieber Institute for Brain Development, the NIH NeuroBioBank at the following repositories: University of Pittsburgh Brain Tissue Donation Program, the University of Maryland Brain and Tissue Bank, the University of Miami Brain Endowment Bank, the Harvard Brain Tissue Resource Center, and the Human Brain and Spinal Fluid Resource Center at the VA West Los Angeles Healthcare Center, and the Mount Sinai School of Medicine.<br><br>For the slice culture experiment using fresh surgically resected human hippocampal tissue: specimens were collected from the Children's Hospital of Philadelphia and the Hospital of the University of Pennsylvania. |
| Ethics oversight | De-identified human tissue specimens were collected and processed under protocols approved by the Institutional Review Boards of the University of Pennsylvania and the Children's Hospital of Philadelphia. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.