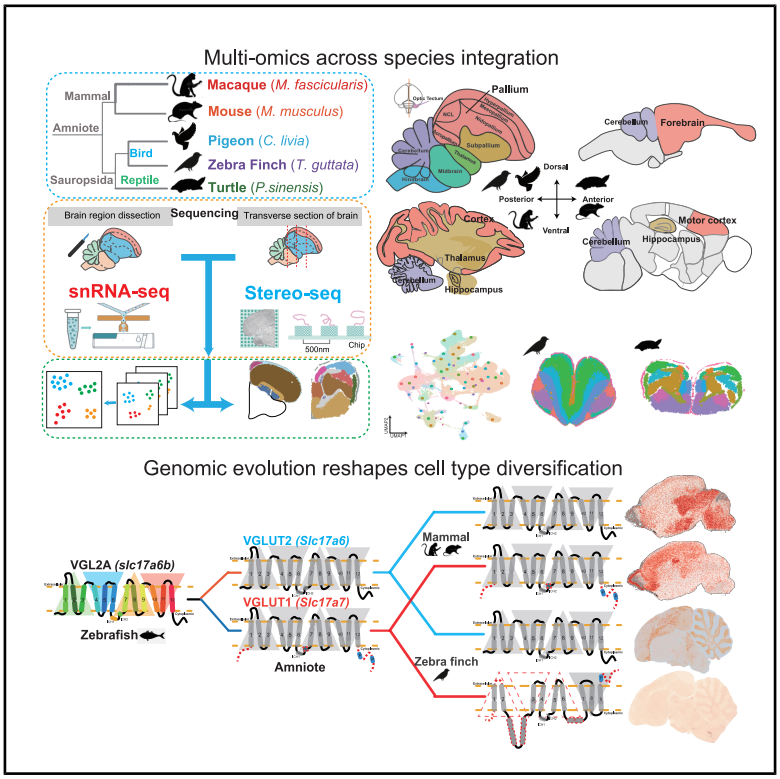
Resource

Developmental Cell

Genomic evolution reshapes cell-type diversification in the amniote brain

Graphical abstract



Highlights

- Comprehensive snRNA-seq atlas of the entire brain in amniote
- Evolution of gene families reshapes diversification in brain cell types
- Mutations in SLC17A6/7 paralogs drive species-specific expression of neurons
- SVIL+ Purkinje cell subtypes promote cerebellar specialization in birds

Authors

Duoyuan Chen, Zhenkun Zhuang, Maolin Huang, ..., Zhenlong Wang, Xun Xu, Shiping Liu

Correspondence

robert@siat.ac.cn (R.K.N.), chun.xu@ion.ac.cn (C.X.), wzl@zzu.edu.cn (Z.W.), xuxun@genomics.cn (X.X.), liushiping@genomics.cn (S.L.)

In brief

Chen et al. construct an over 1.3 million cells cross-species single-nucleus RNA-seq atlas of amniotic brains, uncovering conserved and divergent cell-type evolution. They identified genomic evolution mechanisms driving neuron diversification in the telencephalon, cerebellum, and other regions through shifts in paralogous and positively selected genes.

Chen et al., 2025, Developmental Cell *60*, 1–16
July 7, 2025
https://doi.org/10.1016/j.devcel.2025.04.014

CellPress

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

Developmental Cell

CellPress

Resource

Genomic evolution reshapes cell-type diversification in the amniote brain

Duoyuan Chen,[1,6,11,16,18] Zhenkun Zhuang,[1,4,6,11,16,18] Maolin Huang,[3,18] Yunqi Huang,[1,18] Yuting Yan,[1,4] Yanru Zhang,[1,4] Youning Lin,[1,6,11] Xiaoying Jin,[1,5,11] Yuanmei Wang,[5,6,8] Jinfeng Huang,[2] Wenbo Xu,[3] Jingfang Pan,[1] Hong Wang,[2] Fubaoqian Huang,[1,4,11] Kuo Liao,[1,4] Mengnan Cheng,[1,11] Zhiyong Zhu,[1,5,11] Yinqi Bai,[1]

(Author list continued on next page)

[1]BGI Research, Hangzhou 310030, China
[2]The Institute of Biomedical and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen, 1068 Xueyuan Avenue, Shenzhen University Town, Nanshan District, Shenzhen 518055, China
[3]School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China
[4]School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China
[5]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[6]BGI Research, Shenzhen 518083, China
[7]Institute of Neuroscience, Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, CAS Center for Excellence in Brain Science and Intelligence Technology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[8]HIM-BGI Omics Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou 310018, China
[9]China National GeneBank, BGI Research, Shenzhen 518120, China
[10]Guangdong Genomics Data Center, BGI research, Shenzhen 518120, China
[11]Key Laboratory of Spatial Omics of Zhejiang Province, BGI Research, Hangzhou 310030, China
[12]James D. Watson Institute of Genome Sciences, Hangzhou 310029, China
[13]College of Life Sciences, Northwest University, Xi'an 710069, China
[14]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[15]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

(Affiliations continued on next page)

SUMMARY

Over 320 million years of evolution, amniotes have developed complex brains and cognition through largely unexplored genetic and gene expression mechanisms. We created a comprehensive single-cell atlas of over 1.3 million cells from the telencephalon and cerebellum of turtles, zebra finches, pigeons, mice, and macaques, employing single-cell resolution spatial transcriptomics to validate gene expression patterns across species. Our study identifies significant species-specific variations in cell types, highlighting their conservation and diversification in evolution. We found pronounced differences in telencephalon excitatory neurons (EXs) and cerebellar cell types between birds and mammals. Birds predominantly express *SLC17A6* in EX, whereas mammals express *SLC17A7* in the neocortex and *SLC17A6* elsewhere, possibly due to loss of function of *SLC17A7* in birds. Additionally, we identified a bird-specific Purkinje cell subtype (SVIL+), implicating the lysine-specific demethylase 11 (LSD1)/KDM1A pathway in learning and circadian rhythms and containing numerous positively selected genes, which suggests an evolutionary optimization of cerebellar functions for ecological and behavioral adaptation. Our findings elucidate the complex interplay between genetic evolution and environmental adaptation, underscoring the role of genetic diversification in the development of specialized cell types across amniotes.

INTRODUCTION

Amniotes, which evolved over 320 million years ago (mya) and include reptiles, birds, and mammals, are among the most abundant and widely distributed terrestrial animals. This transition from aquatic to terrestrial habitats accelerated the evolution of amniote brain information-processing capabilities, thereby enhancing adaptation to terrestrial environments.[1] Extensive research has identified significant variations in the morphology and connectivity of the brain throughout the evolution of amniotes.[2–4] Among the amniote taxa, avians represent an independently differentiated branch, characterized by highly specialized

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

Zhiwei Niu,[1,5,11] Ze Zhang,[1,14,11] Ya Xiang,[1,11,13] Xiaofeng Wei,[9,10] Tao Yang,[9,10] Tao Zeng,[1,6] Yuliang Dong,[1,6] Ying Lei,[1,11,17] Yangang Sun,[7] Jian Wang,[1] Huanming Yang,[5,6,8,12] Yidi Sun,[7] Gang Cao,[15] Muming Poo,[7] Longqi Liu,[1,11,17] Robert K. Naumann,[2,*] Chun Xu,[7,*] Zhenlong Wang,[3,*] Xun Xu,[1,6,17,*] and Shiping Liu[1,6,11,16,19,*]

[16]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Hangzhou 310030, China
[17]Shanxi Medical University, BGI Collaborative Center for Future Medicine, Shanxi Medical University, Taiyuan 030001, China
[18]These authors contributed equally
[19]Lead contact
*Correspondence: robert@siat.ac.cn (R.K.N.), chun.xu@ion.ac.cn (C.X.), wzl@zzu.edu.cn (Z.W.), xuxun@genomics.cn (X.X.), liushiping@genomics.cn (S.L.)
https://doi.org/10.1016/j.devcel.2025.04.014

brains. Relative to body weight, the avian brain is notably large and features complex structures and functions.[5,6] The avian brain uniquely exhibits higher neuron and synapse densities compared with other animal groups, as well as distinct telencephalic structure and connectivity patterns relative to mammals.[6,7] The distinctive structure and functionalities of the avian brain are intrinsically linked to their unique survival strategies, serving as a crucial neural foundation for adaptation and reproduction in dynamic and complex environments. By contrast, understanding the mechanism underlying these distinctive structures and functionalities identified a significant gap between complex function and cellular or molecular changes. Gene expression datasets have been employed to address the challenges of comparative analysis for the evolutionary trajectories of brain cell types across multiple species.[2,8,9]

Recently, single-cell RNA sequencing (scRNA-seq) has emerged as the prominent technique for detailed comparative investigations of cell-type composition.[1,9–13] Through comparative analysis, we found that gene family evolution and shifts in paralog expression contribute to cellular diversity.[10] Comparative single-cell sequencing research in birds and mammals has predominantly focused on well-defined and few limited brain subregions.[14–17] Furthermore, studies have generally targeted a narrow range of closely related species.[18] The absence of a comprehensive single-cell atlas of the avian brain limits our detailed understanding of its cellular specificities. Moreover, investigations into more diverse and distantly related species, crucial for understanding the evolutionary dynamics of cell types and brain functions, remain scarce. Challenges in this field partly stem from the absence of a unified data platform and standardized technologies to produce cross-species datasets. Additionally, the field lacks robust integration analysis methods for comparing single-cell data across distantly related species.

Using a standardized single-nucleus RNA sequencing (snRNA-seq) platform, DNBelab C4,[19] we generated a comprehensive cell-type atlas for reptiles (turtles), birds (pigeons and zebra finches), and mammals (macaques and mice). And further validated expressing spatial pattern using spatial transcriptomics with Stereo-seq[20] across several critical brain regions. Our global analysis identified that functional divergence or loss of paralogous genes has significantly driven the evolution of brain cell types. Notably, we identified ~3,000 differentially expressed homologous genes between birds and mammals, particularly the paralogous gene pair *SLC17A6* and *SLC17A7* in cortical excitatory neurons (EXs). These genes exhibit significant expression differences associated with genomic variations between species. Structural analyses of the transmembrane proteins encoded by these genes identified that minor mutations could induce substantial changes in their transmembrane domains. Additionally, genomic evolutionary events were found to significantly influence EX regulatory networks across species. Beyond *SLC17A6* and *SLC17A7*, numerous other homologous genes contribute to the species-specific differences in EX regulation, highlighting the complex molecular mechanisms driving functional divergence in brain cell types. We also identified a distinct Purkinje cell type (SVIL+) in birds, marked by significant differentiation and unique gene expression profiles compared with ALDOC+ and PLCB4+ Purkinje cells in mammals. This cell type displays pronounced differences in gene expression, suggesting a distinct evolutionary trajectory that likely reflects unique evolutionary pressures in birds. These changes in expression patterns may indicate how the avian cerebellum has adapted to new behavioral lifestyles, such as flight. These findings suggest that these cell types are crucial for adaptation to unique ecological niches and the demands of aerial mobility in birds. Our results highlight a correlation between genomic evolution and cell-type differentiation in amniotes, supporting the theory that genomic evolution through natural selection is crucial for the evolutionary differences in brain function. This resource is interactively accessible at https://db.cngb.org/cdcp/absna/.

RESULTS

A cell-type atlas of the amniote brain

To better understand the evolution of cell types in the amniote brain, we constructed a comprehensive single-cell atlas using a droplet-based DNBelab C4 snRNA-seq[19] across species, complemented by several sections of spatial transcriptomic (Stereo-seq) data for validation.[20] This included three generated datasets in this study from the Sauropsida branch: Chinese soft-shell turtle (*Pelodiscus sinensis*), zebra finch (*Taeniopygia guttata*) (hereafter referred to as "finch"), and pigeon (*Columba livia*). Additionally, we incorporated published data from two mammals: mice (*Mus musculus*) and cynomolgus monkeys (*Macaca fascicularis*)[14,21–23] (Figure 1A). Given the complex architecture of avian brains, we conducted snRNA-seq with detailed parcellation in zebra finches and pigeons, encompassing the pallium, subpallium, optic tectum (OT), cerebellum, thalamus, midbrain, and hindbrain (here, "hindbrain" includes the pons and medulla but excludes the cerebellum here). The avian pallium was subdivided into the hyperpallium, mesopallium, arcopallium, nidopallium, and nidopallium caudolaterale (NCL) (Figure 1B; Table S1).

To improve the gene annotation quality in turtles and pigeons, we aligned the raw reads to the reference genomes, integrating a refined gene set annotated via bulk RNA-seq of brain tissue

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

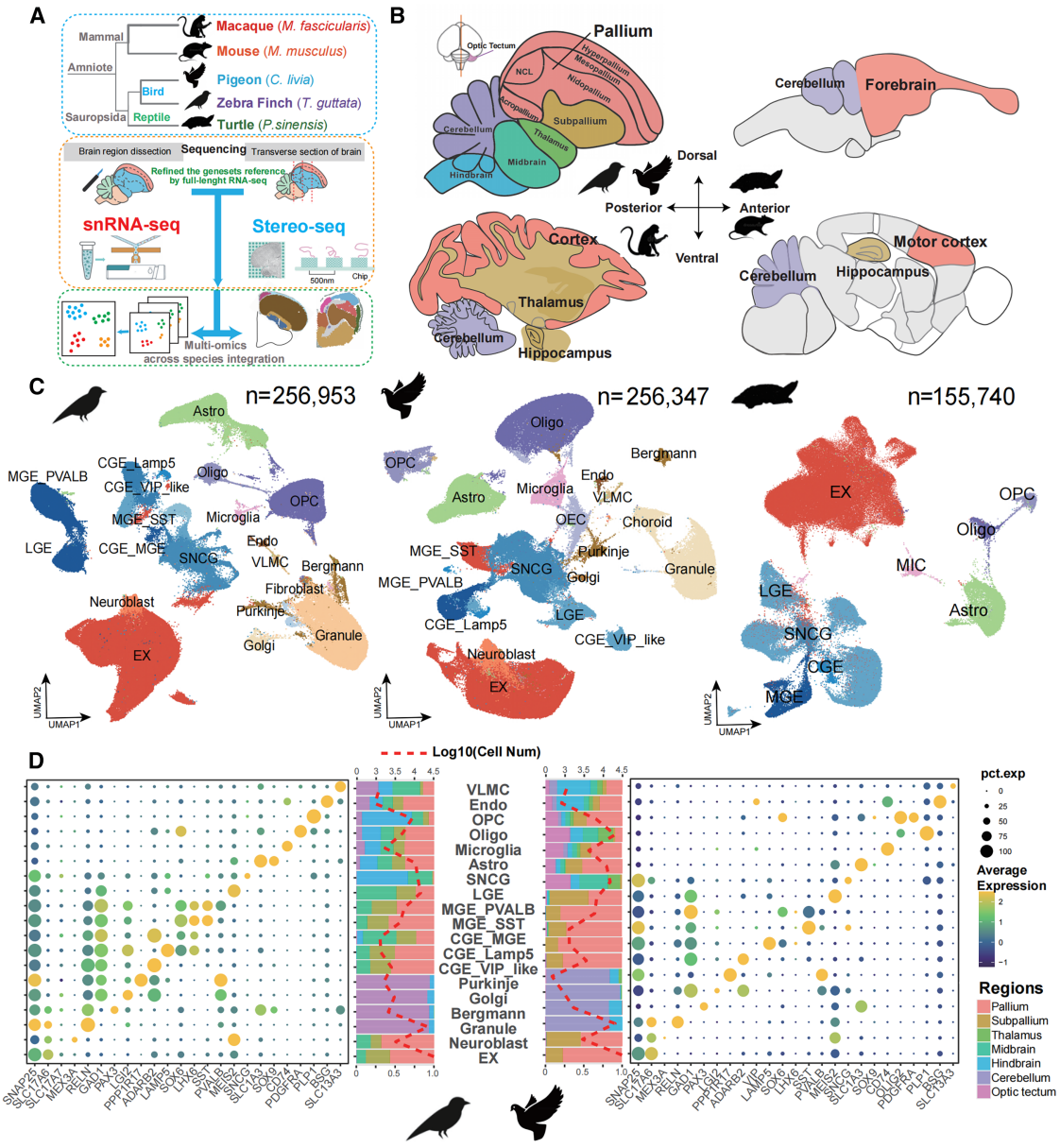


Figure 1. Cell-type conservation and diversity in amniotic brain

(A) Schematic diagram of the complete design analysis model in this paper. Species selection: Sauropsida branch: Chinese soft-shell turtle (Pelodiscus sinensis), zebra finch (Taeniopygia guttata), and pigeon (Columba livia). Mammals: mice (Mus musculus) and cynomolgus monkeys (Macaca fascicularis), multi-omics sequencing, and integration and comparison across species.

(B) Brain regions sampled in detail for snRNA-seq with all five species. For the pigeon and finch, the brain was divided into 11 areas, which were integrated into 7 major areas in subsequent analyses, including pallium, subpallium, optic tectum (OT), cerebellum, thalamus, midbrain, and hindbrain (hindbrain includes the pons and medulla but excludes cerebellum here). The pallium was subdivided into hyperpallium, mesopallium, arcopallium, and nidopallium caudolaterale (NCL).

(C) UMAP of single-cell atlas of the brain derived from pigeons, finches, and turtles, colored by cell type. Some cell-type names were abbreviated: EXs, excitatory neurons; Astro, astrocytes; OPCs, oligodendrocyte precursor cells; Oligo, oligodendrocytes; Endo, endothelial cells; VLMCs, vascular leptomeningeal cells. Since there is no well-defined VIP gene in birds, the cell type homologous to the mammalian CGE_VIP was named CGE_VIP_like.

(D) Dot plot depicting the expression of marker genes across cell types, with the proportion of cell sampling sources represented by variably colored bars. The dashed line indicated the logarithm of the total number of cells for each cell type.

using long-read sequencing such as PacBio[24] and Cyclone (a third-generation sequencing platform from BGI) (see STAR Methods). We increased the total number of annotated genes from 24,856 to 25,892 in turtles and from 15,392 to 19,201 in pigeons and enhanced the completeness of gene structure, updated with a complete gene structure with 97% of genes in turtles and 98% in pigeons in the refined gene sets. Using these refined gene sets as a reference, we significantly improved mapping efficiency, with the proportion of reads aligned to gene regions increasing from approximately 45% to 75% for both

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

species. This enhancement led to an increase in the number of genes detected per cell from approximately 800 to 1,000 and in unique molecular identifiers (UMIs) per cell from about 1,200 to 1,800. After quality filtering, we obtained a total of 1,391,446 cells, of which 669,040 cells were generated in this study for the following species: turtles (155,740), pigeons (256,347), and finches (256,953) (Figure 1C; Table S1).

To annotate the cell types across species, we refined the single-cell graph convolutional neural network model (scGCN),[25] which we termed "entropy-scGCN" (see STAR Methods). By employing the mouse as a reference, we considered several factors for annotating cell types in other species: (1) to include more homologous genes in the amniotes with significant speciation time in the subsequent analysis, we calculated the gene expression of the homologous genes within gene families (Table S2), extending beyond simple one-to-one orthologous relationships. (2) To incorporate entropy theory to evaluate cell-type diversity, we identified 19 main interspecies cell types by applying Louvain clustering followed by the entropy-scGCN approach (Figure 1C). The expression patterns of well-established marker genes aligned with expectations and demonstrated high conservation across species (Figure 1D). (3) Based on the individual clustering by each species and differential gene expression patterns, we subdivided the major cell types into further subtypes (24 in turtle, 45 in pigeon, and 48 in finch) (Figures S1A and S1B).

In conclusion, we have constructed a single-cell atlas of amniote brains using a consistent technology platform and generated datasets for unbiased cross-species comparisons.

Integration and comparison across amniotes

The identification of homologous cell types across species is crucial for understanding cell-type evolution and is influenced by various factors. Traditionally, conserved cell types across species have been defined by canonical marker genes, as demonstrated in model species such as primates and rodents.[18,20,26,27] However, this approach frequently encounters significant limitations, particularly for non-conserved cell types and when addressing a broader range of species. Therefore, it has been proposed that using the historical continuity of gene regulatory networks rather than the expression of individual homologous genes will be more effective.[28] For comparative analysis across multiple species, especially those that diverged long ago, employing comprehensive gene information for identification is particularly crucial. To address this issue, as previously described, we primarily aligned homologous genes across species based on gene family information (see STAR Methods). To minimize biases introduced by varying cell counts, we conducted integrated clustering of single cells using downsampling to ensure equality across five species, totaling 195,162 cells. We conducted scGCN all-against-all species comparisons (Figure S2B) and detected recognized cell marker genes, identifying 25 major cell types (Figure 2A) and 63 cell subtypes (Figures 2D and S2A). To minimize biases from random sampling, we conducted four replicate samplings with an equal proportion of cells, ensuring comprehensive representation of the entire cell population. Both scGCN and integrated analyses consistently identified cross-species cell-type features (Figure S2C). Batch integration of tortoise and zebra finch data with published annotations[9,15] identified strong correspondence with reference cell types, except for cerebellar cell types absent from the reference dataset (Figure S2D). To validate the accuracy of the scGCN model, we performed SAMap[29] to map cell types across the same datasets (see STAR Methods). Correspondence rates from SAMap (Figure S2E) showed strong agreement with scGCN results for closely related species, further supporting scGCN's findings. However, SAMap failed to accurately map cell types for distantly related species, such as between turtles and mice, highlighting the scGCN's advantage in such cases (Figure S2E). Collectively, these benchmarking analyses confirmed the superior accuracy and broader applicability of the scGCN-generated cross-species atlas.

As expected, conservation among major cell types was significantly higher than among subtypes, indicating that interspecies differences in cell types are primarily manifested in the differentiation of cell subtypes (Figures 2A and 2D). Particularly in EXs, where almost all cell subtypes displayed clade- or species-specific characteristics (Figure 2D). To deeply investigate the diversity of cell subtypes across species and identify the emergence of sister cell subtypes, we calculated information entropy values based on the all-against-all species scGCN mapping of each cell subtype (Figure S2B; see STAR Methods). A low entropy score for a specific cell type indicated that it was conserved and could be uniquely mapped to a corresponding cell type in other species via scGCN, signifying evolutionary conservation across species. Conversely, a high entropy score suggested a complex many-to-many mapping relationship across species and predicted the emergence of sister sub-cell types (Figure S2B; Table S4). We observed that EX groups with low entropy and conserved types, such as 5_EX and 11_EX, exhibited relatively uniform proportions across species. By contrast, EX groups with high entropy that varied across species either showed a biased species proportion (22_Granule and 54_EX) or divergent gene expression patterns (e.g., 24_EX expressed *SLC17A6* and *SLC17A7* in birds and mammals, respectively) (Figures 2B and 2D). Most non-neuronal types displayed conserved low entropy, with exceptions including premature oligodendrocytes (oligodendrocyte precursor cells [OPCs] and olfactory ensheathing cells [OECs]) and cerebellar Bergmann cells (Figure 2B). To our surprise, most inhibitory neuron (IN) groups followed a high entropy pattern, although these cell types exhibited no significant biased proportion across species (Figures 2B and 2D). Consequently, we proposed that IN cells exhibit diverse gene expression patterns and might be subdivided into more species-specific subtypes, a hypothesis we confirmed in subsequent analyses (section regarding the IN groups).

Furthermore, based on the distribution of cell-type proportions, we identified several clade-enriched cell subtypes, specifically in birds or mammals (Figure 2D). Most EXs, including three bird-enriched types (30_EX, 35_EX, and 43_EX) and all mammal-enriched EX types (36_EX, 39_EX, and 42_EX), along with certain inhibitory and non-neuronal cells in the cerebellum (38_Purkinje, 40_Purkinje, 62_Purkinje, 41_Bergmann, and 46_Bergmann), exhibited significant clade enrichment. Notably, by analyzing species-specific gene expression, we identified two distinct groups of EX cell types: avian-specific and mammal-specific. Avian EX cells predominantly express *SLC17A6*, while mammalian EX cells primarily express *SLC17A7*, with a minority of

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

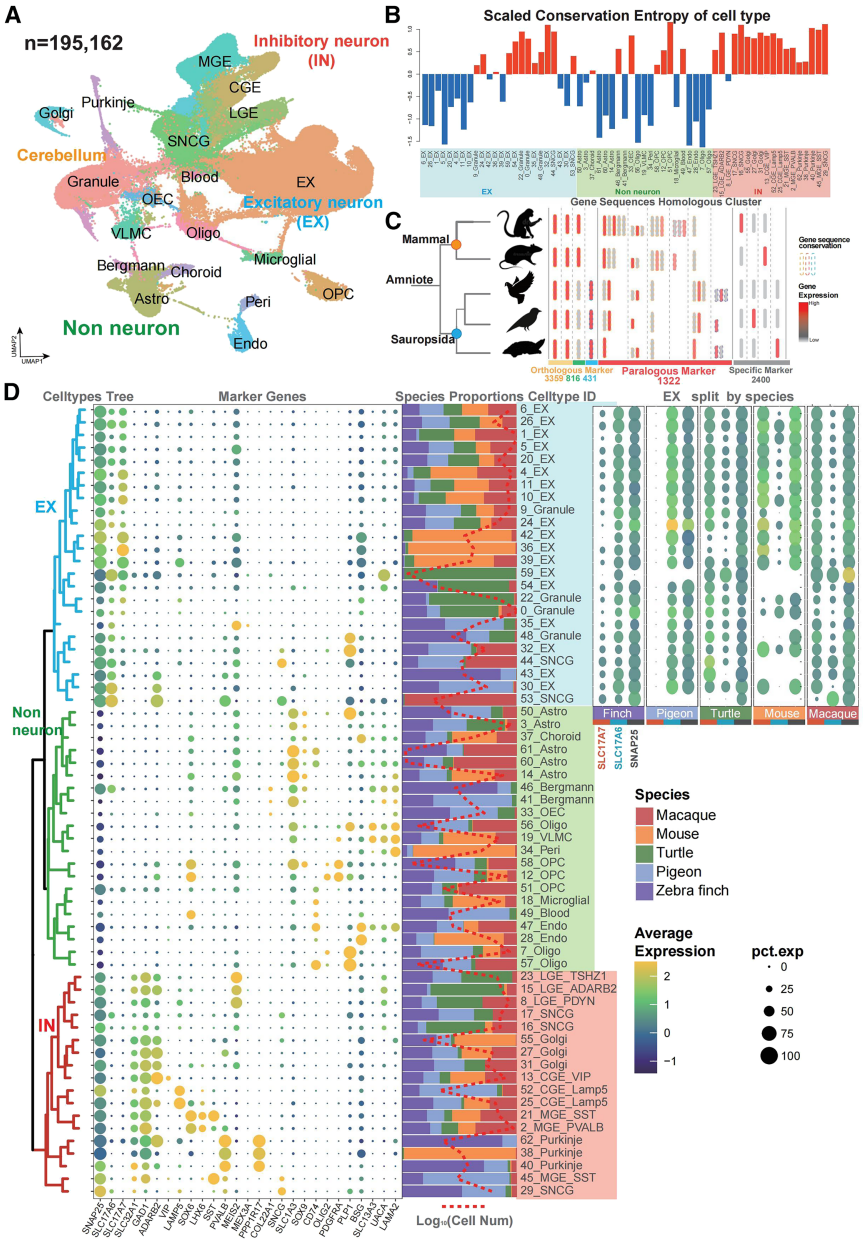


Figure 2. Integration and comparison of the single-cell atlas across amniotes

(A) Integrated cross-species UMAP atlases of the whole brain, downsampling to ~40,000 cells per species (Table S1), covering five species. On the basis of Figure 1D, 3 additional cell types were finely differentiated: pericytes (Peri), olfactory ensheathing cells (OECs), and choroid cells.

(B) Scaled conservation entropy values for 63 cell-type clusters across species, determined using the scGCN algorithm. Sub-cell type with larger values indicate significant differences.

(C) Based on the characteristics of cell-type marker gene families, gene expression was classified into three categories: one-to-one conservation across species, homologous gene family conservation, and species-specific. The schematic plot displays the classification and statistics. The number of genes for each type is indicated below the corresponding color.

(D) Overview of cell taxonomy, marker gene expression, and cell-type abundance across species. Left: cell taxonomy of 63 cell clusters, framed by color according to cell-type categories including excitatory neurons (EXs, blue), inhibitory neurons (INs, red), and non-neuronal cells (non-neuron, green). Middle: dot plots illustrate the expression of marker genes used for annotating cell clusters shown in (A), where dot size and color represent the proportion of cells expressing each gene and the mean expression level within each cluster, respectively. Right: percentage bar plots showing the proportion of five species in each cell type. The red dashed line represents the logarithm of the cell number. Top right: EX cell types were split into five species sources. Dot plot shows the expression of *SLC17A7*, *SLC17A6*, and *SNAP25* genes.

subtypes also expressing *SLC17A6*, and turtle EX cells expressing both genes (Figure 2D).

Within the integrated dataset, we identified sub-cell types exhibiting unique expression patterns specific to certain clades or species. Therefore, we decided to explore how genes contribute to this evolutionary progress. Under consistent conditions, we quantified differentially expressed genes (DEGs) and differentially expressed gene families (DEGFs) across cell types in each species (see STAR Methods). We found that only 19% of DEGs were conserved across evolutionary branches, while the majority were species-specific. By contrast, 33% of DEGFs were conserved across evolutionary clades, with the largest proportion shared among all five species. Those findings suggest that the expression of gene families is more conserved across species than that of one-to-one homologous genes (Figure S2F). To investigate the gain, loss, expansion, or contraction of genes or gene families contributing to cell-type divergence across species, we first categorized the gene families of five species into types that were conserved and distinct for each branch (Figure 2C). Finally, we compared the expansion or contraction of gene families with marker genes across clades (Figures 2C and S3B; Table S3). Using this method, we categorized the marker genes into three categories (see STAR Methods): orthologous markers (55% of all marker genes, with 40% conserved across all amniote species), paralogous markers (16%), and species-specific markers (29%) (Figure 2C).

To investigate the evolutionary dynamics of gene categories, we categorized genes into transcription factors (TFs) and effect genes and conducted differential expression analysis for these groups. Additionally, we compiled a list of 688 positively selected genes identified through genome-wide scans, including 144 avian-specific genes[30] (Table S6), and identified 4,667 TFs from the database[30] (Table S5). Our analysis identified that TFs are more conserved across species, while positively selected genes exhibit greater expression variability (Figure S3C).

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

Conserved orthologous genes predominantly drive uniform gene expression patterns, whereas species-specific expression frequently involves paralogous genes. For instance, *SLC17A6* and *SLC17A7* are specifically expressed in birds and mammals, respectively (Figure 2D). Using Tree2gd,[31] we identified gene duplication events across various evolutionary nodes of amniotes (Figure S3A). By integrating these events with SCENIC-derived[32] regulatory networks, we found that EX networks were most affected by gene duplications, with 20%–35% of genes in differential networks showing duplication, approximately 90% of which originated at ancestral nodes of amniotes or vertebrates (Figures S3D and S3E). These findings demonstrate that TF regulation and gene duplication events have been pivotal drivers of cell-type diversification, particularly in EXs, balancing conserved and species-specific expression patterns.

EXs in the cortex of amniotes

EXs exhibit increased cellular heterogeneity brain region specificity and are more prevalent within specific clades in terms of sub-cell types, as previously described (Figure 2D). For systematic characterization and comparison, we isolated EX neurons from the pallium regions, specifically the hyperpallium in birds, cortex in macaques, primary motor cortex in mice, and forebrain in turtles, for comparative analysis (Figures 3A–3D and S4A). We observed that similar cell types cluster together rather than by species, indicating an elimination of batch effects. However, when examining specific subtypes, we observed in the UMAP results that cells from phylogenetically closer species tend to cluster together (Figure 3B). Notably, several avian cells cluster together with the upper layers (L2-3) and deeper layers (L5-6) of the cortex of both macaques and mice (Figures 3C and 3D), suggesting that birds, like mammals, also harbor EX cells specialized for short- and long-range connections, respectively. Moreover, a significant spatial pattern was observed for avian EX cells, distinct from the laminar structure typically found in mammalian neocortex (Figures S4B and S4C). Furthermore, L4 cells in mammals did not have a significant counterpart in birds, indicating that L4 EX cells may be specific to mammals (Figures 3C and 3D).

SLC17A6+ and SLC17A7+ have been identified predominantly in avian-enriched and mammal-enriched EX neurons, respectively (Figure 2D). To delineate these two distinct clade-enriched groups of EX cell types, we clustered all 39 EX cell subtypes using canonical marker genes (e.g., *SLC17A6* and *SLC17A7*), layer-specific marker genes (e.g., *CUX2*, *RORB*, and *CDH9*), and other species-specific genes (Figure 3E). As expected, we identified two distinct groups of cell types: mammal-specific and sauropsid-specific, characterized by *SLC17A7* and *SLC17A6* expression, respectively. Notably, we identified several genes with expression patterns similar to those of the two canonical genes. Five genes, namely *TMSB4X, HS3ST4, MGAT4C, FTH1*, and *DCC*, were highly expressed in SLC17A7+ EX neurons, associated with functions of axon growth (*DCC*),[34] neural circuit formation, and synaptic plasticity (*TMSB4X*). Additionally, cortex layer marker genes and five other genes (namely, *NT5DC2, PHRF1, LIPT1, MFSD4B*, and *PSMA3*), associated with functions of neuron energy metabolism (such as *LIPT1*) and neurotransmitter transport (such as *MFSD4B*),[35,36] were found to be highly expressed in SLC17A6+ cells. Cluster 24 (Figure 3E), part of the SLC17A6+ cell group and primarily located in the finch brain, showed high expression of deep-layer markers (L5-6) such as *TSHZ2* (related to neural circuitry development),[21] *CDH9* (associated with neuronal connectivity and plasticity),[37] and *RSPO3* (linked to synaptic plasticity and notably concentrated in the finch hyperpallium)[38] (Figures 3I and 3J). Our findings suggested an evolutionary adaptation toward long-range synaptic connections and enhanced neural communication and plasticity.[37,39] Interestingly, cluster 31, also within the SLC17A6+ group, unexpectedly expressed *SLC17A7* in turtles but not in finches and showed the expression of the deep-layer marker *CDH9*, similar to cluster 24 (Figure 3E). This marker is notably enriched in the hyperpallium of birds (Figures 3I and 3J), suggesting that sauropsids may retain some EX cells expressing *SLC17A7*, a trait typically specific to mammalian cortex. These findings suggested that EX neurons in sauropsids exhibit significant divergence and retain functions analogous to those of deep-layer, long-range connection neurons in mammals.[40]

To elucidate the evolutionary mechanisms underlying these two cell-type groups, we performed a comparative analysis on protein sequences and the protein domain folding. *SLC17A6* and *SLC17A7* encode the proteins VGLUT2 and VGLUT1, respectively, both involved in the transport of the neurotransmitter glutamate into synaptic vesicles.[41–43] Utilizing the UniProt database[44] to confirm protein structures and integrating the gene family dataset, we reconstructed the evolutionary trajectory of these proteins from zebrafish to birds and mammals (Figure 3G). *SLC17A6* underwent a duplication event in the ancestral lineage of amniotes, resulting in the paralog *SLC17A7*. Motif sequence analysis of these two genes across amniotic species (Figure 3F) identified that *SCL17A6* motifs, conserved across amniotes, experienced positive selection pressure in the bird clade (Ka/Ks = 4.14 compared with mice). *SLC17A7* either lost the entire gene or specific motifs in birds and turtles, consistent with the observed lack of gene expression in pigeons and limited partial expression in finches.

Upon comparing the motifs of these two genes, we observed that in mammals, *SLC17A7* harbors a duplication of motif-14 positioned closer to the 3′ end compared with *SLC17A6*, whereas in finches, this motif was absent (Figure 3F). Multiple alignment of motif-14 identified a conserved gene structure in mammals and divergence in sauropsids (Figure 3H). The *SLC17A6* and *SLC17A7* genes encode the VGLUT2 and VGLUT1 proteins, and those protein structures (Figures 3G and S4D) elucidated the associations between structural mutation sites and the protein's transmembrane domain.

Overall, significant differences in gene expression exist between avian and mammalian cortical EXs. These disparities could be attributed to gene duplication, site mutations, and other genomic evolutionary events, suggesting that comparative genomics analysis could effectively elucidate single-cell data to highlight differences between cell types.

Evolution of distinct cerebellar cell types and expression characteristics in birds

The cerebellum, thought to have originated early in vertebrate evolution, is present across a broad spectrum of vertebrates, including fish, reptiles, and mammals.[45–48] However, the

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

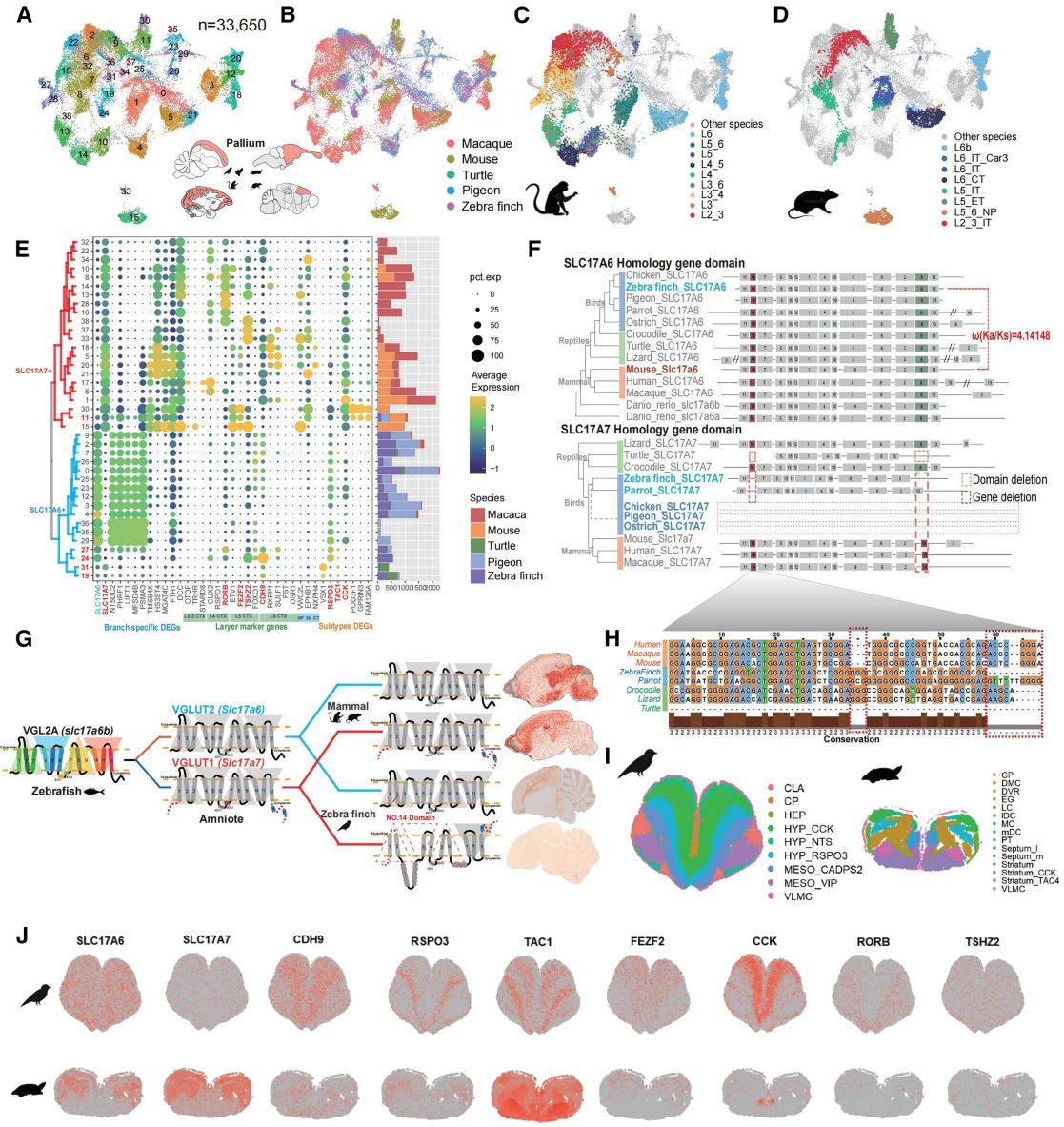


Figure 3. Evolutionary differences in EXs in the cortex and subcortex of amniotes

(A–D) UMAP plots of re-clustered EXs from the cortical areas of five species, with an integrated UMAP for all species shown in (A), color by species in (B), and EX neurons across cortical layers in macaques (C) and mice (D).

(E) Expression of selected marker genes and the clustering relationships among all subtypes of EX neurons. Automatically clustered into two groups: SLC17A6+ (lower group in blue font) and SLC17A7+ (upper group in red font). Marker genes were selected based on clade- or species-specific expressions (blue), layer-specific marker genes in the neocortex (green), and DEGs of the two major groups (orange).

(F) Gene tree illustrating the motifs of the *SLC17A6* (upper panel) and *SLC17A7* (lower panel) gene families in amniotes, detected using Multiple Em for Motif Elicitation (MEME).[33] The location where the domain mutation occurred is marked with a red dashed line. Species without corresponding genes of the same name are marked by gray dashed lines. Mouse and finch *SLC17A6* genes were detected by Tree2gd[31] with Ka/Ks significantly greater than 1, marked on the side.

(G) Schematic diagram showing the structural changes in VGLUT2 (encoded by *SLC17A6*) and VGLUT1 (encoded by *SLC17A7*) proteins throughout the evolution of zebrafish, predicted in the common ancestor of amniotes, mammals, and birds. The right panel depicts the spatial distribution of gene expression (sagittal) in a brain section.

(H) Difference sequence alignment diagram for the no.14 domain in (F) and the corresponding protein variations between species. Amino acid sites showing significant differences between birds and mammals are highlighted in red dashed boxes.

(I) Brain domains of two sections from finches and turtles analyzed using the spatial transcriptomic (Stereo-seq)[20] atlas (see data and code availability).

(J) Spatial patterns of gene expression for several marker genes (E) in two sections from (I).

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

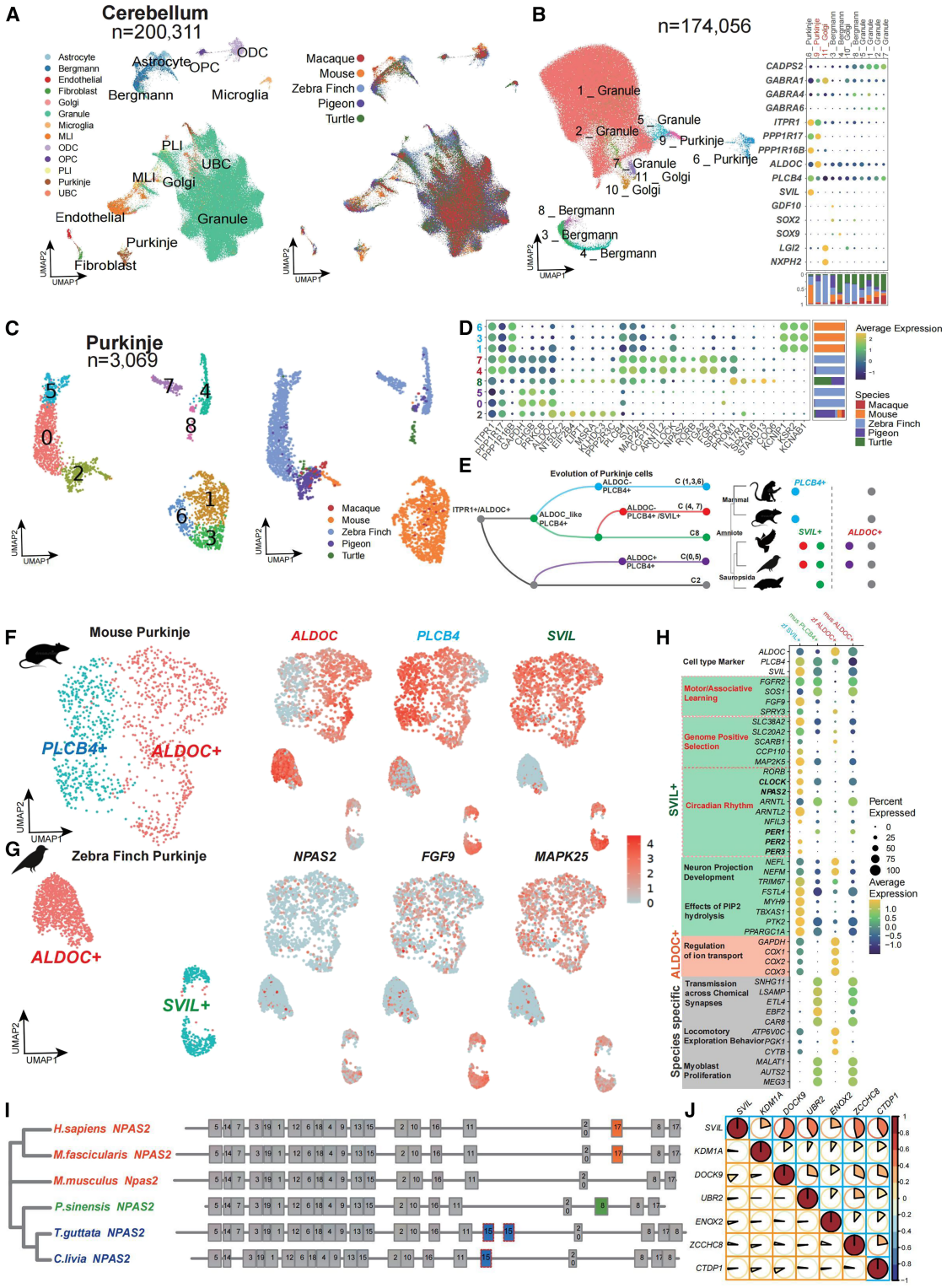


Figure 4. Cell-type diversity in amniote cerebellum

(A) Clustered UMAP plot showing the cell annotation results of the entire cerebellum cells from five species. The species origin distribution is shown on the right.

(B) UMAP of the clustering of cerebellar-specific cell types (granule cells, Bergmann cells, Purkinje cells, and Golgi cells) in five species. Dot plots of marker gene expression for each cluster are shown on the right. The proportion of species is indicated in the bar plot with the corresponding color.

(C) Clustered UMAP plot showing Purkinje cell classification for five species. Subtype numbers are shown on the left and species origin on the right.

(legend continued on next page)

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

complexity and structure of the cerebellum have evolved variably across species.[49,50] To examine the specific cell types present in sauropsids and mammals, we isolated 200,311 cells from their cerebellum, identifying a total of thirteen distinct cell types (Figures S5A and S5C). As expected, the granule cells constituted the majority of the cell population (Figures 4A and S5B). To further analyze cerebellum-specific cell types, we isolated an additional 174,056 cells, comprising granule, Bergmann, Golgi, and Purkinje cells, distinguished by their unique characteristics in the cerebellum compared with other brain regions (Figure 4B). Compared with other brain regions, cerebellar cell types, particularly Bergmann, Golgi, and Purkinje cells, showed greater interspecies variability (Figures 2D and 4B), suggesting a possible functional adaptation within the cerebellum of amniotes to their environment. Specifically, *SOX2* and *SOX9* were predominantly expressed in mammalian and avian Bergmann cells, respectively (Figure S5D). Marker genes of granule cells, such as *GABRA4* and *GABRA6*, demonstrated differential expression between birds and mammals (Figure S5E). Furthermore, genes including *PCLO*, *LGI2*, *NXPH2*, and other genes are differentially expressed in cerebellum-specific cells between different species (Figures S5D and S5F), which were involved in the formation and maintenance of synapses and the transmission of neural signals.[51–53]

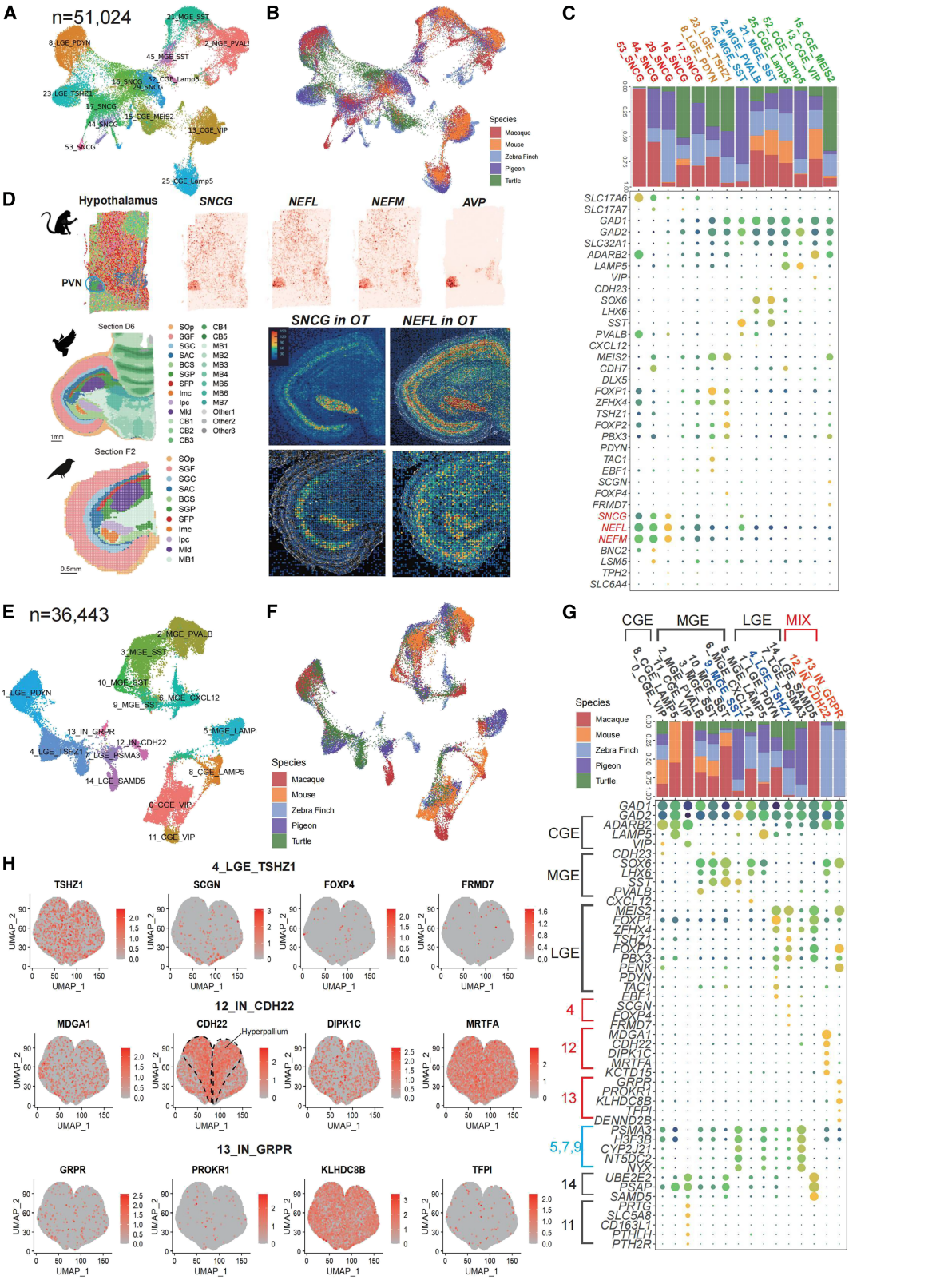
Notably, Purkinje cells, which exhibited the most pronounced species differences (Figure 4B), comprised two distinct subtypes: 6_Purkinje and 9_Purkinje, each characterized by the expression of specific markers, including *ITPR1, PPP1R17, ALDOC, PLCB4*, and *SVIL* (Figure 4B). We re-clustered the 3,069 Purkinje cells and identified nine cell subtypes, which were subsequently categorized into five groups based on species composition and UMAP distributions: C2 (all species), C8 (sauropsids), C(0,5) (bird-specific), C(4,7) (finch-specific), and C (1,3,6) (mammal-specific) (Figures 4C and 4D). We reconstructed the evolutionary history of Purkinje cell types across amniotes based on marker expression patterns (Figure 4E), identifying C2 as the common cell type among amniotes and C(1,3,6) as the mammal-specific subtype (marked by *KCNIP1*, *KCNAB1*, and *KSR2* expression). Notably, C(0,5) (marked by *ALDOC* expression) and C(4,7) (marked by *SVIL* expression) were bird-specific but exhibited distinct expression patterns. As reported by Chen et al., two major Purkinje subtypes were identified in mice: one predominantly expressed *ALDOC* and the other *PLCB4*, findings that our study corroborates.[54] By contrast, birds also possessed ALDOC+ Purkinje cells and also harbored a bird-specific subtype, SVIL+ Purkinje. Subsequent DEG analysis of these Purkinje subtypes identified that ALDOC+ cells from both finch and mouse share conserved marker genes (*ALDOC*, *HSPA8*, *RTN4, THY1*, and *ATP1B1*) (Figures 4F, 4G, and 4H) with association of regulation of ion transport, suggesting that ALDOC+ represents an ancestral Purkinje subtype in early cerebellar evolution.

To deeply investigate the SVIL+ Purkinje cells in birds, we focused on the upregulated genes and their changing regulatory networks. SVIL+ cells were characterized by the expression of genes associated with circadian rhythms in the KEGG database, including *RORB*, *CLOCK*, *NPAS2*, *PER2/3*, and *ARNTL* (Figures 4H and S5G; Table S7). Previous studies in primates have found that PER1 and PER2 immunoreactive (IR) cells were observed in the Purkinje cells of the cerebellum, suggesting a possible relationship with the rhythmic processing in the cerebellum.[55] In birds, PER2/3 is specifically highly expressed in SVIL+ cells, while PER1 is lowly expressed (Figure 4H), which not only validates the association between Purkinje cells and rhythm regulation but also highlights the selective expression of homologous genes across different species. Comparison with positively selected genes in birds identified overlaps, including *NPAS2, SLC38A2, SLC20A2, CCP110*, and *MAP2K5* (Figure 4H). These results suggest that the specialization of SVIL+ cell types may be related to gene sequence evolution. We identified mutations in protein sequences, including motif domain gains or losses, in bird SVIL+ cells. For example, motif alignment for the marker gene *NPAS2* across amniotes identified a motif-15 gain in birds, a motif-8 duplication in turtles, and a motif-17 duplication in mammals (Figure 4I). In the functional enrichment of DEGs, the SVIL+ cells also exhibited enriched functions related to neuron projection development and effects of PIP2 hydrolysis (Figures 4H and S5I). We sought to elucidate the specificity of SVIL+ Purkinje cells in their regulatory mechanisms. Previous studies suggest that SVIL interacts with lysine-specific demethylase 1 (LSD1), modulating the differential expression of downstream genes.[56] To assess whether this regulatory mechanism operates in Purkinje cells, we analyzed the expression patterns of both upregulated and unchanged downstream genes anticipated to be influenced by LSD1 (encoded by *KDM1A*) activity (Figure 4J). The results indicate that in birds, LSD1 and several downstream genes, including *DOCK9*, *UBR2*, and *ENOX2*, are specifically upregulated in SVIL+ cells. This pattern mirrors the neuronal differentiation regulatory role of LSD1.[56] By contrast, none of these downstream genes are co-expressed in Purkinje cells in mice, suggesting that the generation of SVIL+ cells in birds may involve a comparable regulatory mechanism (Figure 4J). Furthermore, the analysis of SCENIC (Single-Cell Regulatory Network Inference and Clustering) coregulatory networks for two Purkinje cell subtypes across species identified differences in protein regulatory networks between zebra finch and mouse Purkinje cells (Figure S5H; see STAR Methods).

Specifically, our data indicate substantial functional evolution within the cerebellum of amniotes, likely driven by diverse environmental and lifestyle adaptations. ALDOC+ Purkinje cells are

(D) Dot plot of conserved and differential gene expression in each Purkinje sub-cell type (C). Species origin proportions are shown on the right.
(E) Diagram of the inferred evolutionary pattern of Purkinje cells in amniotes, based on the conservation of cell subtypes. Right: distribution of the corresponding isoforms in each species, with color corresponding to (D).
(F and G) Expression profiles of marker genes that show differential expression in the two subtypes of Purkinje cells between the mouse and finch.
(H) Dot plot illustrating DEGs between the two subtypes of Purkinje cells in both species. Genes are categorized based on their function.
(I) The circadian gene NPAS2, which is highly expressed in the SVIL+ group of zebra finches, domains the divergence tree obtained by MEME[33] between species.
(J) Plot of the expression association between the SVIL gene and the downstream regulatory genes of LSD1 (KDM1A) in mouse (orange) and zebra finch (blue). The size and color of the pie chart uniformly indicate the co-expression correlation of a pair of genes in the corresponding species, and a higher value indicates that two genes tend to be highly expressed together.

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014



(legend on next page)

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

hypothesized to have emerged before the common ancestor of amniotes, with subsequent divergence into distinct subtypes in avian and mammalian lineages. In mammals, PLCB4+ Purkinje subtypes developed, while in birds, SVIL+ Purkinje subtypes evolved. The observed differences in regulatory networks among avian Purkinje sub-cell types may reflect specific survival strategies and ecological niches. These results lay the groundwork for further experimental validation and deeper exploration of the evolutionary mechanisms driving these differences.

Conservation and divergence of non-EXs in birds

In both the individual species and cross-species integration atlas, we found a population of non-EXs (Figures 1C and 2A). These include well-defined definitions derived from medial, lateral, and caudal ganglionic eminences (MGEs, LGEs, and CGEs, respectively), INs, and a cluster of cells with high specific expression of *SNCG*. To investigate evolutionary cell-type patterns, we isolated 51,024 non-excitatory cells. Integrated analysis identified consistent species conservation and co-clustering across species at the level of major cell types (Figures 5A and 5B). However, the SNCG+ cell groups, upregulated with *SNCG*, *NEFL*, and *NEFM*, showed the greatest divergence in both the pattern of marker gene expression and the proportion of species represented (Figure 5C). SNCG+ cell groups were mainly sampled from the hypothalamus of each species and the OT of pigeons. We analyzed the spatial transcriptomic data from macaques,[57] zebra finches, and pigeons[58] to validate the derived brain regions (Figure 5D). According to the results of spatial transcriptomics, the gene expression of the SNCG+ cluster (*NEFL* and *NEFM*) in the hypothalamus region was consistent with the characteristics of arginine vasopressin (AVP) neurons distributed in the paraventricular nucleus (PVN) (Figure 5D). The SNCG+ cell types in bird OT were only distributed in the stratum griseum centrale (SGC) and pars parvocellularis (Ipc) regions (Figure 5D). These two nucleus isthmi have been verified in finches to be able to connect with retinorecipient structures by interneuron,[59] and the correlation between their features is worthy of further attention.

GABAergic neurons (inhibitory cells, IN), which are more conserved across species such as mice, macaques, and humans than glutamatergic neurons, exhibit less regional brain specificity.[23,60] This conservation is likely due to functional stability. It remains to be determined if this trait persists across a broader range of species. We analyzed 36,443 IN cells and identified 15 subtypes of IN cells (Figures 5E and 5F), originating developmentally from the CGE (c0, c8, and c11), MGE (c2, c3, c5, c6, c9, and c10), and LGE (c1, c4, c7, and c14). Cell types of MGE origin were the most abundant among amniotes, encompassing six subtypes and comprising 45% of the cells. Furthermore, MGE-derived cells exhibited the highest conservation of proportion across species. CGE-derived IN cells included two mammal-specific subtypes, 8_CGE_LAMP5 and 11_CGE_VIP, whereas LGE-derived IN cells, such as 14_LGE_SAMD5, were also specific to mammals. All three mammal-specific IN subtypes were characterized by upregulated genes, including *UBE2E2*, *PSAP*, and *SAMD5*, associated with cellular lipid metabolism and basal processes.

Notably, we identified two cell subtypes with complex origins, 12_IN_CDH22 and 13_IN_GRPR, which were particularly prevalent in sauropsids and abundant in birds (Figures 5E and 5G). Subtype 12_IN_CDH22, characterized by the expression of *MDGA1*, *CDH22*, and *KCTD15*, was associated with genes potentially involved in synapse formation and neuronal communication. *CDH22* was predominantly expressed in the hyperpallium of finches (Figure 5H), a telencephalic region considered homologous to the neocortex in mammals.[61] However, 13_IN_GRPR exhibited a widespread spatial distribution in the finch brain region (Figure 5H), expressing *GRPR*, a neuropeptide-related gene, and notably upregulating *FOXP2*, associated with language or speech in humans, and *PENK*, crucial for neural communication.[62–65]

DISCUSSION

We explored cell-type diversification in the amniote brain, focusing on differences between sauropsids (birds and reptiles) and mammals (primates and rodents) species via single-cell transcriptome analyses. Our study began with a comprehensive single-cell atlas, incorporating approximately 1.3 million cells from the telencephalon and cerebellum of turtles, zebra finches, pigeons, mice, and macaques (Figure 1A; Table S1). This investigation identified species-specific variations across a broad range of cell types and subtypes, closely tied to the conservation and diversification of genomes across species. Notably, we observed significant diversification in telencephalic EXs and various cerebellar cell types between birds and mammals. For instance, most telencephalic EXs in birds express *SLC17A6*, while mammals exhibit greater regional heterogeneity with *SLC17A7* predominantly in the neocortex and *SLC17A6* in other areas (Figures 2D and 3E). Furthermore, we identified a subcellular type of Purkinje cell (SVIL+) in birds (Figure 4G), characterized by high expression of genes enriched in learning and circadian rhythm-related pathways (Figures 4H and S5G). These genes also include many positively selected ones, indicating the association between gene evolution and cell-type differentiation (Figures 4H and 4I). Additionally, we utilized single-cell resolution spatial transcriptomics (Stereo-seq) to analyze the spatial distribution of genes and cell types across amniotes, effectively

Figure 5. Non-EX differentiation across species

(A and B) Clustered UMAP plot showing the cell annotation results of the non-excitatory cells containing SNCG from five species. The species origin distribution is shown in (B).

(C) Dot plot representation of marker genes in each cell cluster in (A). Species origin proportions are identified below with bar plots. The three genes specifically co-expressed by SNCG are highlighted in red.

(D) Expression of SNCG signature genes in monkey hypothalamus and bird (optic tectal) OT region.

(E and F) Clustered UMAP plot showing the cell annotation results of the IN cells from five species. The species origin distribution is shown in (F).

(G) Dot plot representation of marker genes in each cell cluster in (E). Species origin proportions are identified below with bar plots.

(H) Gene expression on the spatial transcriptome, showing bird-specific enriched marker genes for C4, C12, and C13 in (G).

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

validating our findings. Collectively, our results establish a robust connection between genetic evolution and environmental adaptation, illustrating how genetic alterations enhance cell-type diversification to meet diverse ecological challenges and evolutionary demands.

More specifically, how have the cell types evolved in relation to the species adaptability? In our observations, we identified common cell-type-specific high-expression marker genes exhibiting homologous gene phenomena across various cell types. Notably, *SLC17A6* was highly expressed across the entire pallium in birds, *SLC17A7* in mammals, and both genes were expressed in reptiles (Figure 2D). In the avian Purkinje cell subtype, there is high expression of numerous genes involved in circadian rhythm and mitogen-activated protein kinase (MAPK) signaling pathways, suggesting that the avian cerebellum may have distinct functions from those in mammals (Figures 4H and S5G). We also observed homologous expression variations of *GABRA4*, *GABRA6*, *SOX1*, *SOX2*, and *SOX9* in granule cells and Bergmann cells (Figures S5D and S5E).

Phylogenetic analysis suggests that different expression patterns may arise from two theoretical frameworks: (1) the Gene Balance Hypothesis[66] posits that sequence mutations following genome duplication lead to some copies retaining key life-stabilizing domains while others lose functionality, resulting in paralogous genes expressed in distinct patterns across species (e.g., *SLC17A7* in birds). (2) The duplication-degeneration-complementation model[67] suggests that homologous genes subfunctionalize, share ancestral functions, and perform these functions through dose compensation, often seen with multiple homologous genes expressing concurrently (e.g., *SLC17A7* and *SLC17A6* in turtles).

In mammals, VGLUT1 is linked to low-probability neurotransmitter release, while VGLUT2 is associated with high-probability release.[68,69] Our results show that these functional differences are due to variations in protein sequence and structure. Natural selection likely drives these mutations. The selective expression of different homologous proteins across species observed in our single-cell data may contribute to the differences in neuronal distribution and function between birds and mammals. For example, studies indicate higher neuronal density in bird brains compared with mammals,[70] but the exact influence of brain structural differences or environmental factors on neurotransmitter release remains unknown. However, our study provides a basis for understanding how evolutionary adaptations may affect these mechanisms.

Similar findings in comparative studies of the hypothalamus indicate that various species regulate cell functions through subfunctionalization and dose-compensation effects of homologous genes.[10] Our study extends these conclusions across amniotes and down to specific gene domains, such as mutation in *SLC17A7* in zebra finch, which may be related to the loss of a key neurotransmitter transmission function and compensatory expression of paralogous *SLC17A6* (Figure 3G). This adaptive expression is observed in various cerebellar cell types as well (Figures S5D and S5E). Furthermore, sequence differentiation results in the retention of DEGs that positively impact survival, highlighting that positively selected genes exhibit the most detailed species differences. Conversely, TFs are more conserved due to their crucial role in influencing downstream gene expression (Figure S2D).

From a comprehensive perspective, gene duplication, mutation, and loss events in genome evolution significantly influence the cross-species evolution of cell types. TF genes demonstrate greater stability due to the universality of their functions.[71] We propose that ancestral key genes underwent duplication events, resulting in multiple homologous genes that may play a more important role in the cell-type evolution. These genes were retained in the genomes of different species during species divergence. They underwent gains and losses of gene motifs and even complete gene deletions, which significantly influenced the differences in the gene regulatory network, ultimately affecting the differentiation and distribution of cell types. While experimental validation of the functions and pathways of the identified genes and regulatory networks is a crucial next step, it is not feasible within the scope of this study. Furthermore, such experimental efforts are challenging and will demand significant time and resources. Nonetheless, this study has made significant progress by thoroughly analyzing genomic and expression data. By conducting phylogenetic reconstructions, domain architecture analysis, and expression variance studies across diverse tissues and species, we uncovered insights and proposed hypotheses on the functions and evolutionary dynamics of these genes. This approach departs from traditional methods, offering fresh perspectives on the underlying genetic mechanisms. Future experimental validation through gene editing in strategic species could substantiate these findings and open more research avenues.

In the end, identifying and recognizing homologous cell types is a pivotal challenge in conducting cross-species single-cell studies to explore the evolution of cell types, structures, and functions. In this study, we employed a gene family homology approach instead of solely relying on one-to-one orthologous gene information, as is typically done. This method enabled us to connect homologous genes across distantly related species on a larger scale, expanding the analysis to thousands of gene families. While this approach mitigates issues related to paralogous relationships, it might overlook functional differences in paralogous genes, a limitation that future research needs to address. Furthermore, we adopted an enhanced method based on Graph Convolutional Networks (GCN), termed entropy-scGCN, which effectively integrates single-cell data from multiple species using canonical marker information, providing a solid basis for identifying homologous cell types. With the increasing application of large language models in life sciences, particularly for cross-species comparisons such as SATURN (Species Alignment Through Unification of Rna and proteiNs)[72] and UCE (Universal Cell Embeddings),[73] there is significant promise for future advancements. As more high-quality single-cell data from various species become available, employing large models in cross-species studies is expected to substantially improve our understanding of homologous cell types and their evolutionary trajectories.

Limitations of the study

While our cross-species atlas provides insights into amniote brain evolution, several limitations should be noted. First, the current sampling of species (n = 5) and brain subregions may not fully represent the diversity of amniote neural cell types, particularly for reptiles and basal mammalian lineages. Second,

single-cell RNA-seq-based paralog expression analysis could be influenced by technical artifacts such as batch effects or incomplete transcriptome annotation across species. Third, although positively selected genes were linked to neuron diversification, functional validation of their regulatory roles in specific cell types (e.g., cerebellar Purkinje cells) remains to be experimentally confirmed.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for the resources and reagents may be directed to and will be fulfilled by the lead contact, Shiping Liu (liushiping@genomics.cn).

Materials availability

All materials used for stereo-seq and snRNA-seq are commercially available.

Data and code availability

The data that support the findings of this study have been deposited into the CNGB Sequence Archive (CNSA)[74] of the China National GeneBank DataBase (CNGBdb)[75] with accession number CNGBdb: CNP0003026 (https://db.cngb.org/search/project/CNP0003026/). Additionally, processed stereo-seq and snRNA-seq data used in this study can be accessed and downloaded via STOmicsDB[76]: https://db.cngb.org/cdcp/absna/. The updated and improved entropy-scGCN based on the scGCN algorithm is available on GitHub (https://github.com/Dee-chen/entropy-scGCN/). All data were publicly available as of the date of publication. Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This paper is from the Mesoscopic Brain Mapping Consortium. The project was supported by the National Key R&D Program of China (2022YFC3400400, 2021YFA0805100, and 2020YFE0205900); the National Science and Technology Innovation 2030 Major Program (STI2030-2021ZD0200100, STI2030-2022ZD0205000, and STI2030-2022ZD0211700); the Natural Science Foundation of Guangdong Province, China (2021B1515120075); the Guangdong Genomics Data Center (2021B1212100001); the Zhejiang Science and Technology Department (no. 2024C03004); and the Hangzhou Leading Innovation Team Project (no. TD2023003). We sincerely thank the China National GeneBank for providing technical support.

AUTHOR CONTRIBUTIONS

S.L., R.K.N., X.X., Z.W., C.X., and D.C. conceived and designed the study. D. C., S.L., R.K.N., and Z. Zhuang wrote the manuscript. F.H., Y.W., K.L., H.W., J. H., Y. Lei, Y.B., C.X., J.W., H.Y., Yidi Sun, G.C., M.P., Yangang Sun, Z. Zhang, L.L., and Y. Lin contributed to the discussion and revision of the manuscript. H. W., Z.W., M.H., and W.X. contributed to sample collection. T.Z. and Y.D. contributed to the sequencing. Y.H., M.C., and J.P. designed and performed snRNA-seq and stereo-seq. D.C., Z. Zhuang, Y.Y., Y.Z., F.H., K.L., X.J., Y. X., Z.N., and Z. Zhu discussed and conducted data analysis. X.W. and T.Y. collected the analyzed data and established the online database. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to improve the language. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animals
- METHOD DETAILS
  - Library preparation and sequencing
  - Brain tissue collection for snRNA-seq
  - Single-nucleus suspension preparation
  - snRNA-seq data processing
  - Gene family clustering and evolutionary event identification
  - Cell downsampling and cross-species integration
  - SAMap cross-species cell type similarity analysis
  - Update turtle and pigeon genome annotations by IsoSeq
  - Cell type differentially expressed gene analysis and species comparison
  - Co-Expression Regulatory Network Analysis Using SCENIC
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Conservative entropy calculation across species cell types

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.devcel.2025.04.014.

Received: June 23, 2024
Revised: March 5, 2025
Accepted: April 18, 2025
Published: May 13, 2025

REFERENCES

1. Woych, J., Ortega Gurrola, A., Deryckere, A., Jaeger, E.C.B., Gumnit, E., Merello, G., Gu, J., Joven Araus, A., Leigh, N.D., Yun, M., et al. (2022). Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. Science *377*, eabp9186. https://doi.org/10.1126/science.abp9186.
2. Belgard, T.G., Montiel, J.F., Wang, W.Z., García-Moreno, F., Margulies, E. H., Ponting, C.P., and Molnár, Z. (2013). Adult pallium transcriptomes surprise in not reflecting predicted homologies across diverse chicken and mouse pallial sectors. Proc. Natl. Acad. Sci. USA *110*, 13150–13155. https://doi.org/10.1073/pnas.1307444110.
3. La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566–580.e19. https://doi.org/10.1016/j.cell.2016.09.027.
4. Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142. https://doi.org/10.1126/science.aaa1934.
5. Ksepka, D.T., Balanoff, A.M., Smith, N.A., Bever, G.S., Bhullar, B.S., Bourdon, E., Braun, E.L., Burleigh, J.G., Clarke, J.A., Colbert, M.W., et al. (2020). Tempo and Pattern of Avian Brain Size Evolution. Curr. Biol. *30*, 2026–2036.e3. https://doi.org/10.1016/j.cub.2020.03.060.
6. Sayol, F., Lapiedra, O., Ducatez, S., and Sol, D. (2019). Larger brains spur species diversification in birds. Evolution *73*, 2085–2093. https://doi.org/10.1111/evo.13811.
7. Herculano-Houzel, S. (2020). Birds do have a brain cortex-and think. Science *369*, 1567–1568. https://doi.org/10.1126/science.abe0536.
8. He, Z., Han, D., Efimova, O., Guijarro, P., Yu, Q., Oleksiak, A., Jiang, S., Anokhin, K., Velichkovsky, B., Grünewald, S., and Khaitovich, P. (2017). Comprehensive transcriptome analysis of neocortical layers in humans,

chimpanzees and macaques. Nat. Neurosci. *20*, 886–895. https://doi.org/10.1038/nn.4548.

9. Tosches, M.A., Yamawaki, T.M., Naumann, R.K., Jacobi, A.A., Tushev, G., and Laurent, G. (2018). Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. Science *360*, 881–888. https://doi.org/10.1126/science.aar4237.
10. Shafer, M.E.R., Sawh, A.N., and Schier, A.F. (2022). Gene family evolution underlies cell-type diversification in the hypothalamus of teleosts. Nat. Ecol. Evol. *6*, 63–76. https://doi.org/10.1038/s41559-021-01580-3.
11. Styfhals, R., Zolotarov, G., Hulselmans, G., Spanier, K.I., Poovathingal, S., Elagoz, A.M., De Winter, S., Deryckere, A., Rajewsky, N., Ponte, G., et al. (2022). Cell type diversity in a developing octopus brain. Nat. Commun. *13*, 7392. https://doi.org/10.1038/s41467-022-35198-1.
12. Hain, D., Gallego-Flores, T., Klinkmann, M., Macias, A., Ciirdaeva, E., Arends, A., Thum, C., Tushev, G., Kretschmer, F., Tosches, M.A., and Laurent, G. (2022). Molecular diversity and evolution of neuron types in the amniote brain. Science *377*, eabp8202. https://doi.org/10.1126/science.abp8202.
13. Gumnit, E., and Tosches, M.A. (2023). A cell type atlas of the lamprey brain. Nat. Ecol. Evol. *7*, 1591–1592. https://doi.org/10.1038/s41559-023-02195-6.
14. Bakken, T.E., Jorstad, N.L., Hu, Q., Lake, B.B., Tian, W., Kalmbach, B.E., Crow, M., Hodge, R.D., Krienen, F.M., Sorensen, S.A., et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature *598*, 111–119. https://doi.org/10.1038/s41586-021-03465-8.
15. Colquitt, B.M., Merullo, D.P., Konopka, G., Roberts, T.F., and Brainard, M. S. (2021). Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. Science *371*, eabd9704. https://doi.org/10.1126/science.abd9704.
16. Franjic, D., Skarica, M., Ma, S., Arellano, J.I., Tebbenkamp, A.T.N., Choi, J., Xu, C., Li, Q., Morozov, Y.M., Andrijevic, D., et al. (2022). Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. Neuron *110*, 452–469.e14. https://doi.org/10.1016/j.neuron.2021.10.036.
17. Ma, S., Skarica, M., Li, Q., Xu, C., Risgaard, R.D., Tebbenkamp, A.T.N., Mato-Blanco, X., Kovner, R., Krsnik, Ž., de Martin, X., et al. (2022). Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. Science *377*, eabo7257. https://doi.org/10.1126/science.abo7257.
18. Wang, J., Sun, H., Jiang, M., Li, J., Zhang, P., Chen, H., Mei, Y., Fei, L., Lai, S., Han, X., et al. (2021). Tracing cell-type evolution by cross-species comparison of cell atlases. Cell Rep. *34*, 108803. https://doi.org/10.1016/j.celrep.2021.108803.
19. Liu, C., Wu, T., Fan, F., Liu, Y., Wu, L., Junkin, M., Wang, Z., Yu, Y., Wang, W., Wei, W., et al. (2019). A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling. Preprint at bioRxiv. https://doi.org/10.1101/818450.
20. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.
21. Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222–3241.e26. https://doi.org/10.1016/j.cell.2021.04.021.
22. Kozareva, V., Martin, C., Osorno, T., Rudolph, S., Guo, C., Vanderburg, C., Nadaf, N., Regev, A., Regehr, W.G., and Macosko, E. (2021). A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature *598*, 214–219. https://doi.org/10.1038/s41586-021-03220-z.
23. Chen, A., Sun, Y., Lei, Y., Li, C., Liao, S., Meng, J., Bai, Y., Liu, Z., Liang, Z., Zhu, Z., et al. (2023). Single-cell spatial transcriptome reveals cell-type organization in the macaque cortex. Cell *186*, 3726–3743.e24. https://doi.org/10.1016/j.cell.2023.06.009.
24. Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics *13*, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002.
25. Song, Q., Su, J., and Zhang, W. (2021). scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. Nat. Commun. *12*, 3826. https://doi.org/10.1038/s41467-021-24172-y.
26. Han, L., Wei, X., Liu, C., Volpe, G., Zhuang, Z., Zou, X., Wang, Z., Pan, T., Yuan, Y., Zhang, X., et al. (2022). Cell transcriptomic atlas of the non-human primate Macaca fascicularis. Nature *604*, 723–731. https://doi.org/10.1038/s41586-022-04587-3.
27. Qu, J., Yang, F., Zhu, T., Wang, Y., Fang, W., Ding, Y., Zhao, X., Qi, X., Xie, Q., Chen, M., et al. (2022). A reference single-cell regulomic and transcriptomic map of cynomolgus monkeys. Nat. Commun. *13*, 4069. https://doi.org/10.1038/s41467-022-31770-x.
28. Wagner, G.P. (2007). The developmental genetics of homology. Nat. Rev. Genet. *8*, 473–479. https://doi.org/10.1038/nrg2099.
29. Tarashansky, A.J., Musser, J.M., Khariton, M., Li, P., Arendt, D., Quake, S. R., and Wang, B. (2021). Mapping single-cell atlases throughout Metazoa unravels cell type evolution. Elife *10*, e66747. https://doi.org/10.7554/eLife.66747.
30. Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., et al. (2010). The genome of a songbird. Nature *464*, 757–762. https://doi.org/10.1038/nature08819.
31. Chen, D., Zhang, T., Chen, Y., Ma, H., and Qi, J. (2022). Tree2GD: a phylogenomic method to detect large-scale gene duplication events. Bioinformatics *38*, 5317–5321. https://doi.org/10.1093/bioinformatics/btac669.
32. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
33. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43*, W39–W49. https://doi.org/10.1093/nar/gkv416.
34. Duman-Scheel, M. (2009). Netrin and DCC: axon guidance regulators at the intersection of nervous system development and cancer. Curr. Drug Targets *10*, 602–610. https://doi.org/10.2174/138945009788680428.
35. Soreze, Y., Boutron, A., Habarou, F., Barnerias, C., Nonnenmacher, L., Delpech, H., Mamoune, A., Chrétien, D., Hubert, L., Bole-Feysot, C., et al. (2013). Mutations in human lipoyltransferase gene LIPT1 cause a Leigh disease with secondary deficiency for pyruvate and alpha-ketoglutarate dehydrogenase. Orphanet J. Rare Dis. *8*, 192. https://doi.org/10.1186/1750-1172-8-192.
36. Perland, E., Hellsten, S.V., Schweizer, N., Arapi, V., Rezayee, F., Bushra, M., and Fredriksson, R. (2017). Structural prediction of two novel human atypical SLC transporters, MFSD4A and MFSD9, and their neuroanatomical distribution in mice. PLoS One *12*, e0186325. https://doi.org/10.1371/journal.pone.0186325.
37. de Wit, J., and Ghosh, A. (2016). Specification of synaptic connectivity by cell surface interactions. Nat. Rev. Neurosci. *17*, 22–35. https://doi.org/10.1038/nrn.2015.3.
38. Lovell, P.V., Clayton, D.F., Replogle, K.L., and Mello, C.V. (2008). Birdsong "transcriptomics": neurochemical specializations of the oscine song system. PLoS One *3*, e3440. https://doi.org/10.1371/journal.pone.0003440.
39. Polanco, J., Reyes-Vigil, F., Weisberg, S.D., Dhimitruka, I., and Brusés, J. L. (2021). Differential Spatiotemporal Expression of Type I and Type II Cadherins Associated With the Segmentation of the Central Nervous System and Formation of Brain Nuclei in the Developing Mouse. Front. Mol. Neurosci. *14*, 633719. https://doi.org/10.3389/fnmol.2021.633719.
40. Hussan, M.T., Sakai, A., and Matsui, H. (2022). Glutamatergic pathways in the brains of turtles: A comparative perspective among reptiles, birds, and mammals. Front. Neuroanat. *16*, 937504. https://doi.org/10.3389/fnana.2022.937504.

41. Park, S.K., Hong, J.H., Jung, J.K., Ko, H.G., and Bae, Y.C. (2019). Vesicular Glutamate Transporter 1 (VGLUT1)- and VGLUT2-containing Terminals on the Rat Jaw-closing gamma-Motoneurons. Exp. Neurobiol. *28*, 451–457. https://doi.org/10.5607/en.2019.28.4.451.

42. Reimer, R.J. (2013). SLC17: a functionally diverse family of organic anion transporters. Mol. Aspects Med. *34*, 350–359. https://doi.org/10.1016/j.mam.2012.05.004.

43. Serrano-Saiz, E., Vogt, M.C., Levy, S., Wang, Y., Kaczmarczyk, K.K., Mei, X., Bai, G., Singson, A., Grant, B.D., and Hobert, O. (2020). SLC17A6/7/8 Vesicular Glutamate Transporter Homologs in Nematodes. Genetics *214*, 163–178. https://doi.org/10.1534/genetics.119.302855.

44. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

45. Butts, T., Green, M.J., and Wingate, R.J.T. (2014). Development of the cerebellum: simple steps to make a 'little brain'. Development *141*, 4031–4041. https://doi.org/10.1242/dev.106559.

46. Hashimoto, M., and Hibi, M. (2012). Development and evolution of cerebellar neural circuits. Dev. Growth Differ. *54*, 373–389. https://doi.org/10.1111/j.1440-169X.2012.01348.x.

47. Hibi, M., Matsuda, K., Takeuchi, M., Shimizu, T., and Murakami, Y. (2017). Evolutionary mechanisms that generate morphology and neural-circuit diversity of the cerebellum. Dev. Growth Differ. *59*, 228–243. https://doi.org/10.1111/dgd.12349.

48. Sugahara, F., Murakami, Y., Pascual-Anaya, J., and Kuratani, S. (2017). Reconstructing the ancestral vertebrate brain. Dev. Growth Differ. *59*, 163–174. https://doi.org/10.1111/dgd.12347.

49. Kebschull, J.M., Richman, E.B., Ringach, N., Friedmann, D., Albarran, E., Kolluru, S.S., Jones, R.C., Allen, W.E., Wang, Y., Cho, S.W., et al. (2020). Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. Science *370*, eabd5059. https://doi.org/10.1126/science.abd5059.

50. Sepp, M., Leiss, K., Murat, F., Okonechnikov, K., Joshi, P., Leushkin, E., Spänig, L., Mbengue, N., Schneider, C., Schmidt, J., et al. (2024). Cellular development and evolution of the mammalian cerebellum. Nature *625*, 788–796. https://doi.org/10.1038/s41586-023-06884-x.

51. Cen, X., Nitta, A., Ibi, D., Zhao, Y., Niwa, M., Taguchi, K., Hamada, M., Ito, Y., Ito, Y., Wang, L., and Nabeshima, T. (2008). Identification of Piccolo as a regulator of behavioral plasticity and dopamine transporter internalization. 451–463. Mol. Psychiatry *13*, 349. https://doi.org/10.1038/sj.mp.4002132.

52. Favuzzi, E., Deogracias, R., Marques-Smith, A., Maeso, P., Jezequel, J., Exposito-Alonso, D., Balia, M., Kroon, T., Hinojosa, A.J., F Maraver, E., et al. (2019). Distinct molecular programs regulate synapse specificity in cortical inhibitory circuits. Science *363*, 413–417. https://doi.org/10.1126/science.aau8977.

53. Petrenko, A.G., Ullrich, B., Missler, M., Krasnoperov, V., Rosahl, T.W., and Südhof, T.C. (1996). Structure and evolution of neurexophilin. J. Neurosci. *16*, 4360–4369. https://doi.org/10.1523/JNEUROSCI.16-14-04360.1996.

54. Chen, X., Du, Y., Broussard, G.J., Kislin, M., Yuede, C.M., Zhang, S., Dietmann, S., Gabel, H., Zhao, G., Wang, S.S.H., et al. (2022). Transcriptomic mapping uncovers Purkinje neuron plasticity driving learning. Nature *605*, 722–727. https://doi.org/10.1038/s41586-022-04711-3.

55. Guissoni Campos, L.M., Hataka, A., Vieira, I.Z., Buchaim, R.L., Robalinho, I.F., Arantes, G.E.P.S., Viégas, J.S., Bosso, H., Bravos, R.M., and Pinato, L. (2018). Circadian Clock Proteins and Melatonin Receptors in Neurons and Glia of the Sapajus apella Cerebellum. Front. Physiol. *9*, 5. https://doi.org/10.3389/fphys.2018.00005.

56. Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., Liu, S., Garcia, B.A., Wu, H., Wu, F., et al. (2015). A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. Mol. Cell *57*, 957–970. https://doi.org/10.1016/j.molcel.2015.01.010.

57. Lei, Y., Liang, X., Sun, Y., Yao, T., Gong, H., Chen, Z., Gao, Y., Wang, H., Wang, R., Huang, Y., et al. (2024). Region-specific transcriptomic responses to obesity and diabetes in macaque hypothalamus. Cell Metab. *36*, 438–453.e6. https://doi.org/10.1016/j.cmet.2024.01.003.

58. Liao, K., Xiang, Y., Huang, F., Huang, M., Xu, W., Lin, Y., Liao, P., Wang, Z., Yang, L., Tian, X., et al. (2024). Spatial and single-nucleus transcriptomics decoding the molecular landscape and cellular organization of avian optic tectum. iScience *27*, 109009. https://doi.org/10.1016/j.isci.2024.109009.

59. Schmidt, A., and Bischof, H.J. (2001). Neurons with complex receptive fields in the stratum griseum centrale of the zebra finch (Taeniopygia guttata castanotis Gould) optic tectum. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *187*, 913–924. https://doi.org/10.1007/s00359-001-0264-8.

60. Wang, S., Ding, P., Yuan, J., Wang, H., Zhang, X., Chen, D., Ma, D., Zhang, X., and Wang, F. (2022). Integrative cross-species analysis of GABAergic neuron cell types and their functions in Alzheimer's disease. Sci. Rep. *12*, 19358. https://doi.org/10.1038/s41598-022-21496-7.

61. Stacho, M., Herold, C., Rook, N., Wagner, H., Axer, M., Amunts, K., and Güntürkün, O. (2020). A cortex-like canonical circuit in the avian forebrain. Science *369*, eabc5534. https://doi.org/10.1126/science.abc5534.

62. Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S.L., Wiebe, V., Kitano, T., Monaco, A.P., and Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. Nature *418*, 869–872. https://doi.org/10.1038/nature01025.

63. Haesler, S., Wada, K., Nshdejan, A., Morrisey, E.E., Lints, T., Jarvis, E.D., and Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. J. Neurosci. *24*, 3164–3175. https://doi.org/10.1523/JNEUROSCI.4369-03.2004.

64. Staes, N., Sherwood, C.C., Wright, K., de Manuel, M., Guevara, E.E., Marques-Bonet, T., Krützen, M., Massiah, M., Hopkins, W.D., Ely, J.J., and Bradley, B.J. (2017). FOXP2 variation in great ape populations offers insight into the evolution of communication skills. Sci. Rep. *7*, 16866. https://doi.org/10.1038/s41598-017-16844-x.

65. Sun, W., Liu, Z., Jiang, X., Chen, M.B., Dong, H., Liu, J., Südhof, T.C., and Quake, S.R. (2024). Spatial transcriptomics reveal neuron-astrocyte synergy in long-term memory. Nature *627*, 374–381. https://doi.org/10.1038/s41586-023-07011-6.

66. Birchler, J.A., and Veitia, R.A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol. *186*, 54–62. https://doi.org/10.1111/j.1469-8137.2009.03087.x.

67. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531–1545. https://doi.org/10.1093/genetics/151.4.1531.

68. Fremeau, R.T., Jr., Troyer, M.D., Pahner, I., Nygaard, G.O., Tran, C.H., Reimer, R.J., Bellocchio, E.E., Fortin, D., Storm-Mathisen, J., and Edwards, R.H. (2001). The expression of vesicular glutamate transporters defines two classes of excitatory synapse. Neuron *31*, 247–260. https://doi.org/10.1016/s0896-6273(01)00344-0.

69. Fremeau, R.T., Jr., Voglmaier, S., Seal, R.P., and Edwards, R.H. (2004). VGLUTs define subsets of excitatory neurons and suggest novel roles for glutamate. Trends Neurosci. *27*, 98–103. https://doi.org/10.1016/j.tins.2003.11.005.

70. Olkowicz, S., Kocourek, M., Lučan, R.K., Porteš, M., Fitch, W.T., Herculano-Houzel, S., and Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. Proc. Natl. Acad. Sci. USA *113*, 7255–7260. https://doi.org/10.1073/pnas.1517131113.

71. Weidemüller, P., Kholmatov, M., Petsalaki, E., and Zaugg, J.B. (2021). Transcription factors: Bridge between cell signaling and gene regulation. Proteomics *21*, e2000034. https://doi.org/10.1002/pmic.202000034.

72. Rosen, Y., Brbić, M., Roohani, Y., Swanson, K., Li, Z., and Leskovec, J. (2024). Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. Nat. Methods *21*, 1492–1500. https://doi.org/10.1038/s41592-024-02191-z.

73. Rosen, Y., Roohani, Y., Agarwal, A., Samotorčan, L., Consortium, T.S., Quake, S.R., and Leskovec, J. (2023). Universal Cell Embeddings: A

Foundation Model for Cell Biology. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.28.568918.

74. Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., Hua, C., Yang, F., Liu, W., Peng, C., et al. (2020). CNSA: a data repository for archiving omics data. Database (Oxford) *2020*, baaa055. https://doi.org/10.1093/database/baaa055.

75. Chen, F.Z., You, L.J., Yang, F., Wang, L.N., Guo, X.Q., Gao, F., Hua, C., Tan, C., Fang, L., Shan, R.Q., et al. (2020). CNGBdb: China National GeneBank DataBase. Yi Chuan *42*, 799–809. https://doi.org/10.16288/j.yczz.20-080.

76. Xu, Z., Wang, W., Yang, T., Li, L., Ma, X., Chen, J., Wang, J., Huang, Y., Gould, J., Lu, H., et al. (2024). STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. Nucleic Acids Res. *52*, D1053–D1061. https://doi.org/10.1093/nar/gkad933.

77. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

78. Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., Li, C., White, S., Xiong, Z., Fang, D., et al. (2013). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat. Genet. *45*, 701–706. https://doi.org/10.1038/ng.2615.

79. Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H., Huff, C.D., Hu, H., Vickrey, A.I., et al. (2013). Genomic diversity and evolution of the head crest in the rock pigeon. Science *339*, 1063–1067. https://doi.org/10.1126/science.1230422.

80. Wu, T.D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M.J. (2016). GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. Methods Mol. Biol. *1418*, 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15.

81. Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Res. *9*, 304. https://doi.org/10.12688/f1000research.23297.2.

82. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

83. Fischer, S., Crow, M., Harris, B.D., and Gillis, J. (2021). Scaling up reproducible research for single-cell transcriptomics using MetaNeighbor. Nat. Protoc. *16*, 4031–4067. https://doi.org/10.1038/s41596-021-00575-5.

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Zebra finch (male and female, 12-month-old)	This study	N/A
Chinese softshell turtle (male and female)	This study	N/A
Pigeon (male and female, 12-month-old)	This study	N/A
Deposited data		
Public macaque snRNA-seq data	Chen et al.[23]	https://db.cngb.org/search/project/CNP0002035
Public mouse snRNA-seq data	Bakken et al.[14]; Yao et al.[21]; Kozareva et al.[22]	https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x; https://assets.nemoarchive.org/dat-ek5dbmu; https://singlecell.broadinstitute.org/single_cell/study/SCP795/
Raw data of stereo-seq and snRNA-seq data	This study	https://db.cngb.org/search/project/CNP0003026
Processed data of snRNA-seq data	This study	https://db.cngb.org/cdcp/dataset/SCDS0000639/
Processed data of stereo-seq data	This study	https://db.cngb.org/stomics/datasets/STDS0000241
Refined gtf for turtle and pigeon	This study	https://github.com/Dee-chen/Refined-gtf-for-turtle-and-pigeon
Oligonucleotides		
cDNA PCR primer: CTGCTGACGTACTGAGAGGC	Sangon	N/A
Stereo-seq-library-F: /5phos/CTGCTGACGTACTGAGAGG*C*A	Sangon	N/A
Stereo-seq-library-R: GAGACGTTCTCGACTCAGCAGA	Sangon	N/A
Stereo-seq-library-splint-oligo: GTACGTCAGCAGGAGACGTTCTCG	Sangon	N/A
Stereo-seq-read1: CTGCTGACGT ACTGAGAGGCATGGCGACCTTATCAG	Sangon	N/A
Stereo-seq-MDA-primer: TCTGCTGAGTCGAGAACGTC	Sangon	N/A
Stereo-seq-read2: GCCATGTC GTTCTGTGAGCCAAGGAGTT	Sangon	N/A
Software and algorithms		
SAW	https://github.com/BGIResearch/SAW	V2.1.0
GMAP	Wu et al.[80]	https://github.com/juliangehring/GMAP-GSNAP; version 2023-04-28
STAR	Dobin et al.[77]	https://github.com/alexdobin/STAR; V2.5.3
PISA	https://github.com/shiquan/PISA	V0.12
MetaNeighbor	Fischer et al.[83]	https://github.com/maggiecrow/MetaNeighbor
entropy-scGCN	https://github.com/Dee-chen/entropy-scGCN	N/A
scGCN	Song et al.[25]	https://github.com/QSong-github/scGCN

(*Continued on next page*)

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Tree2gd	Chen et al.[31]	https://github.com/Dee-chen/Tree2gd; V1.0.40
ggtree	https://github.com/YuLab-SMU/ggtree	V3.0.4
R	https://cran.r-project.org	V4.0.3
Seurat	https://satijalab.org/seurat/	V4.1.1
clusterProfiler	http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html	V4.2.2
ggplot2	https://cran.rstudio.com/web/packages/ggplot2/index.html	V3.3.6
org.Hs.eg.db	https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html	V1.14.2
pysal	https://pysal.org/pysal/	V2.7.0
python	https://www.python.org/	V3.9.0
anndata	https://anndata.readthedocs.io/en/latest/	V0.9.2
pySCENIC	https://github.com/aertslab/pySCENIC	0.12.1

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

Experimental procedures were performed in accordance with the guidelines of the Animal Care and Use Committees at the Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), China (permit number SIAT-IACUC-210326-NS-WH-A1881). Experiments were performed using 8 male Chinese softshell turtles (Pelodiscus sinensis) weighing 450-600 g, 8 female and male adult (>12 months old) zebra finches (Taeniopygia guttata), and 2 pigeons (Columba livia) at 12 months of age. Turtles were obtained from external breeders and zebra finches from the colony at Zhengzhou university, China.

METHOD DETAILS

Library preparation and sequencing

One-hundred ng of cDNA (20 μl) from each sample were tagmented with Tn5 transposases (Vazyme) at 55°C for 10 mins, followed by quenching the reaction using 5 μL of 0.02% SDS. Subsequently, a PCR reaction mix (75 μL, Library HIFI Master Mix, Library PCR primer mix) was added to each fragmented cDNA sample. The samples were then subjected to thermal cycling for amplification using the following protocol: an initial cycle at 95°C for 5 min, followed by 13 cycles of tri-temperature reactions (98°C for 20 s, 58°C for 20 s and 72°C for 30 s), and a final cycle at 72°C for another five minutes. After amplification, the PCR products were purified using VAHTSTM DNA Clean Beads (VAZYME, N411-03) in two steps with concentrations of beads set as follows: first step with a ratio of bead volume to sample volume as low as approximately six-tenths and second step with a ratio around two-tenths; subsequently these purified products were utilized in DNB (DNA Nano Ball) generation process. Finally, the generated DNBs underwent sequencing on the DNBSEQ™ T10 platform (MGI, Shenzhen, China), employing read1 length of fifty base pairs and read2 length of one hundred base pairs.

Brain tissue collection for snRNA-seq

The snRNA-seq samples were collected from frozen sections adjacent to those for Stereo-seq. These sections were cut at 50-μm thickness, and 3 to 5 sections for each coronal coordinate were collected for snRNA-seq analysis. Sections were transferred into plastic wells on dry ice and stored in a -80 °C refrigerator. Each section was further segmented into distinct areas on dry ice using tissue punchers (5 - 8 mm in diameter). Tissues at the same brain regions were combined in a prechilled pipe as one sample. In the cryostat, the cortical areas were segmented at 1 - 2 mm depth using tissue punchers (2.5 - 4 mm in diameter). After dissection, the samples were immediately frozen in liquid nitrogen and then kept in dry ice or -80 °C refrigerator. Throughout the sampling manipulation, the tissues were carefully transferred to pre-cold tube without thawing.

Single-nucleus suspension preparation

Single nucleus suspension was prepared as previously described.[14] Briefly, frozen brain tissue pieces were placed in Dounce homogenizer with 2 ml pre-chilled homogenization buffer and kept the Dounce homogenizer on ice during grinding. Tissue was homogenized with 10-15 strokes of the pestle A and followed by 10-15 strokes of the pestle B, then added 2 ml homogenization buffer to the Dounce homogenizer and filtered the homogenate through 30μm MACS SmartStrainers (Miltenyi Biotech, #130-110-915) into 15 ml

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

conical tube and centrifuged at 500 g for 5 mins at 4°C to pellet nuclei, then the pellet was resuspended in 1.5 ml of blocking buffer and centrifuged at 500 g for 5 mins at 4°C to pellet nuclei. Nuclei were resuspended with cell resuspension buffer for subsequent snRNA-seq library preparation.

snRNA-seq data processing

Initially, bead barcodes and unique molecular identifier (UMI) sequences were extracted using the parse function in PISA (https://github.com/shiquan/PISA). For cDNA libraries, read1 encompassed bead barcodes from positions 1-10bp and 11-20bp, while the UMI sequence was located at position 21-30bp. The entire read2 (100bp) was utilized for downstream alignment analysis. In the case of Droplet Index libraries, read1 contained bead barcodes at positions 1-10bp and 11-20bp, with the UMI sequence present at position 1-10bp of read2. Additionally, droplet index barcodes were found at positions 11-20bp and 21-30bp of read2. Reads with incorrect barcodes based on the barcode list were excluded from further analysis. Subsequently, STAR[77] was employed to align snRNA-seq data against the genome reference. To estimate the actual number of beads, we used the ''barcodeRanks'' function of DropletUtils tool (https://github.com/MarioniLab/DropletUtils/) to find the threshold value of sharp transition in total UMI counts distribution. Beads with UMI counts less than the threshold were removed. We merged the beads considered to be one cell, and counted the gene expression of cells by PISA.

Gene family clustering and evolutionary event identification

In order to determine the homology relationship of gene sequences of amniotic animals and locate evolutionary events such as gene duplication, acquisition and loss, and positive selection, we downloaded CDS and protein sequences of mice, humans, cynomolgus monkeys, and zebra finches from NCBI data. We used third-generation sequencing to re-annotate existing gene annotations for pigeons and Chinese soft-shell turtles. After that, We performed a complete genetic sequence evolution analysis using the default parameters of Tree2gd[31] on zebrafish with sequence of each species for the outgroup. It includes interspecies blastp, gene gain and loss detection, gene duplication detection and ka/ks gene natural selection pressure analysis. Through Tree2gd's gene family clustering, a total of 38,948 gene families were obtained.

Cell downsampling and cross-species integration

Basic processing and visualization of the scRNA-seq data were performed with the Seurat (v.4.3.0) in R (v.4.2.2). We used biomart to convert gene names between species and found 13,835 genes with the same name across species for integration analysis. We discarded cells with the number of genes (nFeatureRNA) less than 2,00, the number of genes (nCount) more than 2,500 and the percentage of mitochondrial genes (percent.mt) larger than 5%. After the first quality control, a total of 1,391,446 cells from five species were used for cell type classification and scGCN annotation. In order to keep the number of cells consistent across species, we subsampled 1000 cells per cluster for pigeons, zebra finches, and turtles (Table S1), and finally got 5 data set of ~200,000 cells per species (Figures 2A and S2C). We integrated the data from mice, macaque, pigeons, zebra finches, and turtles using the "IntegrateData" function of the Seurat R package (v 4.3.0), with marker genes as anchors.Then principal component analysis (PCA) was carried out and the top 30 components were used for downstream analysis. The integrated datasets were re-clustered with k.param = 20 and resolution = 0.5 using ''FindNeighbors'' and ''FindClusters'' functions. Through integrated analysis, the cell type annotation results were consistent and accurate in all five datasets (Figures 2A and S2C). The dataset shown in Figure 2A contained 195,162 cells, and subsequent cross-species analyses were based on this dataset. We categorized the 63 cell clusters into 19 categories based on classical marker genes, including glutamatergic neurons, GABAergic neurons and non-neuronal cells. Then the 19 categories of cells were further iteratively classified into more cell clusters with higher resolution.

SAMap cross-species cell type similarity analysis

To further validate the accuracy and robustness of our cell type annotations derived from the entropy-scGCN method, we employed SAMap[29] to independently map cell types across the five studied species: turtle, zebra finch, pigeon, mouse, and macaque. SAMap was chosen for its capability to align homologous genes between species, thereby facilitating a comparative analysis of cell type conservation and divergence.

For each species, we processed the single-cell RNA-seq data through a standardized pipeline to ensure consistency in library preparation and sequencing depth. Homologous genes were identified using reciprocal best hit BLAST searches against a concatenated database of all species, followed by filtering for 1-to-1 orthologs. The input gene sequence data was consistent with scGCN and Tree2gd. These orthologous gene sets served as input for SAMap, enabling the construction of cross-species gene expression matrices.

Using SAMap, we aligned the expression profiles of these orthologous genes across species, generating a series of pairwise mappings. The alignment scores were then used to infer the degree of similarity between cell types, allowing us to determine whether a given cell type in one species had a corresponding counterpart in another.

To evaluate the performance of SAMap, we calculated the mapping consistency scores between all species pairs. High consistency scores indicated that SAMap effectively captured conserved cell types across closely related species, whereas lower scores pointed to difficulties in mapping between more distantly related species, such as turtles and mice (Figure S2E). These results were used to corroborate the cell type annotations obtained from the entropy-scGCN method, providing an independent benchmark for the accuracy of our cross-species comparisons.

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

Update turtle and pigeon genome annotations by IsoSeq

To capture the full-length transcriptome, we utilized PacBio sequencing technology. For this purpose, we cryopreserved whole brain samples from turtles and pigeons. High-quality RNA was extracted from these samples under stringent conditions to ensure minimal degradation. The RNA integrity was assessed using a Bioanalyzer, and only samples with an RNA Integrity Number (RIN) above 8.0 were processed further. The PacBio data were processed using IsoSeq v3.4.0 (https://github.com/PacificBiosciences/IsoSeq). Only sequence reads containing both 5' and 3' adaptors were retained to cover the entire transcript. Subsequently, the corrected SMRT reads were aligned to the reference genome PelSin_1.0[78] and Cliv_2.1[79] using GMAP[80] to locate the position of the predicted genes on the chromosomes.A total of 60,167 and 48,142 transcripts were successfully aligned to the genome. We used Gffcompare (https://ccb.jhu.edu/software/stringtie/gffcompare.shtml)[81] to compare the assembled transcript models with the transcript models of the reference genome and concordant transcripts (class code "=") were merged. For cross-species comparisons, the combined annotated protein sequences were aligned using BLAST+[82] and proteins from other species in the article. The genes on the alignment were renamed to their corresponding gene names, and the others retained the gene ID given by the software. The snRNA-seq ratio was significantly improved in the data in this paper (table below). The gtf file is available at https://github.com/Dee-chen/Refined-gtf-for-turtle-and-pigeon.

Species	Turtle		Pigeon	
Gtf version	PelSin_1.0	Refined gtf	Cliv_2.1	Refined gtf
Reads mapped to genome	88.50%	88.50%	91.00%	91.00%
Reads mapped to exonic regions	18.30%	27.90%	26.60%	35.80%
Reads mapped to intronic regions	28.60%	31.90%	21.80%	40.40%
Reads mapped across exon and intron regions	0.90%	1.00%	1.70%	1.70%
Reads mapped antisense to gene	14.20%	15.40%	6.70%	7.80%
Reads mapped to intergenic regions	37.80%	23.70%	42.90%	13.90%
Unidentified type	0.10%	0.20%	0.30%	0.30%

Cell type differentially expressed gene analysis and species comparison

In this study, the data was split by species, and differential gene expression analysis was performed using the "FindAllMarkers" function. To identify genes that were significantly differentially expressed, we applied a filter with the criterion avg_log2FC > 1, resulting in a range of 43,864 to 56,490 differentially expressed genes of each species.To investigate the commonalities among the different species, we assessed the intersection of these differentially expressed genes. The resulting gene intersections were visualized using UpSetR v1.4.0, and an UpSet plot was generated to illustrate the overlapping gene sets.

Furthermore, we extended our analysis to gene families by following a similar approach. We obtained gene families for each species, encompassing a range of 34,151 to 51,547 gene families. To identify the intersections of gene families among the different species, we employed UpSetR v1.4.0.

Co-Expression Regulatory Network Analysis Using SCENIC

To investigate differences in transcriptional regulation of cell type cores across species, we utilized SCENIC[32] for co-expression regulatory network analysis (Figure S3D).

For excitatory neurons across five amniote species, we compared and analyzed significantly divergent transcription factor (TF) regulatory networks, identifying 43 distinct gene sets. The transcription factors and target genes in each regulatory network were intersected with gene duplication events identified using Tree2gd. This approach enabled us to quantify the proportion of genes that underwent duplication and pinpoint the evolutionary nodes at which these events occurred (Figure S3E).

For Purkinje cell types in mice and zebra finches, we performed a differential TF regulatory network analysis focusing on two distinct subtypes in each species. After subdividing Purkinje cells based on differential gene expression, SCENIC was applied to identify potential transcriptional regulators and their target genes.

To ensure cross-sample comparability, normalized gene expression matrices were prepared for Purkinje cells. Differential expression analysis was conducted to identify genes significantly upregulated or downregulated in each Purkinje cell subtype. These identified genes were then used by SCENIC to infer co-expression networks, where nodes represented genes and edges denoted significant correlations between their expression levels (Figure S5H).

SCENIC subsequently identified putative regulons by integrating the inferred co-expression networks with known transcription factors. This analysis highlighted the master regulators responsible for orchestrating gene expression patterns within each Purkinje cell subtype. The resulting networks were visualized to identify modules of co-regulated genes, offering detailed insights into the functional pathways and regulatory mechanisms active in each subtype.

Please cite this article in press as: Chen et al., Genomic evolution reshapes cell-type diversification in the amniote brain, Developmental Cell (2025), https://doi.org/10.1016/j.devcel.2025.04.014

QUANTIFICATION AND STATISTICAL ANALYSIS

Conservative entropy calculation across species cell types

We have innovated and improved on scGCN[25] by developing a multi-omics cross-species cell-type annotation and conservation assessment software sc-Entropy based on information entropy theory. Firstly, the FASTA format protein sequence files of the desired species were downloaded, and subsequently the Tree2GD[31] tool was used to identify the gain, loss, and duplication events of all genes based on the relationship between the sequences to construct the gene family evolutionary lineage tree. Therefore, entropy-scGCN is not only suitable for cell type mapping of single genes, but also for cell type mapping at the gene family level, solving the limitation of using only one-to-one homologous genes. In order to measure how conservative the prediction results of different subgroups are, we introduce the calculation of information entropy in entropy-scGCN.

Firstly, by mapping the cell types of target species and reference species through scGCN, the predicted cell types of each cell can be obtained. Then, according to the test species self-clustering results, the proportion matrix P_{ij} of cells predicted as reference cell types (j) in each target specie's cluster (i) was calculated.

We refer to the calculation formula of information entropy, and then transform the probability matrix(P_{ij}) to get the basic similarity entropy (E_i). The basic information entropy could be obtained by using the calculation formula of information entropy (see Equation 1).

$$E\text{-}raw_i = P_{ij} * Ln(P_{ij}) \quad \text{(Equation 1)}$$

The original entropy($E\text{-}raw_i$) only contains information about whether the probability distribution of mapped cell types obtained by entropy-scGCN is reasonable, without considering the influence of the similarity relationship between cell types in the reference data set.

In order to make the entropy more theoretical in biological science, we consider the similarities and differences between cell types. MetaNeighbor[83] was used to calculate the similarity of cell types in the reference data set, and the similarity distance score(Dis_{ij}) between reference cell types(j) was obtained. The matrix multiplication of the distance score(Dis_{ij}) with the probability matrix(P_{ij}) was performed, and the cell type similarity matrix(C_{ij}) was obtained by normalization(see Equation 2).

$$C_{ij} = P_{ij} \times Dis_{jj} \quad \text{(Equation 2)}$$

The formula for calculating entropy value after adding the distance factor between cell types in the reference data set is as follows (see Equation 3):

$$E_i = C_{ij} * Ln(C_{ij}) \quad \text{(Equation 3)}$$

The smaller the entropy value, the more homogeneous the cell type obtained by entropy-scGCN annotation, that is, a higher proportion of the cell type was annotated as one cell type rather than predicted as multiple cell types.