



OPEN

A scATAC-seq atlas of chromatin accessibility in axolotl brain regions

DATA DESCRIPTOR

Weimin Feng^{1,2,3,14}, Shuai Liu^{2,3,4,14}, Qiuting Deng^{1,3,14}, Sulei Fu^{5,6,14}, Yunzhi Yang^{2,3,4}, Xi Dai^{1,2,3}, Shuai Wang^{1,2,3}, Yijin Wang^{2,3,7}, Yang Liu^{2,3}, Xiumei Lin^{1,2,3}, Xiangyu Pan^{8,9}, Shijie Hao^{1,2,3}, Yue Yuan^{2,3}, Ying Gu³, Xiuqing Zhang³, Hanbo Li^{3,10,11}, Longqi Liu^{2,3}, Chuanyu Liu³ ✉, Ji-Feng Fei^{5,12,13} ✉ & Xiaoyu Wei^{2,3} ✉

Axolotl (*Ambystoma mexicanum*) is an excellent model for investigating regeneration, the interaction between regenerative and developmental processes, comparative genomics, and evolution. The brain, which serves as the material basis of consciousness, learning, memory, and behavior, is the most complex and advanced organ in axolotl. The modulation of transcription factors is a crucial aspect in determining the function of diverse regions within the brain. There is, however, no comprehensive understanding of the gene regulatory network of axolotl brain regions. Here, we utilized single-cell ATAC sequencing to generate the chromatin accessibility landscapes of 81,199 cells from the olfactory bulb, telencephalon, diencephalon and mesencephalon, hypothalamus and pituitary, and the rhombencephalon. Based on these data, we identified key transcription factors specific to distinct cell types and compared cell type functions across brain regions. Our results provide a foundation for comprehensive analysis of gene regulatory programs, which are valuable for future studies of axolotl brain development, regeneration, and evolution, as well as on the mechanisms underlying cell-type diversity in vertebrate brains.

Background & Summary

The axolotl (*Ambystoma mexicanum*) has remarkable regenerative abilities, including the ability to regenerate its limbs, heart, tail, and spinal cord^{1–5}, making it as an ideal model for studying vertebrate regeneration, comparative genomics, and evolution. With the development of technology, great progress has been made in axolotl genome and transcriptome research. In brief, recent advances in single-cell and spatial transcriptomics assays have led to profiling the distribution of cell types of the axolotl telencephalon *in situ* during development and regeneration⁶. Furthermore, by using either species-shared differentially expressed genes or transcription factors (TFs), previous axolotl telencephalon studies have shown the conservation between the medial pallium neurons and amniote hippocampal neurons⁷. Interestingly, the axolotl genome has conserved coding regions with the human genome but is ten times its size, with the majority of the expansion in non-coding regions. With a 32-gigabase-pair genome⁸, non-coding regulatory DNA sequences in axolotls may be proposed to play

¹College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China. ²BGI-Hangzhou, Hangzhou, 310012, China. ³BGI-Shenzhen, Shenzhen, 518103, China. ⁴BGI College & Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, 450000, China. ⁵Department of Pathology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, 510080, China. ⁶Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education; Institute for Brain Research and Rehabilitation, South China Normal University, Guangzhou, 510631, China. ⁷College of Life Sciences, Nankai University, Tianjin, 300071, China. ⁸Medical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China. ⁹Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, 510080, China. ¹⁰BGI-Qingdao, Qingdao, 266555, China. ¹¹Lars Bolund Institute of Regenerative Medicine, Qingdao-Europe Advanced Institute for Life Sciences, BGI-Qingdao, Qingdao, 266555, China. ¹²School of Medicine, South China University of Technology, Guangzhou, Guangdong, 510006, China. ¹³School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, 510515, China. ¹⁴These authors contributed equally: Weimin Feng, Shuai Liu, Qiuting Deng, Sulei Fu. ✉e-mail: liuchuan@genomics.cn; jifengfei@gdph.org.cn; weixiaoyu@genomics.cn

a distinctive role in environmental adaptation. However, a comprehensive understanding of *cis*-regulatory elements in the axolotl genome, especially in complex organs like the brain, has lagged.

Using techniques such as Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and chromatin immunoprecipitation followed by sequencing (ChIP-seq), candidate *cis*-regulatory elements (cCREs) have been mapped to identify and exploit epigenetic features^{9,10}. Due to the extreme cellular heterogeneity of the brain, conventional assays that use bulk tissue samples have difficulty identifying cCREs in specific cell types. Single-cell ATAC sequencing (scATAC-seq) is a powerful technique that allows us to identify the regulatory regions of the genome that are active in individual cells^{11–14}. It was developed to determine chromatin accessibility in various biological scenarios and the corresponding transcriptional programs. In recent years, scATAC-seq has gained significant attention in the field of neuroscience, because it allows for the investigation of the epigenetic landscape of cell-type-specific transcriptional regulatory sequences in brain regions^{15–17}. For example, regulatory elements linked to *Cux2* and *Foxp2*, which are highly expressed only in layers II–IV and layer VI of mice, respectively, are only accessible in cells of the corresponding neuron clusters, reflecting the heterogeneity of chromatin accessibility in different layers of the prefrontal cortex¹⁵.

In this study, we profiled a comprehensive chromatin accessibility landscape across the axolotl brain with scATAC-seq^{16,18,19}, including the olfactory bulb (OB), telencephalon (Tel), diencephalon and mesencephalon (DM), hypothalamus and pituitary (HP), and rhombencephalon (Rho). We were able to obtain single-cell chromatin accessibility for 81,199 cells after applying quality control. By analysing the chromatin accessibility profiles of different cell types, we have identified unique sets of regulatory elements that are specific to each cell type, including microglia, GABAergic, glutamatergic, ependymoglia cell (EGC), and so on. Overall, our findings provide insights into the regulatory mechanisms that underlie the diverse functions of different cell types in axolotl brain regions, and serve as a foundation for future studies aimed at understanding the transcriptional programs that drive the distinguishing features of axolotl brain regions.

Methods

Tissue collection. The Biomedical Research Ethics Committee of Guangdong Provincial People's Hospital (license number: KY-Q-2022-395-01) approved the use of animals in this study. The *d/d* strain of axolotl was used for all experiments without sex bias in the research. Animals were bred and maintained in freshly dechlorinated tap water at 18–20 °C with a 12 h/12 h light-dark cycle in the Laboratory of Neural Development and Regeneration, Guangdong Provincial People's Hospital. Totally 8 adult axolotls were sacrificed, and OB, Tel, DM, HP, and Rho were harvested in this study. In brief, axolotls were deeply anaesthetized using 0.03% ethyl-p-aminobenzoate. Each brain region from 8 animals was dissected from the axolotl head. Then, each brain region was pooled into a separate tube for single-nucleus isolation, respectively. The samples were snap-frozen in liquid nitrogen and then transferred to a –80 °C freezer for storage before dissociation.

Nuclear isolation from frozen brain tissue. The axolotl brain regions focused on this study were the OB, Tel, DM, HP, and Rho (Supplementary Table 1 and Fig. 1a). The single-nucleus preparations were obtained using the Omni-ATAC protocol with certain modifications, as previously described²⁰. In brief, nuclei from the frozen brain regions were isolated^{21,22}. Tissue was cut into pieces and ground in 2 mL of chilled homogenization buffer [HB; 120 mM Tris pH 7.8 (Sigma), 150 mM KCl (Sigma), 30 mM MgCl₂ (Invitrogen), 250 mM sucrose (Sigma), 0.1% NP-40 (Roche), 1 × Protease inhibitor cocktail (Roche), 1 mM DTT (Thermo Fisher Scientific), 1% BSA/0.8 × PBS], and incubated on ice for 5 min. The tissue was homogenized by stroking (grinding) to release the nuclei. Then, nuclei were filtered through a 30 μm cell strainer (Sysmex Partec) into a 15 mL centrifuge tube. After centrifugation at 500 g and 4 °C for 5 min, the nuclei were collected and washed twice with 1 mL of chilled blocking buffer [BB, 1% BSA/0.8 × PBS]. After another round of centrifugation, the nuclei were resuspended in 50 μl of 1% BSA/0.8 × PBS and counted by staining with DAPI.

scATAC-Seq library preparation and sequencing. For the preparation of single-cell ATAC-seq libraries, we employed the DNBelab C Series Single-Cell ATAC Library Prep Set (MGI, #1000021878)¹⁹. To investigate the transcriptional regulation of different brain regions in axolotl with sufficient and high-quality data, a total of 39 libraries were generated. Specifically, there were 8 libraries for OB, DM, HP, and Rho, and 7 libraries for Tel. After quality control, 25 high-quality libraries were retained, including 6 from the OB, 4 from the Tel, 3 from the DM, 6 from the HP, and 6 from the Rho. The barcoded scATAC-seq libraries were generated from the transposed single-cell suspensions. In summary, the protocol involved droplet encapsulation, pre-amplification, emulsion breakage, collection of capture beads, DNA amplification, and purification. Following this, an indexed sequencing library was prepared as per the user guide. Finally, we used the Qubit ssDNA Assay Kit (Thermo Fisher Scientific) to measure the concentrations of the sequencing libraries. The library was sequenced on the MGI DNBSEQ-T1 platform using the following read lengths: 50 bp for read 1, 70 bp for read 2, and 10 bp for the sample index sequencing scheme of the China National GeneBank²³ (CNGB).

scATAC-seq raw data processing. The processing of scATAC-seq data involved the following steps. First, the raw reads were separated into insertions and barcodes and then filtered using PISA (version 1.1) with a minimum sequencing quality of 20; the software is available at <https://github.com/shiquan/PISA>. The sequencing reports for the scATAC-seq datasets are summarized in Table 1. Next, the filtered reads were aligned to the axolotl genome⁸ using BWA (version 0.7.17-r1188)²⁴, and the resulting BAM files were processed using *bap2* (version 0.6.2)²⁵ to identify barcodes from the same cell (Fig. 1b).

Quality control (QC) of the scATAC-seq downstream analysis. ArchR (version 1.0.2)²⁶ was used to filter low quality cells with the following criteria: unique fragments (nFrag) ≥ 1000 and transcription start site

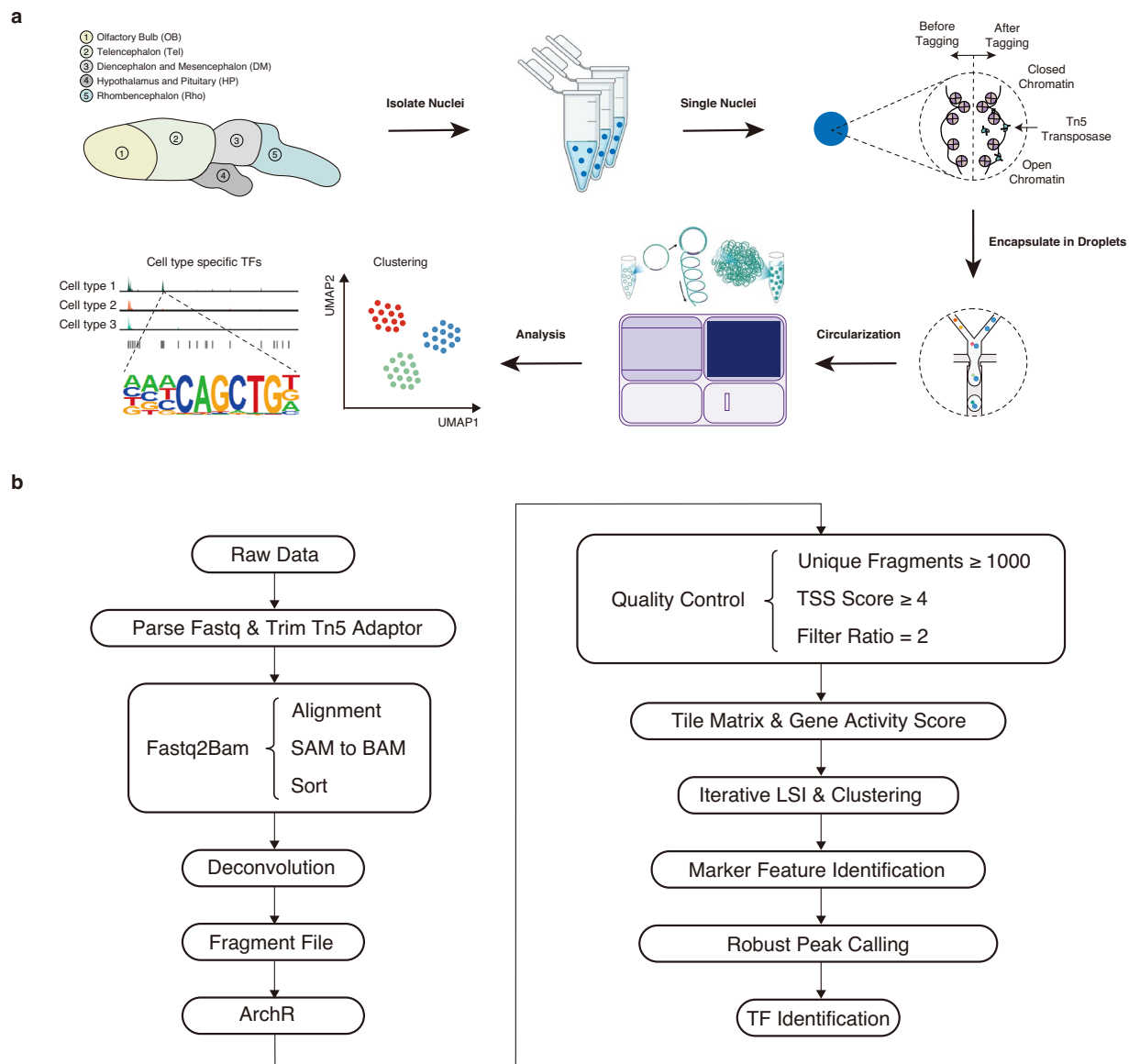


Fig. 1 Schematic diagram illustrating the experimental and data analysis processes of scATAC-seq in the axolotl brain. **(a)** Cartoon illustrates the main experimental and analytical processes. Five different regions from the brain of adult axolotl were collected for single-cell Assay for Transposase Accessible Chromatin (scATAC-seq). Region 1, olfactory bulb (OB); region 2, telencephalon (Tel); region 3, diencephalon and mesencephalon (DM); region 4, hypothalamus and pituitary (HP); and region 5, rhombencephalon (Rho). Colors correspond to the five regions. **(b)** Analysis of the scATAC-seq workflow.

(TSS) score ≥ 4 of each library. Then, we calculated the doublet score by the 'addDoubletScores' function with the parameters of filterRatio = 2 and potential doublets were removed based on the ArchR²⁶.

scATAC-seq gene activity scores. We used ArchR (version 1.0.2)²⁶ to calculate gene activity scores. Briefly, considering the 32-gigabase-pair axolotl genome⁸, to determine whether each cell was accessible within each window, we created 100000 bp windows across the genome. ArchR (version 1.0.2)²⁶ uses a weighted average of the accessibility of the peaks, where the weight of each peak is determined by its proximity to the transcription start site. We calculated gene activity scores using the 'addGeneScoreMatrix' function. Furthermore, due to the sparsity of scATAC-seq data, we employed the MAGIC (Markov Affinity-based Graph Imputation of Cells) imputation technique to impute gene activity scores from bolstering signal strength and reducing noise by sharing information with similar nearby cells.

Latent semantic indexing (LSI) clustering of scATAC-seq. Analyses of the scATAC-seq data were conducted using ArchR²⁶. LSI dimensionality reduction was performed with 'addIterativeLSI' function in ArchR²⁶. By using Seurat's 'FindClusters' function, we clustered the data with parameters: reducedDims = 'IterativeLSI',

Sample ID	Total reads	Reads Pass QC	Estimated number of cells	Mean fragments per cell	TSS Enrichment	Accessible peaks
OB-1	375,820,925	339,441,460	3,651	5,854	20.86	107,713
OB-2	330,716,617	297,016,594	3,594	5,331	20.52	112,426
OB-3	160,832,997	145,264,363	2,922	4,013	19.17	113,486
OB-4	236,547,927	213,626,433	3,243	4,073	17.38	113,332
OB-5	167,666,804	152,023,492	2,793	4,332	18.52	108,282
OB-6	428,341,216	389,533,502	3,441	6,160	20.71	107,890
Tel-1	282,998,946	256,340,446	3,975	5,298	20.98	112,898
Tel-2	170,791,145	154,361,037	3,040	4,385	21.03	114,833
Tel-3	227,647,481	206,999,855	3,376	5,319	20.82	116,451
Tel-4	322,666,500	287,979,852	2,792	6,257	22.21	115,214
DM-1	413,928,279	370,755,560	4,158	5,026	14.44	112,805
DM-2	457,122,784	409,216,317	3,862	5,342	13.62	112,953
DM-3	423,715,798	379,098,525	3,530	6,385	13.85	112,170
HP-1	359,720,792	322,561,635	3,271	6,693	21.59	114,015
HP-2	660,417,261	591,007,407	3,112	8,748	23.20	116,237
HP-3	311,679,700	280,355,891	2,983	6,476	20.93	111,791
HP-4	266,567,225	239,190,771	2,634	5,705	21.74	113,598
HP-5	614,646,512	551,399,386	3,140	8,680	22.72	116,423
HP-6	355,725,790	315,208,623	3,197	5,812	20.71	114,941
Rho-1	604,099,350	539,158,670	2,955	7,493	21.32	114,483
Rho-2	720,833,075	649,038,101	3,338	8,572	20.53	114,006
Rho-3	493,338,543	443,708,686	3,025	8,727	20.87	115,696
Rho-4	427,845,614	386,259,021	2,994	8,268	20.22	115,198
Rho-5	338,051,493	303,840,682	3,190	6,164	19.44	116,404
Rho-6	721,120,639	648,648,015	2,983	8,896	21.75	116,036
Total/Average	9,872,843,413	8,872,034,324	81,199	6,320	19.97	113,571

Table 1. scATAC-seq metadata and mapping statistics.

method = 'Seurat', and resolution = 0.8. Markers of each cluster were identified by the 'getMarkerFeatures' function²⁶ (FDR \leq 0.05 & Log₂FC \geq 0.25), and cell type annotation was procured using well-known cell-type-specific markers obtained from the MAGIC gene activity score.

Integration of library. As our data were produced from the same batch, no significant batch effects were observed. The integration of our libraries was performed by simply merging them using ArchR²⁶. To accomplish this, we first created a project using 'ArchRProject' function. Next, we applied dimensionality reduction to our data using the 'addIterativeLSI' function with 'TileMatrix' matrix. Subsequently, we incorporated UMAP embedding of reduced dimensions object using 'addUMAP' function, allowing us to visualize different brain regions by 'plotEmbedding' function. Finally, we calculated the Pearson correlation coefficients between different libraries by 'cor' function with the gene activity matrix.

Peak calling of scATAC-seq. The model-based analysis of ChIP-seq (MACS2)²⁷ (<https://github.com/hbc-training/Intro-to-ChIPseq>) callpeak command was used to perform peak calling on the Tn5-corrected insertions (representing each end of the Tn5-corrected fragments) for each cell type. In brief, the 'addGroupCoverages' function was used to create pseudo-bulk replicates for each cell type and the 'addReproduciblePeakSet' function (parameters: groupBy = 'Clusters', pathToMA-CS = pathToMacs2, '- nomodel', genomeSize = 2,782,028,915) was used to call peaks. Next, the 'getMarkerFeatures' function was utilized to define the marker peaks for each cell type, and the 'getMarkers' function was applied to get the marker peaks.

Motif enrichment analysis. We annotated motif using the 'addMotifAnnotations' function of ArchR with the default parameters: motifSet = 'cisbp', name = 'Motif', species = 'homo sapiens', version = 2. Then, with the marker peaks, motif enrichment was performed by the 'peakAnnoEnrichment' function, and the top 18 cell-type-specific motifs were shown in the heatmap.

Assigning gene to the scATAC-seq peaks. To assign genes to scATAC-seq peaks, it required axolotl genome annotation object and gene annotation object. Firstly, we forged a BSgenome package for the axolotl genome annotation following the official tutorial (<https://github.com/Bioconductor/BSgenome>). Briefly, we forged axolotl BSgenome package by the 'forgeBSgenomeDataPkg' function with the seed file and the published axolotl genome. Then, we created axolotl genome annotation by 'createGenomeAnnotation' function (parameters: genome = BSgenome.Axolotl). Secondly, to create gene annotation object, we utilized 'createGeneAnnotation' function and 3 GRanges object (GRanges object containing gene coordinates; GRanges object containing gene exon coordinates; GRanges object containing stranded transcription start site coordinates for computing TSS enrichment scores downstream) as input. Finally, scATAC-seq peaks were assigned genes by creating an ArchR

project from the provided ArrowFiles with ‘ArchRProject’ function (parameters: geneAnnotation = axolotl gene annotation, genomeAnnotation = axolotl genome annotation). With the annotated ArchR project, scATAC-seq peaks were assigned genes by using the peak-to-gene assignment algorithm of ArchR based on their proximity to the transcription start site (TSS) of the nearest gene.

Gene ontology (GO) enrichment analysis. Gene set enrichment analysis was processed with ‘enrichGO’ function (parameters: OrgDb = org.Hs.eg.db, ont = ‘BP’) of the clusterProfiler package (version 4.4.4). The marker peaks of cell type used as input ($FDR \leq 0.05$ & $\text{Log}_2FC \geq 0.25$) were annotated based on proximity to the TSS of the nearest genes.

Data Records

We presented chromatin accessibility landscapes for different regions of the axolotl brain, providing valuable insights into the epigenetic regulation mechanisms of brain function and cell heterogeneity. Our data set consists of chromatin accessibility landscapes for 81,199 high-quality single-cells from the OB, Tel, DM, HP, and Rho regions of the axolotl brain, including 3–6 libraries of each region. Figure 1 provides an overview of the laboratory bioinformatical and analysis workflow. Table 1 displays the quality of each library. The raw fastq and fragments data have been deposited in the CNGB Nucleotide Sequence Archive (CNP0004118²⁸). The raw fastq data have also been submitted to the NCBI Sequence Read Archive (PRJNA990676²⁹). Additionally, the peak matrices and metadata has been uploaded to Figshare (<https://doi.org/10.6084/m9.figshare.22548400.v7>)³⁰. The Supplementary Table 1 (available at Figshare)³⁰ exhibited libraries correlation between the accession IDs in the CNGB or NCBI and the sample IDs.

Technical Validation

Quality control of scATAC-seq. In this study, nuclei were extracted from the OB, Tel, DM, HP, and Rho of adult axolotl and were prepared for scATAC-seq (see Methods) (Supplementary Table 1 and Fig. 1a)³⁰. The total number of raw reads was 9,872,843,413, with 8,872,034,324 reads passing QC (Table 1). The raw data were processed via the standard pipeline (Fig. 1b).

After quality control, we obtained a satisfying set of single-cell chromatin accessibility profiles for axolotl brain regions (Fig. 2a–c). We eliminated doublets by applying a filterRatio parameter of 2 in ArchR²⁶ (Fig. 2d), and subsequently examined the chromatin landscape of the 81,199 single cells to investigate cell-type-specific regulatory elements. The average unique fragments of cells remaining after quality filtering was 6,281; the average of TSS enrichment was 19.97, the average accessible peaks of each library were 113,571 (Table 1). To identify reproduced peaks between technical replicates, we computed Pearson correlation coefficients to assess the similarity between technical replicates across regions, based on gene activity (Fig. 2e). UMAP showed the heterogeneity of gene activity in different brain regions, which is probably indicative of differential functioning (Fig. 2f,g). For example, the Tel serves as the hub for sensory processing, oversees voluntary movements and activities, and is linked to sophisticated cognitive skills such as emotions, learning, and memory³¹, whereas the primary role of the HP is to manage neuroendocrine processes, with the paraventricular region housing cell bodies that project to the hypophysis and secrete hormones³².

Identification of accessible cell-type chromatin and comparison of regions. We generated scATAC-seq data across the axolotl OB, Tel, DM, HP, and Rho brain regions. Projecting cells into a low-dimensional embedding (optimized iterative LSI) and estimating gene expression by aggregating the accessibility across the regulatory region of the gene²⁶ (gene activity score), we identified 20 cell types (Fig. 3a). In brief, we identified three homeostatic EGCs, named asclEGC (*Ascl1*⁺, *Gfap*⁺, *Krt18*⁺, *Sfrp1*⁺), chdEGC³³ (*Chd7*⁺, *Olig1*⁺, *Sox10*⁺, *S100a10*⁺) and wntEGC (*Nrep*⁺, *Wnt3a*⁺). We also detected GABAergic (*Gad1*⁺, *Gad2*⁺), glutamatergic (*Slc17a*⁺) neurons, with GABAergic neurons subtypes of obIN (*Foxp2*⁺), scgnIN (*Scgn*⁺), nosIN^{34,35} (*Nos1*⁺), sstIN (*Sst*⁺), and glutamatergic subtypes, including nptxEX-1 (*Nptx1*⁺, *Tbr1*⁺) and nptxEX-2 (*Nptx2*⁺)^{6,7}. We also observed HP neuron^{36,37} (HPN, *Nefm*⁺, *Nova1*⁺) in our data. The brain-resident non-neuronal cell-types microglia (MCG, *C1qb*⁺, *Cd74*⁺), intermediate progenitor cell (IPC, *Olig1*⁺), oligodendrocytes (Oligo, *Eomes*⁺, *Olig2*⁺), vascular leptomeningeal cell (VLMC, *Lum*⁺, *Dcn*⁺), neuronal ciliary^{38,39} (NC, *Ssx2ip*⁺, *Ccdc66*⁺) and two secreting cell types of GEMs (possibly neurosecretory oxytocin or vasopressin neuron terminals⁴⁰, *Nkx2-1*⁺, *Nr5a1*⁺) and corticotropic cells (*Cor*, *Adh1*⁺, *Rab6a*⁺, *Tll2*⁺) were distinguished by their markers (Fig. 3b, Supplementary Fig. 1, Supplementary Tables 2, 3, Supplementary Tables available at Figshare)³⁰. To better understand the functions of these cell types, we performed GO analysis on the differential peaks specific to each cell type. The results of the GO analysis revealed a significant correlation between the identified cell types and their corresponding GO pathways. For example, asclEGC was enriched in the function of differentiation or proliferation pathways, while MGC was enriched in immune-related functions, such as the GO terms of cell activation involved in immune response and so on (Fig. 3d).

Chromatin accessibility analysis was used to identify the differential accessibility regions (DARs) of those 20 cell types (Fig. 3c). For example, peaks were identified specifically within the TSS region of the *Gfap* marker genes that were accessible in both asclEGC and IPC. In addition, signals around *Chd7* were specifically enriched in the chdEGC region but not in other cell types. These phenomena strongly suggest that our data are of adequate quality. To identify responding cells across regions, we investigated the varied cell-type compositions of 25 libraries by percent constituency (Fig. 3e). Interestingly, asclEGC increased gradually from the anterior to the posterior axis, and wntEGC and chdEGC were specifically enriched in Rho. Next, the diversity of cellular compositions across different regions of the axolotl brain was analysed in detail and summarized in Fig. 3f. We found asclEGC, HPN, and GEMs have more distinct gene regulation in regions with a large number of different up-regulated gene activities. As *Gfap* is a stem-cell associated protein that, like stem cells, exhibits a similar

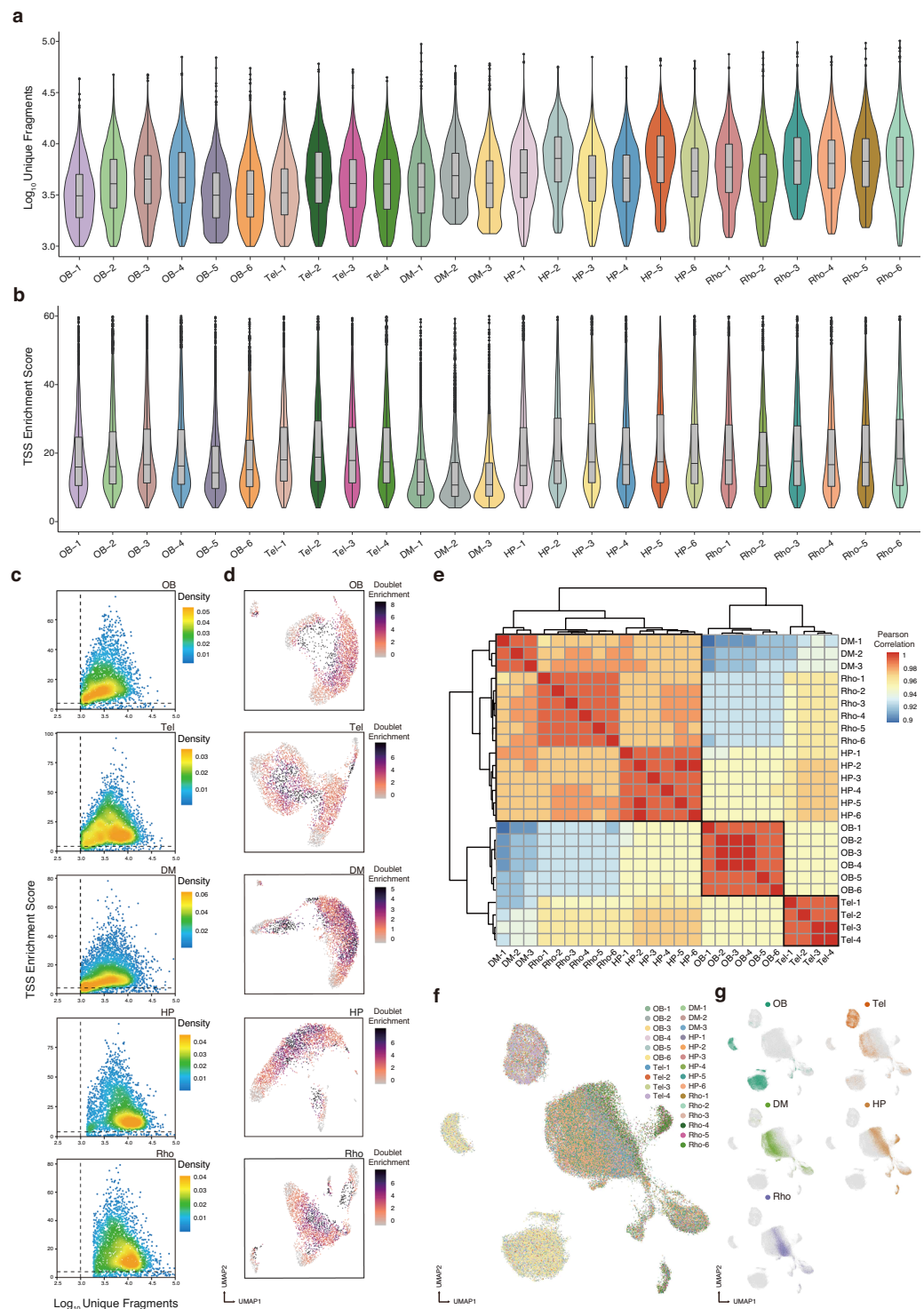


Fig. 2 Quality metrics for scATAC-seq. **(a)** Violin plot illustrating the distribution of \log_{10} (unique fragments) for each library. **(b)** Violin plot of the transcription start site (TSS) enrichment score for each library. **(c)** Quality control (QC) filtering plots of the TSS enrichment fractions vs unique cell fragments for each cell from the olfactory bulb (OB), telencephalon (Tel), diencephalon and mesencephalon (DM), hypothalamus and pituitary (HP), and rhombencephalon (Rho). **(d)** UMAP of scATAC-seq data showing simulated doublet enrichment over expectation. **(e)** Heatmap clustering of correlation coefficients across 25 libraries of scATAC-seq profiles. **(f)** A UMAP visualization of the 25 libraries, each color of which represents a per library interaction of the five brain regions. **(g)** A UMAP visualization representing the five brain regions; Cells are colored by brain region in the plots.

response to injury^{31,32,41,42}, we analysed asclEGC (with *Gfap* chromatin accessible) and found *Egr1* and *Meis1* are accessible in OB; *Mef2c*, *Slc1a6*, *Wnt5b*, *Lhx2* in Tel; *Col* and the *Nd* family in DM; *Chd6*, *Nkx2-3* in HP; *Pax7*

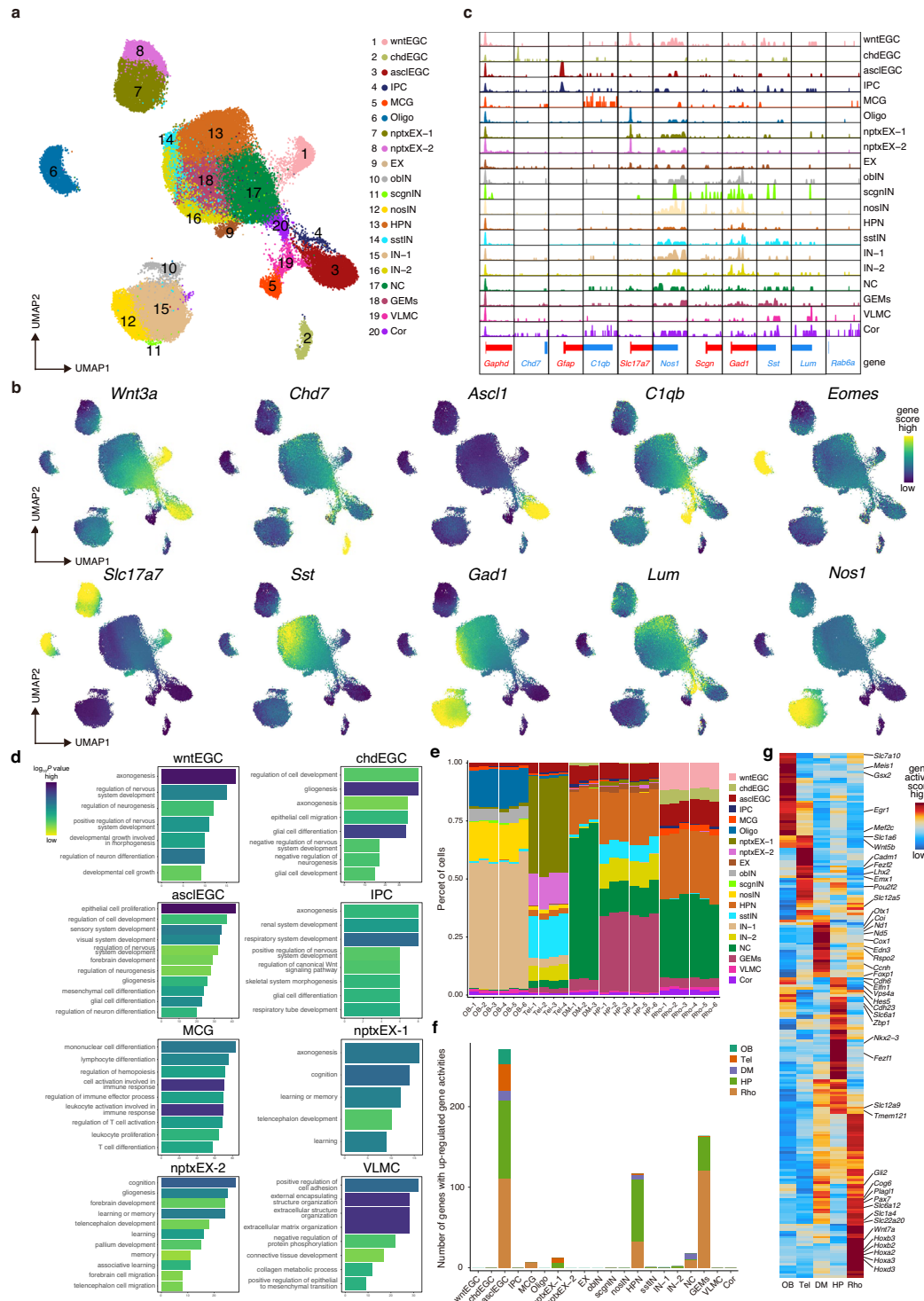


Fig. 3 Single-cell chromatin accessibility and diversity between assessed regions of the axolotl brain. **(a)** UMAP of 20 clusters as identified by scATAC-seq. Cells coloration indicates annotation. **(b)** UMAP visualization of cell-type-specific gene activity score. **(c)** Genome browser view of the aggregated scATAC-seq chromatin accessibility profiles of the housekeeping gene (*Gapdh*) and the cell-type-specific gene loci. **(d)** Bar plot exhibiting the representative GO enrichment pathways of cell types. **(e)** Histogram displaying the distribution of cell types across various brain regions in different libraries. **(f)** Histogram of the number of up-regulated genes of the same cell type in different axolotl brain regions. **(g)** Heatmap of *asclEGC*-specific gene activity scores in the five brain regions.

and the Hox family in Rho, respectively (Fig. 3g). Therefore, this analysis strongly suggests that the scATAC-seq profiles of axolotl brain regions can effectively and accurately identify the accessible chromatin regions within the axolotl genome and provides a direct theoretical basis reference that can be exploited by future studies.

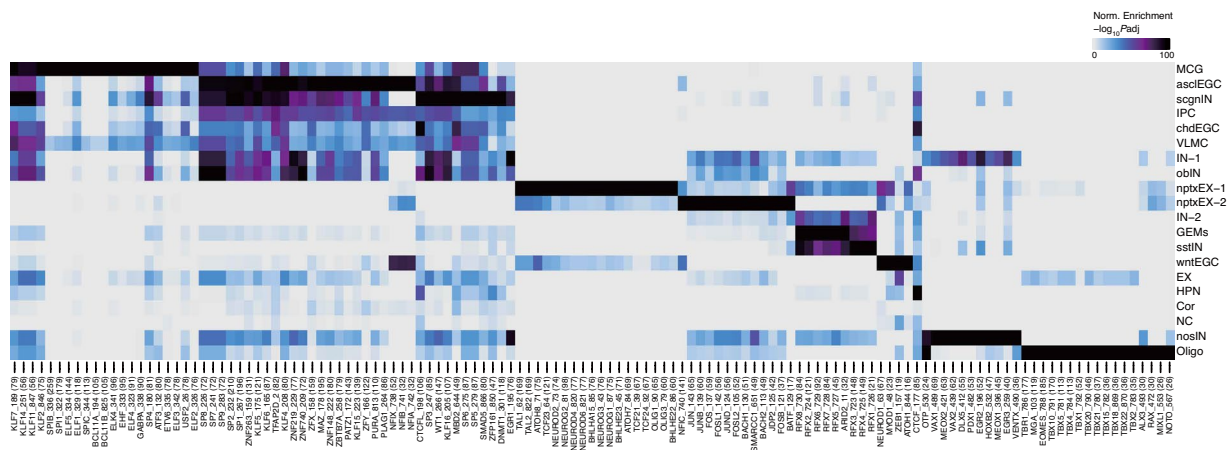


Fig. 4 Identification of cell-type-specific chromatin transcription factors (TFs). Heat map clusters indicate cell-type-specific TFs.

Inferring cell type-specific transcription factors. To gain a better understanding of the regulatory mechanisms governing the chromatin landscape, we implemented peak calling to generate a set of 121,697 cell-type differential peaks based on pseudobulk chromatin accessibility. Then, we used ArchR²⁶ to identify transcription factors that exhibit strong correlation with cell-type-specific open chromatin and compared these results with previous studies (Fig. 4). Conformably, the master regulators have been shown to be closely related to specific functions of cell types. For example, MCG is enriched for motifs of the BCL11 family (BCL11A⁴³, BCL11B⁴⁴). This was supported by the previous study of BCL11A deletion causes apoptosis in immature B cells and common lymphoid progenitors, and also results in delayed or deficient lymphoid development of hematopoietic stem cells to B, T, and NK cells in adult mice⁴³. Besides, BCL11B prevents autoimmune disorders by controlling multiple regulatory T cell gene expression programs⁴⁴. EGCs are enriched for motifs that distinguish progenitor cells from other cell types, in the agreement of the characterization of proliferation in EGCs^{31,32,41,42}. In brief, we observed strong enrichment of PLAG1 in asclEGC, which is consistent with regulation of progenitor cell proliferation and neurogenesis during telencephalic development of mice⁴⁵. We also found NEUROD1 is enriched in wntEGC in our dataset. NEUROD1 promotes transit-amplifying progenitors that contribute to the generation of the majority of the excitatory neurons of the neuroepithelium of the dorsal telencephalon in early development^{46–48}. Moreover, chdEGC is enriched with CTCF, which regulates functional neural development and neuronal diversity⁴⁹. Other well-known cell-type-specific motifs such as the Oligo (TBR1⁵⁰, EOMES⁵¹), nptxEX-1 (NEUROG2, NEUROD2, NEUROD4, NEUROD6⁵², BHLHE22), nptxEX-2 (BACH1, BACH2, the JUN⁵³, NFIC, and FOSL families), nosIN (DLX family), and GEMs (RFX family) were enriched in our data. Taken together, cell-type-specific TFs that have been reported in previous studies are also present in our data, exhibiting similar patterns of enrichment, which indicates the accuracy of our cell type identification and the high-quality of our data. Therefore, we provide a resource-rich and high-quality data of chromatin landscape of the axolotl brain.

Usage Notes

The scATAC-seq data processing pipeline, including read mapping and peak calling, were run on the Linux operating system. All R source code used for downstream data analyses and visualization are provided online (<https://doi.org/10.6084/m9.figshare.22548400.v7>)³⁰.

Code availability

The R code used to identify cell subclusters and profile cell type-specific chromatin accessible regions of the axolotl brain are available online (<https://doi.org/10.6084/m9.figshare.22548400.v7>)³⁰.

Received: 18 May 2023; Accepted: 1 September 2023;

Published online: 14 September 2023

References

1. Brockes, J. P. & Kumar, A. Appendage regeneration in adult vertebrates and implications for regenerative medicine. *Science* **310**, 1919–1923, <https://doi.org/10.1126/science.1115200> (2005).
2. Simon, A. & Tanaka, E. M. Limb regeneration. *Wiley Interdiscip Rev Dev Biol* **2**, 291–300, <https://doi.org/10.1002/wdev.73> (2013).
3. Joven, A., Elewa, A. & Simon, A. Model systems for regeneration: salamanders. *Development* **146**, <https://doi.org/10.1242/dev.167700> (2019).
4. McHedlishvili, L. *et al.* Reconstitution of the central and peripheral nervous system during salamander tail regeneration. *Proc Natl Acad Sci USA* **109**, E2258–E2266, <https://doi.org/10.1073/pnas.1116738109> (2012).
5. Yun, M. H. Changes in Regenerative Capacity through Lifespan. *Int J Mol Sci* **16**, 25392–25432, <https://doi.org/10.3390/ijms161025392> (2015).
6. Wei, X. *et al.* Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444, <https://doi.org/10.1126/science.abp9444> (2022).

7. Lust, K. *et al.* Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. *Science* **377**, eabp9262, <https://doi.org/10.1126/science.abp9262> (2022).
8. Nowoshilow, S. *et al.* The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55, <https://doi.org/10.1038/nature25458> (2018).
9. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213–1218, <https://doi.org/10.1038/nmeth.2688> (2013).
10. Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55, <https://doi.org/10.1016/j.cell.2013.09.011> (2013).
11. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914, <https://doi.org/10.1126/science.aab1601> (2015).
12. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490, <https://doi.org/10.1038/nature14590> (2015).
13. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925–936, <https://doi.org/10.1038/s41587-019-0206-z> (2019).
14. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70–80, <https://doi.org/10.1038/nbt.4038> (2018).
15. Cusanovich, D. A. *et al.* A Single-Cell Atlas of *In Vivo* Mammalian Chromatin Accessibility. *Cell* **174**, 1309–1324 e1318, <https://doi.org/10.1016/j.cell.2018.06.052> (2018).
16. Preissl, S. *et al.* Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci* **21**, 432–439, <https://doi.org/10.1038/s41593-018-0079-3> (2018).
17. Graybuck, L. T. *et al.* Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron* **109**, 1449–1464 e1413, <https://doi.org/10.1016/j.neuron.2021.03.011> (2021).
18. Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat Commun* **9**, 5345, <https://doi.org/10.1038/s41467-018-07771-0> (2018).
19. Yu, Y. *et al.* Single-Nucleus Chromatin Accessibility Landscape Reveals Diversity in Regulatory Regions Across Distinct Adult Rat Cortex. *Front Mol Neurosci* **14**, 651355, <https://doi.org/10.3389/fnmol.2021.651355> (2021).
20. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962, <https://doi.org/10.1038/nmeth.4396> (2017).
21. Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648, <https://doi.org/10.1371/journal.pone.0209648> (2018).
22. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, 108189, <https://doi.org/10.1016/j.celrep.2020.108189> (2020).
23. Huang, J. *et al.* A reference human genome dataset of the BGISEQ-500 sequencer. *Gigascience* **6**, 1–9, <https://doi.org/10.1093/gigascience/gix024> (2017).
24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324> (2009).
25. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* **37**, 916–924, <https://doi.org/10.1038/s41587-019-0147-6> (2019).
26. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411, <https://doi.org/10.1038/s41588-021-00790-6> (2021).
27. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, <https://doi.org/10.1186/gb-2008-9-9-r137> (2008).
28. CNGB Nucleotide Sequence Archive. <https://db.cngb.org/search/project/CNP0004118/> (2023).
29. NCBI Sequence Read Archive. <https://identifiers.org/ncbi/insdc.sra:SRP448692> (2023).
30. Feng, W. *et al.* A scATAC-seq atlas of chromatin accessibility in axolotl brain regions. *Figshare* <https://doi.org/10.6084/m9.figshare.22548400.v7> (2023).
31. Joven, A. & Simon, A. Homeostatic and regenerative neurogenesis in salamanders. *Prog Neurobiol* **170**, 81–98, <https://doi.org/10.1016/j.pneurobio.2018.04.006> (2018).
32. Joven, A. *et al.* Cellular basis of brain maturation and acquisition of complex behaviors in salamanders. *Development* **145**, <https://doi.org/10.1242/dev.160051> (2018).
33. Ohta, S. *et al.* CHD7 promotes proliferation of neural stem cells mediated by MIF. *Mol Brain* **9**, 96, <https://doi.org/10.1186/s13041-016-0275-6> (2016).
34. Fuentealba, P. *et al.* Ivy cells: a population of nitric-oxide-producing, slow-spiking GABAergic neurons and their involvement in hippocampal network activity. *Neuron* **57**, 917–929, <https://doi.org/10.1016/j.neuron.2008.01.034> (2008).
35. Jinno, S. & Kosaka, T. Immunocytochemical characterization of hippocamposeptal projecting GABAergic nonprincipal neurons in the mouse brain: a retrograde labeling study. *Brain Res* **945**, 219–231, [https://doi.org/10.1016/s0006-8993\(02\)02804-4](https://doi.org/10.1016/s0006-8993(02)02804-4) (2002).
36. Sainio, M. T. *et al.* Absence of NEFL in patient-specific neurons in early-onset Charcot-Marie-Tooth neuropathy. *Neurol Genet* **4**, e244, <https://doi.org/10.1212/NXG.0000000000000244> (2018).
37. Tajima, Y. *et al.* NOVA1 acts on Impact to regulate hypothalamic function and translation in inhibitory neurons. *Cell Rep* **42**, 112050, <https://doi.org/10.1016/j.celrep.2023.112050> (2023).
38. Klinger, M. *et al.* The novel centriolar satellite protein SSX2IP targets Cep290 to the ciliary transition zone. *Mol Biol Cell* **25**, 495–507, <https://doi.org/10.1091/mbc.E13-09-0526> (2014).
39. Conkar, D., Bayraktar, H. & Firat-Karalar, E. N. Centrosomal and ciliary targeting of CCDC66 requires cooperative action of centriolar satellites, microtubules and molecular motors. *Sci Rep* **9**, 14250, <https://doi.org/10.1038/s41598-019-50530-4> (2019).
40. Fletcher, P. A. *et al.* Cell Type- and Sex-Dependent Transcriptome Profiles of Rat Anterior Pituitary Cells. *Front Endocrinol (Lausanne)* **10**, 623, <https://doi.org/10.3389/fendo.2019.00623> (2019).
41. Maden, M., Manwell, L. A. & Ormerod, B. K. Proliferation zones in the axolotl brain and regeneration of the telencephalon. *Neural Dev* **8**, 1, <https://doi.org/10.1186/1749-8104-8-1> (2013).
42. Kirkham, M., Hameed, L. S., Berg, D. A., Wang, H. & Simon, A. Progenitor cell dynamics in the Newt Telencephalon during homeostasis and neuronal regeneration. *Stem Cell Reports* **2**, 507–519, <https://doi.org/10.1016/j.stemcr.2014.01.018> (2014).
43. Yu, Y. *et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med* **209**, 2467–2483, <https://doi.org/10.1084/jem.20121846> (2012).
44. Hasan, S. N. *et al.* Bcl11b prevents catastrophic autoimmunity by controlling multiple aspects of a regulatory T cell gene expression program. *Sci Adv* **5**, eaaw0706, <https://doi.org/10.1126/sciadv.aaw0706> (2019).
45. Adnani, L. *et al.* Plag1 and Plag2 have overlapping and distinct functions in telencephalic development. *Biol Open* **7**, <https://doi.org/10.1242/bio.038661> (2018).
46. Liu, M. *et al.* Loss of BETA2/NeuroD leads to malformation of the dentate gyrus and epilepsy. *Proc Natl Acad Sci USA* **97**, 865–870, <https://doi.org/10.1073/pnas.97.2.865> (2000).
47. Tutukova, S., Tarabykin, V. & Hernandez-Miranda, L. R. The Role of NeuroD Genes in Brain Development, Function, and Disease. *Front Mol Neurosci* **14**, 662774, <https://doi.org/10.3389/fnmol.2021.662774> (2021).

48. Schwab, M. H. *et al.* Neuronal basic helix-loop-helix proteins (NEX, neuroD, NDRF): spatiotemporal expression and targeted disruption of the NEX gene in transgenic mice. *J Neurosci* **18**, 1408–1418, <https://doi.org/10.1523/JNEUROSCI.18-04-01408.1998> (1998).
49. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. & Yagi, T. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep* **2**, 345–357, <https://doi.org/10.1016/j.celrep.2012.06.014> (2012).
50. Chuang, H. C., Huang, T. N. & Hsueh, Y. P. Neuronal excitation upregulates Tbr1, a high-confidence risk gene of autism, mediating Grin2b expression in the adult brain. *Front Cell Neurosci* **8**, 280, <https://doi.org/10.3389/fncel.2014.00280> (2014).
51. Yao, Z. *et al.* A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development. *Cell Stem Cell* **20**, 120–134, <https://doi.org/10.1016/j.stem.2016.09.011> (2017).
52. Khan, S. *et al.* Survival of a Novel Subset of Midbrain Dopaminergic Neurons Projecting to the Lateral Septum Is Dependent on NeuroD Proteins. *J Neurosci* **37**, 2305–2316, <https://doi.org/10.1523/JNEUROSCI.2414-16.2016> (2017).
53. Schlingensiefen, K. H. *et al.* The role of Jun transcription factor expression and phosphorylation in neuronal differentiation, neuronal cell death, and plastic adaptations *in vivo*. *Cell Mol Neurobiol* **14**, 487–505, <https://doi.org/10.1007/BF02088833> (1994).

Acknowledgements

We thank the China National GeneBank for offering the services of sequencing for this project. This Project was supported by China Postdoctoral Science Foundation (Grant No. 2022M720096, 2022M722196), National Natural Science Foundation of China (31970782, 92268114, 32070819), National Key R&D Program of China (2021YFA0805000, 2019YFE0106700), and High-Level Hospital Construction Project of GDPH (DFJHBF202103, KJ012021012), Shenzhen Science and Technology Program (Grant No. RCJC20221008092804002).

Author contributions

W.F., X.W. conceived the idea. S.L. collected the sample with the guidance of J.F. Q.D. generated the data with assistance of Y.Y., C.L. W.F., S.L. analysed the data and W.F. wrote the manuscript. S.L., Z.Y. draw a flowchart. X.W., J.F. oversaw the study and revised the manuscript. A final manuscript review and approval was completed by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02533-0>.

Correspondence and requests for materials should be addressed to C.L., J.-F.F. or X.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023