Nucleic Acids Research, 2025, **53**, D1055–D1065
https://doi.org/10.1093/nar/gkae911
Advance access publication date: 18 October 2024
Database issue

MAPbrain: a multi-omics atlas of the primate brain

Liangchen Zhuo[1,4,†], Mengdi Wang[1,4,†], Tingrui Song[1,†], Suijuan Zhong[2], Bo Zeng[3], Zeyuan Liu[3], Xin Zhou[2], Wei Wang[1], Qian Wu[2,*], Shunmin He[1,*] and Xiaoqun Wang[1,2,3,*]

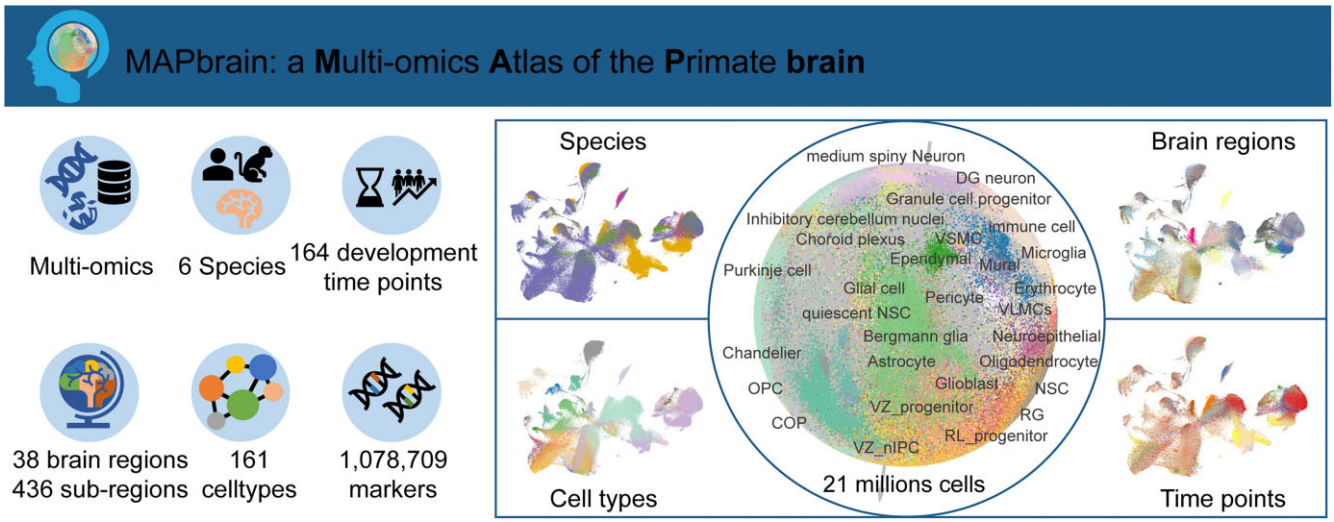
[1]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[2]State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, New Cornerstone Science Laboratory, Beijing Normal University, Beijing 100875, China
[3]Changping Laboratory, Beijing 102206, China
[4]University of Chinese Academy of Sciences, Beijing 100049, China

[*]To whom correspondence should be addressed. Tel: +86 10 64887994; Fax: +86 10 64887994; Email: xiaoqunwang@ibp.ac.cn
Correspondence may also be addressed to Shunmin He. Email: heshunmin@ibp.ac.cn
Correspondence may also be addressed to Qian Wu. Email: qianwu@bnu.edu.cn
[†]The first three authors should be regarded as Joint First Authors.

Abstract

The brain is the central hub of the entire nervous system. Its development is a lifelong process guided by a genetic blueprint. Understanding how genes influence brain development is critical for deciphering the formation of human cognitive functions and the underlying mechanisms of neurological disorders. Recent advances in multi-omics techniques have now made it possible to explore these aspects comprehensively. However, integrating and analyzing extensive multi-omics data presents significant challenges. Here, we introduced MAPbrain (http://bigdata.ibp.ac.cn/mapBRAIN/), a multi-omics atlas of the primate brain. This repository integrates and normalizes both our own lab's published data and publicly available multi-omics data, encompassing 21 million brain cells from 38 key brain regions and 436 sub-regions across embryonic and adult stages, with 164 time points in humans and non-human primates. MAPbrain offers a unique, robust, and interactive platform that includes transcriptomics, epigenomics, and spatial transcriptomics data, facilitating a comprehensive exploration of brain development. The platform enables the exploration of cell type- and time point-specific markers, gene expression comparison between brain regions and species, joint analyses across transcriptome and epigenome, and navigation of cell types across species, brain regions, and development stages. Additionally, MAPbrain provides an online integration module for users to navigate and analyze their own data within the platform.

Graphical abstract



Introduction

The human brain is the origin of thoughts, emotions, perceptions, and memories, functioning as the control center for the entire nervous system. From its earliest developmental stages to the formation of complex functions, the brain undergoes substantial transformations in gene expression, chromatin

Received: August 15, 2024. **Revised:** September 26, 2024. **Editorial Decision:** October 1, 2024. **Accepted:** October 7, 2024

structure, cellular diversity and circuit formation (1). Understanding the mechanisms that govern this process is crucial for gaining deeper insights into the brain, especially in humans and non-human primates, which have evolved advanced and complex functions (2). Recent technological advancements have brought researchers closer to a comprehensive understanding of the brain, extending from molecular and cellular levels to circuits and behavior. Innovations such as single-cell RNA sequencing (3), spatial transcriptomics (4) and single-cell epigenomics (5) are opening promising new avenues for research. These technologies enable the delineation of cell diversity, measurement of chromatin accessibility at the single-cell level and mapping transcript spatial locations, providing unprecedented opportunities to analyze the complex nature of mammalian brains, particularly in humans and non-human primates.

With the rapid growth of high-throughput sequencing technologies and the increasing interest in neuroscience, publicly available brain sequencing datasets have expanded significantly. Recently established databases offer researchers easy access to large-scale brain genomics resources. For example, the single-omics databases like scBrainMap provides single-cell transcriptomic data across 124 brain regions, covering 128 developmental periods of 14 species (6); STAB2 showcases single-cell transcriptomics data across 101 brain regions, covering 45 developmental stages of humans and mice (7) and ZEBRA presents single cell transcriptomics data across 39 brain regions of human and mouse during aging and disease (8). Additionally, StomicsDB explores spatial transcriptomics data of 17 species (9). Furthermore, the databases such as NeMO (10), TEDD (11) and SCAN (12) collect multi-omics data across various brain regions and provide frameworks for showcasing each dataset individually.

To our knowledge, there are no databases dedicated to the cohesive integration and interpretation of multi-omics data, specifically focused on human and non-human brain development across prenatal, postnatal, adult and aging stages. To address this gap, we have developed the MAPbrain database. This repository gathers both our own lab's published and publicly available multi-omics data from 38 key brain regions and 436 subregions across embryonic and adult stages, with 164 time points in humans and non-human primates. MAPbrain offers a unique and robust platform that includes transcriptomics, epigenomics, and spatial transcriptomics data, facilitating a comprehensive exploration of brain development. MAPbrain integrates the most recent published data, providing a larger population of cells and a more extensive and finely classified array of brain regions and developmental stages. Our database uniquely focuses on human and non-human primate brain development and offers analysis modules that enable interactive data visualization, differentially expressed gene searches, cross-species comparisons and inter-omics joint analysis. Furthermore, the interactive visualization of the integrated datasets facilitates precise navigation of cell types and provides anchors for accurate cell annotation of user-uploaded datasets. Through extensive analysis of publicly available brain datasets, MAPbrain serves as a versatile online platform for exploring brain development, enabling users to unravel the spatiotemporal complexities of the brain at regional, cellular and molecular levels.

Materials and methods

Data collection

Data were collected by searching PubMed and other public databases, such as Gene Expression Omnibus (GEO) (13), the UCSC Cell Browser (14), NeMO (10), Descartes (15,16), EMBL-EBI (https://www.ebi.ac.uk/) (17) and Allen Brain Map (https://portal.brain-map.org/) (18), using the following keywords: (i) scRNA, scRNA-seq, single-cell RNA seq, single-cell transcriptome, single-nucleus, ATAC-seq, scATAC-seq, epigenome and spatial transcriptome, in combination with (ii) human brain, monkey brain, primate brain, and specific brain regions. Importantly, datasets generated from cell lines, brain organoids, or disease samples, as well as datasets without cell type annotation or supporting information for cell annotation were explicitly excluded. Metadata were manually annotated, unified, and verified to ensure consistency. We standardized cell-type, brain region and age information, mapping similar cell types or brain regions onto each other. We also provided a categorical scale for age information across prenatal, postnatal, and adult brain stages, standardizing postnatal ages to years. For cells without existing cell type annotation, we assigned cell types according to the expression of marker genes and integrated them with the annotated datasets to verify cell type information. Annotated datasets are available for download from our website, with datasets having access restrictions excluded from the downloadable files.

Preprocessing and integration

We collected gene expression matrices from the available datasets, all of which have been aligned to the reference genome. As noted in the original studies, the data were aligned to the following genome assemblies: hg19 or hg38 (Human), *Macaca fascicularis* genome version 5.0 (*Macaca fascicularis*), rheMac10 or Mmul10 (*Macaca mulatta*), Clint_PTRv2 or panTro6 (Chimpanzee), calJac3 or CJ1700 (Marmoset), Susie3 (Gorilla). For preprocessing the single cell RNA-seq data, we used the Seurat package (v5.0.1) (19), setting parameters to $150 < \text{nFeatures} < 8000$ and mt-ratio $< 10\%$ for cell quality control. To standardize feature names, we used biomaRt (20) to convert ENSEMBL IDs (21) to gene symbols. For datasets containing >15 000 cells, we randomly sampled each dataset down to approximately 15 000 cells for downstream integration. Within each dataset, we utilized 'NormalizeData' and 'FindVariableFeatures' functions to normalize gene expression matrices and identify highly variable genes. The PCA embeddings, calculated using the 'RunPCA' function, were then input into the 'RunHarmony' function from the harmony package (22) to integrate single-cell data across different datasets. The cell type navigation for all the datasets was calculated using the 'RunUMAP' function with harmony embeddings as input. To address the issue of cell crowding in the UMAP latent space and achieve a more readily interpretable representation, we embedded single cell data in a low-dimensional hype-spherical using scPhere (23). The gene expression matrix, containing only variable genes, was imported into the function 'SCPHERE' function and Trainer for model building and training. After training, we mapped the cells into the hyperspherical space using the 'model.encode' function (Supplementary Figure S1). For cross-species data integration, we utilized 1:1 orthologs with the BioMart interface, ensuring

that each gene from one species has a direct counterpart in the other species being studied.

Processing of spatial transcriptomics data

Space Ranger (version 1.0.0) software from 10X Genomics was utilized for processing, alignment and counting barcodes/UMIs against the human hg38 reference genome for each spot on the Visum spatial transcriptomic array. The resulting raw UMI count matrix, images, spot-image coordinates, and scale factors were imported into Seurat. The 'SCTransform' function was used to normalize expression values for total UMI counts per cell, followed by dimensionality reduction and clustering using the 'RunPCA', 'FindNeighbors', 'FindClusters' and 'RunUMAP' functions. The 'SpatialDimPlot' function was employed to visualize spatial gene expression from the 10X Visium assay overlaid on tissue images. The gene expression count matrices and the mapped spot coordinates onto the tissue images were exported for spatial visualization.

Processing of ATAC-seq data

For single-cell ATAC-seq data with available peak files, we converted the peak files in BED format into the bigwig file format using the bedGraphToBigWig tool (24) for visualization in the Genome Browser. For those ATAC-seq data without available peak BED files, we utilized an ATAC-seq analyzing pipeline for data processing. Specifically, reads filtering, mapping (mapping to the human reference genome hg19) and transposase cut site identification were performed with the Cell Ranger ATAC Software (v1.2.0). The generated fragment files were then processed with the R package SnapATAC (v 1.0.0) and snap file with a bin size of 2.5 kb was generated. The resulting cell-by-bin count matrix was first converted into a binary matrix by replacing the non-zero items with 1. Then, cells that were outliers for any of the following criterion were considered to be of low quality and discarded: (i) $\log_{10}$-transformed number of unique fragments between 2 and 5 (log10(UMI)); (ii) fragments in promoter ratio between 0.15 and 0.6; (iii) the number of fragments in peaks between 1000 and 40 000; (iv) the fraction of fragments in peaks >30% and (v) transcriptional start site (TSS) enrichment score >1.8. Bins overlapping with the ENCODE blacklist were excluded. The filtered binarized cell-by-bin matrix was the loaded into R package Signac (v0.2.5) for subsequent dimensionality reduction analysis and cell type annotations. For each cluster, peak calling was performed with fragments from cells in the cluster by using the command MACS2 and the parameters '–nomodel –shift 100 –extsize 200 –qval 5e-2 -B -f BED –nolambda –keep-dup all –call-summits' callpeak in MACS2 software (v 2. 2. 7. 1) (25). Peaks from each cluster were then merged and a binarized cell-by-peak matrix was constructed by converting non-zero counts to 1. The resulting peak files in BED format were then converted into the bigwig file format using the bedGraphToBigWig tool.

Differentially expressed gene (DEG) identification

Differentially expressed genes among cell types were identified using the 'FindAllMarkers' function in Seurat (v5.0.1). Genes with an adjusted *P*-value <0.05 and log fold change >0.25 were considered significantly elevated in specific cell types. The genes were then arranged by log fold change in descending order, and the top 10 markers were selected for DEG heatmap visualization.

Cell mapping

To establish a precise and effective framework for predicting cell types in user-uploaded datasets, using cell annotation data from the MAPbrain datasets as a reference, we implemented the following pipeline. First, we merged the user-uploaded dataset with the MAPbrain dataset, normalized gene expression with the 'NormalizeData' function and selected variable genes with the 'FindVariableFeatures' function. Next, we conducted PCA-based dimensionality reduction with the 'RunPCA' function. The resulted pca-embeddings were then imported into the 'RunHarmony' function from the harmony package (22) for dataset integration. UMAP-based dimensionality reduction was subsequently performed using the 'RunUMAP' function with the harmony embeddings as input. To infer cell identities of the user-uploaded dataset, we conducted k-nearest neighbors analysis (knn) in the UMAP latent space using the R function knn, with parameters k and l. For each query cell, we counted the identities of its k closest reference cells. If at least $(k - l)$ reference cells shared the same cell identities, the query cell was assigned to that cell identity. The cell mapping results were then exported as a .csv file and sent to the user via e-mail.

Website implementation

MAPbrain is a user-friendly website developed using JavaScript, PHP, and the ThinkPHP5.0 (https://github.com/kongrp/thinkphp5.0) framework. It supports browsing, querying, and various analysis functionalities. The site integrates the visualization tool Plotly (https://plotly.com/) and the genome browser plugin JBrowse 2 (26), providing users with an intuitive data exploration experience. Users can easily explore available data, perform queries, and access additional analysis tools without the need for username and password authentication.

Results

Database overview

The database consolidates a wide array of multi-omics datasets, including single-cell transcriptomics, epigenomics and spatial transcriptomics, from human and five non-human primate species including Macaca fascicularis, Macaca mulatta, Chimpanzee, Gorilla and Marmoset. The collected studies exhibit heterogeneity in several aspects: collection methods vary, annotations are inconsistent, and sample regions differ. After data standardization and integration of over 60 public datasets (Supplementary Table S1) (15,18,27–73), we assembled a corpus of 13 712 470 cells from the human brain and 8 028 186 cells from non-human primate brains, spanning prenatal, postnatal, and adult developmental stages with 164 time points. The human brain samples cover 36 key brain regions and 371 subregions of human, while the non-human primate brain samples cover 22 key brain regions and 82 subregions (Figure 1, Supplementary Table S2). Notably, MAPbrain encompasses most major brain regions and developmental time points from both human and non-human primate brains. Using the Harmony integration method, we mapped different groups of cells from diverse datasets into a shared embedding, effec-

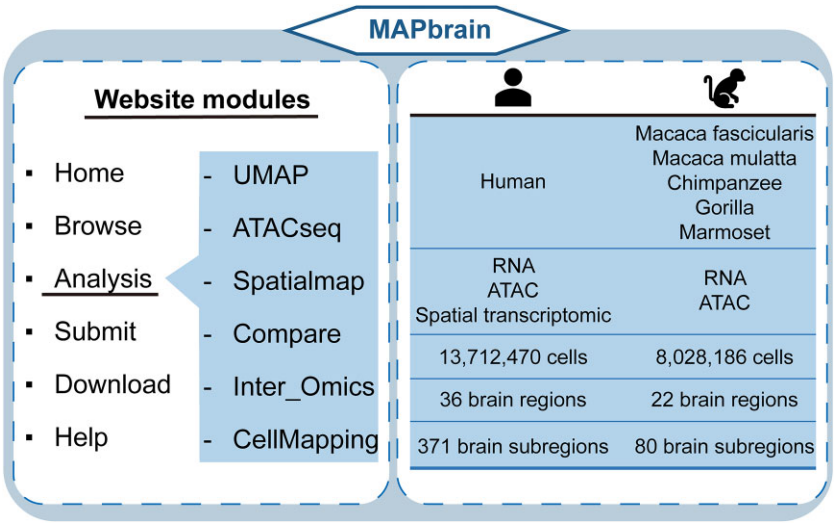


Figure 1. The MAPbrain database utilizes an open-data approach, offering unrestricted access to over 21 million high-quality cells. These cells are derived from 436 sub-regions across 38 brain regions and 164 developmental time points, ensuring extensive coverage and utility for a wide range of research applications.

tively correcting for batch effects across the diverse datasets. This mapping reflects the levels of gene expression similarity among cell types by their proximity on the map.

The MAPbrian website features 6 main sections: 'Home' section, 'Browse' section, 'Analysis' section, 'Submit' section, 'Download' section and 'Help' section. The 'Analysis' section contains six analysis modules: 'Analysis-UMAP', 'Analysis-Compare', 'Analysis-ATACseq', 'Analysis-Spatialmap', 'Analysis-Inter_Omics' and 'Analysis-CellMapping', facilitating interactive exploration and analysis at multi-omics levels and integrating a self-constructed pipeline for cell annotation (Figure 1). The web application and search engines of MAPbrain are installed on a high-performance Linux server, ensuring unrestricted access to the MAPbrain database and the analysis modules for interactive data exploration and customized analyses. Overall, MAPbrain serves as a user-friendly database offering an extensive cross-study summary of the human and non-human primate brains in the context of brain development.

Search engine and Quick start of MAPbrain

On the 'Home' page of the website, we introduce the main content of the data collection and the functionality of the database, along with a cell type navigation map. The interactive cell type navigation map enables precise cell positioning and annotation on the integrated sphere map. Users can also search for information related to brain regions, cell types, and genes of interest through the query box in the 'Home' section (Figure 2A). Additionally, the Home page offers quick link to specific sections, allowing users to access them by clicking on the categorized icons or entries in the data information box below quick search section.

Browse significant differences cross data

The 'Browse' section provides an overview of the data collection, where entries are classified into a hierarchical structure, starting from species to brain regions, and organized to present the statistical parameters for each calculated marker within specific cell types, brain regions and developmental time points. Overall, an extensive collection of 1 078 709 markers for 161 cell types spanning 164 developmental time points across different brain regions were provided. Moreover, each marker was hyperlinked to the analysis module for detailed visualization of the selected gene (Figure 2B).

Interactive exploration of MAPbrain transcriptomics data

Analysis-UMAP modules offers interactive visualization of single cell transcriptomics data (Figure 2C). On this page, user can access the gene expression profiles of specific genes by simply searching the gene symbol on the 'Gene expression' search box on the UMAP-analysis module (Figure 2D). Furthermore, the Analysis-UMAP module includes statistical boxplots of each dataset across six species, based on metadata such as cell-type, time, and sub-region. Another sub-module within the Analysis-UMAP provides heatmaps that show the mean expression distribution of the top 10 differentially expressed markers, ranked by fold change values, and grouped by diverse cell types from the selected data set (Figure 2E). In addition, all plots are interactive, enabling zooming, downloading, and toggling of the presentation of categories, and user can easily access cell positioning, cell type, brain region, developmental time point information by hovering the mouse over data points (Figure 2F).

Geometrical visualization of MAPbrain spatial transcriptomics data

For spatial transcriptomic data, the Analysis-Spatialmap module displays the gene expression on interactive images of different tissue slices. Upon selecting species, tissue, and submitting a gene symbol in the 'Genes/Ensembl ID' search box, the spatial gene expression profiles of the corresponding genes are projected on the spatial image (Figure 2G).

Comparative investigation of MAPbrain epigenomics data

For snATAC-seq data, the Analysis-ATACseq module displays chromatin accessibility loci across various cell groups

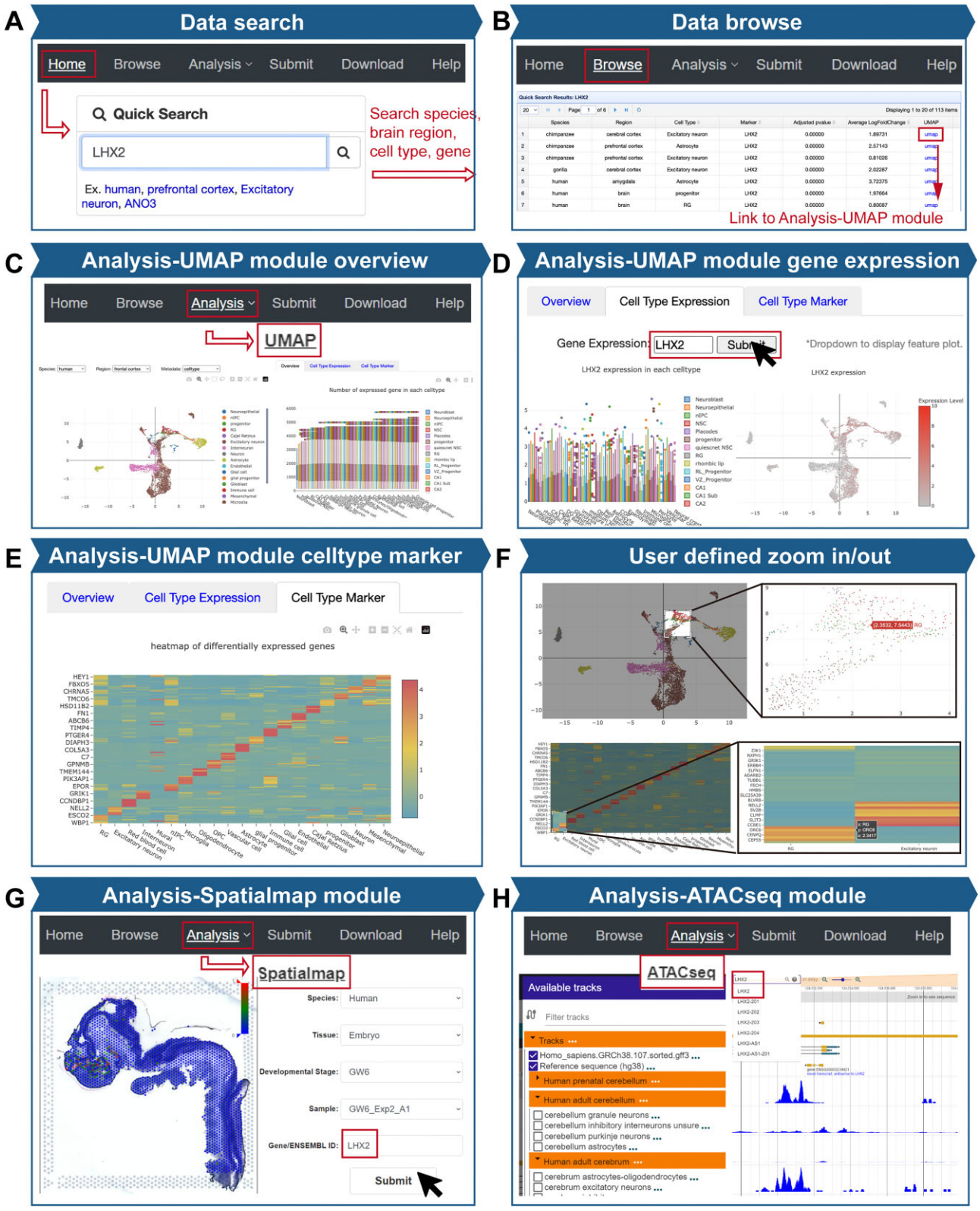


Figure 2. Demonstration of the search logic and Analysis-UMAP, Analysis-Spatialmap and Analysis-ATAC-seq modules. (**A–D**) Procedure for querying information and expression profiles of genes in single-cell transcriptomics data. (**E**) The Analysis-UMAP module allows access to cell type markers. (**F**) Illustration of the interactive usage in the Analysis-UMAP module. (**G**) Procedure for querying spatial gene expression profiles in spatial transcriptomics data. (**H**) Procedure for querying chromatin accessibility in snATAC-seq data

and developmental stages, supporting comparative views of multiple datasets by selecting them in the drop-down menu (Figure 2H).

Overall, these modules facilitate effortless integrative exploration of specific genes at multi-omics levels.

Comparative analysis across conditions

Cross-condition analysis is also a major focus for researches, as it is critical for understanding how different genetic programs affect biological systems or processes. MAPbrain provides an Analysis-Compare module, offering cross-species and cross-region pair-wise comparison analysis in gene expression, cell type composition, and developmental time points. Users can go to the Analysis-Compare module by clicking 'Compare' in the drop-down menu. By selecting species and tissue types and submitting gene symbol in the 'Gene Expression' search box for both datasets, users can easily access the quality metrics of the datasets, the cell type composition, the expression profiles of the selected genes, and the heatmap showing differentially expressed genes among cell populations in each dataset (Figure 3), which provides valuable tools for cross-condition analysis. Apart from the visual comparison of gene expression data, MAPbrain provides an online differential expression analysis tool that allows users to directly compute differentially expressed genes between two selected datasets, showing with data visualization and results download options. This feature enables customized cross-species, cross-region or cross-cell-type comparison of the data within the MAPbrain database.

Joint multi-omics analysis

RNA-seq reveals gene expression levels, while ATAC-seq identifies accessible chromatin regions. Linking RNA-seq and ATAC-seq provides a more comprehensive understanding of gene regulation by connecting changes in chromatin accessibility with gene expression data, offering deeper insights into the mechanisms underlying cellular processes and brain development. To achieve this, MAPbrain provides a joint analysis of gene expression profiles and the accessibility of the regulatory elements in chromatin, combining the RNA transcriptional profiles with chromatin accessibility data from related regulatory elements and supports an interconnected visualization of gene expression and chromatin accessibility of associate gene elements. In this module, users can access chromatin accessibility and genes expression by querying a gene symbol in the 'Gene Expression' search box. Additionally, by selecting multiple ATAC-seq datasets with different developmental time points, users could compare the dynamic changes of the regulatory elements throughout brain development. The search results are displayed in two separate windows for each omics (Figure 4).

Cell mapping analysis

To maximize the utility of our integrated data, MAPbrain includes an online tool that accepts user sequencing data as input. By harmonizing user-provided data with our annotated data, this tool provides reliable prediction of cell type annotations and generates a cell navigation map for the input data, enabling users to explore their own data using MAPbrain. Users can submit their own data by clicking the 'Submit' button in the 'CellMapping' module. It's important to note that the input file format for this analysis should be a Seurat object. Users can specify the parameters 'para1' and 'para2'. 'para1' corresponds to the number of neighbors used to make the prediction. 'para2' corresponds to minimum votes for a definite decision. Once the CellMapping analysis is complete, the cell annotation results will be sent to the user via E-mail (Figure 5).

Accessible data and website

In the 'Submit' section, users can upload their own datasets to the database. In the 'Download' section, users can download all processed datasets as SeuratObject in RDS file format, enabling further analysis with the R Seurat package. In the 'Help' section, we have provided extensive guidance and detailed descriptions for navigating the MAPbrain platform.

Discussion

With the rapid advancement of single-cell sequencing and multi-omics techniques, the availability of transcriptome, epigenome, and spatial transcriptome studies on the primate brain has been expanding continuously (51,69). However, the absence of universal standards for data annotation and integration significantly limits researchers' ability to extract in-depth information from this vast data repository. Specifically, the compatibility among various data formats across different omics and datasets is often inadequate, which hinders the widespread application and research of these valuable data collections.

To date, various databases have been developed to explore human brain data, most of which are limited to showcasing individual datasets. By applying a recently developed and widely adopted integration pipeline to a large data collection, MAPbrain provides a broader spectrum of brain regions, a finer classification of sub-regions, and a wider range of developmental periods compared to other existing databases. Unlike model animals like mice, human and non-human primates brain samples are more valuable and rarer to acquire. To our knowledge, MAPbrain incorporates the largest assemblage of transcriptome data from human and non-human primates brain, with over 21 million cells, enabling comparisons of ongoing transformations during brain formation at spatial region and temporal stage levels. Furthermore, the integration of multi-omics data along with the integrative and interactive analysis functions provides a valuable platform for studying brain development.

The collection of both human and non-human primates brain data, facilitating cross-species comparisons. Understanding the differences between human and non-human primates brains is crucial for gaining insights into developmental and evolutionary variations in cognition. Meanwhile, the similarities between them make non-human primates brains ideal research subjects for invasive experiments and data acquisition.

We also notice that there is still room for improvement in MAPbrain. Currently, most metadata are collected using different annotation standards across different data sets, as dictated by their original publications. This variation can affect the consistency of integration results. Selecting a reference dataset for each brain region and mapping the curated annotation on other dataset might be a suitable solution to the problem. Additionally, variations in sequencing depth among sc/snRNA-seq datasets make it difficult to eliminate batch effects and complicate data integration analysis.

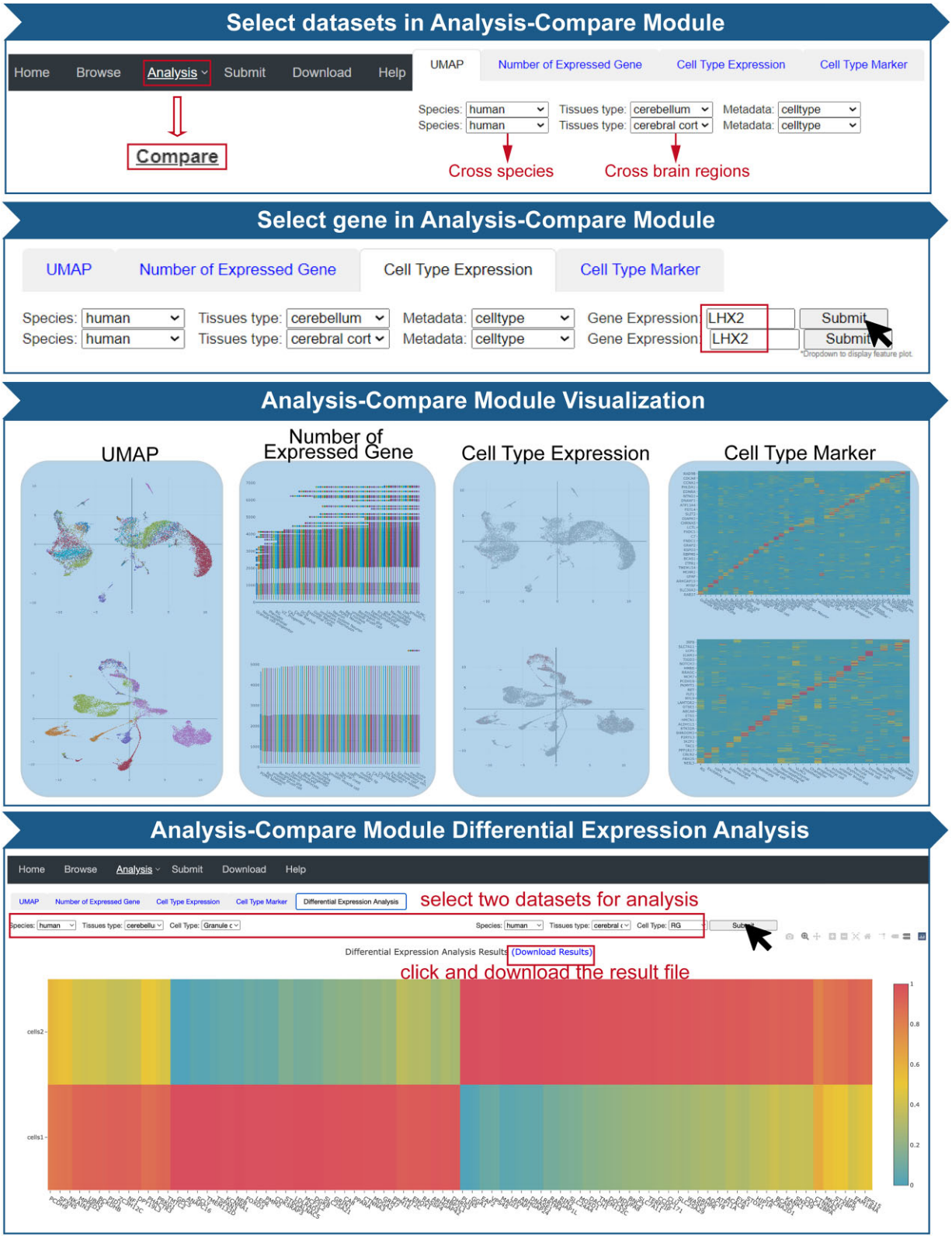


Figure 3. Illustration of the usage of the Analysis-Compare module.

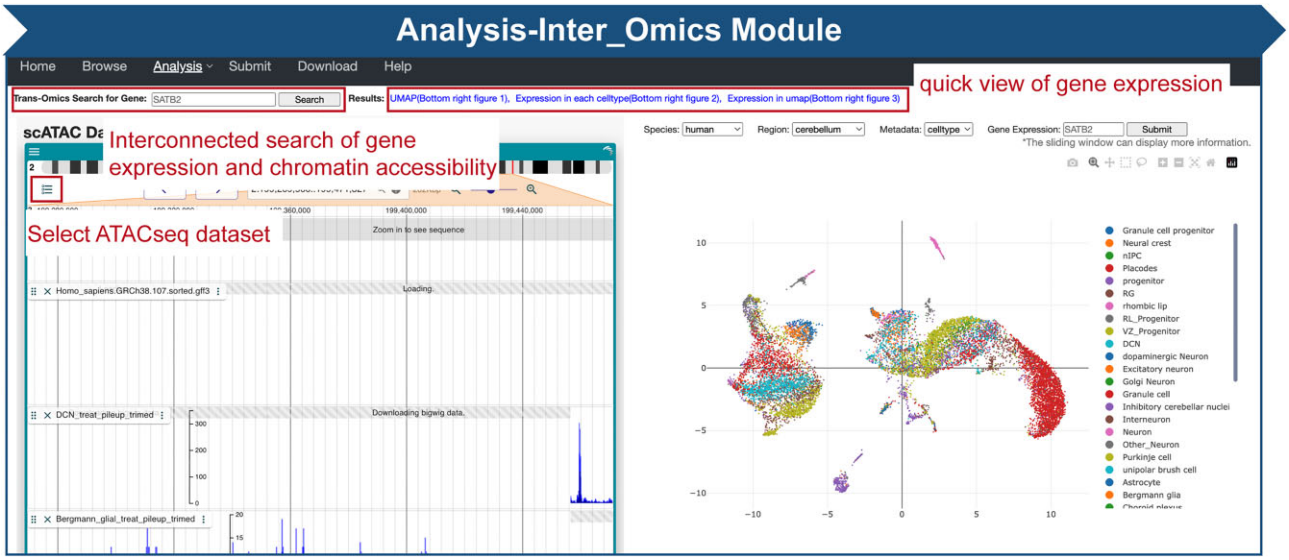


Figure 4. Illustration of the usage of the Analysis-Inter_Omics module.

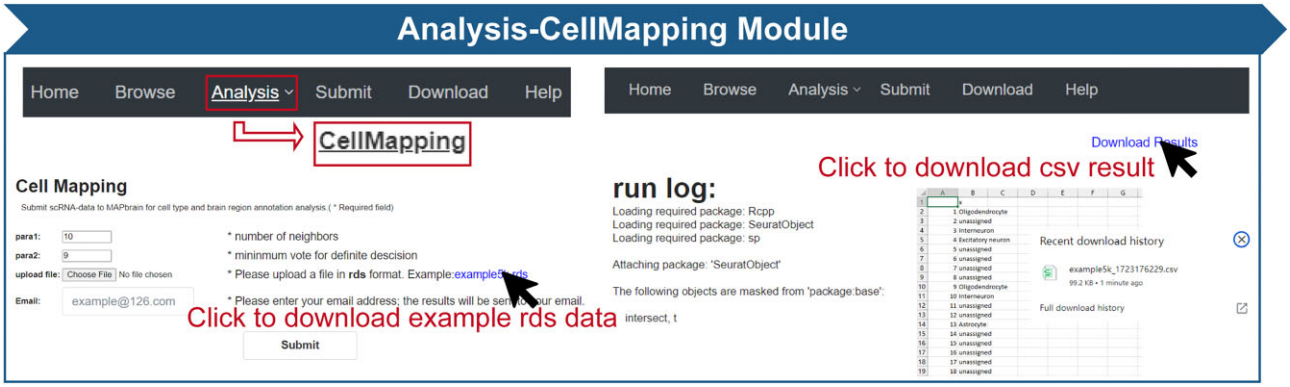


Figure 5. Illustration of the usage of the Analysis-CellMapping module.

The emergence of spatial transcriptomic technologies at single-cell level-such as Stereo-seq (74), STARmap (75), seqFISH (76) and MERFISH (77)- has led to a surge in data published and demonstrated in diverse formats and frameworks. Future work for MAPbrain could involve incorporating a significant archive of spatial transcriptomic data, expanding the range and enhancing the spatial resolution of our database. Building on the previously established tool AllenDigger (78), MAPbrain could map deep sequencing data onto the original spatial location within the brain, creating a precise cell navigation brain map. This would facilitate the further exploration of spatio-temporal dynamics and functional heterogeneity within the brain at the single-cell level.

Overall, we will continuous to incorporate newly published data, particularly the spatial transcriptomics data, to create a comprehensive multi-omics database on brain development in humans and non-human primates, serving as valuable resource for neuroscience research.

Data availability

MAPbrain is freely available at http://bigdata.ibp.ac.cn/mapBRAIN/.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank the National Resource Center for Non-Human Primates for their support. We acknowledge all data contributors who made their raw or processed sequencing data freely accessible.

Author contributions: X.W., Q.W. and S.H. conceived and supervised the project. M.W. and L.Z. searched and collected data sets. M.W. and L.Z. performed the preprocessing and integrating analysis of all data sets. T.S., M.W. and L.Z. established the MAPbrain website. X.W., Q.W., M.W. and L.Z. wrote the manuscript, and all authors edited and proofread the manuscript.

Funding

Science and Technology Innovation 2030—Brain Science and Brain-inspired Intelligence Project of China [2022ZD0207700, 2022YEF0203200, 2021ZD0200102, 2022ZD0211900, 2021ZD0202403]; National Natural

Science Foundation of China (NSFC) [32192411, 32122037, 32222032]; CAS Project for Young Scientists in Basic Research [YSBR-013, YSBR-097]; Changping Laboratory; Fundamental Research Funds for the Central Universities; New Cornerstone Science Foundation. Funding for open access charge: New Cornerstone Science Foundation.

Conflict of interest statement

None declared.

References

1 Levitt,P. (2003) Structural and functional maturation of the developing primate brain. *J. Pediatr.*, **143**, 35–45.
2 Leigh,S.R. (2004) Brain growth, life history, and cognition in primate and human evolution. *Am. J. Primatol.*, **62**, 139–164.
3 Kolodziejczyk,A.A., Kim,J.K., Svensson,V., Marioni,J.C. and Teichmann,S.A. (2015) The technology and biology of single-cell RNA sequencing. *Mol. Cell*, **58**, 610–620.
4 Moses,L. and Pachter,L. (2022) Museum of spatial transcriptomics. *Nat. Methods*, **19**, 534–546.
5 Kelsey,G., Stegle,O. and Reik,W. (2017) Single-cell epigenomics: recording the past and predicting the future. *Science*, **358**, 69–75.
6 Chi,Y., Qi,R., Zhou,Y., Tong,H., Jin,H., Turck,C.W., Chen,W.-H. and Wang,G.-Z. (2023) scBrainMap: a landscape for cell types and associated genetic markers in the brain. *Database*, **2023**, baad035.
7 Yang,Y.T., Gan,Z., Zhang,J., Zhao,X., Yang,Y., Han,S., Wu,W. and Zhao,X.-M. (2024) STAB2: an updated spatio-temporal cell atlas of the human and mouse brain. *Nucleic Acids Res.*, **52**, D1033–D1041.
8 Flotho,M., Amand,J., Hirsch,P., Grandke,F., Wyss-Coray,T., Keller,A. and Kern,F. (2024) ZEBRA: a hierarchically integrated gene expression atlas of the murine and human brain at single-cell resolution. *Nucleic Acids Res.*, **52**, D1089–D1096.
9 Xu,Z., Wang,W., Yang,T., Li,L., Ma,X., Chen,J., Wang,J., Huang,Y., Gould,J. and Lu,H. (2024) STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res.*, **52**, D1053–D1061.
10 Ament,S.A., Adkins,R.S., Carter,R., Chrysostomou,E., Colantuoni,C., Crabtree,J., Creasy,H.H., Degatano,K., Felix,V., Gandt,P., *et al.* (2022) The Neuroscience Multi-Omic Archive: a BRAIN Initiative resource for single-cell transcriptomic and epigenomic data from the mammalian brain. *Nucleic Acids Res.*, **51**, D1075–D1085.
11 Zhou,Z., Tan,C., Chau,MatthewHoiK., Jiang,X., Ke,Z., Chen,X., Cao,Y., Kwok,Y.K., Bellgard,M., Leung,TakY, *et al.*2022) TEDD: a database of temporal gene expression patterns during multiple developmental periods in human and model organisms. *Nucleic Acids Res.*, **51**, D1168–D1178.
12 Deng,Y., Lu,Y., Li,M., Shen,J., Qin,S., Zhang,W., Zhang,Q., Shen,Z., Li,C., Jia,T., *et al.* (2023) SCAN: spatiotemporal Cloud Atlas for Neural cells. *Nucleic Acids Res.*, **52**, D998–D1009.
13 Clough,E., Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H. and Sherman,P.M. (2024) NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.*, **52**, D138–D144.
14 Speir,M.L., Bhaduri,A., Markov,N.S., Moreno,P., Nowakowski,T.J., Papatheodorou,I., Pollen,A.A., Raney,B.J., Seninge,L. and Kent,W.J. (2021) UCSC Cell Browser: visualize your single-cell data. *Bioinformatics*, **37**, 4578–4580.
15 Cao,J., O'day,D.R., Pliner,H.A., Kingsley,P.D., Deng,M., Daza,R.M., Zager,M.A., Aldinger,K.A., Blecher-Gonen,R. and Zhang,F. (2020) A human cell atlas of fetal gene expression. *Science*, **370**, eaba7721.
16 Cao,J., Spielmann,M., Qiu,X., Huang,X., Ibrahim,D.M., Hill,A.J., Zhang,F., Mundlos,S., Christiansen,L. and Steemers,F.J. (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, **566**, 496–502.
17 Thakur,M., Bateman,A., Brooksbank,C., Freeberg,M., Harrison,M., Hartley,M., Keane,T., Kleywegt,G., Leach,A. and Levchenko,M. (2023) EMBL's European bioinformatics Institute (EMBL-EBI) in 2022. *Nucleic Acids Res.*, **51**, D9–D17.
18 Hodge,R.D., Bakken,T.E., Miller,J.A., Smith,K.A., Barkan,E.R., Graybuck,L.T., Close,J.L., Long,B., Johansen,N., Penn,O., *et al.* (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature*, **573**, 61–68.
19 Hao,Y., Stuart,T., Kowalski,M.H., Choudhary,S., Hoffman,P., Hartman,A., Srivastava,A., Molla,G., Madad,S., Fernandez-Granda,C., *et al.* (2024) Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.*, **42**, 293–304.
20 Durinck,S., Spellman,P.T., Birney,E. and Huber,W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.*, **4**, 1184–1191.
21 Martin,F.J., Amode,M.R., Aneja,A., Austine-Orimoloye,O., Azov,AndreyG., Barnes,I., Becker,A., Bennett,R., Berry,A., Bhai,J., *et al.* (2022) Ensembl 2023. *Nucleic Acids Res.*, **51**, D933–D941.
22 Korsunsky,I., Millard,N., Fan,J., Slowikowski,K., Zhang,F., Wei,K., Baglaenko,Y., Brenner,M., Loh,P.-r. and Raychaudhuri,S. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*, **16**, 1289–1296.
23 Ding,J. and Regev,A. (2021) Deep generative model embedding of single-cell RNA-Seq profiles on hyperspheres and hyperbolic spaces. *Nat. Commun.*, **12**, 2554.
24 Kent,W.J., Zweig,A.S., Barber,G., Hinrichs,A.S. and Karolchik,D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204–2207.
25 Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M. and Li,W. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
26 Diesh,C., Stevens,G.J., Xie,P., De,Jesus, Martinez,T., Hershberg,E.A., Leung,A., Guo,E., Dider,S., Zhang,J., *et al.* (2023) JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.*, **24**, 74.
27 Trevino,A.E., Müller,F., Andersen,J., Sundaram,L., Kathiria,A., Shcherbina,A., Farh,K., Chang,H.Y., Paşca,A.M., Kundaje,A., *et al.* (2021) Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell*, **184**, 5053–5069.
28 Nowakowski,T.J., Bhaduri,A., Pollen,A.A., Alvarado,B., Mostajo-Radji,M.A., Di Lullo,E., Haeussler,M., Sandoval-Espinosa,C., Liu,S.J., Velmeshev,D., *et al.* (2017) Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*, **358**, 1318–1323.
29 Bhaduri,A., Andrews,M.G., Mancia Leon,W., Jung,D., Shin,D., Allen,D., Jung,D., Schmunk,G., Haeussler,M., Salma,J., *et al.* (2020) Cell stress in cortical organoids impairs molecular subtype specification. *Nature*, **578**, 142–148.
30 Bhaduri,A., Sandoval-Espinosa,C., Otero-Garcia,M., Oh,I., Yin,R., Eze,U.C., Nowakowski,T.J. and Kriegstein,A.R. (2021) An atlas of cortical arealization identifies dynamic molecular signatures. *Nature*, **598**, 200–204.
31 Velmeshev,D., Perez,Y., Yan,Z., Valencia,J.E., Castaneda-Castellanos,D.R., Wang,L., Schirmer,L., Mayer,S., Wick,B., Wang,S., *et al.* (2023) Single-cell analysis of prenatal and postnatal human cortical development. *Science*, **382**, eadf0834.
32 Kim,C.N., Shin,D., Wang,A. and Nowakowski,T.J. (2023) Spatiotemporal molecular dynamics of the developing human thalamus. *Science*, **382**, eadf9941.
33 van Bruggen,D., Pohl,F., Langseth,C.M., Kukanja,P., Lee,H., Albiach,A.M., Kabbe,M., Meijer,M., Linnarsson,S., Hilscher,M.M., *et al.* (2022) Developmental landscape of human forebrain at a single-cell level identifies early waves of oligodendrogenesis. *Dev. Cell*, **57**, 1421–1436.

34 Ayhan,F., Kulkarni,A., Berto,S., Sivaprakasam,K., Douglas,C., Lega,B.C. and Konopka,G. (2021) Resolving cellular and molecular diversity along the hippocampal anterior-to-posterior axis in humans. *Neuron*, **109**, 2091–2105.

35 Su,Y., Zhou,Y., Bennett,M.L., Li,S., Carceles-Cordon,M., Lu,L., Huh,S., Jimenez-Cyrus,D., Kennedy,B.C., Kessler,S.K., *et al.* (2022) A single-cell transcriptome atlas of glial diversity in the human hippocampus across the postnatal lifespan. *Cell Stem Cell*, **29**, 1594–1610.

36 Domcke,S., Hill,A.J., Daza,R.M., Cao,J., O'Day,D.R., Pliner,H.A., Aldinger,K.A., Pokholok,D., Zhang,F., Milbank,J.H., *et al.* (2020) A human cell atlas of fetal chromatin accessibility. *Science*, **370**, eaba7612.

37 Zhu,K., Bendl,J., Rahman,S., Vicari,J.M., Coleman,C., Clarence,T., Latouche,O., Tsankova,N.M., Li,A., Brennand,K.J., *et al.* (2023) Multi-omic profiling of the developing human cerebral cortex at the single-cell level. *Sci. Adv.*, **9**, eadg3754.

38 Ament,S.A., Cortes-Gutierrez,M., Herb,B.R., Mocci,E., Colantuoni,C. and McCarthy,M.M. (2023) A single-cell genomic atlas for maturation of the human cerebellum during early childhood. *Sci. Transl. Med.*, **15**, eade1283.

39 Braun,E., Danan-Gotthold,M., Borm,L.E., Lee,K.W., Vinsland,E., Lönnerberg,P., Hu,L., Li,X., He,X., Andrusivová,Ž., *et al.* (2023) Comprehensive cell atlas of the first-trimester developing human brain. *Science*, **382**, eadf1226.

40 Herb,B.R., Glover,H.J., Bhaduri,A., Colantuoni,C., Bale,T.L., Siletti,K., Hodge,R., Lein,E., Kriegstein,A.R., Doege,C.A., *et al.* (2023) Single-cell genomics reveals region-specific developmental trajectories underlying neuronal diversity in the human hypothalamus. *Sci. Adv.*, **9**, eadf6251.

41 Herring,C.A., Simmons,R.K., Freytag,S., Poppe,D., Moffet,J.J.D., Pflueger,J., Buckberry,S., Vargas-Landin,D.B., Clément,O., Echeverría,E.G., *et al.* (2022) Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell*, **185**, 4428–4447.

42 Aldinger,K.A., Thomson,Z., Phelps,I.G., Haldipur,P., Deng,M., Timms,A.E., Hirano,M., Santpere,G., Roco,C., Rosenberg,A.B., *et al.* (2021) Spatial and cell type transcriptional landscape of human cerebellar development. *Nat. Neurosci.*, **24**, 1163–1175.

43 Eze,U.C., Bhaduri,A., Haeussler,M., Nowakowski,T.J. and Kriegstein,A.R. (2021) Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.*, **24**, 584–594.

44 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.

45 Xu,Y., Zhang,T., Zhou,Q., Hu,M., Qi,Y., Xue,Y., Nie,Y., Wang,L., Bao,Z. and Shi,W. (2023) A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat. Cell Biol.*, **25**, 604–615.

46 Hao,Z.-Z., Wei,J.-R., Xiao,D., Liu,R., Xu,N., Tang,L., Huang,M., Shen,Y., Xing,C., Huang,W., *et al.* (2022) Single-cell transcriptomics of adult macaque hippocampus reveals neural precursor cell populations. *Nat. Neurosci.*, **25**, 805–817.

47 Mannens,C.C.A., Hu,L., Lönnerberg,P., Schipper,M., Reagor,C.C., Li,X., He,X., Barker,R.A., Sundström,E., Posthuma,D., *et al.* (2024) Chromatin accessibility during human first-trimester neurodevelopment. *Nature*, https://doi.org/10.1038/s41586-024-07234-1.

48 Siletti,K., Hodge,R., Mossi Albiach,A., Lee,K.W., Ding,S.-L., Hu,L., Lönnerberg,P., Bakken,T., Casper,T., Clark,M., *et al.* (2023) Transcriptomic diversity of cell types across the adult human brain. *Science*, **382**, eadd7046.

49 Jorstad,N.L., Song,J.H.T., Exposito-Alonso,D., Suresh,H., Castro-Pacheco,N., Krienen,F.M., Yanny,A.M., Close,J., Gelfand,E., Long,B., *et al.* (2023) Comparative transcriptomics reveals human-specific cortical features. *Science*, **382**, eade9516.

50 Micali,N., Ma,S., Li,M., Kim,S.-K., Mato-Blanco,X., Sindhu,S.K., Arellano,J.I., Gao,T., Shibata,M., Gobeske,K.T., *et al.* (2023) Molecular programs of regional specification and neural stem cell fate progression in macaque telencephalon. *Science*, **382**, eadf3786.

51 Li,Y., Li,Z., Wang,C., Yang,M., He,Z., Wang,F., Zhang,Y., Li,R., Gong,Y., Wang,B., *et al.* (2023) Spatiotemporal transcriptome atlas reveals the regional specification of the developing human brain. *Cell*, **186**, 5892–5909.

52 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.

53 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.

54 Krienen,F.M., Goldman,M., Zhang,Q., C. H. del Rosario,R., Florio,M., Machold,R., Saunders,A., Levandowski,K., Zaniewski,H., Schuman,B., *et al.* (2020) Innovations present in the primate interneuron repertoire. *Nature*, **586**, 262–269.

55 Krienen,F.M., Levandowski,K.M., Zaniewski,H., del Rosario,R.C.H., Schroeder,M.E., Goldman,M., Wienisch,M., Lutservitz,A., Beja-Glasser,V.F., Chen,C., *et al.* (2023) A marmoset brain cell census reveals regional specialization of cellular identities. *Sci. Adv.*, **9**, eadk3986.

56 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.

57 Zhu,Y., Sousa,A.M.M., Gao,T., Skarica,M., Li,M., Santpere,G., Esteller-Cucala,P., Juan,D., Ferrández-Peral,L., Gulden,F.O., *et al.* (2018) Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science*, **362**, eaat8077.

58 Maynard,K.R., Collado-Torres,L., Weber,L.M., Uytingco,C., Barry,B.K., Williams,S.R., Catallini,J.L., Tran,M.N., Besich,Z., Tippani,M., *et al.* (2021) Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.*, **24**, 425–436.

59 Fan,X., Fu,Y., Zhou,X., Sun,L., Yang,M., Wang,M., Chen,R., Wu,Q., Yong,J., Dong,J., *et al.* (2020) Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci. Adv.*, **6**, eaaz2978.

60 Fan,X., Dong,J., Zhong,S., Wei,Y., Wu,Q., Yan,L., Yong,J., Sun,L., Wang,X., Zhao,Y., *et al.* (2018) Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.*, **28**, 730–745.

61 Zhong,S., Zhang,S., Fan,X., Wu,Q., Yan,L., Dong,J., Zhang,H., Li,L., Sun,L., Pan,N., *et al.* (2018) A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*, **555**, 524–528.

62 Zhong,S., Ding,W., Sun,L., Lu,Y., Dong,H., Fan,X., Liu,Z., Chen,R., Zhang,S., Ma,Q., *et al.* (2020) Decoding the development of the human hippocampus. *Nature*, **577**, 531–536.

63 Zhou,X., Lu,Y., Zhao,F., Dong,J., Ma,W., Zhong,S., Wang,M., Wang,B., Zhao,Y., Shi,Y., *et al.* (2022) Deciphering the spatial-temporal transcriptional landscape of human hypothalamus development. *Cell Stem Cell*, **29**, 328–343.

64 Zeng,B., Liu,Z., Lu,Y., Zhong,S., Qin,S., Huang,L., Zeng,Y., Li,Z., Dong,H., Shi,Y., *et al.* (2023) The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell Stem Cell*, **30**, 851–866.

65 Shi,Y., Wang,M., Mi,D., Lu,T., Wang,B., Dong,H., Zhong,S., Chen,Y., Sun,L., Zhou,X., *et al.* (2021) Mouse and human share conserved transcriptional programs for interneuron development. *Science*, **374**, eabj6641.

66 Wang,W., Wang,M., Yang,M., Zeng,B., Qiu,W., Ma,Q., Jing,X., Zhang,Q., Wang,B., Yin,C., *et al.* (2022) Transcriptome dynamics

of hippocampal neurogenesis in macaques across the lifespan and aged humans. *Cell Res.*, **32**, 729–743.
67 Zhong,S., Wang,M., Huang,L., Chen,Y., Ge,Y., Zhang,J., Shi,Y., Dong,H., Zhou,X., Wang,B., *et al.* (2023) Single-cell epigenomics and spatiotemporal transcriptomics reveal human cerebellar development. *Nat. Commun.*, **14**, 7613.
68 Lei,Y., Cheng,M., Li,Z., Zhuang,Z., Wu,L., sun,Y., Han,L., Huang,Z., Wang,Y., Wang,Z., *et al.* (2022) Spatially resolved gene regulatory and disease-related vulnerability map of the adult Macaque cortex. *Nat. Commun.*, **13**, 6747.
69 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.
70 Chiou,K.L., Huang,X., Bohlen,M.O., Tremblay,S., DeCasien,A.R., O'Day,D.R., Spurrell,C.H., Gogate,A.A., Zintel,T.M., Unit,C.B.R, *et al.*2023) A single-cell multi-omic atlas spanning the adult rhesus macaque brain. *Sci. Adv.*, **9**, eadh1914.
71 Schmitz,M.T., Sandoval,K., Chen,C.P., Mostajo-Radji,M.A., Seeley,W.W., Nowakowski,T.J., Ye,C.J., Paredes,M.F. and Pollen,A.A. (2022) The development and evolution of inhibitory neurons in primate cerebrum. *Nature*, **603**, 871–877.
72 Markenscoff-Papadimitriou,E., Whalen,S., Przytycki,P., Thomas,R., Binyameen,F., Nowakowski,T.J., Kriegstein,A.R., Sanders,S.J., State,M.W., Pollard,K.S., *et al.* (2020) A chromatin accessibility atlas of the developing human telencephalon. *Cell*, **182**, 754–769.
73 Wirthlin,M., Kaplow,I.M., Lawler,A.J., He,J., Phan,B.N., Brown,A.R., Stauffer,W.R. and Pfenning,A.R. (2020) The Regulatory Evolution of the Primate Fine-Motor System. bioRxiv doi: https://doi.org/10.1101/2020.10.27.356733, 27 October 2020, preprint: not peer reviewed.
74 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.
75 Shi,H., He,Y., Zhou,Y., Huang,J., Maher,K., Wang,B., Tang,Z., Luo,S., Tan,P., Wu,M., *et al.* (2023) Spatial atlas of the mouse central nervous system at molecular resolution. *Nature*, **622**, 552–561.
76 Eng,C.-H.L., Lawson,M., Zhu,Q., Dries,R., Koulena,N., Takei,Y., Yun,J., Cronin,C., Karp,C., Yuan,G.-C., *et al.* (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, **568**, 235–239.
77 Xia,C., Babcock,H.P., Moffitt,J.R. and Zhuang,X. (2019) Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.*, **9**, 7721.
78 Wang,M., Zhuo,L., Ma,W., Wu,Q., Zhuo,Y. and Wang,X. (2023) AllenDigger, a tool for spatial expression data visualization, spatial heterogeneity delineation, and single-cell registration based on the Allen Brain Atlas. *J. Phys. Chem. A*, **127**, 2864–2872.

Received: August 15, 2024. **Revised:** September 26, 2024. **Editorial Decision:** October 1, 2024. **Accepted:** October 7, 2024