Supplementary Information

Single-nucleus and spatial transcriptome identify brain landscape of gene regulatory networks associated with behavioral maturation in honeybees

Xiaohuan Mu[1,#], Zijing Zhang[2,#], Qun Liu[3], Jie Ma[3], Yating Qin[3], Haoyu Lang[1], Yingying Zhang[3], Nannan Zhang[3], Qunfei Guo[4,5], Pei Zhang[4], Denghui Li[3], Ruihua Zhang[3,5], Qianyue Ji[3], Aijun Jiang[3], Yang Wang[4], Shanshan Pan[3], Xiawei Liu[3], Xuemei Liu[3], Jiahui Sun[3], Yan Liu[6], Hao Chen[6], Li Zheng[6], Liang Meng[3], Haorong Lu[4], He Zhang[4], Yifan Zhai[6], Qiye Li[4,5], Junnian Liu[3], Huanming Yang[4], Jian Wang[4], Xiaosong Hu[1], Xun Xu[4,*], Shanshan Liu[3,*], Hao Zheng[1,*]

[1]College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China

[2]College of Life Sciences, Hebei Normal University, Shijiazhuang 050024, China

[3]BGI Research-Qingdao, BGI, Qingdao 266555, China

[4]BGI Research-Shenzhen, Shenzhen 518083, China

[5]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[6]Institute of Plant Protection, Shandong Academy of Agricultural Sciences, Jinan 250100, China

*Corresponding authors:

Hao Zheng, hao.zheng@cau.edu.cn

Shanshan Liu, liushanshan@genomics.cn

Xun Xu, xuxun@genomics.cn

[#]These authors contributed equally to this work.

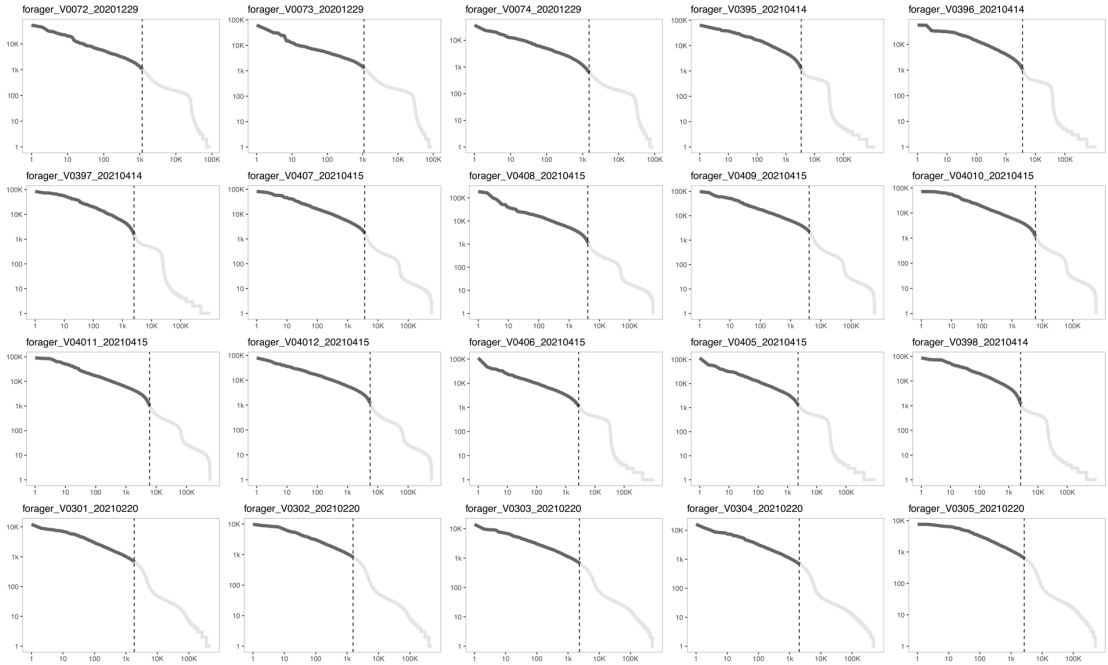

Fig. S1. Ambient RNA noise reduction in forager libraries. The barcode rank plots represent the cumulative numbers of UMIs attributed to real cells and empty barcodes for each library from forager bees.

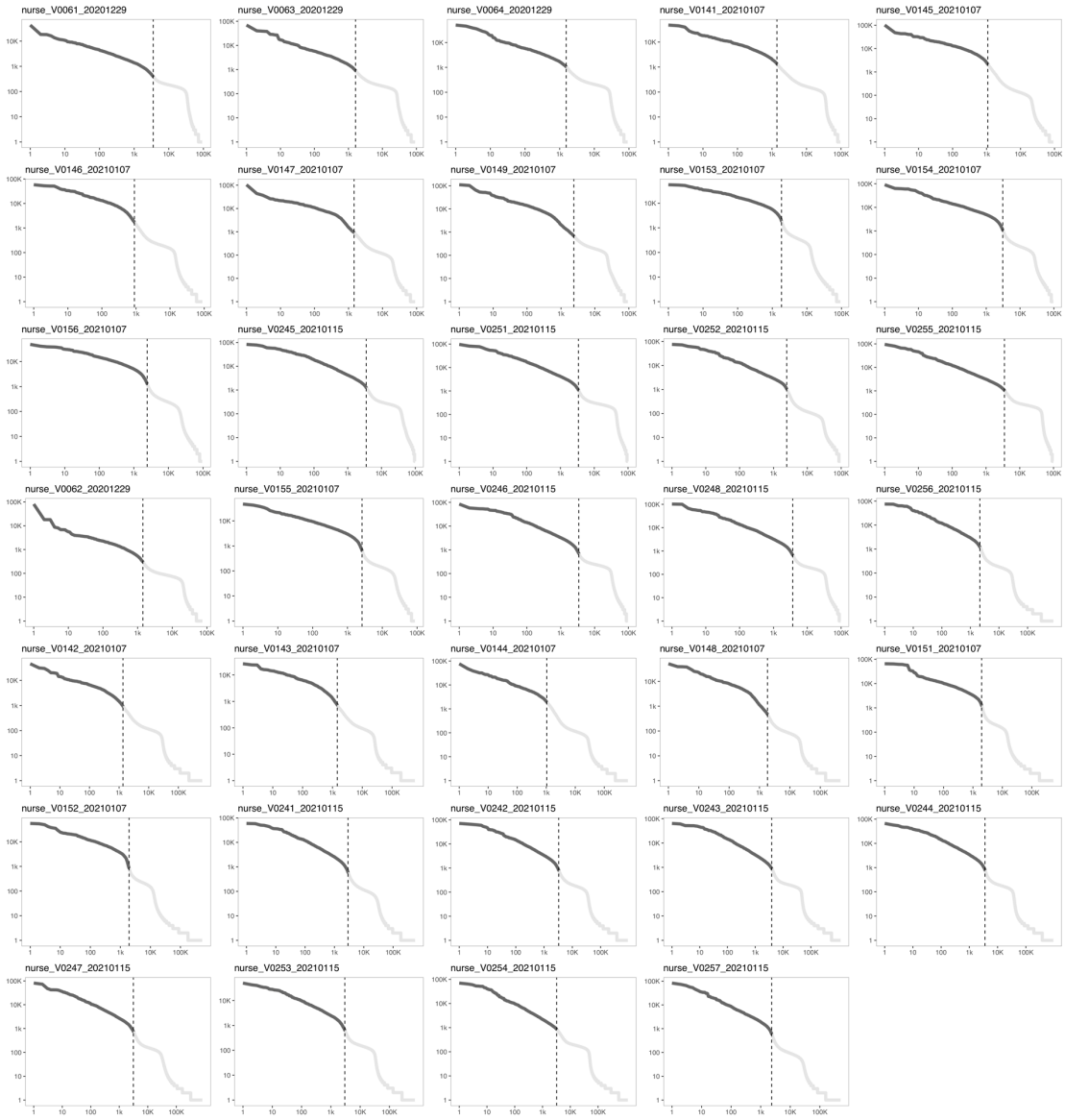

Fig. S2. Ambient RNA noise reduction in nurse libraries. The barcode rank plots represent the cumulative numbers of UMIs attributed to real cells and empty barcodes for each library from nurse bees.

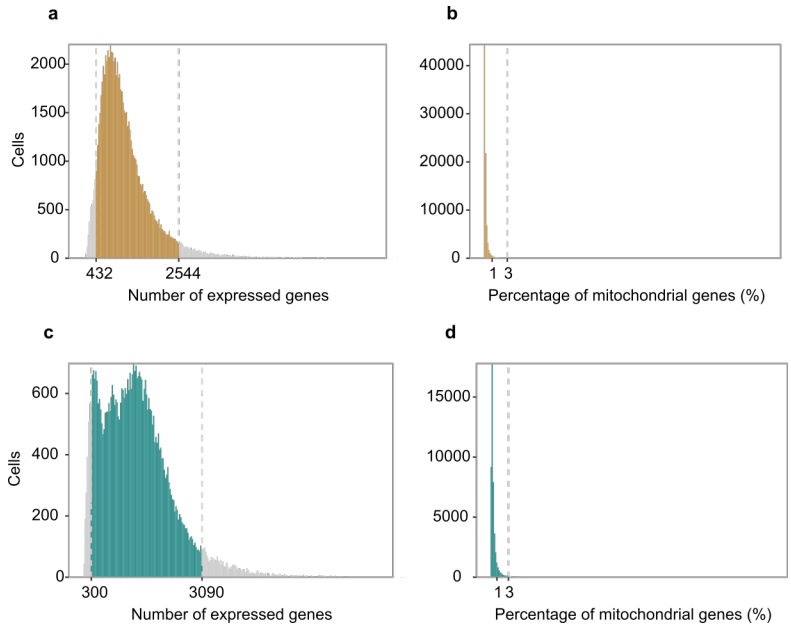

Fig. S3. Single-cell sequencing quality control. (a-b) Nurse bees: **(a)** Number of expressed genes. Cutoff: 5th to 95th. **(b)** Percentage of mitochondrial genes. Cutoff: < 0.03 (8,739 cells failed). **(c-d)** Forager bees: **(c)** Number of expressed genes. Cutoff: 5th to 95th. **(d)** Percentage of mitochondrial genes. Cutoff: < 0.03 (6,860 cells filtered out).

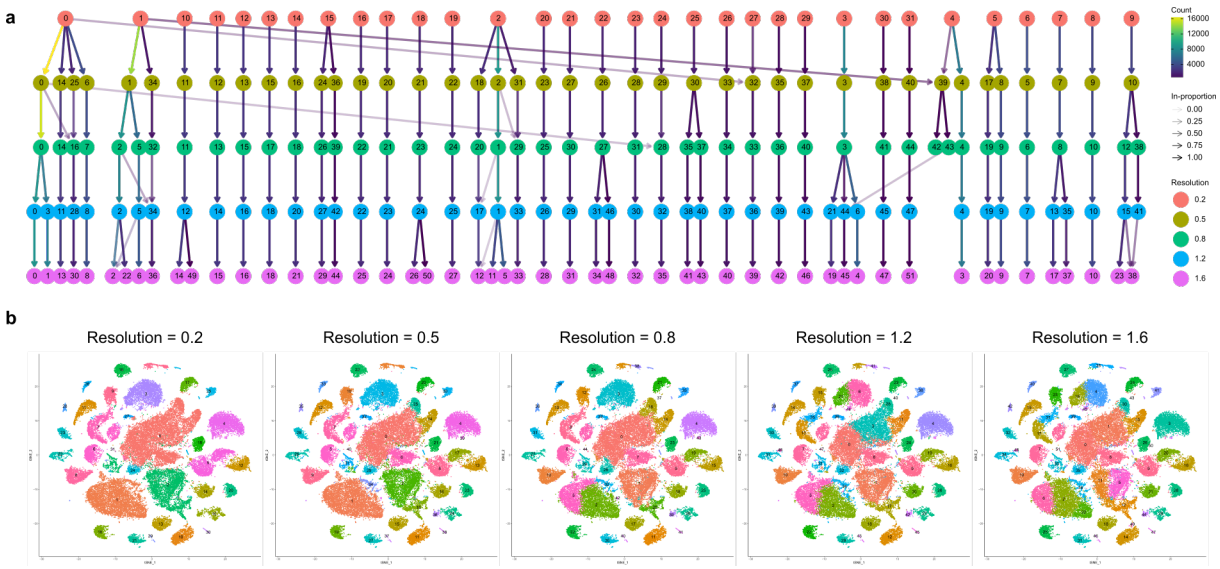

Fig. S4. Clustering tree based on unsupervised clustering of integrated UMI matrix. (a) The nodes are colored according to the resolution parameters from 0.2 to 1.6. The lines are colored according to the number of cells. The transparency is adjusted according to the in-proportion, with stronger lines showing links more important to the higher-resolution cluster. **(b)** tSNE visualization of the clustering of cells from nurse and forager bees with resolution parameters from 0.2 to 1.6.

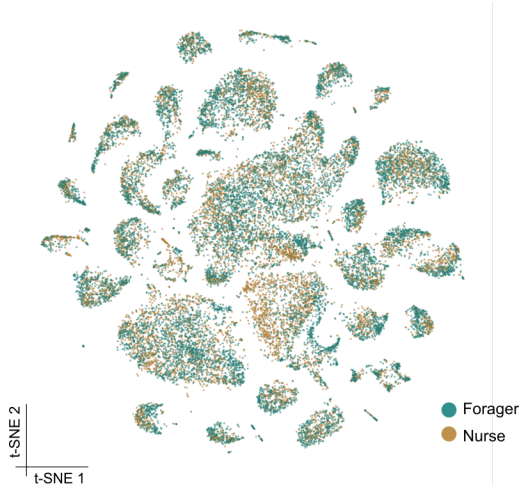

Fig. S5. Cell clustering in worker bee brains. tSNE visualization of cell clustering based on transcriptomes from nurse and forager bees. Each dot represents a cell, with colors indicating the labor division of workers.

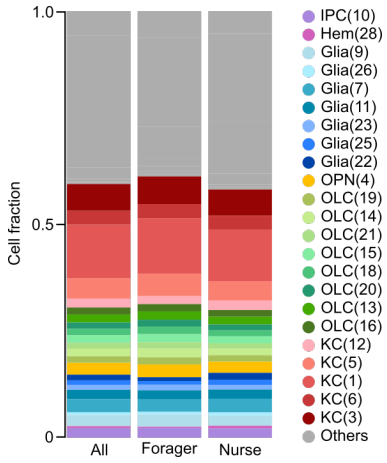

Fig. S6. Cell composition in worker bee brains. Fraction of cells of each subtype from all (n = 54), nurse (n = 34), and forager (n = 20) bees. All the annotations of the cell subtypes refer to the clusters shown in Fig. 1. Others: unknown cell type. Source data are provided as a Source Data file.

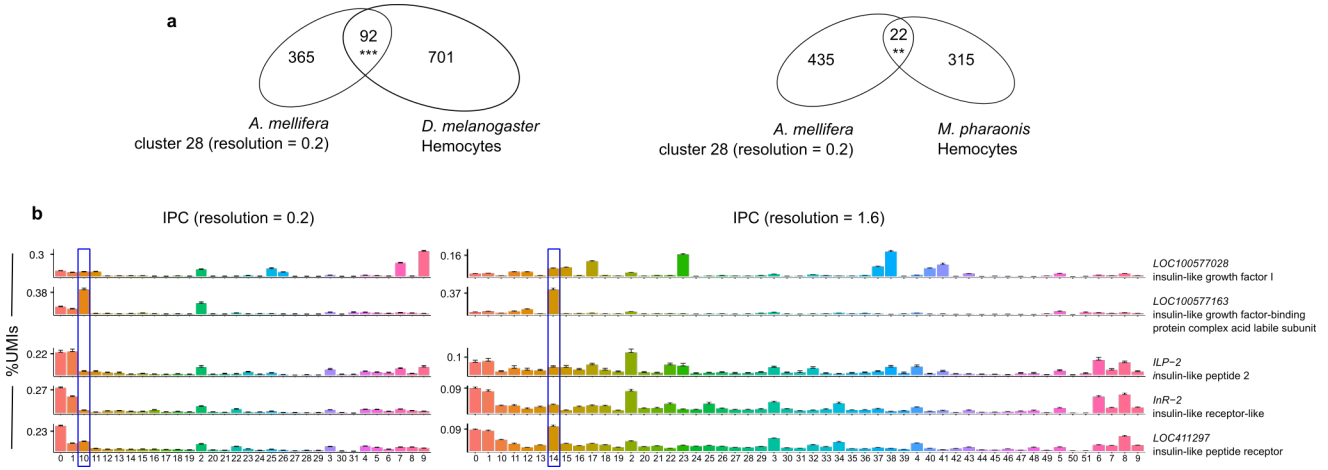

Fig. S7. Annotations of hemocyte and insulin-producing cells at different resolutions. (a) Overrepresentation analysis within marker genes of cluster 28 (resolution = 0.2, calculated by Seurat) in *A. mellifera* and hemocytes markers of *D. melanogaster* (Davie *et al.*, 2018) and *M. pharaonic* (Li *et al.*, 2022). Significance evaluated by Fisher's exact test (two-sided), adjusted by Benjamini-Hochberg correction: ** = 0.001 < adj. P < 0.01; ***= adj. P < 0.001. **(b)** Selected marker genes for the clusters of insulin-related (IPC) cells annotated in nurse and forager bee brains on two different resolutions. The expression levels of indicated genes were evaluated by the collapsed pseudobulk expression in each cluster as a percentage of unique molecular identifiers (UMIs) of the total cluster. Bars represent the means of 54 biological replicates + SEM. Source data are provided as a Source Data file.

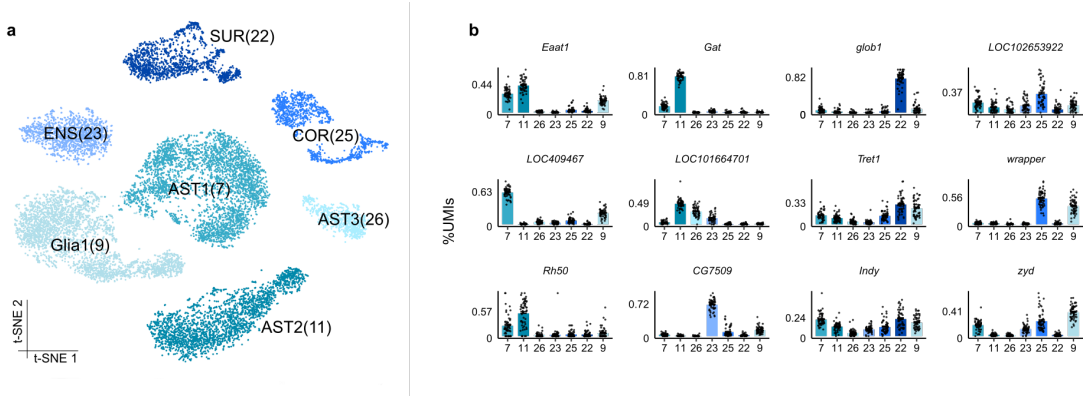

Fig. S8. Subclustering of glial cells from honey bee worker brains. (a) tSNE visualization of the subclustering of glial cells from the brains of nurse and forager bees. AST: astrocyte-like glia; COR: cortex glia; ENS: ensheathing glia; SUR: surface glia. Cluster identifiers from Seurat clustering at a resolution of 0.2 are indicated in parentheses. **(b)** Selected marker genes for the subclusters annotated in glial cells. Bars represent the means of 54 biological replicates + SEM. Source data are provided as a Source Data file.

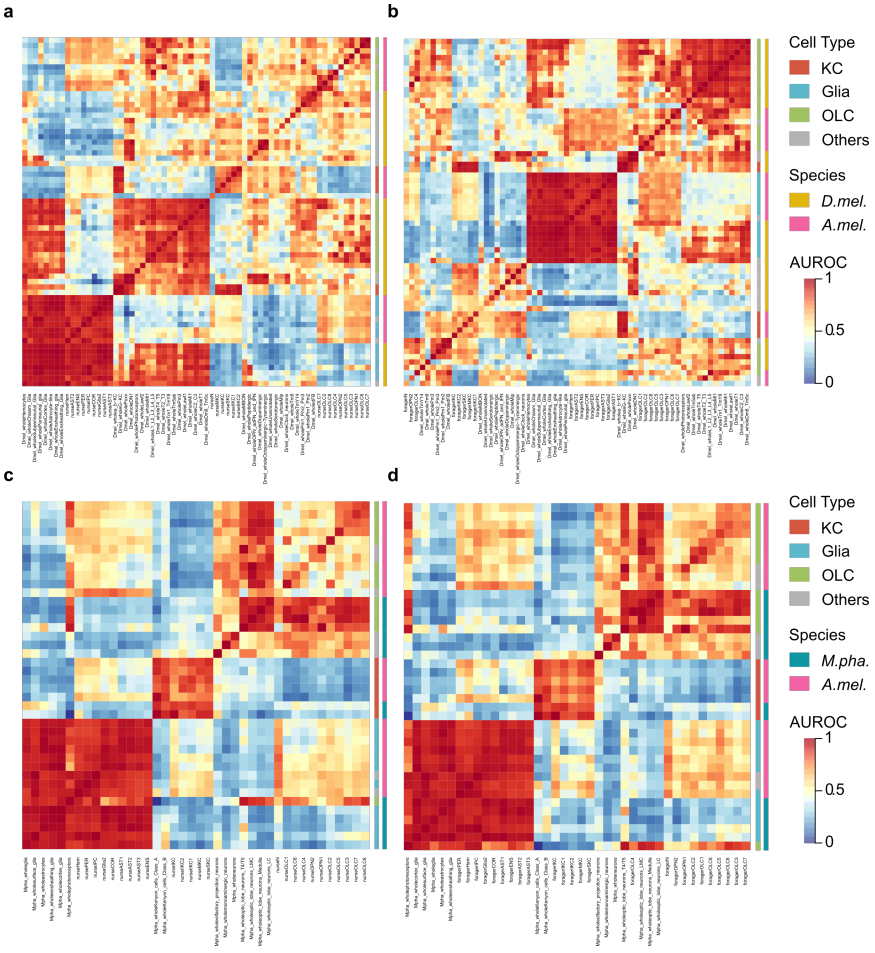

Fig. S9. Classification of the *A. mellifera* cell clusters compared with *D. melanogaster* and *M. pharaonic*. (a-b) Pairwise transcriptional similarity (measured by AUROC scores) of cell clusters from *D. melanogaster* (Davie *et al.*, 2018) to nurse **(a)** and forager honeybees **(b)**. **(c-d)** Pairwise transcriptional similarity of cell clusters from *M. pharaonic* (Li *et al.*, 2022) to nurse **(c)** and forager bees **(d)**. Source data are provided as a Source Data file.

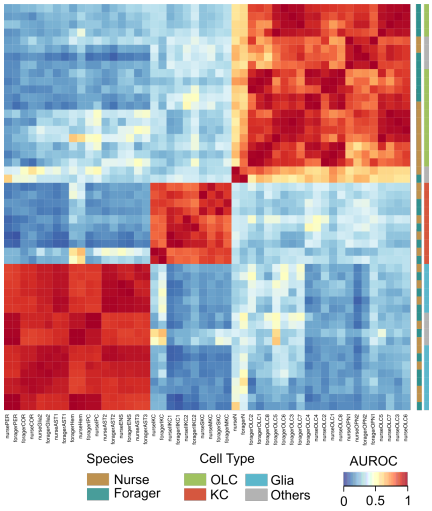

Fig. S10. Consistency of cell types across worker castes. Clustered heatmap showing the pairwise transcriptional similarity (measured by AUROC scores) of cell clusters between nurse and forager bees. Source data are provided as a Source Data file.

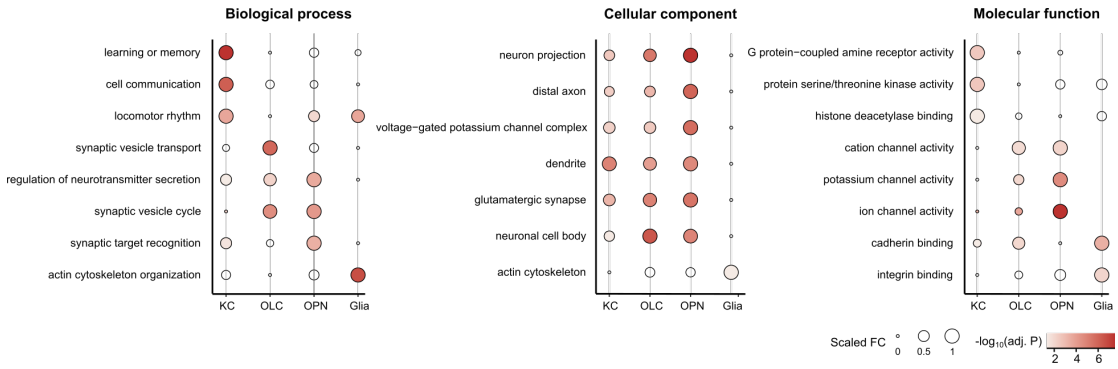

Fig. S11. Functional profiles of the brain cell populations. Gene Ontology enrichment analysis of differentially expressed genes for different cell types. Fold change is determined by the number of detected genes compared to the total number of genes within the GO term. Significance evaluated by two-sided Hypergeometric test, adjusted by Benjamini-Hochberg correction. Related to Dataset S4.

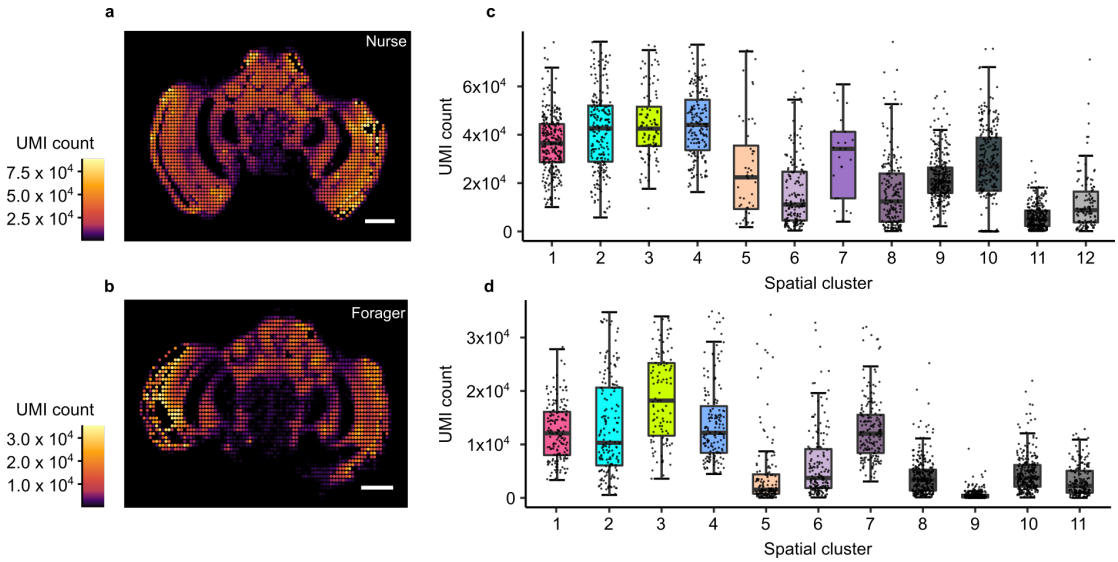

Fig. S12. Quality control metrics of the spatial transcriptomic datasets. (a-b) Spatial visualization of the distribution of the number of transcripts (UMI) in the brain of nurse **(a)** and forager **(b)** bees. Scale bars, 0.3 mm. **(c-d)** Box plots showing the number of UMI per spatial cluster from the section of nurse **(c)** and forager **(d)** bee brains. Each point represents a bin 50. Each box represents the interquartile range (IQR), spanning the 25th to 75th percentiles. The central line within the box indicates the median. Whiskers extend to the minima and maxima within 1.5 × IQR. Source data are provided as a Source Data file.

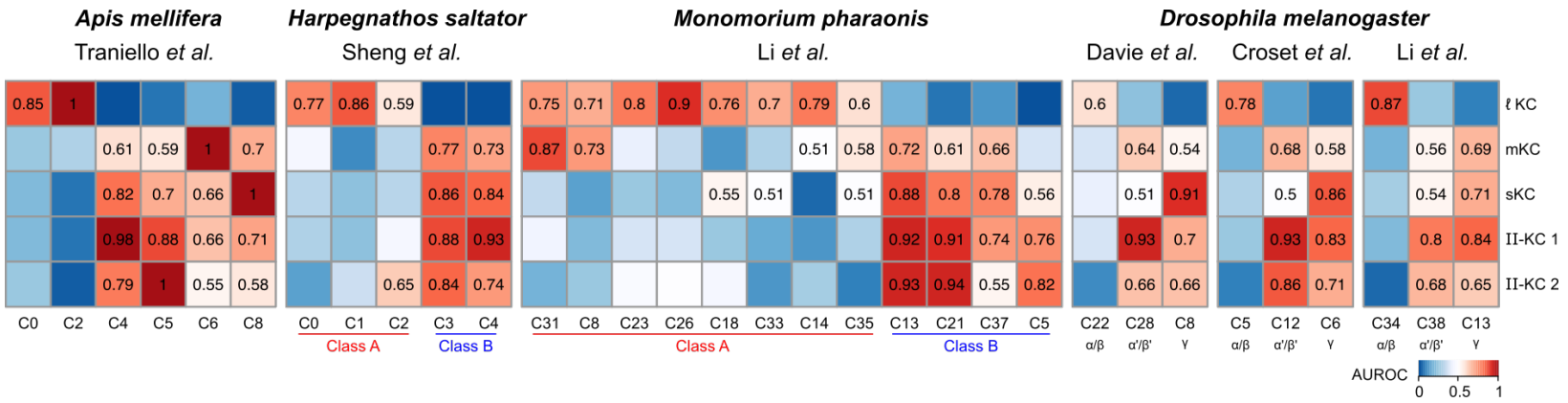

Fig. S13. Comparison of Kenyon cells between honeybees, ants, and *Drosophila*. Pairwise AUROC scores showing the cross-species transcriptional similarity of the KC subtypes from three hymenopteran species, including *A. mellifera* (Traniello *et al.*, 2020), *H. saltator* (Li *et al.*, 2020), *M. pharaonic* (Li *et al.*, 2022) and *D. melanogaster*. *D. melanogaster* KCs from three independent studies by Davie *et al.*, Croset *et al.*, and Li *et al.*. Comparisons with AUROC scores > 0.5 are indicated with exact values. Source data are provided as a Source Data file.

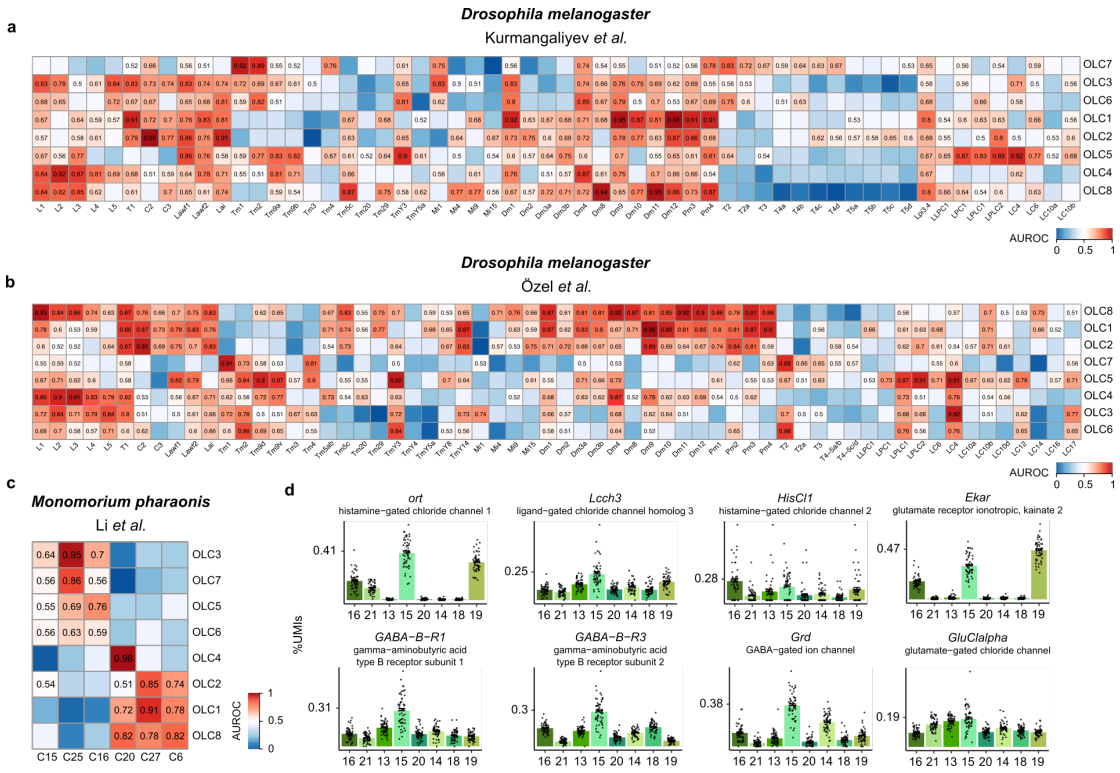

Fig. S14. Comparison of optic lobe cells between honeybees, ants, and *Drosophila*. (a-b) Pairwise AUROC scores showing the cross-species transcriptional similarity of OLC clusters between *A. mellifera* and *D. melanogaster*. *D. melanogaster* OL cell clusters from two independent studies focusing on the *Drosophila* optic lobes, Kurmangaliyev *et al.* and Özel *et al.*, were incorporated. Comparisons with AUROC scores > 0.5 are shown with exact values. **(c)** Pairwise AUROC scores showing the cross-species transcriptional similarity of OL clusters between *A. mellifera* and *M. pharaonic* (Li *et al.*, 2022). **(d)** Selected marker genes for the OLC4 of honeybee. The y-axis shows the collapsed pseudobulk expression in each cluster (%UMIs). Bars represent the means of 54 biological replicates + SEM. Source data are provided as a Source Data file.

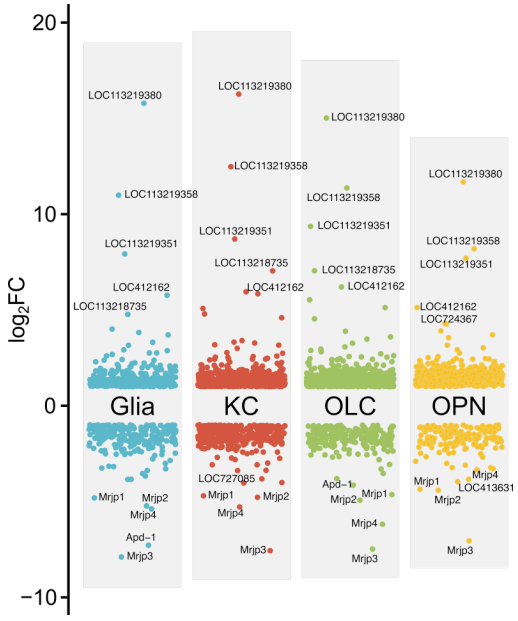

Fig. S15. Differentially expressed genes in the brain cells of nurse and forager bees. Volcano plots showing differentially expressed genes (DEGs) in each cell type between worker castes. Significance assessed by two-sided quasi-likelihood F-tests in edgeR, adjusted by Benjamini-Hochberg. DEGs were identified as genes with an absolute $log_2FC > 1$ and an adjusted P value < 0.05. The top 10 differentially expressed genes between nurses and foragers are labeled accordingly. Source data are provided as a Source Data file.

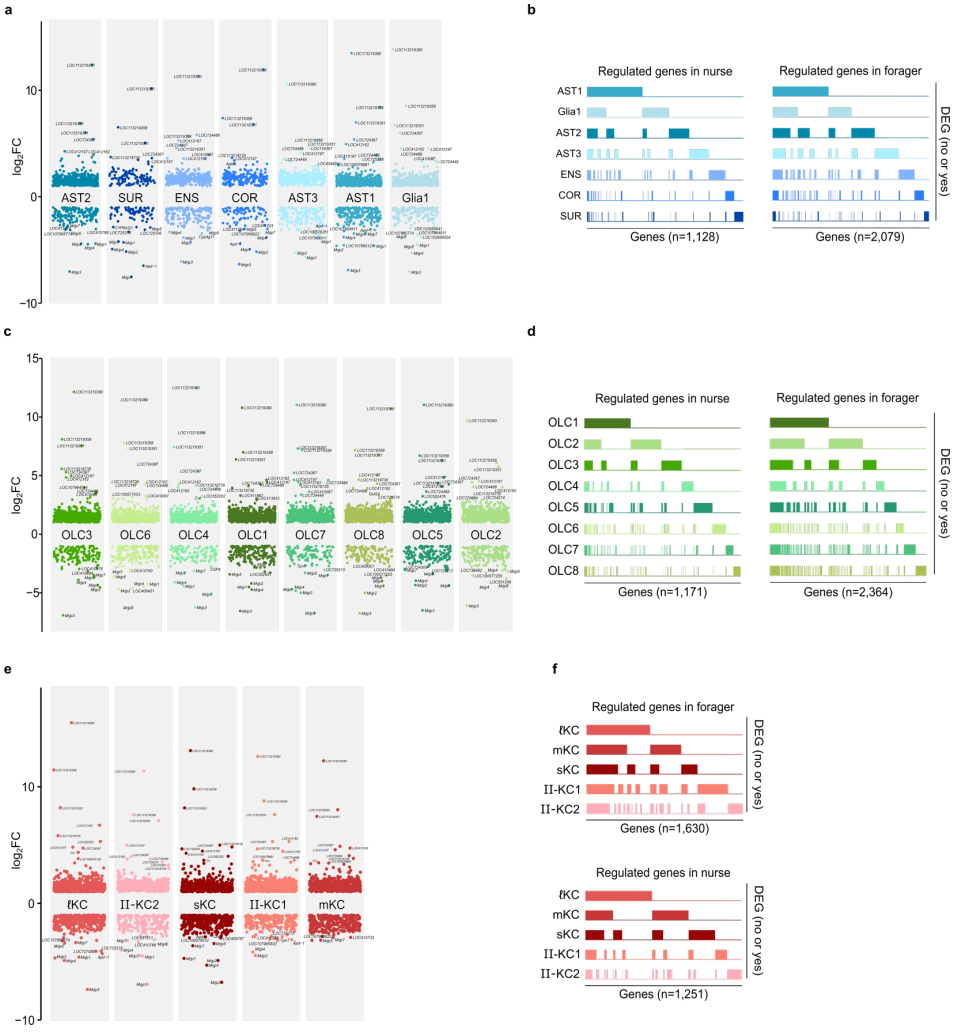

Fig. S16. Distinctive gene expression profiles of different cell subtypes from nurse and forager bees. (a) Volcano plots showing DEGs in seven cell subtypes of glial cells between nurse and forager bees. Significance assessed by two-sided quasi-likelihood F-tests in edgeR, adjusted by Benjamini-Hochberg. DEGs were identified as genes with an absolute $\log_2$FC >1 and an adjusted P value < 0.05. **(b)** DEGs identified from seven glial cell subtypes of nurse (n=1,128) and forager (n=2,079) bees. Bars indicate whether a gene is a DEG in a given cell type or not. **(c)** Volcano plots showing DEGs in eight cell subtypes of optic lobe cells between nurse and forager bees. **(d)** DEGs identified from eight optic lobe cell subtypes of nurse (n=1,171) and forager (n=2,364) bees. **(e)** Volcano plots showing DEGs in five cell subtypes of Kenyon cell between nurse and forager bees. **(f)** DEGs identified from five Kenyon cell subtypes of nurse (n=1,251) and forager (n=1,630) bees. Source data are provided as a Source Data file.

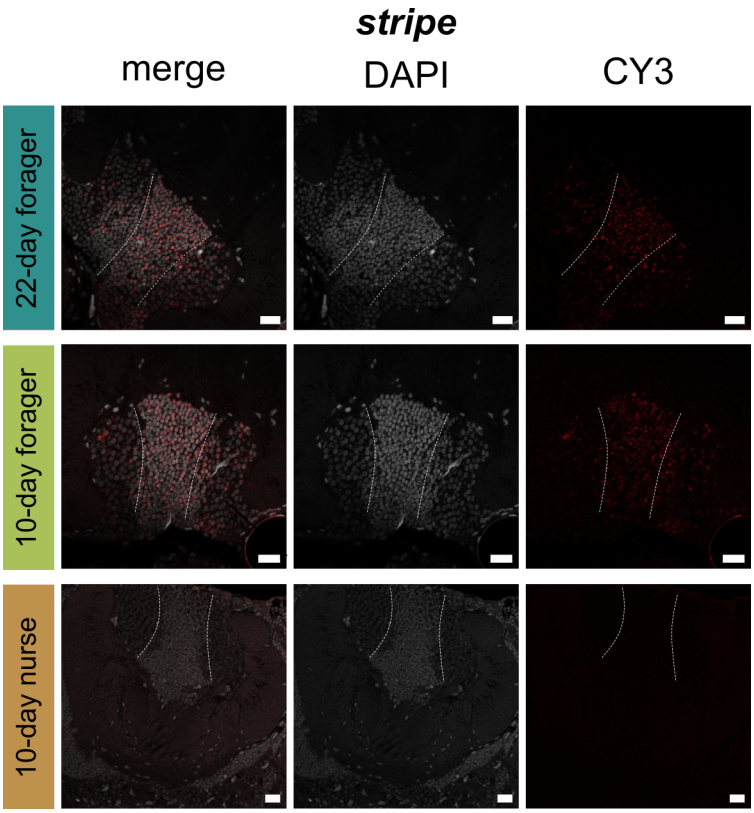

Fig. S17. The localization of *stripe* in the MB. Confocal microscopy images showing simultaneous detection of *stripe* mRNA (red) and (4', 6-diamidino-2-phenylindole) DAPI-stained nuclei (white) in worker brains. brain sections are from 22-day-old forager, 10-day-old forager, and 10-day-old nurse bee. Scale bars = 20 um.

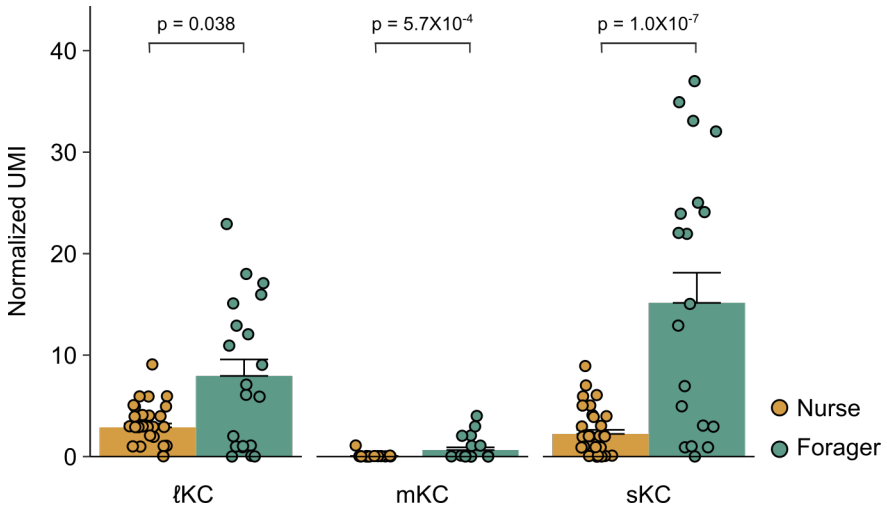

Fig. S18. Differential expression of *caveolin-3* in KC subtypes. The bar plot shows the expression of *caveolin-3* in LKC, mKC, and sKC across nurse and forager bees. Each dot represents a biological replicate (nurse: n = 34, forager: n = 20). Significance assessed by two-sided quasi-likelihood F-tests in edgeR. Bars indicate the mean expression + SEM. Source data are provided as a Source Data file.