Cell Reports, Volume *18*

Supplemental Information

Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity

Renchao Chen, Xiaoji Wu, Lan Jiang, and Yi Zhang

Supplemental Figures

Figure S1

Figure S1 – Related to Figure 1. Workflow of single-cell RNA-seq of mouse hypothalamus

(**A**) Schematic diagram showing the hypothalamic regions used for sample preparation. Adult mouse brain was first cut into 1mm-thick coronal sections and then hypothalamic tissues (shown in red contours) were dissected from 4 successive slices along the rostral-caudal axis.

(**B**) Box plots showing the distribution of gene/transcript number detected in each single cell.

(**C**) Workflow of cell type classification. The entire dataset was analyzed to identify 3,319 cells with > 2000 different transcripts in each cell, which were then subjected to R package Seurat for classification. After each round of clustering, the cells in the largest cluster were subject to next round cluster analysis for a total of four rounds. All cell clusters were then pooled together and clusters with less than 10 cells, or representing double-droplets or without a marker identified or out of hypothalamus were filtered out. At the end, 45 cell clusters with distinct transcriptional features were identified.

(**D**) tSNE plots showing the results of different rounds of clustering. Distinct clusters are shown with different colors in each round.

(**E**) SC3 and Seurat generated similar clustering results. Heap map showing the pair-wise comparison of the clusters generated using Seurat and SC3. The x-axis represents clusters generated with SC3 and the y-axis represents clusters generated with Seurat. The number in each intersection represents the overlap between the two clustering results, which is color-coded.

(**F**) Distribution of cells from different treatments and different batches. tSNE plot showing the distributions of cells from different treatments (left panel) and different experimental batches (right panel). Different treatments or batches are represented with different colors. The two clusters indicated with arrows only contain cells from one control animal, which are excluded in our final clustering result because they were derived from brain regions out of hypothalamus.

Figure S2

1 POPC
2 OPC
3 NFO
4 MO
5 Astro
6 Ependy
7 Tany
8 Endo 1
9 Endo 2
10 Micro
11 Macro

-2 0 2

Figure S2 – Related to Figure 2. Non-neuronal cell clusters feature distinct gene expression patterns

Heatmap showing the cell type-specific genes are differentially expressed across the 11 non-neuronal subtypes. Differentially expressed genes with power > 0.4, fold change > 2 among the 11 non-neuronal cell clusters were used to generate the heatmap. Columns represent individual cells and rows represent individual genes. The gene expression level is color-coded. POPC: proliferating oligodendrocyte progenitor cell; OPC: oligodendrocyte progenitor cell; NFO: newly formed oligodendrocytes; MO: myelinating oligodendrocyte; Astro: astrocyte; Ependy: ependymocyte; Tany: tanycyte; Endo: endothelial cell; Micro: microglia; Macro: macrophage.

Figure S3

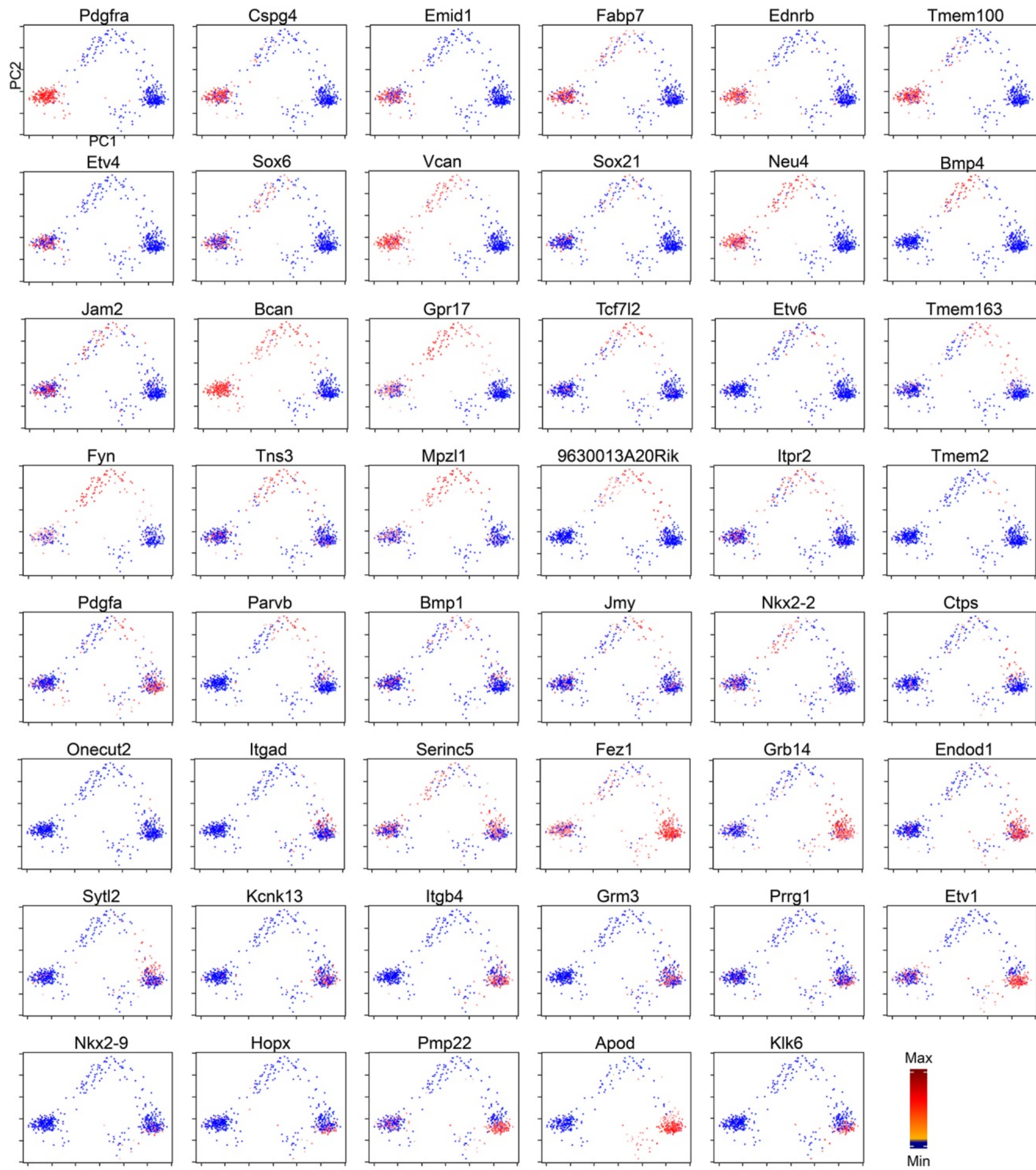


Figure S3 - Related to Figure 3. Expression patterns of stage-specific genes during oligodendrocyte maturation

tSNE plots showing the expression patterns of stage-specific genes during oligodendrocyte maturation. The expression level is color-coded. For each gene, if the expression level in a cell is less than 5% of the max value of that gene in the whole cell population, the gene is considered as not expressed in that cell and is represented as a blue dot. The results are very similar to that obtained in a recent study (Marques et al., 2016).

Figure S4

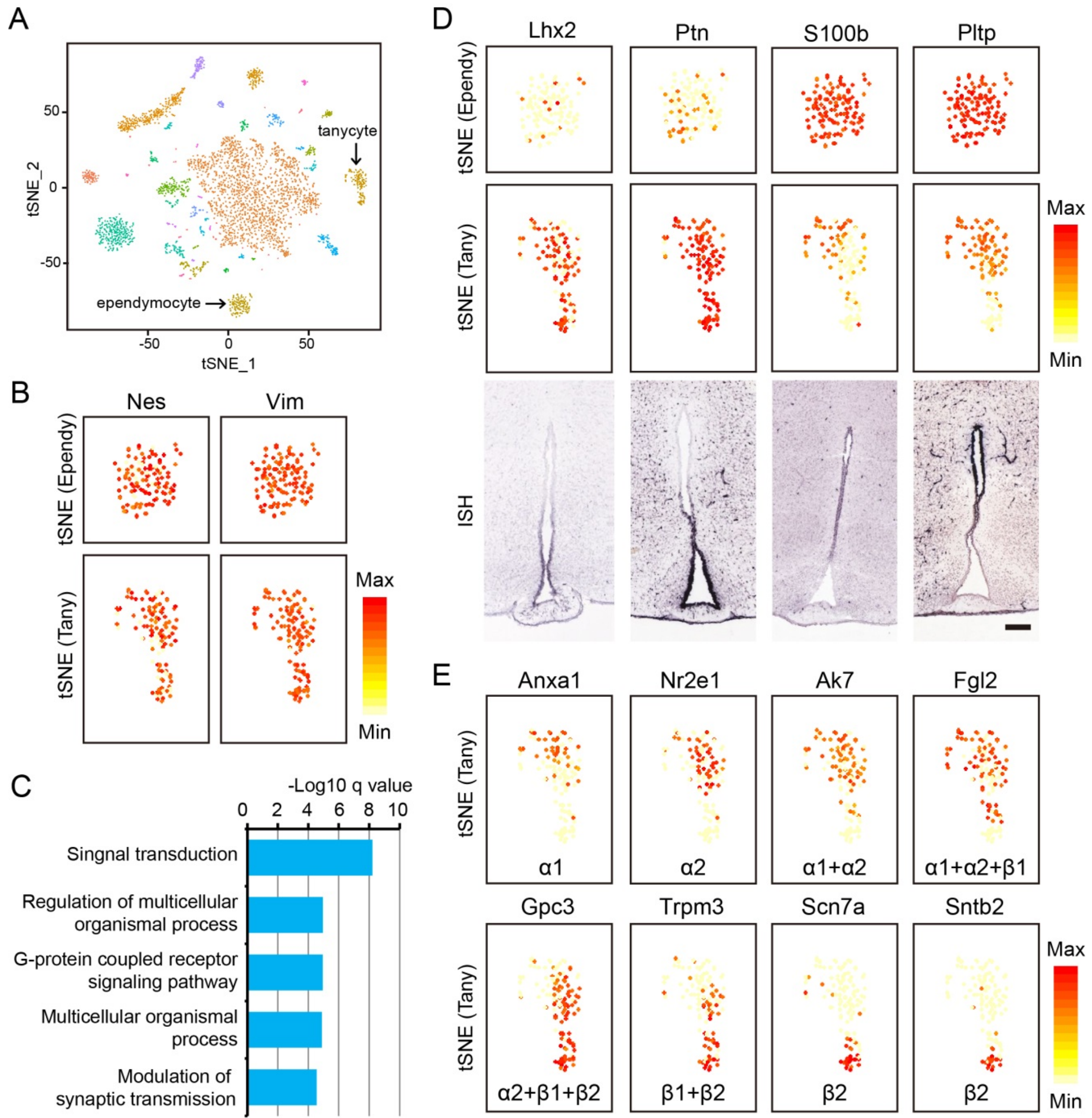


Figure S4 - Related to Figure 4. Transcriptional features and gene expression heterogeneity of tanycytes

(**A**) tSNE plot presentation of the 3319 hypothalamic cells based on their transcriptomes indicates that tanycyte form a cell cluster that can be easily distinguished from the ependymocyte cluster.

(**B**) tSNE plots showing that *Nes* (nestin) and *Vim* (vimentin) are highly expressed in both ependymocyte (Ependy) and tanycyte (Tany). Gene expression level is color-coded.

(**C**) Enriched terms of the tanycyte cluster revealed by GO analysis. The hypergeometric-test was used.

(**D**) Expression patterns of selected tanycyte- and ependymocyte-specific genes. tSNE plots (upper and middle panels) and ISH data (lower panels, from Allen Brain Atlas) showing the differential expression patterns of genes in tanycytes (*Lhx2* and *Ptn*) and ependymocytes (*S100b* and *Pltp*). Gene expression level is color-coded. Scale bars, 200 μm.

(**E**) tSNE plots showing the expression of selected genes enriched in different subsets of tanycytes. The genes are ordered according to their expression level along the vertical axis of the tSNE map. For each gene, corresponding tanycyte subtype(s) are listed. Gene expression level is color-coded.

Figure S5

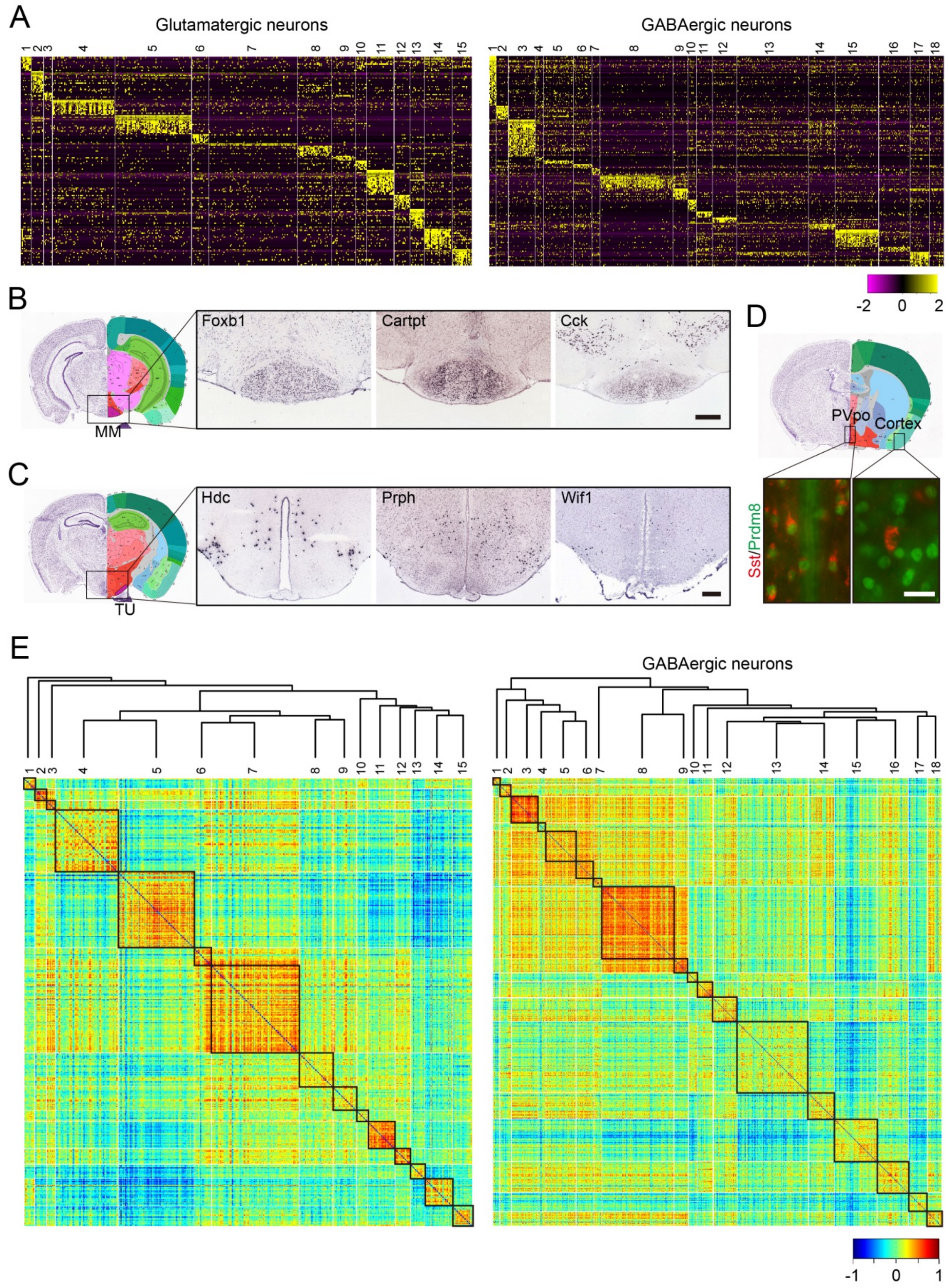


Figure S5 - Related to Figure 5. Distinct expression patterns of hypothalamic neuronal cluster-specific genes

(**A**) Heatmaps showing the expression pattern of cell-type specific genes in the 15 glutamatergic and 18 GABAergic neurons clusters. Differentially expressed genes with power > 0.4, fold change > 2 among the glutamatergic or GABAergic clusters were used to generate the heatmap. Columns correspond to individual cells and the numbers above the heatmaps indicate cluster identity. Rows represent individual genes and expression level is color-coded.

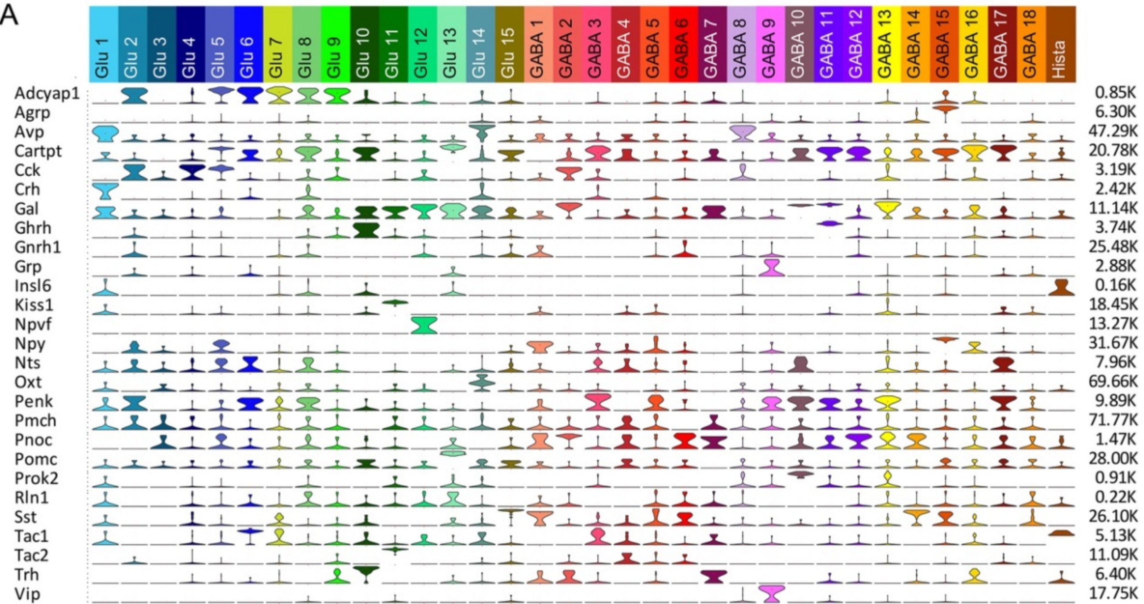
(**B**) ISH data (from Allen Brain Atlas) showing the expression of *Foxb1*, *Cartpt* and *Cck* in MM. The contour region in the left panel is enlarged and shown in the right three panels. MM, mammillary body. Scale bars, 300 µm.

(**C**) ISH data (from Allen Brain Atlas) showing the expression of *Hdc*, *Prph* and *Wif1* in TU. The contour region in the left panel is enlarged and shown in the right three panels. TU, tuberal nucleus. Scale bars, 300 µm.

(**D**) Immunostaining showing co-localization of Sst and Prdm8 in PVpo but not in cortex. The lower panels represent corresponding boxed regions in the upper panel. Scale bars, 40 µm.

(**E**) Clustering relationship of the 15 glutamatergic and 18 GABAergic neuron subtypes. The dendrograms indicate the relatedness among neuron subtypes based on gene expression. Differentially expressed genes with SD > 2 are used to generate the hierarchical clustering trees and cell-cell similarity heatmaps. The numbers on the top of the heatmaps indicate cluster identity.

Figure S6



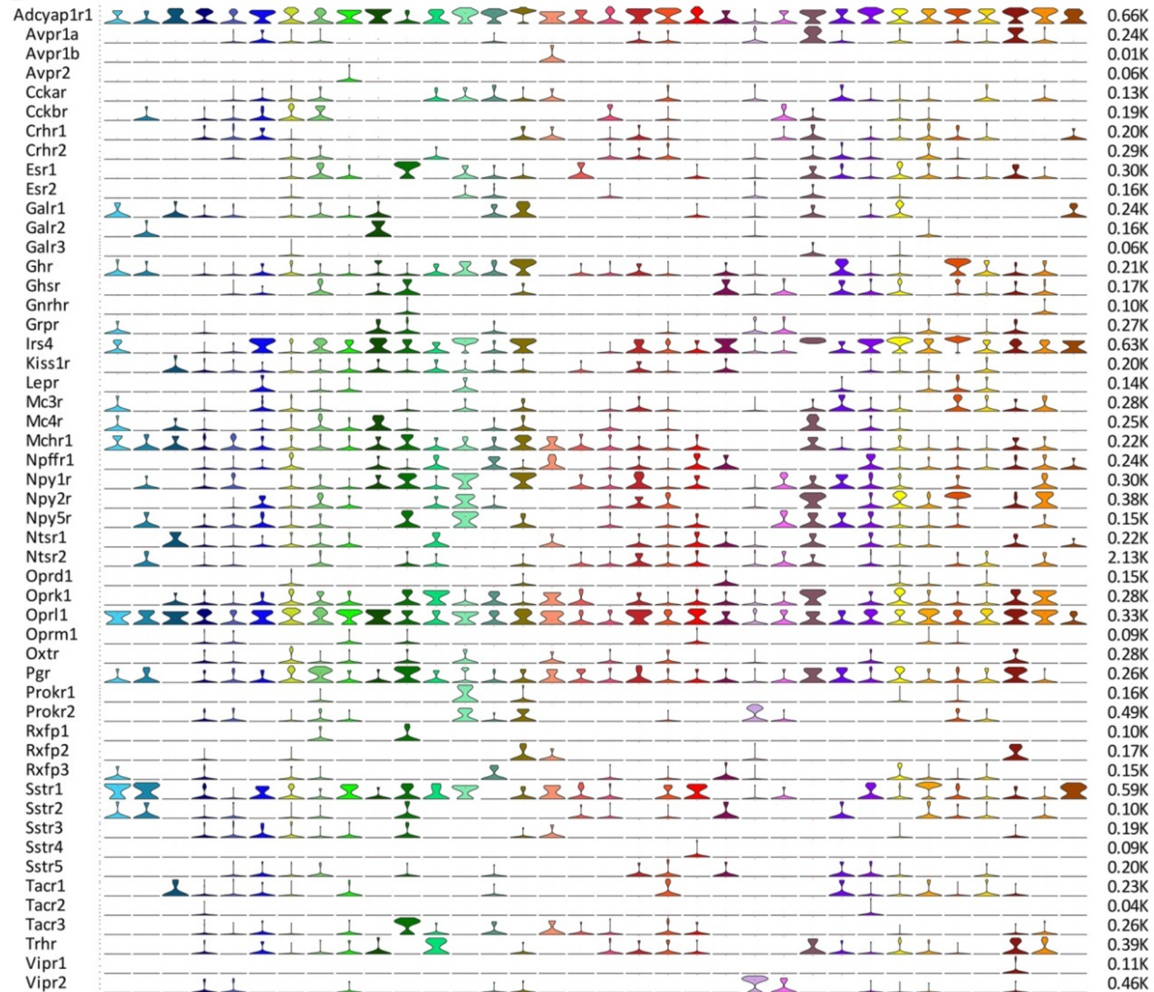


Figure S6 - Related to Figure 6. Expression of neuropeptides and receptors across neuronal subtypes

(**A**) Violin plot showing the expression profile of different neuropeptides (rows) across the hypothalamic neuronal subtypes (columns). Gene expression level is shown on a linear scale and adjusted for each gene with the maximum TPM value indicated on right.

(**B**) Violin plot showing the expression profile of different neuropeptide receptors (rows) among the hypothalamic neuronal subtypes (columns).

Figure S7

A

Glu5: 2900097C17Rik, Hcn1, Deb1, Tmp1, Hivep2, Asna1
Glu8: Higd2a, Puf60, H2afz, Aldoc, Rbfox2, Stk25
Glu12: Ass1, Fis1, Mrpl53, Prdx6, Psma1, Tmsb10
GABA1: Trappc6b, Mgst3, 2510002D24Rik, Vdac2, Gpm6a, Cope
GABA11: Pdgfa, Selm, Ctxn1, Cirbp, Eif2b3, Dbc1
GABA15: Txndc12, Lgmn, Arf5, Clstn3, Ram18b, Faim2
GABA18: Nmt1, Ddost, Vamp2, Cops6, Ubxn6, Psma4

log2(TPM+1)

Fed FD

B

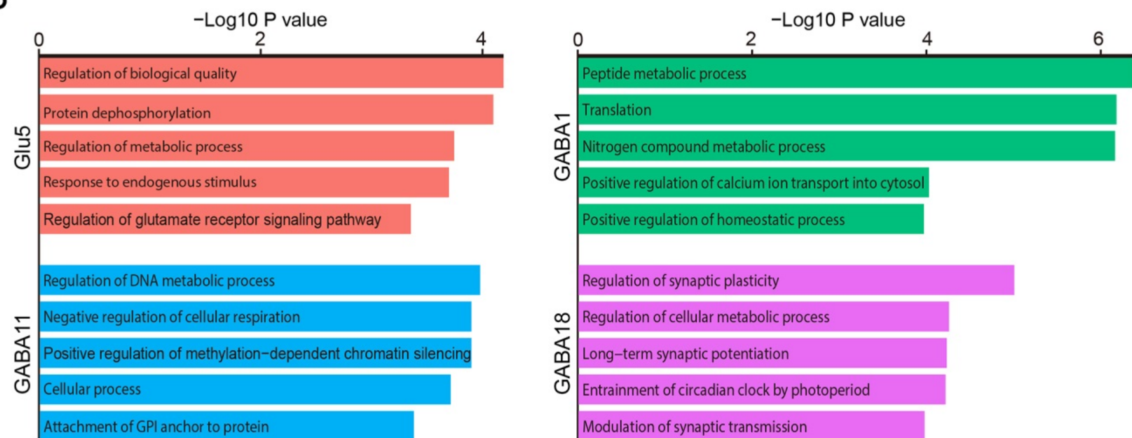


Figure S7 - Related to Figure 7. Food deprivation-induced transcription changes in different hypothalamic neuron clusters

(**A**) Violin plots showing representative genes in different neuronal subtypes that are induced by food deprivation. Each column represents one neuron cluster. Gene expression level is shown on a log scale. For the fold change and p-value of each gene, see Table S7. Fed, normal feeding; FD, food deprivation.

(**B**) GO enrichment analysis of representative neuronal clusters that exhibit high numbers of gene transcriptional changes. The hypergeometric-test was used. Also see Table S8.

Supplemental Tables

Table S1 - Related to Figure 1 - Summary of the clustering results

Table S2 - Related to Figure 3 - Six groups of genes dynamically expressed during oligodendrocyte differentiation

Table S3 - Related to Figure 3 - GO results of the six groups of genes with dynamic expression during oligodendrocyte differentiation

Table S4 - Related to Figure 4 - Differentially expressed genes in tanycyte and ependymoctye

Table S5 - Related to Figure 5 - Maker genes of neuronal clusters

Table S6 - Related to Figure 7 - Summary of scRNA-seq comparing transcriptome of hypothalamic neuron clusters from normal and food-deprived mice

Table S7 - Related to Figure 7 - GO analyses of genes in different hypothalamic neuron subtypes affected by food deprivation

Supplemental Experimental Procedures

Animals

All animal experiments followed the guidelines of the Institutional Animal Care and Use Committee at Harvard Medical School. Young adult female (8 - 10 weeks) B6D2F1 mice (C57B6 female × DBA2 male) were used. One day before the experiments, each animal was separated into individual fresh cages. For 24 h food deprivation treatment, only water was provided. In total 4 control animals and 3 food-deprived mice were used for single-cell RNA sequencing. All single-cell RNA-seq experiments were performed in five batches within 3 months. Each animal was processed separately (cell dissociation, library preparation) and regarded as a biological replicate.

Tissue dissection and dissociation

For hypothalamus dissection, the mice were anesthetized and the entire brain was removed and transferred into ice-cold Hibernate A/B27 medium (60 ml Hibernate A medium with 1 ml B27 and 0.15 ml Glutamax). The coronal sections from Bregma -0.22 to – 3.16 mm were then cut with brain matrix and further sliced into 1 mm slices. Hypothalamic tissue was then dissected from each slice under dissection microscope and subjected to tissue dissociation (Figure S1A). The hypothalamic tissues were dissociated into single-cell suspension using a papain-based dissociation protocol (Brewer and Torricelli, 2007) with some modifications. Briefly, the hypothalamic tissues from each mouse were cut into small pieces and incubated in Hibernate A-Ca medium with 2 mg/ml papain at 30^0C for 40 min with constant agitation. After washing with 5 ml Hibernate A/B27 medium, the tissues were triturated with fire polished glass Pasteur pipettes into a single-cell suspension within 6 ml Hibernate A/B27 medium. To remove debris, the single-cell suspension was loaded on a 4-layer OptiPrep gradient and centrifuged at 800 g for 15 min at 4^0C. Fractions 2 – 4 were then collected and washed with 5 ml Hibernate A/B27 medium and 5 ml DPBS with 0.01% BSA. In some experiments (batch 4 and 5), only fraction 3 is collected to enrich neurons. The cells were spun down at 200 g for 3

min and re-suspended in 0.4 ml DPBS with 0.01% BSA. A 10 µl cell suspension was stained with Trypan Blue and the live cells were counted. During the entire procedure, the tissues or cells were kept in ice-cold solutions except for the papain digestion.

Single cell capture, library preparation, and sequencing

Single cells and barcoded beads were captured into nanoliter-sized droplets as previously described (Macosko et al., 2015). The hypothalamic cell suspension was diluted to 100 cells/µl with DPBS containing 0.01% BSA and 0.6 – 1 ml cell suspension was loaded in each experiment. After cell capture, reverse transcription, cDNA amplification and sequencing library preparation were perform as described (Macosko et al., 2015). Sequencing was performed using Illumina Hiseq 2500 sequencer. Raw sequencing reads were analyzed using the Drop-seq software (Macosko et al., 2015) (http://mccarrolllab.com/dropseq/). Transcript count for each gene was converted to transcripts per million (TPM) for downstream analysis.

Cell clustering

The R package Seurat was used for cell clustering analysis (Macosko et al., 2015). To determine the optimal cut-off that balances data quality and cell numbers, we tried different cut-offs (detect expression of 800, 2000, and 2500 genes in each single cell) and found that 2000-gene cut-off gave the best results, thus we used the 3319 cells with 2000 or more gene expression detectable for clustering analysis. The highly variable genes were identified from these cells using Seurat with the default setting followed by principle component analysis (PCA). Then the statistically significant PCs ($p<0.05$) were used for 2-dimension tSNE. Based on the tSNE map, density-based clustering (DBSCAN) was used to cluster cells based on their proximity (G.use=3), resulting in 40 clusters with a large neuronal cluster containing 1574 cells. We reasoned that the PCs used for the first round clustering might be mainly comprised of non-neuronal cells and some specific neuronal subtypes. The majority of neurons, despite having subtype-specific genes, may have more closely related transcriptomes that form the large neuronal cluster. To separate the large cluster, we extracted cells within the cluster for further clustering using the

same strategy described above. The same analysis was repeated for an additional two times. After four rounds of clustering, a total of 73 cell clusters were identified.

Filter the initial clustering results

We applied the follow criteria to filter out initial clustering results: 1) marker genes were identified for all of the clusters with the function markers.all in R package Seurat; If a positive marker (marker gene enriched in a certain cluster compared with other clusters) cannot be found for a cluster, the cluster was excluded; 2) If a cluster co-expresses both non-neuronal and neuron markers, or non-neuronal markers of different subtypes, or glutamatergic and GABAergic markers (*Slc17a6* and *Slc32a1*), then the cluster is excluded as the cluster likely represent double-droplets; 3) Clusters with less than 10 cells were excluded, which can further remove very small neuron clusters which may represent neuron-neuron double-droplets; 4) By manually checking the expression pattern of makers for each cluster in mouse brain (Allen Brain Atlas), clusters from brain regions out of hypothalamus (e.g. thalamus) were excluded. This criterion was used to exclude non-hypothalamic cell clusters from non-hypothalamic tissue. After filtering, a total of 45 cell clusters were identified.

Seurat and SC3 clustering results comparisom

We applied SC3 to classify the 3319-cell dataset, which calculates consensus for each value of k and averages the clustering results of k-means using a consensus approach (Kiselev et al., 2016). Briefly, UMI counts of 3319-cell dataset were considered as count data to construct a new object with R package scater. QC metrics were computed for the created object and data filtering was performed with default parameters. To get clustering results with SC3, we started with function sc3_prepare. This method prepares an object of SCESet class for SC3 clustering. Function sc3_calc_distscalculates in SC3 package calculates the distances, including Euclidean, Pearson and Spearman distances. Function sc3_calc_transfs in SC3 package calculates transformation of the distance matrices corresponding to PCA and graph Laplacian transformations. K-means was then performed on the transformed distance matrices with function sc3_kmeans in SC3 package. Function Sc3_calc_consens in SC3 package calculates consensus matrices

based on the clustering solutions. To estimate the best K, function sc3_estimate_k in SC3 package which utilizes the Tracy-Widom theory on random matrices was used. The preliminary analysis with optimal K=27 shows that the majority of the cluster are not stable based on Silhouette plot or contain subpopulation structure based on manually checking the consensus matrix. Then a wide range of K from 27 to 150 was used to get different clustering solutions. For a given cluster from Seurat the maximum recovery rate was estimated across the different K values. Then a non-overlapping maximum recovered consensus cluster solution was generated as final clustering results. By calculating the overlap of cells across different clusters generated by Seurat and SC3, we found these two methods generate very similar clustering results (**Figure S1E**).

Predict cell identities for each of the 14,437 sequenced cells with Svm

Function markers.all in R package Seurat was used to identify cluster marker genes using the 3319 cell-derived dataset (ROC test). All marker genes with power less than 0.4 were discarded. The 3319 cells were fed to the function svm as the training dataset to build a classifier based on 3587 marker genes. Then function prediction was used to assign each of the 14437 cells (>= 800 transcripts in each cell) to one cluster based on transcriptional similarity. A total of 14437 cells were then subject to identified marker genes using function markers.all in R package Seurat. In total 1430 genes with at least 5-fold change were identified.

Analysis of the oligodendrocyte and tanycyte clusters

To analyze the timing of differentiation in oligodendrocytes, we extracted cells belonging to oligodendrocyte precursor cell, newly formed oligodendrocyte and myelinating oligodendrocyte clusters from the 3319-cell dataset (2000-gene cut-off). The three subpopulations were pooled for analysis. The 768 most variable genes among all the single cells were identified by Seurat (ROC test, power>0.4) (Macosko et al., 2015). A pseudo developmental timeline of single cells was then calculated with the package Monocle (Trapnell et al., 2014), using the most variable genes as time ordering genes. Based on the established differentiation direction of oligodendrocyte (from OPC to NFO to MO), the direction of pseudotime axis was determined. In Figure 3C, the six groups of

genes displaying six temporal patterns during the three developmental stages were identified by Seurat (ROC test, power>0.4). GO Term analysis of these six groups of genes was performed using Gorilla (Eden et al., 2009). All analyses for tanycyte and ependymal cells were based on the 3319-cell dataset. Identification of the differentially expressed genes between tanycyte and ependymocyte were carried out using Seurat (ROC test, power>0.4) (Macosko et al., 2015). GO term analysis were performed using Gorilla (Eden et al., 2009). Identification of the differentially expressed genes within tanycyte population and principle component analysis (PCA) were performed with Seurat (Macosko et al., 2015).

tSNE plot, dendrogram and heatmap for glutamatergic and GABAergic clusters

Neurons belonging to clusters Glu1-15 and GABA1-18 were extracted from the 3319-cell dataset and new Seurat objects were built using function setup for glutamatergic neurons and GABAergic neurons, respectively. The marker genes were identified by using the function find_all_markers of Seurat package (ROC test, requiring power > 0.4, fold change > 2), which was used for constructing a gene expression matrix for tSNE plot and heatmap. Rtsne with parameters of 'pca = TRUE, max_iter = 2000, perplexity = 30' was used to build a 2D map for visualization. Each dot represented a single cell and each cluster was colored with scheme consistent with other figures. Heatmaps in Figure S5A was generated by function doHeatMap. To generate the dendrograms presented in Figure S5E, variable genes with SD > 2 within glutamatergic or GBABergic neurons were selected, and the mean values of those genes in each cluster were then calculated. Hierarchical clustering was performed with R function dist and hclust (method="ward.D2") based on mean values of variable genes. To generate the cell-cell similarity heatmap in Figure S5E, the gene set which were used to generate Figures 5A and 5B were used. The expression TPM value of each gene was transformed to log2 value and scaled. Spearman correlation of each pair of cells was calculated by R function RCORR. Then R function heatmap.2 was used to generate the heatmap.

Identification of food deprivation affected genes

Base on the clustering results of 14,437 cells, within each neuronal cluster, the number of cells from "Fed" (ad libitum fed) and "FD" (food-deprived) mice are calculated. Clusters have at least 4 cells from both "Fed" and "FD" group are subjected to further analysis. lrTest function (likelihood ratio test) in R package MAST was used to calculate P value with "hurdle" model for comparing gene expression level. Genes with fold change > 1.5 and P value < 0.01 are defined as differentially expressed genes. GO analysis are performed with Gorilla (Eden et al., 2009) with default parameters. All gene symbols in mouse genome are used as background gene set. Redundant GO terms are manually removed.

Immunostaining

Anesthetized mice (8 -10 week) were perfused with PBS followed by 4% paraformaldehyde (PFA) in PBS. The whole brain was removed and fixed in 4% PFA overnight followed by 30% sucrose in PBS for 24h at 4^0C. Coronal sections were cut at 30 µm with cryostat and stored in PBS before use. A free-floating procedure was used for immunostaining. Briefly, brain slices containing the regions of interest were blocked in 5% BSA and 0.3% Triton-X 100 in PBS for 1 h at room temperature before being incubated in primary antibodies in blocking buffer overnight at 4^0C. After washing 3 times for 15 min in PBS, sections were incubated in blocking buffer containing secondary antibodies and DAPI at room temperature for 2 h, followed by washing 3 times with PBS. Brain slices were then attached to slides and coverslips were applied. The Sst antibody (rabbit, 1:100) was from Invetrogen, the Pax6 antibody (rabbit, 1:300) was from Covance, the Prdm8 antibody (guinea pig, 1:1000) was a gift from Dr. Sarah E. Ross (Ross et al., 2012), the Agrp antibody (goat, 1:100) was from Santa Cruz, the Cirbp (goat, 1:100), Crabp1 (mouse, 1:200) and Trim28 (mouse, 1:100) antibodies were from Abcam.