Reproducibility in the Application to the Mouse Hypothalamic Preoptic Region MERFISH Data

This file contains instructions for reproducing the results and figures in the application to the mouse hypothalamic preoptic region MERFISH data. The codes are in "code" folder. Please set the working directory to the source file location ("code" folder).

The following code files can be directly run as their input data are in the "input\_data" and "result\_data" folders: Step3\_BACT.R and Step8\_Figure3\_S1.R.

**Data**

The mouse hypothalamic preoptic region MERFISH data consist of five tissue sections from the mouse hypothalamic preoptic region collected by Chen et al. (2015). MERFISH\_0.19 is used for visualizing the cell typing performances of all methods. The data file “MERFISH\_0.19.h5ad” is downloaded from the link

<http://sdmbench.drai.cn/tcm/download/?file_path=/mnt/JINGD/data/file/sdmbench/db/MERFISH_0.19.h5ad>.

In the data preprocessing procedure, since the raw data have been normalized, we directly took log and selected 50 top principal components.

The preprocessed data includes:

* "gene\_matrix\_raw.csv": The raw ST data matrix saved from the h5ad file.
* "coordinates.csv": The spatial coordinates saved from the h5ad file.
* "adata\_obs.csv": The cell information file saved from the h5ad file.
* "coord\_and\_pc.RData": The spatial coordinates and preprocessed gene data matrix.

**Code**

**Step 1: "Step1\_read\_data.py"**

Read data from the h5ad file. The output of this step is "gene\_matrix\_raw.csv," "coordinates.csv," and "adata\_obs.csv."

**Step 2: "Step2\_preprocessing.R"**

Preprocess the ST raw count data matrix using the function DataPreprocess() in the R package BACT. The output of this step is "coord\_and\_pc.RData."

**Step 3: "Step3\_BACT.R"**

Implement BACT on the data from Step 2, and conduct the MCMC posterior inference. The output of this step is "merfish\_0.19\_result\_BACT.csv" which saves the estimated cluster labels for all the cells.

**Step 4: "Step4\_SpaGCN.R"**

Implement SpaGCN model. The output of this step is "merfish\_0.19\_result\_SpaGCN.csv".

In the "merfish\_0.19\_result\_SpaGCN.csv":

* "coord\_x": first dimension coordinate.
* "coord\_y": second dimension coordinate.
* "pred": cell typing labels.

**Step 5: "Step5\_STAGATE.py"**

Implement STAGATE model. The output of this step is "merfish\_0.19\_result\_STAGATE.csv".

In the "merfish\_0.19\_result\_STAGATE.csv":

* "coord\_x": first dimension coordinate.
* "coord\_y": second dimension coordinate.
* "mclust": cell typing labels obtained by R package mclust.

**Step 6: "Step6\_BANKSY.py"**

Implement BANKSY model. The output of this step is "merfish\_0.19\_result\_BANKSY.csv".

In the "merfish\_0.19\_result\_BANKSY.csv":

* "coord\_x": first dimension coordinate.
* "coord\_y": second dimension coordinate.
* "mclust": cell typing labels obtained by R package mclust.

**Step 7: "Step7\_BASS.R"**

Implement BASS model. The output of this step is "merfish\_0.19\_result\_BASS.csv".

In the "merfish\_0.19\_result\_BASS.csv":

* "x": first dimension coordinate.
* "y": second dimension coordinate.
* "c": estimated cell typing labels obtained via the MCMC samples.
* "z": estimated region labels obtained via the MCMC samples.

**Step 8: "Step8\_Figure3\_S1.R"**

Draw Figure 3 in the manuscript and Supplementary Figure S1 in the supplementary materials using the cell typing results of all methods.