Reproducibility in the Application to the Mouse Visual Cortex STARmap\* Data

This file contains instructions for reproducing the results and figures in the application to the mouse visual cortex STARmap\* 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\_Figure2.R.

**Data**

The mouse visual cortex STARmap\* data is collected by Wang et al. (2018). The data file “STARmap\_star.h5ad” is downloaded from the link

<http://sdmbench.drai.cn/tcm/download/?file_path=/mnt/JINGD/data/file/sdmbench/db/STARmap_20180505_BY3_1k.h5ad>,

and the cell type annotation file “cell\_annotation\_file.txt” is downloaded from the link

<https://drive.google.com/drive/folders/1I1nxheWlc2RXSdiv24dex3YRaEh780my?usp=sharing>

provided by the tutorial of STAGATE (Dong and Zhang, 2022):

<https://stagate.readthedocs.io/en/latest/T9_STARmap.html>.

In the data preprocessing procedure, we log-normalized the count data and selected 50 top principal components.

The preprocessed data includes:

* "gene\_matrix\_raw.csv": The raw ST count data 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 "starmap\_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 "starmap\_result\_SpaGCN.csv".

In the "starmap\_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 "starmap\_result\_STAGATE.csv".

In the "starmap\_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 "starmap\_result\_BANKSY.csv".

In the "starmap\_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 "starmap\_result\_BASS.csv".

In the "starmap\_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\_Figure2.R"**

Draw Figure 2 in the manuscript using the cell typing results of all methods.