Readme document for manuscript

“Identification of Cell-Type-Specific Spatially Variable Genes Accounting for Excess Zeros”

1. **Data**
   1. RealApplication

The data are publicly available for download via the online data portal at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111672. No registration is required.

Citation: Moncada R, Barkley D, Wagner F, Chiodin M et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol 2020 Mar;38(3):333-342.

1. **Code**
   1. Abstract

All of the data preprocessing and analysis in this paper were completed using R. The code is provided to conduct preprocessing on the raw data, apply proposed CTSV method, compared with SPARK-X, SPARK, SpatialDE, BOOST-GP, SOMDE, and trendsceek, and generate descriptive plots.

* 1. Description

All of the R scripts are available as the supplementary code.

License information: MIT.

For R and R packages, we use R version 3.6.2 (2019-12-12, “Dark and Stormy Night”). The used R packages are:

* readr, version 1.3.1 (https://CRAN.R-project.org/package=readr)
* dplyr, version 0.8.5 (https://CRAN.R-project.org/package= dplyr)
* plotly, version 4.9.2 (https://CRAN.R-project.org/package=plotly)
* stringr, version 1.4.0 (https://CRAN.R-project.org/package= stringr)
* grid, version 3.6.2 (https://CRAN.R-project.org/package= grid)
* ggpubr, version 0.3.0 (https://CRAN.R-project.org/package= ggpubr)
* gridExtra, version 2.3 (https://CRAN.R-project.org/package= gridExtra)
* gplots, version 3.0.3 (https://CRAN.R-project.org/package= gplots)
* ggplot2, version 3.3.0 (https://CRAN.R-project.org/package= ggplot2)
* umap, version 0.2.5.0 (https://CRAN.R-project.org/package= umap)
* Rcpp, version 1.0.4 (https://CRAN.R-project.org/package= Rcpp)
* RcppArmadillo, version 0.9.850.1.0 (https://CRAN.R-project.org/package= RcppArmadillo)
* aricode, version 1.0.0 (https://CRAN.R-project.org/package= aricode)

The computing platform was used for the analyses in this paper. The details of the computing platform are:

* Operating system: CentOS 6.6
* CPU: Intel Xeon E5 2650 2.2G Hz
* RAM: 64GB
  1. Instructions for Use

All data preprocessing and analysis as well as Figures 2 and 3 in the manuscript can be reproduced.

Detailed workflow information is contained in the "README.pdf" in "Simulation", "RealApplication\_A" and "RealApplication\_B" directories.

The general steps in the simulation are:

1. Generate the data and apply the proposed model to the data.

2. Generate figure 2 in the paper.

The general steps in the real application 1 are:

1. Conduct data preprocessing.

2. Apply the proposed model to the preprocessed data.

3. Generate the PDAC-A parts of figure 3 in the paper.

The general steps in the real application 2 are:

1. Conduct data preprocessing.

2. Apply the proposed model to the preprocessed data.

3. Generate the PDAC-B parts of figure 3 in the paper.