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 4.0.2 (2020-06-22). The used R packages are:

* MASS, version 7.3-54 (https://CRAN.R-project.org/package= MASS)
* DIRECT, version 1.0.1 (https://CRAN.R-project.org/package= DIRECT)
* Rcpp, version 4.9.2 (https://CRAN.R-project.org/package= Rcpp)
* RcppDist, version 4.9.2 (https://CRAN.R-project.org/package= RcppDist)
* RcppArmadillo, version 1.4.0 (https://CRAN.R-project.org/package= RcppArmadillo)
* pscl, version 3.6.2 (https://CRAN.R-project.org/package= pscl)
* doSNOW, version 0.3.0 (https://CRAN.R-project.org/package= doSNOW)
* ggplot2, version 3.3.5 (https://CRAN.R-project.org/package= ggplot2)
* pheatmap, version 3.0.3 (https://CRAN.R-project.org/package= pheatmap)
* pROC, version 3.3.0 (https://CRAN.R-project.org/package= pROC)
* dplyr, version 1.0.7 (https://CRAN.R-project.org/package= dplyr)
* VennDiagram, version 1.0.4 (https://CRAN.R-project.org/package= VennDiagram)
* stringr, version 0.9.850.1.0 (https://CRAN.R-project.org/package= stringr)
* doParallel, version (https://CRAN.R-project.org/package= doParallel)
* foreach, version (https://CRAN.R-project.org/package= foreach)
* Seurat, version 1.0.0 (https://CRAN.R-project.org/package= Seurat)
* sctransform, version (https://github.com/ChristophH/sctransform)
* qvalue, version (http://bioconductor.org/packages/release/bioc/html/qvalue.html)
* SPOTlight, version (https://github.com/MarcElosua/SPOTlight)
* SPARK, version 1.1.1 (https://github.com/xzhoulab/SPARK)
* trendsceek, version (https://github.com/edsgard/trendsceek)
* SOMDE, version (https://github.com/XuegongLab/somde)
* sptialde, version (https://github.com/Teichlab/SpatialDE)

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

* Operating system: macOS Catalina 10.15.4
* CPU: 2.3 GHz 8-Core Intel Core i9
* RAM: 16GB
  1. Instructions for Use

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

Detailed workflow information is contained in the "README.pdf" in "Simulation" and "Real data analysis" directories.

The general steps in the simulation are:

1. Generate the data.
2. Apply the proposed model to the data.
3. Generate figure 2, figure 3, and figure 4 in the paper.

The general steps in the real application are:

1. Conduct data preprocessing.

2. Apply the proposed model and existing methods to the preprocessed data.

3. Generate the figures1, figure5, and figure6 in the paper.