Package 'SPADE'

March 22, 2023

Type Package

Title Spatially Expressed Pattern Investication Method With Spatial Transcriptomics Data

Version 1.0

Date 2023-03-22

Authors Fei Qin, Feifei Xiao, Guoshuai Cai

Maintainer Fei Qin <fqin@email.sc.edu>

Description
    We developed SPADE, a spatial pattern investigation method, to identify spatially expressed genes within groups and between groups using spatial transcriptomics data.

License GPL-3

LazyData TRUE

Depends R (¿= 4.0),
    mvtnorm,
    stats

URL https://github.com/thecailab/SPADE

RoxygenNote 7.1.2

VignetteBuilder knitr

Encoding UTF-8

Language en-GB

Imports Rcpp (¿= 1.0.8)

LinkingTo Rcpp, RcppArmadillo

Suggests rmarkdown

R topics documented:

ComputeGaussianPL   *Hyperparameter range.*

Description

The range for the hyperparameter in the Gaussian Kernel, which was calculated from the coordinate of all spots.

Usage

```
ComputeGaussianPL(X, compute_distance = TRUE)
```

Arguments

X   Cell coordinates matrix n x 2 or kernel matrix computed already

compute_distance

Compute the distance matrix using generic function dist, default=TRUE

Value

This function returns the range for the hyperparameter in the Gaussian Kernel.

lrang   The range for the hyperparameter in the Gaussian kernel

---

Delta_fit   *Estimation of the optimal delta*

Description

Estimation of the optimal delta in the variance function between groups.

Usage

```
Delta_fit(location, y, L)
```

Arguments

location   Spatial coordinates for all spots.
y   Read counts data for the gene.
L   The length-scale hyperparameter in the kernel function.

Value

This function returns the optimal log transformed delta value.

results   The optimal log delta value.

Get_Lambda	*Eigenvalue for Kernel matrix*

Description

Calculating the eigenvalue for the kernel matrix K.

Usage

```
Get_Lambda(K)
```

Arguments

K   Kernel matrix.

Value

This function returns eigenvalue for the Kernel matrix.

lambda   Eigenvalue for Kernel matrix.

lengthscale_fit	*Optimization of likelihood function within groups*

Description

Optimization of likelihood function within groups.

Usage

```
lengthscale_fit(location, y, L)
```

Arguments

location   Spatial coordinates for all spots.

y   Read counts data for a gene.

L   The length-scale hyperparameter in the kernel function.

Value

This function returns likelihood at the point of optimal delta.

results   The likelihood with the optimal delta value.

---

lengthscale_fit_DE   *Optimization of likelihood function between groups*

Description

Optimization of likelihood function between groups.

Usage

```
lengthscale_fit_DE(location1, location2, y1, y2, L)
```

Arguments

location1	Spatial coordinates for all spots in group 1.
location2	Spatial coordinates for all spots in group 2.
y1	Read counts data for the gene in group 1.
y2	Read counts data for the corresponding gene in group 2.
L	The length-scale hyperparameter in the kernel function.

Value

This function returns likelihood at the point of optimal delta.

results	The likelihood with the optimal delta value for identifying SE genes between groups.

LL   *Log likelihood function within groups*

Description

The log transformed multivariate normally distributed marginal likelihood to identify SE genes within groups.

Usage

```
LL(log_delta, UTy, UT1, S, n)
```

Arguments

log_delta	Log transformed delta in the variance function.
UTy	The expression UTy will need to be re-computed for each gene.
UT1	The expression UT1 only depends on the coordinates X and can be pre-computed and reused for every gene.
S	A diagonal matrix used in the spectral decomposition of the covariance.
n	The number of spots for each gene.

Value

This function returns the log likelihood for each delta.

LL_res	The log likelihood for each delta value.

LL_combin	*Log likelihood function between groups*

Description

The log transformed multivariate normally distributed marginal likelihood to identify SE genes between groups.

Usage

```
LL_combin(log_delta, UTy1, UTy2, UT1, UT2, S1, S2, n1, n2)
```

Arguments

log_delta	Log transformed delta in the variance function.
UTy1	The expression UTy will need to be re-computed for the gene in group 1.
UTy2	The expression UTy will need to be re-computed for the gene in group 2.
UT1	The expression UT1 only depends on the coordinates X and can be pre-computed and reused for the gene in group 1.
UT2	The expression UT1 only depends on the coordinates X and can be pre-computed and reused for the gene in group 2.
S1	A diagonal matrix used in the spectral decomposition of the covariance in group 1.
S2	A diagonal matrix used in the spectral decomposition of the covariance in group 2.
n1	The number of spots for each gene in group 1.
n2	The number of spots for each gene in group 2.

Value

This function returns the log likelihood for each delta.

LL_res	The log likelihood for each delta value.

LL_DE	*Log likelihood function*

Description

The log transformed multivariate normally distributed marginal likelihood .

Usage

```
LL_DE(log_delta, location, y, L)
```

Arguments

`log_delta`	Log transformed delta in the variance function.
`location`	Spatial coordinates for all spots.
`y`	Read counts data for the gene.
`L`	The length-scale hyperparameter in the kernel function.

Value

This function returns the log likelihood for each delta.

`LL_res`	The log likelihood for each delta value.

`mu_hat`	*The estimate of mu*

Description

The estimate of mean expression level for each gene.

Usage

```
mu_hat(delta, UTy, UT1, S)
```

Arguments

`delta`	Delta value in the variance function.
`UTy`	The expression UTy will need to be re-computed for each gene,
`UT1`	The expression UT1 only depends on the coordinates X and can be pre-computed and reused for each gene.
`S`	A diagonal matrix used in the spectral decomposition of the covariance.

Value

This function returns the estimate of mean expression level for each gene.

`mu_h`	The estimate of mean expression level for each gene.

SPADE_DE	*SPADE estimation and test between groups*

Description

Estimation of optimal parameter in the kernel function to identify SE genes between groups. P values will be provided based on a crossed likelihood ratio test.

Usage

```
SPADE_DE(readcounts1, readcounts2, location1, location2)
```

Arguments

readcounts1	Read counts data for identifying spatially expressed genes in group 1. Each row is a gene and each column is a spot.
readcounts2	Read counts data for identifying spatially expressed genes in group 2. Each row is a gene and each column is a spot.
location1	Spatial coordinates for all spots in group 1.
location2	Spatial coordinates for all spots in group 2.

Value

This function returns the estimated parameters and some statistics of the SPADE method.

geneid	Gene index.
theta_Gau1	The estimated optimal length-scale hyperparameter in the Gaussian kernel for group 1.
theta_Gau2	The estimated optimal length-scale hyperparameter in the Gaussian kernel for group 2.
logLik11	The corresponding log likelihood calculated with optimal hyperparameter estimated above for group 1.
logLik21	The corresponding log likelihood calculated with optimal hyperparameter estimated above for group 2.
logLik10	The log likelihood calculated with expression value from group 1 and optimal hyperparameter estimated for group 2.
logLik20	The log likelihood calculated with expression value from group 2 and optimal hyperparameter estimated for group 1.
Diff	The likelihood ratio test statistic is given by 2*(logLik11 + logLik21 - logLik10 - logLik20).
Pvalue	P values calculated with likelihood ratio test statistic using F test with degree freedom of one.
Adjust.Pvalue	Adjusted P values calculated with the Benjamini and Hochberg method.

SPADE_estimate   *SPADE estimation for hyperparameter within groups*

Description

Estimation of optimal hyperparameter in the kernel function to identify SE genes within groups.

Usage

```
SPADE_estimate(expr_data, info)
```

Arguments

expr_data	Read counts data for identifying spatially expressed genes. Each row is a gene and each column is a spot.
info	Spatial coordinates for all spots.

Value

This function returns the estimated parameters and some statistics of the SPADE method.

GeneID	Gene index.
theta_Gau	The estimated optimal length-scale hyperparameter in the Gaussian kernel
Lik_Gau	The log likelihood using optimal hyparameter estimated above.

Examples

```
Y <- matrix(rnorm(10000, 10, 2),100, 100)
info <- matrix(runif(200, 1, 100), 100, 2)
output <- SPADE_estimate(expr_data=Y, info=)
output
```

SPADE_norm   *SPADE normalization*

Description

Normlization of read counts data into continuous data.

Usage

```
SPADE_norm(readcounts, info)
```

Arguments

readcounts	Read counts data for identifying spatially expressed genes. Each row is a gene and each column is a spot.
info	Spatial coordinates for all spots.

Value

This function returns the normalized continuous data.

regdata   The normlized continuous data

SPADE_test   *SPADE test within groups*

Description

Test for identifying SE genes within groups

Usage

```
SPADE_test(object, location, para)
```

Arguments

object   Read counts data for identifying spatially expressed genes. Each row is a
                gene and each column is a spot.

location   Spatial coordinates for all spots.

para   Parameter data frame estimated from the SPADE_estimate function.

Value

This function returns the statistics and P values for the test

GeneID   Gene index.

Q   The Q statistics calculated using the SKAT method.

Pvalue   P value calculated from the Q statistics based on the Davies method.

Adjust.Pvalue   Adjusted P values calculated with the Benjamini and Hochberg method.

Index