Package: scBio (via r-universe)

September 10, 2024

Type Package

Title Single Cell Genomics for Enhancing Cell Composition Inference from Bulk Genomics Data

Version 0.1.6

Author Amit Frishberg [aut, cre], Naama Peshes-Yaloz [aut], Irit Gat-Viks [aut]

Maintainer Amit Frishberg <amfrishberg@gmail.com>

URL https://github.com/amitfrish/scBio

BugReports https://github.com/amitfrish/scBio/issues

Description Cellular population mapping (CPM) a deconvolution algorithm in which single-cell genomics is required in only one or a few samples, where in other samples of the same tissue, only bulk genomics is measured and the underlying fine resolution cellular heterogeneity is inferred.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

biocViews limma

Depends R (>= 2.10)

Imports sp, foreach, parallel, doSNOW, raster, fields, LiblineaR, limma

Repository https://amitfrish.r-universe.dev

RemoteUrl https://github.com/amitfrish/scbio

RemoteRef HEAD

RemoteSha 0b89273888c2a616976149bca515a8e649391dfe

Contents

`BulkFlu` *Gene expression profiles of flu and pbs sample.*

Description

A dataset containing the RNA-seq profiles of colaborative-cross (CC) mice 2 days after the infection with either the flu virus or PBS.

Usage

```
BulkFlu
```

Format

A matrix with 1858 rows (genes) and 74 columns (samples).

`CPM` *The Cellular Population Mapping (CPM) algorithm.*

Description

This function initiate the Cellular Population Mapping (CPM) algorithm - a deconvolution algorithm in which single-cell genomics is required in only one or a few samples, where in other samples of the same tissue, only bulk genomics is measured and the underlying fine resolution cellular heterogeneity is inferred. CPM predicts the abundance of cells (and cell types) ranging monotonically from negative to positive levels. Using a relative framework these values correspond to decrease and increase in cell abundance levels, respectively. On the other hand, in an absolute framework lower values (including negatives) correspond to lower abundances and vise versa. These values are comparable between samples.

Usage

```
CPM(SCData, SCLabels, BulkData, cellSpace, no_cores = NULL,
  neighborhoodSize = 10, modelSize = 50, minSelection = 5,
  quantifyTypes = F, typeTransformation = F, calculateCI = F)
```

Arguments

SCData	A matrix containing the single-cell RNA-seq data. Each row corresponds to a certain gene and each column to a certain cell. Importantly, CPM relies on many iterative processes and therefore might take a long running time. For extremely large single cell datasets, we suggest to use only a portion of the data, using random sampling of the cells.
SCLabels	A vector containing the labels of each of the cells.
BulkData	A matrix containing heterogenous RNA-seq data for one or more samples. Each row corresponds to a certain gene and each column to a certain sample.
cellSpace	The cell state space corresponding to the single-cell RNA-seq data. It can be a vector for a 1-dim space or a 2D matrix for a two space where each column represents a different dimension. The cell space should incorporate the similarities of cells within cell types. Similarities between cells from different cell types, based on the cell space, are not taken into account in CPM.
no_cores	A number for the amount of cores which will be used for the analysis. The defalt (NULL) is total number of cores minus 1.
neighborhoodSize	Cell neighborhood size which will be used for the analysis. This should be lower than the number of cells in the smallest cell type. The defalt is 10.
modelSize	The reference subset size in each iteration of CPM. This should be lower than the total number of cells. The defalt is 50.
minSelection	The minimum number of times in which each reference cell is selected. Increasing this value might have a large effect on the algorithm's running time. The defalt is 5.
quantifyTypes	A boolean parameter indicating whether the prediction of cell type quantities is needed. This is recommended only in the case of homogenicity within cell types. Cell types with high inner cellular variability will recieve less reliabe values. The default is FALSE.
typeTransformation	This parameter will have an effect only if quantifyTypes = TRUE. A boolean parameter indicating whether cell type deconvolution should be provided in fractions. This is done by substracting all cell types by values of the minimal cell type and dividing in their sum. This is not recommended, since it reduces the comparability between sample. The default is FALSE.
calculateCI	A boolean parameter indicating whether the calculation of confidence itervals is needed. The default is FALSE.

Value

A list including:

predicted	CPM predicted cell abundance matrix. Each row represents a sample and each column a single cell.
cellTypePredictions	CPM predicted cell-type abundance matrix. Each row represnts a sample and each column a single cell-type. This is calculated if quantifyTypes = TRUE.

confIntervals	A matrix containing the confidence iterval for each cell and sample. Each row represnts a sample and each column a single cell. This is calculated if calculateCI = TRUE.
numOfRuns	The number of deconvolution repeats preformed by CPM.

References

Frishberg, A., Peshes-Yaloz, N., Cohn, O., Rosentul, D., Steuerman, Y., Valadarsky, L., Yankovitz, G., Mandelboim, M., Iraqi, F.A., Amit, I. et al. (2019) Cell composition analysis of bulk genomics using single-cell data. Nature Methods, 16, 327-332.

Examples

```
data(SCLabels)
data(SCFlu)
data(BulkFlu)
data(SCCellSpace)

# Creating relative bulk data (Infleunza infection compared to PBS):
BulkFluReduced = BulkFlu - rowMeans(BulkFlu[,grep("pbs",colnames(BulkFlu))])
BulkFluReduced = BulkFluReduced[,grep("flu",colnames(BulkFluReduced))]

# Running CPM using only a single cell-type:
oneCellTypeIndexes = which(SCLabels == "MPS")
res = CPM(SCData = SCFlu[,oneCellTypeIndexes], SCLabels = SCLabels[oneCellTypeIndexes],
       BulkData = BulkFluReduced, cellSpace = SCCellSpace[oneCellTypeIndexes,], no_cores = 2)

## Not run:

# Running CPM using a variety of cell-types:
res = CPM(SCFlu, SCLabels, BulkFluReduced, SCCellSpace, no_cores = 2)
### Full multi-threading : CPM(SCFlu, SCLabels, BulkFluReduced, SCCellSpace)

## End(Not run)
```

SCCellSpace *Single-cell cell space.*

Description

A dataset containing a 2-dim cell space of all single-cells in the SCFlu dataset.

Usage

```
SCCellSpace
```

Format

A matrix with 349 rows (cells) and 2 columns (dimensions).

SCFlu — *Gene expression profiles of lung cells after influenza infection.*

Description

A dataset containing the RNA-seq profiles of lung cells from multiple cell types, taken from two mice 2 days after the infection with either the flu virus or PBS.

Usage

```
SCFlu
```

Format

A matrix with 1858 rows (genes) and 349 columns (cells).

SCLabels — *Single-cell classification into cell types.*

Description

A dataset containing the classification of each cell in the SCFlu dataset to a specific cell type.

Usage

```
SCLabels
```

Format

A vector with 349 values.

Index