Package: STREAK (via r-universe)

September 13, 2024

Type Package

Title Receptor Abundance Estimation using Feature Selection and Gene Set Scoring

Version 1.0.0

Maintainer Azka Javaid <azka.javaid.gr@dartmouth.edu>

Description Performs receptor abundance estimation for single cell RNA-sequencing data using a supervised feature selection mechanism and a thresholded gene set scoring procedure. Seurat's normalization method is described in: Hao et al., (2021) <doi:10.1016/j.cell.2021.04.048>, Stuart et al., (2019) <doi:10.1016/j.cell.2019.05.031>, Butler et al., (2018) <doi:10.1038/nbt.4096> and Satija et al., (2015) <doi:10.1038/nbt.3192>. Method for reduced rank reconstruction and rank-k selection is detailed in: Javaid et al., (2022) <doi:10.1101/2022.10.08.511197>. Gene set scoring procedure is described in: Frost et al., (2020) <doi:10.1093/nar/gkaa582>. Clustering method is outlined in: Song et al., (2020) <doi:10.1093/bioinformatics/btaa613> and Wang et al., (2011) <doi:10.32614/RJ-2011-015>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

LazyDataCompression xz

Imports Ckmeans.1d.dp, Matrix, Seurat, SPECK, stats, VAM

RoxygenNote 7.2.3

Depends R (>= 2.10)

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author H. Robert Frost [aut], Azka Javaid [aut, cre]

Repository CRAN

Date/Publication 2023-11-17 21:10:02 UTC

Contents

`receptorAbundanceEstimation` *Receptor abundance estimation for single cell RNA-sequencing (scRNA-seq) data using gene set scoring and thresholding.*

Description

Performs receptor abundance estimation for mxn scRNA-seq target data using gene set scoring and thresholding. scRNA-seq target counts are normalized and reduced rank reconstructed (RRR) using the `SPECK::randomizedRRR()` function. Gene set scoring is next performed leveraging expression from the top most weighted genes based on the gene sets weights membership matrix with the `VAM::vam()` function. The resulting cell-specific gene set scores are then thresholded utilizing the `Ckmeans.1d.dp::Ckmeans.1d.dp()` function. Note that this function only performs normalization and does not perform any quality control (QC) checks on the inputted target scRNA-seq counts matrix. Any QC needed can be performed on the target matrix before passing it as an input to the function.

Usage

```
receptorAbundanceEstimation(
  target.rnaseq,
  receptor.geneset.matrix,
  num.genes = 10,
  rank.range.end = 100,
  min.consec.diff = 0.01,
  rep.consec.diff = 2,
  manual.rank = NULL,
  seed.rsvd = 1,
  max.num.clusters = 4,
  seed.ckmeans = 2
)
```

Arguments

`target.rnaseq` mxn scRNA-seq counts matrix for m cells and n genes.

`receptor.geneset.matrix` nxh Gene sets weights membership matrix.

num.genes Number of top most weighted genes for subsequent gene set scoring and thresholding.

rank.range.end See documentation for the randomizedRRR function from the SPECK package.

min.consec.diff
See documentation for the randomizedRRR function from the SPECK package.

rep.consec.diff
See documentation for the randomizedRRR function from the SPECK package.

manual.rank See documentation for the randomizedRRR function from the SPECK package.

seed.rsvd See documentation for the randomizedRRR function from the SPECK package.

max.num.clusters
See documentation for the ckmeansThreshold function from the SPECK package.

seed.ckmeans See documentation for the ckmeansThreshold function from the SPECK package.

Value

- receptor.abundance.estimates - A mxh matrix consisting of abundance estimates for m cells in h receptors.

Examples

```
data("train.malt.rna.mat")
data("train.malt.adt.mat")
receptor.geneset.matrix.out <- receptorGeneSetConstruction(train.rnaseq =
                                        train.malt.rna.mat[1:100,1:80],
                                        train.citeseq =
                                        train.malt.adt.mat[1:100,1:2],
                                        rank.range.end = 70,
                                        min.consec.diff = 0.01,
                                        rep.consec.diff = 2,
                                        manual.rank = NULL,
                                        seed.rsvd = 1)
dim(receptor.geneset.matrix.out)
head(receptor.geneset.matrix.out)
data("target.malt.rna.mat")
receptor.abundance.estimates.out <- receptorAbundanceEstimation(target.rnaseq =
                                    target.malt.rna.mat[1:200,1:80],
                                    receptor.geneset.matrix =
                                    receptor.geneset.matrix.out,
                                    num.genes = 10, rank.range.end = 70,
                                    min.consec.diff = 0.01,
                                    rep.consec.diff = 2,
                                    manual.rank = NULL, seed.rsvd = 1,
                                    max.num.clusters = 4, seed.ckmeans = 2)
dim(receptor.abundance.estimates.out)
head(receptor.abundance.estimates.out)
```

`receptorGeneSetConstruction` *Gene sets weights membership matrix construction for receptor abundance estimation.*

Description

Computes nxh gene sets weights membership matrix using associations learned between log-normalized and reduced rank reconstructed (RRR) mxn scRNA-seq training data and mxh CITE-seq ADT training counts normalized using the centered log ratio (CLR) transformation. scRNA-seq counts are normalized and RRR using the `SPECK::randomizedRRR()` function while CITE-seq counts are normalized using the `Seurat::NormalizeData()` function with the `normalization.method` parameter set to `CLR`. Spearman rank correlations are computed between the normalized CITE-seq data and the normalized and RRR scRNA-seq data.

Usage

```
receptorGeneSetConstruction(
  train.rnaseq,
  train.citeseq,
  rank.range.end = 100,
  min.consec.diff = 0.01,
  rep.consec.diff = 2,
  manual.rank = NULL,
  seed.rsvd = 1
)
```

Arguments

`train.rnaseq`	mxn scRNA-seq counts matrix for m cells and n genes.
`train.citeseq`	mxh CITE-seq ADT counts matrix for m cells (same cells as the `train.rnaseq matrix`) and h cell-surface proteins.
`rank.range.end`	See documentation for the `randomizedRRR` function from the `SPECK` package.
`min.consec.diff`	See documentation for the `randomizedRRR` function from the `SPECK` package.
`rep.consec.diff`	See documentation for the `randomizedRRR` function from the `SPECK` package.
`manual.rank`	See documentation for the `randomizedRRR` function from the `SPECK` package.
`seed.rsvd`	See documentation for the `randomizedRRR` function from the `SPECK` package.

Value

- `receptor.geneset.matrix` - A nxh gene sets weights membership matrix where a column i from h corresponds to the weights for n genes from the scRNA-seq matrix trained against the corresponding CITE-seq ADT transcript h.

Examples

```
data("train.malt.rna.mat")
data("train.malt.adt.mat")
receptor.geneset.matrix.out <- receptorGeneSetConstruction(train.rnaseq =
                                        train.malt.rna.mat[1:100,1:80],
                                        train.citeseq =
                                        train.malt.adt.mat[1:100,1:2],
                                        rank.range.end = 70,
                                        min.consec.diff = 0.01,
                                        rep.consec.diff = 2,
                                        manual.rank = NULL, seed.rsvd = 1)
dim(receptor.geneset.matrix.out)
head(receptor.geneset.matrix.out)
```

---

target.malt.rna.mat *Single cell RNA-sequencing (scRNA-seq) target subset of the 10X Genomics MALT counts.*

Description

A random subset of joint scRNA-seq/CITE-seq 10X Genomics human extranodal marginal zone B-cell tumor/mucosa-associated lymphoid tissue (MALT) target data. See the `dataProcessing.R` file from `data-raw` folder for code to recreate data subset.

Usage

```
target.malt.rna.mat
```

Format

A scRNA-seq counts matrix of `dgCMatrix-class` from the `Matrix` package with 4000 cells and 33538 genes.

Source

<https://www.10xgenomics.com/resources/datasets/10-k-cells-from-a-malt-tumor-gene-expression-and-cell-surface-protein-3-standard-3-0-0>

`train.malt.adt.mat` *Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq) training subset of the 10X Genomics MALT counts.*

Description

A random subset of joint scRNA-seq/CITE-seq 10X Genomics human extranodal marginal zone B-cell tumor/mucosa-associated lymphoid tissue (MALT) training data. See the `dataProcessing.R` file from `data-raw` folder for code to recreate data subset.

Usage

```
train.malt.adt.mat
```

Format

A CITE-seq counts matrix of `dgeMatrix-class` from the `Matrix` package with 1000 cells and 17 genes.

Source

<https://www.10xgenomics.com/resources/datasets/10-k-cells-from-a-malt-tumor-gene-expression-and-cell-surface-protein-3-standard-3-0-0>

`train.malt.rna.mat` *Single cell RNA-sequencing (scRNA-seq) training subset of the 10X Genomics MALT counts.*

Description

A random subset of joint scRNA-seq/CITE-seq 10X Genomics human extranodal marginal zone B-cell tumor/mucosa-associated lymphoid tissue (MALT) training data. See the `dataProcessing.R` file from `data-raw` folder for code to recreate data subset.

Usage

```
train.malt.rna.mat
```

Format

A scRNA-seq counts matrix of `dgCMatrix-class` from the `Matrix` package with 1000 cells and 33538 genes.

Source

<https://www.10xgenomics.com/resources/datasets/10-k-cells-from-a-malt-tumor-gene-expression-and-cell-surface-protein-3-standard-3-0-0>

Index