

# Package ‘scSorter’

March 17, 2021

**Type** Package

**Title** Implementation of 'scSorter' Algorithm

**Version** 0.0.2

**Description** Implements the algorithm described in  
Guo, H., and Li, J., ``scSorter: assigning cells to known cell types accord-  
ing to known marker genes".  
Cluster cells to known cell types based on marker genes specified for each cell type.

**Depends** R (>= 3.6.0)

**Imports** stats (>= 3.6.0)

**License** GPL-3

**LazyData** true

**RoxygenNote** 7.1.1

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Hongyu Guo [aut],  
Jun Li [aut, cre]

**Maintainer** Jun Li <jun.li@end.edu>

**Repository** CRAN

**Date/Publication** 2021-03-17 06:40:02 UTC

## R topics documented:

scSorter . . . . .	2
xfindvariable_genes . . . . .	3
xnormalize_scData . . . . .	3

<b>Index</b>	<b>5</b>
--------------	----------

---

 scSorter

*scSorter*


---

## Description

This is the main function that implements the scSorter method.

## Usage

```
scSorter(
  expr,
  anno,
  default_weight = 2,
  n_start = 10,
  alpha = 0,
  u = 0.05,
  max_iter = 100,
  setseed = 0
)
```

## Arguments

expr	A matrix of the input expression data. Each row represents a gene and each column represents a cell. Each row of this matrix should be named by the gene name it represents.
anno	A matrix or data frame that contains marker genes specified for cell types of interest. It should contain three columns named "Type", "Marker", and "Weight" that records the name and weight of marker genes specified for each cell type. "Weight" column is optional. If it is not specified, the default_weight will be applied to all marker genes.
default_weight	The default weight assigned to marker genes. The default value is 2.
n_start	The number of possible cluster initializations. The default value is 10.
alpha	The parameter determines the cutoff whether the cell type of a cell should be considered as undecided during unknown cell calling. The default value is 0.
u	The parameter determines whether undecided cells are further processed. The default value is 0.05.
max_iter	The maximum number of iterations for the algorithm to update parameters. The default value is 100.
setseed	Random seed for cluster initialization. The default value is 0.

## Value

A list contains the elements: Pred\_Type: The predicted cell types. Pred\_param: The parameter estimates of mu and delta.

**Examples**

```
load(system.file('extdata', 'example_data.RData', package = 'scSorter'))
result = scSorter(expr, anno)
misclassification_rate = 1 - mean(result$Pred_Type == true_type)
table(result$Pred_Type, true_type)
```

---

xfindvariable\_genes     *Select Highly Variable Genes*

---

**Description**

Select Highly Variable Genes following the vst approach. Please only use this function when you do not have access to Seurat package. More details are available in the vignette of this package.

**Usage**

```
xfindvariable_genes(expr, ngenes = 2000)
```

**Arguments**

expr	A matrix of input scRNA-seq data. Rows correspond to genes and columns correspond to cells.
ngenes	The number of most variable genes to be selected.

**Value**

A vector of top highly variable genes with the total number determined by @ngenes option.

---

xnormalize\_scData     *Normalize scRNA-seq Data*

---

**Description**

Normalize scRNA-seq data. Please only use this function when you do not have access to Seurat package. More details are available in the vignette of this package.

**Usage**

```
xnormalize_scData(expr)
```

**Arguments**

expr	A matrix of input scRNA-seq data. Rows correspond to genes and columns correspond to cells.
------	---

**Value**

A matrix of normalized expression data.

# Index

scSorter, [2](#)

xfindvariable\_genes, [3](#)

xnormalize\_scData, [3](#)