

Package: mixhvg (via r-universe)

March 2, 2025

Type Package

Title Mixture of Multiple Highly Variable Feature Selection Methods

Version 1.0.1

Maintainer Ruzhang Zhao <ruzhangzhao@gmail.com>

Description Highly variable gene selection methods, including popular public available methods, and also the mixture of multiple highly variable gene selection methods,
<<https://github.com/RuzhangZhao/mixhvg>>. Reference:
<[doi:10.1101/2024.08.25.608519](https://doi.org/10.1101/2024.08.25.608519)>.

Imports scran, Seurat, Matrix, methods, SingleCellExperiment, scuttle

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.1

Config/pak/sysreqs libglpk-dev make libicu-dev libpng-dev libxml2-dev
libssl-dev python3 zlib1g-dev

Repository <https://ruzhangzhao.r-universe.dev>

RemoteUrl <https://github.com/ruzhangzhao/mixhvg>

RemoteRef HEAD

RemoteSha 482bab83ca7fff1ec869fee5ef7d18ce319ecc5

Contents

FindVariableFeaturesMix	2
Index	5

 FindVariableFeaturesMix

FindVariableFeaturesMix

Description

FindVariableFeaturesMix

Usage

```
FindVariableFeaturesMix(
  object,
  method.names = c("scran", "scran_pos", "seuratv1"),
  nfeatures = 2000,
  loess.span = 0.3,
  clip.max = "auto",
  num.bin = 20,
  binning.method = "equal_width",
  extra.rank = NULL,
  verbose = FALSE
)
```

Arguments

object	An object, SeuratObject and matrix(including sparse matrix) are both acceptable
method.names	The following methods can be directly used for highly variable feature selection. The mixture of methods take a vector of method list, e.g. c("scran","scran_pos","seuratv1"), which is also default. scran: Use mean-variance curve adjustment on lognormalized count matrix, which is scran ModelGeneVar. mv_ct: Use mean-variance curve adjustment on count matrix, inherited from scran ModelGeneVar. mv_nc: Use mean-variance curve adjustment on normalized count matrix, inherited from scran ModelGeneVar. mv_lognc: The same as scran. mv_PFlogPF: Use mean-variance curve adjustment on PFlog1pPF matrix, inherited from scran ModelGeneVar. scran_pos: Use scran poisson version, modelGeneVarByPoisson. seuratv3: Use logmean-logvariance curve adjustment on count matrix, which is vst, Seurat FindVariableFeatures Function(https://satijalab.org/seurat/reference/findvariablefeatures). logmv_ct: The same as seuratv3. logmv_nc: Use logmean-logvariance curve adjustment on normalized count matrix, inherited from seuratv3(vst).

	logmv_lognc: Use logmean-logvariance curve adjustment on lognormalized count matrix, inherited from <code>seuratv3(vst)</code> .
	logmv_PFlogPF: Use logmean-logvariance curve adjustment on PFlog1pPF matrix, inherited from <code>seuratv3(vst)</code> .
	seuratv1: Use dispersion on lognormalized count matrix, which is dispersion (<code>disp</code>), Seurat FindVariableFeatures Function(https://satijalab.org/seurat/reference/findvariablefeatures).
	disp_lognc: The same as <code>seuratv1</code> .
	disp_PFlogPF: Use dispersion on PFlog1pPF matrix, inherited from <code>seuratv1(disp)</code> .
	mean_max_ct: Highly Expressed Features with respect to count matrix.
	mean_max_nc: Highly Expressed Features with respect to normalized count matrix.
	mean_max_lognc: Highly Expressed Features with respect to lognormalized count matrix
<code>nfeatures</code>	Number of features to select as top variable features.
<code>loess.span</code>	(Only work for logmv based methods like <code>seuratv3</code>). Loess span parameter used when fitting the variance-mean relationship
<code>clip.max</code>	(Only work for logmv based methods like <code>seuratv3</code>). After standardization values larger than <code>clip.max</code> will be set to <code>clip.max</code> ; default is 'auto' which sets this value to the square root of the number of cells
<code>num.bin</code>	(Only work for logmv or dispersion based methods)Total number of bins to use in the scaled analysis (default is 20)
<code>binning.method</code>	Specifies how the bins should be computed. Available methods are: equal_width: each bin is of equal width along the x-axis[default]. equal_frequency: each bin contains an equal number of features (can increase statistical power to detect overdispersed features at high expression values, at the cost of reduced resolution along the x-axis).
<code>extra.rank</code>	We support the user to input customized gene rank. Provide the best gene with rank "1", and the worst gene with largest number.
<code>verbose</code>	Whether to show progress bar for calculations. Default is FALSE.

Details

The function inherits from FindVariableFeatures function of Seurat Package. Refer to <https://github.com/RuzhangZhao/mixhvg> for user manual.

Value

object: If the input is SeuratObject, the return is also SeuratObject; if the input is matrix(including sparse matrix), the return is the highly variable feature names.

Examples

```
if(0){
simple_matrix<-matrix(1:2e4,nrow=4000,ncol=5)
rownames(simple_matrix)<-1:nrow(simple_matrix)
```

```
colnames(simple_matrix)<-1:ncol(simple_matrix)
simple_matrix_HVG<-FindVariableFeaturesMix(simple_matrix)
}
```

Index

FindVariableFeaturesMix, [2](#)