

# Package ‘HIPPO’

March 20, 2023

**Type** Package

**Title** Heterogeneity-Induced Pre-Processing tOol

**Version** 1.10.0

**Description** For scRNA-seq data, it selects features and clusters the cells simultaneously for single-cell UMI data. It has a novel feature selection method using the zero inflation instead of gene variance, and computationally faster than other existing methods since it only relies on PCA+Kmeans rather than graph-clustering or consensus clustering.

**License** GPL (>=2)

**Depends** R (>= 3.6.0)

**Encoding** UTF-8

**LazyData** true

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**URL** <https://github.com/tk382/HIPPO>

**BugReports** <https://github.com/tk382/HIPPO/issues>

**Imports** ggplot2, graphics, stats, reshape2, gridExtra, Rtsne, umap, dplyr, rlang, magrittr, irlba, Matrix, SingleCellExperiment, ggrepel

**RoxygenNote** 7.1.0

**biocViews** Sequencing, SingleCell, GeneExpression, DifferentialExpression, Clustering

**git\_url** <https://git.bioconductor.org/packages/HIPPO>

**git\_branch** RELEASE\_3\_16

**git\_last\_commit** a7c28ba

**git\_last\_commit\_date** 2022-11-01

**Date/Publication** 2023-03-20

**NeedsCompilation** no

**Author** Tae Kim [aut, cre],  
Mengjie Chen [aut]

**Maintainer** Tae Kim <tk382@uchicago.edu>

## R topics documented:

|                                     |    |
|-------------------------------------|----|
| ensg_hgnc . . . . .                 | 2  |
| get_data_from_sce . . . . .         | 3  |
| get_hippo . . . . .                 | 3  |
| get_hippo_diffexp . . . . .         | 4  |
| hippo . . . . .                     | 4  |
| hippo_diagnostic_plot . . . . .     | 5  |
| hippo_diffexp . . . . .             | 6  |
| hippo_dimension_reduction . . . . . | 7  |
| hippo_feature_heatmap . . . . .     | 8  |
| hippo_pca_plot . . . . .            | 9  |
| hippo_tsne_plot . . . . .           | 9  |
| hippo_umap_plot . . . . .           | 10 |
| nb_prob_zero . . . . .              | 11 |
| pois_prob_zero . . . . .            | 11 |
| preprocess_heterogeneous . . . . .  | 12 |
| preprocess_homogeneous . . . . .    | 12 |
| toydata . . . . .                   | 13 |
| zero_proportion_plot . . . . .      | 13 |
| zinb_prob_zero . . . . .            | 15 |
| %>% . . . . .                       | 15 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>16</b> |
|--------------|-----------|

---

|           |   |
|-----------|---|
| ensg_hgnc | <i>A reference data frame that matches ENSG IDs to HGNC symbols</i> |
|-----------|---|

---

### Description

A reference data frame that matches ENSG IDs to HGNC symbols

### Usage

```
ensg_hgnc
```

### Format

A data frame with 46606 rows and 2 columns

**ensg** Ensembl ENSG IDs

**hgnc** HGNC symbols

### Source

<http://www.biomart.org/>

---

|                   |                                    |
|-------------------|------------------------------------|
| get_data_from_sce | <i>Access data from SCE object</i> |
|-------------------|------------------------------------|

---

**Description**

Access data from SCE object

**Usage**

```
get_data_from_sce(sce)
```

**Arguments**

sce                    SingleCellExperiment object

**Value**

count matrix

**Examples**

```
data(toydata)
X = get_data_from_sce(toydata)
```

---

|           |  |
|-----------|--|
| get_hippo | <i>Access hippo object from SingleCellExperiment object.</i> |
|-----------|--|

---

**Description**

Access hippo object from SingleCellExperiment object.

**Usage**

```
get_hippo(sce)
```

**Arguments**

sce                    SingleCellExperiment object

**Value**

hippo object embedded in SingleCellExperiment object

**Examples**

```
data(toydata)
set.seed(20200321)
toydata = hippo(toydata,K = 10,z_threshold = 1,outlier_proportion = 0.01)
hippo_object = get_hippo(toydata)
```

---

`get_hippo_diffexp`      *Return hippo\_diffexp object*

---

### Description

Return hippo\_diffexp object

### Usage

```
get_hippo_diffexp(sce, k = 1)
```

### Arguments

|                  |  |
|------------------|--|
| <code>sce</code> | SingleCellExperiment object with hippo |
| <code>k</code>   | integer round of result of interest    |

### Value

data frame of differential expression test

### Examples

```
data(toydata)
set.seed(20200321)
toydata = hippo(toydata,K = 10,z_threshold = 1,outlier_proportion = 0.01)
toydata = hippo_diffexp(toydata)
result1 = get_hippo_diffexp(toydata)
```

---

`hippo`      *HIPPO's hierarchical clustering*

---

### Description

HIPPO's hierarchical clustering

### Usage

```
hippo(sce, K = 20, z_threshold = 2, outlier_proportion = 0.001, verbose = TRUE)
```

### Arguments

|                                 |   |
|---------------------------------|---|
| <code>sce</code>                | SingleCellExperiment object   |
| <code>K</code>                  | number of clusters to ultimately get  |
| <code>z_threshold</code>        | numeric > 0 as a z-value threshold for selecting the features   |
| <code>outlier_proportion</code> | numeric between 0 and 1, a cut-off so that when the proportion of important features reach this number, the clustering terminates |
| <code>verbose</code>            | if set to TRUE, it shows progress of the algorithm  |

**Value**

a list of clustering result for each level of  $k=1, 2, \dots K$ .

**Examples**

```
data(toydata)
toydata = hippo(toydata,K = 10,z_threshold = 1,outlier_proportion = 0.01)
```

---

hippo\_diagnostic\_plot *Conduct feature selection by computing test statistics for each gene*

---

**Description**

Conduct feature selection by computing test statistics for each gene

**Usage**

```
hippo_diagnostic_plot(sce, show_outliers = FALSE, zvalue_thresh = 10)
```

**Arguments**

- sce                SingleCellExperiment object with count matrix
- show\_outliers    boolean to indicate whether to circle the outliers with given zvalue\_thresh
- zvalue\_thresh    a numeric  $v$  for defining outliers

**Value**

a diagnostic plot that shows genes with zero inflation

**Examples**

```
data(toydata)
hippo_diagnostic_plot(toydata, show_outliers=TRUE, zvalue_thresh = 2)
```

---

hippo\_diffexp                    *HIPPO's differential expression*

---

## Description

HIPPO's differential expression

## Usage

```
hippo_diffexp(  
  sce,  
  top.n = 5,  
  switch_to_hgnc = FALSE,  
  ref = NA,  
  k = NA,  
  plottitle = ""  
)
```

## Arguments

|                |   |
|----------------|---|
| sce            | SingleCellExperiment object with hippo  |
| top.n          | number of markers to return   |
| switch_to_hgnc | if the current gene names are ensemble ids, and would like to switch to hgnc                                      |
| ref            | a data frame with columns 'hgnc' and 'ensg' to match each other, only required when switch_to_hgnc is set to TRUE |
| k              | number of rounds of clustering that you'd like to see result. Default is 1 to K                                   |
| plottitle      | title of the resulting plot   |

## Value

list of differential expression result

## Examples

```
data(toydata)  
set.seed(20200321)  
toydata = hippo(toydata,K = 10,z_threshold = 1,outlier_proportion = 0.01)  
result = hippo_diffexp(toydata)
```

---

hippo\_dimension\_reduction  
*compute t-SNE or umap of each round of HIPPO*

---

## Description

compute t-SNE or umap of each round of HIPPO

## Usage

```
hippo_dimension_reduction(  
  sce,  
  method = c("umap", "tsne"),  
  perplexity = 30,  
  featurelevel = 1  
)
```

## Arguments

|              |  |
|--------------|--|
| sce          | SingleCellExperiment object with hippo object in it.                                 |
| method       | a string that determines the method for dimension reduction: either 'umap' or 'tsne' |
| perplexity   | numeric perplexity parameter for Rtsne function                                      |
| featurelevel | the round of clustering that you will extract features to reduce the dimension       |

## Value

a data frame of dimension reduction result for each k in 1, ..., K

## Examples

```
data(toydata)  
set.seed(20200321)  
set.seed(20200321)  
toydata = hippo(toydata, K = 10, z_threshold = 1, outlier_proportion = 0.01)  
toydata = hippo_dimension_reduction(toydata, method="tsne")  
hippo_tsne_plot(toydata)
```

---

hippo\_feature\_heatmap *HIPPO's feature heatmap*

---

## Description

HIPPO's feature heatmap

## Usage

```
hippo_feature_heatmap(  
  sce,  
  switch_to_hgnc = FALSE,  
  ref = NA,  
  top.n = 50,  
  kk = 2,  
  plottitle = ""  
)
```

## Arguments

|                |   |
|----------------|---|
| sce            | SingleCellExperiment object with hippo  |
| switch_to_hgnc | if the current gene names are ensemble ids, and would like to switch to hgnc                                      |
| ref            | a data frame with columns 'hgnc' and 'ensg' to match each other, only required when switch_to_hgnc is set to TRUE |
| top.n          | number of markers to return   |
| kk             | integer for the round of clustering that you'd like to see result. Default is 2                                   |
| plottitle      | title for the plot  |

## Value

list of differential expression result

## Examples

```
data(toydata)  
set.seed(20200321)  
toydata = hippo(toydata,K = 10,z_threshold = 1,outlier_proportion = 0.01)  
hippo_feature_heatmap(toydata)
```



---

|                |  |
|----------------|--|
| hippo_pca_plot | <i>visualize each round of hippo through t-SNE</i> |
|----------------|--|

---

**Description**

visualize each round of hippo through t-SNE

**Usage**

```
hippo_pca_plot(sce, k = NA, pointsize = 0.5, pointalpha = 0.5, plottitle = "")
```

**Arguments**

|            |   |
|------------|---|
| sce        | SingleCellExperiment object with hippo and t-SNE result in it                   |
| k          | number of rounds of clustering that you'd like to see result. Default is 1 to K |
| pointsize  | size of the point for the plot (default 0.5)                                    |
| pointalpha | transparency level of points for the plot (default 0.5)                         |
| plottitle  | title for the ggplot  |

**Value**

ggplot for pca in each round

**Examples**

```
data(toydata)
set.seed(20200321)
toydata = hippo(toydata, K = 10, z_threshold = 1)
hippo_pca_plot(toydata, k = 2:3)
```

---

|                 |  |
|-----------------|--|
| hippo_tsne_plot | <i>visualize each round of hippo through t-SNE</i> |
|-----------------|--|

---

**Description**

visualize each round of hippo through t-SNE

**Usage**

```
hippo_tsne_plot(sce, k = NA, pointsize = 0.5, pointalpha = 0.5, plottitle = "")
```

**Arguments**

|            |   |
|------------|---|
| sce        | SingleCellExperiment object with hippo and t-SNE result in it                   |
| k          | number of rounds of clustering that you'd like to see result. Default is 1 to K |
| pointsize  | size of the point for the plot (default 0.5)                                    |
| pointalpha | transparency level of points for the plot (default 0.5)                         |
| plottitle  | title for the ggplot output   |

**Value**

ggplot object for t-SNE in each round

**Examples**

```
data(toydata)
set.seed(20200321)
toydata = hippo(toydata, K = 10, z_threshold = 1, outlier_proportion = 0.01)
toydata = hippo_dimension_reduction(toydata, method="tsne")
hippo_tsne_plot(toydata)
```

---

|                 |   |
|-----------------|---|
| hippo_umap_plot | <i>visualize each round of hippo through UMAP</i> |
|-----------------|---|

---

**Description**

visualize each round of hippo through UMAP

**Usage**

```
hippo_umap_plot(sce, k = NA, pointsize = 0.5, pointalpha = 0.5, plottitle = "")
```

**Arguments**

|            |   |
|------------|---|
| sce        | SingleCellExperiment object with hippo and UMAP result in it                    |
| k          | number of rounds of clustering that you'd like to see result. Default is 1 to K |
| pointsize  | size of the point for the plot (default 0.5)                                    |
| pointalpha | transparency level of points for the plot (default 0.5)                         |
| plottitle  | title of the resulting plot   |

**Value**

ggplot object for umap in each round

**Examples**

```

data(toydata)
set.seed(20200321)
toydata = hippo(toydata,K = 10,z_threshold = 1,outlier_proportion = 0.01)
toydata = hippo_dimension_reduction(toydata, method="umap")
hippo_umap_plot(toydata)

```

---

nb\_prob\_zero

*Expected zero proportion under Negative Binomial*


---

**Description**

Expected zero proportion under Negative Binomial

**Usage**

```
nb_prob_zero(lambda, theta)
```

**Arguments**

lambda            numeric vector of means of negative binomial  
theta             numeric vector of the dispersion parameter for negative binomial, 0 if poisson

**Value**

numeric vector of expected zero proportion under Negative Binomial

**Examples**

```
nb_prob_zero(3, 1.1)
```

---

pois\_prob\_zero

*Expected zero proportion under Poisson*


---

**Description**

Expected zero proportion under Poisson

**Usage**

```
pois_prob_zero(lambda)
```

**Arguments**

lambda            numeric vector of means of Poisson

**Value**

numeric vector of expected proportion of zeros for each lambda

**Examples**

```
pois_prob_zero(3)
```

---

preprocess\_heterogeneous

*Preprocess UMI data without cell label so that each row contains information about each gene*

---

**Description**

Preprocess UMI data without cell label so that each row contains information about each gene

**Usage**

```
preprocess_heterogeneous(X)
```

**Arguments**

X a matrix object with counts data

**Value**

data frame with one row for each gene.

**Examples**

```
data(toydata)
df = preprocess_heterogeneous(get_data_from_sce(toydata))
```

---

preprocess\_homogeneous

*Preprocess UMI data with inferred or known labels*

---

**Description**

Preprocess UMI data with inferred or known labels

**Usage**

```
preprocess_homogeneous(sce, label)
```

**Arguments**

sce                SingleCellExperiment object with counts data  
label             a numeric or character vector of inferred or known label

**Value**

data frame with one row for each gene.

**Examples**

```
data(toydata)
labels = SingleCellExperiment::colData(toydata)$phenoid
df = preprocess_homogeneous(toydata, label = labels)
```

---

|         |  |
|---------|--|
| toydata | <i>A sample single cell sequencing data subsetted from Zheng2017</i> |
|---------|--|

---

**Description**

A sample single cell sequencing data subsetted from Zheng2017

**Usage**

```
toydata
```

**Format**

Single Cell experiment object with 10,000 genes and 100 cells

**Source**

<https://www.nature.com/articles/ncomms14049>

---

|                      |   |
|----------------------|---|
| zero_proportion_plot | <i>visualize each round of hippo through zero proportion plot</i> |
|----------------------|---|

---

**Description**

visualize each round of hippo through zero proportion plot

**Usage**

```
zero_proportion_plot(  
  sce,  
  switch_to_hgnc = FALSE,  
  ref = NA,  
  k = NA,  
  plottitle = "",  
  top.n = 5,  
  pointsize = 0.5,  
  pointalpha = 0.5,  
  textsize = 3  
)
```

**Arguments**

|                |   |
|----------------|---|
| sce            | SingleCellExperiment object with hippo element in it  |
| switch_to_hgnc | boolean argument to indicate whether to change the gene names from ENSG IDs to HGNC symbols |
| ref            | a data frame with hgnc column and ensg column   |
| k              | select rounds of clustering that you would like to see result. Default is 1 to K            |
| plottitle      | Title of your plot output   |
| top.n          | number of top genes to show the name  |
| pointsize      | size of the ggplot point  |
| pointalpha     | transparency level of the ggplot point  |
| textsize       | text size of the resulting plot   |

**Value**

a ggplot object that shows the zero proportions for each round

**Examples**

```
data(toydata)  
set.seed(20200321)  
toydata = hippo(toydata, K = 10, z_threshold = 1, outlier_proportion = 0.01)  
data(ensg_hgnc)  
zero_proportion_plot(toydata, switch_to_hgnc = TRUE, ref = ensg_hgnc)
```

---

|                |   |
|----------------|---|
| zinb_prob_zero | <i>Expected zero proportion under Negative Binomial</i> |
|----------------|---|

---

**Description**

Expected zero proportion under Negative Binomial

**Usage**

```
zinb_prob_zero(lambda, theta, pi)
```

**Arguments**

|        |  |
|--------|--|
| lambda | gene mean  |
| theta  | dispersion parameter, 0 if zero-inflated poisson |
| pi     | zero inflation, 0 if negative binomial           |

**Value**

Expected zero proportion under Zero-Inflated Negative Binomial

**Examples**

```
zinb_prob_zero(3, 1.1, 0.1)
```

---

|     |   |
|-----|---|
| %>% | <i>re-export magrittr pipe operator</i> |
|-----|---|

---

**Description**

re-export magrittr pipe operator

# Index

## \* datasets

ensg\_hgnc, [2](#)  
toydata, [13](#)  
%>%, [15](#)

ensg\_hgnc, [2](#)

get\_data\_from\_sce, [3](#)  
get\_hippo, [3](#)  
get\_hippo\_diffexp, [4](#)

hippo, [4](#)  
hippo\_diagnostic\_plot, [5](#)  
hippo\_diffexp, [6](#)  
hippo\_dimension\_reduction, [7](#)  
hippo\_feature\_heatmap, [8](#)  
hippo\_pca\_plot, [9](#)  
hippo\_tsne\_plot, [9](#)  
hippo\_umap\_plot, [10](#)

nb\_prob\_zero, [11](#)

pois\_prob\_zero, [11](#)  
preprocess\_heterogeneous, [12](#)  
preprocess\_homogeneous, [12](#)

toydata, [13](#)

zero\_proportion\_plot, [13](#)  
zinb\_prob\_zero, [15](#)