Package 'GTEs'

February 27, 2025

Type Package

Title Group Technical Effects

Version 1.0.0

Language en-US

Date 2025-02-20

Maintainer Yang Zhou <yangz@stu.hit.edu.cn>

Description Implementation of the GTE (Group Technical Effects) model for single-cell data. GTE is a quantitative metric to assess batch effects for individual genes in single-cell data. For a single-cell dataset, the user can calculate the GTE value for individual features (such as genes), and then identify the highly batch-sensitive features. Removing these highly batch-sensitive features results in datasets with low batch effects.

License GPL-3

Encoding UTF-8

Depends R (>= 4.0.0)

Imports stats, Matrix, matrixStats, Rcpp, RcppEigen, dplyr

LinkingTo Rcpp (>= 1.0.8), RcppEigen

RoxygenNote 7.2.3

NeedsCompilation yes

URL https://github.com/yzhou1999/GTEs, https://yzhou1999.github.io/GTEs/

BugReports https://github.com/yzhou1999/GTEs/issues

Author Yang Zhou [aut, cre]

Repository CRAN

Date/Publication 2025-02-27 16:50:10 UTC

Contents

`group_onehot` *Compute one-hot matrix for given data frame and variable (s)*

Description

Compute one-hot matrix for given data frame and variable (s)

Usage

```
group_onehot(x, ivar)
```

Arguments

`x`	Input data frame.
`ivar`	Variable (s) for one-hot computation.

`Run.GroupTechEffects` *Compute the group technical effects.*

Description

Compute the group technical effects.

Usage

```
Run.GroupTechEffects(X, meta, g_factor, b_factor, do.scale = FALSE)
```

Arguments

`X`	Input data matrix.
`meta`	Input metadata (data.frame).
`g_factor`	Group variable (s).
`b_factor`	Batch variable (s).
`do.scale`	Whether to perform scaling.

Value

A list containing the overall GTE ($OverallTechEffects) and the GTE ($GroupTechEffects) of each subgroup under the group variable.

Examples

```
# X is a normalized expression matrix with rows as features and columns as cells.

# meta is a data.frame with columns containing metadata such as cell type, batch, etc.

data_file <- system.file("extdata", "example_data.rds", package = "GTEs")
example_data <- readRDS(data_file)
meta_file <- system.file("extdata", "example_meta.rds", package = "GTEs")
example_meta <- readRDS(meta_file)
GTE_ct <- Run.GroupTechEffects(example_data, example_meta,
                               g_factor = "CellType",
                               b_factor = "Batch")
```

---

scale_data *Scale data matrix*

Description

Scale data matrix

Usage

```
scale_data(
  data.x,
  do.center = TRUE,
  do.scale = TRUE,
  row.means = NULL,
  row.sds = NULL
)
```

Arguments

data.x	Input data matrix.
do.center	Whether center the row values. (default TRUE)
do.scale	Whether scale the row values. (default TRUE)
row.means	The provided row means to center. (default NULL)
row.sds	The provided row standard deviations to scale. (default NULL)

Select.HBGs *Select highly batch-sensitive genes (HBGs) under a group variable.*

Description

Select highly batch-sensitive genes (HBGs) under a group variable.

Usage

```
Select.HBGs(GTE, bins = 0.1, gte.ratio = 0.95)
```

Arguments

GTE	GTE result.
bins	Bins.
gte.ratio	Ratio of selected HBGs to the total GTE.

Value

Identified HBGs.

Examples

```
# GTE is the result of Run.GroupTechEffects function.
data_file <- system.file("extdata", "GTE_ct.rds", package = "GTEs")
GTE_ct <- readRDS(data_file)
HBGs <- Select.HBGs(GTE_ct)
```

select_hbgs *Select HBGs using GTE vector.*

Description

Select HBGs using GTE vector.

Usage

```
select_hbgs(gte, bins = 0.1, gte.ratio = 0.95, is.sort = TRUE)
```

Arguments

gte	Named GTE vector.
bins	Bins.
gte.ratio	Ratio of selected HBGs to overall GTE.
is.sort	Whether to sort genes by GTE from largest to smallest.

Index