Package 'precision.subtype'

October 4, 2022

Type Package

Title On Data Normalization for Tumor Subtyping with microRNA data

Version 0.99.00

Date 2020-10-4

Author Becky Wing-Yan Yuen

Maintainer Becky Wing-Yan Yuen <wyyuen@terpmail.umd.edu>

Description precision.subtype provides the function for normalization, batch effect correction, and clustering workflow on microRNA data that we implemented in our manuscript.

Depends R (>= 3.5.0)

License GPL (>= 2)

Imports sva, vsn, preprocessCore, cluster, fpc, sparcl, som, mclust

Keywords MicroRNA experiments; Batch effects; Normalization; Clustering; Integrative analysis

R topics documented:

---

cluster_other	*Clustering by Self-Organizing Map (SOM), K-Means, or Model-Based Clustering*

Description

The function gives clustering results from a data frame or matrix by by Self-Organizing Map (SOM), K-Means, or Model-Based Clustering.

Usage

```
cluster_other(dat, true_cluster, clust_method = c("SOM", "kmeans", "MNM"),
ncluster = 2, normalize = FALSE,
norm_method = c("quantile", "median", "vsn"), combat = FALSE,
batch = NULL, eliminate = TRUE, eliminate_lv = 6 )
```

Arguments

dat	data matrix or data frame, each row corresponds to a probe and each column corresponds to a sample.
true_cluster	numeric vector indicating the true cluster labels corresponding to the samples, prespecified by the user.
clust_method	character string specifying the clustering method to be used. The available options are "SOM", "kmeans" and "MNM".
ncluster	an integer indicating the number of clusters. The default value is 2.
normalize	logical indicator. If normalize = TRUE, the data will be normalized by the chosen normalization method. The default value is normalize = FALSE.
norm_method	character string specifying the normalization method to be used. The available options are "quantile", "median" and "vsn".
combat	logical indicator. If combat = TRUE, ComBat() will be applied after normalization. The default value is combat = FALSE.
batch	numeric vector indicating the batch labels for ComBat(), prespecified by the user.
eliminate	logical indicator. If eliminate = TRUE, probes with more than 95% expressions smaller than eliminate_lv. The default value is eliminate = TRUE.
eliminate_lv	default value is 6.

Value

List with the following elements:

clustering	numeric vector indicating the estimated cluster labels corresponding to the samples.
cluster	nclust character string vectors storing the names of samples in each cluster.
ARI	corrected Rand index evaluating the accuracy of clustering results.

Author(s)

Becky Wing-Yan Yuen

References

Yilin Wu*, Becky Wing-Yan Yuen*, Yingying Wei, Li-Xuan Qin^. On data normalization for tumor subtyping with microRNA data, Manuscript.

Examples

```
data("example_data", package = "precision.subtype")

uni_true_cluster = rep(NA,dim(example_data)[2])
uni_true_cluster[which(substr(colnames(example_data),7,7)=="E")] = 1
uni_true_cluster[which(substr(colnames(example_data),7,7)=="V")] = 2

results = cluster_other(dat = example_data, true_cluster = uni_true_cluster,
clust_method = "kmeans")

str(results)
```

---

cluster_pam | *Clustering by Partition Around Medoids (PAM)*

Description

The function gives clustering results from a data frame or matrix by Partition Around Medoids (PAM) method.

Usage

```
cluster_pam(dat, true_cluster, distance = c("euclidean", "spearman", "pearson"),
ncluster = 2, normalize = FALSE,
norm_method = c("quantile", "median", "vsn"),
combat = FALSE, batch = NULL, eliminate = TRUE,
eliminate_lv = 6 )
```

Arguments

dat	data matrix or data frame, each row corresponds to a probe and each column corresponds to a sample.
true_cluster	numeric vector indicating the true cluster labels corresponding to the samples, prespecified by the user.
distance	character string specifying the method to be used for calculating dissimilarities between observations in pam(). The available options are "euclidean", "spearman" and "pearson".
ncluster	an integer indicating the number of clusters. The default value is 2.
normalize	logical indicator. If normalize = TRUE, the data will be normalized by the chosen normalization method. The default value is normalize = FALSE.
norm_method	character string specifying the normalization method to be used. The available options are "quantile", "median" and "vsn".
combat	logical indicator. If combat = TRUE, ComBat() will be applied after normalization. The default value is combat = FALSE.
batch	numeric vector indicating the batch labels for ComBat(), prespecified by the user.
eliminate	logical indicator. If eliminate = TRUE, probes with more than 95% expressions smaller than eliminate_lv. The default value is eliminate = TRUE.
eliminate_lv	default value is 6.

Value

List with the following elements:

clustering	numeric vector indicating the estimated cluster labels corresponding to the samples.
cluster	nclust character string vectors storing the names of samples in each cluster.
ARI	corrected Rand index evaluating the accuracy of clustering results.

Author(s)

Becky Wing-Yan Yuen

References

Yilin Wu*, Becky Wing-Yan Yuen*, Yingying Wei, Li-Xuan Qin^. On data normalization for tumor subtyping with microRNA data, Manuscript.

Examples

```
data("example_data", package = "precision.subtype")

uni_true_cluster = rep(NA,dim(example_data)[2])
uni_true_cluster[which(substr(colnames(example_data),7,7)=="E")] = 1
uni_true_cluster[which(substr(colnames(example_data),7,7)=="V")] = 2

results = cluster_pam(dat = example_data, true_cluster = uni_true_cluster,
distance = "euclidean")

str(results)
```

---

cluster_Skmeans   *Clustering by Sparse K-Means Clustering*

Description

The function gives clustering results from a data frame or matrix by sparse k-means clustering.

Usage

```
cluster_Skmeans(dat, true_cluster, ncluster = 2, normalize = FALSE,
norm_method = c("quantile", "median", "vsn"), combat = FALSE,
batch = NULL, eliminate = TRUE, eliminate_lv = 6 )
```

Arguments

dat	data matrix or data frame, each row corresponds to a probe and each column corresponds to a sample.
true_cluster	numeric vector indicating the true cluster labels corresponding to the samples, prespecified by the user.
ncluster	an integer indicating the number of clusters. The default value is 2.
normalize	logical indicator. If normalize = TRUE, the data will be normalized by the chosen normalization method. The default value is normalize = FALSE.
norm_method	character string specifying the normalization method to be used. The available options are "quantile", "median" and "vsn".
combat	logical indicator. If combat = TRUE, ComBat() will be applied after normalization. The default value is combat = FALSE.
batch	numeric vector indicating the batch labels for ComBat(), prespecified by the user.
eliminate	logical indicator. If eliminate = TRUE, probes with more than 95% expressions smaller than eliminate_lv. The default value is eliminate = TRUE.
eliminate_lv	default value is 6.

Value

List with the following elements:

clustering   numeric vector indicating the estimated cluster labels corresponding to the samples.

cluster   nclust character string vectors storing the names of samples in each cluster.

ARI   corrected Rand index evaluating the accuracy of clustering results.

Author(s)

Becky Wing-Yan Yuen

References

Yilin Wu*, Becky Wing-Yan Yuen*, Yingying Wei, Li-Xuan Qin^. On data normalization for tumor subtyping with microRNA data, Manuscript.

Examples

```
data("example_data", package = "precision.subtype")

uni_true_cluster = rep(NA,dim(example_data)[2])
uni_true_cluster[which(substr(colnames(example_data),7,7)=="E")] = 1
uni_true_cluster[which(substr(colnames(example_data),7,7)=="V")] = 2

results = cluster_Skmeans(dat = example_data, true_cluster = uni_true_cluster)

str(results)
```

---

example_data   *A data set for running examples in user manual*

Description

This data set is the biological-effects-only data with amplification constant 1.6.

Format

A data frame with 3,523 rows by 192 columns.

Details

The data set contains 192 tumor samples (columns) with 3,523 probes (rows).

References

Yilin Wu*, Becky Wing-Yan Yuen*, Yingying Wei, Li-Xuan Qin^. On data normalization for tumor subtyping with microRNA data, Manuscript.

Index