

# Package ‘PRECISION.array.subtyping’

October 17, 2022

**Type** Package

**Title** On Data Normalization for Tumor Subtyping with microRNA data

**Version** 0.99.0

**Date** 2020-10-17

**Author** Becky Wing-Yan Yuen, Yilin Wu, Yingying Wei, Li-Xuan Qin

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

**Description** PRECISION.array.subtyping 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 3

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

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

**RoxygenNote** 7.2.1

## R topics documented:

cluster_other . . . . .	1
cluster_pam . . . . .	3
cluster_Skmeans . . . . .	4
example_data . . . . .	5

<b>Index</b>	<b>7</b>
--------------	----------

---

cluster_other	<i>Clustering by Self-Organizing Map (SOM), K-Means, or Model-Based Clustering</i>
---------------	--

---

## 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**

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

**Value**

List with the following elements:

<code>clustering</code>	numeric vector indicating the estimated cluster labels corresponding to the samples.
<code>cluster</code>	<code>nclust</code> character string vectors storing the names of samples in each cluster.
<code>ARI</code>	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<sup>^</sup>. On data normalization for tumor subtyping with microRNA data, Manuscript.

## Examples

```
data("example_data", package = "PRECISION.array.subtyping")

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<sup>^</sup>. On data normalization for tumor subtyping with microRNA data, Manuscript.

**Examples**

```
data("example_data", package = "PRECISION.array.subtyping")

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	<i>Clustering by Sparse K-Means Clustering</i>
-----------------	--

---

**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.array.subtyping")

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

## \* **datasets**

example\_data, [5](#)

cluster\_other, [1](#)

cluster\_pam, [3](#)

cluster\_Skmeans, [4](#)

example\_data, [5](#)