Package: cbmc (via r-universe)

January 8, 2025

Title Raw and Processed Matrices of the CITE-seq CBMC Dataset

Version 0.1.0

Description Provides expression matrices from 8,617 cord blood mononuclear cells generated with CITE-seq. The CITE-seq technique and CBMC dataset are described in <doi:10.1038/nmeth.4380>

URL `https://mojaveazure.r-universe.dev/cbmc`, `https://github.com/mojaveazure/cbmc`

BugReports `https://github.com/mojaveazure/cbmc/issues`

License CC BY 4.0

Encoding UTF-8

BuildResaveData best

RoxygenNote 7.3.2

Depends Matrix (>= 1.7.1), R (>= 4.4)

Imports methods, utils

Suggests data.table, R.utils, rprojroot

Enhances Seurat, SeuratObject

Repository https://mojaveazure.r-universe.dev

RemoteUrl https://github.com/mojaveazure/cbmc

RemoteRef HEAD

RemoteSha 6e19a6e101f455a4fdb81771e792660bfe9c54c2

Contents

`cbmc` *CBMC*

Description

CITE-seq data from 8,617 cord blood mononuclear cells (CBMC); the dataset is split into two matrices: one for the RNA expression data and one for the cell-surface protein (ADT) expression data

Usage

```
cbmc.rna

cbmc.adt
```

Format

`cbmc.rna`: A `dgCMatrix` with the raw RNA expression data

`cbmc.adt`: A `dgCMatrix` with the raw ADT expression data

Source

NIH GEO accession: `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100866`

Nature Methods article: `http://www.nature.com/nmeth/journal/v14/n9/full/nmeth.4380.html`

References

Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 2017 Sep;14(9):865-868.

`cbmc.annotations` *CBMC Annotations*

Description

Cell-type annotations for the CBMC dataset, as generated using the **Seurat** standard workflow

Usage

```
cbmc.annotations
```

Format

A data frame with annotations for the CBMC dataset; the data frame contains the following columns as `factor`s:

- "`rna_annotations`": annotations derived from the RNA expression matrix
- "`protein_annotations`": annotations derived from the ADT expression matrix; note some values are `NA`

Cell barcodes are present in the row names of the data frame

`cbmc.norm` *CBMC Normalized*

Description

Normalized versions of the CBMC dataset using standard **Seurat** normalization techniques. The collapsed RNA expression matrix is normalized using log-normalization. The ADT expression matrix is normalized using centered log ratio normalization

Usage

```
cbmc.rna.norm

cbmc.adt.norm
```

Format

`cbmc.rna.norm`: A `dgCMatrix` with the log-normalized RNA expression data

`cbmc.adt.norm`: A `dgCMatrix` with the CLR-normalized ADT expression data

Source

The **Seurat** multimodal vignette

NIH GEO accession: `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100866`

Nature Methods article: `http://www.nature.com/nmeth/journal/v14/n9/full/nmeth.4380.html`

See Also

`Seurat::NormalizeData()` for details on log- and CLR-normalization

`cbmc.rna.collapsed` *CBMC Collapsed*

Description

Species-collapsed version of the CBMC dataset. Species-collapsing consists of:

- stripping the prefix “`HUMAN_`” from all feature names where it is present
- determining the top 100 negative control features prefixed with “`MOUSE_`” and removing all other negative control features

For the CBMC dataset, only the RNA matrix needs to be collapsed

Usage

```
cbmc.rna.collapsed
```

Format

A `dgCMatrix` with the species-collapsed RNA expression matrix

Source

The **Seurat** multimodal vignette

NIH GEO accession: `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100866`

Nature Methods article: `http://www.nature.com/nmeth/journal/v14/n9/full/nmeth.4380.html`

See Also

`Seurat::CollapseSpeciesExpressionMatrix()`

Index