Package 'EDecExampleData'

September 27, 2016

Title Example dataset for use with the EDec package

Version 0.9

Description This package contains an example dataset used to illustrate the use of the EDec method.

Depends R (>= 3.1.2)

License MIT + file LICENSE

Encoding UTF-8

LazyData TRUE

RoxygenNote 5.0.1.9000

NeedsCompilation no

Author Vitor Onuchic [aut, cre]

Maintainer Vitor Onuchic <vitor.onuchic@bcm.edu>

R topics documented:

---

gene_exp_mixtures *Expression profiles of cell type mixtures*

Description

A matrix containing simulated mixtures of gene expression profiles of 6 different cell types. There are two types of mixtures: those representing methylation profiles of breast tumor samples (200), and those representing normal breast samples (100).

1

Usage

```
gene_exp_mixtures
```

Format

A matrix containing the gene expression profiles of simulated mixtures of cell types. Rows represent different genes. Columns represent differen mixture samples.

Details

Proportions of constituent cell types used to generate these 300 expression profiles were the same ones used to generate the 300 methylation profiles in meth_mixtures, so that for each expression profile there is a matching methylation profile. Proportions of constituent cell types in the 200 samples representing breast tumors were generated from a Dirichlet distribution with parameters that led to mean proportions of 59 epithelial, 9 normal stroma, and 1 the 100 samples representing normal breast were generated from a Dirichlet distribution with parameters that led to mean proportions of 1 epithelial, 1 normal stroma, and 9 log-normal random variables were used to generate noisy versions of the expression profiles of each constituent cell type. Once this procedure was performed for every locus of all expression profiles that would constitute a particular mixture sample, a linear combination of the noisy expression profiles with the set of proportions of each constituent cell type generated for that sample gave us the expression profile of that mixture.

Source

The original gene expression profiles used to create these mixtures came from the deconvoluted expression profiles generated by applying EDec to the TCGA breast tumor dataset. These profiles can be found at: http://genboree.org/theCommons/documents/571. The original profiles can also be found in: true_cell_type_gene_exp.

---

meth_mixtures   *Methylation profiles of cell type mixtures*

Description

A matrix containing simulated mixtures of experimentally derived methylation profiles of 6 different cell types. There are two types of mixtures: those representing methylation profiles of breast tumor samples (200), and those representing normal breast samples (100).

Usage

```
meth_mixtures
```

Format

A matrix containing the methylation profiles of simulated mixtures of cell types. Rows represent Illumina Infinium HumanMethylation27 BeadChip probes. Columns represent different mixture samples.

Details

The mixture samples in this dataset were created by mixing the methylation profiles of representing cancer epithelial, cancer stroma, cancer immune, normal epithelial, normal stroma, and normal immune cell types. Proportions of constituent cell types used to generate these 300 methylation profiles were the same ones used to generate the 300 expression profiles in gene_exp_mixtures, so that for each methylation profile there is a matching expression profile. Proportions of constituent cell types in the 200 samples representing breast tumors were generated from a Dirichlet distribution with parameters that led to mean proportions of 59 epithelial, 9 normal stroma, and 1 the 100 samples representing normal breast were generated from a Dirichlet distribution with parameters that led to mean proportions of 1 epithelial, 1 normal stroma, and 9 random variables were used to generate noisy versions of the methylation profiles of each constituent cell type. Once this procedure was performed for every locus of all methylation profiles that would constitute a particular mixture sample, a linear combination of the noisy methylation profiles with the set of proportions of each constituent cell types generated for that sample gave us the methylation profile of that mixture.

Source

The original methylation profiles used to create the mixtures in this dataset were gathered from different GEO datasets: GSE44837, GSE40699, GSE35069, GSE39981, GSE40699, GSE16368. We used different methylation profiles to represent different cell types in this simulated dataset. Cancer epithelial cell type methylation profile came from HCC38 cell line. Normal epithelial profile came from purified breast luminal epithelial cells. The cancer stromal profile came from IMR90 cell line. The normal stromal profile came from AG04449 cell line. The cancer immune came from purified CD8+ T-cells. The normal immune came from purified monocytes. The original profiles can also be found in: true_cell_type_meth

---

reference_meth *Reference methylation profiles*

Description

This dataset contains methylation profiles for four different classes of cell types: breast cancer cell lines, purified normal breast epithelial cells, stromal cells, and immune cells.

Usage

```
reference_meth
```

Format

A matrix containing methylation profiles for various cell types. Rows represent Illumina Infinium HumanMethylation27 BeadChip probes. Columns represent different cell types.

Source

These methylation profiles were gathered from different GEO datasets: GSE44837, GSE40699, GSE35069, GSE39981, GSE40699, GSE16368.

reference_meth_class	*Cell type classes of reference methylation profiles*

Description

This dataset contains the cell type class associated with each of the reference methylation profiles in the reference_meth dataset.

Usage

```
reference_meth_class
```

Format

A vector of 4 possible strings: "cancer_ep", "normal_ep", "stromal", and "immune". Each string represents a different class of cell type. Each element in this vector corresponds to a column in reference_meth.

true_cell_type_gene_exp	
	True cell type specific gene expression

Description

This dataset contains the gene expression profiles of each constituent cell type used to build the gene_exp_mixtures dataset.

Usage

```
true_cell_type_gene_exp
```

Format

A matrix containing the gene expression profiles of each cell type. Rows represent different genes. Columns represent different cell types.

true_cell_type_meth *True cell type specific methylation*

Description

This dataset contains the methylation profiles of each constituent cell type used to build the meth_mixtures dataset.

Usage

    true_cell_type_meth

Format

A matrix containing the methylation profiles of each cell type. Rows represent Illumina Infinium HumanMethylation27 BeadChip probes. Columns represent different cell types.

true_cell_type_props *True cell type proportions*

Description

This dataset contains the cell type proportions used to build the 300 mixtures in both the meth_mixtures and the gene_exp_mixtures datasets.

Usage

    true_cell_type_props

Format

A matrix containing proportions of constituent cell types in each mixture sample. Rows represent different samples. Columns represent different cell types.

Index