# Package 'sampleClassifierData'

October 27, 2026

**Type** Package

**Title** Pre-processed data for use with the sampleClassifier package

**Version** 1.36.0

**Date** 2021-11-21

**Author** Khadija El Amrani

**Maintainer** Khadija El Amrani <khadija.el-amrani@charite.de>

**Description** This package contains two microarray and two RNA-seq datasets that have been preprocessed for use with the sampleClassifier package. The RNA-seq data are derived from Fagerberg et al. (2014) and the Illumina Body Map 2.0 data. The microarray data are derived from Roth et al. (2006) and Ge et al. (2005).

**License** Artistic-2.0

**Depends** R (>= 3.4), SummarizedExperiment

**Suggests** BiocStyle

**biocViews** ExperimentData, ExpressionData, MicroarrayData, SequencingData, RNASeqData, ArrayExpress

**NeedsCompilation** no

**git_url** https://git.bioconductor.org/packages/sampleClassifierData

**git_branch** RELEASE_3_23

**git_last_commit** c4427e5

**git_last_commit_date** 2026-04-28

**Repository** Bioconductor 3.23

**Date/Publication** 2026-10-27

## Contents

sampleClassifierData-package . . . . . 2
se_micro_refmat . . . . . 3
se_micro_testmat . . . . . 3
se_rnaseq_refmat . . . . . 4
se_rnaseq_testmat . . . . . 5

**Index** **6**

---

`sampleClassifierData-package`
*Pre-processed data for use with the sampleClassifier package*

---

## Description

This package contains two microarray and two RNA-seq datasets that have been preprocessed for use with the sampleClassifier package. The RNA-seq data are derived from Fagerberg et al. (2014) and the Illumina Body Map 2.0 data. The microarray data are derived from Roth et al. (2006) and Ge et al. (2005).

## Details

This package contains a collection of publicly available microarray and RNA-seq datasets that have been pre-processed for use with the R-package 'sampleClassifier'.

The dataset `se_rnaseq_refmat` contains 71 samples from 24 tissues, derived from Fagerberg et al. (2014). Each tissue is represented by 3 replicates (except ovary which is represented by 2 replicates).

The dataset `se_micro_refmat` contains 78 samples from 26 tissues, derived from Roth et al. (2006). Each tissue is represented by 3 replicates.

The dataset `se_rnaseq_testmat` contains 12 samples from the Illumina Body Map 2.0 data.

The dataset `se_micro_testmat` contains 16 samples from Ge et al. (2005).

See the package vignette for details on the pre-processing of the data.

## Author(s)

Khadija El Amrani

Maintainer: Khadija El Amrani <khadija.el-amrani@charite.de>

## References

Fagerberg, L., Hallstroem, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjoestedt, E., Lundberg, E., Szigyarto, C.A.-K., Skogs, M., Takanen, J.O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J.M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., Uhlen, M.: Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular & cellular proteomics : MCP 13(2), 397-406 (2014).

Roth, R.B., Hevezi, P., Lee, J., Willhite, D., Lechner, S.M., Foster, A.C., Zlotnik, A.: Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 7(2), 67-80 (2006). doi:10.1007/s10048-006-0032-6

The Illumina Body Map 2.0 data (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/).

Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, et al. Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics 2005; 86:127-141. doi: 10.1016/j.ygeno.2005.04.008. pmid:15950434

---

`se_micro_refmat` *microarray gene expression data set*

---

## Description

Microarray gene expression data set with 78 samples from 26 tissues, derived from Roth et al. (2006). Each tissue is represented by 3 replicates.

## Usage

```
data(se_micro_refmat)
```

## Format

A `matrix` with 54675 probesets and 78 samples.

## Value

microarray data matrix as SummarizedExperiment

## References

Roth, R.B., Hevezi, P., Lee, J., Willhite, D., Lechner, S.M., Foster, A.C., Zlotnik, A.: Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 7(2), 67-80 (2006). doi:10.1007/s10048-006-0032-6

## Examples

```
data(se_micro_refmat)
```

---

`se_micro_testmat` *microarray gene expression data set*

---

## Description

Microarray gene expression data set with 16 samples from Ge et al. (2005).

## Usage

```
data(se_micro_testmat)
```

## Format

A matrix with 22283 probesets and 16 samples.

## Value

microarray data matrix as SummarizedExperiment

## References

Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, et al. Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics 2005; 86:127-141. doi: 10.1016/j.ygeno.2005.04.008. pmid:15950434

## Examples

```
data(se_micro_testmat)
```

---

# se_rnaseq_refmat *RNA-seq gene expression data set*

---

## Description

RNA-seq gene expression data set with 71 samples from 24 tissues, derived from Fagerberg et al. (2014). Each tissue is represented by 3 replicates (except ovary which is represented by 2 replicates).

## Usage

```
data(se_rnaseq_refmat)
```

## Format

A matrix with 43819 genes and 71 samples.

## Value

RNA-seq data matrix as SummarizedExperiment

## References

Fagerberg, L., Hallstroem, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjoestedt, E., Lundberg, E., Szigyarto, C.A.-K., Skogs, M., Takanen, J.O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J.M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., Uhlen, M.: Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular & cellular proteomics : MCP 13(2), 397-406 (2014).

## Examples

```
data(se_rnaseq_refmat)
```

---

`se_rnaseq_testmat` *RNA-seq gene expression data set*

---

## Description

RNA-seq gene expression data set with 12 samples from the Illumina Body Map 2.0 data.

## Usage

```
data(se_rnaseq_testmat)
```

## Format

A `matrix` with 43819 genes and 12 samples.

## Value

RNA-seq data matrix as SummarizedExperiment

## References

The Illumina Body Map 2.0 data (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/).

## Examples

```
data(se_rnaseq_testmat)
```

# Index

∗ **dataset**
    se_micro_refmat, 3
    se_micro_testmat, 3
    se_rnaseq_refmat, 4
    se_rnaseq_testmat, 5

matrix, *3–5*

sampleClassifierData
        *(*sampleClassifierData-package*)*, 2
sampleClassifierData-package, 2
se_micro_refmat, 3
se_micro_testmat, 3
se_rnaseq_refmat, 4
se_rnaseq_testmat, 5