

# Package ‘mcsurvdata’

July 24, 2025

**Type** Package

**Version** 1.26.0

**Date** 2023-07-20

**Title** Meta cohort survival data

**Description** This package stores two merged expressionSet objects that contain the gene expression profile and clinical information of -a- six breast cancer cohorts and -b- four colorectal cancer cohorts. Breast cancer data are employed in the vignette of the hrnbiased package for survival analysis of gene signatures.

**VignetteBuilder** knitr

**License** GPL (>=2)

**NeedsCompilation** no

**biocViews** ExperimentData, Homo\_sapiens\_Data, GEO, MicroarrayData

**URL** <https://github.com/adricaba/mcsurvdata>

**Depends** R (>= 3.5), ExperimentHub

**Suggests** BiocStyle, knitr

**Imports** AnnotationHub, Biobase

**git\_url** <https://git.bioconductor.org/packages/mcsurvdata>

**git\_branch** RELEASE\_3\_21

**git\_last\_commit** ee330c9

**git\_last\_commit\_date** 2025-04-15

**Repository** Bioconductor 3.21

**Date/Publication** 2025-07-24

**Author** Adria Caballe Mestres [aut, cre],  
Antoni Berenguer Llergo [aut],  
Camille Stephan-Otto Attolini [aut]

**Maintainer** Adria Caballe Mestres <[adria.caballe@irbbarcelona.org](mailto:adria.caballe@irbbarcelona.org)>

Contents

|                      |          |
|----------------------|----------|
| mcsurvdata . . . . . | 2        |
| nda.brca . . . . .   | 2        |
| nda.crc . . . . .    | 4        |
| <b>Index</b>         | <b>6</b> |

---

|            |                                                                                                                       |
|------------|-----------------------------------------------------------------------------------------------------------------------|
| mcsurvdata | <i>Processed gene expression data and clinical information of several breast cancer and colorectal cancer cohorts</i> |
|------------|-----------------------------------------------------------------------------------------------------------------------|

---

Description

Contains the processed gene expression data and clinical data from six breast cancer studies [nda.brca](#) as well as from four colon cancer studies [nda.crc](#).

Details

ExpressionSet objects with merged data

Examples

```
library(ExperimentHub)
eh <- ExperimentHub()
nda.brca <- query(eh, "mcsurvdata")[["EH1497"]]
nda.crc <- query(eh, "mcsurvdata")[["EH1498"]]
```

---

|          |                                                                                             |
|----------|---------------------------------------------------------------------------------------------|
| nda.brca | <i>Processed gene expression data and clinical information of six breast cancer cohorts</i> |
|----------|---------------------------------------------------------------------------------------------|

---

Description

Processed gene expression data and clinical data of 2294 patients from six breast cancer studies. These include GSE1456, GSE2034, GSE2990, GSE3494, GSE7390 and the metabric. Normalization is done by adjusting each gene by technical covariates such as Eklund metrics and scanning day in basis of a mixed effects model. Data merging is done by standardizing the gene expression matrix of GSE1456, GSE2034, GSE2990, GSE3494 and GSE7390 to a reference dataset (metabric). Only ER+ samples are included. Data in this package are used for the hrunbiased R package vignette

## Details

ExpressionSet object with merged data from studies:

- GSE1456
- GSE2034
- GSE2990
- GSE3494
- GSE7390
- metabric

## Author(s)

Adria Caballe Mestres, Antoni Berenguer Llergo, Camille Stephan-Otto Attolini.

## References

- Caballe Mestres A, Berenguer Llergo A and Stephan-Otto Attolini C. Adjusting for systematic technical biases in risk assessment of gene signatures in transcriptomic cancer cohorts. *bioRxiv* (2018).
- Eklund A. and Szallasi Z. Correction of technical bias in clinical microarray data improves concordance with known biological information. *Genome Biology* 9, R26 (2008).
- Pawitan, Y. et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Research* 7, R953 (2005).
- Wang, Y. et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* 365, 671-679 (2005).
- Sotiriou, C. et al. Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. *JNCI: Journal of the National Cancer Institute* 98, 262-272 (2006).
- Miller, L. D. et al. From The Cover: An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proceedings of the National Academy of Sciences* 102, 13550-13555 (2005).
- Desmedt, C. et al. Strong Time Dependence of the 76-Gene Prognostic Signature for Node-Negative Breast Cancer Patients in the TRANSBIG Multicenter Independent Validation Series. *Clinical Cancer Research* 13, 3207-3214 (2007).
- Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486, 346-352 (2012).

## Examples

```
library(ExperimentHub)
eh <- ExperimentHub()
nda.brca <- query(eh, "mcsurvdata")[[ "EH1497" ]]

# survival info
cbind(nda.brca$evn, nda.brca$tev)
```

---

nda.crc*Processed gene expression data and clinical information of four colon cancer cohorts*

---

## Description

Processed gene expression data and clinical data of 914 patients from four colorectal cancer studies. These include GSE14333, GSE33113, GSE37892 and GSE39582. Normalization is done by adjusting each gene by technical covariates such as Eklund metrics and scanning day in basis of a mixed effects model. Data merging is done by standardizing the gene expression matrix of GSE14333, GSE33113 and GSE37892 to a reference dataset (GSE39582). Only MSS samples are included.

## Details

ExpressionSet object with merged data from studies:

- GSE14333
- GSE33113
- GSE39582
- GSE37892

## Author(s)

Adria Caballe Mestres, Antoni Berenguer Llergo, Camille Stephan-Otto Attolini.

## References

- Caballe Mestres A, Berenguer Llergo A and Stephan-Otto Attolini C. Adjusting for systematic technical biases in risk assessment of gene signatures in transcriptomic cancer cohorts. *bioRxiv* (2018).
- Eklund A. and Szallasi Z. Correction of technical bias in clinical microarray data improves concordance with known biological information. *Genome Biology* 9, R26 (2008)
- Jorissen, R. N. et al. Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. *Clinical Cancer Research* 15, 7642-7651 (2009).
- De Sousa E Melo, F. et al. Methylation of cancer-stem-cell-associated wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* 9, 476-485 (2011).
- Marisa, L. et al. Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLoS Medicine* 10 (2013).
- Laibe, S. et al. A seven-gene signature aggregates a subgroup of stage II colon cancers with stage III. *Omics : a journal of integrative biology* 16, 560-5 (2012).

**Examples**

```
library(ExperimentHub)
eh <- ExperimentHub()
nda.crc <- query(eh, "mcsurvdata")[["EH1498"]]

# survival info
cbind(nda.crc$evn,nda.crc$tev)
```

# Index

## \* **datasets**

mcsurvdata, [2](#)

nda.brca, [2](#)

nda.crc, [4](#)

## \* **package**

mcsurvdata, [2](#)

nda.brca, [2](#)

nda.crc, [4](#)

mcsurvdata, [2](#)

mcsurvdata-package (mcsurvdata), [2](#)

nda.brca, [2](#), [2](#)

nda.crc, [2](#), [4](#)