

# Package ‘RUCova’

April 24, 2025

**Title** Removes unwanted covariance from mass cytometry data

**Version** 1.0.0

## Description

Mass cytometry enables the simultaneous measurement of dozens of protein markers at the single-cell level, producing high dimensional datasets that provide deep insights into cellular heterogeneity and function. However, these datasets often contain unwanted covariance introduced by technical variations, such as differences in cell size, staining efficiency, and instrument-specific artifacts, which can obscure biological signals and complicate downstream analysis. This package addresses this challenge by implementing a robust framework of linear models designed to identify and remove these sources of unwanted covariance. By systematically modeling and correcting for technical noise, the package enhances the quality and interpretability of mass cytometry data, enabling researchers to focus on biologically relevant signals.

**License** GPL-3

**Encoding** UTF-8

**biocViews** Software, SingleCell

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.2

**Suggests** knitr, rmarkdown, BiocManager, BiocStyle, remotes, ggpubr, ggcorrplot, ggh4x, testthat (>= 3.0.0)

**Imports** dplyr, fastDummies, ggplot2, stringr, tibble, Matrix, ComplexHeatmap, grid, circlize, SingleCellExperiment, SummarizedExperiment, tidyverse, tidyr, magrittr, S4Vectors

**VignetteBuilder** knitr

**Depends** R (>= 4.4.0)

**SystemRequirements** GNU make

**BugReports** <https://github.com/molysbio/RUCova/issues>

**URL** <https://github.com/molysbio/RUCova>

**LazyData** true

**Config/testthat/edition** 3

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

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

**git\_last\_commit** 6769220

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

**Repository** Bioconductor 3.21

**Date/Publication** 2025-04-23

**Author** Rosario Astaburuaga-García [aut, cre] (ORCID:  
<<https://orcid.org/0000-0003-1179-4080>>)

**Maintainer** Rosario Astaburuaga-García <[rosario.astaburuaga@charite.de](mailto:rosario.astaburuaga@charite.de)>

## Contents

|                                |    |
|--------------------------------|----|
| RUCova-package . . . . .       | 2  |
| calc_mean_BC . . . . .         | 3  |
| calc_mean_DNA . . . . .        | 4  |
| compare_corr . . . . .         | 5  |
| heatmap_compare_corr . . . . . | 6  |
| HNSCC_data . . . . .           | 7  |
| rucova . . . . .               | 9  |
| sce . . . . .                  | 10 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>11</b> |
|--------------|-----------|

---

|                |  |
|----------------|--|
| RUCova-package | <i>RUCova: Removing Unwanted Covariance in Mass Cytometry Data</i> |
|----------------|--|

---

## Description

The RUCova package provides tools for removing unwanted covariance in mass cytometry (CyTOF) data. It is designed to help researchers preprocess and analyze CyTOF data by identifying and correcting for technical and biological sources of unwanted variation.

## Key Features

- Correction of technical and biological variation.
- Visualization tools for assessing data quality.

## Getting Started

To get started with RUCova, load the package and explore the example dataset:

```
library(RUCova)
data(HNSCC_data) # Example dataset
head(HNSCC_data)
```

For a detailed workflow, see the vignette:

```
vignette("RUCova")
```

**Author(s)**

- Rosario Astaburuaga-García (maintainer)

**References**

For more information on the methods used in this package, see:

- Astaburuaga-García, R. et al. (2023). "RUCova: A tool for removing unwanted covariance in CyTOF data." *Bioinformatics*.

**Author(s)**

**Maintainer:** Rosario Astaburuaga-García <rosario.astaburuaga@charite.de> ([ORCID](#))

**See Also**

Useful links:

- <https://github.com/molsysbio/RUCova>
- Report bugs at <https://github.com/molsysbio/RUCova/issues>

---

calc\_mean\_BC

*Calculated mean of normalised highest BC per cell*

---

**Description**

Calculated mean of normalised highest BC per cell

**Usage**

```
calc_mean_BC(sce, name_assay = "counts", bc_channels, n_bc, q = 0.95)
```

**Arguments**

|             |   |
|-------------|---|
| sce         | A SingleCellExperiment object with markers and SUCs in linear scale stored in the assay "name_assay". A sinh transformation is applied within the function. |
| name_assay  | A string specifying the name of the assay including the BC channels in linear scale. Default is "counts".   |
| bc_channels | Vector specifying the names of the BC channels  |
| n_bc        | number of barcoding isotopes per cell. n_bc = 3 for the Fluidigm kit.   |
| q           | Quantile for normalisation. Default is 0.95.  |

**Value**

The SingleCellExperiment object with an extra column "mean\_BC" in the corresponding assay.

**Examples**

```
sce <- RUCova::sce
bc_channels <- c(c("Pd102Di", "Pd104Di", "Pd105Di", "Pd106Di", "Pd108Di", "Pd110Di"),
  c("Dead_cells_194Pt", "Dead_cells_198Pt")
)
sce <- RUCova::calc_mean_BC(sce, name_assay = "counts", bc_channels, n_bc = 4, q = 0.95)
```

---

calc\_mean\_DNA

*Calculated mean of normalised Iridium isotopes*


---

**Description**

Calculated mean of normalised Iridium isotopes

**Usage**

```
calc_mean_DNA(sce, name_assay = "counts", dna_channels, q)
```

**Arguments**

|              |  |
|--------------|--|
| sce          | A SingleCellExperiment object with markers and SUCs in linear scale stored in the assay "name_assay". Asinh transformation is applied within the function. |
| name_assay   | A string specifying the name of the assay including the DNA channels in linear scale.  |
| dna_channels | Vector specifying the names of the DNA channels  |
| q            | Quantile for normalisation.  |

**Value**

The SingleCellExperiment object with an extra column "mean\_BC" in the corresponding assay.

**Examples**

```
sce <- RUCova::sce
dna_channels <- c("DNA_191Ir", "DNA_193Ir")
sce <- RUCova::calc_mean_DNA(sce, name_assay = "counts", dna_channels, q = 0.95)
```

---

|              |  |
|--------------|--|
| compare_corr | <i>Get pearson correlation coefficients between markers on a double triangular matrix for comparison (lower triangle: before RUCova, upper triangle: after RUCova). If RUCova has not been applied, the output is a symmetric matrix</i> |
|--------------|--|

---

### Description

Get pearson correlation coefficients between markers on a double triangular matrix for comparison (lower triangle: before RUCova, upper triangle: after RUCova). If RUCova has not been applied, the output is a symmetric matrix

### Usage

```
compare_corr(
  sce,
  name_assay_before = "counts",
  name_assay_after = NULL,
  name_reduced_dim = NULL
)
```

### Arguments

|                   |  |
|-------------------|--|
| sce               | A SingleCellExperiment object with markers and SUCs in linear scale stored in the assay "name_assay". Asinh transformation is applied within the function. |
| name_assay_before | A string specifying the name of the assay before RUCova (with original counts in linear scale).  |
| name_assay_after  | A string specifying the name of the assay after RUCova (with original counts in linear scale).   |
| name_reduced_dim  | A string specifying the name of the dimensionality reduction data stored under reducedDim(). If "PCA", then PCs will be included in the heatmap.           |

### Value

#A matrix with pearson correlation coefficients.

### Examples

```
sce <- RUCova::sce
bc_channels <- c("Pd102Di", "Pd104Di", "Pd105Di", "Pd106Di", "Pd108Di", "Pd110Di",
"Dead_cells_194Pt", "Dead_cells_198Pt")
sce <- RUCova::calc_mean_BC(sce, name_assay = "counts", bc_channels, n_bc = 4, q = 0.95)
dna_channels <- c("DNA_191Ir", "DNA_193Ir")
sce <- RUCova::calc_mean_DNA(sce, name_assay = "counts", dna_channels, q = 0.95)
```

```

# Markers:
m <- c("pH3", "IdU", "Cyclin_D1", "Cyclin_B1", "Ki.67", "pRb", "pH2A.X", "p.p53", "p.p38",
"pChk2", "pCDC25c", "cCasp3", "cPARP", "pAkt", "pAkt_T308", "pMEK1.2", "pERK1.2", "pS6", "p4e.BP1",
"pSmad1.8", "pSmad2.3", "pNFkB", "IkBa", "CXCL1", "Lamin_B1", "pStat1", "pStat3", "YAP", "NICD")
# SUCs::
x <- c("total_ERK", "pan_Akt", "mean_DNA", "mean_BC")
sce <- RUCova::rucova(sce = sce, name_assay_before = "counts", markers = m, SUCs = x,
apply_asinh_SUCs = TRUE, model = "interaction", center_SUCs = "across_samples",
col_name_sample = "line", name_assay_after = "counts_interaction")
compare_corr(sce[,sce$line == "Cal33"], name_assay_before = "counts", name_assay_after = "counts_interaction")
heatmap_compare_corr(sce[,sce$line == "Cal33"], name_assay_before = "counts", name_assay_after = "counts_interact

```

---

`heatmap_compare_corr` *Plot pearson correlation coefficients between markers on a double triangular heatmap (lower triangle: before RUCova, upper triangle: after RUCova). If RUCova has not been applied, the output is a symmetric heatmap.*

---

## Description

Plot pearson correlation coefficients between markers on a double triangular heatmap (lower triangle: before RUCova, upper triangle: after RUCova). If RUCova has not been applied, the output is a symmetric heatmap.

## Usage

```

heatmap_compare_corr(
  sce,
  name_assay_before = "counts",
  name_assay_after = NULL,
  name_reduced_dim = NULL
)

```

## Arguments

`sce` A `SingleCellExperiment` object with markers and SUCs in linear scale stored in the assay "name\_assay". Asinh transformation is applied within the function.

`name_assay_before` A string specifying the name of the assay before RUCova (with original counts in linear scale).

`name_assay_after` A string specifying the name of the assay before RUCova (with original counts in linear scale).

`name_reduced_dim` A string specifying the name of the dimensionality reduction data stored under `reducedDim()`. If "PCA", then PCs will be included in the heatmap.

**Value**

#A heatmap with pearson correlation coefficients.

**Examples**

```
sce <- RUCova::sce
bc_channels <- c("Pd102Di", "Pd104Di", "Pd105Di", "Pd106Di", "Pd108Di", "Pd110Di",
"Dead_cells_194Pt", "Dead_cells_198Pt")
sce <- RUCova::calc_mean_BC(sce, name_assay = "counts", bc_channels, n_bc = 4, q = 0.95)
dna_channels <- c("DNA_191Ir", "DNA_193Ir")
sce <- RUCova::calc_mean_DNA(sce, name_assay = "counts", dna_channels, q = 0.95)
# Markers:
m <- c("pH3", "IdU", "Cyclin_D1", "Cyclin_B1", "Ki.67", "pRb", "pH2A.X", "p.p53", "p.p38",
"pChk2", "pCDC25c", "cCasp3", "cPARP", "pAkt", "pAkt_T308", "pMEK1.2", "pERK1.2", "pS6", "p4e.BP1",
"pSmad1.8", "pSmad2.3", "pNFkB", "IkBa", "CXCL1", "Lamin_B1", "pStat1", "pStat3", "YAP", "NICD")
# SUCs:
x <- c("total_ERK", "pan_Akt", "mean_DNA", "mean_BC")
sce <- RUCova::rucova(sce = sce, name_assay_before = "counts", markers = m, SUCs = x,
apply_asinh_SUCs = TRUE, model = "interaction", center_SUCs = "across_samples",
col_name_sample = "line", name_assay_after = "counts_interaction")
heatmap_compare_corr(sce[,sce$line == "Cal33"], name_assay_before = "counts", name_assay_after = "counts_interact
```

---

HNSCC\_data

*HNSCC data set*


---

**Description**

Is a tibble containing mass cytometry data of single-cell marker signals (rows = cells, columns = markers and metadata) in linear scale. This data set should be clean, meaning you excluded beads, debris, doublets, dead cells, and single-cells are demultiplexed (important if you want to adapt the linear fits to the samples). In this example we offer a mass cytometry data set consisting of 8 Head-and-Neck Squamous Cell Carcinoma (HNSCC) lines in irradiated (10 Gy) and control (0 Gy) conditions (Figure 2 and Figure 3 in the manuscript).

**Usage**

```
HNSCC_data
```

**Format**

A data frame with 108649 rows and 59 variables: #

**CXCL1** Signal for the CXCL1 marker, a cytokine associated with inflammatory responses.

**Ce140Di** Signal for the Cerium 140 marker, typically used as a control or calibration marker.

**Cyclin\_B1** Marker for the G2/M cell cycle phase regulator Cyclin B1.

**Cyclin\_D1** Marker for the G1 cell cycle phase regulator Cyclin D1.

**DNA\_191Ir, DNA\_193Ir** Signals for DNA-intercalating markers labeled with Iridium isotopes, used to assess nuclear content.

**Dead\_cells\_194Pt, Dead\_cells\_195Pt, Dead\_cells\_196Pt, Dead\_cells\_198Pt** Signals for dead cell markers labeled with Platinum isotopes, used to identify and exclude dead cells.

**Event\_length** Measure of event duration in the mass cytometer, used for quality control.

**GDF15** Signal for GDF15, a marker associated with stress and inflammation.

**IdU** Signal for IdU (Iododeoxyuridine), used to assess DNA synthesis.

**IkBa** Signal for IkBa, an inhibitor of the NF-kB signaling pathway.

**Ki.67** Signal for Ki-67, a marker of cell proliferation.

**Lamin\_B1** Signal for Lamin B1, a nuclear lamina protein.

**NICD** Signal for the Notch Intracellular Domain, a marker of Notch signaling activity.

**Pd102Di, Pd104Di, Pd105Di, Pd106Di, Pd108Di, Pd110Di** Signals for Palladium isotopes, used as barcodes for cell multiplexing.

**Pt194Di\_norm** Normalized signal for Platinum 194 isotope.

**Time** Timestamp for the acquisition of each event.

**YAP** Signal for Yes-associated protein (YAP), a transcriptional co-activator in the Hippo pathway.

**bc** Barcode channel used for multiplexing.

**beadsOut** Indicator for exclusion due to bead contamination.

**cCasp3** Signal for cleaved Caspase-3, a marker of apoptosis.

**cPARP** Signal for cleaved PARP, another marker of apoptosis.

**dose** Treatment condition: either 0Gy (control) or 10Gy (irradiated).

**irradiated** Indicator for whether the sample was irradiated.

**line** HNSCC cell lines included in the data: "Cal27", "Cal33", "UPCISCC099", "UPCISCC131", "UTSCC16A", "UDSCC2", "UPCISCC154", and "VUSCC147".

**lowPt** Indicator of low Platinum signal.

**p.p38, p.p53, p4e.BP1** Signals for phosphorylated p38, p53, and 4E-BP1, markers of stress response and translation regulation.

**pAkt, pAkt\_T308** Signals for phosphorylated Akt, a marker of PI3K/Akt pathway activity.

**pCDC25c, pChk2** Signals for phosphorylated CDC25c and Chk2, markers of DNA damage response.

**pERK1.2** Signal for phosphorylated ERK1/2, a marker of MAPK pathway activity.

**pH2A.X, pH3** Signals for phosphorylated H2A.X and H3, markers of DNA damage and mitosis, respectively.

**pMEK1.2, pNFkB** Signals for phosphorylated MEK1/2 and NF-kB, markers of MAPK and inflammatory signaling.

**pRb** Signal for phosphorylated Rb, a marker of cell cycle regulation.

**pS6** Signal for phosphorylated S6, a marker of protein synthesis.

**pSmad1.8, pSmad2.3** Signals for phosphorylated Smad1/8 and Smad2/3, markers of TGF-beta signaling.

**pStat1, pStat3** Signals for phosphorylated Stat1 and Stat3, markers of JAK/STAT pathway activity.

**pan\_Akt** Total Akt protein signal.

**replicate** Identifier for technical replicates.

**singlets** Indicator for singlet events, excluding doublets.

**total\_ERK** Total ERK protein signal.



---

rucova *Remove unwanted covariance*

---

### Usage

```
rucova(
  sce,
  name_assay_before = "counts",
  markers,
  SUCs = c("mean_DNA", "mean_BC", "total_ERK", "pan_Akt"),
  name_reduced_dim = "PCA",
  apply_asinh_SUCs = TRUE,
  model = "interaction",
  col_name_sample = "line",
  center_SUCs = "across_samples",
  keep_offset = TRUE,
  name_assay_after = "counts_rucova"
)
```

### Arguments

|                   |   |
|-------------------|---|
| sce               | A SingleCellExperiment object with markers and SUCs in linear scale stored in the assay "name_assay_before". Asinh transformation is applied within the function. |
| name_assay_before | A string specifying the name of the assay before RUCova (with original counts in linear scale).   |
| markers           | Vector of marker names to normalise, y (in linear scale).   |
| SUCs              | Vector of surrogates of unwanted covariance to use for normalisation, x (in linear scale).  |
| name_reduced_dim  | string specifying the name of the dimensionality reduction result in the SingleCellExperiment sce.  |
| apply_asinh_SUCs  | Apply (TRUE) or not (FALSE) asinh transformation to the SUCs. TRUE if SUCs are the measured surrogates, FALSE if SUCs are PCs.                                    |
| model             | A character: "simple", "offset" or "interaction" defining the model.  |
| col_name_sample   | A character indicating the column name in "data" defining each sample.  |
| center_SUCs       | A character "across_samples" or "per_sample" defining how to center the SUCs in zero.   |
| keep_offset       | Keep (TRUE) or not (FALSE) the offset intercept between samples.+   |
| name_assay_after  | A string specifying the name of the assay after RUCova (with regressed counts in linear scale).   |

**Value**

The input `SingleCellExperiment` object with an additional assay (`name_assay_after`) and a list in the metadata containing all the model details.

**Examples**

```
sce <- RUCova::sce
bc_channels <- c("Pd102Di", "Pd104Di", "Pd105Di", "Pd106Di", "Pd108Di", "Pd110Di",
"Dead_cells_194Pt", "Dead_cells_198Pt")
sce <- RUCova::calc_mean_BC(sce, name_assay = "counts", bc_channels, n_bc = 4, q = 0.95)
dna_channels <- c("DNA_191Ir", "DNA_193Ir")
sce <- RUCova::calc_mean_DNA(sce, name_assay = "counts", dna_channels, q = 0.95)
# Markers:
m <- c("pH3", "IdU", "Cyclin_D1", "Cyclin_B1", "Ki.67", "pRb", "pH2A.X", "p.p53", "p.p38", "pChk2",
"pCDC25c", "cCasp3", "cPARP", "pAkt", "pAkt_T308", "pMEK1.2", "pERK1.2", "pS6", "p4e.BP1", "pSmad1.8",
"pSmad2.3", "pNFkB", "IkBa", "CXCL1", "Lamin_B1", "pStat1", "pStat3", "YAP", "NICD")
# SUCs::
x <- c("total_ERK", "pan_Akt", "mean_DNA", "mean_BC")
sce <- RUCova::rucova(sce = sce, name_assay_before = "counts", markers = m, SUCs = x,
apply_asinh_SUCs = TRUE, model = "interaction", center_SUCs = "across_samples",
col_name_sample = "line", name_assay_after = "counts_interaction")
```

---

sce

*SingleCellExperiment Object with HNSCC Data Set*

---

**Description**

A `SingleCellExperiment` object containing mass cytometry data of single-cell marker signals. The data includes one or more assays (e.g., "counts") with signals in linear scale. Rows represent markers, and columns represent cells. The data is clean, excluding calibration beads, debris, doublets, and dead cells. Single cells are demultiplexed, which is important for adapting linear fits to samples. Metadata such as samples and treatment conditions are stored in the `colData`.

**Usage**

```
sce
```

**Format**

A `SingleCellExperiment` object with the following components:

**assays** One or more assays, such as "counts", containing the marker signals.

**colData** Column metadata, including cell annotations such as `cell_id`, `line`, and `dose`.

**rowData** Row metadata, including marker annotations.

# Index

## \* **datasets**

HNSCC\_data, [7](#)

sce, [10](#)

## \* **internal**

RUCova-package, [2](#)

calc\_mean\_BC, [3](#)

calc\_mean\_DNA, [4](#)

compare\_corr, [5](#)

heatmap\_compare\_corr, [6](#)

HNSCC\_data, [7](#)

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

rucova, [9](#)

RUCova-package, [2](#)

sce, [10](#)