

Package: RPBC (via r-universe)

October 12, 2026

Title Region-Preserving Batch Correction for Paired Region-Resolved Single-Cell Data

Version 0.99.0

Description Implements Region-Preserving Batch Correction (RPBC), an empirical-Bayes location-scale framework for paired region-resolved single-cell and spatial-omics data. RPBC preserves a patient-balanced population-average regional contrast while harmonizing patient-wide location shifts, patient-specific regional deviations, and patient-region scale heterogeneity. The package also provides Adaptive RPBC for group-size-dependent stabilization of patient-region scale variation and covariate-adjusted RPBC for preserving prespecified patient-level covariate-associated variation during harmonization. RPBC is designed to interoperate with Bioconductor data structures for single-cell and spatial experiments.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

Config/roxygen2/version 8.1.0

biocViews BatchEffect, Preprocessing, SingleCell, Spatial, StatisticalMethod

Suggests BiocStyle, knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports matrixStats, SingleCellExperiment, SpatialExperiment, SummarizedExperiment

Depends R (>= 4.5.0)

VignetteBuilder knitr

Config/pak/sysreqs libmagick++-dev gsfonts libicu-dev libssl-dev zlib1g-dev

Repository <https://biocstaging.r-universe.dev>

Date/Publication 2026-10-12 19:37:57 UTC

RemoteUrl `https://github.com/BiocStaging/RPBC`
RemoteRef `HEAD`
RemoteSha `8db3a51a4014c1ffe88f70994e83ab089ef9a40`

Contents

haley_gbm_example	2
rpbc	3
Index	6

haley_gbm_example	<i>Haley GBM IMC Example Dataset</i>
-------------------	--------------------------------------

Description

A representative example dataset for demonstrating Region-Preserving Batch Correction (RPBC) using glioblastoma imaging mass cytometry (IMC) data. The dataset retains all seven patients and the paired Core/Edge design, with 30 cells sampled from each patient-region combination.

Usage

`data(haley_gbm_example)`

Format

A `SpatialExperiment` object with 10 markers (rows) and 420 cells (columns). The "counts" assay contains raw IMC intensities.

Cell-level metadata include:

- cell_id** Unique cell identifier.
- ROI** Region-of-interest identifier.
- patient_id** Patient identifier.
- region_type** Paired tumor region, coded as "Core" or "Edge".
- age** Patient age.
- sex** Patient sex.
- mgmt_status** MGMT status.

Details

The dataset contains 10 representative biological markers measured across 420 cells. Raw IMC intensities are stored in the "counts" assay, allowing preprocessing transformations to be applied directly through `rpbcb()`.

This example dataset was constructed by sampling 30 cells from each patient-region combination while retaining all seven patients and both Core and Edge regions. It is intended for package examples, testing, and vignettes and does not represent the complete Haley GBM IMC dataset.

The 10 included markers are CD45, CD3, Cd8a, CD68, CD163, GFAP, Sox2, Olig2, Ki67, and Vimentin.

Source

Haley MJ, et al. (2024). Hypoxia coordinates the spatial landscape of myeloid cells within glioblastoma to affect survival. *Science Advances*, 10(20), eadj3301. <https://doi.org/10.1126/sciadv.adj3301>

Haley M, Couper K. (2024). Imaging mass cytometry data from IDH wildtype glioblastomas. Dryad. <https://doi.org/10.5061/dryad.2z34tmprw>

rpbcb

Region-Preserving Batch Correction

Description

Apply Region-Preserving Batch Correction (RPBC) to paired region-resolved single-cell or spatial data.

Usage

```
rpbcb(  
  object,  
  patient,  
  region,  
  assay_name = "counts",  
  transform = c("none", "asinh", "log1p"),  
  cofactor = 5,  
  covariates = NULL,  
  adaptive = FALSE,  
  output_assay = "rpbcb",  
  core_level = "Core",  
  edge_level = "Edge",  
  conv = 1e-04,  
  max_iter = 100L,  
  eps = 1e-08,  
  delta_adaptive_k = 1,  
  delta_bounds = c(0.5, 2),  
  return_fit = FALSE  
)
```

Arguments

<code>object</code>	A <code>SingleCellExperiment</code> or <code>SpatialExperiment</code> object.
<code>patient</code>	Character string naming the column in <code>colData(object)</code> containing patient identifiers.
<code>region</code>	Character string naming the column in <code>colData(object)</code> containing the paired region labels.
<code>assay_name</code>	Character string naming the assay to use as input. Default is "counts". For count-based sequencing data, an appropriately normalized and log-transformed assay such as "logcounts" (e.g., $\log_2(\text{CPM} + 1)$) may be supplied with <code>transform = "none"</code> .
<code>transform</code>	Character string specifying the transformation applied to the selected assay before RPBC harmonization. One of "none", "asinh" or "log1p". Default is "none".
<code>cofactor</code>	Positive numeric value used when <code>transform = "asinh"</code> . The transformation is $\text{asinh}(x / \text{cofactor})$. Default is 5.
<code>covariates</code>	Optional character vector naming patient-level covariates in <code>colData(object)</code> . Covariates supplied at the cell level must be constant within patient. If <code>NULL</code> , ordinary RPBC is used.
<code>adaptive</code>	Logical. If <code>TRUE</code> , use Adaptive RPBC. Adaptive RPBC cannot currently be combined with covariate adjustment.
<code>output_assay</code>	Character string giving the name of the assay in which the harmonized expression matrix will be stored. Default is "rpbc".
<code>core_level</code>	Character string identifying the Core region.
<code>edge_level</code>	Character string identifying the Edge region.
<code>conv</code>	Positive convergence tolerance for empirical-Bayes estimation.
<code>max_iter</code>	Positive integer giving the maximum number of empirical-Bayes iterations.
<code>eps</code>	Small positive numerical constant used for numerical stability.
<code>delta_adaptive_k</code>	Positive Adaptive RPBC stabilization constant. The validated default is 1.
<code>delta_bounds</code>	Numeric vector of length two giving the lower and upper bounds for Adaptive RPBC patient-region variance multipliers. The validated default is <code>c(0.5, 2.0)</code> .
<code>return_fit</code>	Logical. If <code>FALSE</code> , return the modified Bioconductor object. If <code>TRUE</code> , return a list containing both the modified object and the full internal RPBC fit.

Details

RPBC harmonizes patient and patient-by-region nuisance variation while preserving the prespecified population-average regional effect.

The function supports:

- ordinary RPBC;
- covariate-adjusted RPBC; and

- Adaptive RPBC.

Ordinary and covariate-adjusted RPBC use the same internal statistical engine. Adaptive RPBC uses a separate internal engine because its patient-region scale update includes group-size-dependent stabilization.

Value

By default, the input object with the harmonized expression matrix stored in `output_assay`. If `return_fit = TRUE`, a list containing object and fit.

Examples

```
data("haley_gbm_example", package = "RPBC")
```

```
# Ordinary RPBC
```

```
example_ordinary <- rpbc(  
  object = haley_gbm_example,  
  patient = "patient_id",  
  region = "region_type",  
  assay_name = "counts",  
  transform = "asinh",  
  cofactor = 5,  
  adaptive = FALSE,  
  output_assay = "rpbc"  
)
```

```
# Adaptive RPBC
```

```
example_adaptive <- rpbc(  
  object = haley_gbm_example,  
  patient = "patient_id",  
  region = "region_type",  
  assay_name = "counts",  
  transform = "asinh",  
  cofactor = 5,  
  adaptive = TRUE,  
  output_assay = "rpbc_adaptive"  
)
```

```
# Covariate-adjusted RPBC
```

```
example_covariate <- rpbc(  
  object = haley_gbm_example,  
  patient = "patient_id",  
  region = "region_type",  
  assay_name = "counts",  
  transform = "asinh",  
  cofactor = 5,  
  covariates = c("age", "sex"),  
  adaptive = FALSE,  
  output_assay = "rpbc_covariate"  
)
```

Index

* **datasets**

haley_gbm_example, [2](#)

haley_gbm_example, [2](#)

rpbc, [3](#)

rpbc(), [3](#)