

# Package ‘batchCorr’

November 5, 2025

**Type** Package

**Title** Within And Between Batch Correction Of LC-MS Metabolomics Data

**Version** 1.0.0

**Date** 2020-05-29

**Description** From the perspective of metabolites as the continuation of the central dogma of biology, metabolomics provides the closest link to many phenotypes of interest. This makes metabolomics research promising in teasing apart the complexities of living systems. However, due to experimental reasons, the data includes non-biological variation which limits quality and reproducibility, especially if the data is obtained from several batches. The batchCorr package reduces unwanted variation by way of between-batch alignment, within-batch drift correction and between-batch normalization using batch-specific quality control samples and long-term reference QC samples. Please see the associated article for more thorough descriptions of algorithms.

**License** GPL-2

**biocViews** BiomedicalInformatics, Metabolomics, MassSpectrometry, BatchEffect, Normalization, QualityControl

**Depends** R (>= 4.4.0), SummarizedExperiment

**Imports** reshape, mclust, BiocParallel, methods

**Suggests** BiocStyle, knitr, testthat

**LazyData** FALSE

**RoxygenNote** 7.3.2

**Encoding** UTF-8

**URL** <https://github.com/MetaboComp/batchCorr>

**BugReports** <https://github.com/MetaboComp/batchCorr/issues>

**VignetteBuilder** knitr

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

**git\_branch** RELEASE\_3\_22

**git\_last\_commit** 25f7e5b

**git\_last\_commit\_date** 2025-10-29

**Repository** Bioconductor 3.22

**Date/Publication** 2025-11-05

**Author** Anton Ribbenstedt [cre] (ORCID:

<<https://orcid.org/0000-0002-9985-5644>>),

Carl Brunius [aut] (ORCID: <<https://orcid.org/0000-0003-3957-870X>>),

Vilhelm Suksi [aut]

**Maintainer** Anton Ribbenstedt <parasitetwin@gmail.com>

## Contents

|                            |    |
|----------------------------|----|
| alignBatches . . . . .     | 2  |
| correctDrift . . . . .     | 4  |
| getBatch . . . . .         | 6  |
| mergeBatches . . . . .     | 7  |
| normalizeBatches . . . . . | 8  |
| peakInfo . . . . .         | 10 |
| ThreeBatchData . . . . .   | 11 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>12</b> |
|--------------|-----------|

---

|              |                                            |
|--------------|--------------------------------------------|
| alignBatches | <i>Between-batch alignment of features</i> |
|--------------|--------------------------------------------|

---

## Description

Multi-batch alignment of features artificially split between batches.

## Usage

```
alignBatches(PeakTabNoFill, PeakTabFilled, ...)
```

```
## S4 method for signature 'ANY,ANY'
```

```
alignBatches(
  peakInfo,
  PeakTabNoFill,
  PeakTabFilled,
  batches,
  sampleGroups,
  selectGroup = "QC",
  NAhard = 0.8,
  mzdifff = 0.002,
  rtdiff = 15,
  report = TRUE,
  reportPath = NULL
)
```

```
## S4 method for signature 'SummarizedExperiment,SummarizedExperiment'
```

```
alignBatches(
  PeakTabNoFill,
  PeakTabFilled,
  batches,
  sampleGroups,
  assay.type1 = NULL,
```

```

    assay.type2 = NULL,
    name = NULL,
    mz_col = "mz",
    rt_col = "rt",
    ...
)

```

## Arguments

|               |                                                                                      |
|---------------|--------------------------------------------------------------------------------------|
| PeakTabNoFill | SummarizedExperiment object or matrix before gap-filling                             |
| PeakTabFilled | SummarizedExperiment object or matrix without missing values                         |
| ...           | parameters with same behavior across methods                                         |
| peakInfo      | matrix with mz and rt in columns 1:2 (see e.g. <code>?peakInfo</code> )              |
| batches       | character scalar or factor, batch labels                                             |
| sampleGroups  | character scalar or character, group labels                                          |
| selectGroup   | character scalar, identifier of QC samples                                           |
| NAhard        | numeric scalar, proportion of NAs within batch for feature to be considered missing  |
| mzdiff        | numeric scalar, tolerance for difference in m/z                                      |
| rtdiff        | numeric scalar, tolerance for difference in retention time                           |
| report        | boolean, whether to export diagnostic plots into your work directory (default: TRUE) |
| reportPath    | character scalar, directory path for report                                          |
| assay.type1   | character scalar, assay of PeakTabNoFill to be used in case of multiple assays       |
| assay.type2   | character scalar, assay of PeakTabFilled to be used in case of multiple assays       |
| name          | character scalar, name of the resultant assay in case of multiple assays             |
| mz_col        | character scalar, name of column containing mass information                         |
| rt_col        | character scalar, name of column for retention time                                  |

## Details

A basic method with matrices (samples as rows) and SummarizedExperiment is supported. For grouping variables such as batches, the basic method expects vectors, while the SummarizedExperiment method expects the names of the respective columns in PeakTabFilled. The basic method returns a list with the aligned peak table and information about the process, whereas the SummarizedExperiment method assigns the aligned peak table to the object supplied to PeakTabFilled.

## Value

a SummarizedExperiment object as PeakTabFill with the aligned matrix or a list of six:

- PTalign: the aligned peak table
- boolAveragedAlign: boolean vector of final features where alignment has been made using feature averaging
- PTfill: peaktable without missing values (indata)
- boolKeep: boolean vector of features kept after alignment
- boolAveragedFill boolean vector of original features where alignment has been made using feature averaging
- aI: alignIndex object (indata)

## Examples

```
data("ThreeBatchData")
# Basic method
## Extract peakinfo (i.e. m/z and rt of features).
peakIn <- peakInfo(PT = PTnofill, sep = "@", start = 3)
## Perform multi-batch alignment
alignBat <- alignBatches(
  peakInfo = peakIn, PeakTabNoFill = PTnofill,
  PeakTabFilled = PTfill, batches = meta$batch,
  sampleGroups = meta$grp, selectGroup = "QC",
  report = FALSE
)
## Extract new peak table
PT <- alignBat$PTalign

# SummarizedExperiment
## Construct SummarizedExperiment
peaks <- SimpleList(t(PTnofill), t(PTfill))
sampleData <- meta
featureData <- peakInfo(PT = PTnofill, sep = "@", start = 3)
rownames(featureData) <- rownames(peaks[[1]])
se <- SummarizedExperiment(assays = peaks, colData = sampleData,
  rowData = featureData)
names(assays(se)) <- c("nofill", "fill")

se <- alignBatches(PeakTabNoFill = se, PeakTabFilled = se,
  batches = "batch", sampleGroups = "grp", report = FALSE,
  assay.type1 = "nofill", assay.type2 = "fill",
  name = "aligned", rt_col = "rt", mz_col = "mz")
```

---

correctDrift

*Within-batch signal intensity drift correction*

---

## Description

Correct drift with cluster-based drift correction.

## Usage

```
correctDrift(peakTable, ...)

## S4 method for signature 'ANY'
correctDrift(
  peakTable,
  injections,
  sampleGroups,
  QCID = "QC",
  RefID = "none",
  modelNames = c("VVV", "VVE", "VEV", "VEE", "VEI", "VVI", "VII"),
  G = seq(5, 35, by = 10),
  smoothFunc = "spline",
  spar = 0.2,
  CVlimit = 0.3,
```

```

    report = TRUE,
    reportPath = NULL
)

## S4 method for signature 'SummarizedExperiment'
correctDrift(peakTable, injections, sampleGroups, assay.type, name, ...)

```

### Arguments

|                |                                                                                                        |
|----------------|--------------------------------------------------------------------------------------------------------|
| peakTable      | SummarizedExperiment object or matrix                                                                  |
| ...            | parameters with same behavior across methods                                                           |
| injections     | character scalar or numeric, injection order                                                           |
| sampleGroups   | character scalar or character, group labels                                                            |
| QCID           | character scalar, identifier of QC samples                                                             |
| RefID          | character scalar, identifier of external reference samples for unbiased assessment of drift correction |
| modelNameNames | character, Which mclust geometries to test                                                             |
| G              | integer, numbers of clusters to test                                                                   |
| smoothFunc     | character scalar, choice of regression function: spline or loess (default: spline)                     |
| spar           | numeric, smoothing parameter value (defaults to 0.2)                                                   |
| CVlimit        | coefficient of variance threshold for filtering (default = 0.3)                                        |
| report         | boolean, whether to print pdf reports of drift models                                                  |
| reportPath     | character scalar, directory path for report                                                            |
| assay.type     | character scalar, assay of to be used in case of multiple assays                                       |
| name           | character scalar, name of the resultant assay in case of multiple assays                               |

### Details

A basic method for matrix (samples as rows) and SummarizedExperiment is supported. For grouping variables such as sampleGroups, the basic method expects vectors, while the SummarizedExperiment method expects the names of the respective columns. The basic method returns a list with the corrected peak table and information about the process, whereas the SummarizedExperiment method assigns the corrected peak table to the object supplied.

### Value

A SummarizedExperiment object with the corrected and filtered matrix or a list, including the corrected and filtered matrix as well as processing information:

- actionInfo: see what happened to each cluster
- TestFeatsCorr: to extract drift-corrected data
- TestFeatsFinal: to extract drift-corrected data which pass the criterion  $QC\ CV < CVlimit$

## Examples

```
data("ThreeBatchData")
set.seed(2024)
# Get batches
batchB <- getBatch(
  peakTable = PTfill, meta = meta,
  batch = meta$batch, select = "B"
)
batchF <- getBatch(
  peakTable = PTfill, meta = meta,
  batch = meta$batch, select = "F"
)
# Drift correction using QCs
BCorr <- correctDrift(
  peakTable = batchB$peakTable,
  injections = batchB$meta$inj,
  sampleGroups = batchB$meta$grp, QCID = "QC",
  G = seq(5, 35, by = 3), modelNames = c("VVE", "VEE"),
  report = FALSE
)
# Using SummarizedExperiment, more unbiased drift correction using QCs &
# external reference samples
## Construct SummarizedExperiment
peaks <- SimpleList(t(PTnofill), t(PTfill))
sampleData <- meta
featureData <- peakInfo(PT = PTnofill, sep = "@", start = 3)
rownames(featureData) <- rownames(peaks[[1]])
se <- SummarizedExperiment(assays = peaks, colData = sampleData,
  rowData = featureData)
names(assays(se)) <- c("nofill", "fill")

## Correct drift for single batch
se <- se[, colData(se)$batch == "F"]
se <- correctDrift(se,
  injections = "inj", sampleGroups = "grp", RefID = "Ref",
  G = seq(5, 35, by = 3), modelNames = c("VVE", "VEE"),
  report = FALSE, assay.type = "fill", name = "corrected")
```

---

getBatch

---

*Extract specific batch from peaktable and metadata*


---

## Description

Extract specific batch from peaktable and metadata

## Usage

```
getBatch(peakTable, meta, batch, select)
```

## Arguments

peakTable      Multi-batch peak table

|        |                                                                                               |
|--------|-----------------------------------------------------------------------------------------------|
| meta   | Multi-batch metadata (including e.g. batch, injection sequence, sample tye, sample name, ...) |
| batch  | vector (length = nSamples) containing batch information (e.g. A, B, C)                        |
| select | which batch to extract                                                                        |

### Value

list object

\$peakTable: Single batch peak table

\$meta: Single batch metadata

### Examples

```
data("ThreeBatchData")
# Get batches for drift correction
batchB <- getBatch(
  peakTable = PTfill, meta = meta,
  batch = meta$batch, select = "B"
)
batchF <- getBatch(
  peakTable = PTfill, meta = meta,
  batch = meta$batch, select = "F"
)
```

---

mergeBatches

*Merge batches after drift correction*

---

### Description

The output of the within-batch drift correction is a correction object. This function merges peak tables from several batches by extracting information from the correction objects. The user must specify a minimum proportion of qualified batches per feature, i.e. such batches where the QC CV is < the specified limit. There is thus a risk that features with poor quality (in certain batches) are present, but the features are present in high quality in sufficient proportion of batches to anyway warrant inclusion.

### Usage

```
mergeBatches(batchList, qualRatio = 0.5)
```

### Arguments

|           |                                                       |
|-----------|-------------------------------------------------------|
| batchList | A list of correction objects (after drift correction) |
| qualRatio | numeric scalar, features with QC CV < the             |

**Value**

A list object containing

‘peakTableOrg’ A merged peak table of original data

‘peakTableCorr’ A merged peak table of drift-corrected data

‘batch’ Batch identifier (per sample)

‘injection’ Injection number (per sample)

**Examples**

```
data("ThreeBatchData")
set.seed(2024)
# Get batches
batchB <- getBatch(
  peakTable = PTfill, meta = meta,
  batch = meta$batch, select = "B"
)
batchF <- getBatch(
  peakTable = PTfill, meta = meta,
  batch = meta$batch, select = "F"
)
# Drift correction using QCs
BCorr <- correctDrift(
  peakTable = batchB$peakTable,
  injections = batchB$meta$inj,
  sampleGroups = batchB$meta$grp, QCID = "QC",
  G = seq(5, 35, by = 3), modelNames = c("VVE", "VEE"),
  report = FALSE
)
# Using SummarizedExperiment, more unbiased drift correction using QCs &
# external reference samples
## Construct SummarizedExperiment
peaks <- SimpleList(t(PTnofill), t(PTfill))
sampleData <- meta
featureData <- peakInfo(PT = PTnofill, sep = "@", start = 3)
rownames(featureData) <- rownames(peaks[[1]])
se <- SummarizedExperiment(assays = peaks, colData = sampleData,
  rowData = featureData)
names(assays(se)) <- c("nofill", "fill")

## Correct drift for single batch
se <- se[, colData(se)$batch == "F"]
se <- correctDrift(se,
  injections = "inj", sampleGroups = "grp", RefID = "Ref",
  G = seq(5, 35, by = 3), modelNames = c("VVE", "VEE"),
  report = FALSE, assay.type = "fill", name = "corrected")
```

---

normalizeBatches

*Between-batch normalisation*

---

**Description**

Batches are feature-wise normalised by Ref samples if passing heuristic criteria (CV and fold change). Otherwise normalized by population median.

**Usage**

```
normalizeBatches(peakTableCorr, ...)

## S4 method for signature 'ANY'
normalizeBatches(
  peakTableCorr,
  batches,
  sampleGroup,
  refGroup = "QC",
  population = "all",
  CVlimit = 0.3,
  FCLimit = 5,
  medianZero = c("mean", "min")
)

## S4 method for signature 'SummarizedExperiment'
normalizeBatches(peakTableCorr, batches, sampleGroup, assay.type, name, ...)
```

**Arguments**

|               |                                                                                                                                                                                        |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| peakTableCorr | SummarizedExperiment object or matrix                                                                                                                                                  |
| ...           | parameters with same behavior across methods                                                                                                                                           |
| batches       | character scalar or factor, batch labels                                                                                                                                               |
| sampleGroup   | character scalar or character, group labels                                                                                                                                            |
| refGroup      | character scalar, identifier of reference samples (default = "QC")                                                                                                                     |
| population    | character scalar, identifier of population samples in sampleGroups (default: "all")                                                                                                    |
| CVlimit       | numeric scalar, coefficient of variance threshold for normalizing each batch by reference samples (default: 0.3)                                                                       |
| FCLimit       | numeric scalar, fold-change between average intensity in batches threshold for normalizing each batch by reference samples (default: 5)                                                |
| medianZero    | character scalar, strategy for substituting median value for population normalization when median is zero (-> Inf). Either 'mean' or 'min' (default: "min, i.e. lowest non-zero value) |
| assay.type    | character scalar, assay to be used in case of multiple assays                                                                                                                          |
| name          | character scalar, name of the resultant assay in case of multiple assays                                                                                                               |

**Details**

A basic method for matrix (samples as rows) and SummarizedExperiment is supported. For grouping variables such as sampleGroup, the basic method expects vectors, while the SummarizedExperiment method expects the names of the respective columns. The basic method returns a list with the normalized peak table and information about the process, whereas the SummarizedExperiment method assigns the normalized peak table to the object supplied.

**Value**

An list object containing: A SummarizedExperiment object with the normalized peak table or a list, including the normalized matrix and processing information:

- peakTable: normalised peak table

- `refCorrected`: boolean matrix with information on which batches were normalised by reference samples; else normalised by population median

### Examples

```
# Note that the example data does not include any biological samples
data("ThreeBatchData")
normData <- normalizeBatches(
  peakTableCorr = PTfill, batches = meta$batch,
  sampleGroup = meta$grp, refGroup = "Ref",
  population = "all"
)

# With SummarizedExperiment
peaks <- SimpleList(t(PTnofill), t(PTfill))
sampleData <- meta
featureData <- peakInfo(PT = PTnofill, sep = "@", start = 3)
rownames(featureData) <- rownames(peaks[[1]])
se <- SummarizedExperiment(assays = peaks, colData = sampleData,
  rowData = featureData)
names(assays(se)) <- c("nofill", "fill")

se <- normalizeBatches(
  peakTableCorr = se, batches = "batch", sampleGroup = "grp",
  refGroup = "Ref", population = "all", assay.type = "fill",
  name = "normalized"
)
```

---

peakInfo

*Extract m/z and rt from peak table*

---

### Description

Extract features from peak table and report their m/z and rt values

### Usage

```
peakInfo(PT, sep = "_", timepos = 2, start = 1)
```

### Arguments

|         |                                                                                      |
|---------|--------------------------------------------------------------------------------------|
| PT      | a peak table with variables as columns                                               |
| sep     | character separating mz from rt, e.g. "_"                                            |
| timepos | Which position carries info about rt (1 for before separator; 2 for after separator) |
| start   | character from which to start the read of peakInfo (from PT colnames)                |

### Value

a matrix with m/z and rt of features as columns

## Examples

```
data("ThreeBatchData")
# Extract peakinfo (i.e. m/z and rt of features). These column names have 2
# leading characters describing LC-MS mode -> start at 3
peakIn <- peakInfo(PT = PTnofill, sep = "@", start = 3)
```

---

|                |                       |
|----------------|-----------------------|
| ThreeBatchData | <i>ThreeBatchData</i> |
|----------------|-----------------------|

---

## Description

The untargeted LC-MS metabolomics data was collected as per the associated publication. In brief, the data was preprocessed using the **xcms** package and biological samples were removed.

## Format

A list with three objects encompassing three batches, 48 QC samples, 42 long-term reference samples and 5000 features

**PTnofill** matrix, peak table with missing values

**PTfill** matrix, peak table without missing values

**meta** data.frame, sample metadata including batch, group and injection order

## Author(s)

Carl Brunius

## References

Carl Brunius, Lin Shi, Rikard Landberg Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. *Metabolomics*, 12:173. <https://doi.org/10.3390/metabo10040135>

# Index

`alignBatches`, [2](#)  
`alignBatches`, ANY, ANY-method  
    (`alignBatches`), [2](#)  
`alignBatches`, `SummarizedExperiment`, `SummarizedExperiment`-method  
    (`alignBatches`), [2](#)  
  
`correctDrift`, [4](#)  
`correctDrift`, ANY-method (`correctDrift`),  
    [4](#)  
`correctDrift`, `SummarizedExperiment`-method  
    (`correctDrift`), [4](#)  
  
`getBatch`, [6](#)  
  
`mergeBatches`, [7](#)  
`meta` (`ThreeBatchData`), [11](#)  
  
`normalizeBatches`, [8](#)  
`normalizeBatches`, ANY-method  
    (`normalizeBatches`), [8](#)  
`normalizeBatches`, `SummarizedExperiment`-method  
    (`normalizeBatches`), [8](#)  
  
`peakInfo`, [10](#)  
`PTfill` (`ThreeBatchData`), [11](#)  
`PTnofill` (`ThreeBatchData`), [11](#)  
  
`ThreeBatchData`, [11](#)