

# Package ‘phosphonormalizer’

October 18, 2022

**Type** Package

**Title** Compensates for the bias introduced by median normalization in

**Version** 1.20.0

**Date** 2016-11-11

**Description** It uses the overlap between enriched and non-enriched datasets to compensate for the bias introduced in global phosphorylation after applying median normalization.

**biocViews** Software, StatisticalMethod, WorkflowStep, Normalization, Proteomics

**License** GPL (>= 2)

**Imports** plyr, stats, graphics, matrixStats, methods

**Suggests** knitr, rmarkdown, testthat

**Enhances** MSnbase

**Depends** R (>= 4.0)

**VignetteBuilder** knitr

**NeedsCompilation** no

**LazyData** true

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

**git\_branch** RELEASE\_3\_15

**git\_last\_commit** bb4bdfd

**git\_last\_commit\_date** 2022-04-26

**Date/Publication** 2022-10-18

**Author** Sohrab Saraei [aut, cre],  
Tomi Suomi [ctb],  
Otto Kauko [ctb],  
Laura Elo [ths]

**Maintainer** Sohrab Saraei <[sohrab.saraei@blueprintgenetics.com](mailto:sohrab.saraei@blueprintgenetics.com)>

## R topics documented:

|                            |   |
|----------------------------|---|
| enriched.rd . . . . .      | 2 |
| non.enriched.rd . . . . .  | 3 |
| normalizePhospho . . . . . | 4 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>6</b> |
|--------------|----------|

---

|             |                         |
|-------------|-------------------------|
| enriched.rd | <i>Enriched dataset</i> |
|-------------|-------------------------|

---

### Description

A dataset containing sequences, modifications and abundances of about 4000 peptides over 5 samples with 3 technical replicates each.

### Usage

```
enriched.rd
```

### Format

A data frame with 4099 rows and 17 variables, all samples are median normalized:

**Sequence** The sequence of the peptide

**Modification** The modification and its location

**gcNorm.ctr12.1** Sample: Control 2 Technical Replicate: 1

**gcNorm.ctr12.2** Sample: Control 2 Technical Replicate: 2

**gcNorm.ctr12.3** Sample: Control 2 Technical Replicate: 3

**gcNorm.ctr11.1** Sample: Control 1 Technical Replicate: 1

**gcNorm.ctr11.2** Sample: Control 1 Technical Replicate: 2

**gcNorm.ctr11.3** Sample: Control 1 Technical Replicate: 3

**gcNorm.CIP2A.1** Sample: CIP2A Technical Replicate: 1

**gcNorm.CIP2A.2** Sample: CIP2A Technical Replicate: 2

**gcNorm.CIP2A.3** Sample: CIP2A Technical Replicate: 3

**gcNorm.RAS.1** Sample: RAS Technical Replicate: 1

**gcNorm.RAS.2** Sample: RAS Technical Replicate: 2

**gcNorm.RAS.3** Sample: RAS Technical Replicate: 3

**gcNorm.OA.1** Sample: OA Technical Replicate: 1

**gcNorm.OA.2** Sample: OA Technical Replicate: 2

**gcNorm.OA.3** Sample: OA Technical Replicate: 3 ...

### Value

Example Non-enriched dataset

**Source**

<http://www.nature.com/articles/srep13099>

---

non.enriched.rd      *Non-enriched dataset*

---

**Description**

A dataset containing sequences, modifications and abundances of about 17000 peptides measured over 5 samples with 3 technical replicates each.

**Usage**

non.enriched.rd

**Format**

A data frame with 16982 rows and 17 variables, all samples are median normalized:

**Sequence** The sequence of the peptide

**Modification** The modification and its location

**gcNorm.ctr12.1** Sample: Control 2 Technical Replicate: 1

**gcNorm.ctr12.2** Sample: Control 2 Technical Replicate: 2

**gcNorm.ctr12.3** Sample: Control 2 Technical Replicate: 3

**gcNorm.ctr11.1** Sample: Control 1 Technical Replicate: 1

**gcNorm.ctr11.2** Sample: Control 1 Technical Replicate: 2

**gcNorm.ctr11.3** Sample: Control 1 Technical Replicate: 3

**gcNorm.CIP2A.1** Sample: CIP2A Technical Replicate: 1

**gcNorm.CIP2A.2** Sample: CIP2A Technical Replicate: 2

**gcNorm.CIP2A.3** Sample: CIP2A Technical Replicate: 3

**gcNorm.RAS.1** Sample: RAS Technical Replicate: 1

**gcNorm.RAS.2** Sample: RAS Technical Replicate: 2

**gcNorm.RAS.3** Sample: RAS Technical Replicate: 3

**gcNorm.OA.1** Sample: OA Technical Replicate: 1

**gcNorm.OA.2** Sample: OA Technical Replicate: 2

**gcNorm.OA.3** Sample: OA Technical Replicate: 3 ...

**Value**

Example Non-enriched dataset

**Source**

<http://www.nature.com/articles/srep13099>

---

|                  |   |
|------------------|---|
| normalizePhospho | <i>Pairwise Normalization of MS-based phosphoproteomic data</i> |
|------------------|---|

---

### Description

This function compensates for the bias introduced in global phosphorylation in the sample after using median normalization.

### Usage

```
normalizePhospho(enriched, non.enriched, phospho = NULL,
  samplesCols, modseqCols, techRep, plot.fc=NULL)
```

### Arguments

|              |   |
|--------------|---|
| enriched     | The enriched data with the type data.frame or MSnSet, which should contain the sequence, modification of the sequence with their phosphorylation site and their abundances across samples                     |
| non.enriched | The non-enriched data with the type data.frame or MSnSet, which should contain the sequence, modification of the sequence with their phosphorylation site and their abundances across samples                 |
| phospho      | a string that shows the term that represents phosphorylation in the modification column of the data. If it is not assigned, "Phospho" will be used as the default value                                       |
| samplesCols  | A data.frame with two columns, with the column names enriched and non.enriched, of type numeric or integer, which must contain the column number of samples that hold the abundances                          |
| modseqCols   | A data.frame with two columns, with the names enriched and non.enriched, of type numeric or integer, which must contain the column number of samples that hold the sequence and modifications of the peptides |
| techRep      | a factor that holds information about columns order and the technical replicates of the samples   |
| plot.fc      | This parameter if set plots the fold change distribution before and after pairwise normalization. controls and samples should be set as named vectors in a list (look at the example)                         |

### Details

It is shown that global median normalization can introduce bias in the fold change of global phosphorylation between samples. It is suggested that by taking the non-enriched data into consideration, this bias could be compensated (Kauko et al. 2015).

### Value

A data.frame with the normalized values and their sequence and modification.

**Author(s)**

Sohrab Saraei, Tomi Suomi, Otto Kauko, Laura L. Elo  
Maintainer: Sohrab Saraei <sohrab.saraei@utu.fi>

**References**

<http://www.nature.com/articles/srep13099>

**See Also**

[MSnbase](#)

**Examples**

```
#Specifying the column numbers of abundances in the original data.frame,  
#from both enriched and non-enriched runs  
samplesCols <- data.frame(enriched=3:17, non.enriched=3:17)  
#Specifying the column numbers of sequence and modification in the original data.frame,  
#from both enriched and non-enriched runs  
modseqCols <- data.frame(enriched = 1:2, non.enriched = 1:2)  
#The samples and their technical replicates  
techRep <- factor(x = c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5))  
#Call the function  
norm <- normalizePhospho(enriched = enriched.rd, non.enriched = non.enriched.rd,  
  samplesCols = samplesCols, modseqCols = modseqCols, techRep = techRep,  
plot.fc = list(control = c(1,2), samples = c(3,4,5)))  
head(norm)
```

# Index

- \* **Mass-spectrometry**
    - normalizePhospho, [4](#)
  - \* **Normalization,**
    - normalizePhospho, [4](#)
  - \* **Phosphoproteomics,**
    - normalizePhospho, [4](#)
  - \* **datasets**
    - enriched.rd, [2](#)
    - non.enriched.rd, [3](#)
- enriched.rd, [2](#)
- non.enriched.rd, [3](#)
- normalizePhospho, [4](#)