

# phosphonormalizer: Phosphoproteomics Normalization

Sohrab Saraei\*, Tomi Suomi, Otto Kauko, Laura L. Elo

\*sohrab.saraei (at) utu.fi

November 11, 2016

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Input data</b>	<b>2</b>
<b>3</b>	<b>Pairwise normalization</b>	<b>2</b>
<b>4</b>	<b>Installation</b>	<b>2</b>
<b>5</b>	<b>References</b>	<b>3</b>

## 1 Introduction

The phosphonormalizer package tries to find phosphopeptides that are present in both enriched and non-enriched datasets with the same sequence and modification. Phosphopeptides considered in the normalization procedure must be quantified in all samples. Additionally, phosphopeptides quantified multiple times are summed together. Moreover, ratios with maximum fold difference more than 1.5x interquartile range will be excluded (Kauko et al. 2015).

After finding the common phosphopeptides, the enriched/non-enriched ratios are calculated and normalized. The median of abundance ratios is used as pairwise normalization factor, which is used to normalize the enriched samples (Kauko et al. 2015).

## 2 Input data

In order to use phosphonormalizer package, we assume that the experiment have been conducted on both enriched and non-enriched samples. These datasets must have the sequence, modification and abundance columns. The sequence and modification columns in the dataframe must be in character format and the abundance columns in numeric. The algorithm expects that the abundances are pre-normalized with median normalization (Kauko et al. 2015). This package also supports MSnSet data type from MSnbase package which is used in data preprocessing step of bioconductor mass spectrometry proteomics workflow (see more: [Bioconductor proteomics workflow](#)).

## 3 Pairwise normalization

The normalization begins by loading the phosphonormalizer package. Here for demonstration, the data used is from "enriched.rd" and "non.enriched.rd" are available with the package (Kauko et al. 2015).

## 4 Installation

To install this package, start R and enter:

```
> ## try http:// if https:// URLs are not supported
> source("https://bioconductor.org/biocLite.R")
> biocLite("phosphonormalizer")

> #Load the library
> library(phosphonormalizer)
> #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
+ )
```

## 5 References

Kauko, O. et al, *Label-free quantitative phosphoproteomics with novel pairwise abundance normalization reveals synergistic RAS and CIP2A signaling* Sci. Rep. 5, 13099; doi: 10.1038/srep13099, 2015.