phosphonormalizer: Phosphoproteomics Normalization

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

*`sohrab.saraei (at) utu.fi`

November 11, 2016

Contents

1   Introduction

Global centering-based normalization is a commonly-used normalization approach in mass spectrometry (MS) -based label-free proteomics. It scales the peptide abundances to have the same median intensities, based on an assumption that the majority of abundances remain the same across the samples. However, especially in phosphoproteomics experiments, this assumption can introduce bias, as the enrichment of phosphopeptides during sample preparation can mask large unidirectional biological changes. Therefore, a novel method called pairwise normalization has been introduced that addresses this possible bias by utilizing phosphopeptides quantified in both enriched and non-enriched samples to calculate factors that mitigate the bias (Kauko et al. 2015). The phosphonormalizer package implements the pairwise normalization (Saraei et al., under review).

The phosphonormalizer package (Saraei et al. under review) normalizes the enriched samples in label-free MS-based phosphoproteomics using phosphopeptides that are present in both enriched and non-enriched data of the same samples. If there are no common phosphopeptides between the enriched and non-enriched data, then the normalization is not possible and an error is generated.

2   Input data

To use the phosphonormalizer package, we assume that the experiments have been conducted on both enriched and non-enriched samples. The input data is assumed to be a data frame whose columns are the sequence, modification and abundances (samples). The sequence and modification columns must be in the character format and the abundance columns in the numeric format. The abundances are assumed to be pre-normalized with median normalization. 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 workow).

3   Pairwise normalization

The normalization begins by loading the phosphonormalizer package, which includes two example datasets for demonstration: "enriched.rd" and "non.enriched.rd".

Phosphopeptides considered in the normalization procedure must be quantified in all samples. Phosphopeptides quantified multiple times are summed together. For each phosphopeptide in the overlap, its abundance in the non-enriched data is divided by its counterpart in the enriched data to calculate peptide abundance ratios. Phosphopeptides with ratios of more than 1.5x interquartile range of overall fold change are excluded to ensure that the method is not sensitive

to outliers. Finally, the median of the abundance ratios is used as the pairwise normalization factor to normalize the enriched samples (Kauko et al. 2015). For convenience, boxplots of fold change distributions before and after pairwise normalization can also be generated by setting the plot parameter.

4 Installation

To install the phosphonormalizer package, start R and enter:

```
> ## try http:// if https:// URLs are not supported
> if (!requireNamespace("BiocManager", quietly=TRUE))
+     install.packages("BiocManager")
> BiocManager::install("phosphonormalizer")
```

5 Example

```
> #Load the library
> library(phosphonormalizer)
> #Specify 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)
> #Specify 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))
> #If the parameter plot.fc is set, then the corresponding plots of sample fold changes are
> #Here, for demonstration, the fold change distributions are shown for samples 3 vs 1
> plot.param <- list(control = c(1), samples = c(3))
> # Call the function to perform the pairwise normalization:
> norm <- normalizePhospho(enriched = enriched.rd, non.enriched = non.enriched.rd,
+               samplesCols = samplesCols, modseqCols = modseqCols, techRep = techRep,
+         plot.fc = plot.param)
```

Peptide log fold changes (sample 3 vs sample 1)
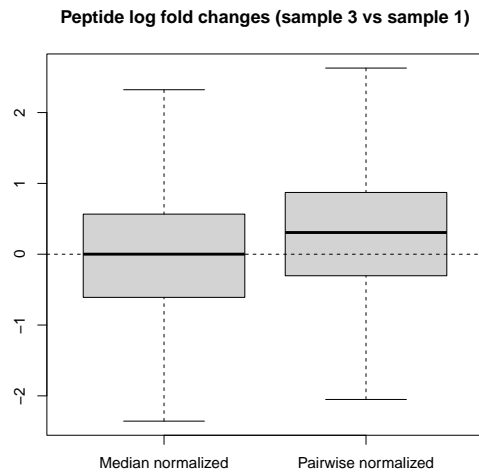
6 References

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.

Saraei, s. et al, *Phosphonormalizer: an R package for normalization of MS-based label-free phosphoproteomics* Bioinformatics, Under review.