Documentation for Normalyzer:

Prerequisites:

R version >3

Packages:

vsn,Rcmdr,PerformanceAnalytics,preprocessCore,limma,MASS,abind,e1071,ape,car,raster

Installation:

Download and install Normalyzer R package from http://quantitativeproteomics.org/normalyzer

Usage:

library(Normalyzer)

normalyzer("filename","projectName")

OR

normalyzer (object,"projectName") #Only objects of type dataframe and matrix are supported with the same format as for the file input.

For instructions on preparing the data file, check http://quantitativeproteomics.org/normalyzer/help.php

Normalyzer can also be used online at http://quantitativeproteomics.org/normalyzer

Evaluation of external data

To evaluate just one dataset normalized outside of normalizer, following function can be used.

analyzeAndPlotForOneObject("filename", dataname, projectname)

File format is similar to Normalyzer format.

Description of methods:

Global: In global methods, normalization is done across all samples without considering any replicate grouping. The basic assumption for any global normalization is that most of the measured variables (genes/peptides/proteins) are not differentially expressed across conditions. This can be confirmed with the DE plots of Log2 transformed data at the end of the Normalyzer report. This hypothesis is generally true for most high-throughput datasets. In case the assumption doesn't stand, local normalization can be optionally done.

Local: Local normalization is only done within the replicate groups without any overall/global normalization. This could be useful in cases where global normalization is not possible and when too few variables are measured.

Log_2

Data is Log_2 transformed.

LOESS

Log2 transformed data is normalized by LOESS method using the function "normalizeCyclicLoess". For the global normalization, each sample is normalized to the average of all samples, whereas in the local method, each sample within the replicate group is normalized to the average of that replicate group. For more information on this method please see the function "normalizeCyclicLoess" in the Limma package.

Robust Linear Regression (RLR)

Log2 transformed data is normalized by robust linear regression ("rlm" function). For the method, each sample is normalized to the median of all samples, while in the local method, each sample in a replicate group is normalized to the median of samples in the replicate group. For more information on this method please see the function rlm in the MASS package.

Variance Stabilization Normalization (VSN)

Log2 transformed data is normalized using the function "justvsn" from the vsn package. Global and local normalization are done as described in the RLR section above.

Total Intensity (TI-G)

Intensity of each variable in a given sample is divided by the sum of intensities of all variables in the sample and then multiplied by the median of 'sum of intensities of all variables in all samples'. The normalized data is then transformed to Log_2.

Median Intensity (MedI-G)

Intensity of each variable in a given sample is divided by the median of intensities of all variables in the sample and then multiplied by the mean of 'median of sum of intensities of all variables in all samples'. The normalized data is then transformed to Log_2.

Average Intensity (AI-G)

Intensity of each variable in a given sample is divided by the mean of sum of intensities of all variables in the sample and then multiplied by the mean of 'mean of sum of intensities of all variables in all samples'. The normalized data is then transformed to Log_2.

Quantile-G

Quantile normalization is done by the function "normalize.quantiles" from the package preprocessCore.

NormFinder (NF-G)

NormFinder normalization is done using the R script obtained from the developers of NormFinder algorithm. The R script was modified for incorporation in the Normalyzer pipeline. Normalization by NormFinder is a global normalization method which can be used even when few variables are measured as long as control variables are included in the assay. In this method, two variables with least local and global variance are selected by NormFinder. The intensity of each variable in a sample is divided by the mean of intensities of the two NormFinder-identified control variables from the corresponding samples and multiplied by the mean of all intensities of the control variables. The data is then log_2 transformed.

Analysis

Total Intensity: Sum of intensities of all variables in a sample.

Missing Values: Total number of 'NA' in a sample

MDS plot: Multidimensional scaling plot with "cmdscale()" function from the stats package.

PCV: Mean of 'intragroup CV of all replicate groups'

PMAD: Mean of 'intragroup Median Absolute Deviation across all replicate groups'

PEV: Mean of 'intragroup Pooled Estimate of Variance across all replicate groups'

Relative PCV, PMAD and PEV compared to log2: Results from PCV, PMAD and PEV from all normalized data relative to those of the log_2 data.

Stable variables plot: 5% of least differentially expressed variables are identified by ANOVA analysis of Log_2 transformed data. Thereafter, global CV of these variables is estimated from different normalized datasets. A plot of global CV of the stable variables from all datasets on the y-axis and PCV-compared to log2 on the x-axis is generated.

CV vs intensity plot: For the first replicate group in each of the normalized dataset, a plot of PCV of each variable vs the average intensity of the variable in the replicate group is plotted.

MA plots: Is plotted with plotMA function of the limma package. The first sample in each dataset is plotted against the average of the replicate group that sample belongs to.

Scatter plots: The first two samples from each dataset are plotted.

QQ plots: QQ plots are plotted for the first sample in each normalized data set.

Boxplots: Boxplots for all samples are plotted and colored according to the replicate grouping.

RLE plots: Relative log expression value plots. Ratio between the expression of the variable and the median expression of this variable across all samples. The samples should be aligned around zero. Any deviation would indicate discrepancies in the data.

Density plots: Done using the density function.

MeanSDplots: Done using the "meanSDplot" function of the vsn package.

Pearsson Corr.: Mean of intragroup pearsson correlation using the "cor" function.

Spearman Corr.: Mean of intragroup Spearman correlation using the "cor" function.

Dendrograms: Generated using the "hclust" function. Data is centered and scaled prior to analysis. Coloring of replicates done using "as.phylo" function of the ape package.

DE plots: ANOVA and Kruskal Wallis tests.