Package 'ProteinPCA'

April 12, 2025

Type Package

Title Principal Component Analysis (PCA) Tool on Protein Expression Data

Version 0.1.0

Author Paul Angelo C. Manlapaz [aut, cre] (<https://orcid.org/0000-0002-1203-2064>)

Maintainer Paul Angelo C. Manlapaz <pacmanlapaz@gmail.com>

Description

Analysis of protein expression data can be done through Principal Component Analysis (PCA), and this R package is designed to streamline the analysis. This package enables users to perform PCA
and it generates biplot and scree plot for advanced graphical visualization. Optionally, it supports grouping/clustering visualization with PCA loadings and confidence ellipses. With this R package, researchers can quickly explore complex protein datasets, interpret variance contributions, and visualize
sample clustering through intuitive biplots. For more details, see Jolliffe (2001) <doi:10.1007/b98835>,
Gabriel (1971) <doi:10.1093/biomet/58.3.453>, Zhang et al. (2024) <doi:10.1038/s41467-024-53239-9>, and
Anandan et al. (2022) <doi:10.1038/s41598-022-07781-5>.

License GPL-3

Encoding UTF-8

Imports stats, ggplot2, gridExtra

Suggests testthat

Config/testthat/edition 3

RoxygenNote 7.3.2

NeedsCompilation no

Repository CRAN

Date/Publication 2025-04-12 08:30:05 UTC

Contents

`pca_analysis` *Principal Component Analysis (PCA) Tool on Protein Expression Data*

Description

This function performs PCA on protein expression data and produces a biplot using ggplot2. Optionally, it supports grouping/clustering visualization with PCA loadings and confidence ellipses.

Arguments

`data`	A numeric matrix or data frame of protein expression data. Rows are samples, columns are features.
`scale`	Logical. Should the data be scaled? Default is TRUE.
`center`	Logical. Should the data be centered? Default is TRUE.
`plot`	Logical. Should a PCA biplot be generated? Default is TRUE.
`groups`	Optional. A factor or character vector specifying group memberships of the samples.

Value

A list containing:

`pca`	The PCA object from `prcomp`.
`explained_variance`	The percentage of variance explained by each principal component.
`plot`	A ggplot2 PCA biplot (if `plot = TRUE`).

Author(s)

Paul Angelo C. Manlapaz

References

Jolliffe, I. (2001). Principal Component Analysis (2nd ed.). Springer. https://doi.org/10.1007/b98835 Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. Biometrika, 58(3), 453–467. https://doi.org/10.1093/biomet/58.3.453 Zhang, Z., Chen, L., Sun, B., Ruan, Z., Pan, P., Zhang, W., Jiang, X., Zheng, S., Cheng, S., Xian, L., Wang, B., Yang, J., Zhang, B., Xu, P., Zhong, Z., Cheng, L., Ni, H., & Hong, Y. (2024). Identifying septic shock subgroups to tailor fluid strategies through multi-omics integration. Nature Communications, 15(1). https://doi.org/10.1038/s41467-024-53239-9 Anandan, A., Nagireddy, R., Sabarinathan, S.,

Bhatta, B. B., Mahender, A., Vinothkumar, M., Parameswaran, C., Panneerselvam, P., Subudhi, H., Meher, J., Bose, L. K., & Ali, J. (2022). Multi-trait association study identifies loci associated with tolerance of low phosphorus in Oryza sativa and its wild relatives. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-07781-5

Examples

```
set.seed(123)
data_matrix <- matrix(rnorm(100 * 20), nrow = 100, ncol = 20)
rownames(data_matrix) <- paste0('Sample_', 1:100)
colnames(data_matrix) <- paste0('Protein_', 1:20)
groups <- sample(c("Group A", "Group B"), 100, replace = TRUE)
result <- pca_analysis(data_matrix, groups = groups)
print(result$explained_variance)
print(result$plot)
```

Index