R and Bioconductor-based Workflow for Downstream Analysis of LFQ Meat Proteomics Data: Enhancing Reproducibility in Meat Proteomics Research

Sudipta Hazra, Claudia Terlouw, Mohammed Gagaoua, Kerry Joe P., Brigitte Picard, Hamill Ruth

HAL Id: hal-04795654

https://hal.inrae.fr/hal-04795654v1

Submitted on 21 Nov 2024

R and Bioconductor-based Workflow for Downstream Analysis of LFQ Meat Proteomics Data: Enhancing Reproducibility in Meat Proteomics Research

Sudipta Hazra[1,2], Claudia Terlouw[3], Mohammed Gagaoua[4], Joe P. Kerry[2], Brigitte Picard[3], and Ruth Hamill[1]

[1] Department of Food Quality and Sensory Science, Teagasc Food Research Centre, Ashtown, D15 DY05 Dublin, Ireland

[2] School of Food and Nutritional Sciences, College of Science, Engineering and Food Science, University College Cork, T12 K8AF Cork, Ireland

[3] Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genès-Champanelle, France

[4]PEGASE, INRAE, Institut Agro, 35590 Saint-Gilles, France

Abstract:

Background:

Proteomics have been extensively used in meat science to understand the biochemistry of meat quality. Identified and quantified proteomics data obtained through processing with analytical software tools, such as Mascot and Progenesis QI, among others, serve as a common starting point for meat scientists to analyze shotgun proteomic data. However, a standardized workflow for the downstream statistical analysis and visualization of these factorial design datasets is lacking.

Methods:

The workflow was developed using several R and Bioconductor packages including Tidyverse, UniprotR, QFeature, limma, EFS, ggplot2, pheatmap, ggvenn, and UpSetR, along with custom functions. The methodology is demonstrated using an experimental shotgun dataset preprocessed via Mascot and Progenesis QI. This dataset was derived from postmortem muscle tissues from a 2×4 factorial study of cattle reared under two slaughter conditions and four feeding regimes.

Results:

This workflow produced a result table with the gene name, UniProt accession, log2 fold change, p-value, and the corresponding adjusted p-value from the LFQ data. Publication-quality Venn diagrams, UpSet plots, volcano plots, and heat maps were generated. Dimensional reduction techniques such as PCA, t-SNE, and UMAP were integrated into the workflow. In addition, eight distinct feature selection techniques based on ensemble learning were applied using the EFS package, and the results were compared with differentially abundant proteins and visualized using a volcano plot. To enhance the readability, we implemented a strategy to present the correlation of large datasets more effectively.

Discussion:

This R and Bioconductor-based workflow presents a consolidated resource of selected tools that facilitates downstream processing, statistical evaluation, and comprehensive visualization of shotgun proteomic data, thereby potentially enabling meat science researchers to address specific biological questions.