promor: An Integrative Approach for Proteomics Data Analysis and Modeling

Chathurani Ranathunge[1], Lubna Pinky[1], Sagar S. Patel[1], Vanessa L. Correll[2], O. John Semmes[2], Robert K. Armstrong[1, 3], C. Donald Combs[1], and Julius O. Nyalwidhe[2]

[1]School of Health Professions, Eastern Virginia Medical School, Norfolk, VA, USA,
[2]Leroy T. Canoles Jr. Cancer Research Center, Eastern Virginia Medical School, Norfolk, VA, USA,
[3]Sentara Center for Simulation and Immersive Learning, Eastern Virginia Medical School, Norfolk, VA, USA

Motivation

- **Label-free quantification (LFQ)** approaches are rapidly gaining importance in proteomics research.
- There is still a great need for specialized tools that can make the downstream statistical analysis of LFQ data **widely accessible** and **reproducible**.
- **promor** is a user-friendly, comprehensive **R package** that streamlines **differential expression analysis** of LFQ data and **building predictive models** with top protein candidates.

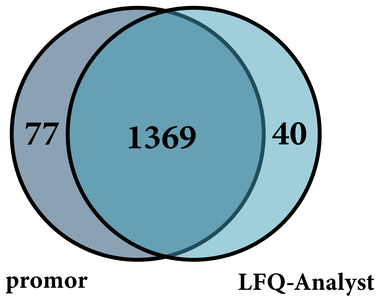
promor includes a suite of functions for:

- **Quality control**
- **Visualization**
- **Missing data imputation**
- **Data normalization**
- **Differential expression analysis**
- **Feature selection**
- **Building models using multiple machine learning algorithms**
- **Model evaluation**

Benchmarking

- **promor** was benchmarked on a publicly available data set (PRIDE ID: PXD000279).
- The results from the differential expression (DE) analysis were compared against those from **LFQ-Analyst[1].**

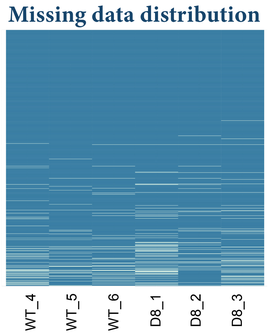
promor and LFQ-Analyst share 92% DE proteins in common



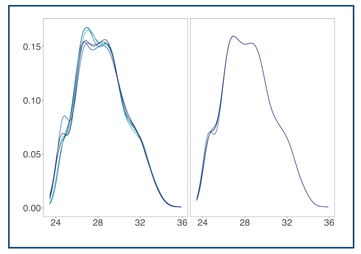
Quality control & Visualization

- **promor** provides tools for:
  - filtering proteins based on different criteria (eg: group level missing data) at multiple levels
  - missing data imputation with a variety of methods
  - data normalization with multiple methods
  - a variety of data visualization options .

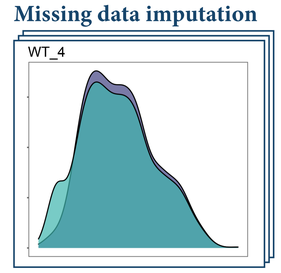
Missing data distribution



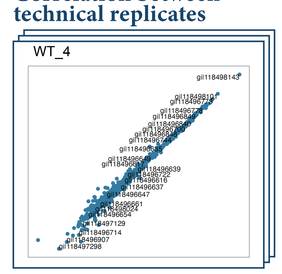
Visualize normalization



Missing data imputation



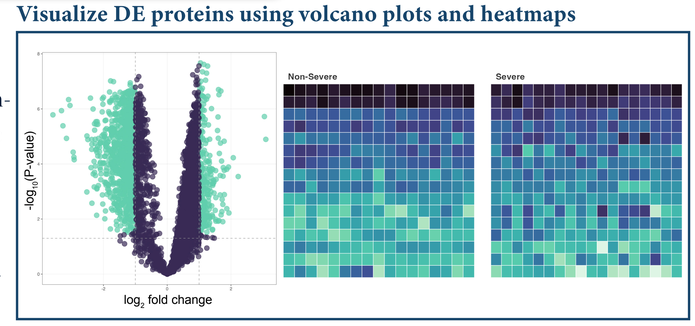
Correlation between technical replicates



Differential expression analysis

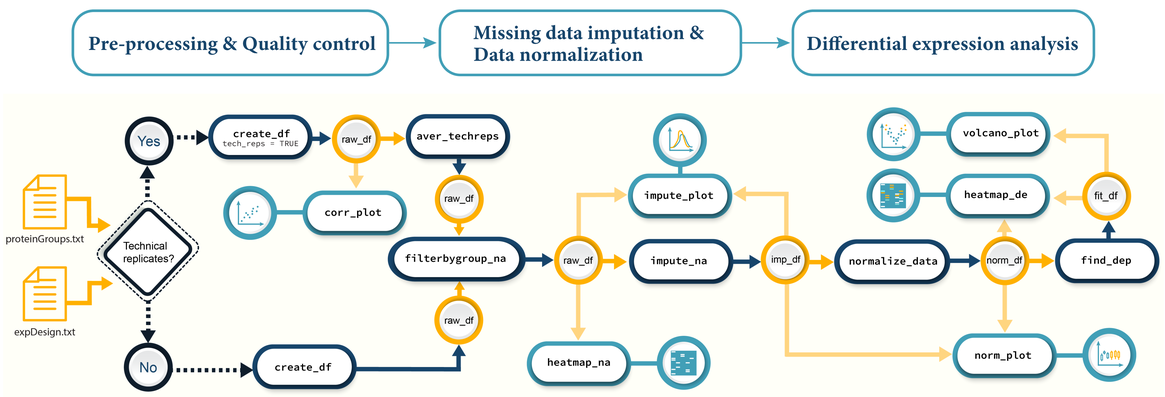
- **promor** uses the empirical Bayes method implemented in the R package, **limma[2]**, for differential expression (DE) analysis.

Visualize DE proteins using volcano plots and heatmaps

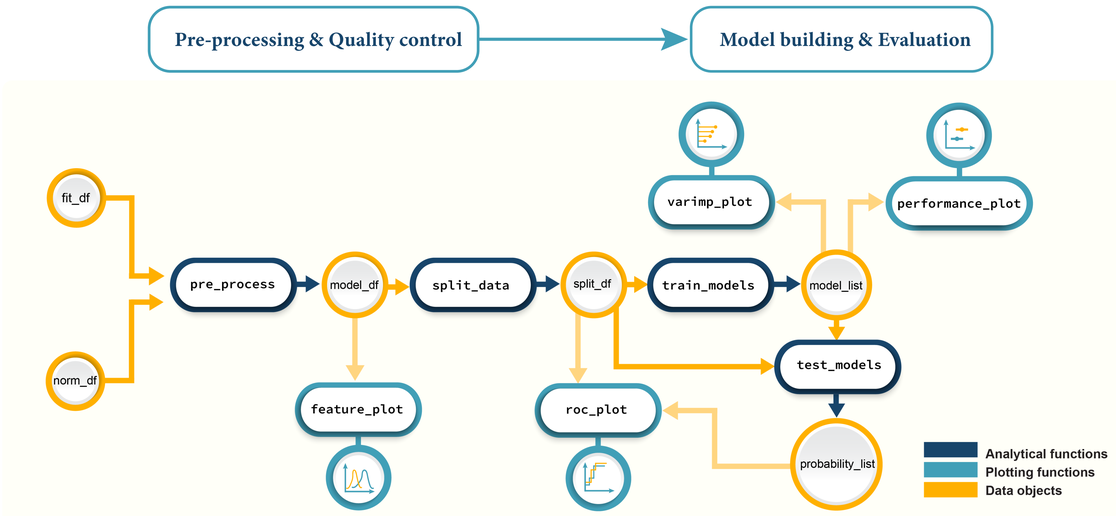


Workflows

promor workflow for LFQ proteomics data analysis



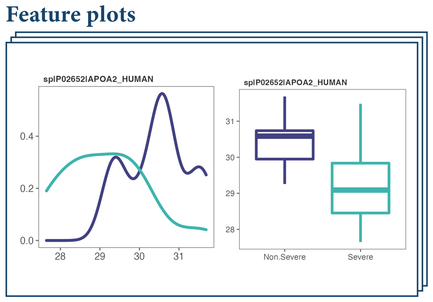
promor workflow for building predictive models with protein candidates



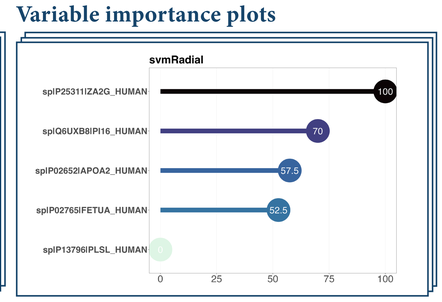
Feature selection

- **Feature plots:** visualize protein variation among groups prior to building models.
- **Variable importance plots**: assess the importance of different proteins in the models built using different machine learning algorithms.

Feature plots



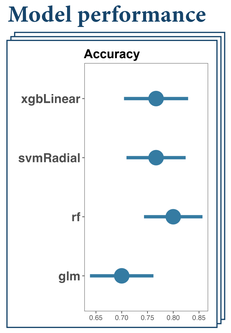
Variable importance plots



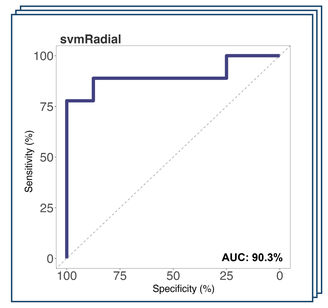
Model building & Evaluation

- **promor** provides a default list of four widely used machine learning algorithms (out of about 200 accessible) to build predictive models.
- **Performance plots:** use multiple metrics to assess the performance of models trained on training data.
- **ROC plots:** build Receiver Operating Characterisitic (ROC) curves to evaluate the diagnostic ability of models built using different machine learning algorithms.

Model performance



ROC curves



Data & code availability

Data and code used for conducting analyses and making plots shown on this poster are available at: **https://github.com/caranathunge/ismb_2022**

References

1. Shah, Anup D., et al. "LFQ-analyst: an easy-to-use interactive web platform to analyze and visualize label-free proteomics data preprocessed with MaxQuant." Journal of proteome research 19.1 (2019): 204-211.
2. Ritchie, Matthew E., et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies." Nucleic acids research 43.7 (2015): e47-e47.

Funding

This research was supported by the **Hampton Roads Biomedical Research Consortium**.

Give promor a try!

@caranathunge

HAMPTON ROADS BIOMEDICAL RESEARCH CONSORTIUM

EVMS Eastern Virginia Medical School