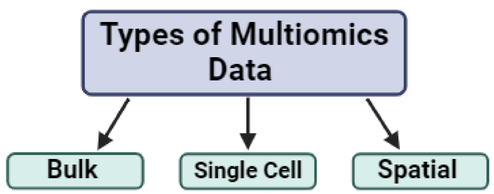
MESSI: A Nextflow pipeline for benchmarking multiomics integration methods for disease classification

Chunqing Tony Liang [1,3] , Amrit Singh [2,3]
[1] UBC: Faculty of Science, Bioinformatics , [2] Faculty of Medicine, Anesthesiology, Pharmacology and Therapeutics, [3]Centre for Heart Lung Innovation

INTRODUCTION

- **Omics** is the comprehensive study of all molecules of a particular type within an organism (**e.g. proteins, metabolites, genes**).
- Multiomics compensates for missing or unreliable information in any single omics data.



- Many integration methods: **which to use?**
- How to **reproduce** method and get same results?
- Create a benchmarking study that is **all encompassing** (many methods and datasets), easy to **setup and replicate**.
- We introduce *Multiomics Experiments with SyStematic Interrogation* (***MESSI***).
- Tool to assist in the development and evaluation of new/existing multiomics integration methods.

METHODS

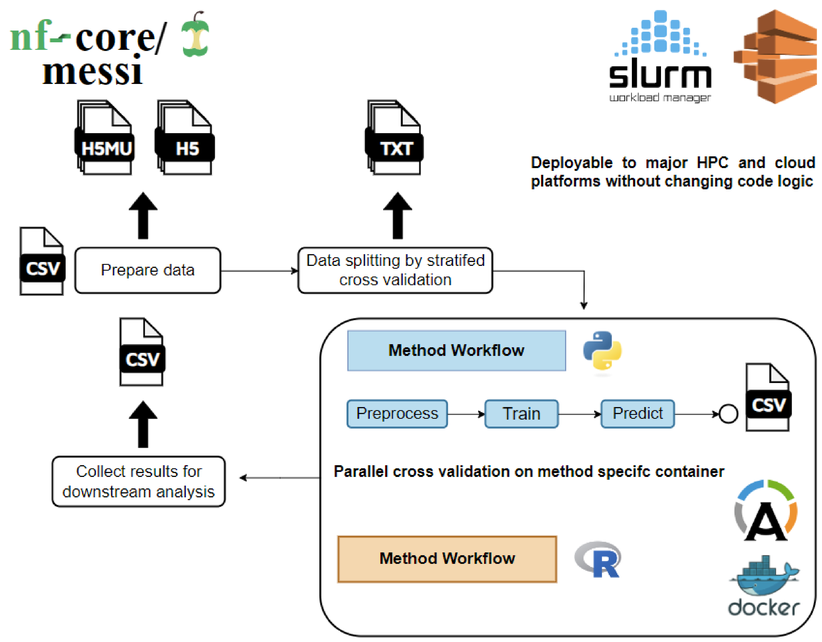


Figure 1. Overview of MESSI. Data are standardized into MuData and MultiAssayExperiemnt, then passed to methods for cross-validation in a parallel and fully isolated containers (each method has its own software environment), results are then collected for downstream meta analysis between methods and datasets combinations. This pipeline could be applied for N methods on 1 dataset or M datasets on 1 method or N methods on M datasets.

RESULTS

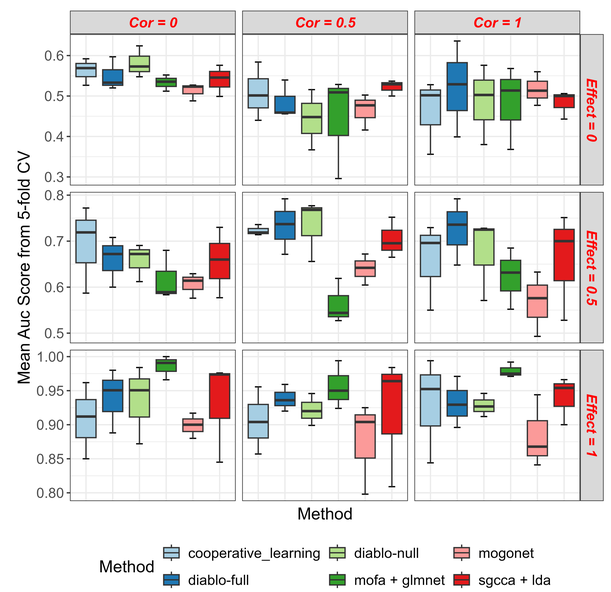


Figure 5. Mean AUC distribution in simulated data stratified by effect and correlation. Score ranges from 0-1. As effect increases, mean auc score shift upwards.

RESULTS

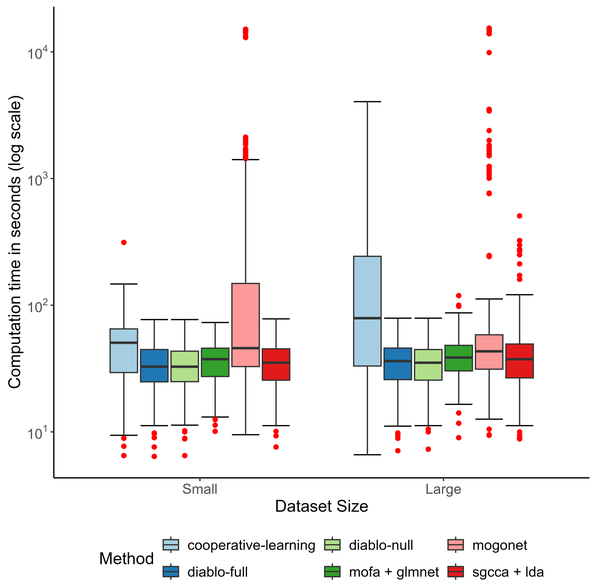


Figure 2. Computational time of multiomics methods of varying data size. Dataset size (Z) is number of subjects x total number of features. Large if Z > median(Z_1,Z_2, ... , Z_M), where M is number of datasets.

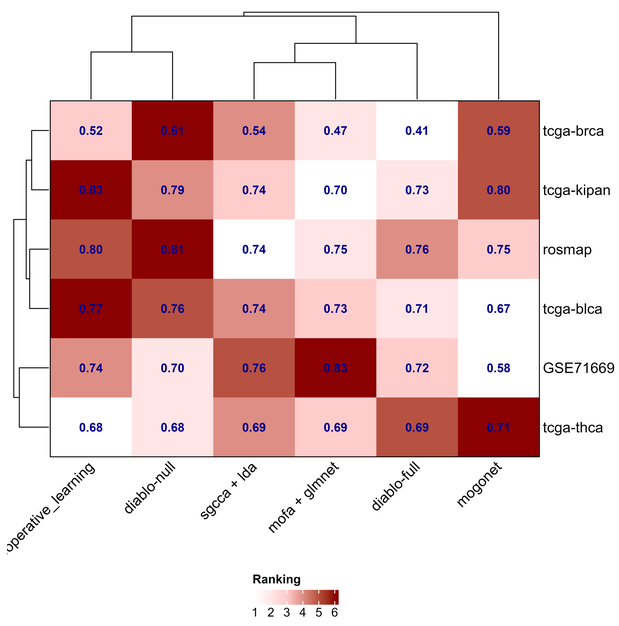


Figure 3. Mean AUC (5-fold CV) ranking in real datasets. The higher is the rank (darker) indicates better performance of multiomics method. Cell number is the mean auc score.

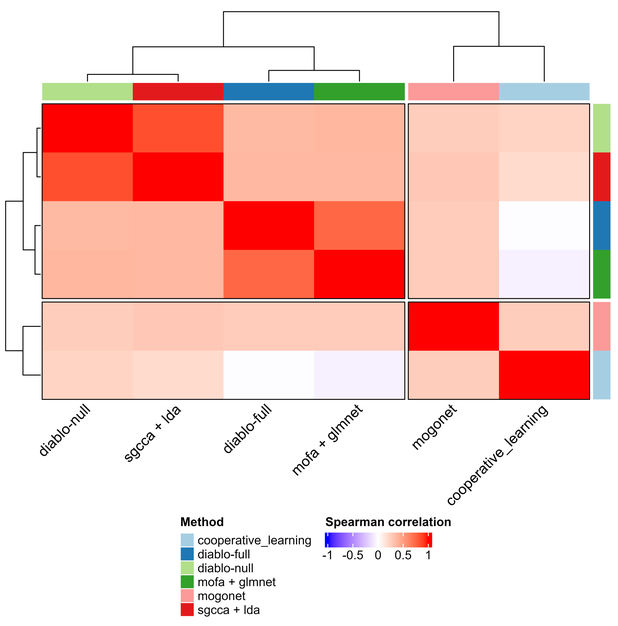


Figure 4. Spearman rank correlation on feature weights across methods. Stronger correlation indicates similar weights for features identified in datasets.

CONCLUSION

- **No method works** universally well at simulated data with noise, often worse than random guessing.
- **GNN like MOGONET, overcomplicated** yet low performance.
- **Cooperative Learning** is the **best** now based on mean auc scores on 5-fold CV, although takes more time to compute.
- Methods overall work better when effect is high, and lesst senstivie to correlation between omics
- Need to generalize for more datasets, at single cell or spatial level
- Also consider implementing more methods (Bayesian, DL)

ACKNOWLEDGEMENTS

THE UNIVERSITY OF BRITISH COLUMBIA
Department of Anesthesiology, Pharmacology & Therapeutics
Faculty of Medicine
Centre for Heart Lung Innovation
UBC and St. Paul's Hospital
CompBio Lab @ HLI
https://cbl-hli.med.ubc.ca/
@lab_comp