# Uncovering Patterns in COVID-19 Pneumonia: Predicting Patient Outcomes with PARAFAC2
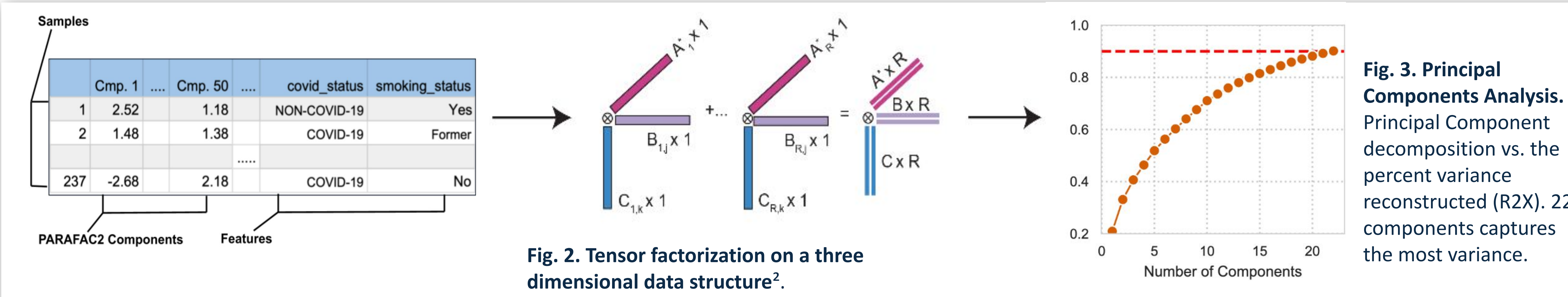
Salina Adhanom[2], Andrew Ramirez[1], Jackson Chin[1], Dr. Aaron Meyer[1]
[1]Department of Bioengineering, University of California, Los Angeles, CA 90024, USA
[2]BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

UCLA

QCBio

Introduction

- Pneumonia is an acute respiratory infection affecting the lungs, classified by the type of pathogen involved (i.e., COVID-19, Methicillin-resistant Staphylococcus aureus, or Cryptococcus)
- Patients with COVID-19 pneumonia spend more time in the ICU compared to those with pneumonia caused by other pathogens.
- The challenge we face is understanding the underlying reasons for the extended ICU stays in COVID-19 patients.
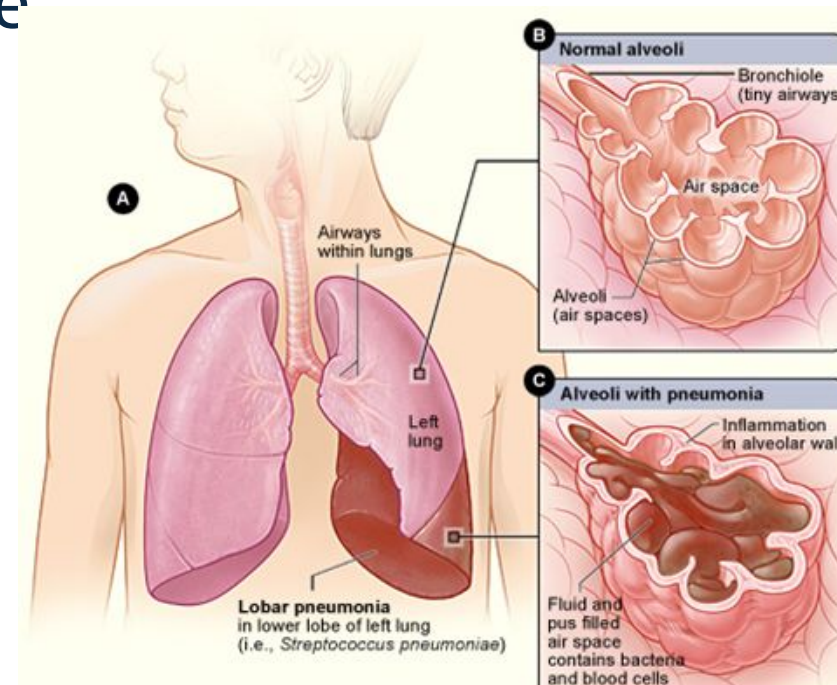
Approach: Principal Components Analysis & Logistic Regression

Fig. 2. Tensor factorization on a three dimensional data structure[2].

Fig. 3. Principal Components Analysis. Principal Component decomposition vs. the percent variance reconstructed (R2X). 22 components captures the most variance.

Methods

Fig. 1. PARAFAC2 (Pf2). PARAFAC2 (Pf2) applied to rank-R mode-3 tensor[2].

- We suggest employing PARAFAC2 (Pf2) as a multidimensional data reduction method for analyzing single-cell (SC) data, non-uniform structures.
- PF2 reduces data complexity, a sum of lower-dimensional tensors, by explaining component associations across each dimension (e.g., cells, conditions, genes), preventing conflation of biological patterns.
- Thus, the execution of this study advances our understanding of COVID-19 pneumonia by utilizing PARAFAC2 to uncover and track patterns of cellular response across different patient conditions, ultimately enhancing our ability to predict patient outcomes.
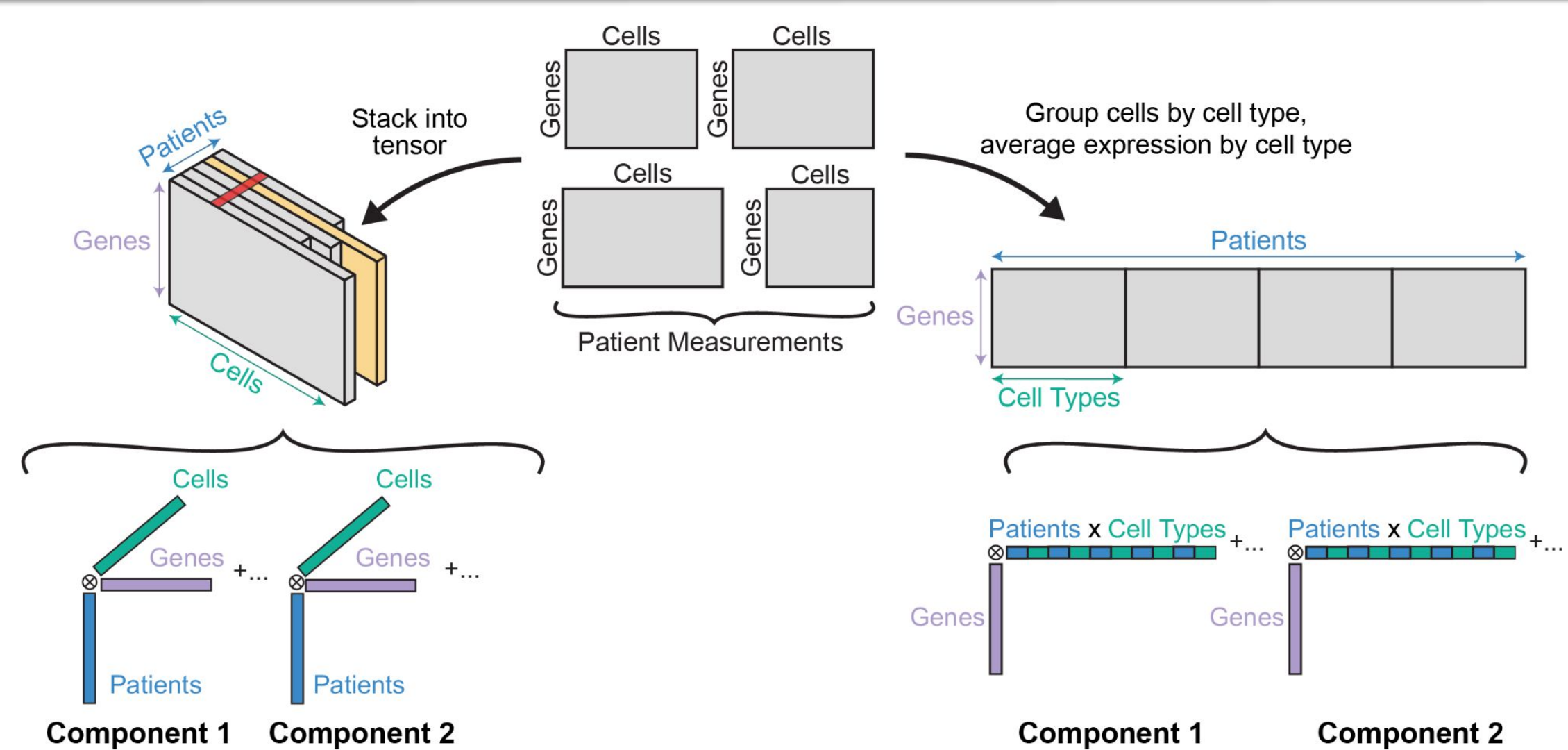
Fig. 4. Loadings and scores for Pf2 Components. Scatterplots labeled by each Pf2 component measuring weights for A) loading coefficients B) scores for each observation.
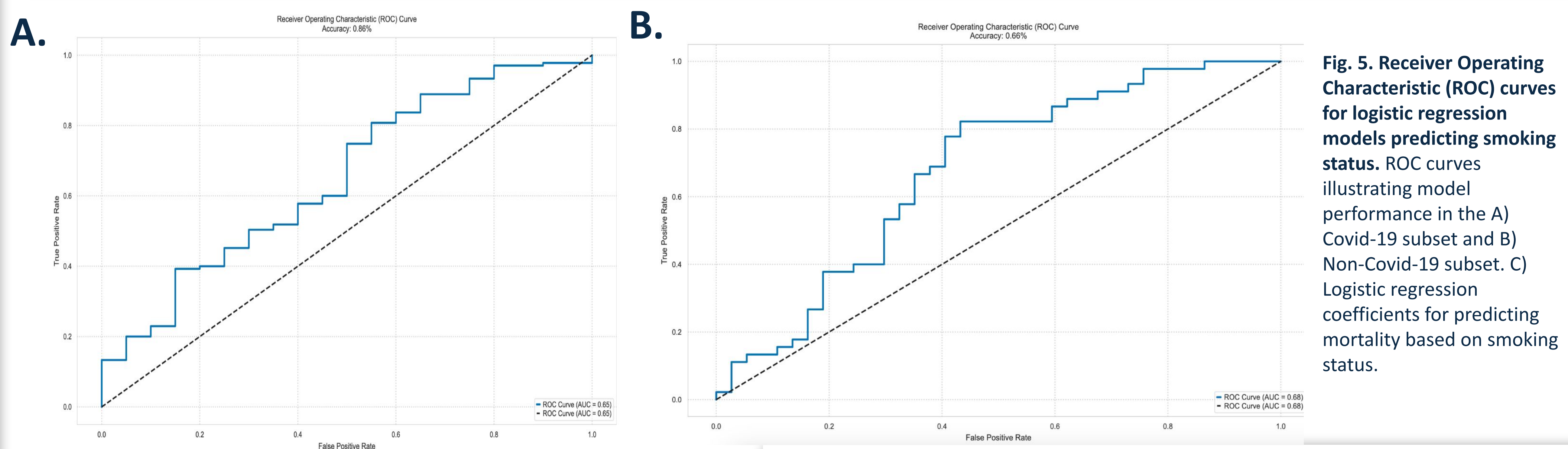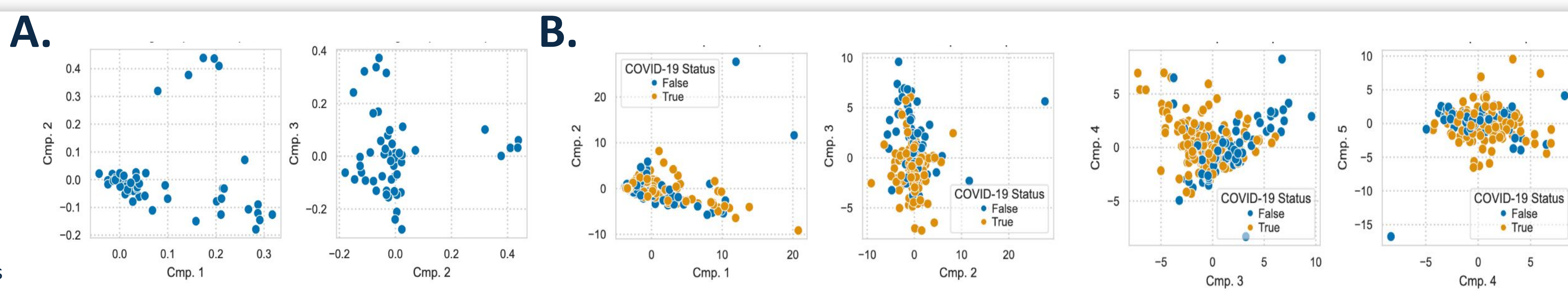

A.

B.

Fig. 5. Receiver Operating Characteristic (ROC) curves for logistic regression models predicting smoking status. ROC curves illustrating model performance in the A) Covid-19 subset and B) Non-Covid-19 subset. C) Logistic regression coefficients for predicting mortality based on smoking status.

Data Collection

- Patients at Northwestern Memorial Hospital from June 2018 and March 2022,
  - 50 - secondary pneumonia
  - 190 - COVID-19 based pneumonia
  - 252 - other pneumonia (bacterial)
  - 93 - respiratory failure unrelated to pneumonia
- All patients admitted to the ICU with some patients collected at multiple, irregular time points.
  - scRNA-seq of bronchoalveolar lavage (BAL) samples, individual cells collected from the lower respiratory tract.
  - Clinical data and outcomes from Electronic Health Record (EHR) and microbiological analysis.

C.
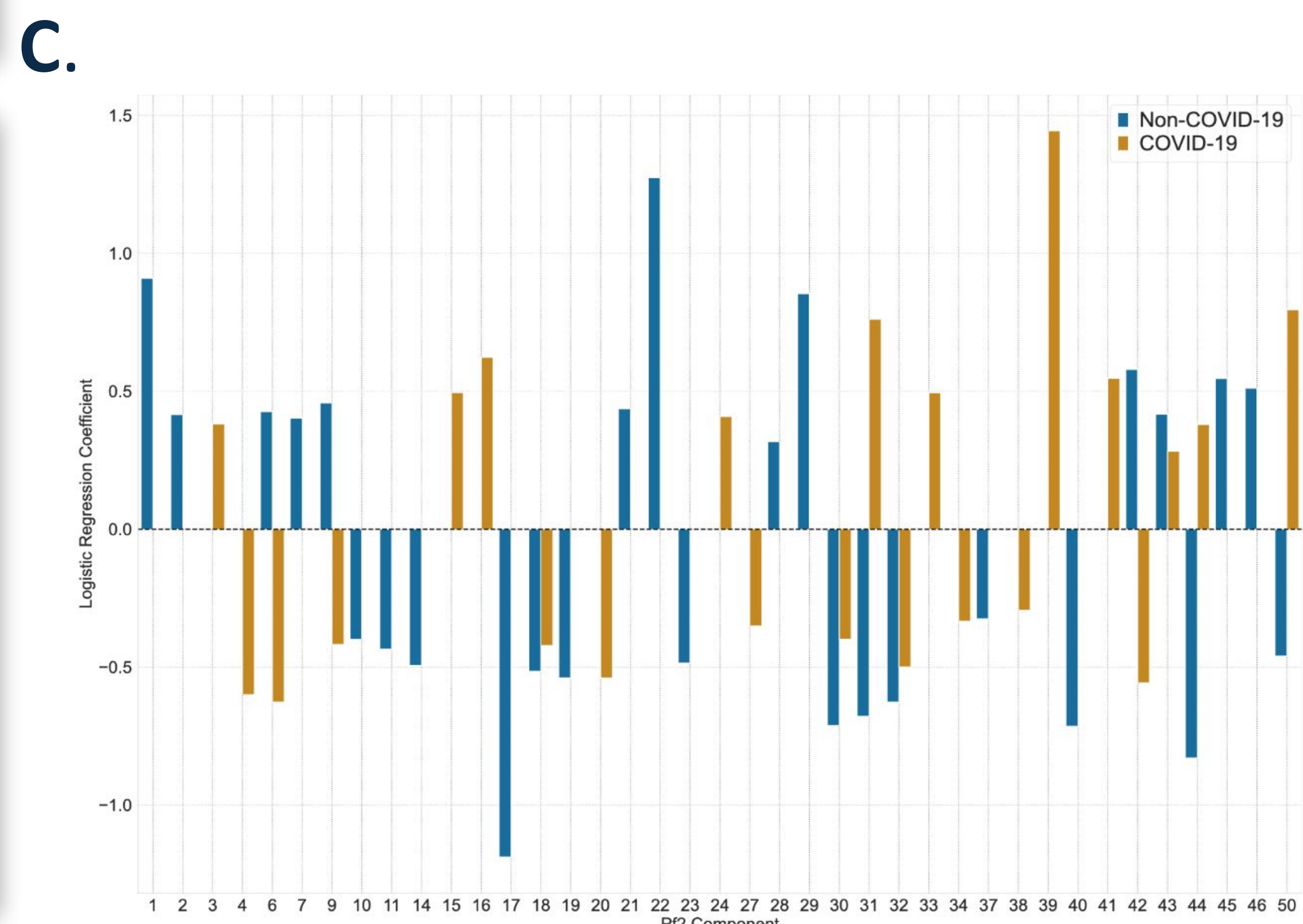
Conclusion

- This study highlights the efficacy of PARAFAC2 in the analysis of scRNA-seq data from patients with COVID-19 pneumonia. The distinct cellular profiles of COVID-19 and Non-COVID-19 patients may elucidate the longer ICU stays and danger compared to other pneumonia types.
- These findings underscore the imperative for precision medicine approaches tailored to the pathophysiology of COVID-19 pneumonia.

Acknowledgements

References

1. Gao, Catherine, et al., (2023). Machine learning links unresolving secondary pneumonia to mortality in patients with severe pneumonia. The Journal of Clinical Investigation, 133(12) https://doi.org/10.1172/JCI170682
2. Ramirez, Andrew, et al., (2024). Integrative, high-resolution analysis of single cells across experimental conditions with PARAFAC2. bioRxiv. https://doi.org/10.1101/2024.07.29.605698