Benchmarking strategies to identify single-cell phenotypic changes

Presented By: Zachary Caterer

Rotated at Fan Zhang Lab at CU Anschutz

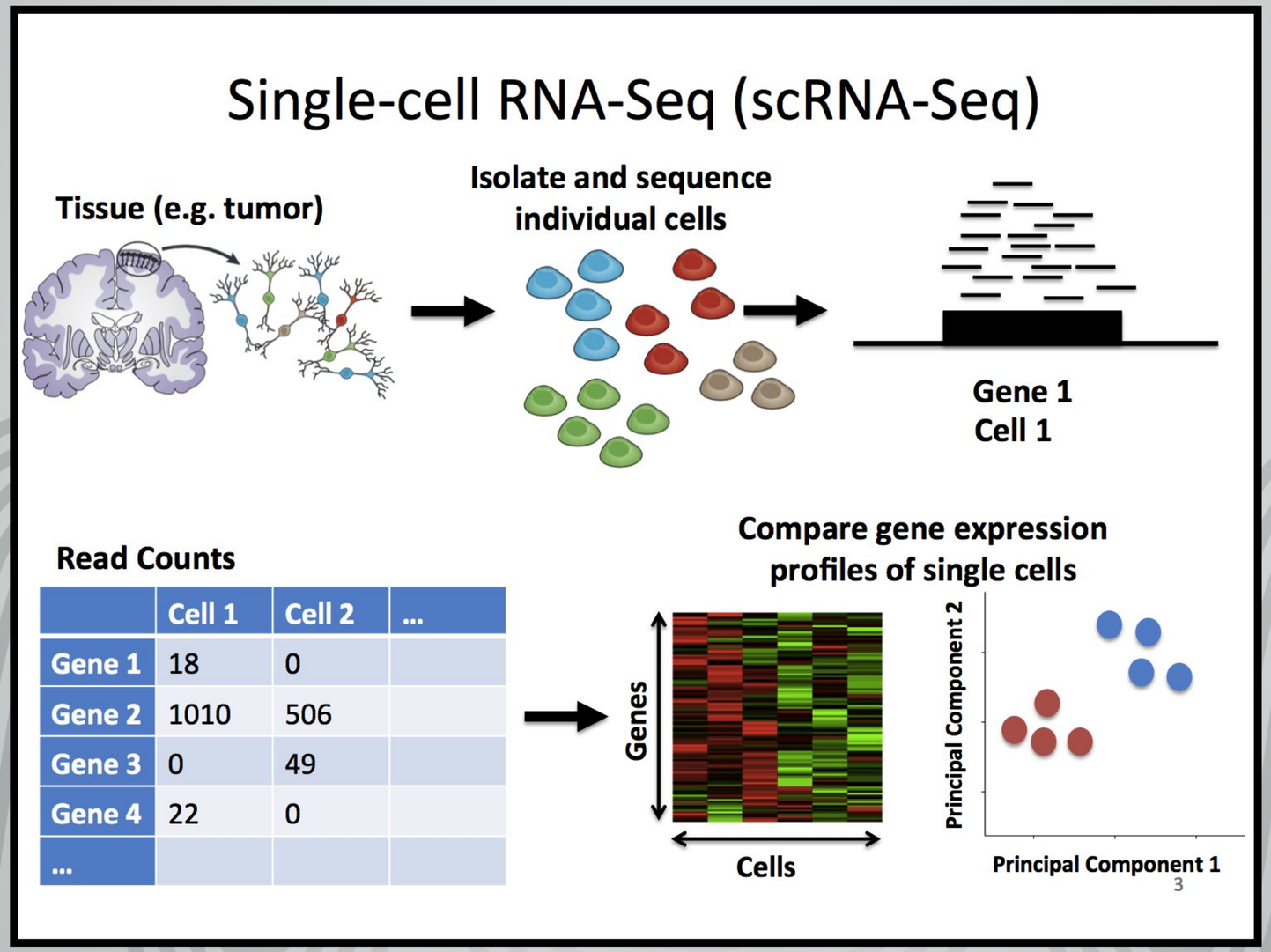
2024 / 10 / 18

fanzhanglab.org

BioFrontiers Institute
UNIVERSITY OF COLORADO **BOULDER**

Background

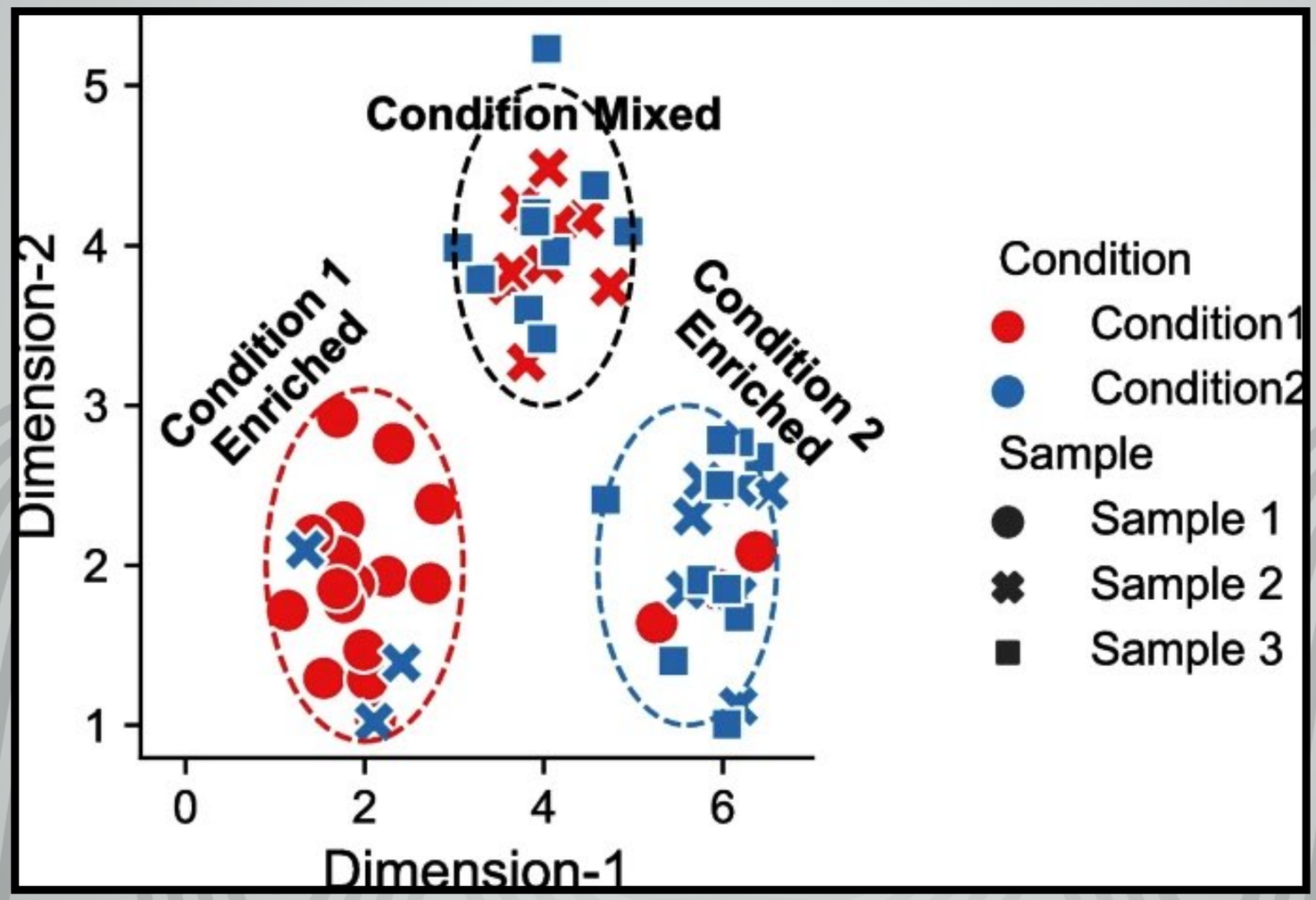
Single-cell RNA-Sequencing



	Cell 1	Cell 2	...
Gene 1	18	0	
Gene 2	1010	506	
Gene 3	0	49	
Gene 4	22	0	
...			

Hicks 2017

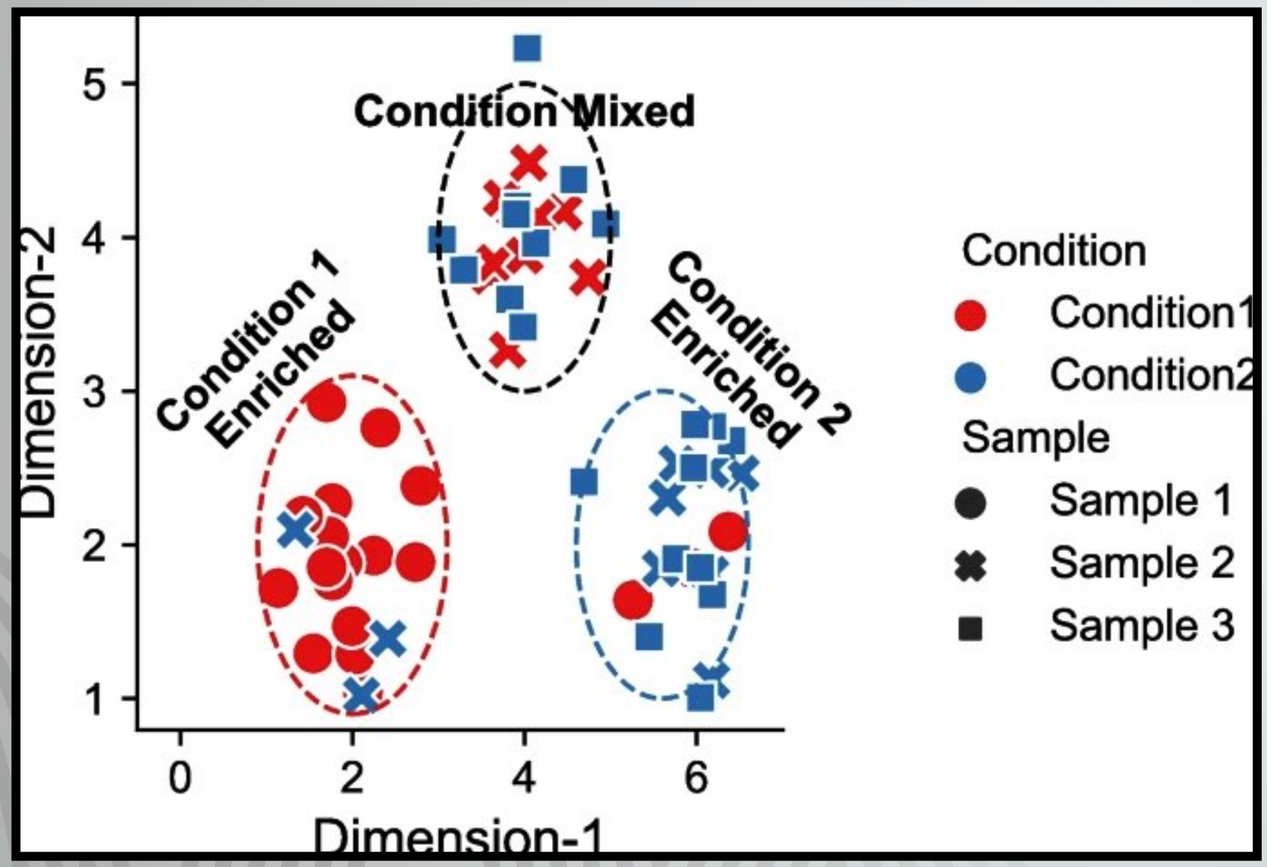
Differential Abundance



Yi 2024

Differential Abundance Clustering

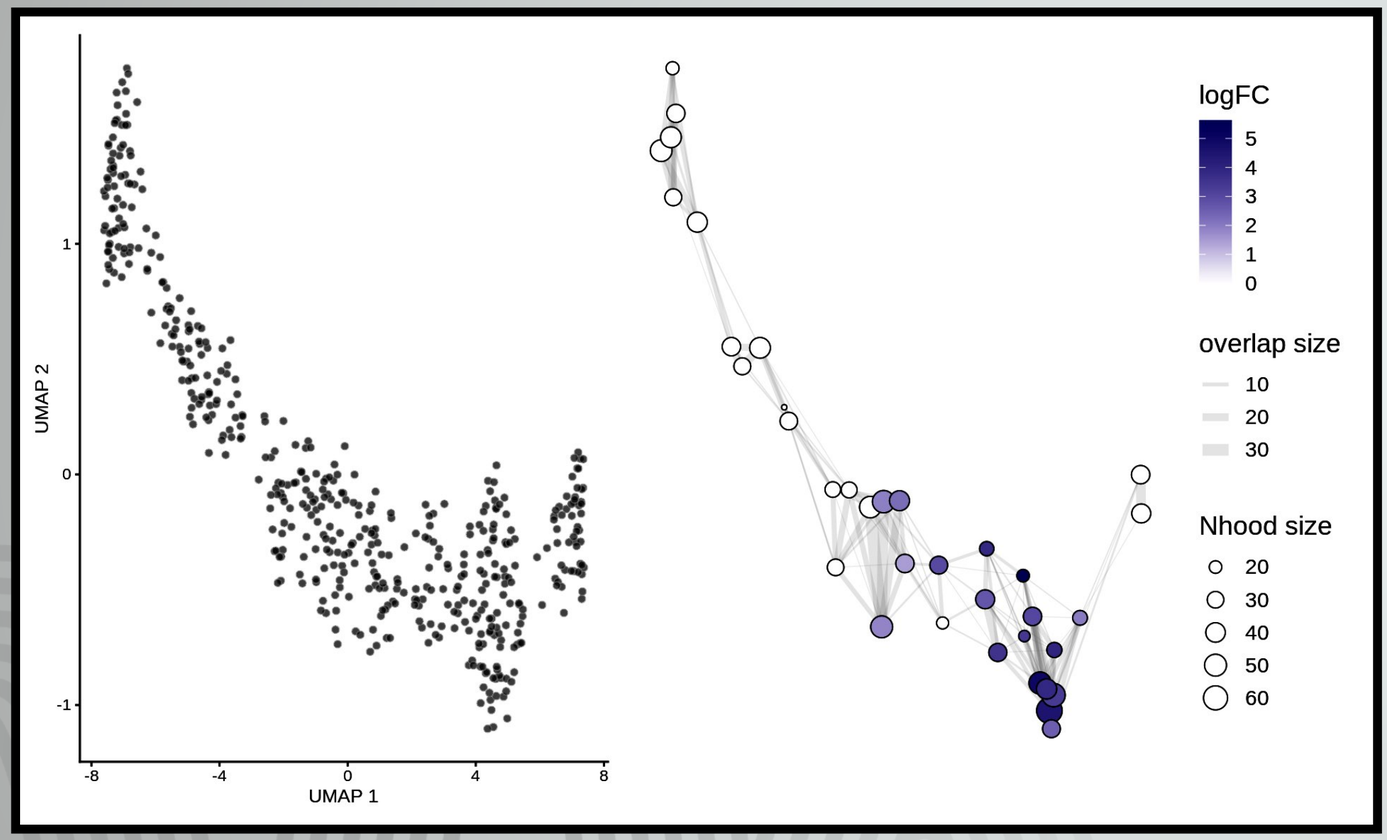
Cluster Based Differential Abundance



Yi 2024

1. Group similar cells into clusters
2. Compare the abundance of clusters

Cluster Free Differential Abundance



1. Identify of differential abundant cells
2. Inferred which cells are associated with different conditions
3. More robust analysis of cellular heterogeneity

miloR

Dann, *Nat Biotechnol*, 2022.
doi.org/10.1038/s41587-021-01033-z

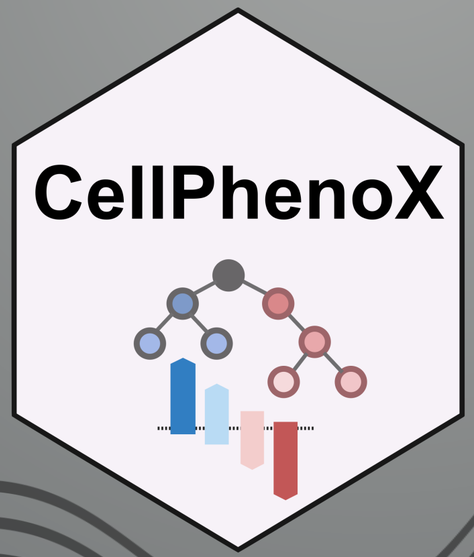
What to do when the difference between healthy and disease cells is small?

Goal

- Identify single cell phenotypic changes
- Understand the factors contributing to these phenotypic changes

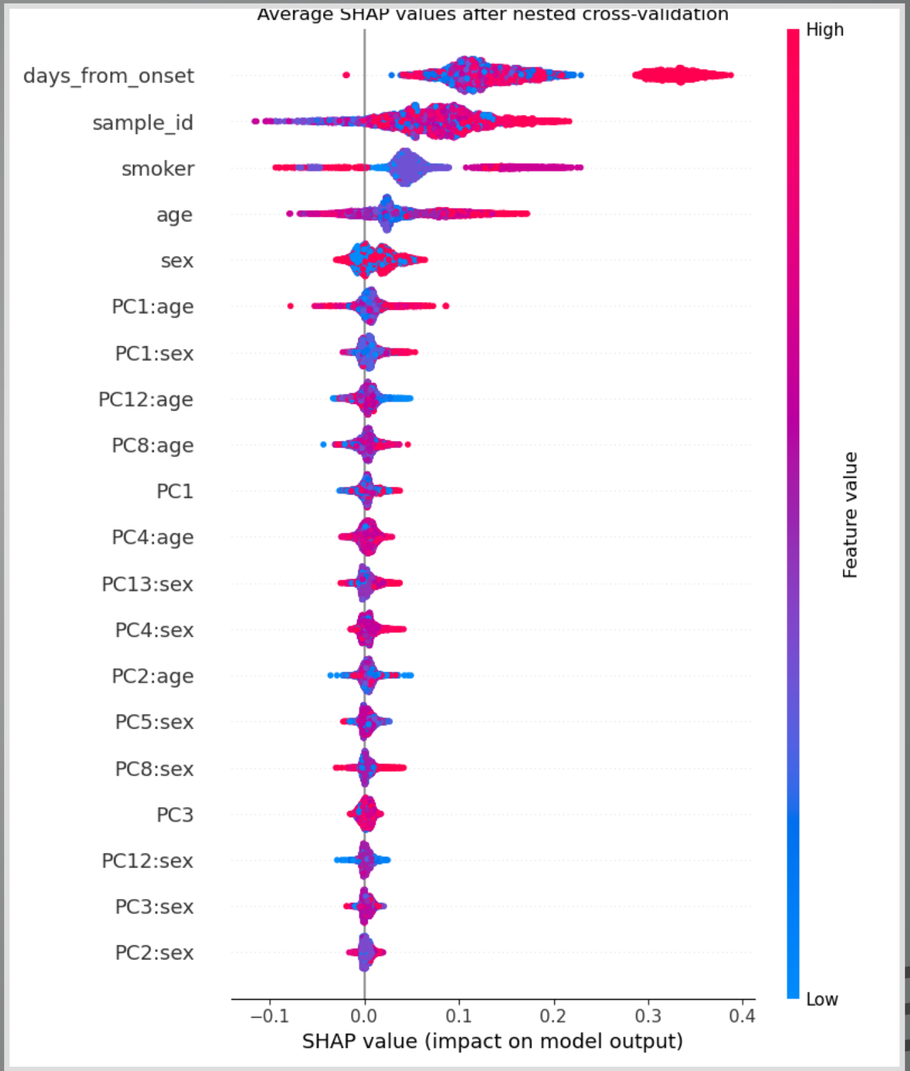
pyCellPhenoX

In development by Jade Young and Zhang Lab at the University of Colorado Anschutz



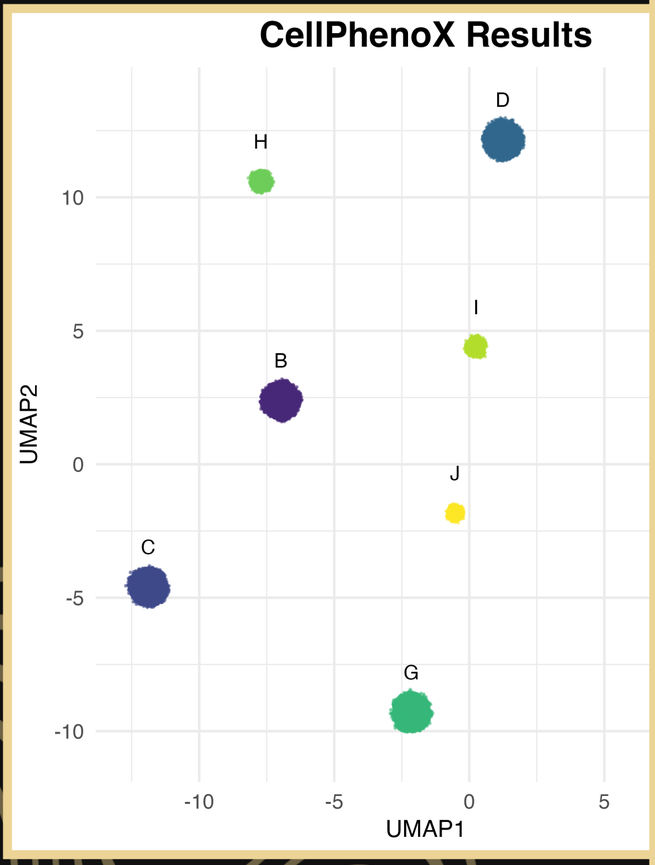
BioFrontiers Institute
UNIVERSITY OF COLORADO **BOULDER**

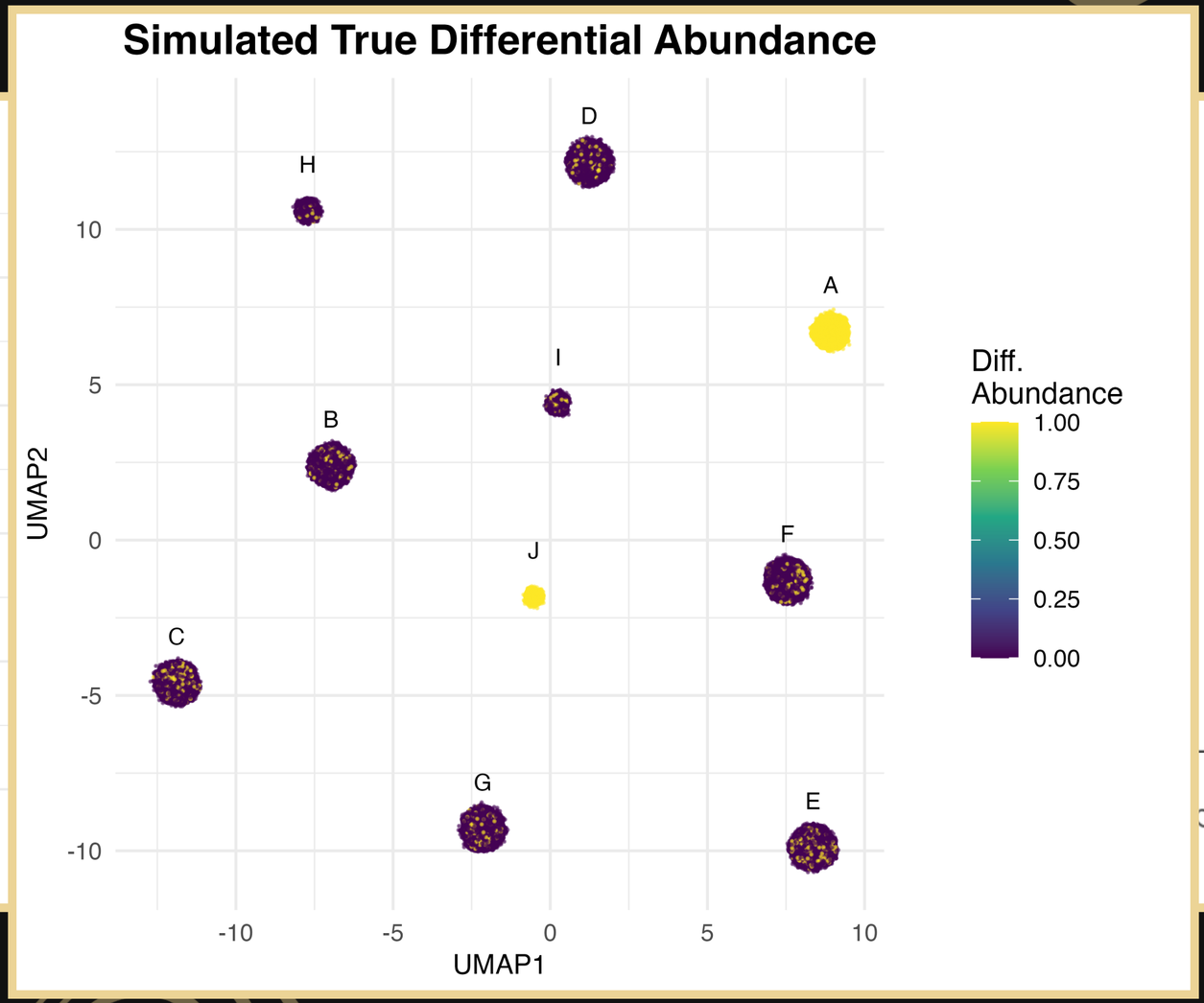
Shapley Additive exPlanations (SHAP) values

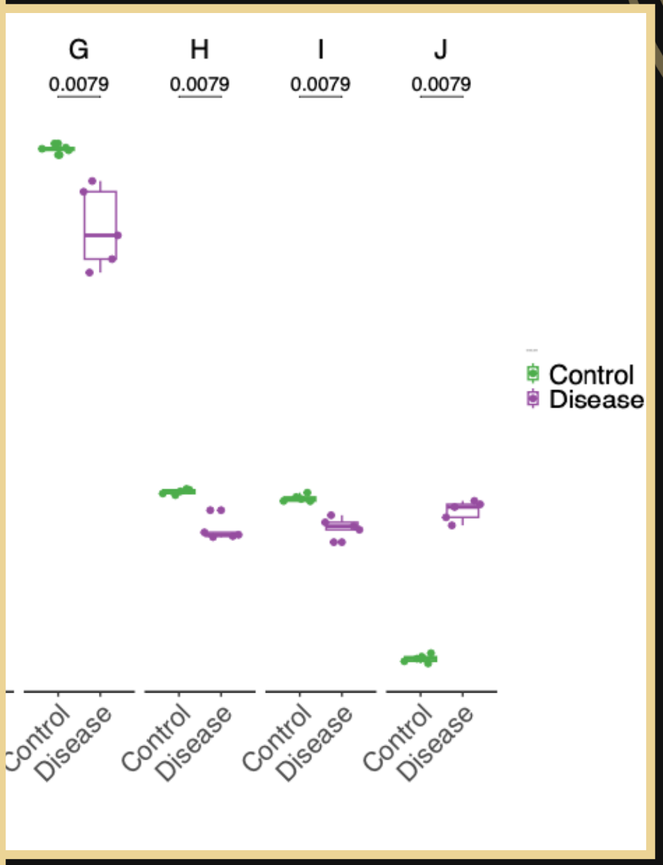


Results

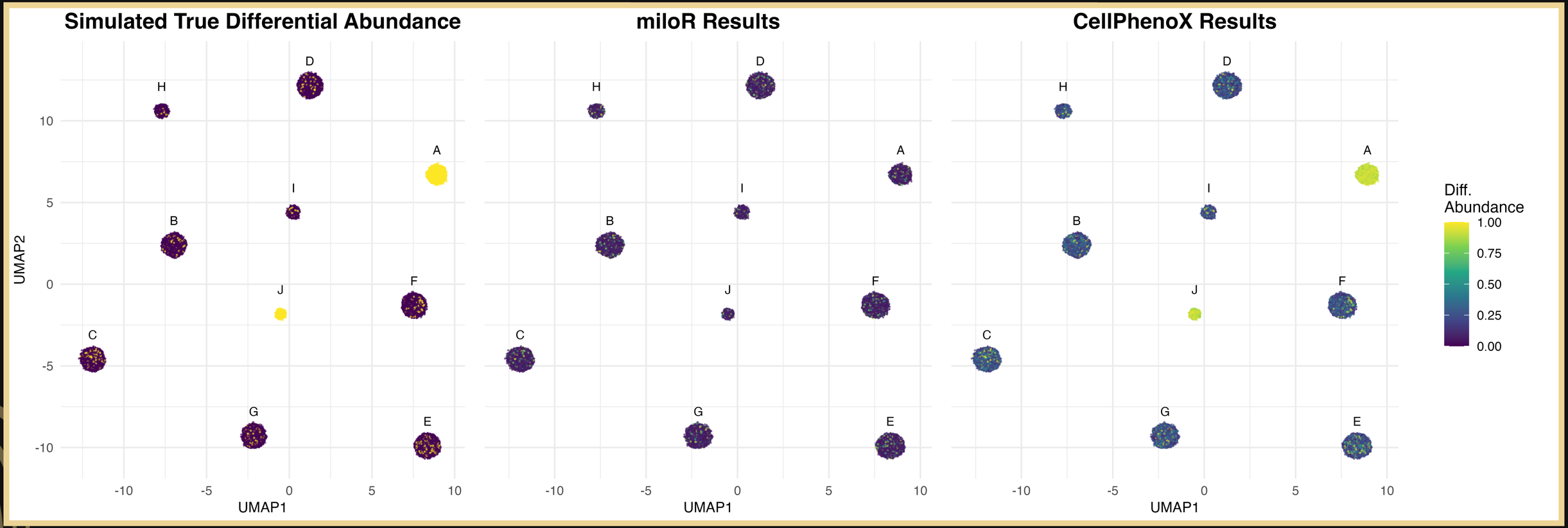
Simulated dataset







miloR vs pyCellPhenoX



Software development of pyCellPhenoX

Importing Dependencies

```
[1]
import pyCellPhenoX

import pandas as pd
```

Python

```
/Users/zhanglab/Documents/Python Projects/pyCellPhenoX/pycpx/lib/python3.12/site-packages/tqdm/auto.py:21: TqdmWarning: IProgress not four
  from .autonotebook import tqdm as notebook_tqdm
```

Step 1: import data

```
[3]
# paths to expression data and meta data files
expression_file = "./../input/uc_fibroblast_exp.csv"
meta_file = "./../input/uc_fibroblast_meta.csv"
output_path = "./../output/"
# read in data
expression_mat = pd.read_csv(expression_file, index_col=0)
meta = pd.read_csv(meta_file, index_col=0)
```

Conclusions

miloR vs CellPhenoX

- Employed **miloR** for differential abundance analysis
- Compared results on same dataset, and CellPhenoX results are promising

Software development

- Refactored **pyCellPhenoX** by cleaning up code and removing unused elements
- Documented with **Sphinx**, deployed to **PyPI, Anaconda** and **Github**.

Acknowledgements

University of Colorado
Anschutz Medical Campus

Department of Medicine Rheumatology
Department of Biomedical Informatics
Center of Health AI

EMAIL fan.3.zhang@cuanschutz.edu

GITHUB github.com/fanzhanglab

WEBSITE fanzhanglab.org

Questions