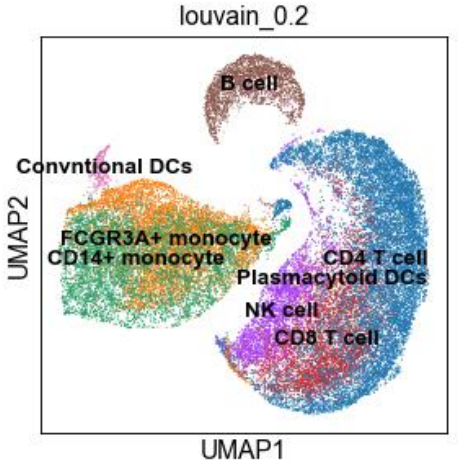
Single cell RNA analysis on PBMCs of COVID-19

WANG, Han- Yi 115513659



Outline

- Background
  - Importance of scRNA analysis on COVID-19 PBMCs
  - Dataset
- Method
- Results
- Conclusion

1. Background

258, 164, 425

Number of confirmed cases

5, 166, 192

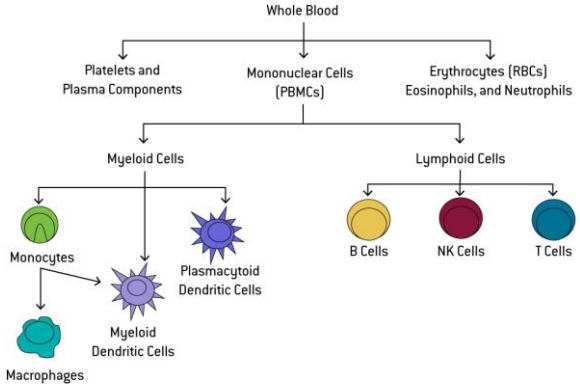
Number of confirmed deaths

scRNA analysis on PBMCs - Importance

- PBMCs - major immune cells in the human body
- Provides a better understanding of **immune response**
- Aid in **vaccine development** or other **therapy**

Dataset - Peripheral Blood Mononuclear Cells of COVID-19

- Extracted **2 healthy and 2 hospitalized** samples from a dataset of 48 samples, including hospitalized, infected, exposed, and healthy controls
- Instrumentation: Illumina NovaSeq 6000; Raw seq-data processed by 10x Genomics Cell Ranger



2. Method – scRNA Pipeline

Quality Control

- Cell filtering
- Gene filtering
- Mitochondiral filtering
- Doublet removal

Dimension Reduction

- Feature selection
- PCA
- UMAP

Clustering

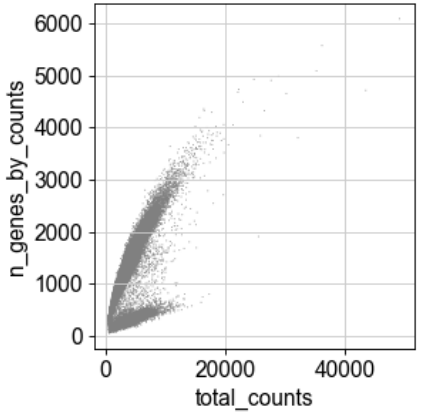
- Louvain

Celltype Prediction

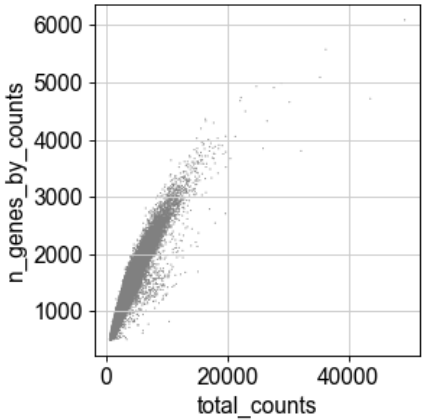
- Differential gene analysis
- Evaluation: Ingest

1. Quality Control

- **Basic filtering**
  - Cells with fewer than 500 genes & genes expressed in less than 10 cells are removed
- **Mitochondiral filtering**
  - High percentage of mitochondrial gene indicates apoptosis
- **Doublet removal**
  - Extremely high number of total counts in cells might be doublets
  - Remove by doublet simulation – *Scrublet* python package



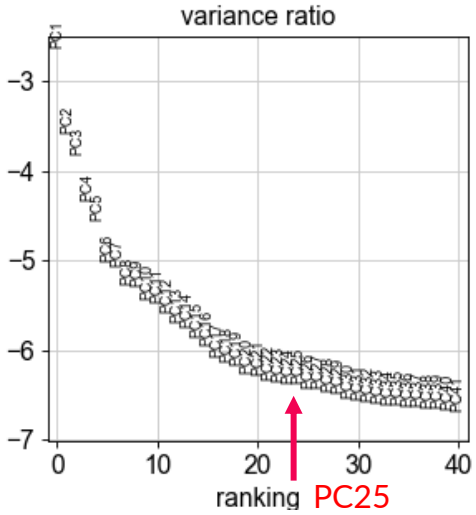
Before filtering



After filtering – better linear relationship between total counts and genes expressed

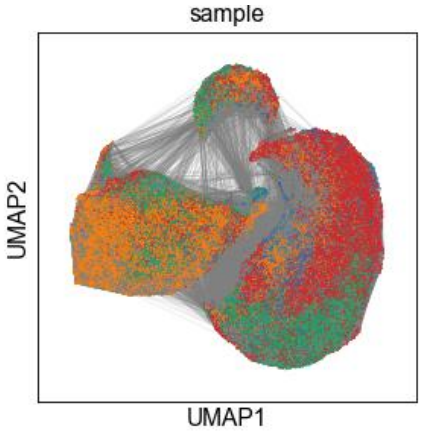
2. Dimension Reduction

- Normalization and logaritmization
- Feature selection: top 2000 variable genes
  - Define genes which provide a good separation of the cell clusters
- PCA
  - Performed on top variable genes
  - Top 25 PCs are extracted – retain most information

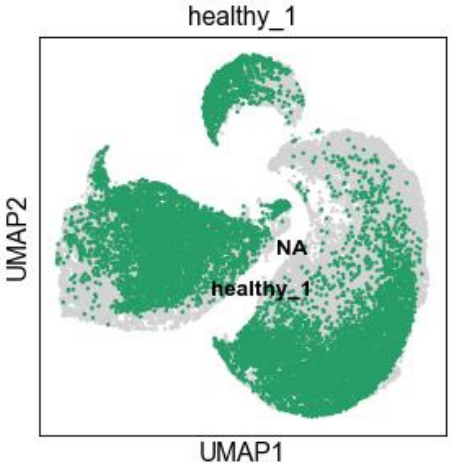


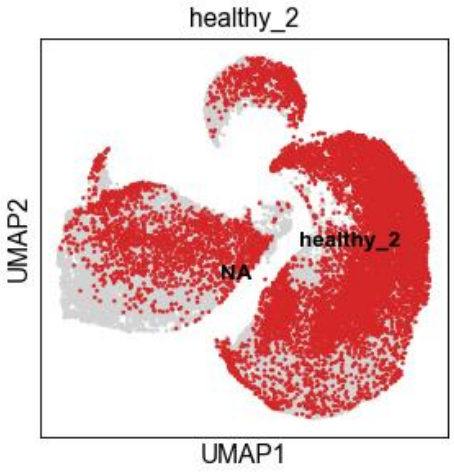
PCA variance ratio graph

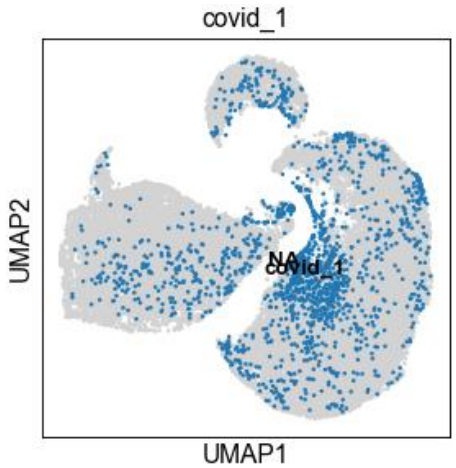
- UMAP – based on neighborhood graph
  - Node: Each node is a cell
  - Edges: Build connection between K nearest neighbors
  - Visualizes cell differences across samples: Healthy - outer, COVID - inner

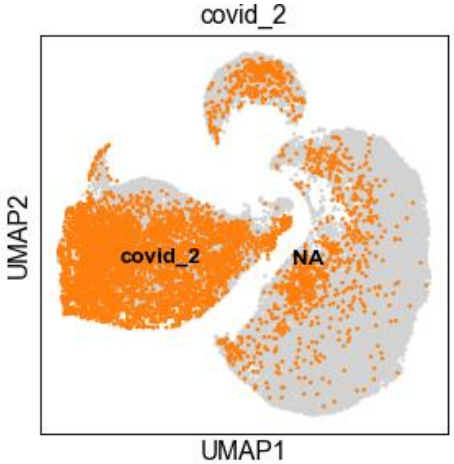


Neighborhood graph







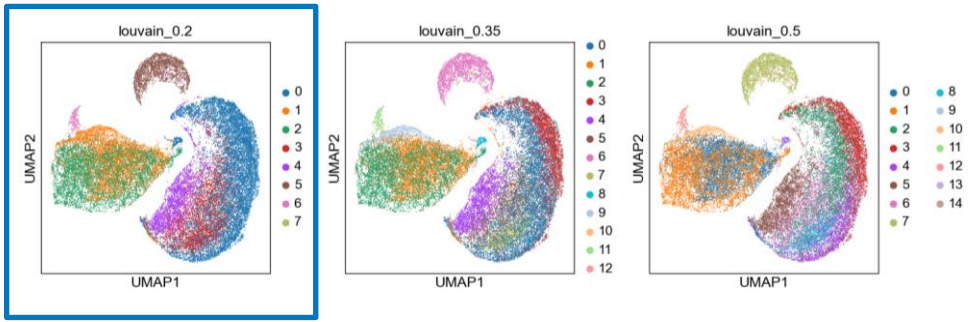


3. Clustering – Graph based clustering

- Louvian - Graph based clustering algorithm
  - Computation of neighborhood graph (done in UMAP)
  - Clustering of neighborhood graph
    - Clusters cell groups that maximizes the connections within the group

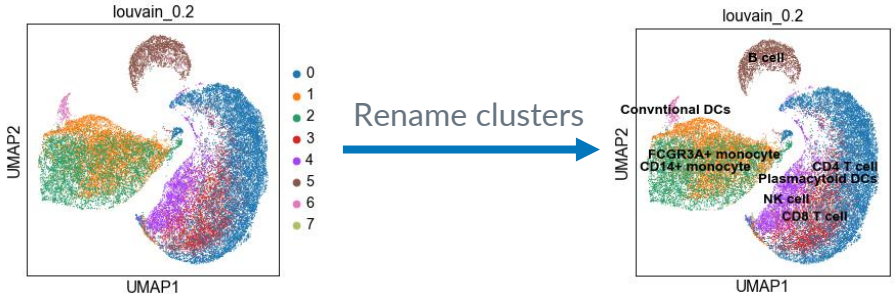
- Resolution parameter
  - Defines the detailedness of the clusters
  - Low: identify major cell types of PBMCs (this project=0.2)
  - High: detailed PBMC subpopulation analysis

Higher resolution, more detailed clusters



4. Celltype Prediction

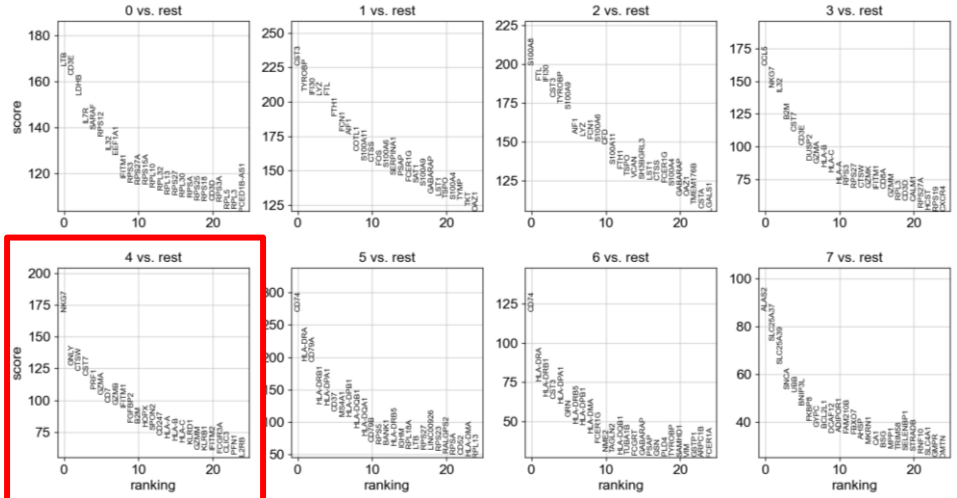
- Identify marker gene in each cluster by differential gene analysis
- Match marker gene with cell type
- Rename the clusters by identified cell types

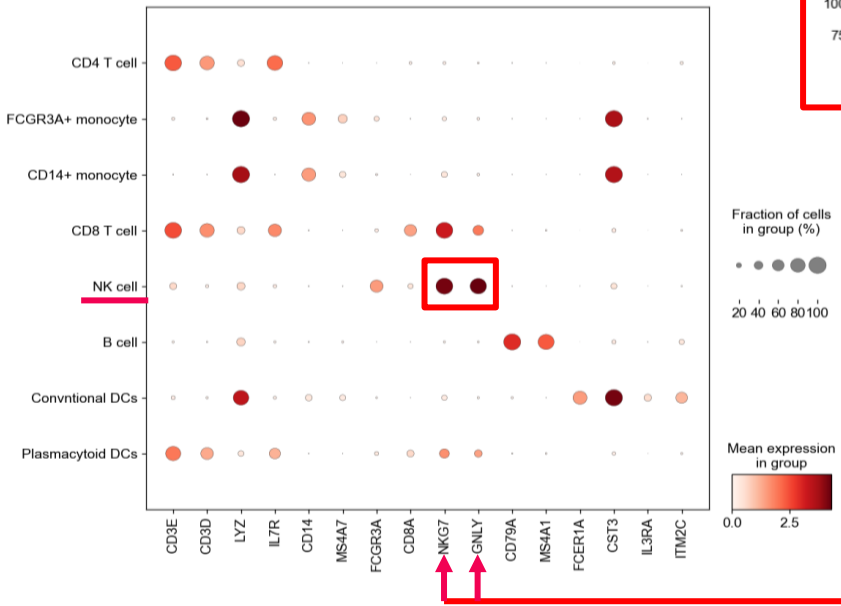


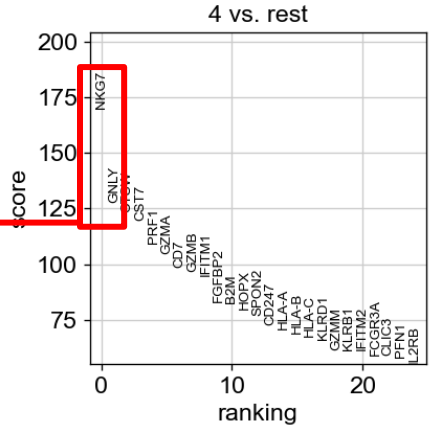
Cell Type	Gene Markers
CD4 T cells	CD3E, CD3D, IL7R
FCGR3A+ Monocytes	LYZ, CST3, FCGR3A
CD14+ Monocytes	LYZ, CST3,CD14
CD8 T cells	CD3E, CD8A
NK cells	NKG7, GNLY
B cells	CD79A, MS4A1
Conventional Dendritic Cells	CST3, FCER1A
Plasmacytoid Dendritic Cells	IL3RA, ITM2C

Gene markers and its correpsonding cell types

- T-test
  - Compute the ranking for highly differential genes in each cluster



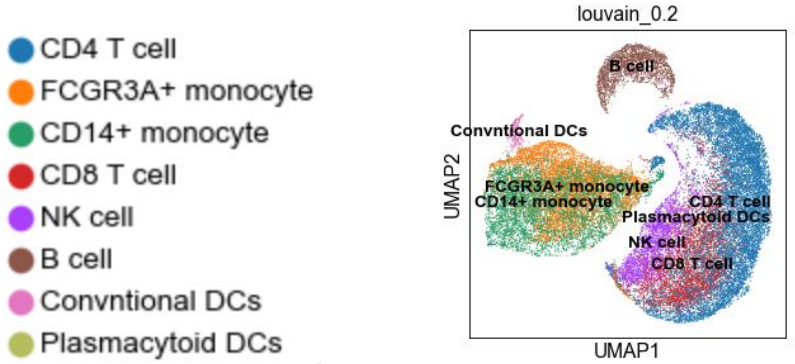


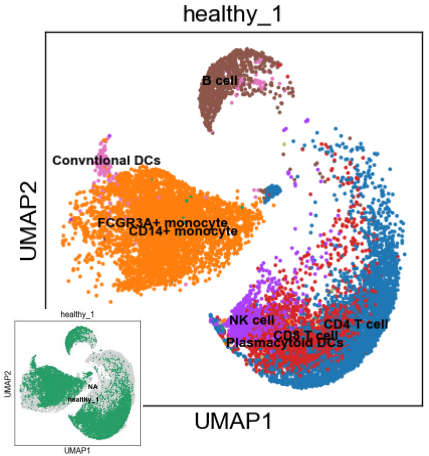


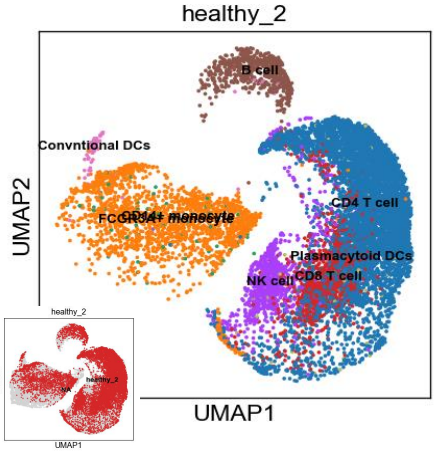
3.
Results & Evaluation

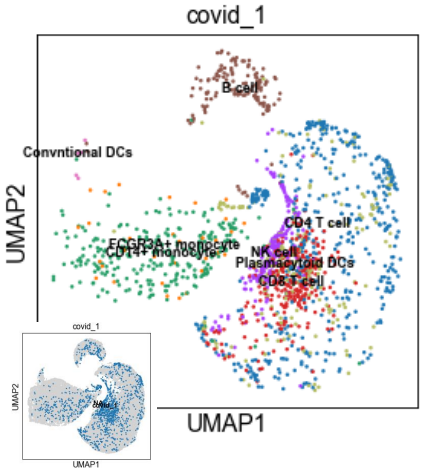
1. Celltype Annotation by Marker Gene

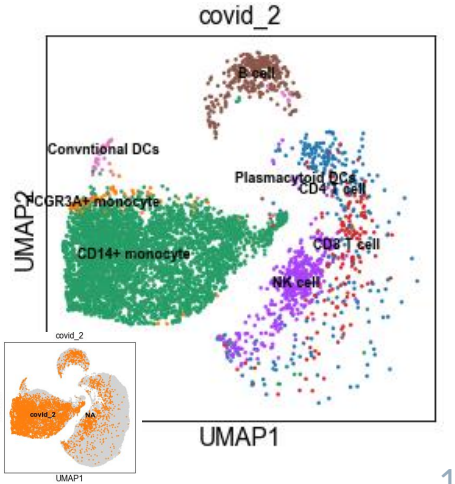
- Decrease of T cells in COVID individuals (blue & red)
- Increase in CD14+ monocyte in patient covid_2 (green)









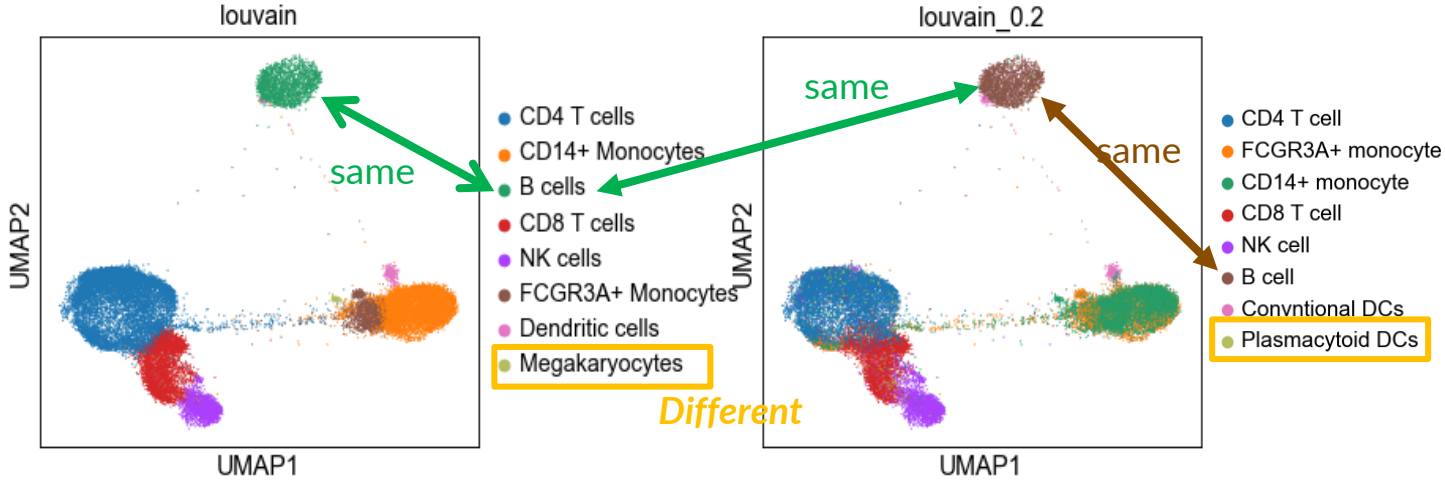


- **Results align with most COVID scRNA analysis: Decreased in T cells population in COVID individuals**
  - T cells are essential in virus clearance
  - Patients with a larger decrease in T cells has a more severe disease status
- **Increase in CD14+ monocytes in covid_2 patient**
  - CD14+ monocytes are responsible for phagocytosis of foreign substances
  - Past analysis demonstrated a positive correlation between expansion of CD14+ monocytes and disease severity
    - Patient covid_2 might be in a more severe disease status

https://www.frontiersin.org/articles/10.3389/fimmu.2021.720109/ful
https://journals.lww.com/bls/fulltext/2020/07000/t_cell_response_in_patients_with_covid_19.2.aspx

2. Evaluation - Celltype Annotation by Ingest

- Dataset is not labelled
- *Ingest python package*: projects labels of reference data to unlabelled data



Transffered annotation from known reference PBMC dataset

Mannual annotation by differential gene analysis

4. Conclusion

- Highlights the major populations of PBMCs
- Provides an insight to pathogenesis of COVID-19
- Future improvements
  - Better quality control: sex filtering, removal of batch effect and dropouts
  - More detailed analysis of subpopulations of PBMC
    - Indicate potential cellular components for **targeted therapy** or **vaccine development**