CBS Project Proposal

Based on Guo, C., Li, B., Ma, H. et al. Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. Nat Commun 11, 3924 (2020)

Michał Stanowski

April 23, 2025

Overview

- monocyte-associated cytokine storm responsive to tocilizumab treatment
- study provides insights into the distribution of immune cells at various stages of COVID-19
- study sheds light on the therapeutic effects of tocilizumab

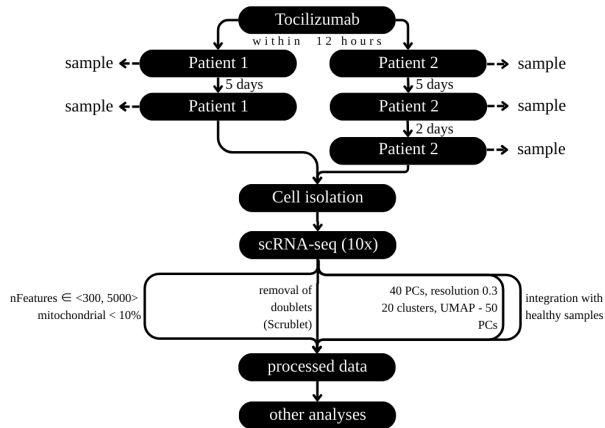Figure: Tocilizumab (Image source: Roche)

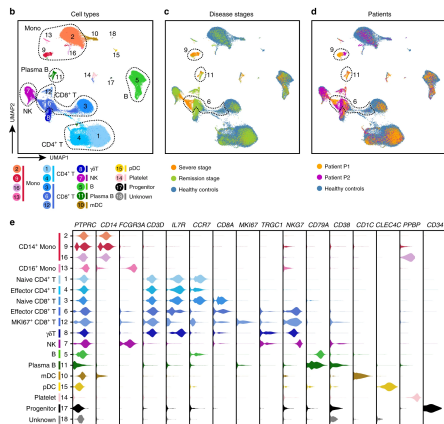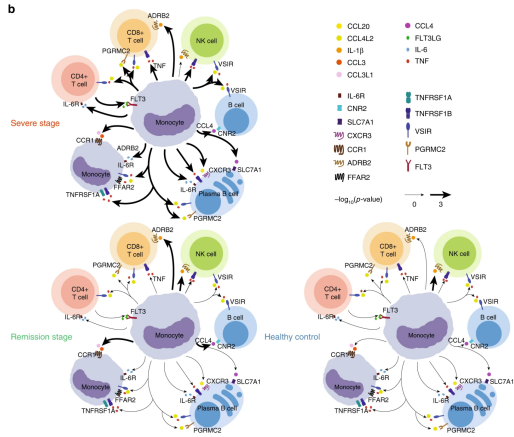Figure: Overview of the methods used in the study

Figure: UMAP representation of cells and violin plots of selected marker genes for multiple cell subpopulations [Guo, 2020]

- identified the **monocyte subpopulation contributing to the cytokine storm**
- they analysed cytokine/recceptor interaction network
- humoral and cellular responses remain untouched

Guo et al. Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. Nat Commun 11, 3924 (2020).

Figure: Potential cytokine/receptor interactions between monocytes and other types of peripheral immune cells [Guo, 2020]

- identified the monocyte subpopulation contributing to the cytokine storm
- **analysed cytokine/receptor interaction network**
- humoral and cellular responses remain untouched

Guo et al. Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. Nat Commun 11, 3924 (2020).

- tried to reproduce the oryginal analysis
- **loaded the data** from severe (GEO) and healthy (10X genomics) patients (count matrices along with barcodes and features)
- took subsets according to the paper
- normalized, scaled and conducted PCA

```r
library(Seurat)

pbmc_exp.data <- Read10X(data.dir =
"C:/Users/Michał/Documents/Studia/StudiaMagisterskie/Bioinformatyka/RokISemI
I/CBS/cbs/project/matrices/")
pbmc_con.data <- Read10X(data.dir =
"C:/Users/Michał/Documents/Studia/StudiaMagisterskie/Bioinformatyka/RokISemI
I/CBS/cbs/project/filtered_feature_bc_matrix")
pbmc_exp <- CreateSeuratObject(counts = pbmc_exp.data, min.cells = 3,
min.features = 200)
pbmc_con <- CreateSeuratObject(counts = pbmc_con.data, min.cell = 3,
min.features = 200)

pbmc_exp[["percent.mt"]] <- PercentageFeatureSet(pbmc_exp, pattern = "^MT-")
pbmc_con[["percent.mt"]] <- PercentageFeatureSet(pbmc_con, pattern = "^MT-")
pbmc_exp[["percent.mt"]]
pbmc_con[["percent.mt"]]
```

Figure: Loading the data

Current state of the analysis (2)

- tried to reproduce the oryginal analysis
- **loaded the data** from severe (GEO) and healthy (10X genomics) patients (count matrices along with barcodes and features)
- **took subsets according to the paper**
- **normalized, scaled and conducted PCA**

```
VlnPlot(pbmc_exp, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"), ncol = 3)
VlnPlot(pbmc_con, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"), ncol = 3)

plot1_exp <- FeatureScatter(pbmc_exp, feature1 = "nCount_RNA", feature2 = "percent.mt")
plot2_exp <- FeatureScatter(pbmc_exp, feature1 = "nCount_RNA", feature2 = "nFeature_RNA")
plot1_exp + plot2_exp

plot1_con <- FeatureScatter(pbmc_con, feature1 = "nCount_RNA", feature2 = "percent.mt")
plot2_con <- FeatureScatter(pbmc_con, feature1 = "nCount_RNA", feature2 = "nFeature_RNA")
plot1_con+ plot2_con

pbmc_exp <- subset(pbmc_exp, subset = nFeature_RNA > 500 & nFeature_RNA < 6000 & percent.mt < 10)
pbmc_con <- subset(pbmc_con, subset = nFeature_RNA > 300 & nFeature_RNA < 5000 & percent.mt < 10)
```

Figure: Taking the subsets of the data

```
pbmc_con <- NormalizeData(pbmc_con)
pbmc_exp <- NormalizeData(pbmc_exp)

pbmc_con <- ScaleData(pbmc_con)
pbmc_exp <- ScaleData(pbmc_exp)

pbmc_con <- FindVariableFeatures(pbmc_con, selection.method = "vst", nfeatures = 5000)
pbmc_exp <- FindVariableFeatures(pbmc_exp, selection.method = "vst", nfeatures = 5000)
```

Figure: Normalization, scaling and PCA

- tried to remove the doublets
- they used **Scrublet** which is wrote in Python
- problems with exporting the data, converting **Seurat to AnnData**
- found many more doublets than original paper (50 vs 374, severe) and (997 vs 25.424, healthy)

```
Preprocessing...
Simulating doublets...
Embedding transcriptomes using PCA...
Calculating doublet scores...
Automatically set threshold at doublet score = 0.58
Detected doublet rate = 2.8%
Estimated detectable doublet fraction = 57.5%
Overall doublet rate:
        Expected   = 25.0%
        Estimated  = 4.9%
Elapsed time: 20.8 seconds
Detected doublet rate = 33.7%
Estimated detectable doublet fraction = 89.8%
Overall doublet rate:
        Expected   = 25.0%
        Estimated  = 37.6%
Wrote 374 doublets to 'doublet_ids.csv' file
```

Figure: Scrublet results on 'severe' matrix

Changes

- run the analysis with different values - e.g. for the mitochondrial cutoff, different doublet finders, different resolutions (obtained 20 clusters via Seurat analysis)

- perform batch effect analysis - is there some technical variation?

- perform thorough analysis of monocyte subclusters via K-means

- comparison with different public datasets (is the monocyte cluster characteristic for COVID-19?)
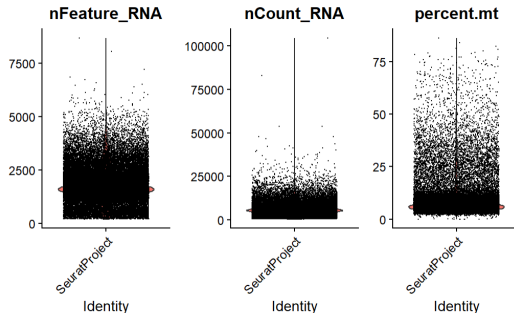
- even trajectory analyses (e. g. Monocle3)

Figure: Healthy sample feature plot

Summary

1. study highlights differences of PBMCs composition between healthy and severe patients and suggests potential tocilizumab molecular mechanism
2. already done: loaded the data, conducted some processing steps, tried to find the doublets with Scrublet
3. to do: find technical variation, conduct analysis with different parameters, explore subclusters of monocyte cluster, compare with other PBMCs datasets

📄 Guo, C. e. a. (2020).
Single-cell analysis of two severe covid-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm.
Nature Communications, 11:3924.