

# Visualizations

Emily Kibbler

2025-08-12

This pipeline is for analysis of scRNA-seq data from a publicly available data set.

Citation:

Phetsouphanh C, [...] Matthews GV. Improvement of immune dysregulation in individuals with long COVID at 24-months following SARS-CoV-2 infection. Nat Commun. 2024 Apr 17;15(1):3315. doi: 10.1038/s41467-024-47720-8. PMID: 38632311; PMCID: PMC11024141.

[Click here to access the data from GEO](#)

This script finds and graphs variable features, plots principal components, and generates UMAP graphs

```
library(anndata)
library(Seurat)
library(MuDataSeurat)
library(BPCells)
library(reticulate)
library(Seurat)
library(Azimuth)
library(tidyverse)
library(ggpubr)
```

## Part 1: 4 month time point

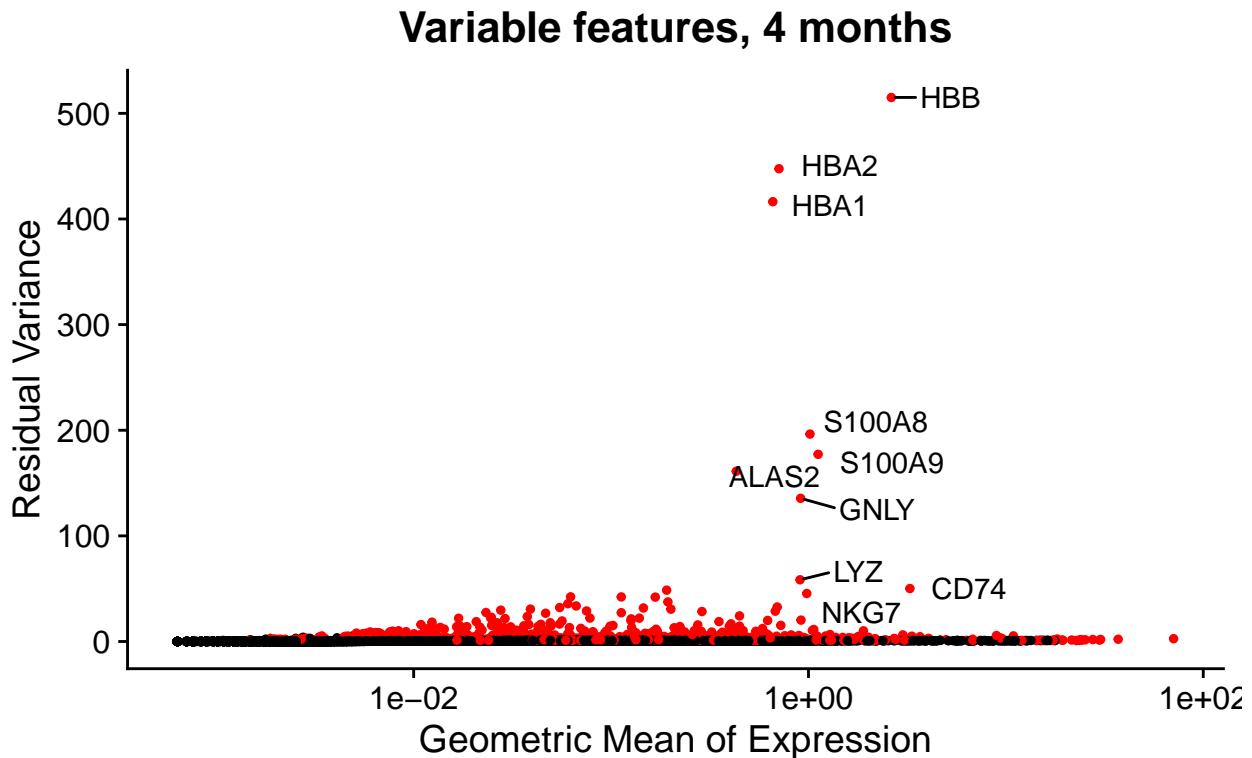
Read data. This R data file is the final output of markdown script titled “Long COVID data analysis: map 4 month data onto reference”

```
combo <- readRDS("./data/covid/4m_mapped_combo.rds")
```

Variable features

```
combo <- FindVariableFeatures(combo)
# Identify the 10 most highly variable genes
top10 <- head(VariableFeatures(combo), 10)
# Plot variable features
VariableFeaturePlot(combo, assay = "integrated") %>%
  LabelPoints(points = top10,
  repel = TRUE) +
  theme(legend.position = "bottom") +
  ggtitle("Variable features, 4 months")

## When using repel, set xnudge and ynudge to 0 for optimal results
```

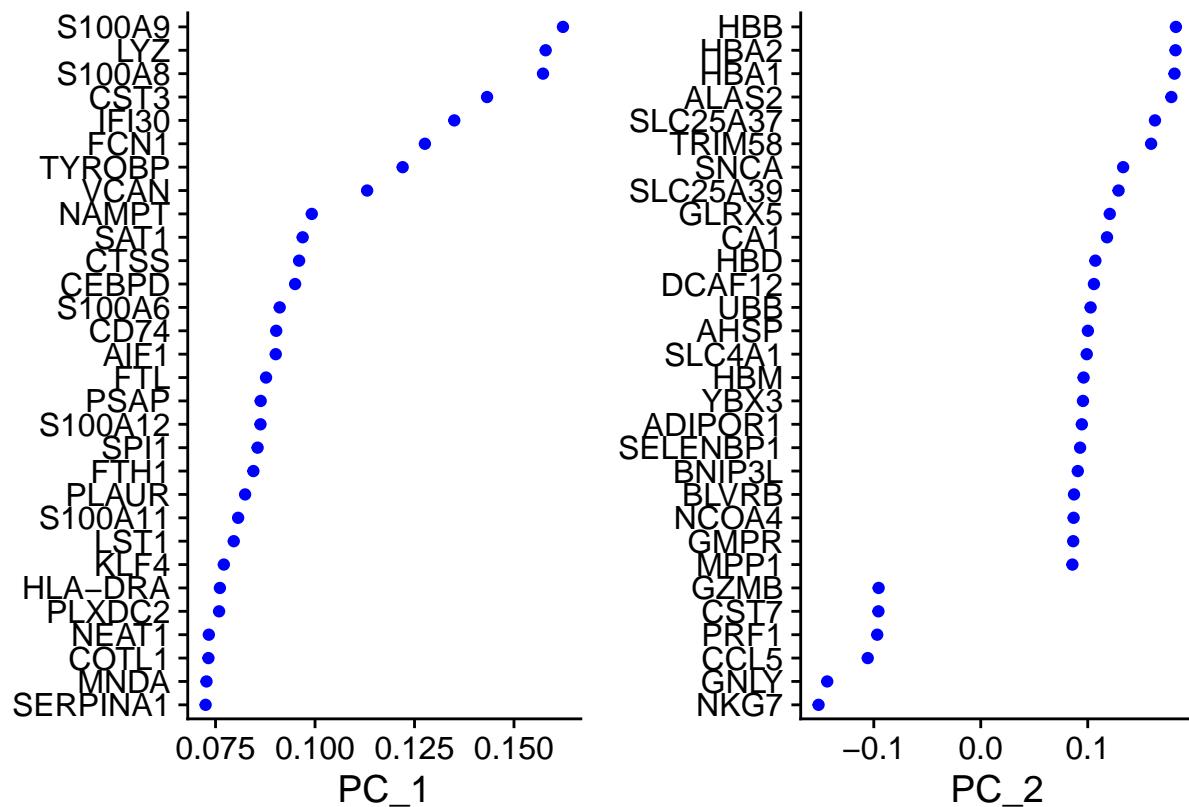


PCA

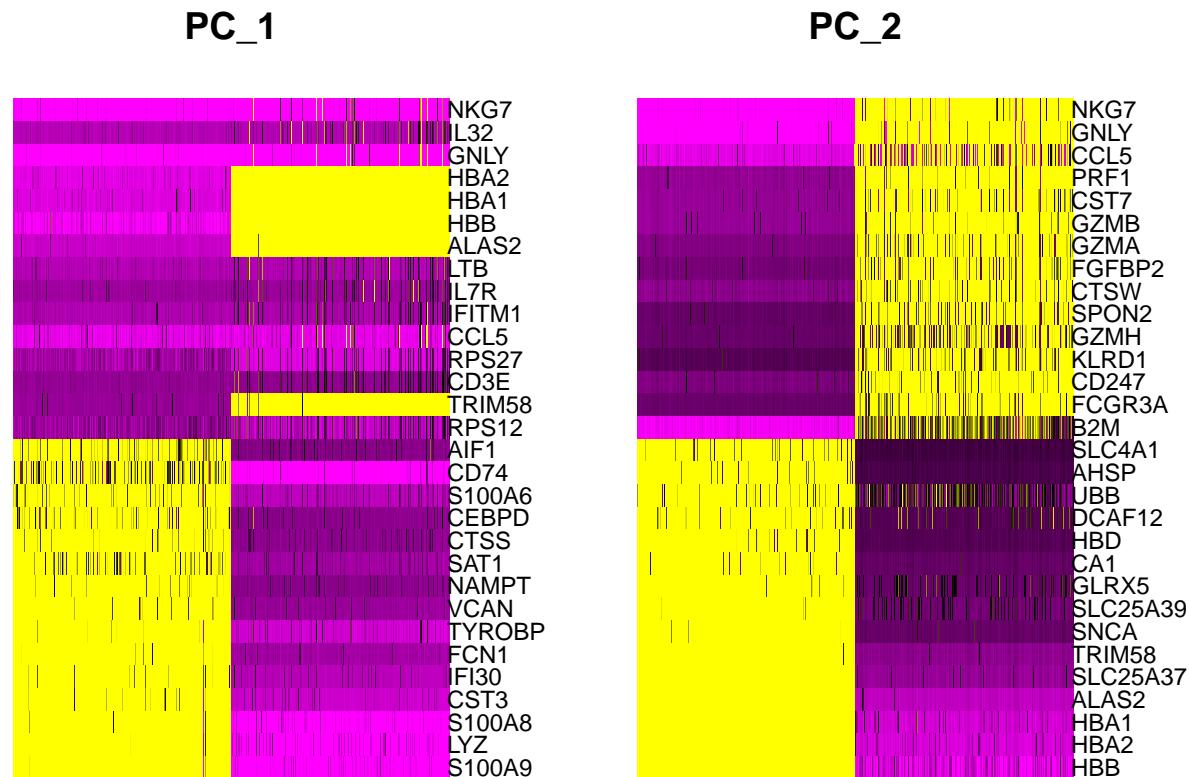
```
print(combo[["pca"]], dims = 1:5, nfeatures = 5)

## PC_ 1
## Positive: S100A9, LYZ, S100A8, CST3, IFI30
## Negative: NKG7, IL32, GNLY, HBA2, HBA1
## PC_ 2
## Positive: HBB, HBA2, HBA1, ALAS2, SLC25A37
## Negative: NKG7, GNLY, CCL5, PRF1, CST7
## PC_ 3
## Positive: LTB, RPS12, EEF1A1, RPL32, RPL13
## Negative: NKG7, GNLY, GZMB, PRF1, CCL5
## PC_ 4
## Positive: S100A8, S100A9, IL7R, LYZ, IL32
## Negative: CD74, HLA-DRA, CD79A, MS4A1, HLA-DRB1
## PC_ 5
## Positive: EEF1A1, NKG7, S100A4, TYROBP, RPL10
## Negative: ARHGAP15, MALAT1, ZBTB20, KLF12, HBB

VizDimLoadings(combo, dims = 1:2, reduction = "pca")
```



```
DimHeatmap(combo, dims = 1:2, cells = 500, balanced = TRUE)
```



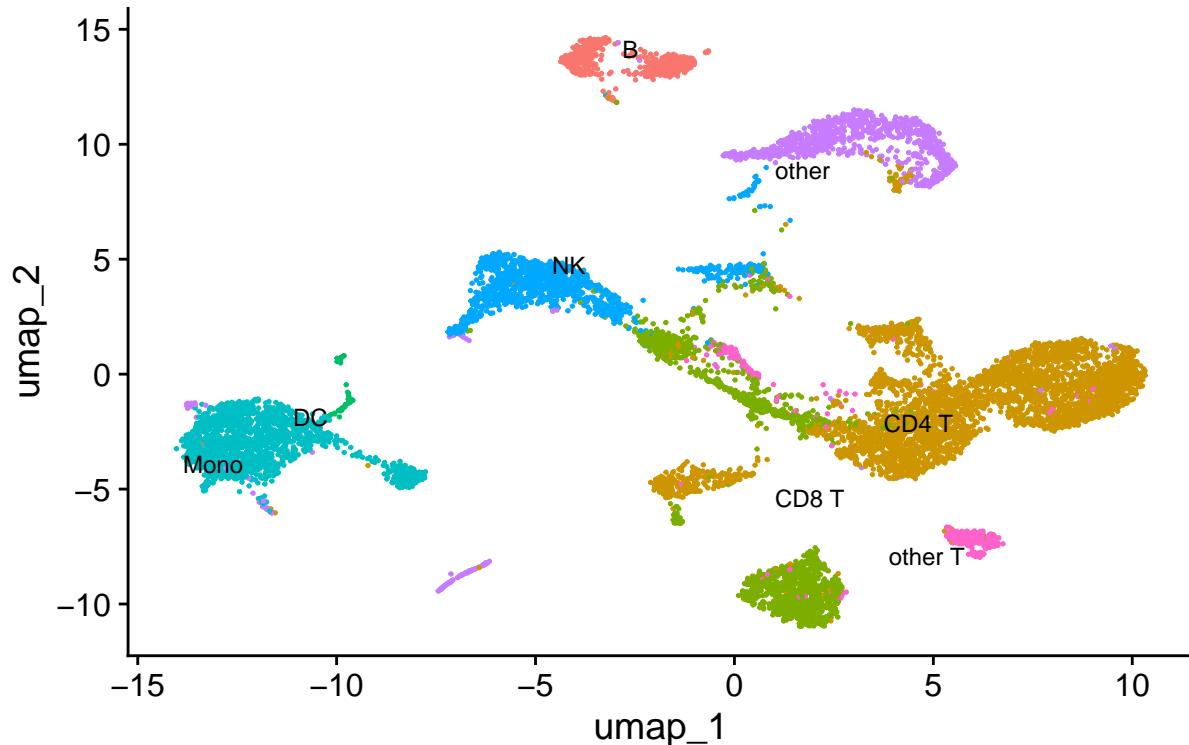
UMAP

```

DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.11",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 1, 4 months")

```

## Cell classification level 1, 4 months

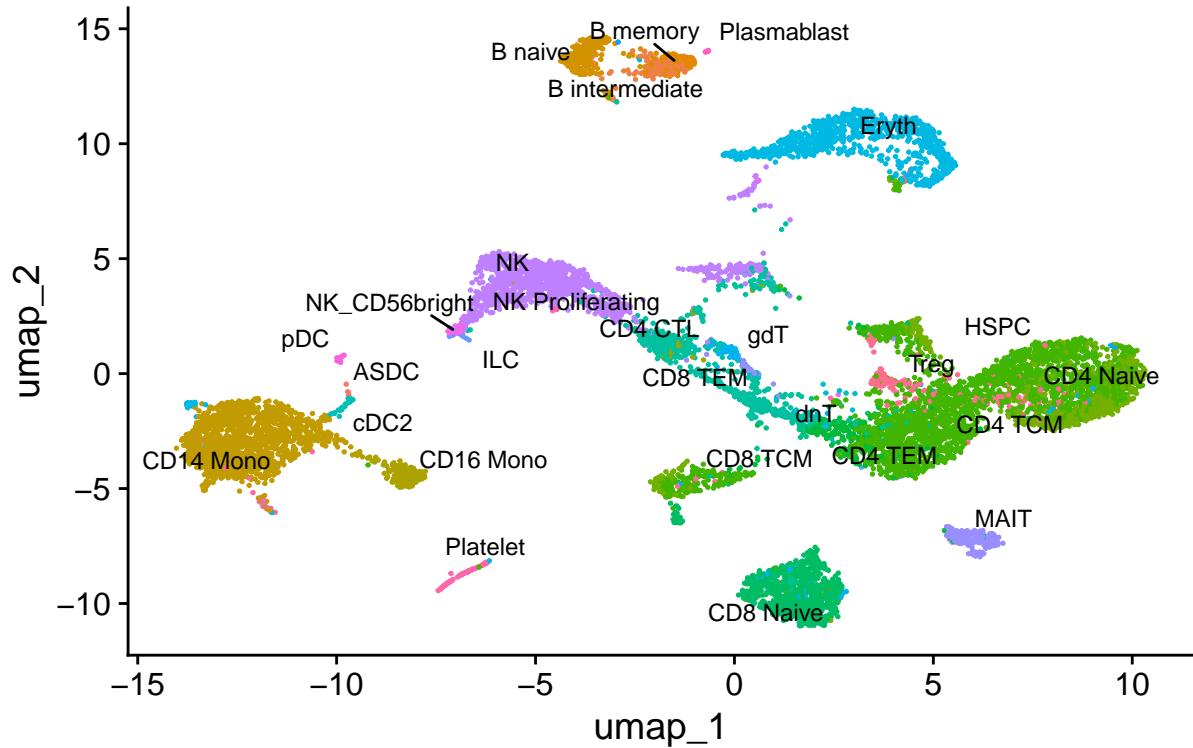


```

DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.12",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 2, 4 months")

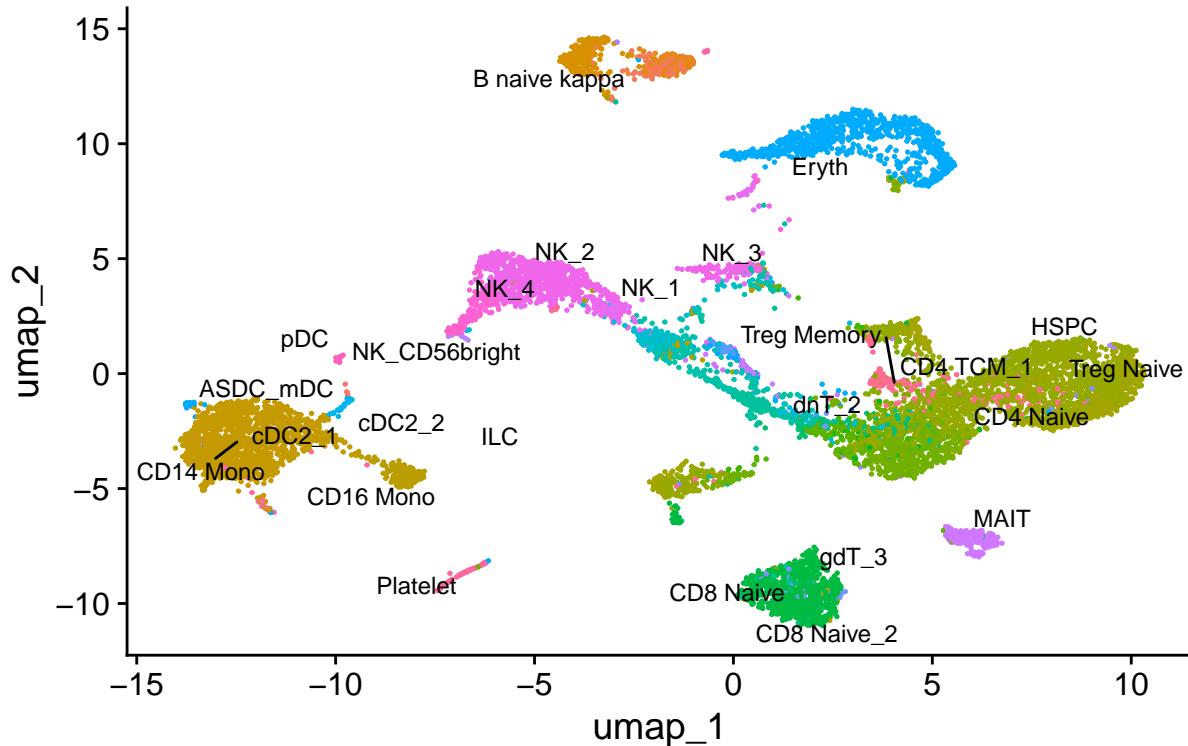
```

## Cell classification level 2, 4 months



```
DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.13",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 3, 4 months")
```

## Cell classification level 3, 4 months



### Part 2: 8 month time point

Read data. This R data file is the final output of markdown script titled “Long COVID data analysis: map 8 month data onto reference”

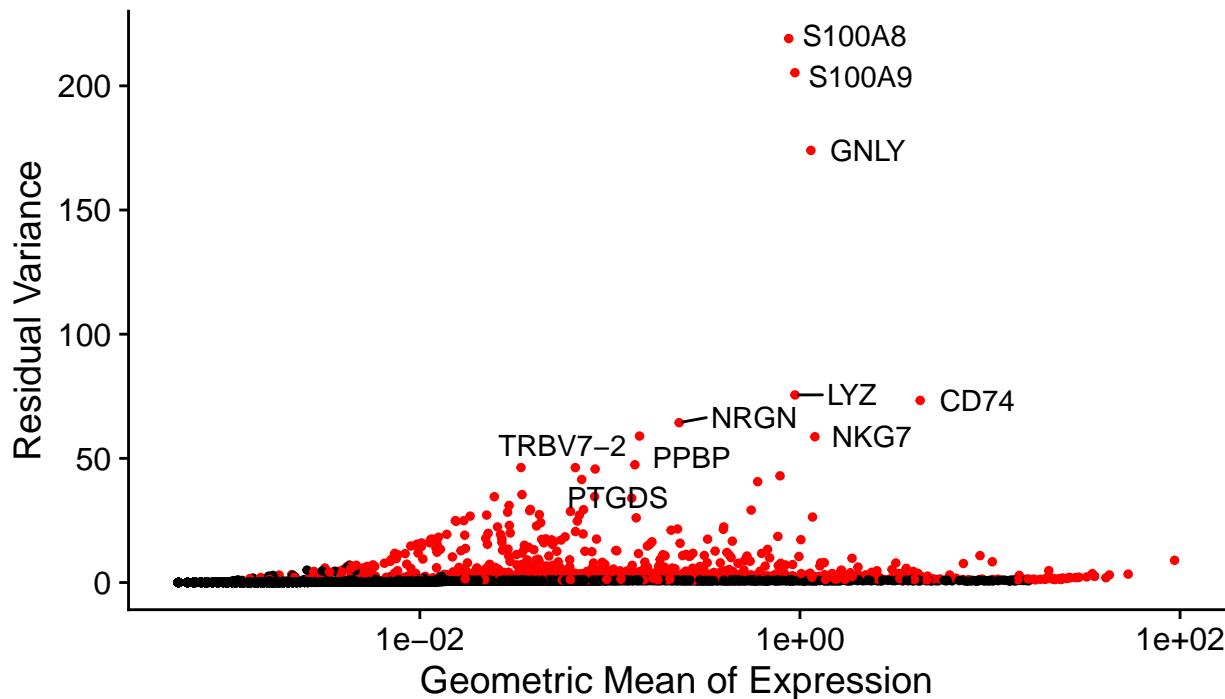
```
combo <- readRDS("./data/covid/8m_mapped_combo.rds")
```

Variable features

```
combo <- FindVariableFeatures(combo)
# Identify the 10 most highly variable genes
top10 <- head(VariableFeatures(combo), 10)
# Plot variable features
VariableFeaturePlot(combo, assay = "integrated") %>%
  LabelPoints(points = top10,
  repel = TRUE) +
  theme(legend.position = "bottom") +
  ggtitle("Variable features, 8 months")
```

## When using repel, set xnudge and ynudge to 0 for optimal results

## Variable features, 8 months



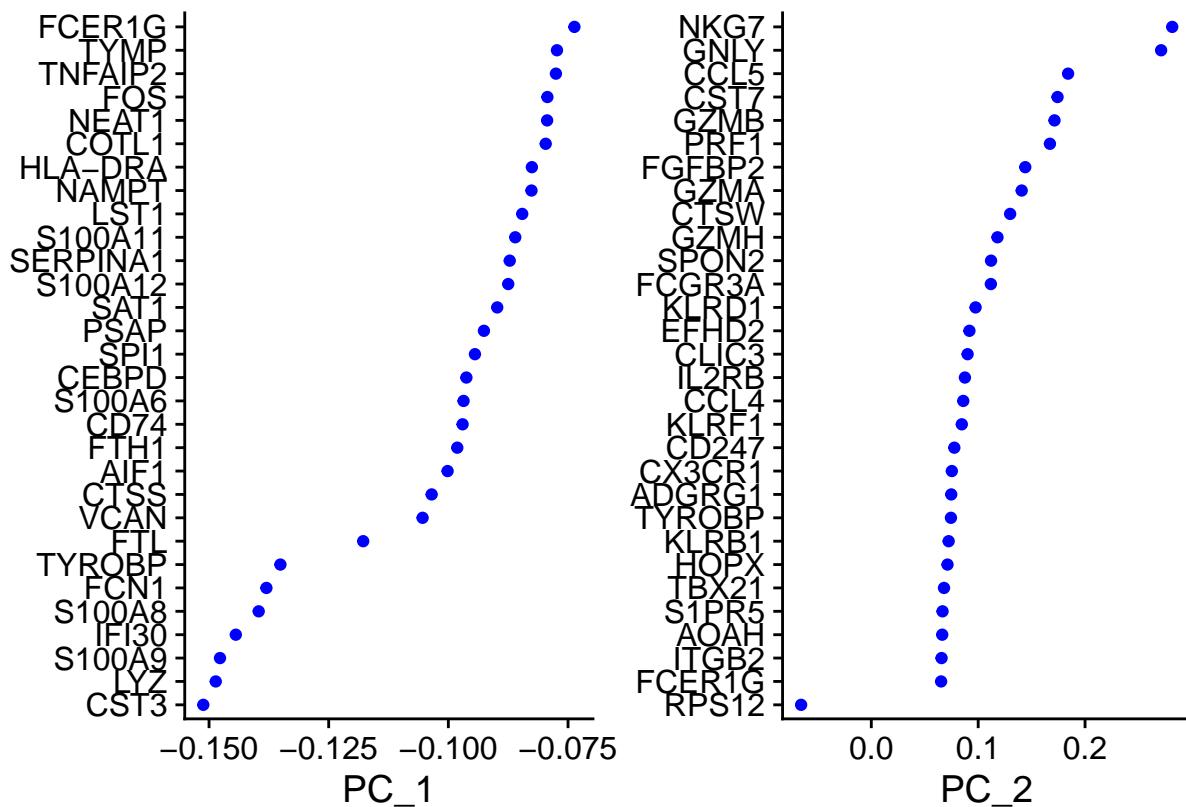
- Non-variable count: 16330    • Variable count: 3000

PCA

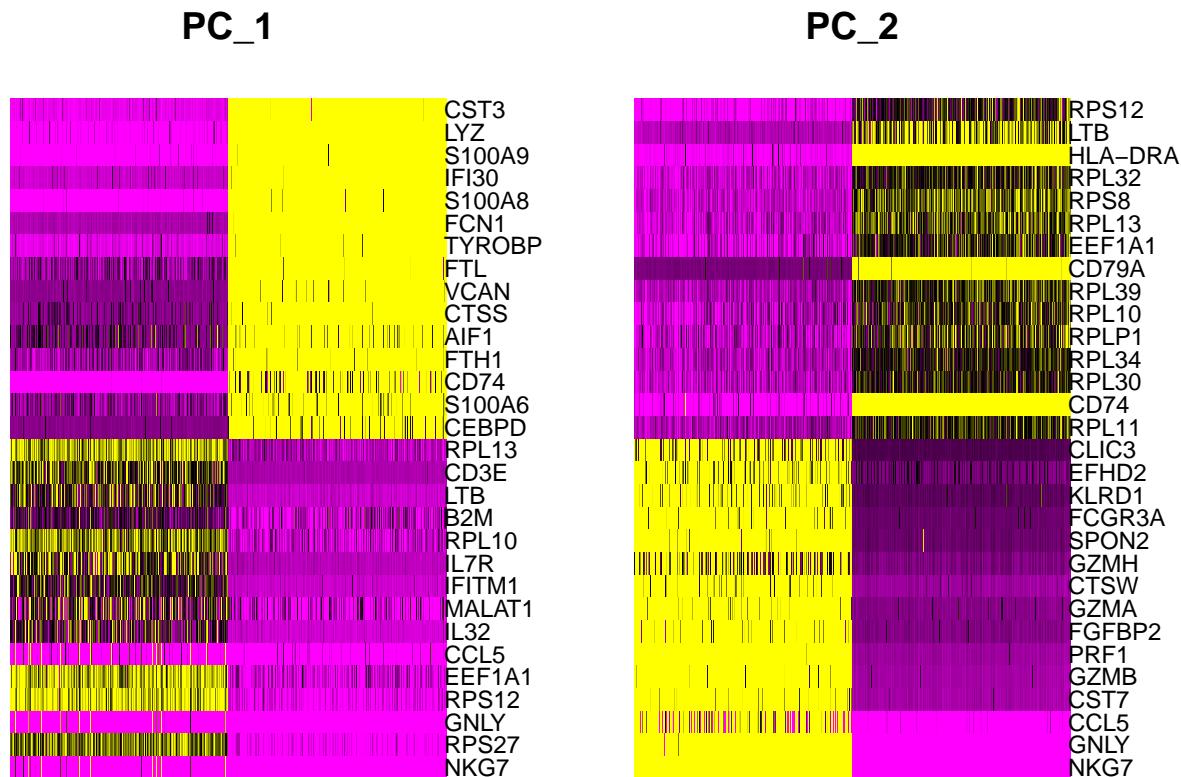
```
print(combo[["pca"]], dims = 1:5, nfeatures = 5)
```

```
## PC_ 1
## Positive: NKG7, RPS27, GONLY, RPS12, EEF1A1
## Negative: CST3, LYZ, S100A9, IFI30, S100A8
## PC_ 2
## Positive: NKG7, GONLY, CCL5, CST7, GZMB
## Negative: RPS12, LTB, HLA-DRA, RPL32, RPS8
## PC_ 3
## Positive: CD74, HLA-DRA, CD79A, MS4A1, HLA-DPA1
## Negative: IL7R, S100A8, S100A9, IL32, S100A6
## PC_ 4
## Positive: NKG7, EEF1A1, CD74, RPL41, ACTB
## Negative: MALAT1, ARHGAP15, FOXP1, PDE3B, KLF12
## PC_ 5
## Positive: NRGN, PPBP, TUBB1, CAVIN2, PF4
## Negative: NKG7, GONLY, MALAT1, TYROBP, EEF1A1
```

```
VizDimLoadings(combo, dims = 1:2, reduction = "pca")
```



```
DimHeatmap(combo, dims = 1:2, cells = 500, balanced = TRUE)
```



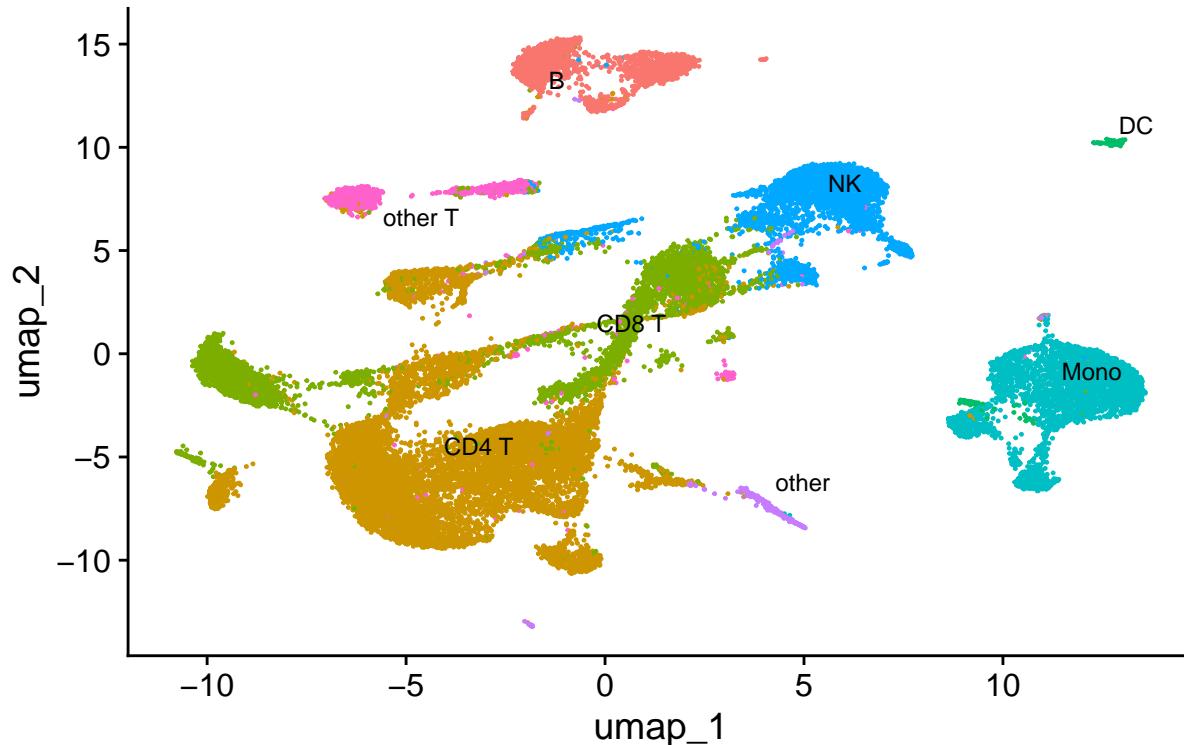
UMAP

```

DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.11",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 1, 8 months")

```

## Cell classification level 1, 8 months

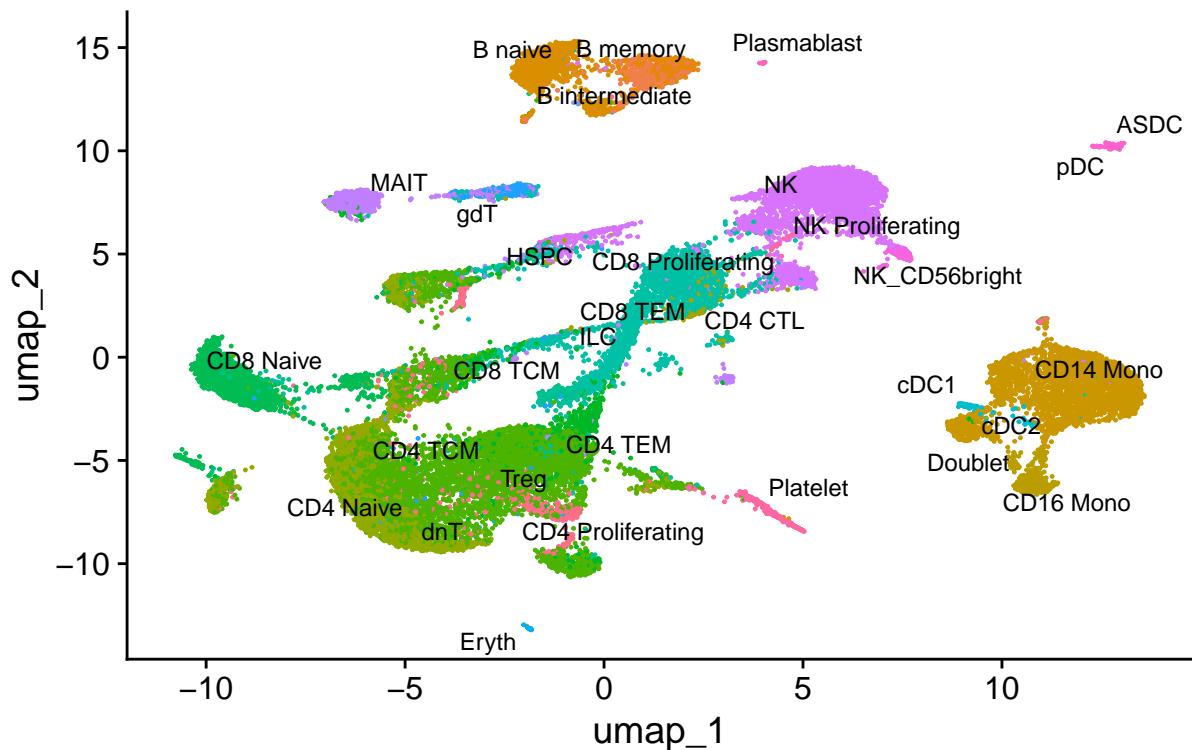


```

DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.12",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 2, 8 months")

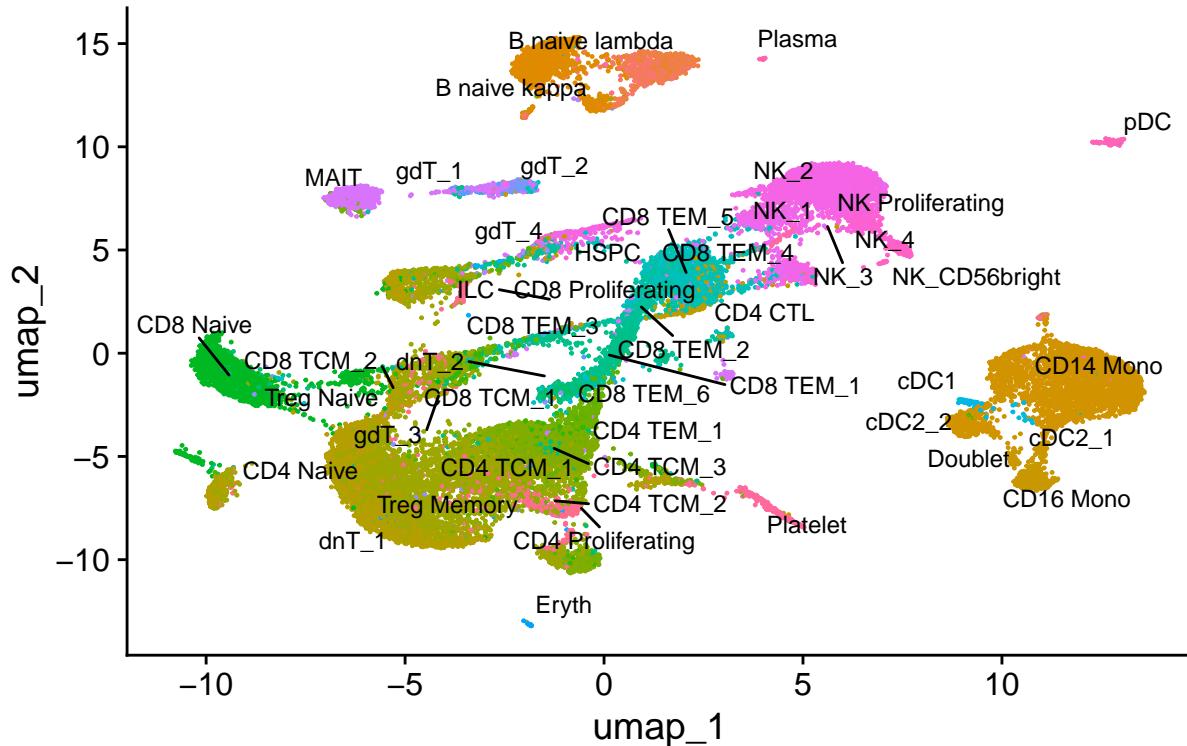
```

## Cell classification level 2, 8 months



```
DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.13",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 3, 8 months")
```

## Cell classification level 3, 8 months



### Part 3: 24 month time point

Read data. This R data file is the final output of markdown script titled “Long COVID data analysis: map 24 month data onto reference”

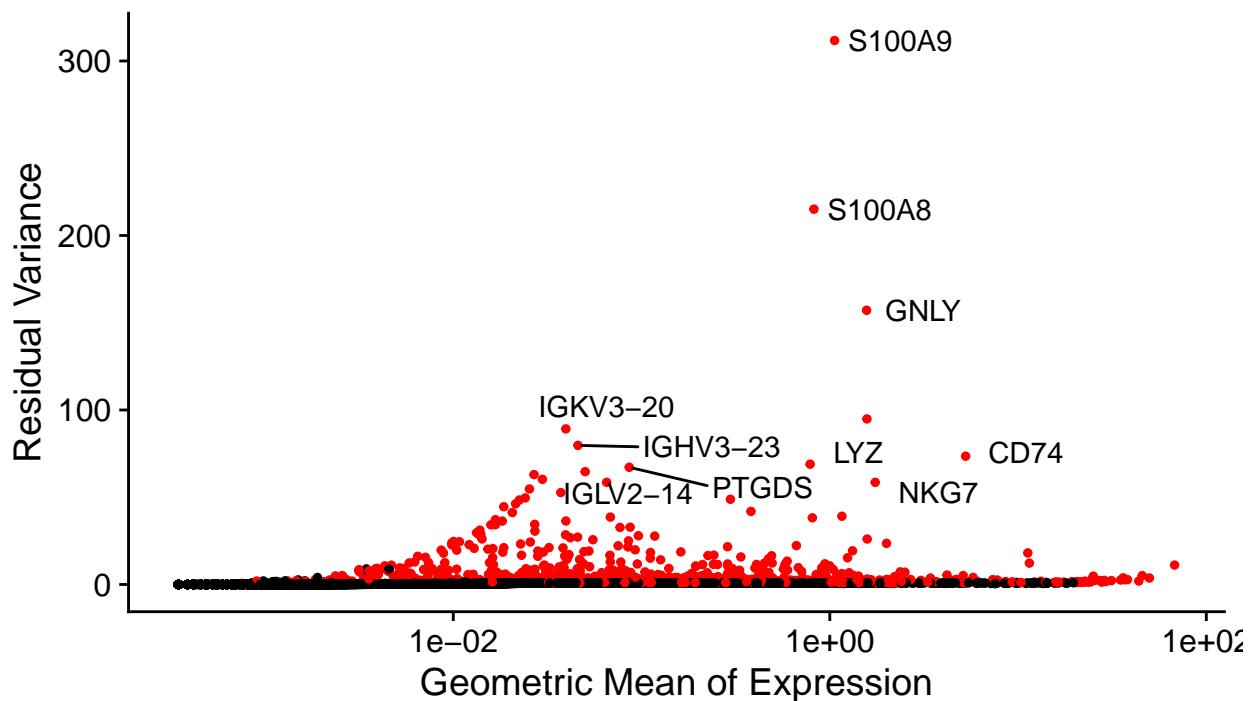
```
combo <- readRDS("./data/covid/24m_mapped_combo.rds")
```

Variable features

```
combo <- FindVariableFeatures(combo)
# Identify the 10 most highly variable genes
top10 <- head(VariableFeatures(combo), 10)
# Plot variable features
VariableFeaturePlot(combo, assay = "integrated") %>%
  LabelPoints(points = top10,
  repel = TRUE) +
  theme(legend.position = "bottom") +
  ggtitle("Variable features, 24 months")
```

## When using repel, set xnudge and ynudge to 0 for optimal results

## Variable features, 24 months



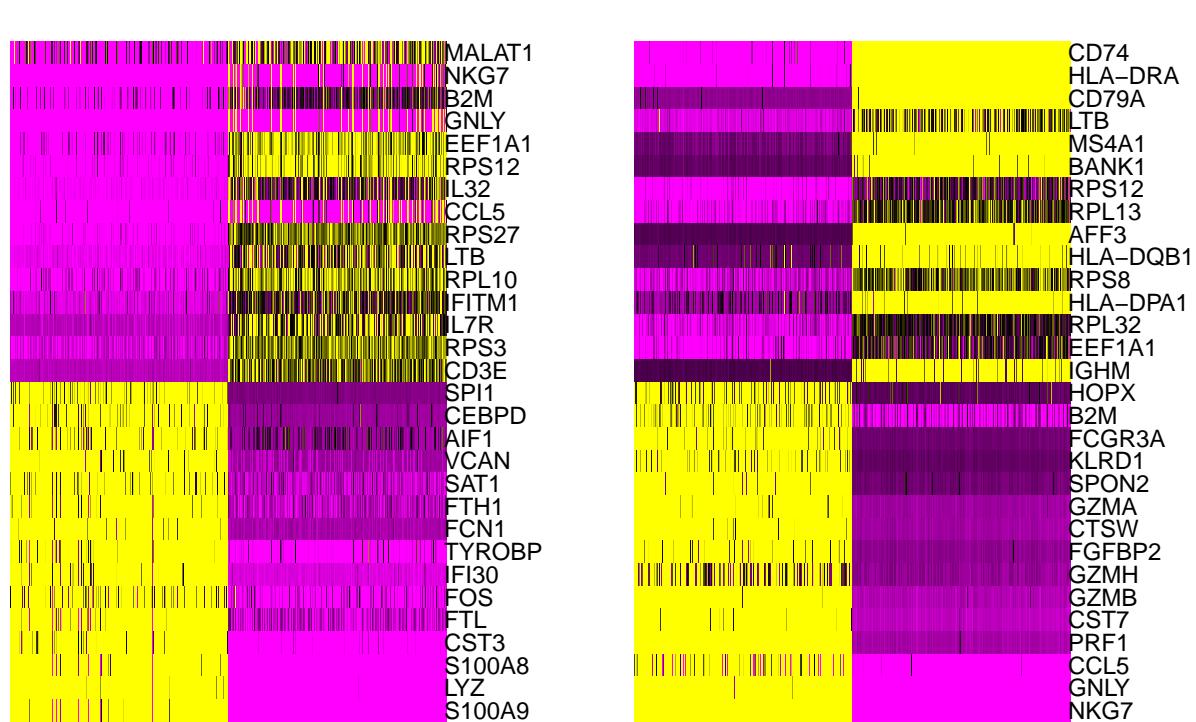
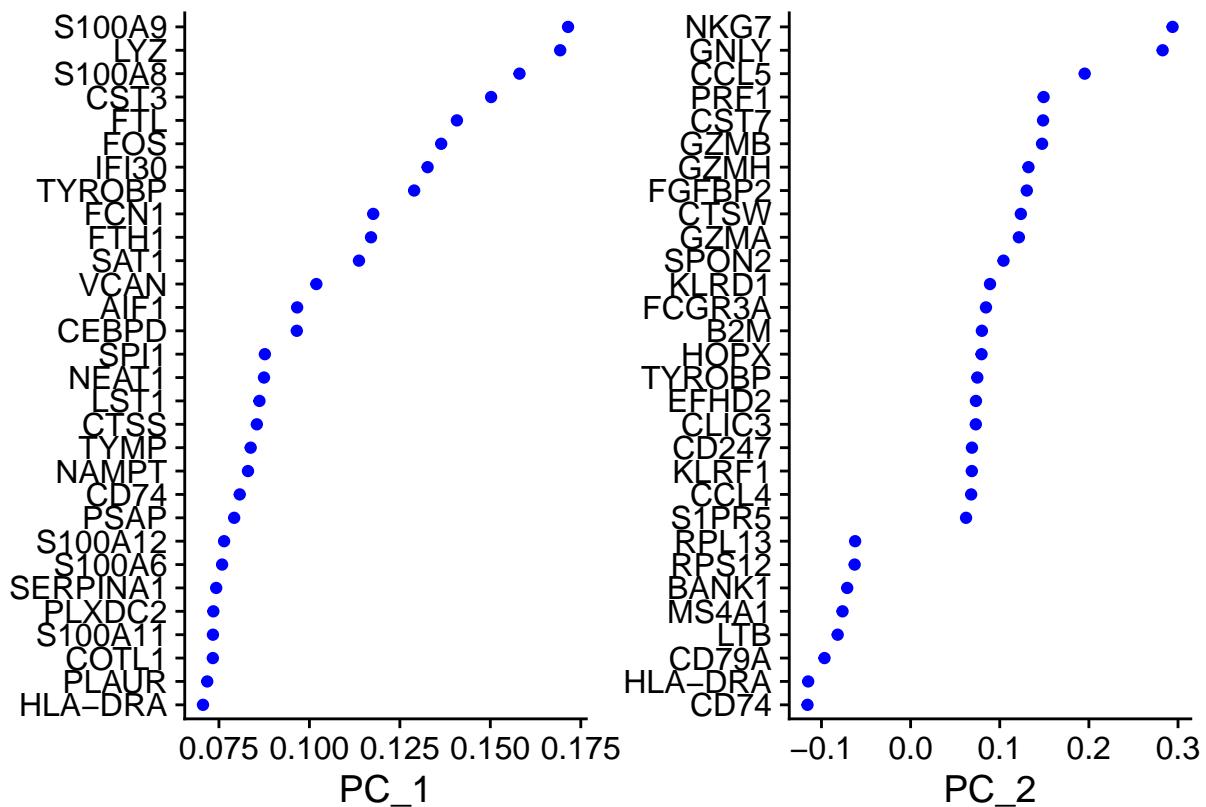
- Non-variable count: 17134
- Variable count: 3000

PCA

```
print(combo[["pca"]], dims = 1:5, nfeatures = 5)

## PC_ 1
## Positive: S100A9, LYZ, S100A8, CST3, FTL
## Negative: MALAT1, NKG7, B2M, GNLY, EEF1A1
## PC_ 2
## Positive: NKG7, GNLY, CCL5, PRF1, CST7
## Negative: CD74, HLA-DRA, CD79A, LTB, MS4A1
## PC_ 3
## Positive: IL7R, RPS12, S100A8, S100A9, TCF7
## Negative: CD74, HLA-DRA, CD79A, MS4A1, HLA-DPA1
## PC_ 4
## Positive: NRGN, PPBP, TUBB1, CAVIN2, PF4
## Negative: NKG7, GNLY, S100A8, S100A9, MALAT1
## PC_ 5
## Positive: CDKN1C, FCGR3A, IFITM3, FCER1G, AIF1
## Negative: S100A8, S100A9, VCAN, LYZ, S100A12

VizDimLoadings(combo, dims = 1:2, reduction = "pca")
```



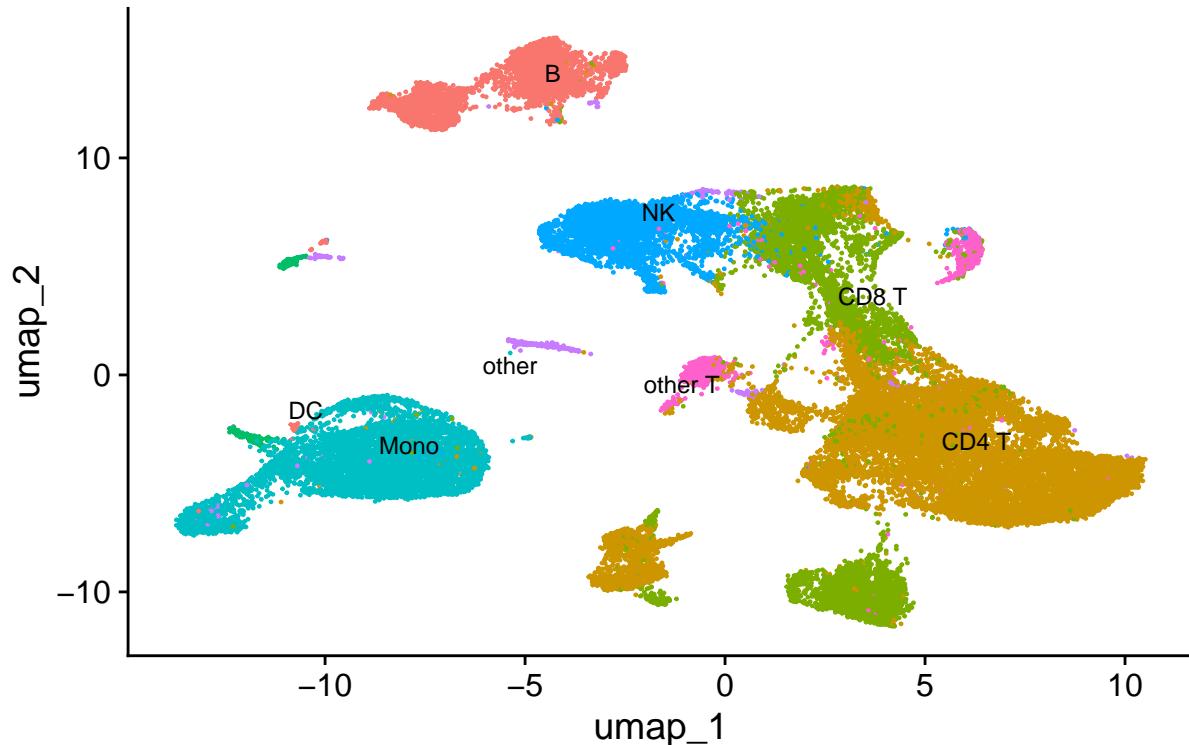
UMAP

```

DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.11",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 1, 24 months")

```

## Cell classification level 1, 24 months

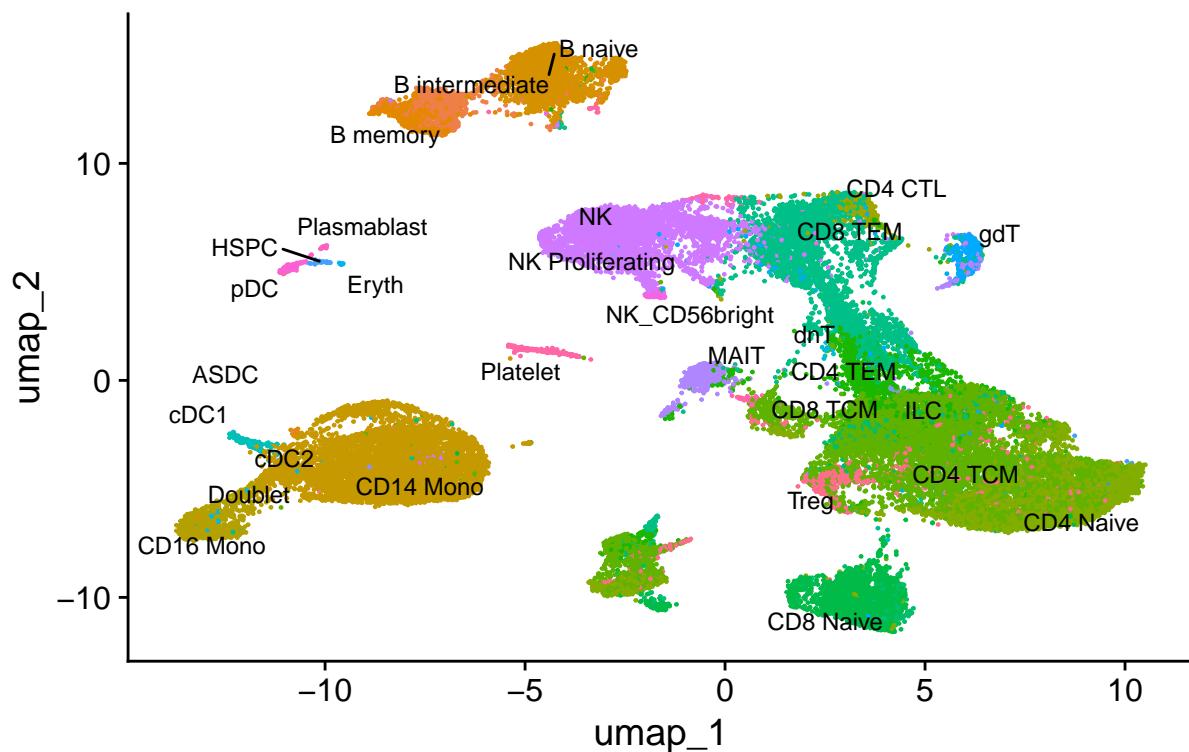


```

DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.12",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 2, 24 months")

```

## Cell classification level 2, 24 months



```
DimPlot(combo,
        reduction = "umap",
        group.by = "predicted.celltype.13",
        label = TRUE,
        label.size = 3,
        repel = TRUE) +
NoLegend() +
ggtitle("Cell classification level 3, 24 months")
```

## Cell classification level 3, 24 months

