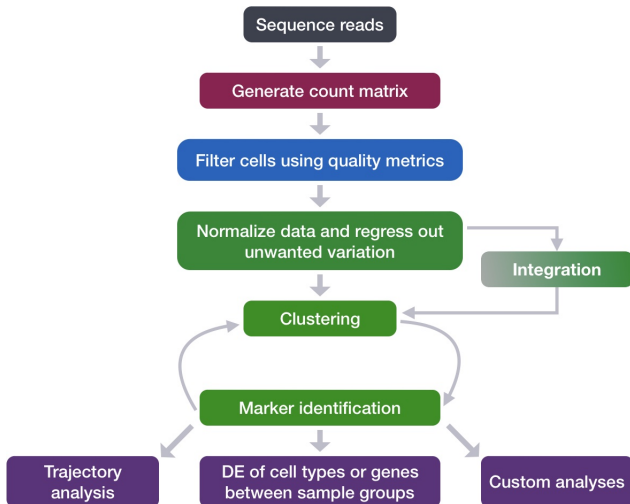


STAT588/BIOL588: Genomic Data Science

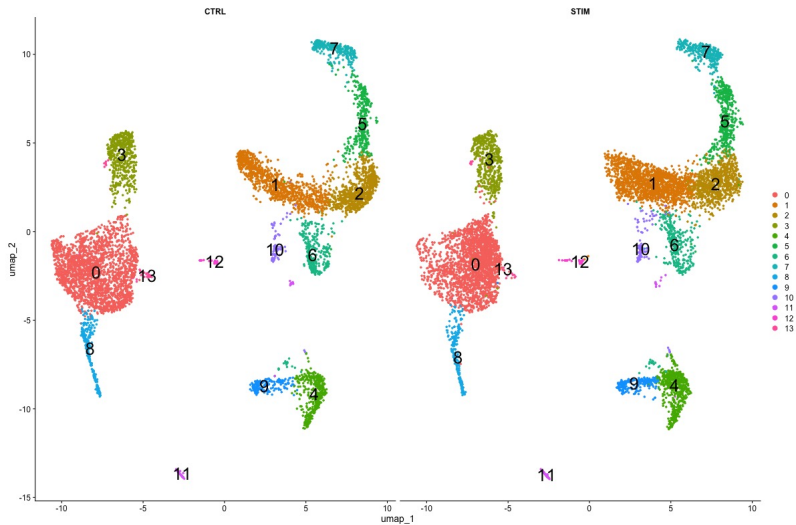
Lecture 23: Single-cell RNA-seq Marker Identification

Yen-Yi Ho (hoyen@stat.sc.edu)

scRNAeq Workflow



Clustering

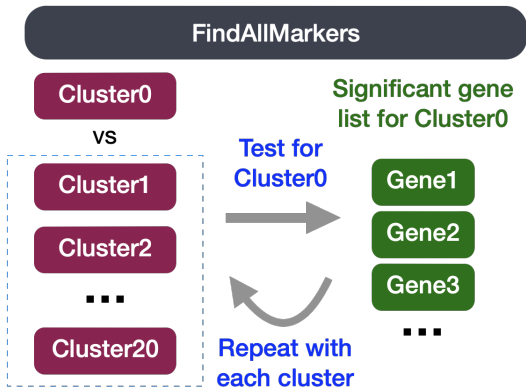


Marker Identification

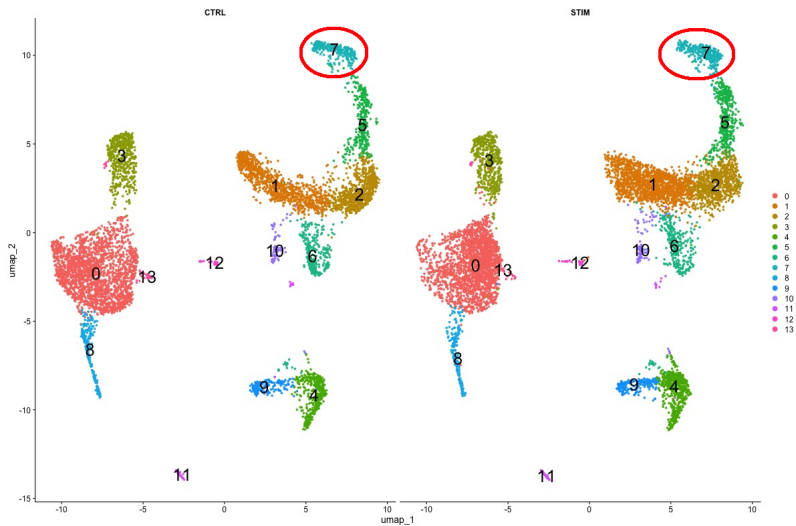
Goal

- ▶ To **determine the gene markers** for each of the clusters
- ▶ To **identify cell types** of each cluster using markers
- ▶ **Re-cluster based on cell type markers**

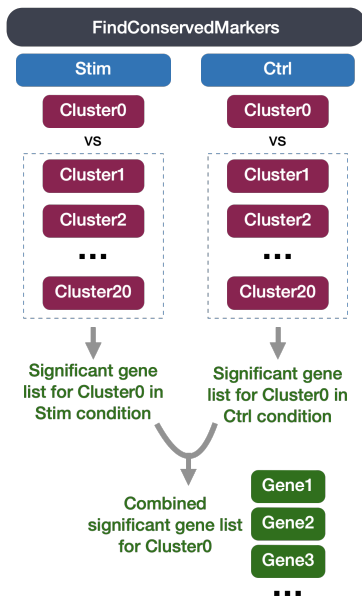
Identifying ALL Markers



Clustering



Identifying Conserved Markers



Identifying Conserved Markers

```
DefaultAssay(immune.combined) <- "RNA"  
immune <- JoinLayers(immune.combined)  
cluster7_conserved_markers <- FindConservedMarkers(immune,  
  ident.1 = 7,  
  grouping.var = "stim",  
  only.pos = TRUE,  
  logfc.threshold = 0.25)
```

- ▶ ident.1: the cluster of interest
- ▶ grouping.var: the "group" variable

Conserved Markers: Control

	CTRL_p_val	CTRL_avg_log2FC	CTRL_pct.1	CTRL_pct.2
GNLY	0	6.854586	0.943	0.046
NKG7	0	5.358880	0.953	0.085
GZMB	0	5.078135	0.839	0.044
CLIC3	0	5.765314	0.601	0.024
CTSW	0	5.307246	0.537	0.030
KLRD1	0	5.261553	0.507	0.019

- ▶ `condition_avg_logFC`: average log fold change for condition. Positive values indicate that the gene is more highly expressed in the cluster.
- ▶ `condition_pct.1`: percentage of cells where the gene is detected in the cluster for condition
- ▶ `condition_pct.2`: percentage of cells where the gene is detected on average in the other clusters for condition

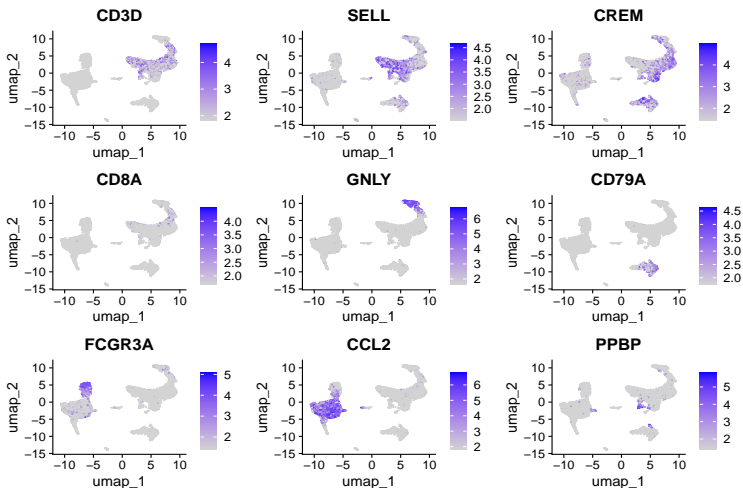
Look for markers with large difference in expression between **pct.1** and **pct.2**.

Conserved Markers: Treatment

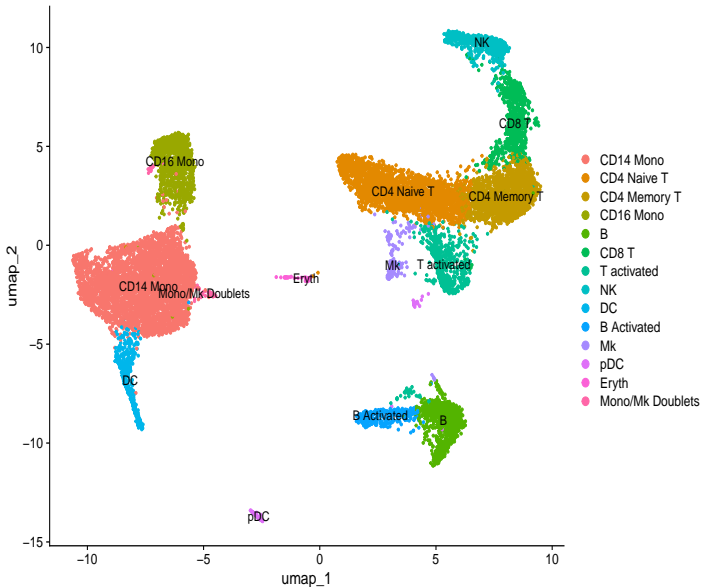
	STIM_p_val	STIM_avg_log2FC	STIM_pct.1	STIM_pct.2
GNLY	0	6.479248	0.954	0.059
NKG7	0	4.977410	0.954	0.081
GZMB	0	5.151799	0.895	0.060
CLIC3	0	5.510236	0.623	0.031
CTSW	0	5.268607	0.593	0.035
KLRD1	0	4.884551	0.556	0.027

Visualizing Marker Genes

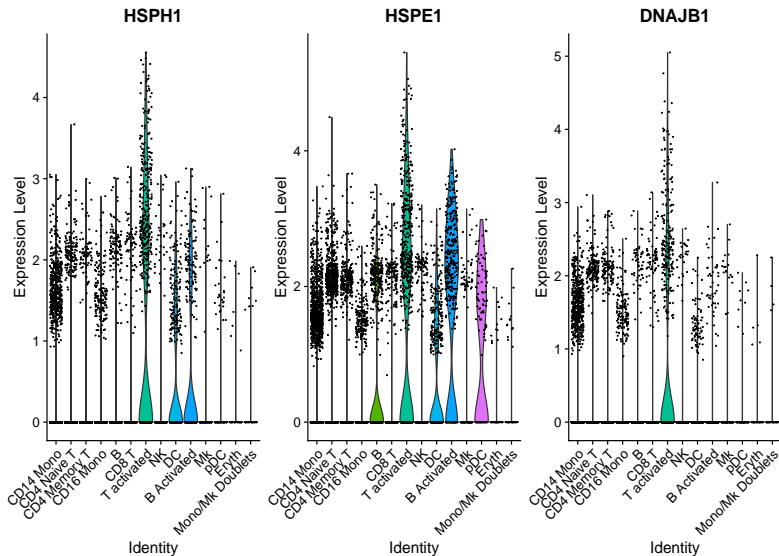
The **FeaturePlot()** function help to explore the expression of different identified markers by clusters.



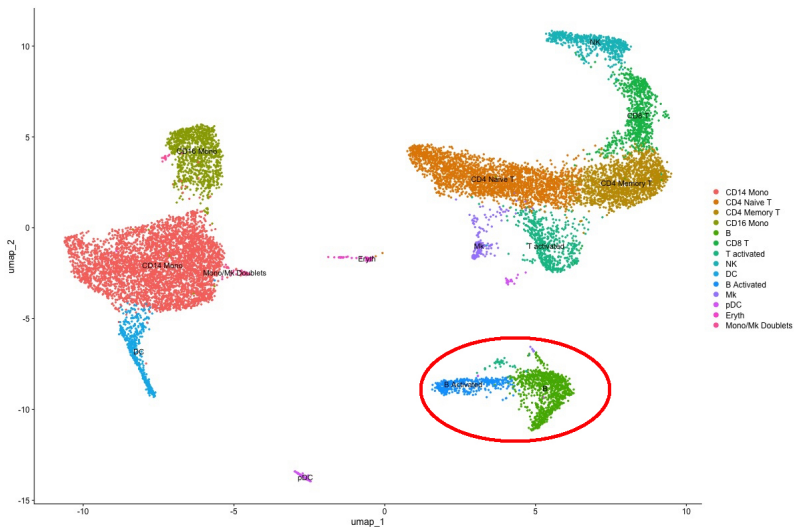
Labeling the Clustering with Markers



Violin Plots



Labeling the Clustering with Markers



Identifying DE Genes Across Conditions

##		p_val	avg_log2FC	pct.1	pct.2	p_val_adj
##	ISG15	9.550061e-159	5.0448219	0.998	0.238	1.342070e-154
##	IFIT3	2.610512e-154	6.1031136	0.965	0.051	3.668553e-150
##	IFI6	2.789158e-152	5.5112449	0.965	0.076	3.919603e-148
##	ISG20	1.131068e-149	3.0557389	1.000	0.667	1.589489e-145
##	IFIT1	5.106433e-140	6.1904950	0.910	0.032	7.176070e-136
##	MX1	4.941048e-123	3.9818856	0.907	0.115	6.943654e-119
##	LY6E	8.660219e-119	3.7675611	0.894	0.152	1.217021e-114
##	TNFSF10	1.050320e-111	6.4123168	0.787	0.022	1.476015e-107
##	IFIT2	5.443946e-108	5.5539615	0.785	0.037	7.650377e-104
##	B2M	1.720154e-97	0.6085489	1.000	1.000	2.417332e-93

DE Genes Feature Plots

