

CAMML Quick Vignette

Courtney Schiebout, H. Robert Frost

11/23/2021

Load Libraries

Libraries “CAMML” (Schiebout and Frost 2022) and “Seurat” (Satija et al. 2015) need to be loaded to carry out this vignette, in addition to several other libraries for data processing and gene set development (Satija et al. 2015; Robinson, McCarthy, and Smyth 2010; Carlson 2020; Liberzon et al. 2011). Packages will also load additional libraries they depend on.

```
library(CAMML)
library(Seurat)
library(SeuratObject)
library(edgeR)
library(org.Hs.eg.db)
library(msigdb)
```

Get Gene Set

Cell type gene sets can be loaded with the `GetGeneSet` function. In this case, we will load “immune.cells” which calls data for 5 immune cell types: T cells, B cells, NK cells, Monocytes, and Hematopoietic Stem Cells (HSCs).

```
gene.set.df <- GetGeneSets(data = "immune.cells")
```

Load Data

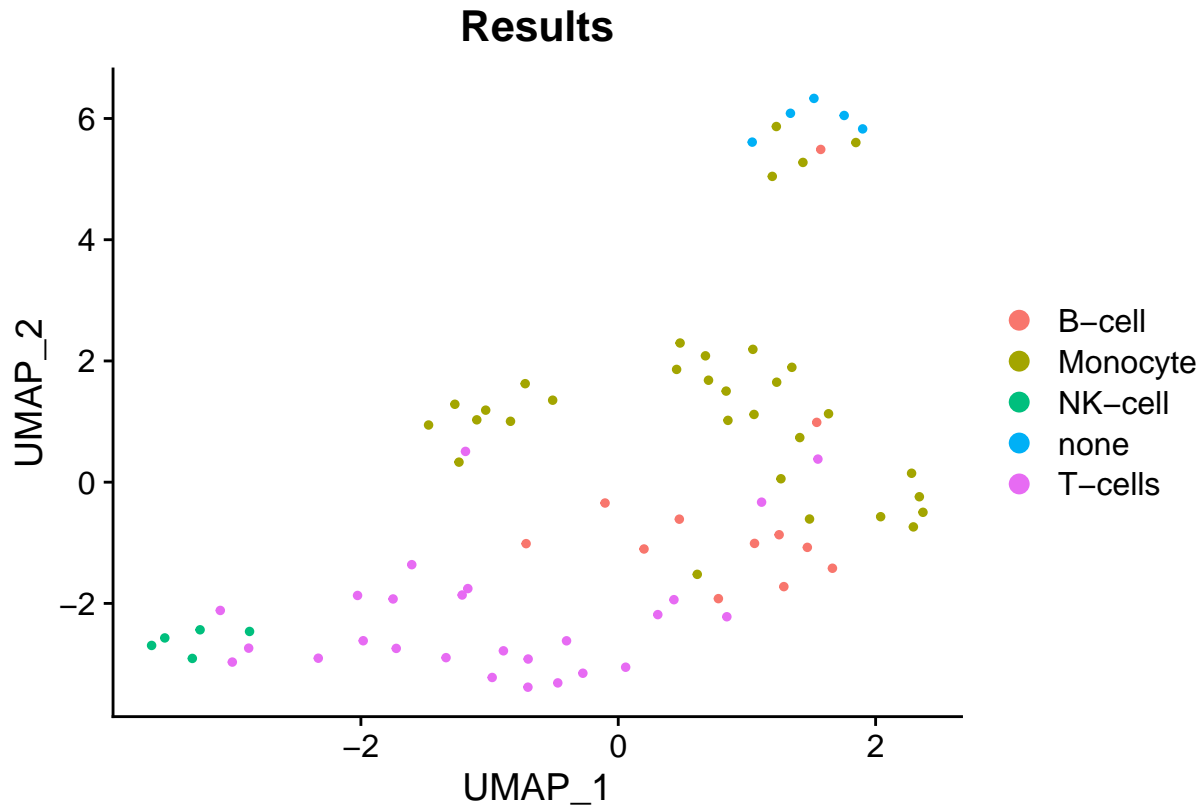
For this quick example, we will use “pbmc_small” from Seurat, which will provide a Seurat Object of 80 peripheral blood mononuclear cells (Satija et al. 2015).

```
seurat <- SeuratObject::pbmc_small
seurat <- RunPCA(seurat)
seurat <- RunUMAP(seurat, dims = 1:10)
```

Run CAMML

Once a gene set data frame and Seurat Object are defined, CAMML can simply be run by inputting both in the CAMML function. Labels can be defined for each cell using `GetCAMMLLabels` and designating the preferred label types.

```
seurat <- CAMML(seurat, gene.set.df)
results <- GetCAMMLLabels(seurat, labels = "top1")
seurat$Results <- unlist(results)
UMAPPlot(object = seurat, group = "Results")
```



References

Carlson, Marc. 2020. *Org.Hs.eg.db: Genome Wide Annotation for Human*.

Liberzon, Arthur, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P. Mesirov. 2011. "Molecular Signatures Database (MSigDB) 3.0." *Bioinformatics* 27 (12): 1739–40. <https://doi.org/10.1093/bioinformatics/btr260>.

Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth. 2010. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1): 139–40. <https://doi.org/10.1093/bioinformatics/btp616>.

Satija, Rahul, Jeffrey A Farrell, David Gennert, Alexander F Schier, and Aviv Regev. 2015. "Spatial Reconstruction of Single-Cell Gene Expression Data." *Nature Biotechnology* 33 (5): 495–502. <https://doi.org/10.1038/nbt.3192>.

Schiebout, Courtney, and H. Robert Frost. 2022. "CAMML: Multi-Label Immune Cell-Typing and Stemness Analysis for Single-Cell RNA-Sequencing." *Pacific Symposium on Biocomputing*.