AML3

jtrincado

2022-02-08 11:30:14

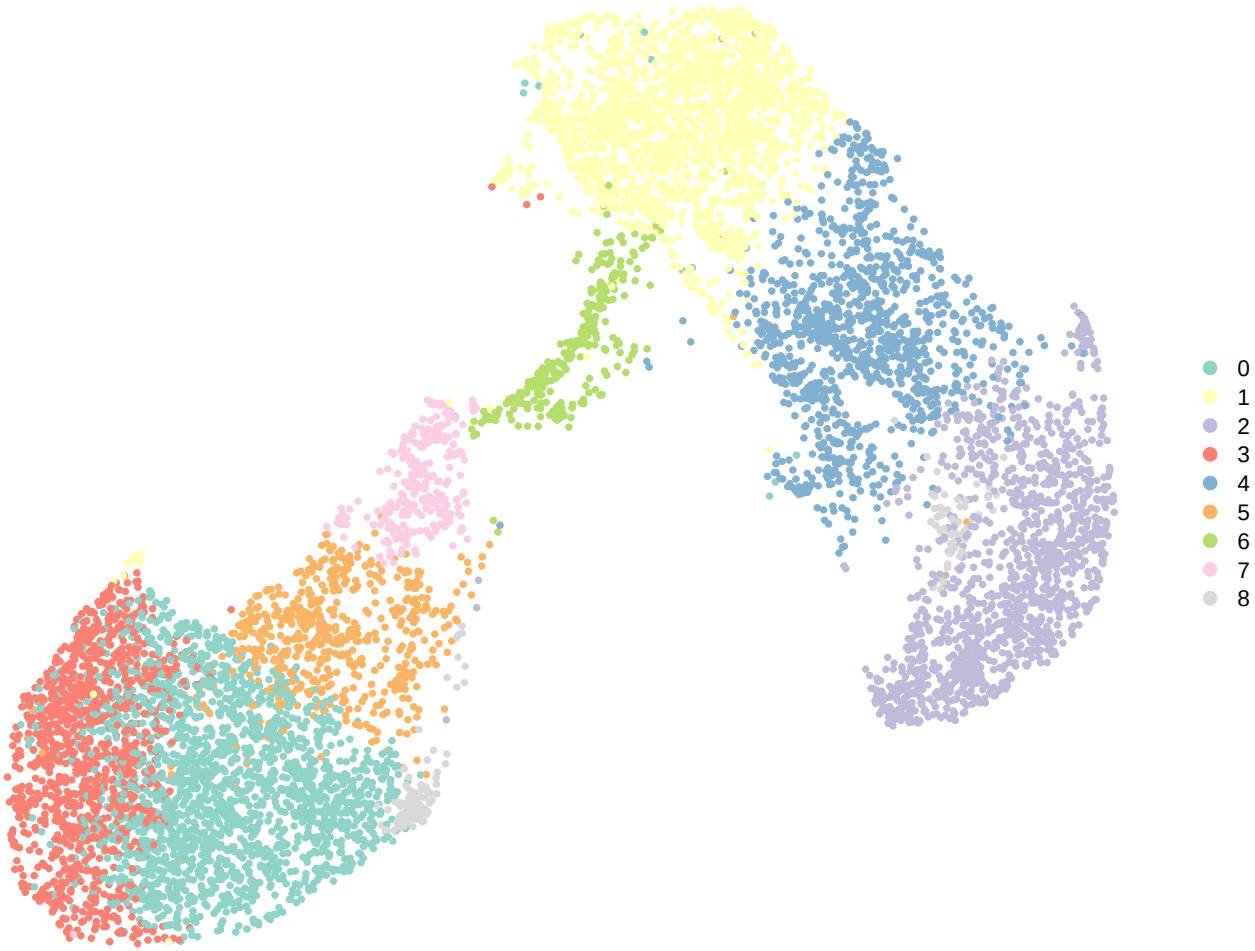
Contents

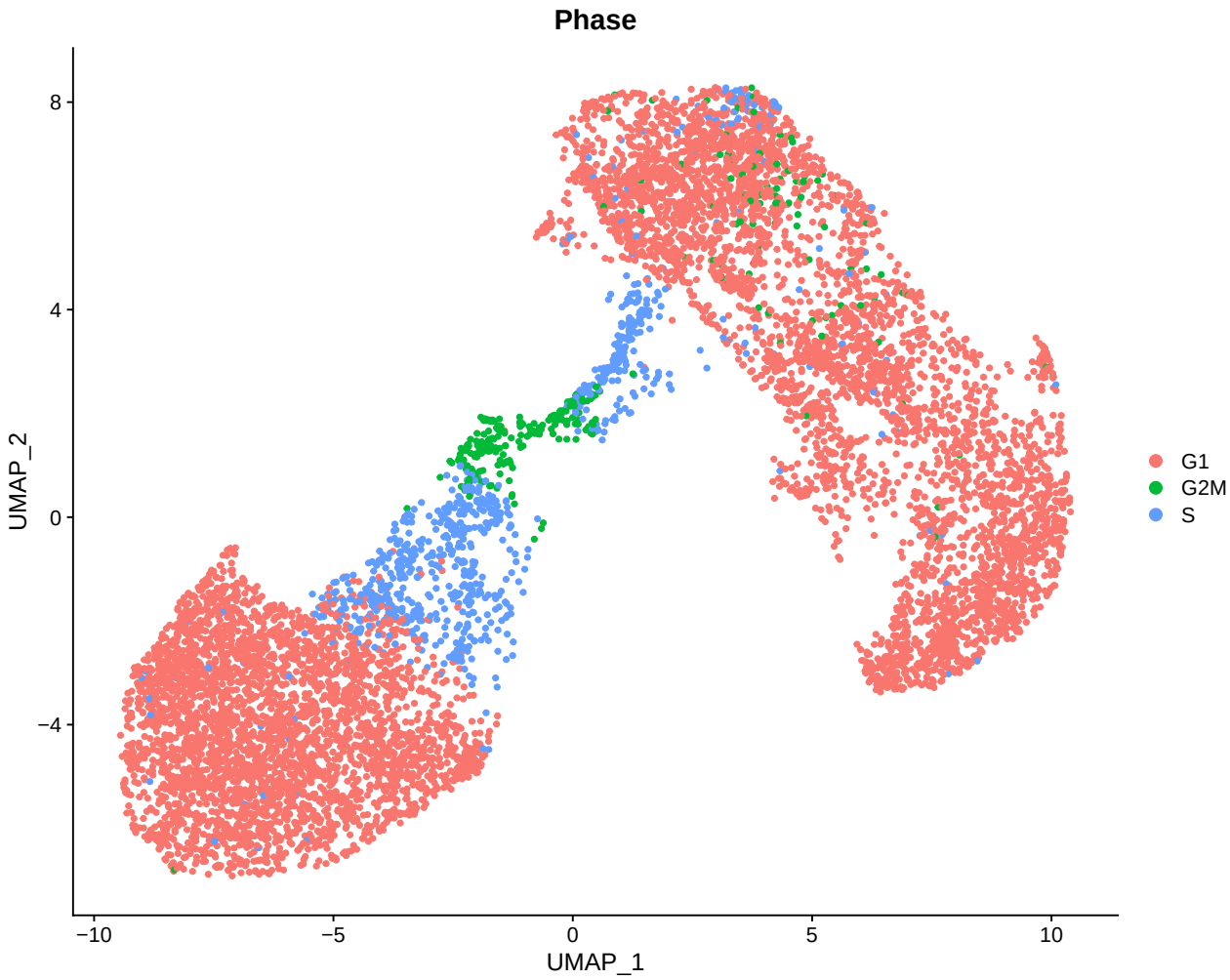
1. Put together both 34 and 38 libraries. Apply QC and dimensionality reduction.

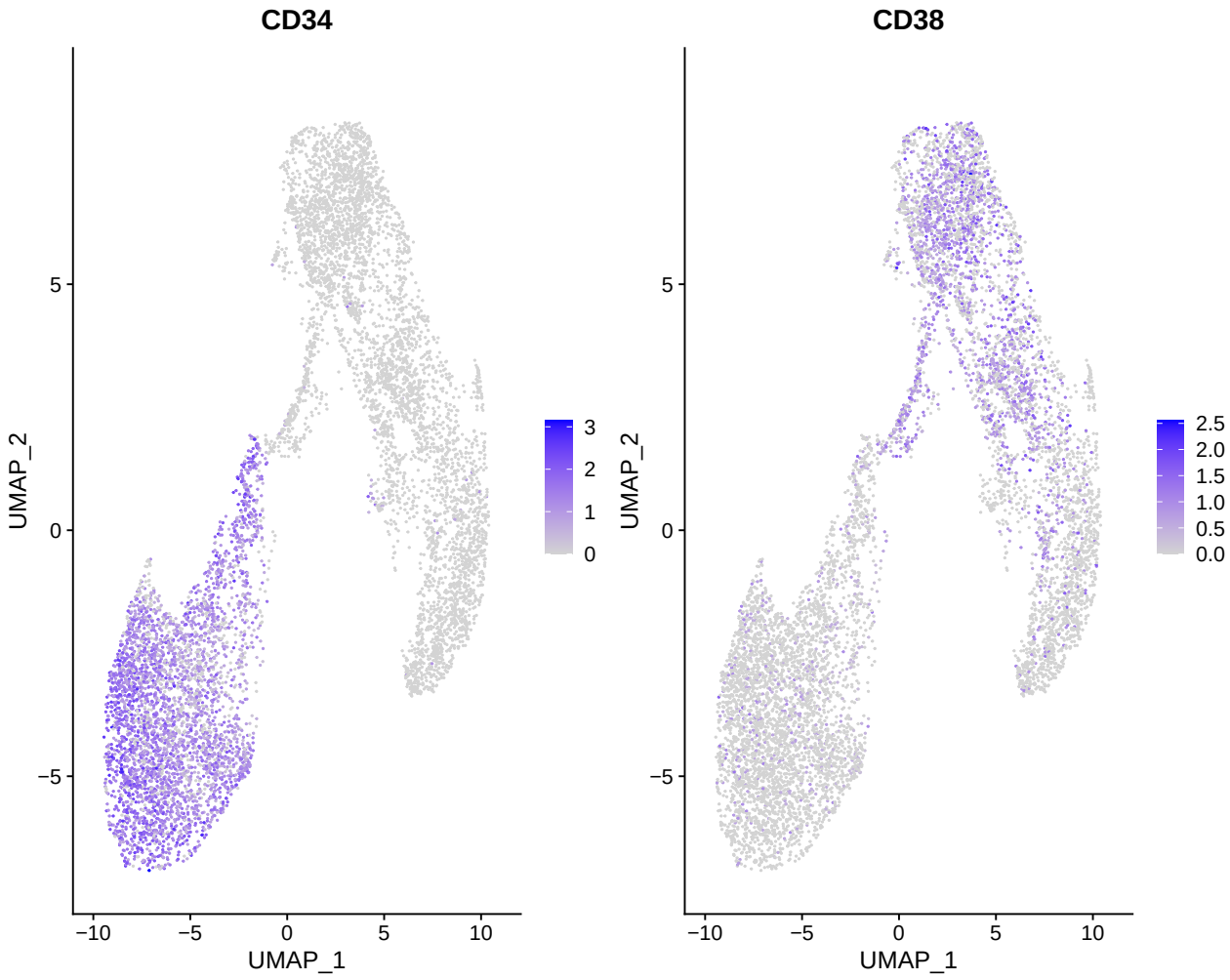
```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 8522
## Number of edges: 279946
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.8667
## Number of communities: 9
## Elapsed time: 1 seconds
```

```
## CD34_AAACCCAGTCCAATCA CD34_AAACCCAGTGGTATGG CD34_AAACCCATCCAGTGTA
##                     0                     0                     0
## CD34_AAACGAAAGCTGTGCC CD34_AAACGAAGTAAGGTCG
##                     0                     3
## Levels: 0 1 2 3 4 5 6 7 8
```

```
## Warning: The default method for RunUMAP has changed from calling Python UMAP via reticulate to the R
## To use Python UMAP via reticulate, set umap.method to 'umap-learn' and metric to 'correlation'
## This message will be shown once per session
```

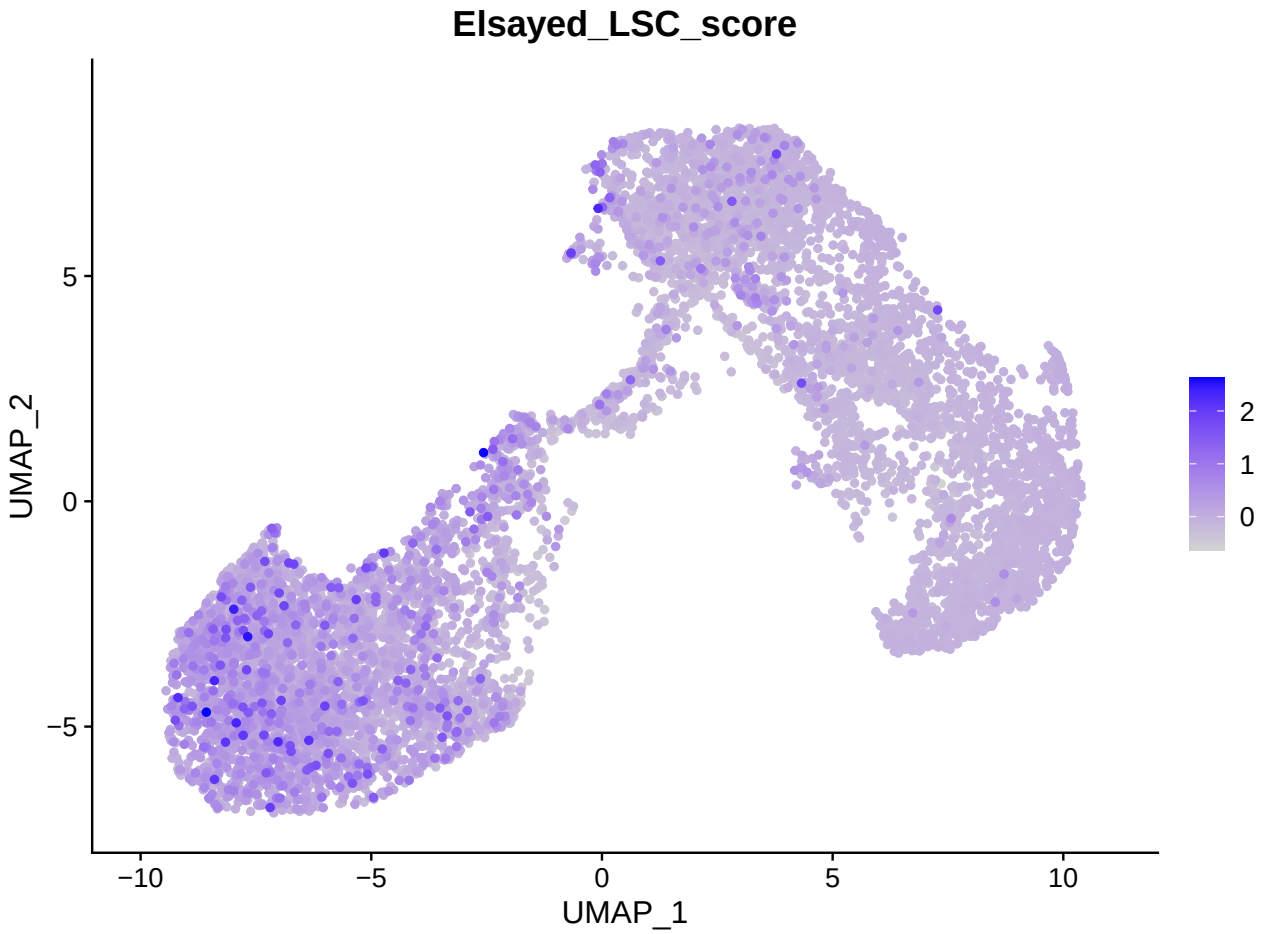

0
1
2
3
4
5
6
7
8

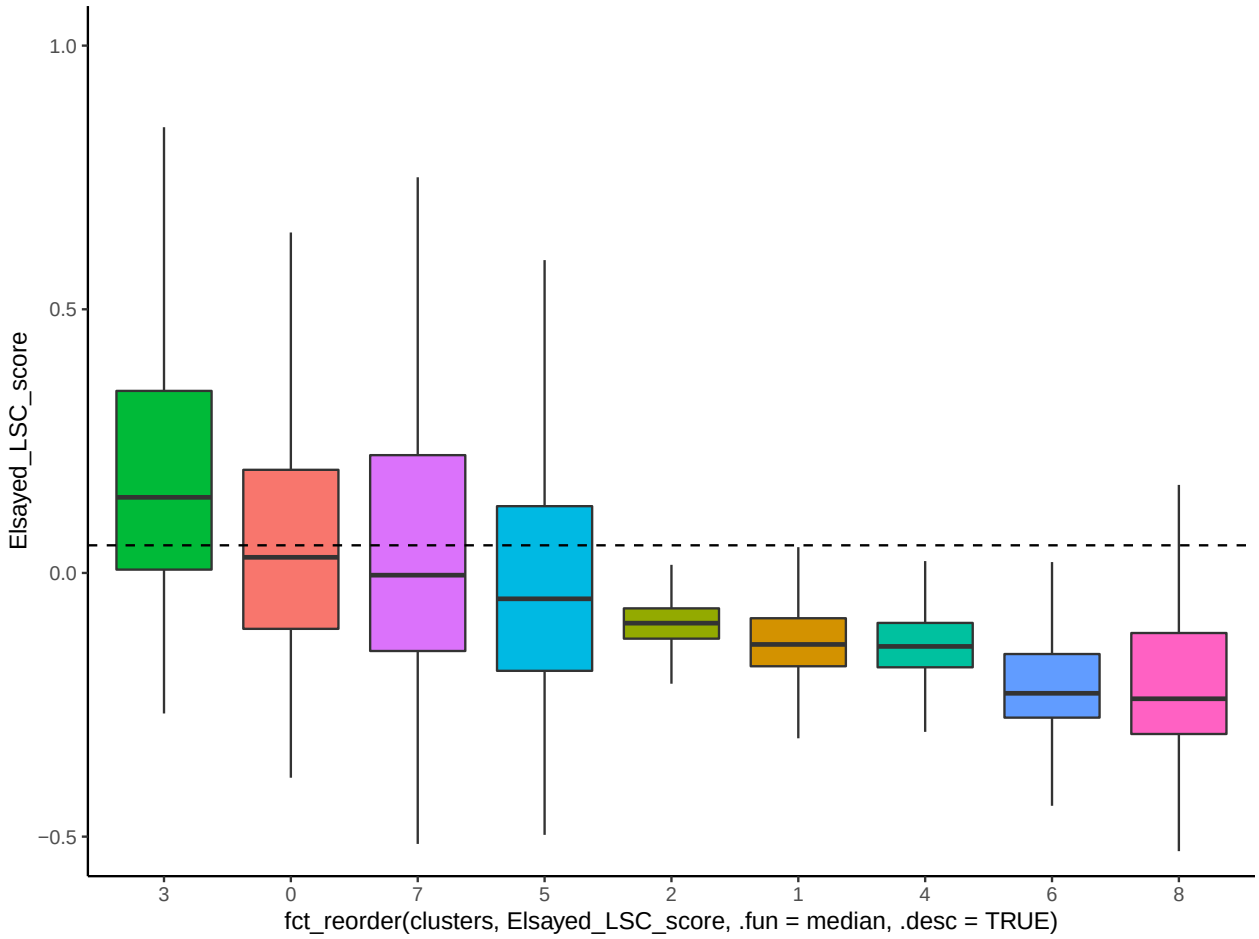

Phase
8
4
0
−4
UMAP_2
−10
−5
0
5
10
UMAP_1
G1
G2M
S



2. Get the LSC6 score

```
## [1] "CD34"   "SPINK2" "SOCS2"  "FAM30A" "ADGRG1" "DNMT3B"
```

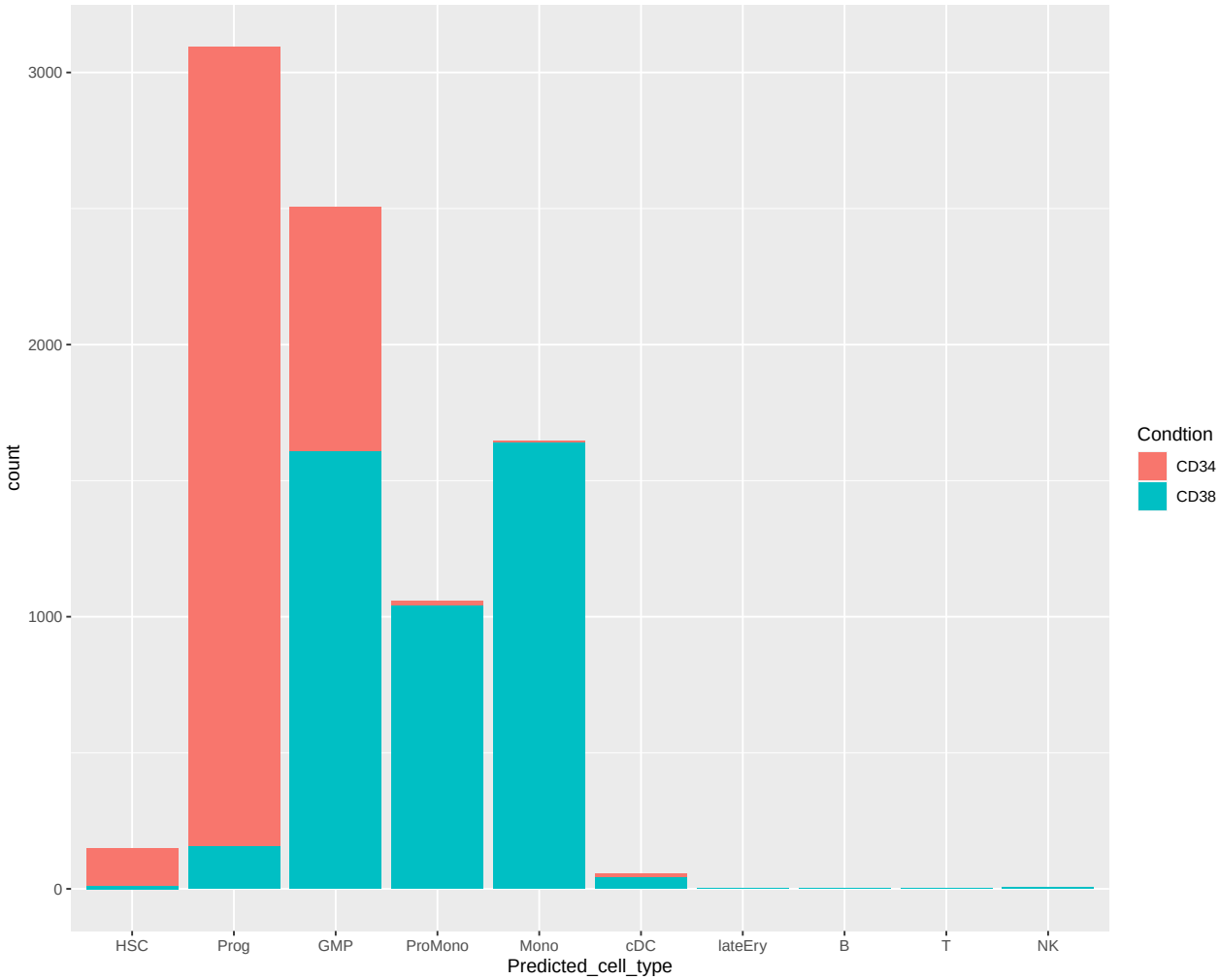

Elsayed_LSC_score
UMAP_2
5
0
−5
−10
−5
0
5
10
UMAP_1
2
1
0

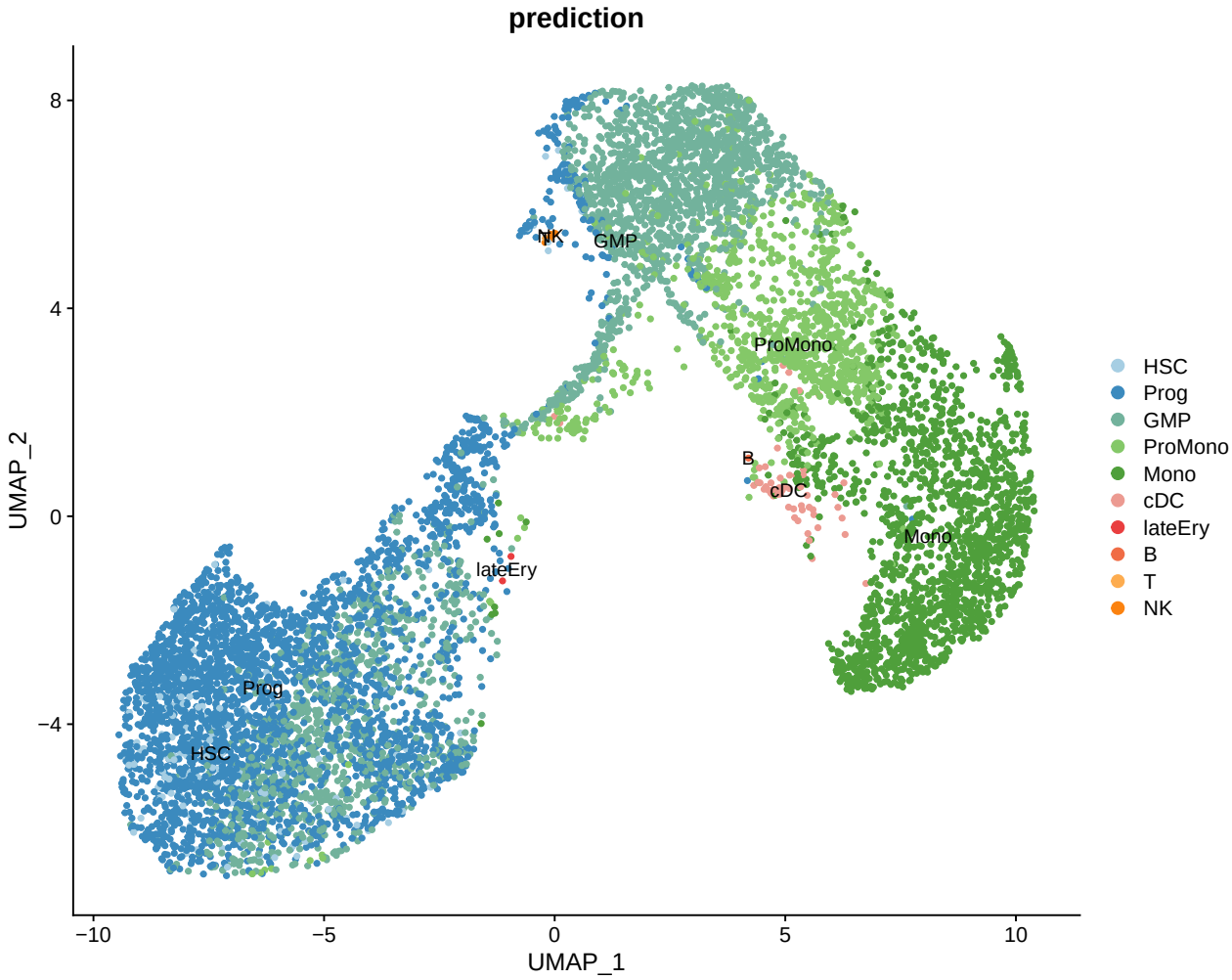


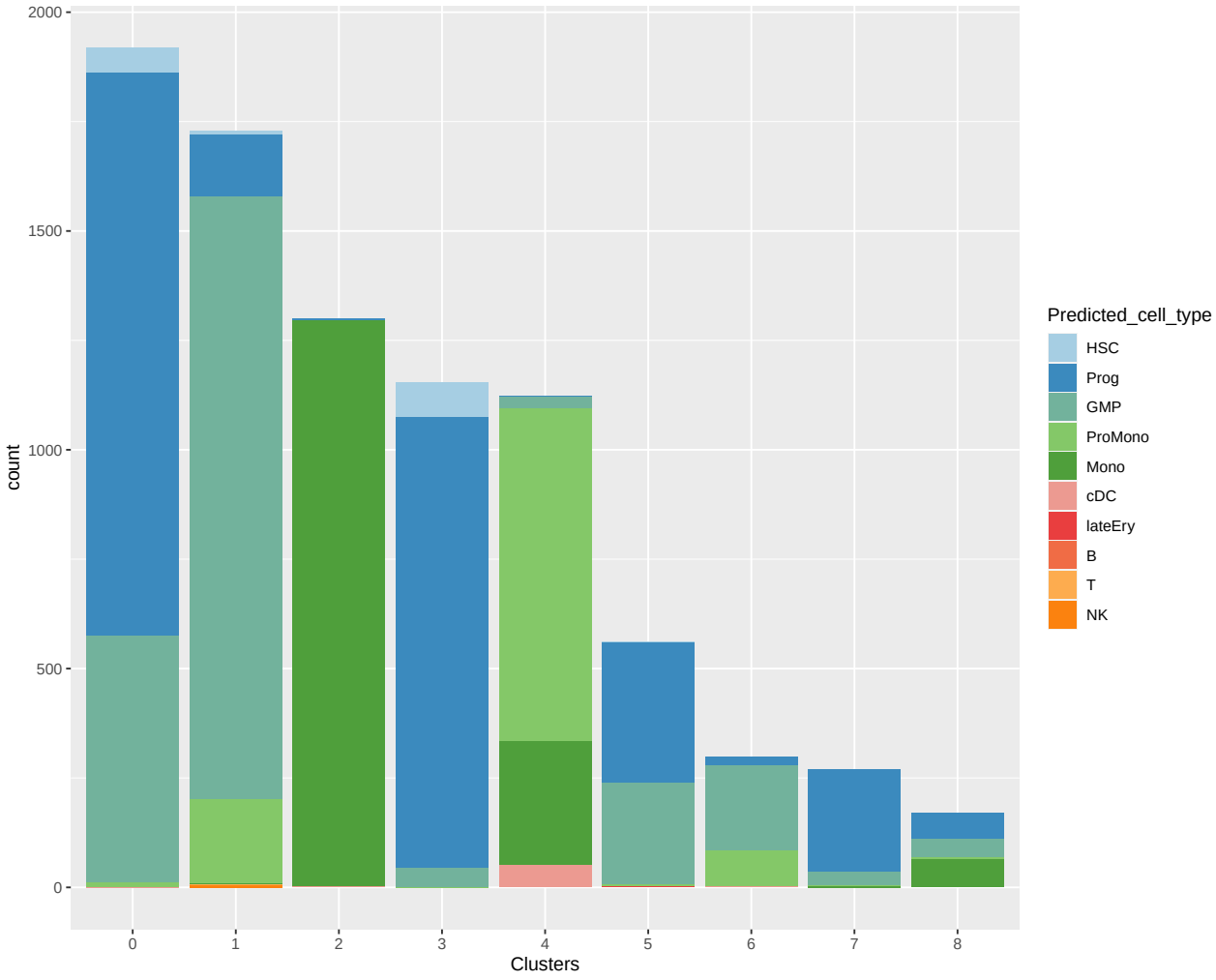
3. Predict the class of the cells using the markers and the expression of the BM cells form Van_Galen paper

```
## Performing PCA on the provided reference using 1821 features as input.
## Projecting cell embeddings
## Finding neighborhoods
## Finding anchors
##  Found 3059 anchors
## Filtering anchors
##  Retained 1475 anchors
## Finding integration vectors
## Finding integration vector weights
## Predicting cell labels
##
##         HSC Prog  GMP ProMono Mono  cDC  pDC earlyEry lateEry ProB    B Plasma
##   CD34  139 2941  901      19    7   15    0        0       0    0    0      0
##   CD38   11  154 1607    1039 1640   40    0        0       2    0    1      0
```

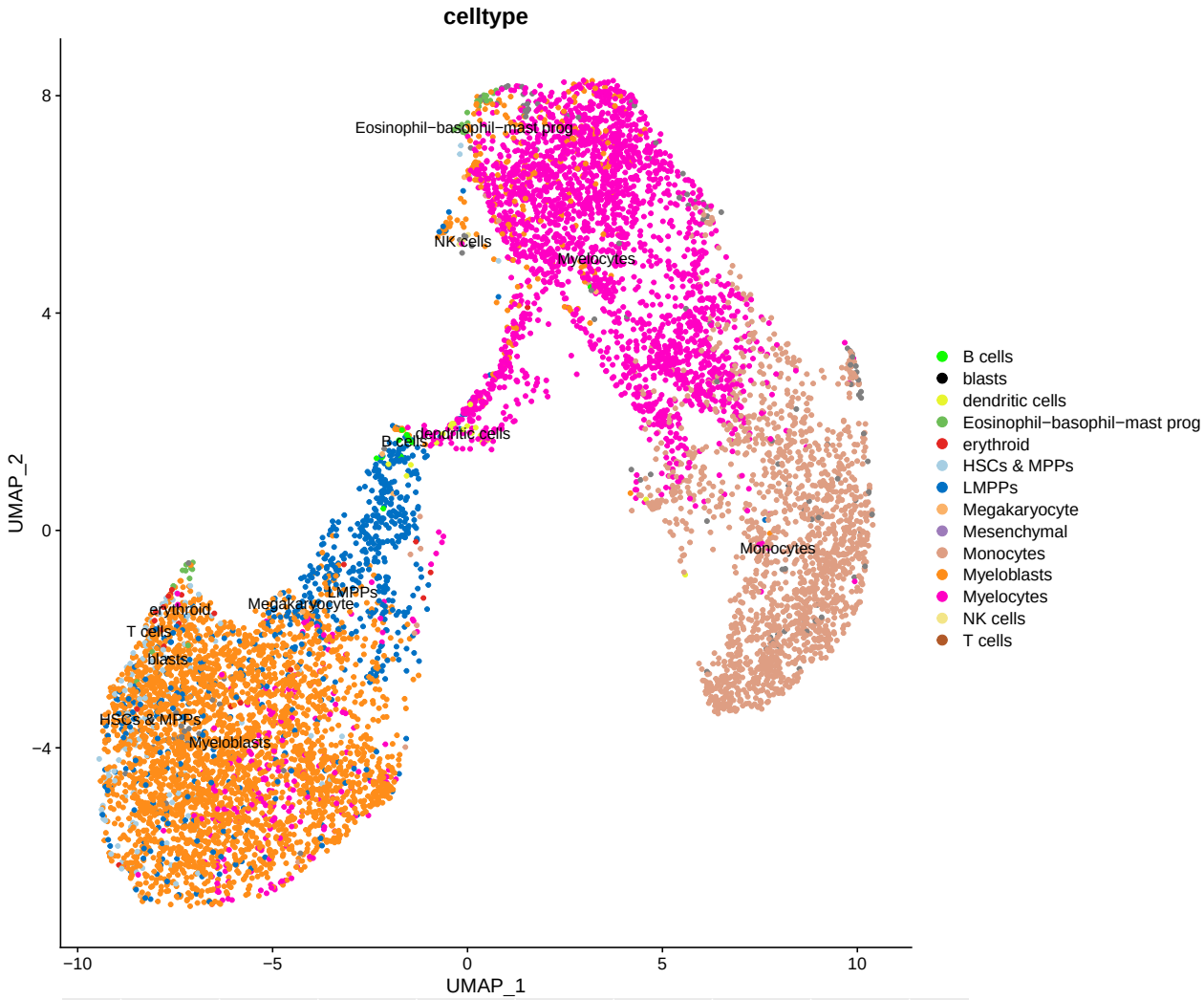
```
##
##              T  CTL   NK
##    CD34      0    0    0
##    CD38      1    0    5
```

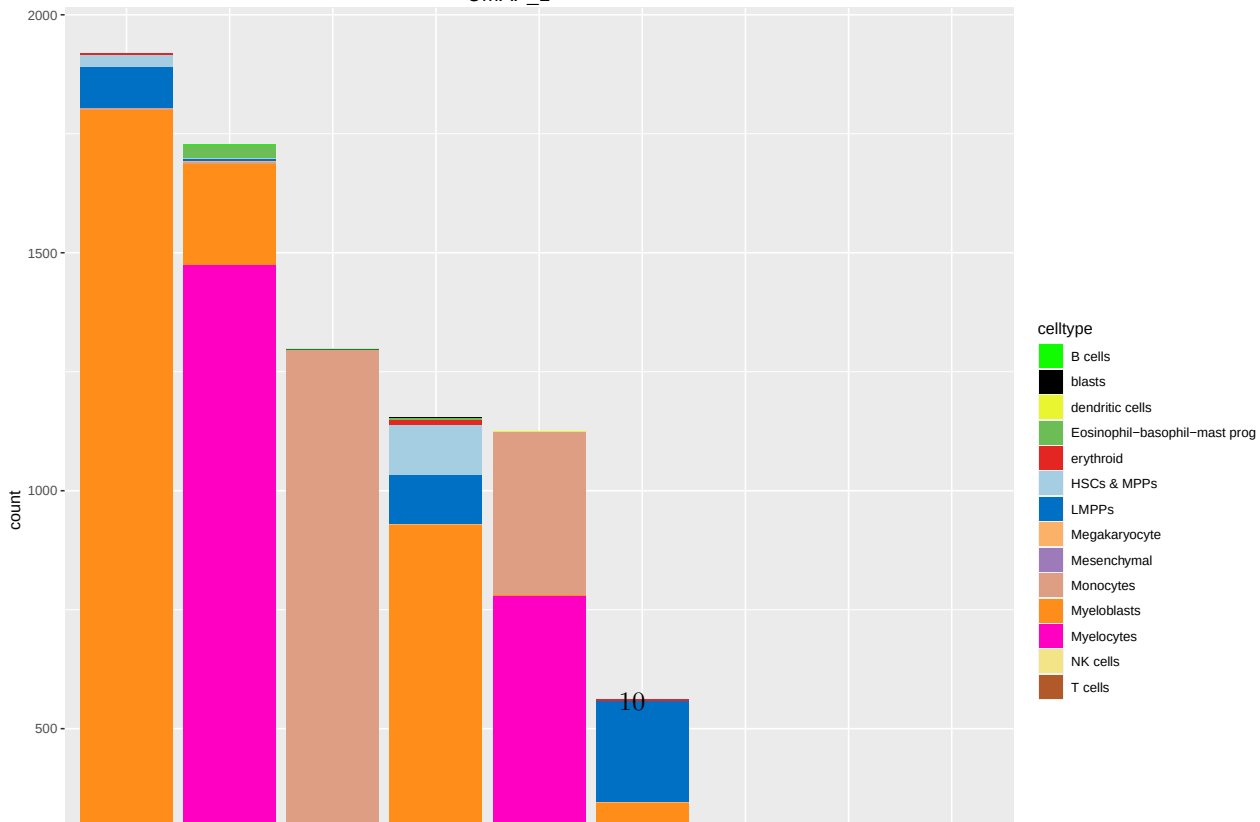



prediction
UMAP_2
UMAP_1
8
4
0
−4
−10
−5
0
5
10
NK
GMP
ProMono
B
cDC
Mono
lateEry
Prog
HSC
HSC
Prog
GMP
ProMono
Mono
cDC
lateEry
B
T
NK


2000
1500
1000
500
0
count
0
1
2
3
4
5
6
7
8
Clusters
Predicted_cell_type
HSC
Prog
GMP
ProMono
Mono
cDC
lateEry
B
T
NK

4. Project the predictions from Velten onto our UMAP





Cluster 0 seems the most likely to be enriched in LSC