AML5

jtrincado

2022-02-08 11:28:36

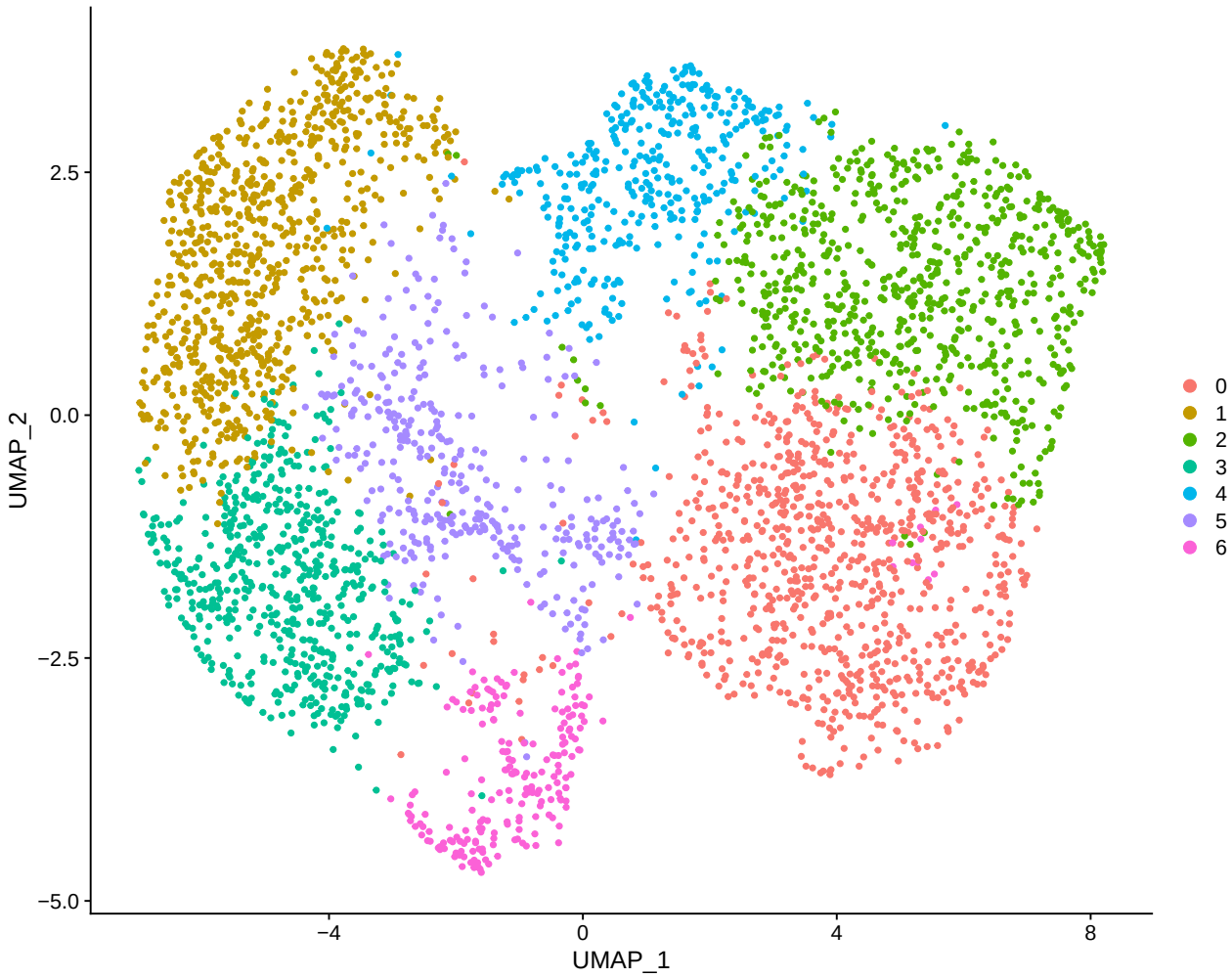
Contents

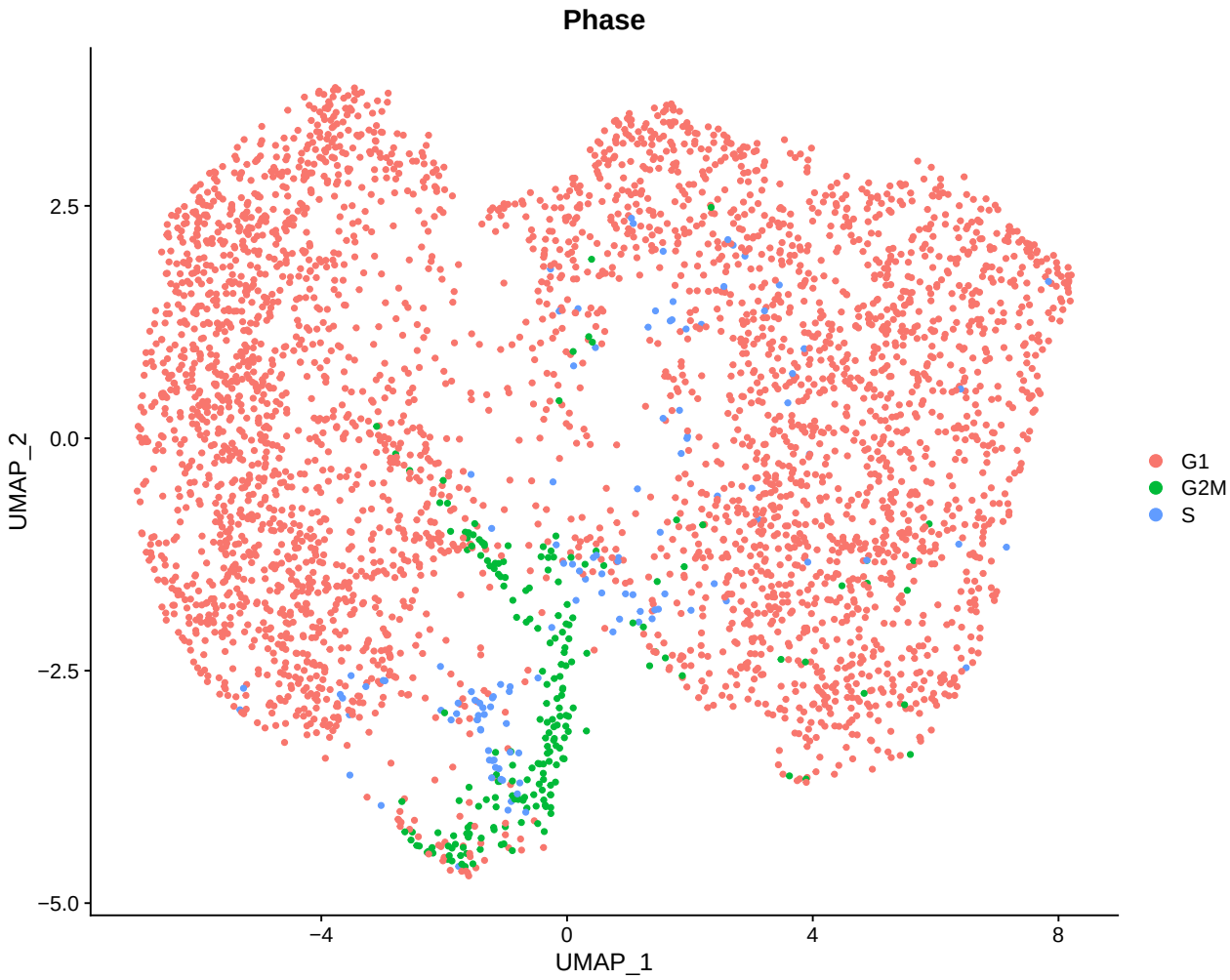
1. Put together both 34 and 38 libraries. Apply QC and dimensionality reduction.

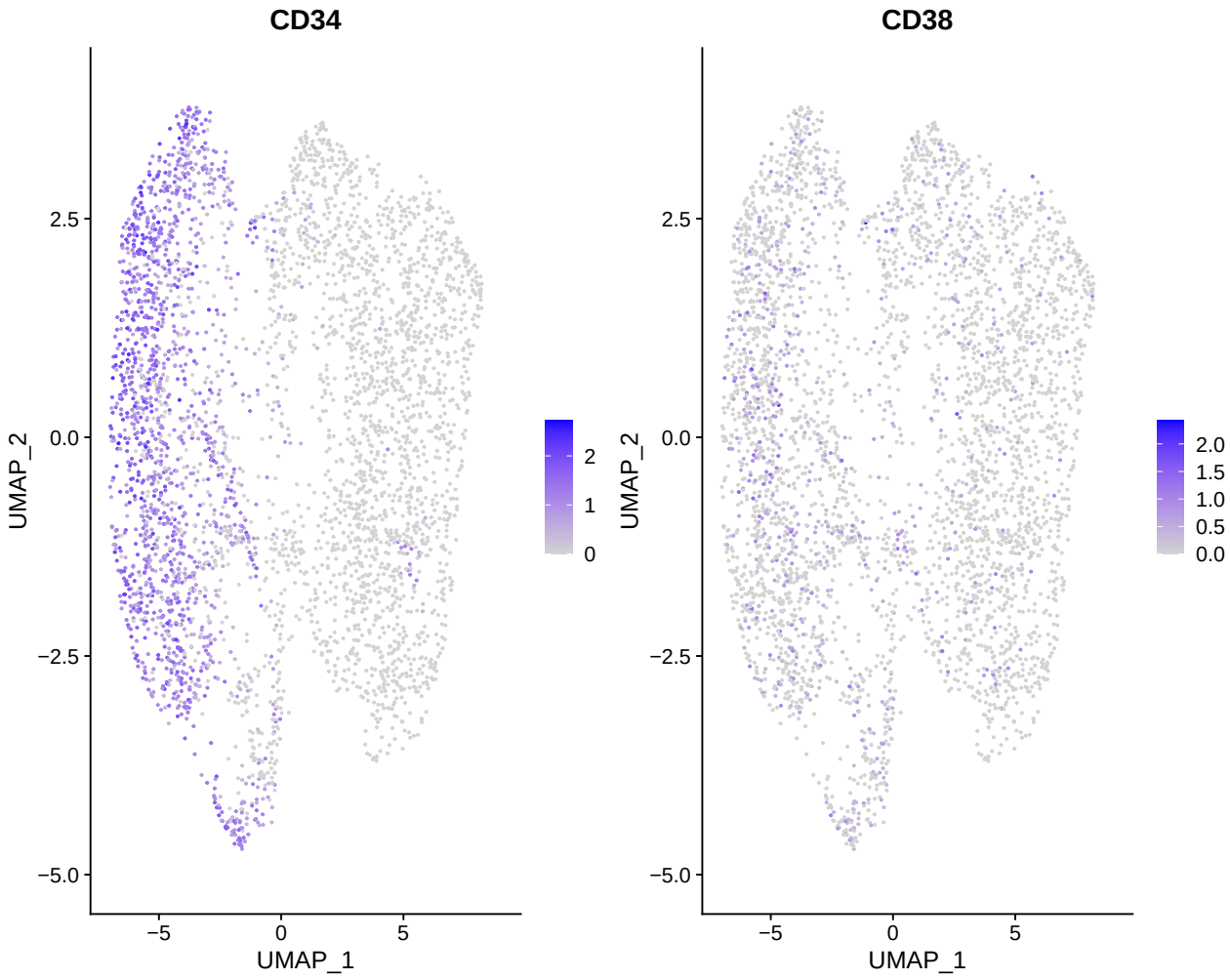
```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 3979
## Number of edges: 132147
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.8250
## Number of communities: 7
## Elapsed time: 0 seconds
```

```
##
##          0   1   2   3   4   5   6
##   CD34  38 824  35 574 207 287 135
##   CD38 835   4 713   3 174  69  81
```

```
## 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
```

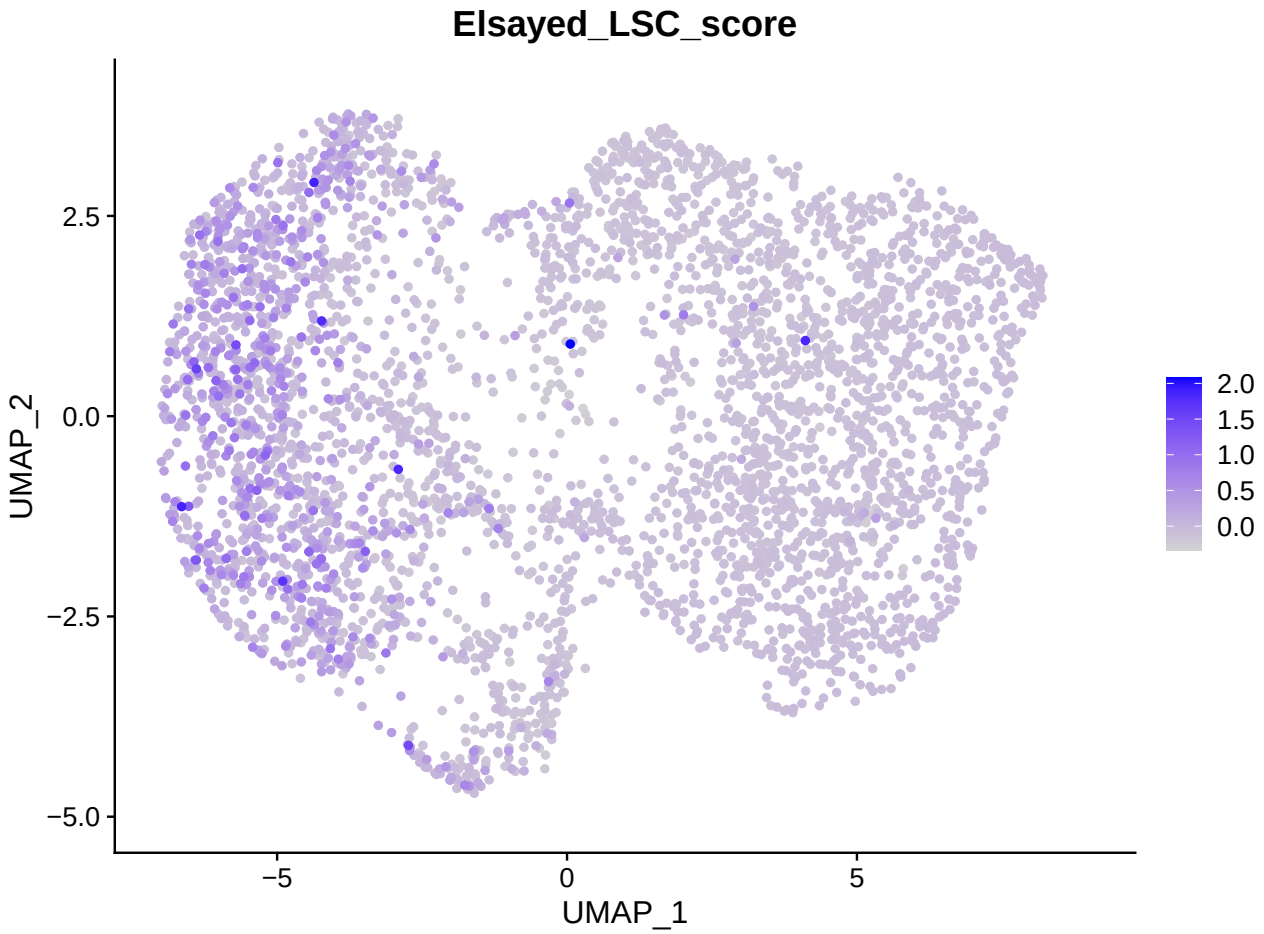

2.5
0.0
−2.5
−5.0
UMAP_2
−4
0
4
8
UMAP_1
0
1
2
3
4
5
6

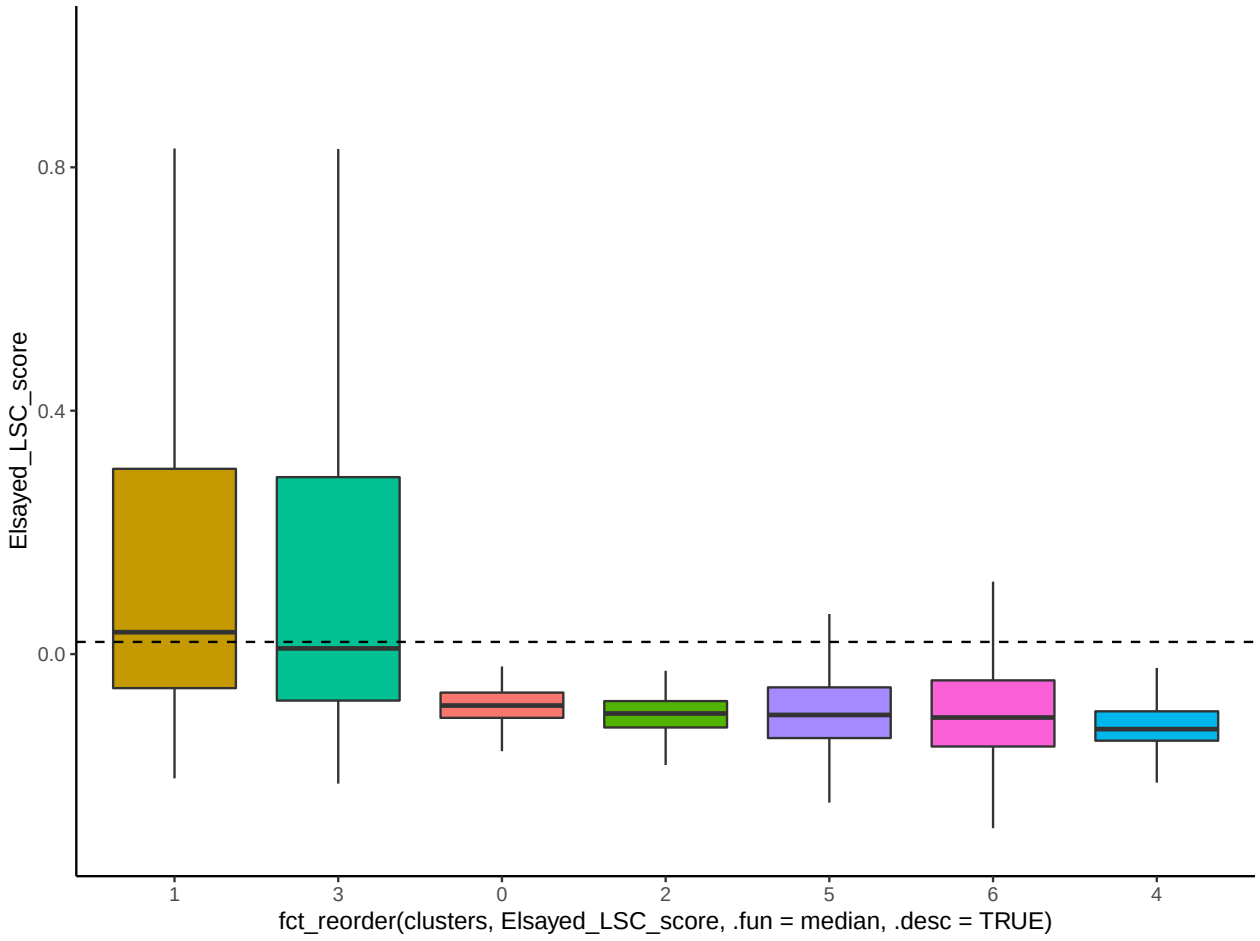

Phase
2.5
0.0
−2.5
−5.0
UMAP_2
−4
0
4
8
UMAP_1
G1
G2M
S



2. Get the LSC6 score

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

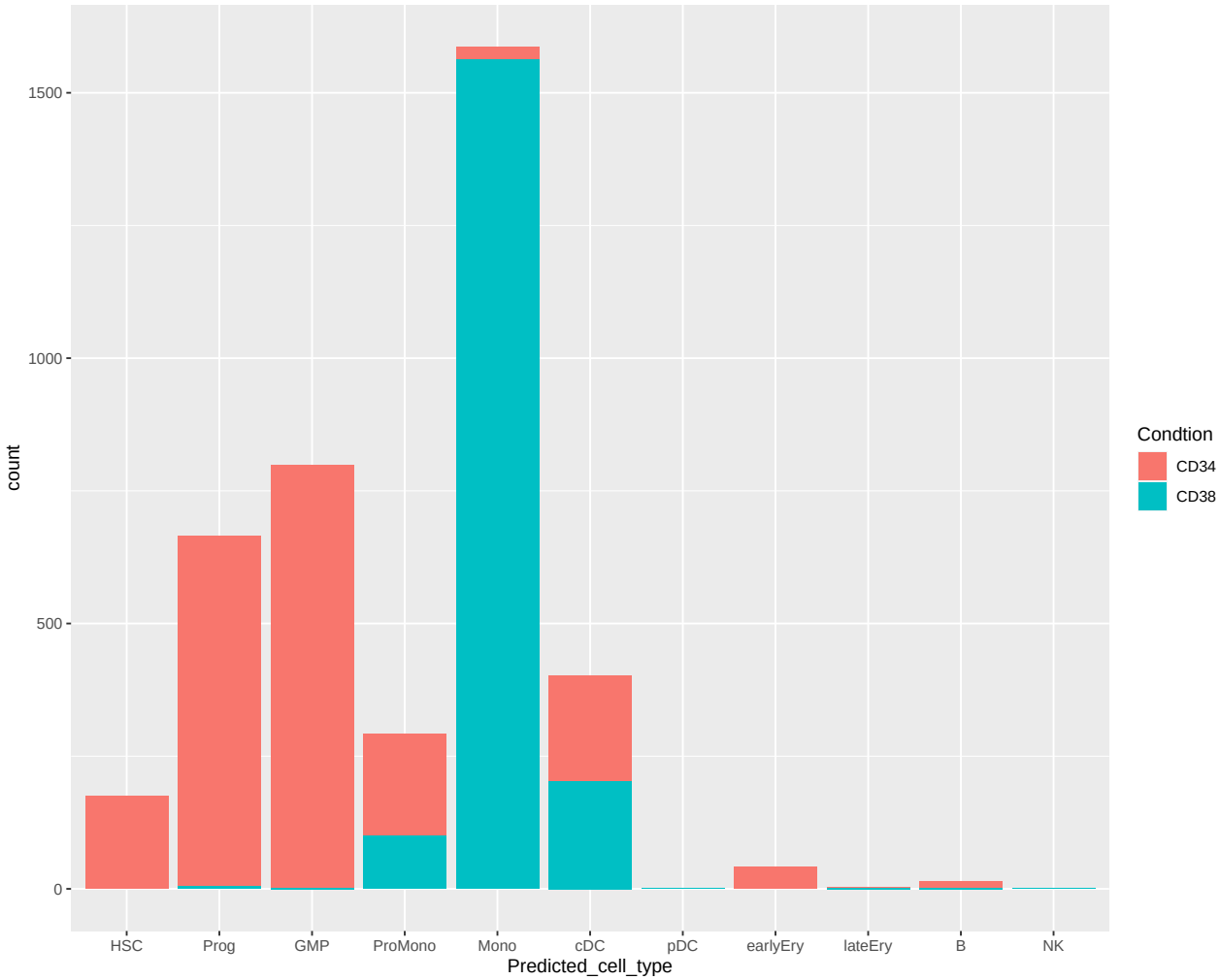

Elsayed_LSC_score
2.5
0.0
−2.5
−5.0
UMAP_2
−5
0
5
UMAP_1
2.0
1.5
1.0
0.5
0.0

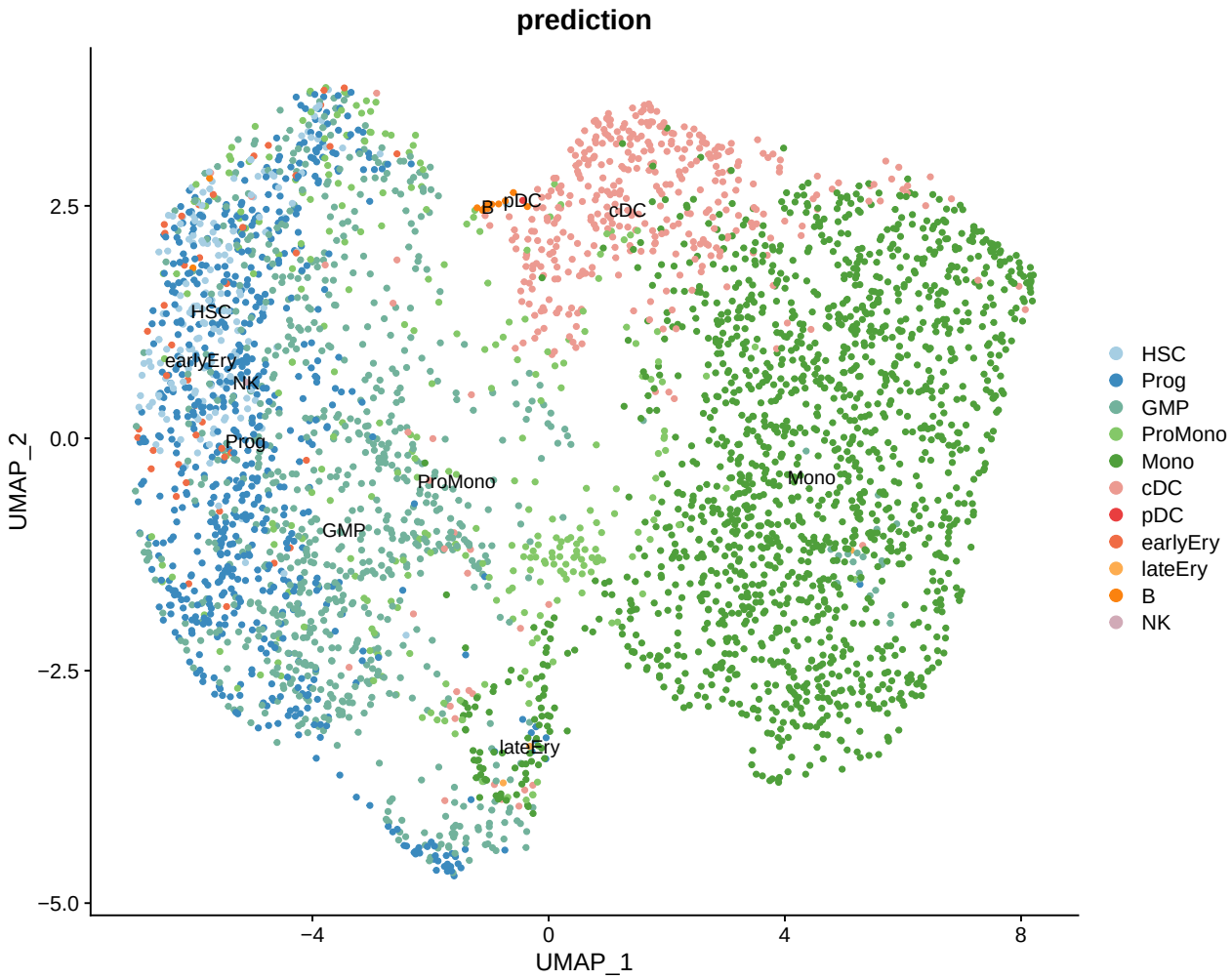


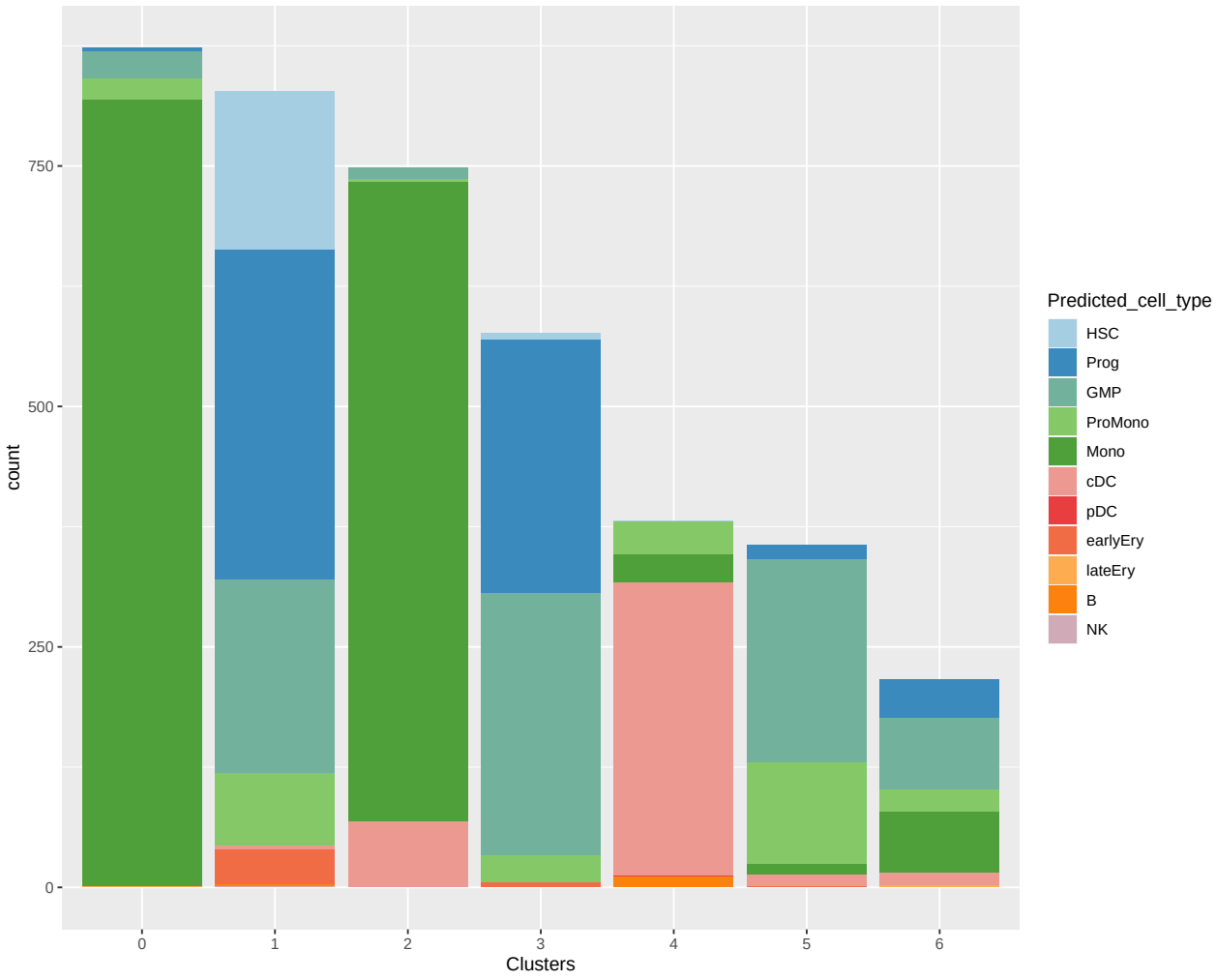
3. Predict the class of the cells using the markers and the expression of the BM cells form Van_Galen paper

```
## Performing PCA on the provided reference using 1821 features as input.
## Projecting cell embeddings
## Finding neighborhoods
## Finding anchors
##  Found 2488 anchors
## Filtering anchors
##  Retained 1601 anchors
## Finding integration vectors
## Finding integration vector weights
## Predicting cell labels
##
##         HSC Prog  GMP ProMono Mono  cDC  pDC earlyEry lateEry ProB    B Plasma
##   CD34  174  660  797     192   24  199    0       42       1    0   11      0
##   CD38    0    5    2     100 1563  203    1        0       2    0    2      0
```

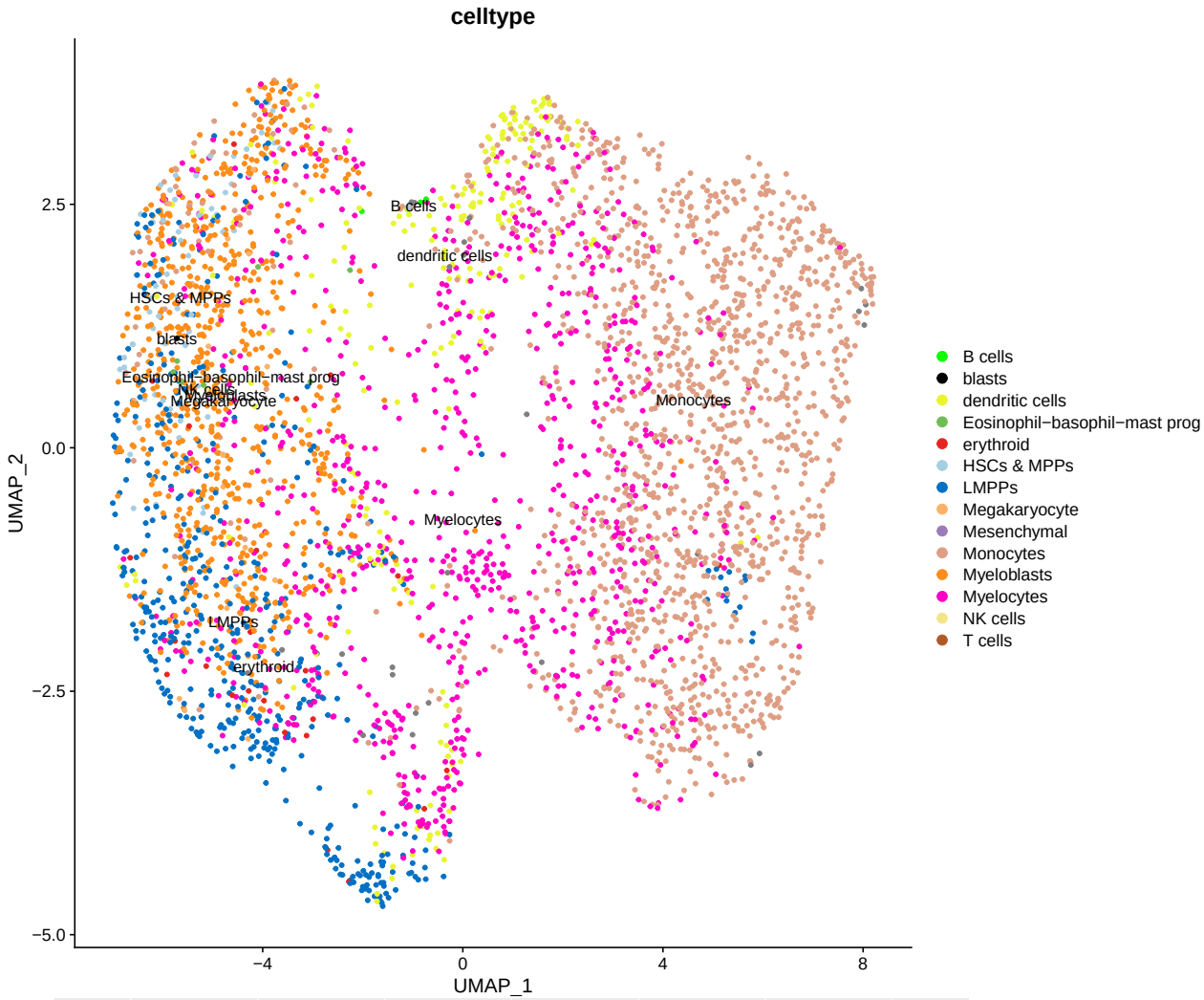
```
##
##           T  CTL   NK
##    CD34   0    0    0
##    CD38   0    0    1
```

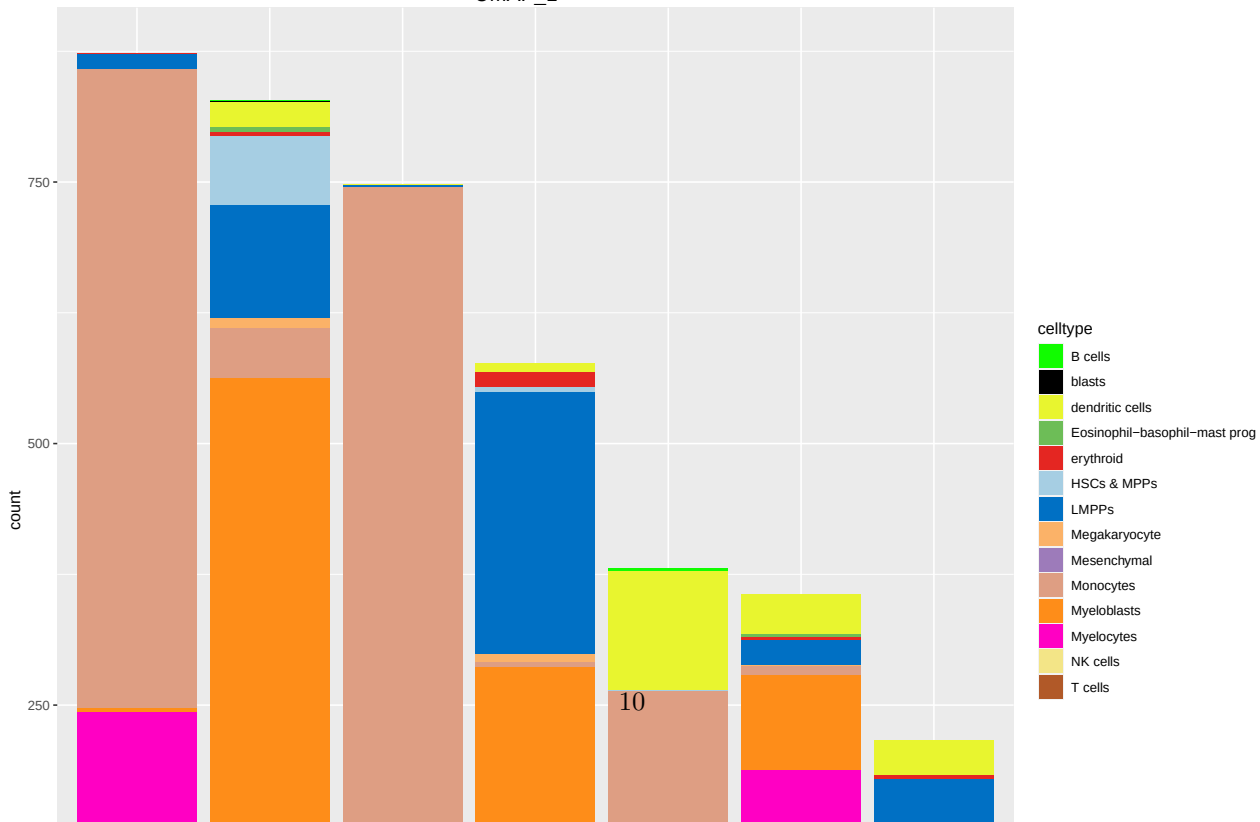



prediction
UMAP_2
UMAP_1
2.5
0.0
−2.5
−5.0
−4
0
4
8
HSC
earlyEry
NK
Prog
GMP
ProMono
B
pDC
cDC
lateEry
Mono
HSC
Prog
GMP
ProMono
Mono
cDC
pDC
earlyEry
lateEry
B
NK


count
750
500
250
0
0
1
2
3
4
5
6
Clusters
Predicted_cell_type
HSC
Prog
GMP
ProMono
Mono
cDC
pDC
earlyEry
lateEry
B
NK

4. Project the predictions from Velten onto our UMAP





Cluster 1 seems the most likely to be enriched in LSC