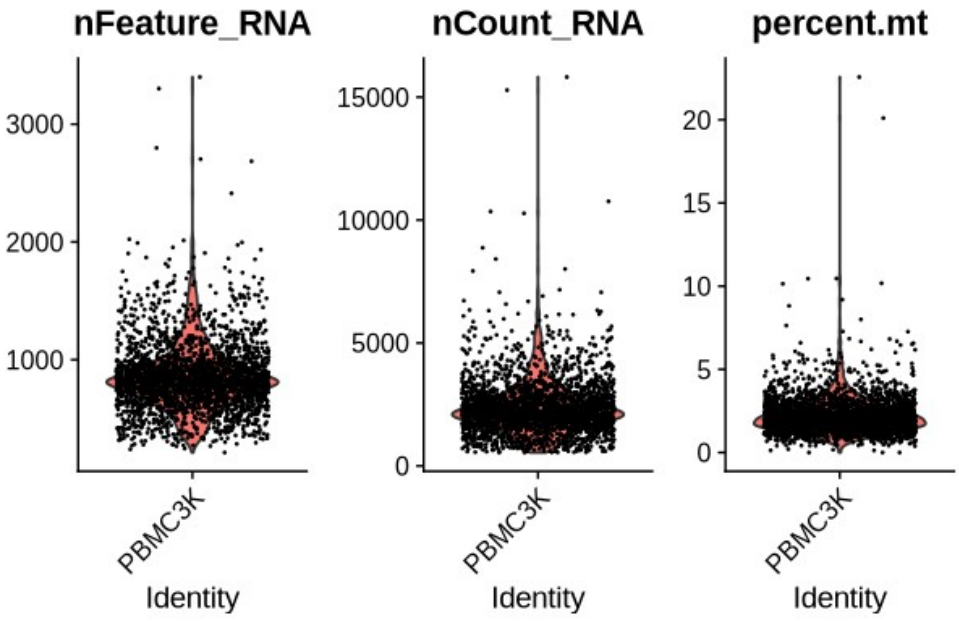
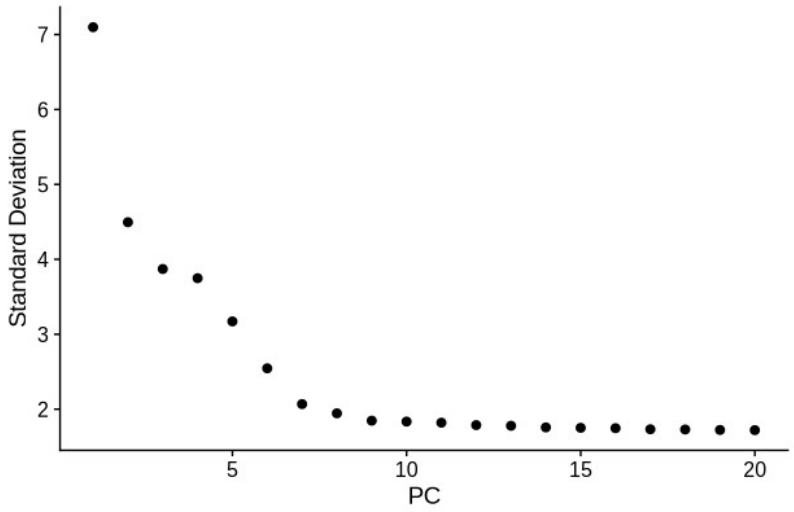
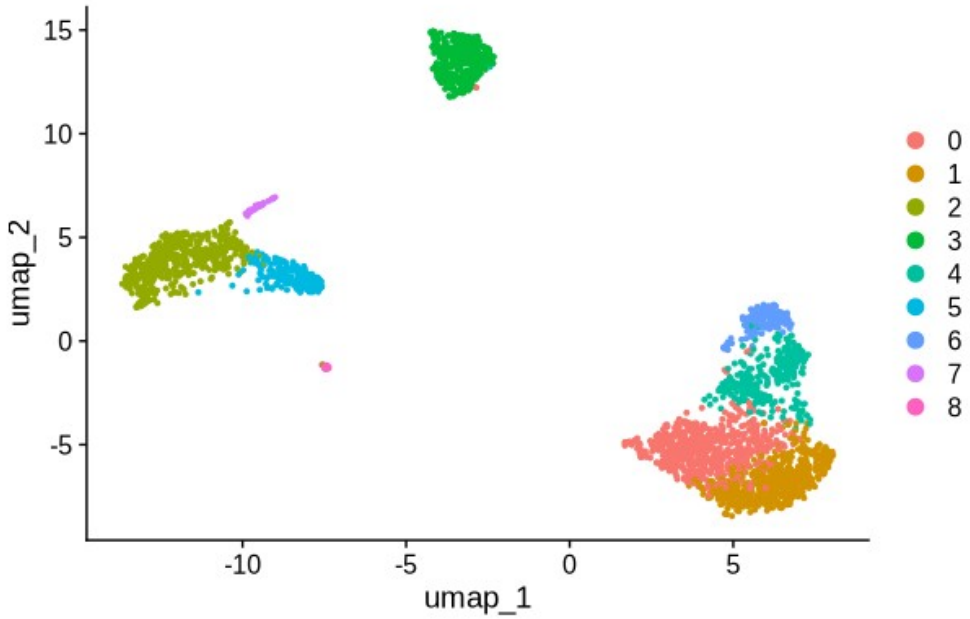
Quality control :
- we filter our data based on 200<feature data < 2500 and less than 5 percent of mitochondrial reads.



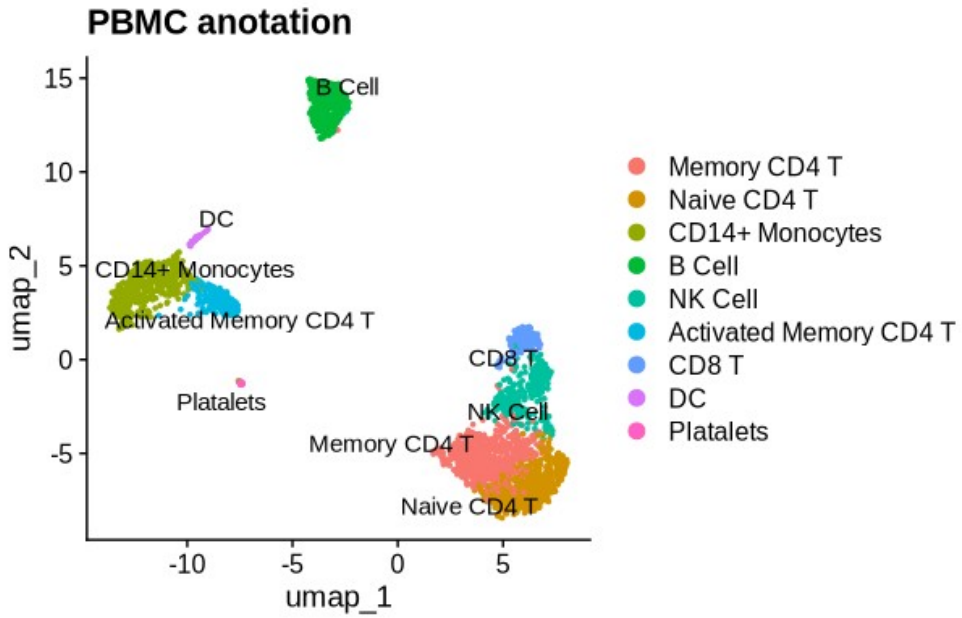
Normalization and Scaling :
- after NormalizData() and ScaleData(), we shift to dimensionality reduction. First by PCA to choose the dimension for down stream analysis. Based on the plot we use 15 pcs.



Clustering :
- after running FindNieghbors() and FindClusters() with 15 pcs, we use umap dimensionality reduction for a better visualization. Finally having the Dimplot()



Cell marker annotation :
manual annotation

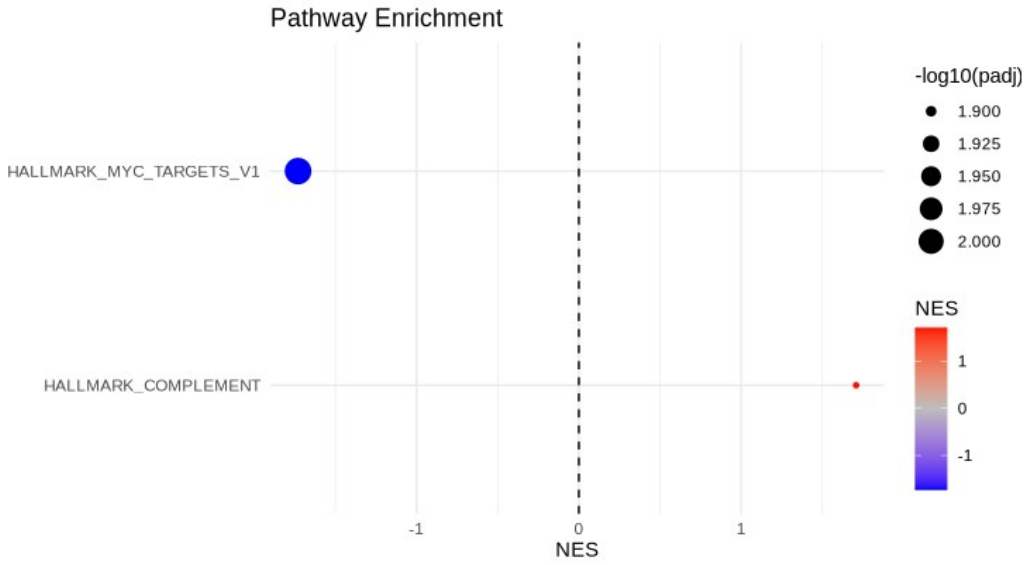


GSEA:

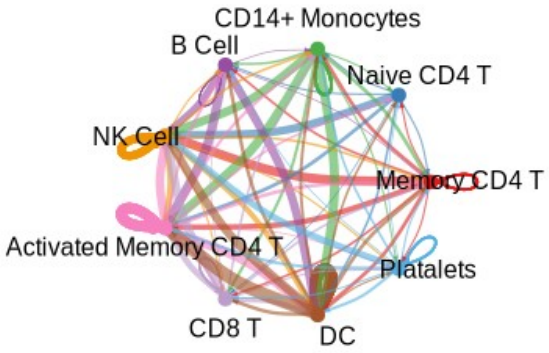
To find what biological process are different between 2 cell types “activated memory CD4 T” vs “Naive CD4T”, we conducted GSEA. After finding markers by **FindMarkers()**, we ranked genes by how different they are between different cell types(log2foldchange).
By **fgsea** we test if pathway genes are in top/bottom of our ranked list or how enriched the pathway is compared to random chance.

Pathways with normalized enrichment score (NES) > 0 indicate enrichment in activated memory CD4 T cells, while NES < 0 indicates enrichment in naive CD4 T cells.
Pathways with adjusted p-value < 0.05 were considered statistically significant.

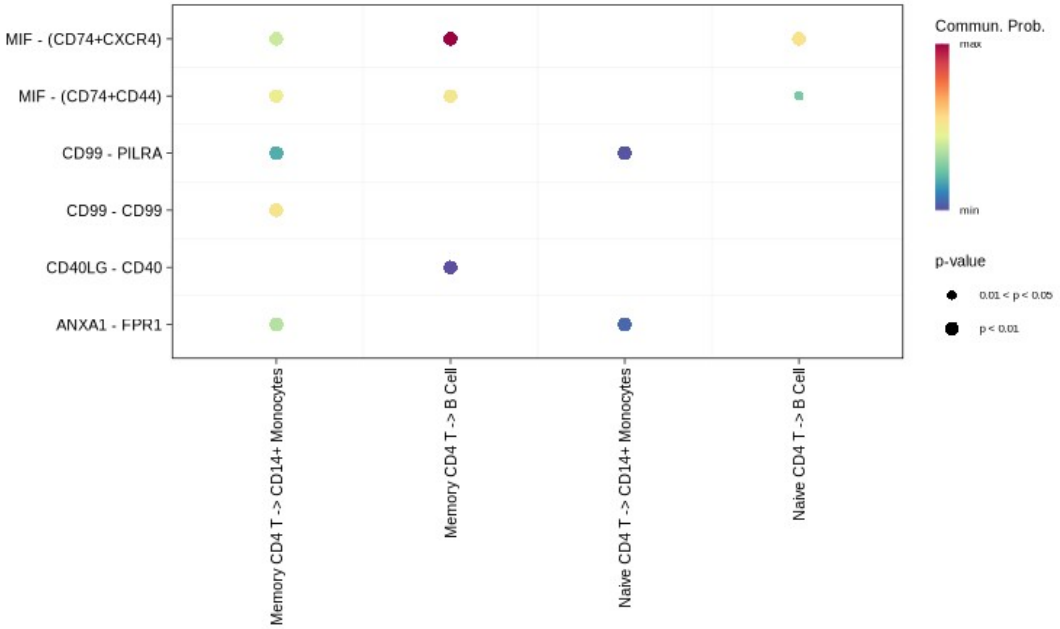
The result is :
When CD4 T cells activate, they turn on proliferation programs (MYC targets) while turning down innate immune functions (complement system).



Cell-Cell communication :
A) overall view
-how do all cell types talk together :



-Bubble plot :
what are top ligand receptor pairs in our reads.

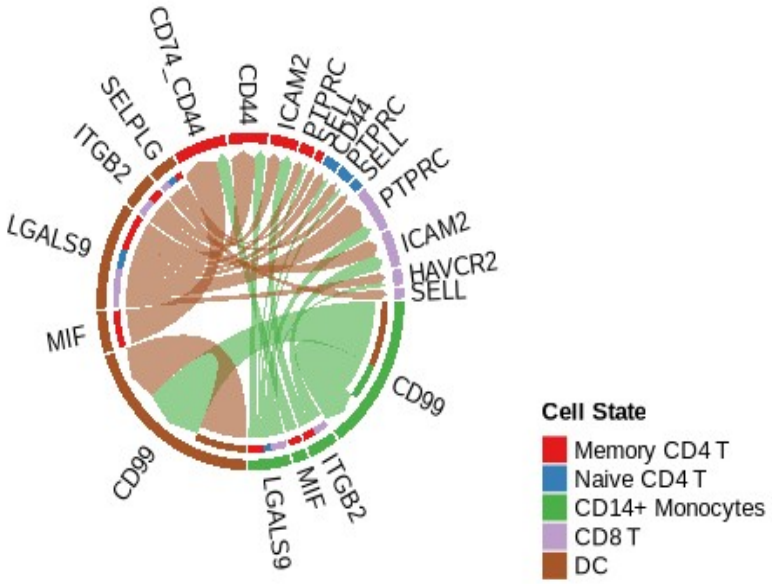


The result : CD4T cells are actively communicating with both monocytes and B cells through Inflammatory signals (MIF), Cell adhesion (CD99),B cell help (CD40-CD40LG) and Resolution signals (ANXA1)

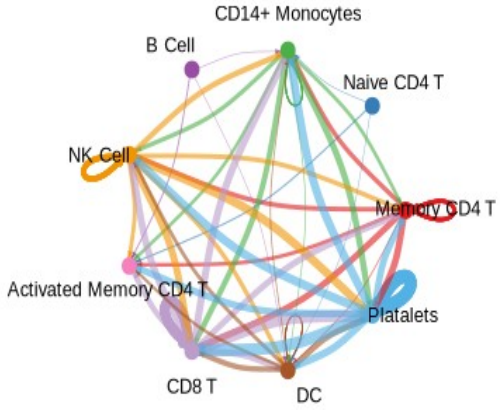
B) specific question

- what signals do **myeloid cells** provide in immune cells ?

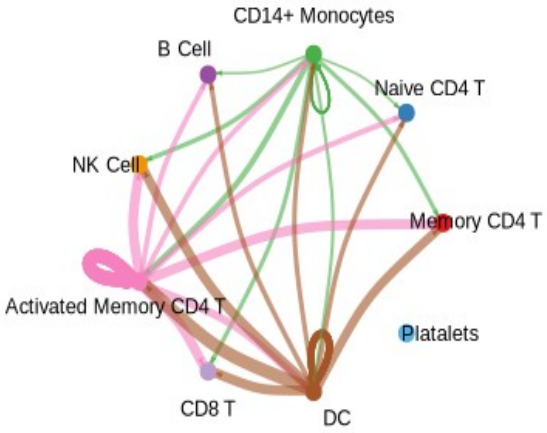
After defining specific cell types as **source.use()** and target.use() ,the plot shows CD99 and LGALS9 as dominant genes in interactions between cell types. CD99 is a key immune cell adhesion molecule and LGALS9 has a role in immune regulation and hemostasis.



- which cells communicate via CD99 pathway ?

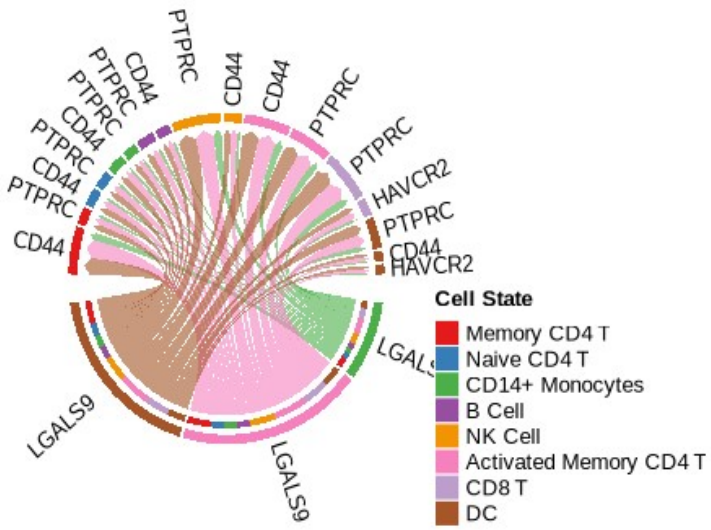


- which cells communicate via GALECTIN pathway (containing LGALS9) ?

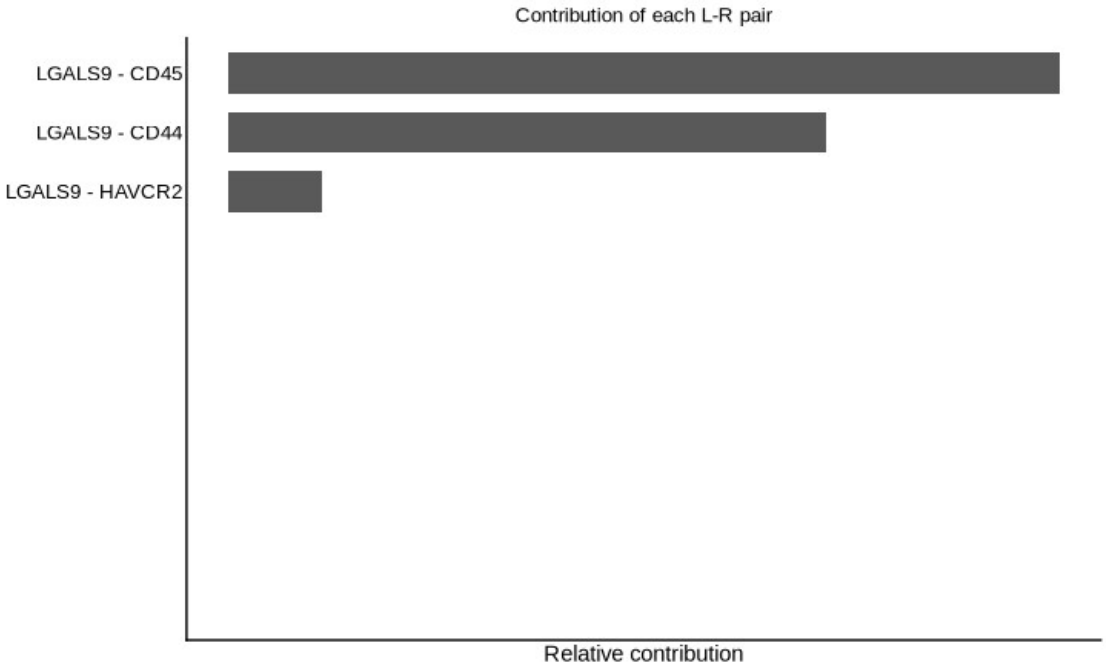


note: platelets are isolated since they have no role in immune regulation and no receptor for LGALS9, for immune regulation.

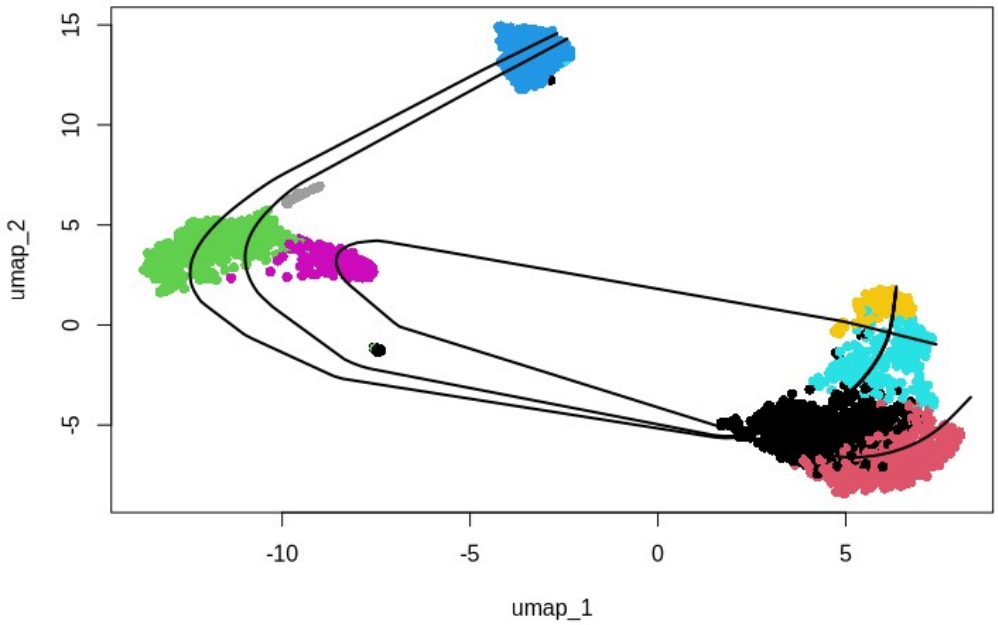
- For GALECTIN as an immune regulating pathway, what are specific ligand receptors ?



The most influential ligand receptors for GALECTIN :

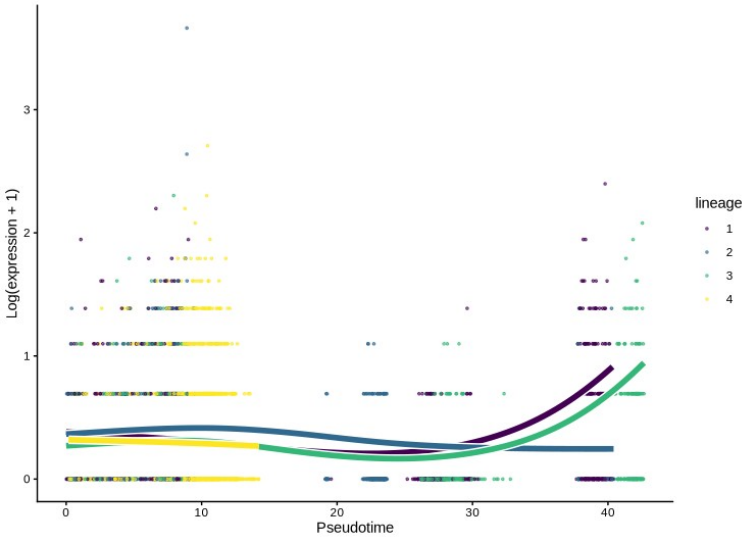


Trajectory analysis :

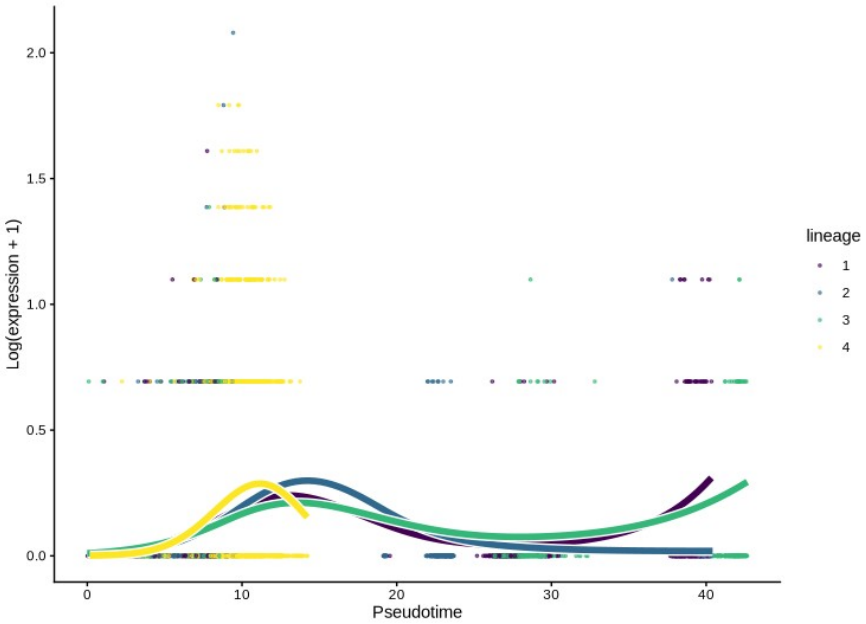


- tracking genes : by categorizing three groups of genes into native markers, early activation and housekeeping genes as control we tracked the expression change during time.

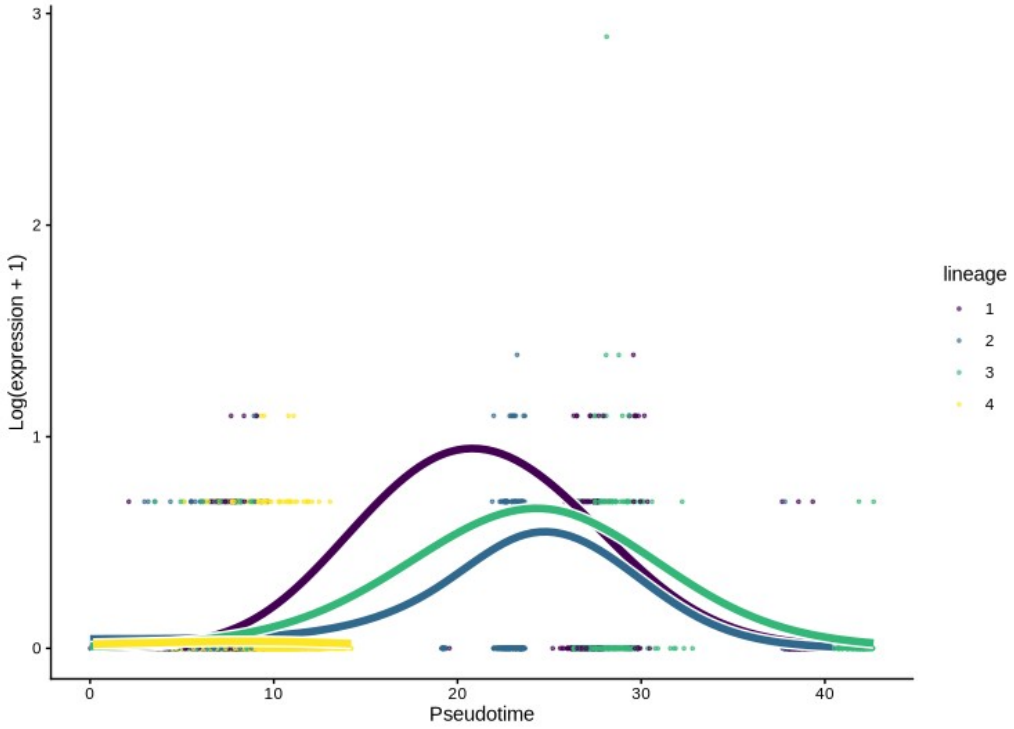
- The smoother plot for CD69 as an early activated gene, the blue line shows the pattern of decreasing overtime.



-The smoother plot for CCR7 as an naive marker, the blue line indicates a maximum followed by a decreasing pattern overtime.
-

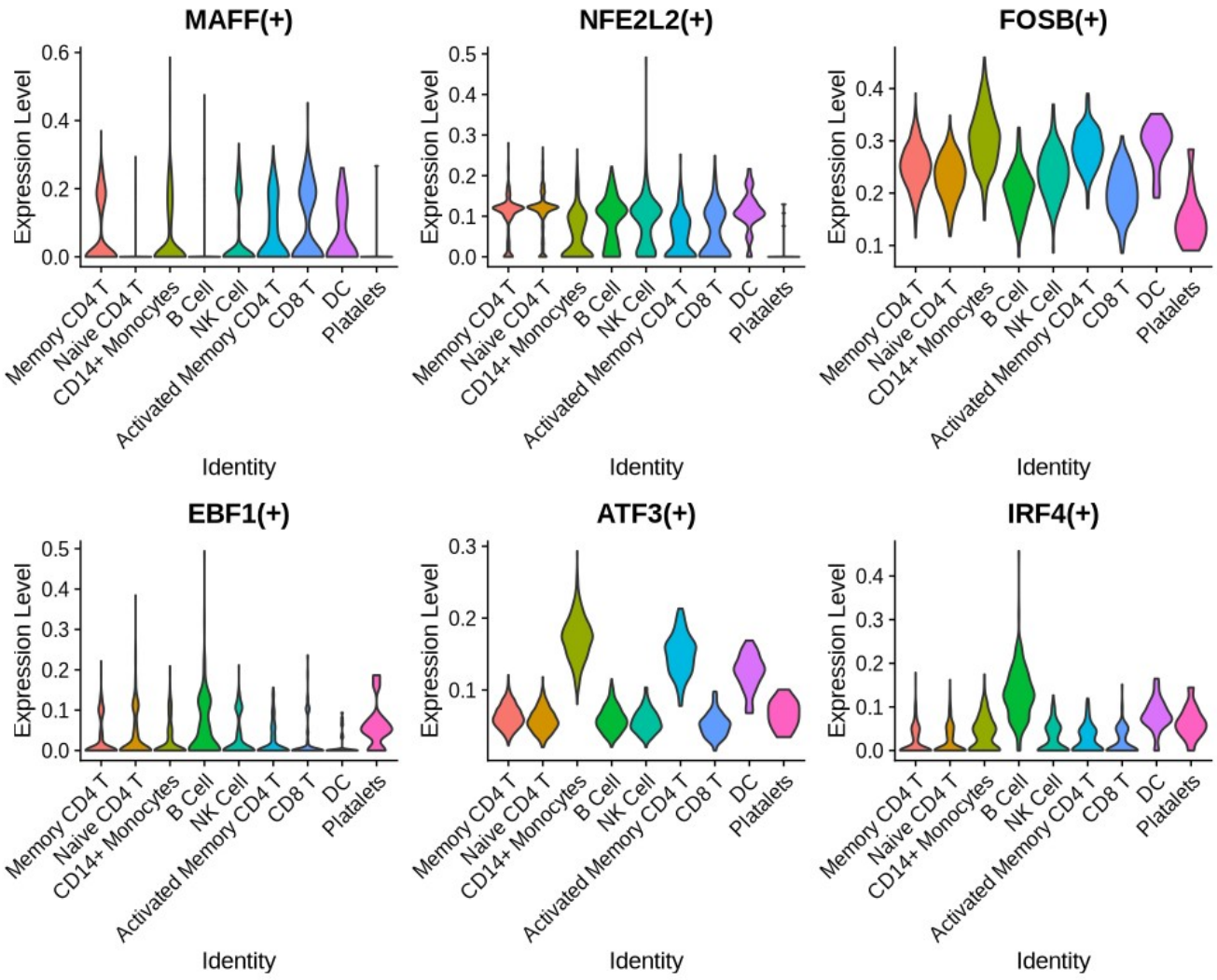


- the expression remains constant both in naive and activated state, confirming the pattern of a housekeeping gene for CD4.
-



Gene regulatory networks by pyscenic :
we preformed a regulatory network and transcription factor (TF) activity analysis by starting from our seurat object expression matrix , which was later transposed for **pyscenic**. First we used GRNBoost2 to create co-expression models resulting in a adjacencies.tsv file. By using the ctx command and the h38 motif database. we annotated our TF-target modules with cis regulatory motifs for top 200 targets and top 500 regulators, resulting in a reg.csv file.
To compute TF activities in each cell , we used AUCell that results in a matrix of regulons as cols, and cells as rows. Which was then transformed and added to the seurat object to visualize TF activities among clusters.

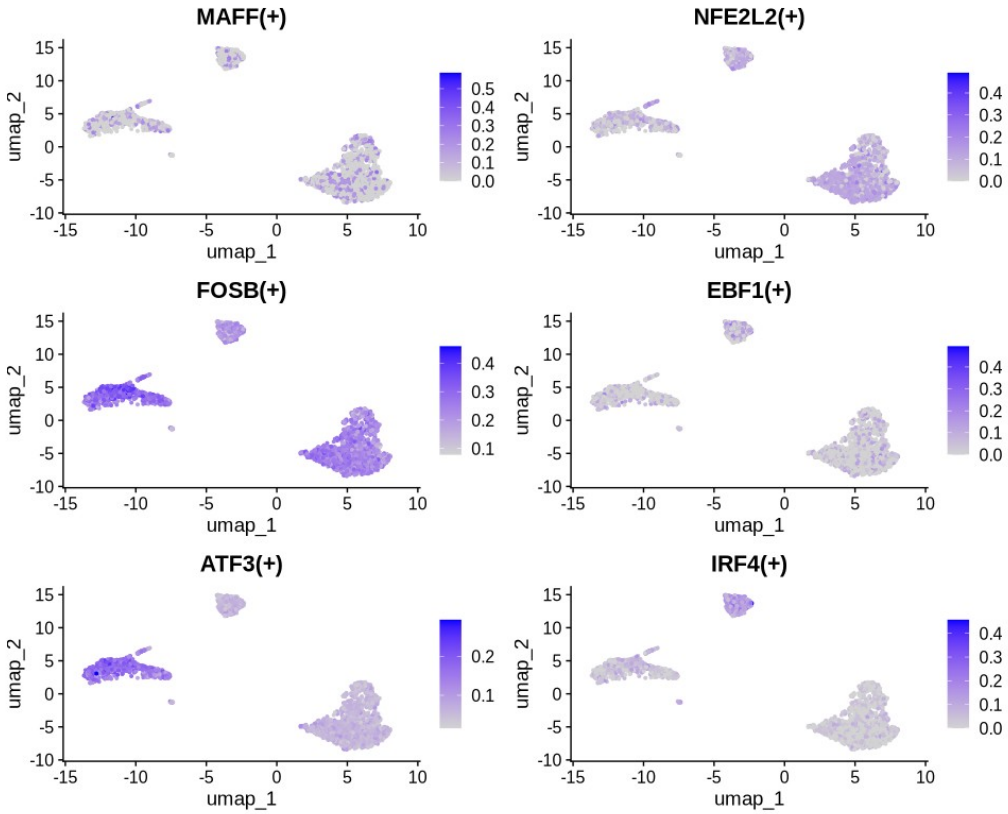
Finally, Top 10 Tfs among clusters were extracted , and then were filtered by their biological importance (house keeping markers as ZNF family were ignored).



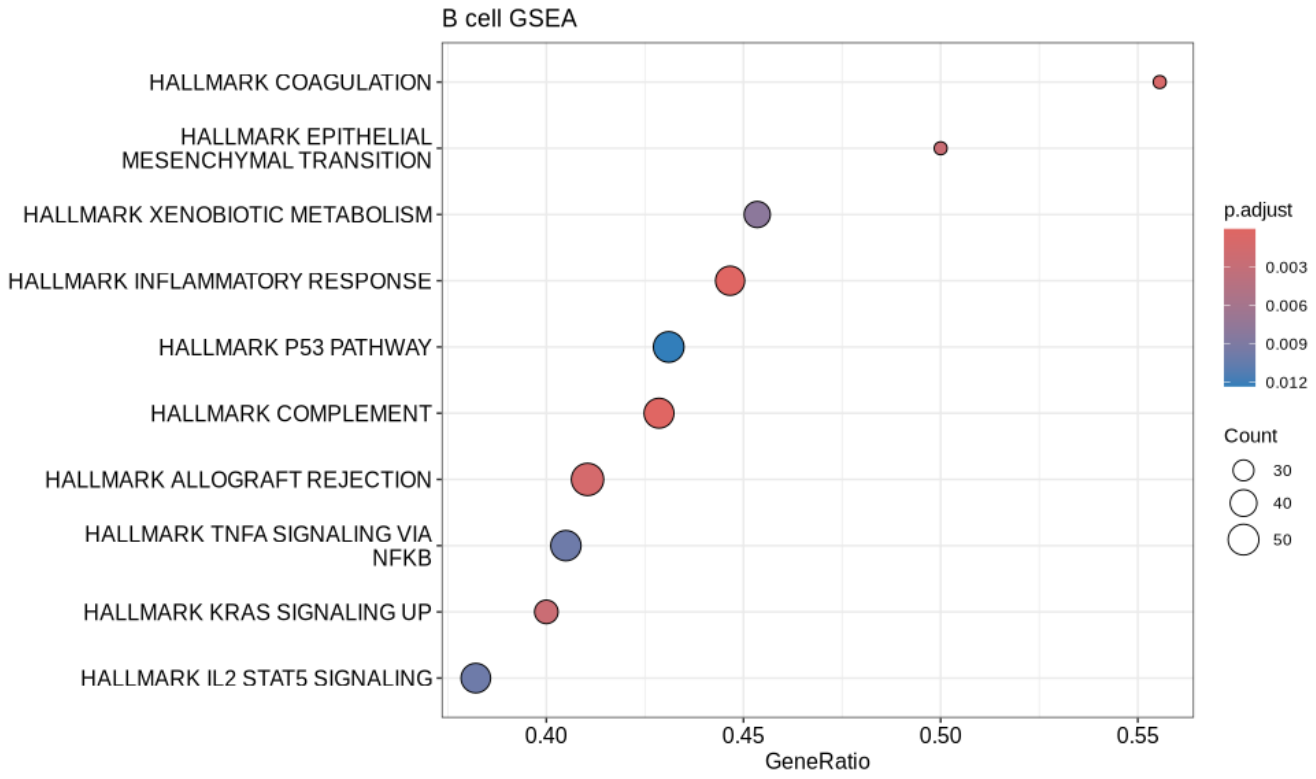
Violin plot of top top Tfs among clusters

Biological interpretations of TFs :

- MAFF : is part of **oxidative stress** response, mostly expressed in **monocytes**
- NFE2L2 : the regulator of **antioxidant and redox genes**, highly expressed in **NK cells** as they are active metabolically and can take part in oxidative stress.
- FOSB : is a signal for **immune activation**, as expected highly expressed in **monocytes and T cells**.
- EBF1 : a critical factor for **B cell differentiation,** as expected highly expressed in **B cells**.
- ATF3 : another **stress response** TF, common in monocytes and **T cells.**
- IRF4 : an **interon regulatory** factor, found in **B cells.**



Feature plot of TF expression



dot plot of significant pathways in the cluster of B cells, indicates related pathway to stress and inflammatory signals and B cell biology except for coagulation and mesenchymal transition that has lees than 20 counts.

Biological interoperate :
inflammatory- signaling and Xenobiotic-metabolism >> **inflammatory and stress**

IL2 STAT, KRAS,TNFA,P53 >> related to B cell biology