

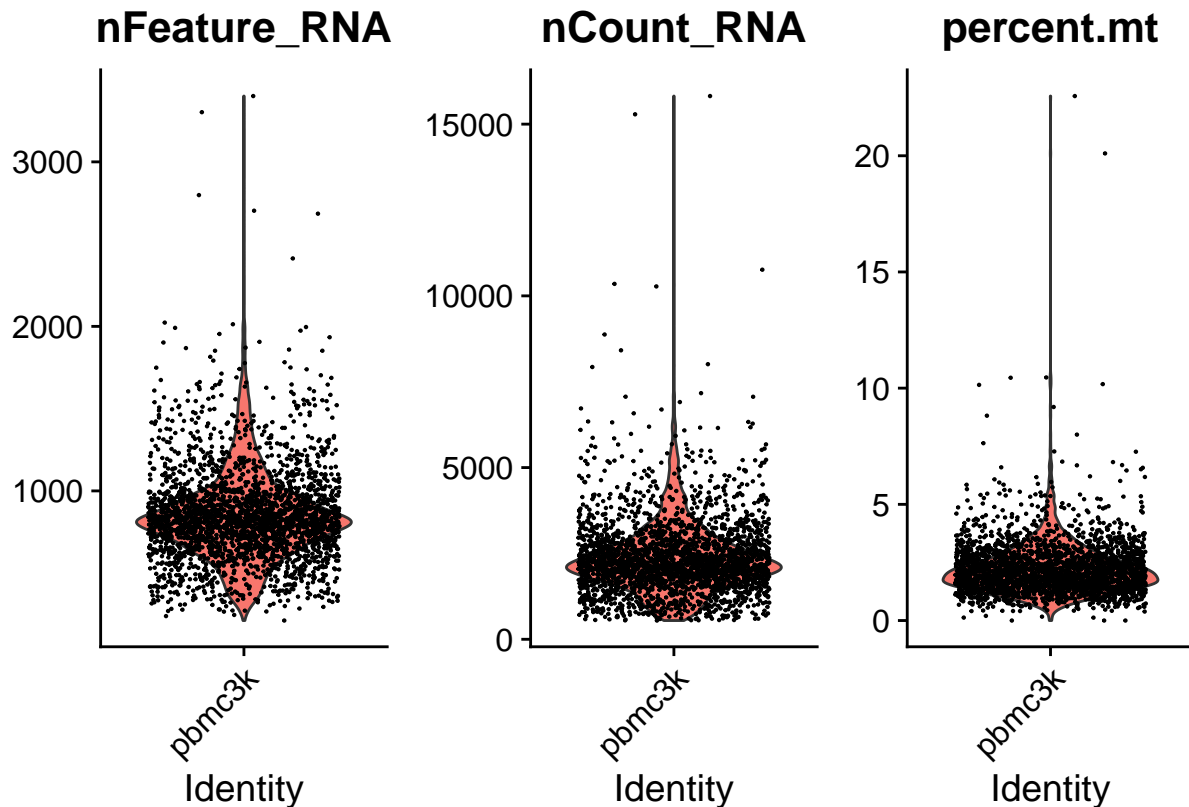
Seurat_tutorial

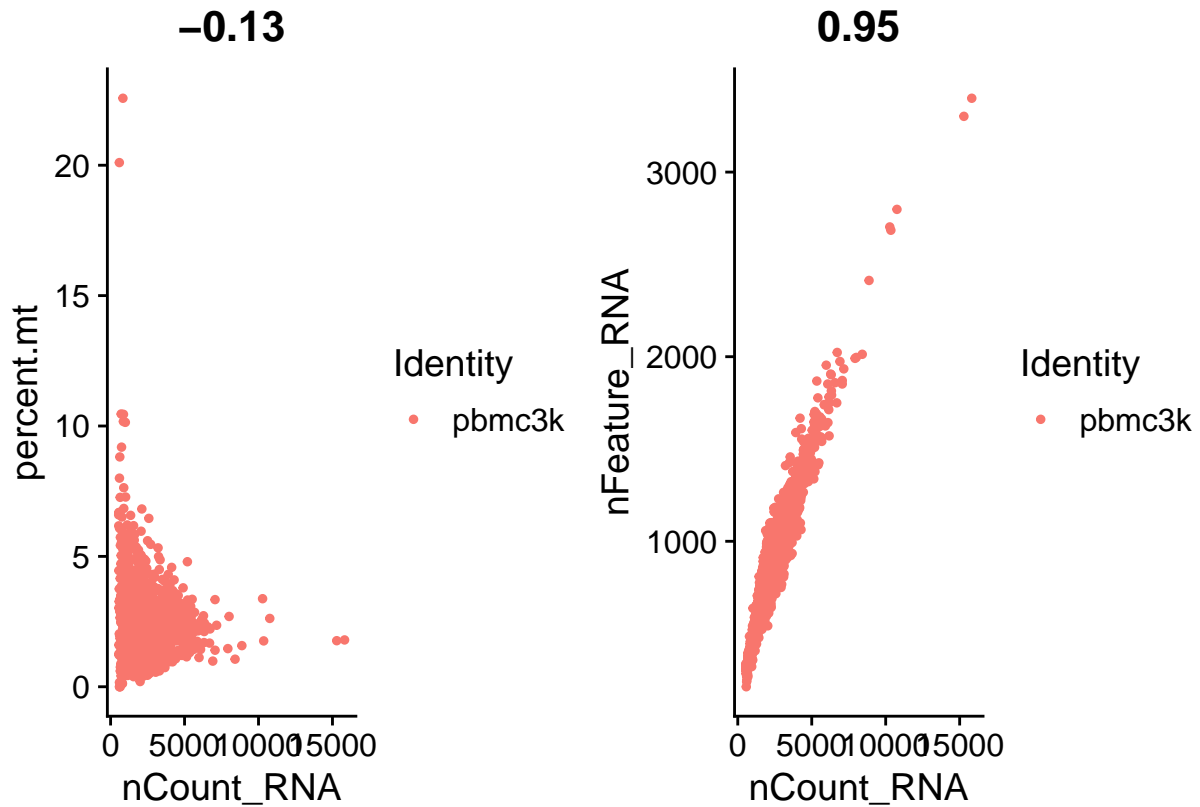
2025-11-20

De dataset die gebruikt wordt voor deze tutorial bestaat uit 2.700 single cells (PBMC's, peripheral blood mononuclear cells) die zijn gesequenced in de Illumina Nextseq 500. De dataset is van 10X Genomics, en de ruwe data kan gevonden worden op deze site: https://cf.10xgenomics.com/samples/cell/pbmc3k/pbmc3k_filtered_gene_bc_matrices.tar.gz.

QC en selecteren van de cellen

In de vioolplot worden de features, counts en percentage mitochondria weergegeven. Van de features blijven tussen de 200 en 2500 over. Van de mitochondria blijft onder de 5% over, daarboven is de cel waarschijnlijk van slechte kwaliteit of is deze stervende.



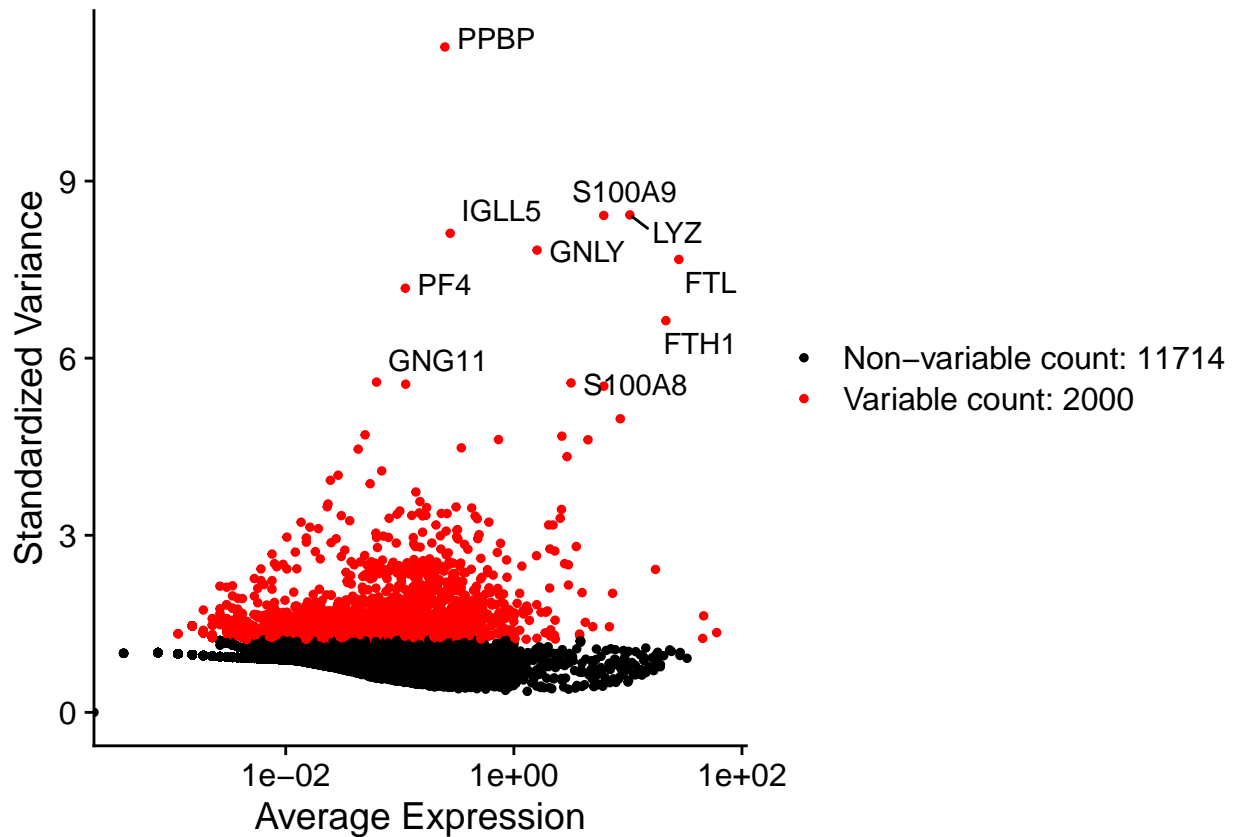


Normaliseren van de gegevens

De data wordt genormaliseerd door de totale expressie van elke cel om te zetten naar 10.000 en vervolgens een log-transformatie te nemen (het getal weer kleiner maken). Hierdoor kunnen de getallen per cel met elkaar vergeleken worden.

Kenmerkselectie

De 10 meest variërende genen worden in onderstaand figuur aangegeven met hun naam.



De gegevens schalen

De gemiddelde expressie over de cellen wordt op 0 gesteld, ook wordt de variantie over de cellen op 1 gesteld.

Lineaire dimensionale reductie uitvoeren

Met behulp van principal component analyses (PCA) worden er van de oude variabelen, nieuwe variabelen gemaakt door lineaire combinaties te maken. Hieruit ontstaan enkele hoofdcomponenten (PC's). PC1 dekt het grootste deel van de variatie binnen de data. PC2 dekt daarna het grootste deel, en zo voort.

Dimensionaliteit

Clusteren

Niet-lineaire dimensionale reductie uitvoeren (UMAP/tSNE)

Het vinden van differentieel tot expressie gebrachte kenmerken (clusterbiomarkers)

Toewijzen van celtype-identiteit aan clusters

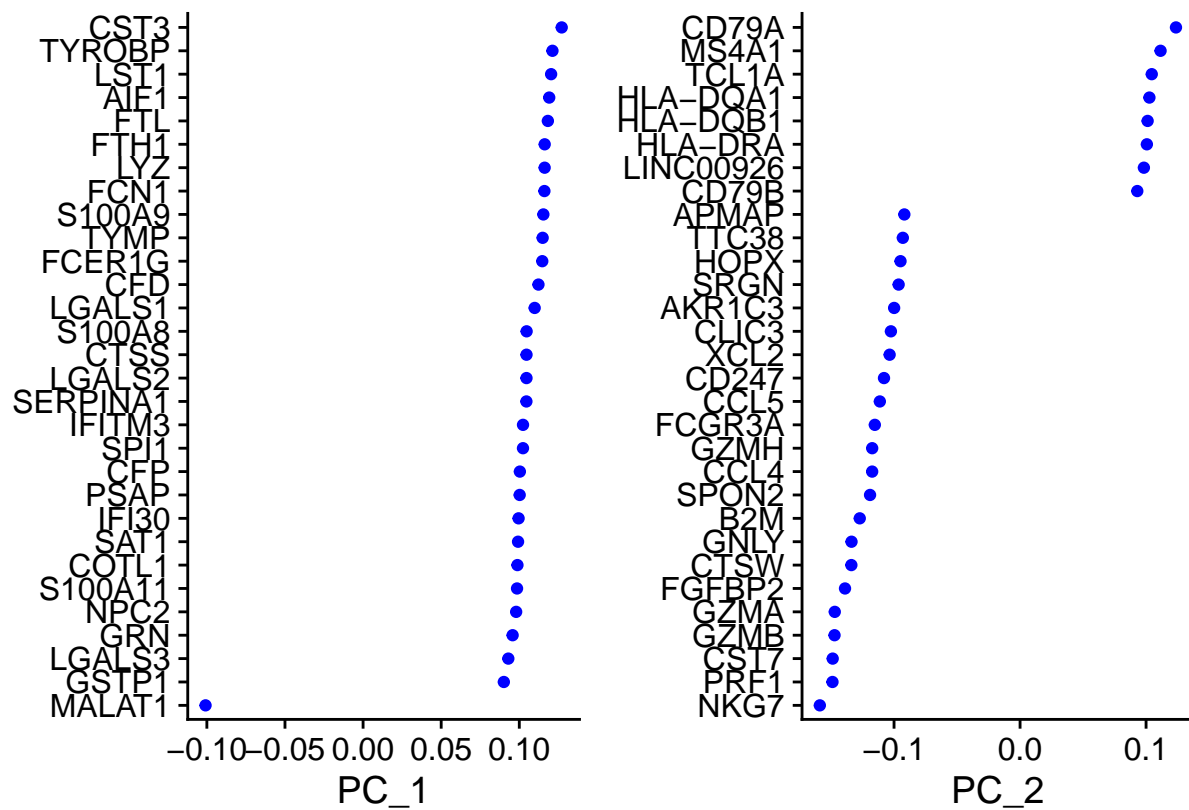


Figure 1: VizDimLoadings. De topgenen worden weergegeven die geassocieerd worden met PC1 of PC2.

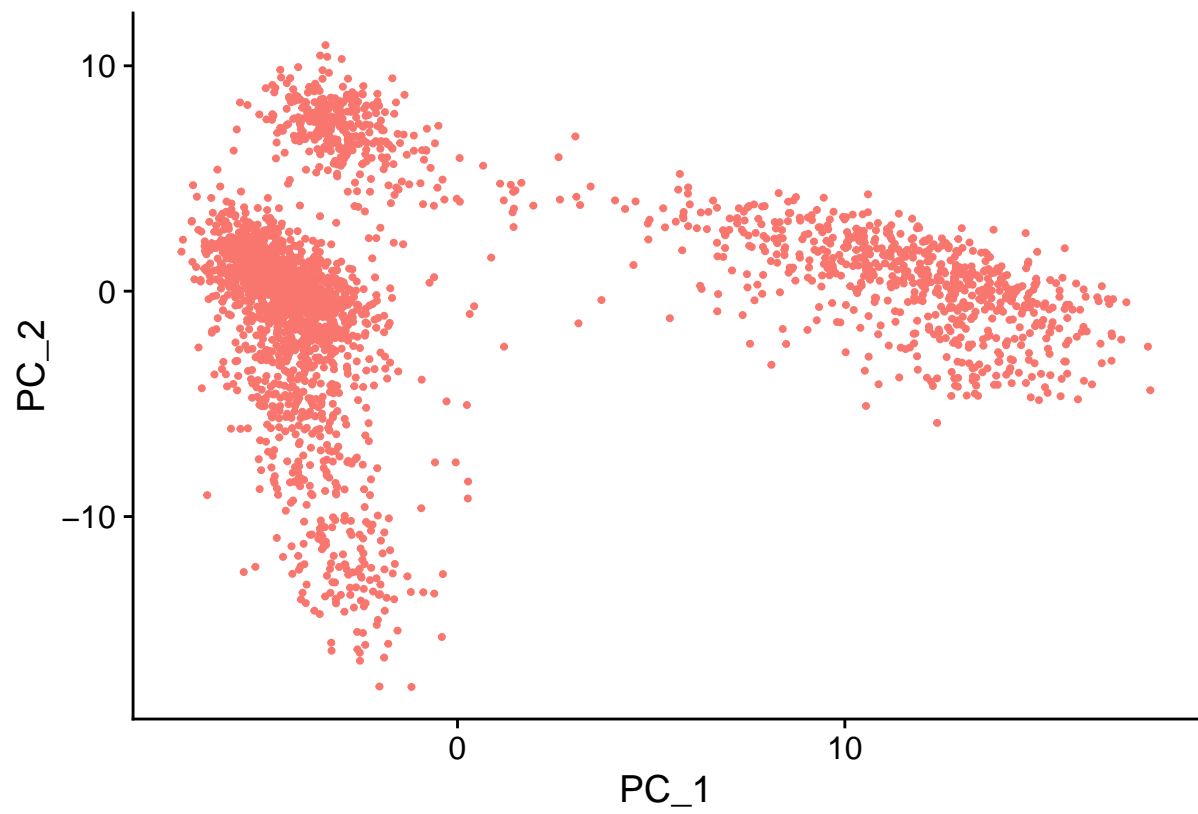


Figure 2: DimPlot, Op de x-as wordt PC1 geplott en op de y-as wordt PC2 geplott. Elke stip representeert een cel.

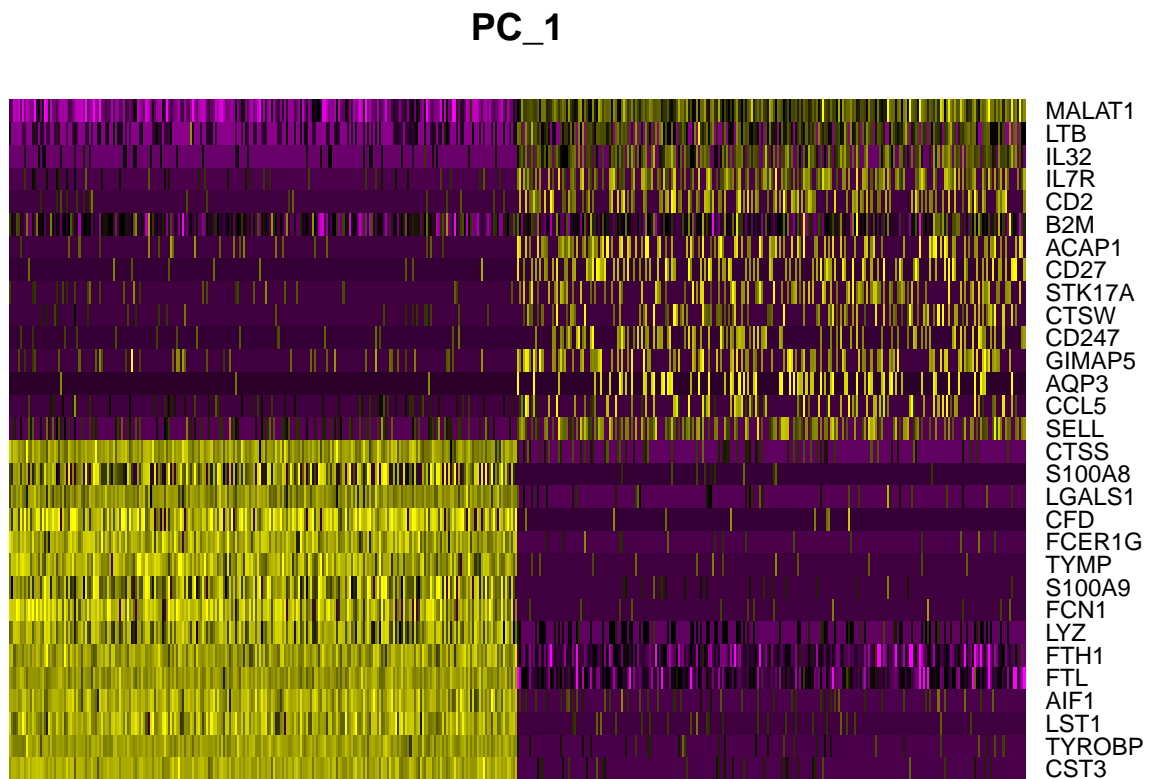


Figure 3: DimHeatmap, PC1 wordt weergegeven met als de verschillende genen in de rijen, en verschillende cellen in de kolommen. Paars is een lage PCA-score, geel is een hoge PCA-score. Clusters worden hiermee gevisualiseerd

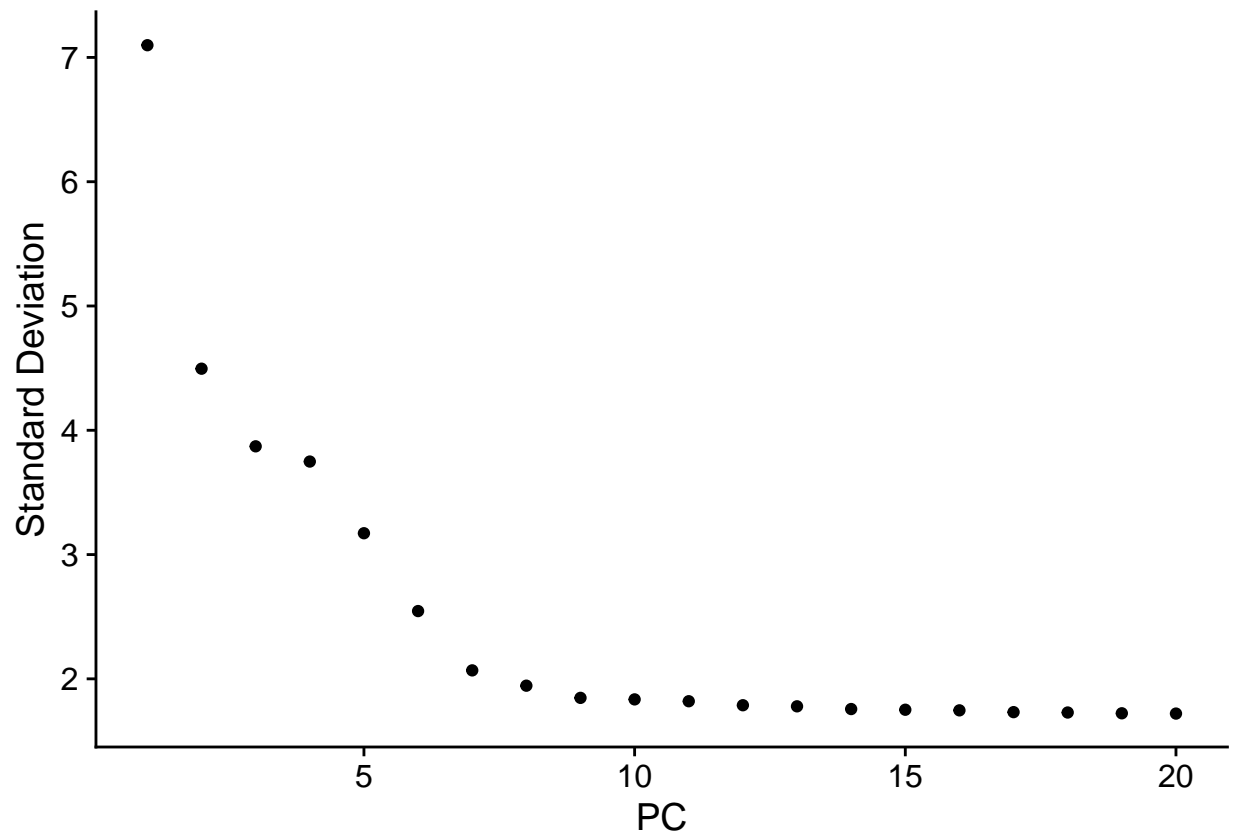


Figure 4: ElbowPlot

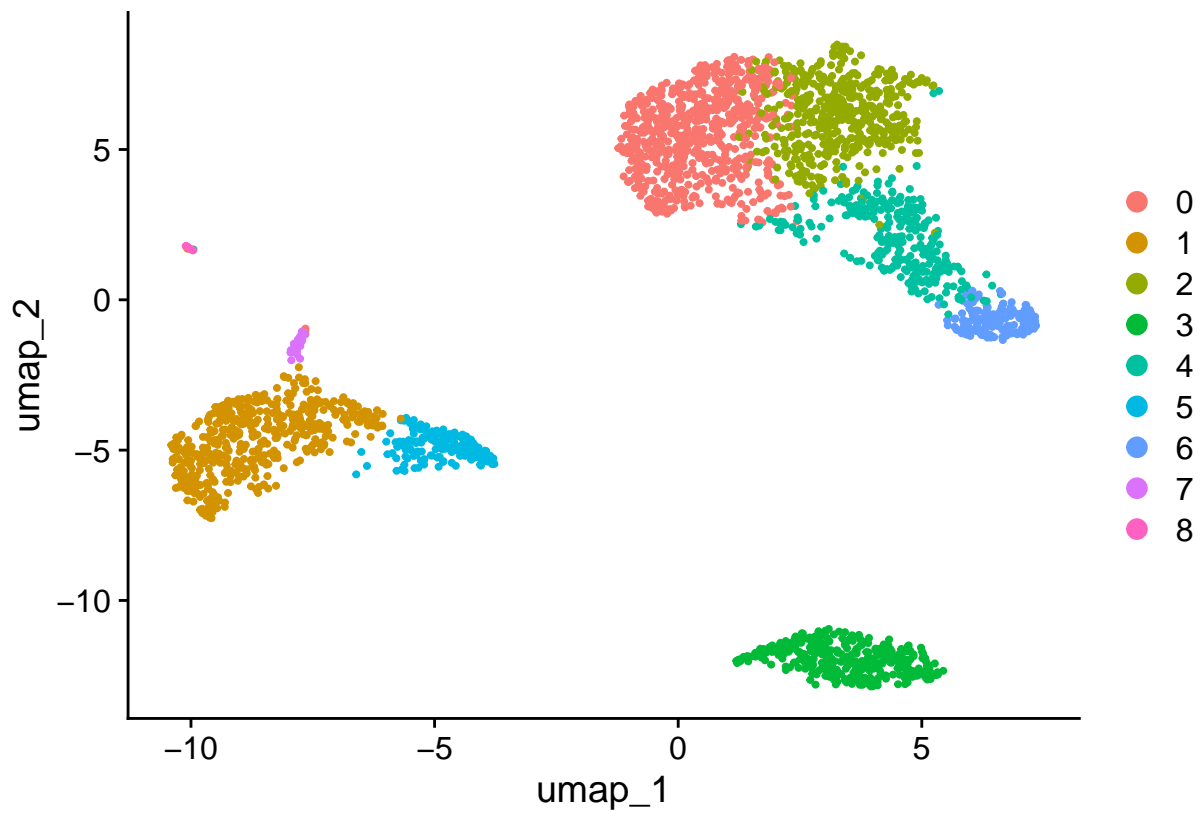


Figure 5: DimPlot: Op de x-as wordt umap_1 geplott en op de y-as wordt umap_2 geplott. Elke stip representeert een cel. De kleur geeft aan bij welk cluster deze cel hoort.

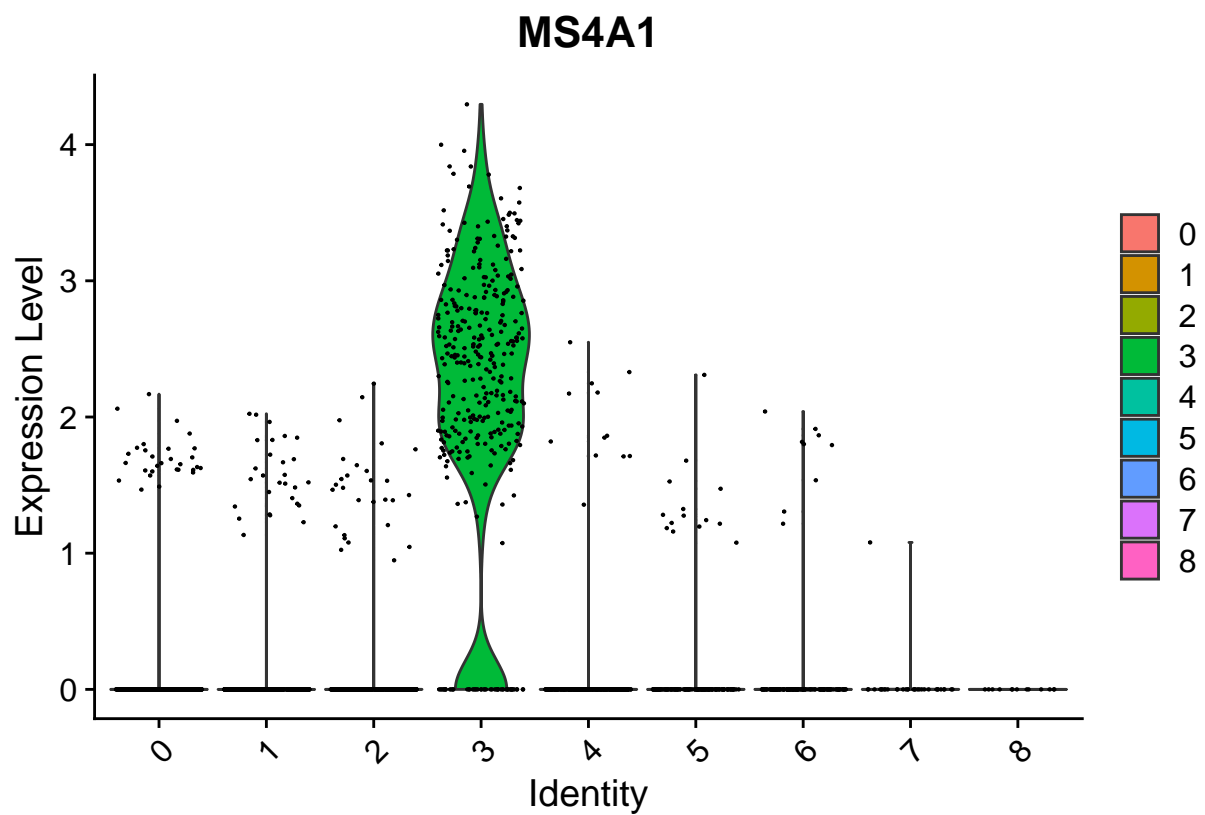


Figure 6: ViolinPlot: Het expressielevel van het gen MS4A1 wordt per cluster weergegeven. In cluster 3 komt dit gen het meest tot expressie.

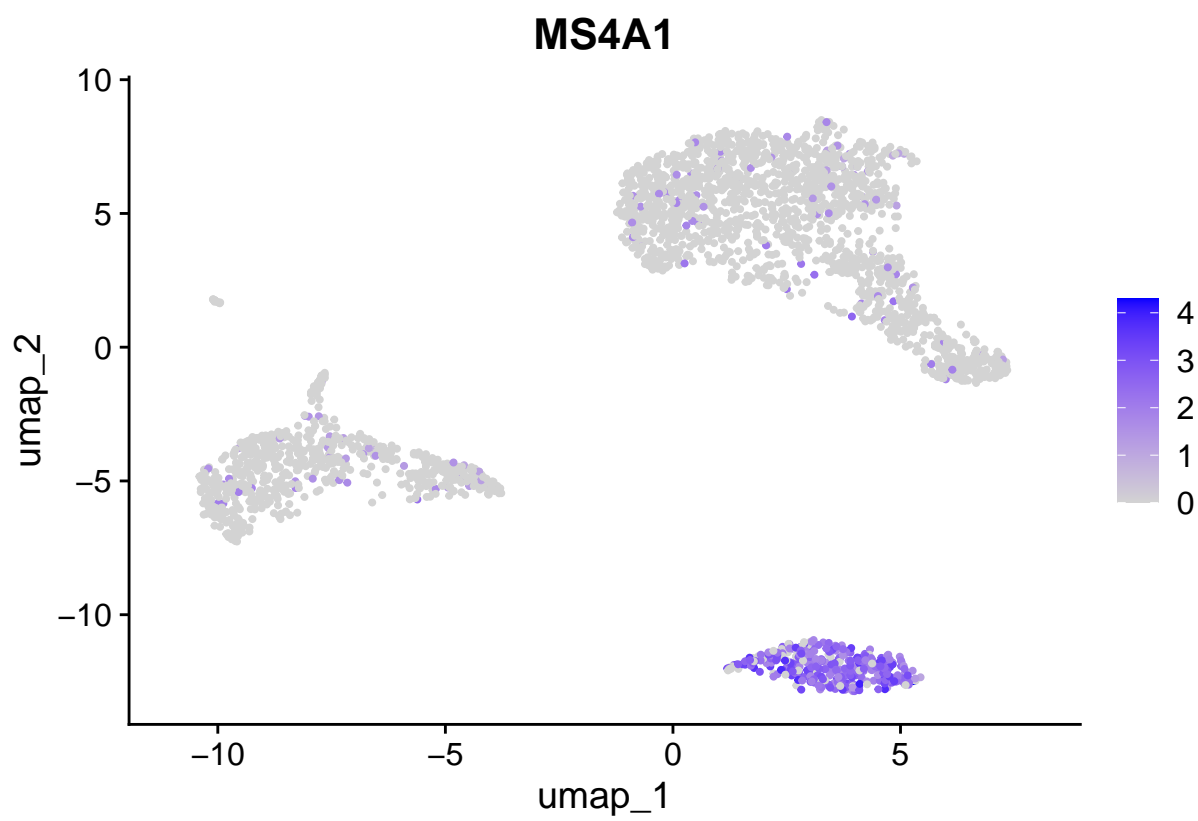


Figure 7: FeaturePlot, MS4A1

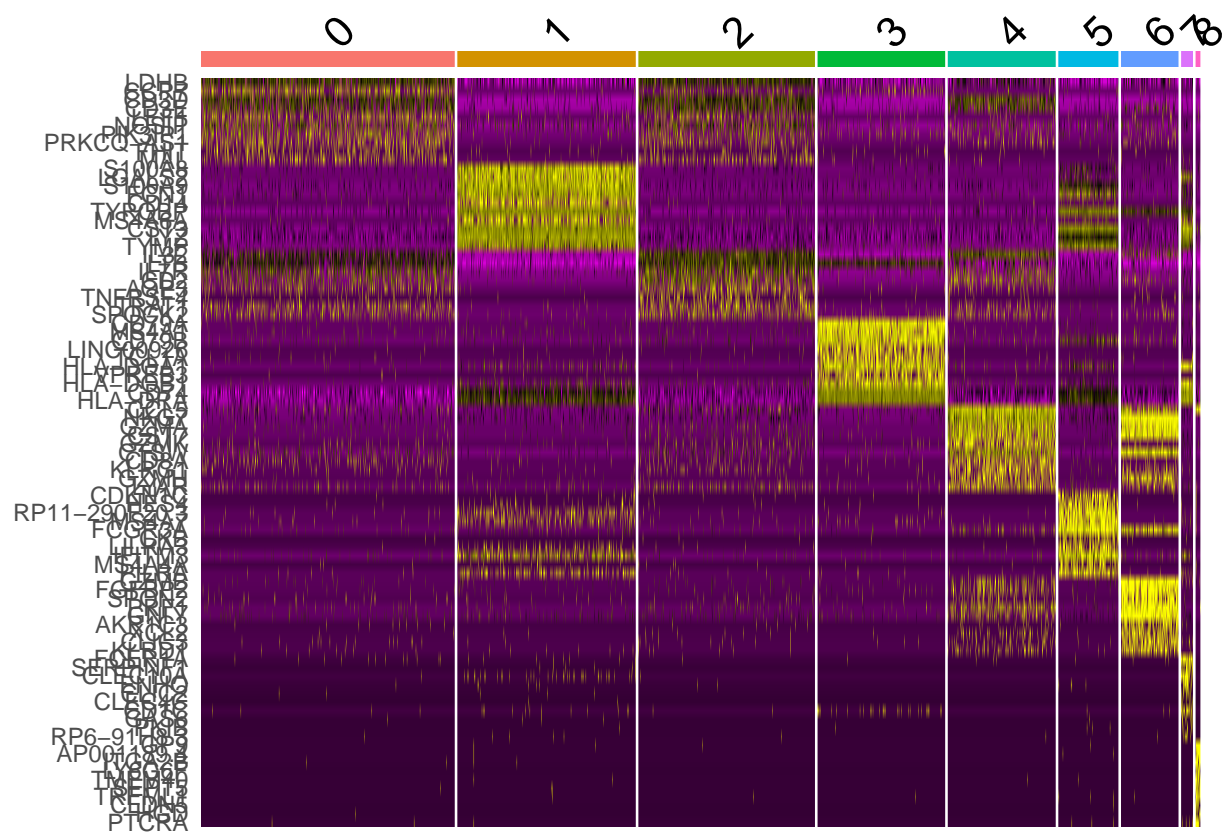


Figure 8: DoHeatMap

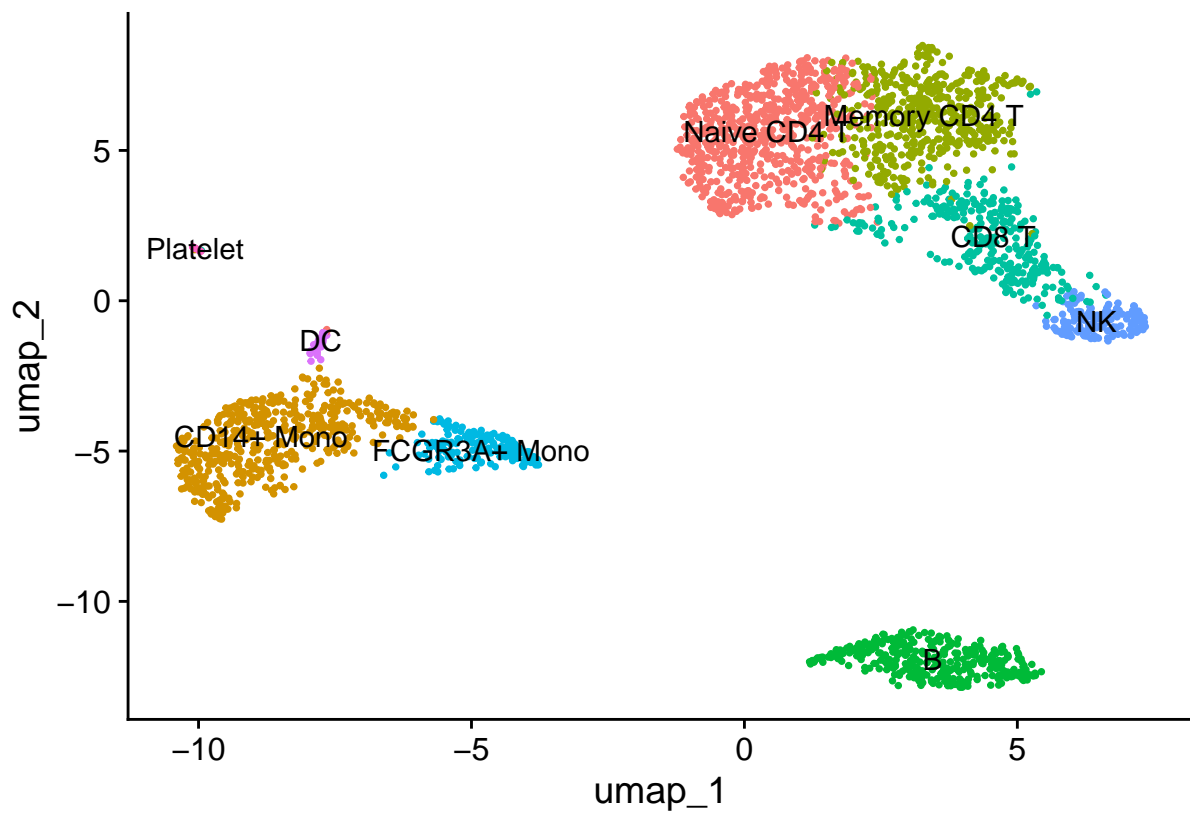


Figure 9: DimPlot