

Seurat pipeline

Anne Brussaard

2025-05-03

In dit bestand zullen de volgende deelvragen behandeld worden. 1. Welke filtering is geschikt voor de data? 2. Hoeveel PCA's worden meegenomen in verdere analyse? 3. Welke clusters kunnen gebruikt worden voor de analyse met BRIE2.

Om deze vragen te kunnen beantwoorden zal de volgende workflow worden gevuld. 1. Data laden 2. Seurat object maken 3. Kwaliteits check data 4. Normalisatie data 5. PCA analyse 6. Clustering 7. UMAP 8. Selectie clusters

Stap 1.1: Packages laden. De packages dplyr, ggplot2, pheatmap, tidyr, RColorBrewer, ggrepel, Seurat, Matrix, patchwork en here worden geladen.

Stap 1.2: Data wordt geladen

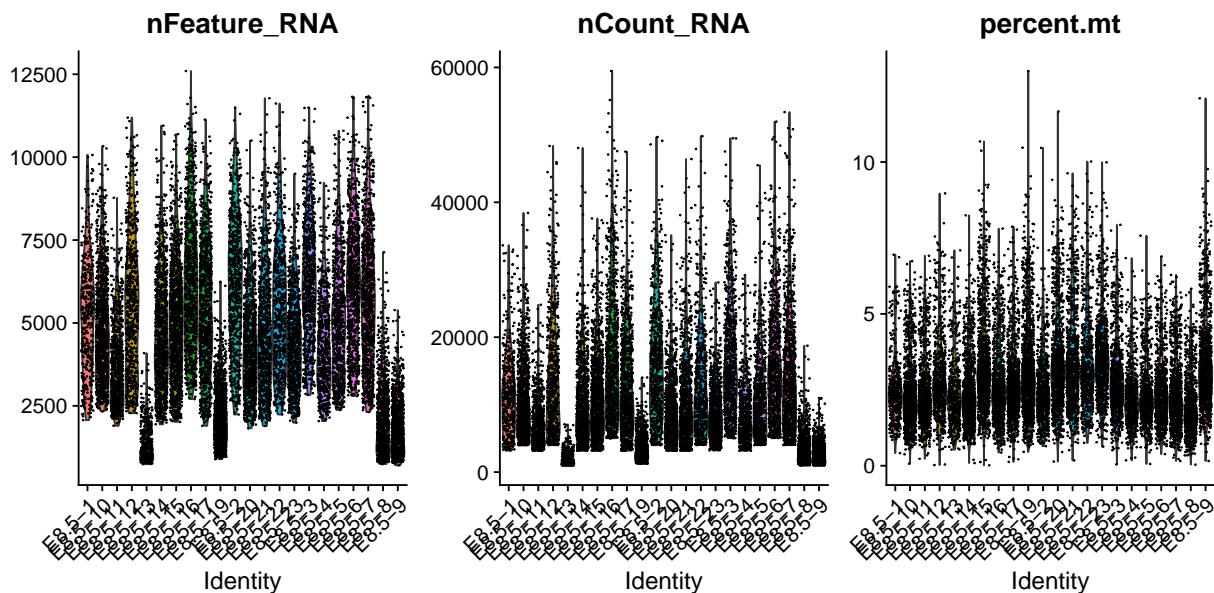
Stap 1.3: MTX files worden geladen

Stap 2.1: Seurat object maken

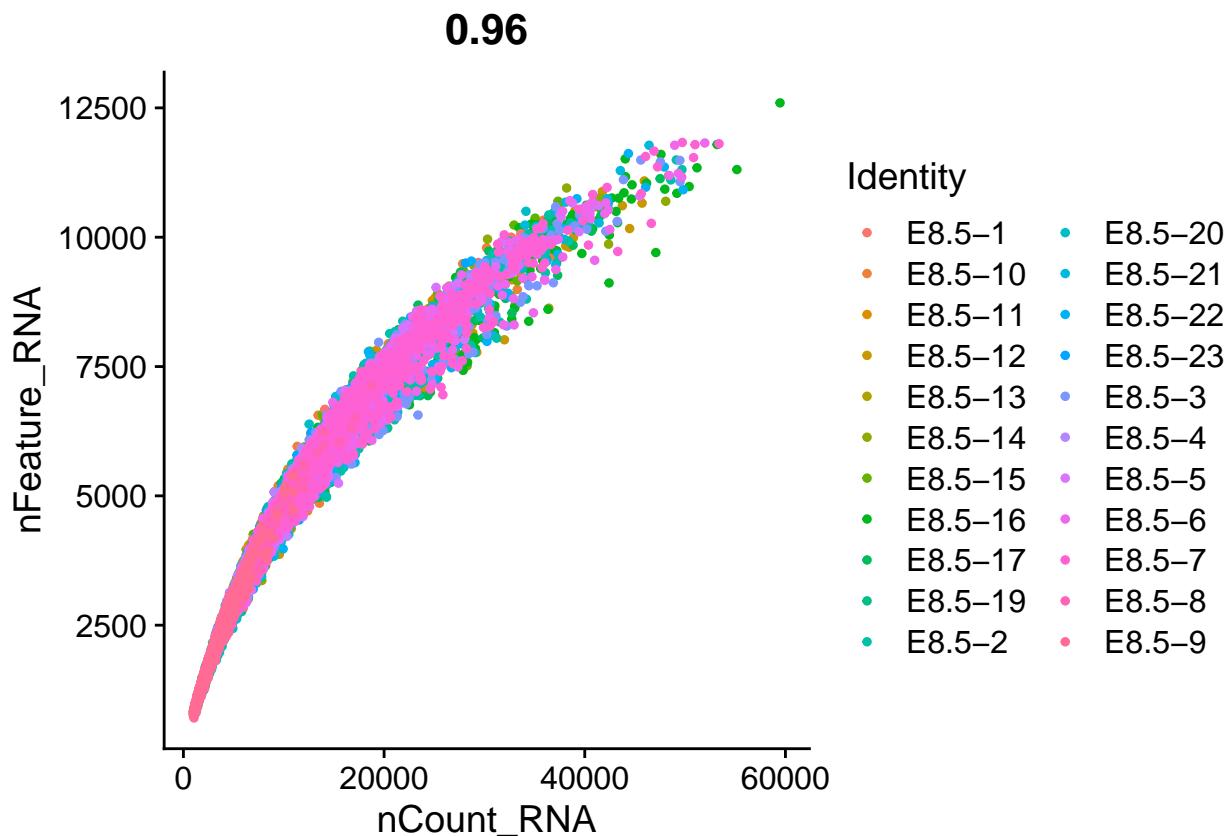
Het percentage mitochondriale expressie wordt handmatig toegevoegd aan het seurat object

Deelvraag: Welke filtering is geschikt voor de data?

Stap 3.1: Kwaliteit check met visualisatie in violin plot van nFeature_RNA (unieke genen per cel), nCount_RNA (totaal aantal moleculen), en percent.mt (mitochondriale expressie).



Stap 3.2: Visualisatie in Shatterplot



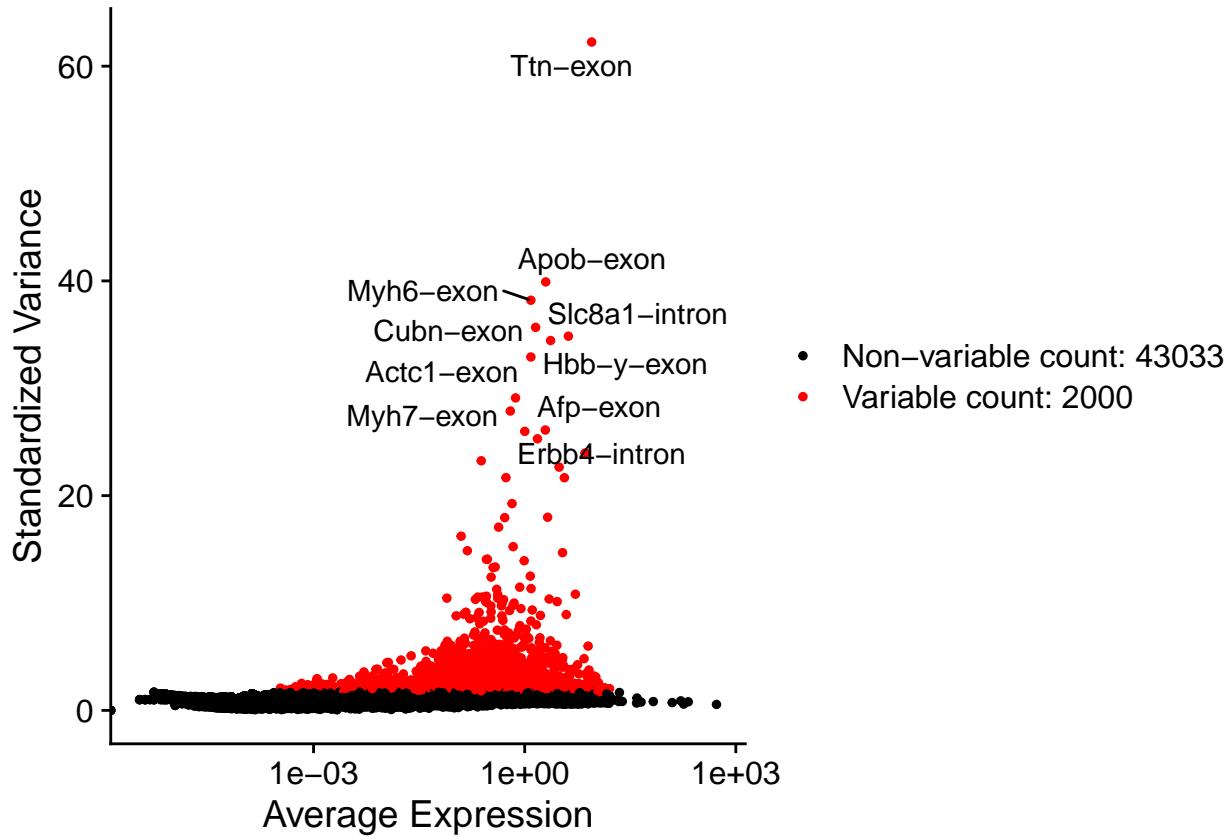
Deel conclusie: Op basis van deze figuren is gekozen voor de volgende filtering $nFeature_RNA < 2000 > 12000$ zo worden de lege en dubbele metingen niet meegenomen. $\text{percent.mt} < 5$ zo worden cellen die dood zijn of stress hebben niet meegenomen.

Stap 3.3: Filteren van de data

Stap 4.1: Normalisatie van de data

Stap 4.2: Variabele genen vinden

Stap 4.3: Variable genen plotten

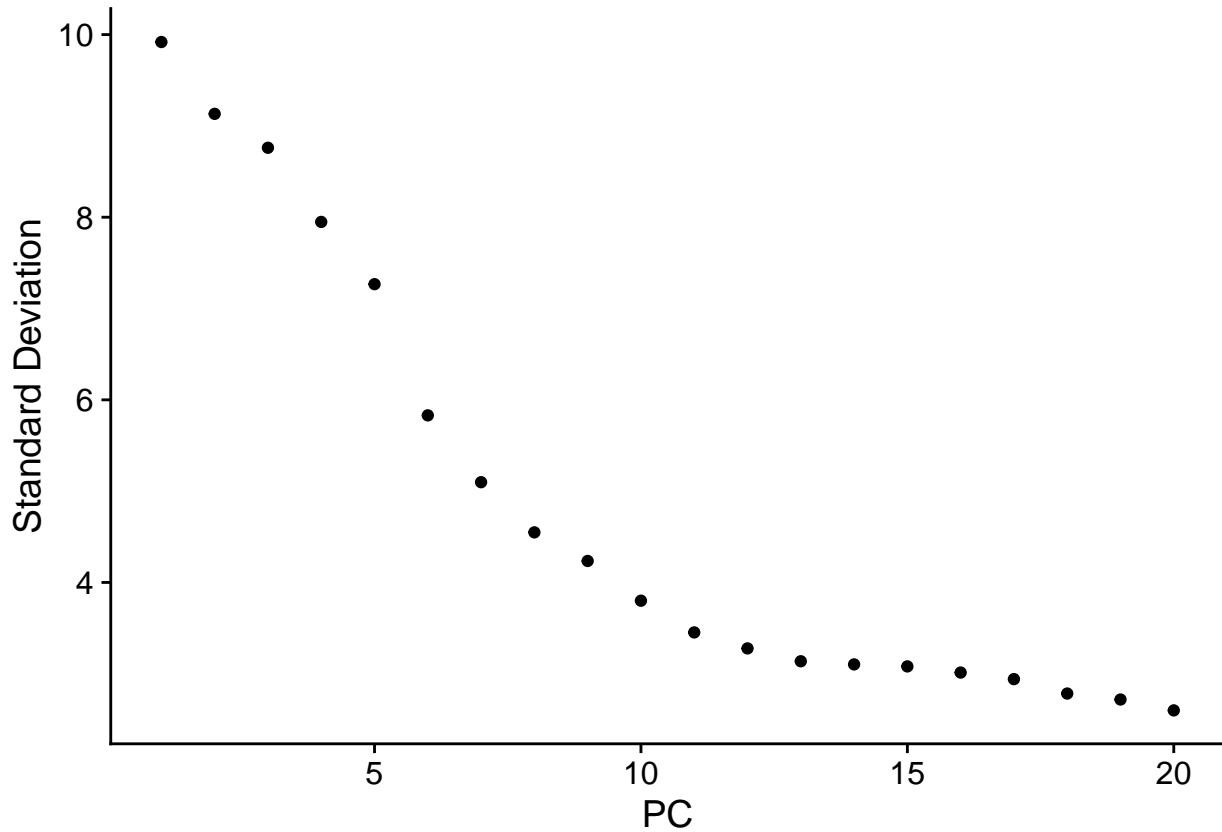


Stap 4.4: Schalen van genen. Er wordt een lijst gemaakt van alle gebruikte genen, deze krijgen allemaal een gemiddelde en standaard deviatie van de expressie

Deelvraag: Hoeveel PCAs worden meegenomen in verdere analyse?

Stap 5.1: PCA analyse uitvoeren

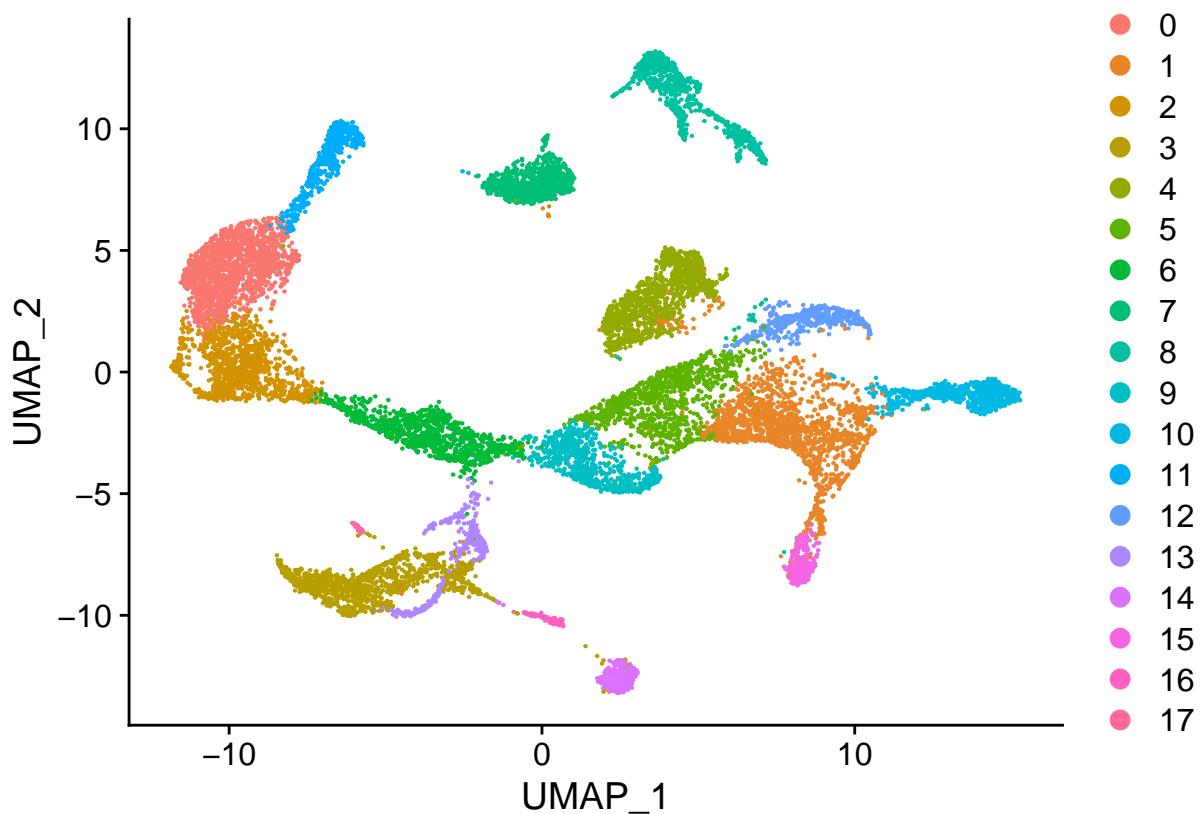
Stap 5.2: PCA analyse plotten



Deel conclusie: Vanuit de eerder gevolgde tutorial is geadviseerd om 10 PC's mee te nemen in cluster analyse voor betrouwbaar resultaat. Ook is te zien in de elbowplot dat de "elbow" afneemt na 10 PC's. Daarom zullen de eerste 10 PC's worden meegenomen in de verder cluster analyse.

Stap 6.1: Clusteren Er zal cluster analyse worden uitgevoerd op de eerste 10 PC's.

Stap 7.1: UMAP clusters



In de UMAP worden 17 verschillende clusters weergegeven.

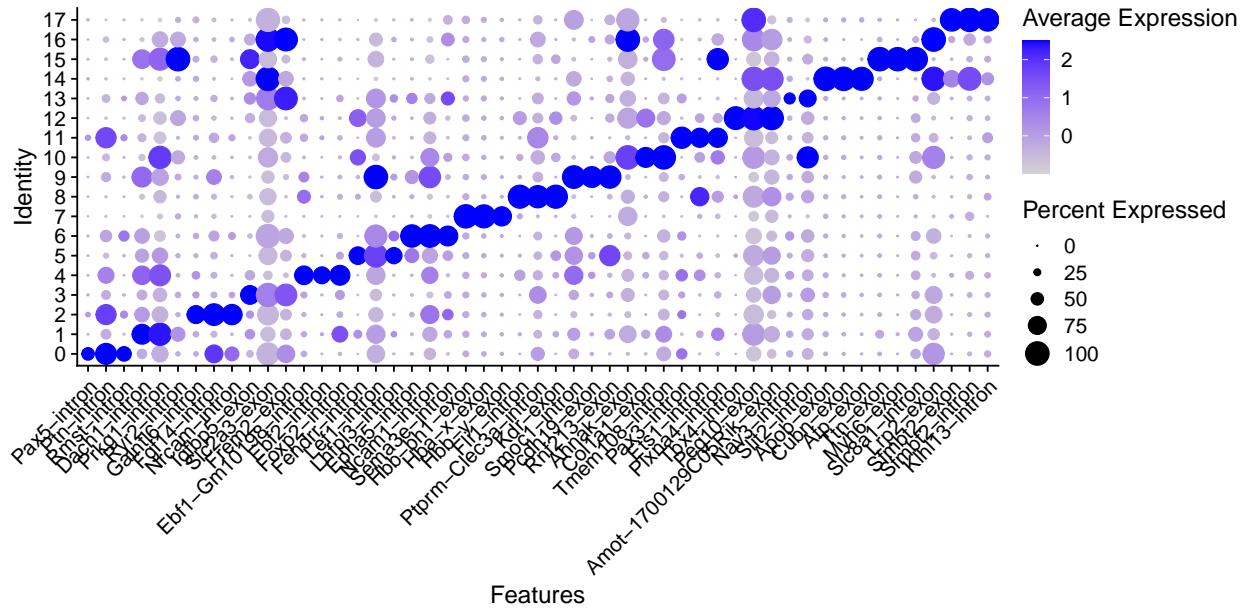
Deelvraag: Welke clusters worden geselecteerd voor analyse in BRIE2?

Om te kunnen bepalen welke clusters meegenomen worden in analyse met BRIE2 zal nu verder gekeken worden naar de biomarkers.

Stap 8.1. Biomarkers markers vinden in alle clusters en alleen de positieve worden gerapporteerd

Stap 8.2 Top 3 genen per cluster selecteren

Stap 8.3: Top 3 genen per cluster plotten



Omdat met BRIE2 gekeken kan worden naar exon skipping zal gekeken worden in welke clusters genen met variatie in expressie in exonen voorkomt

Stap 8.4: Exon genen selecteren

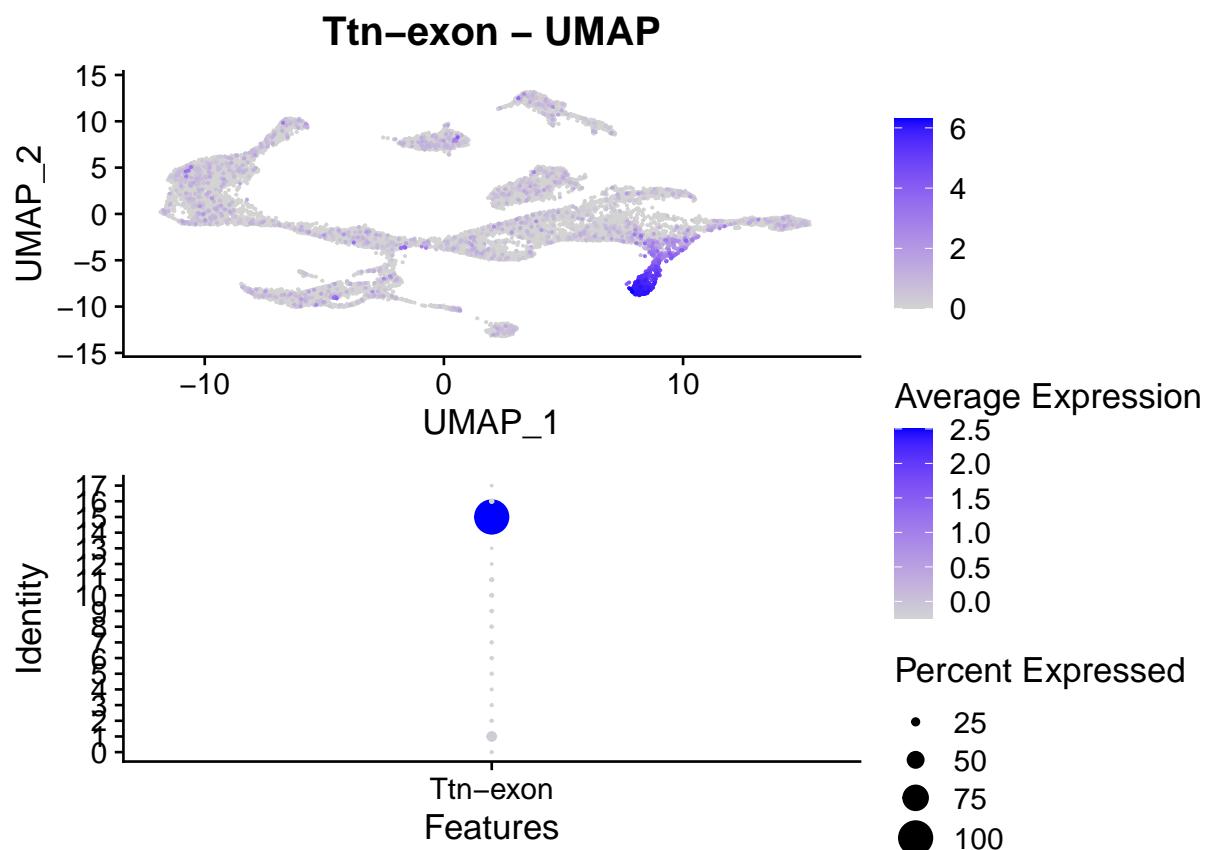
Stap 8.5: Gemiddelde expressie per cluster berekenen

Stap 8.6: Variatie per exon tussen clusters berekenen

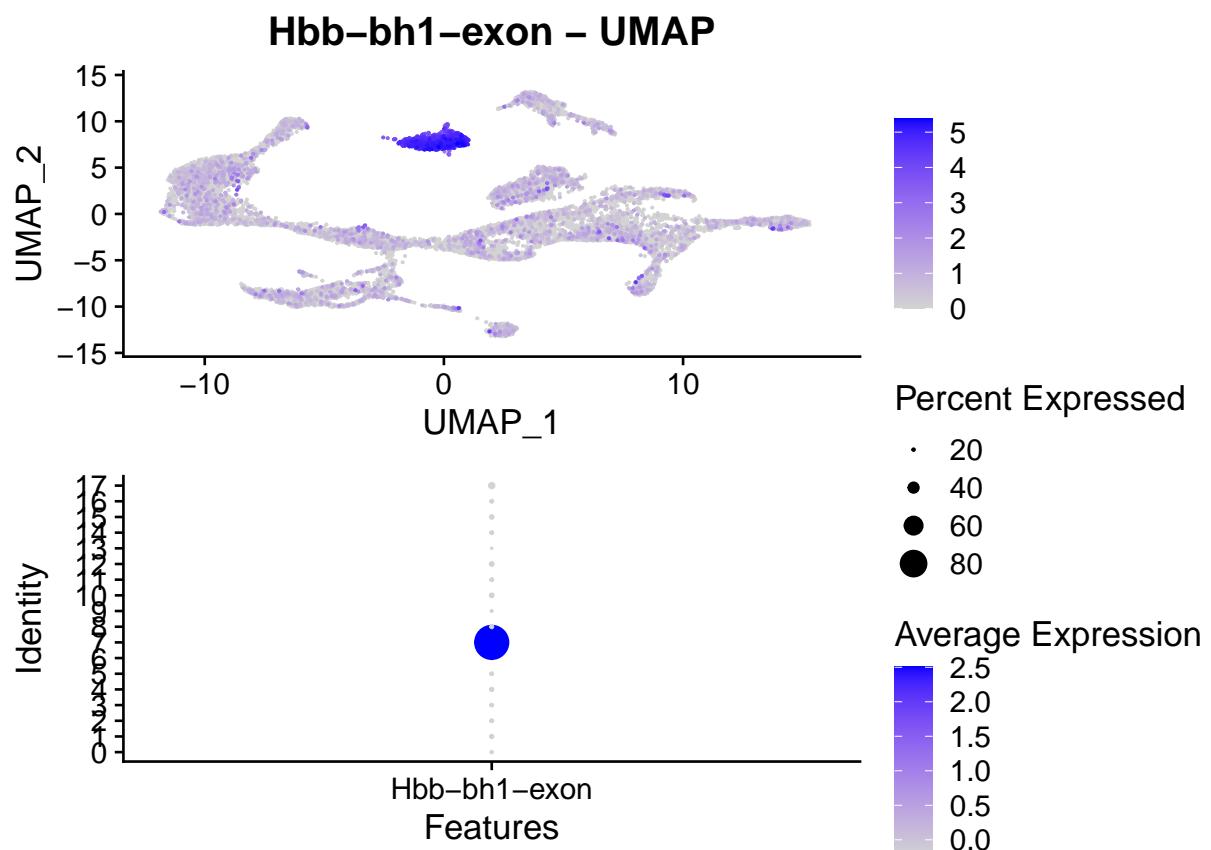
Stap 8.7: Top 5 exonen selecteren

Stap 8.7: Verschil tussen clusters aantonen en visualiseren

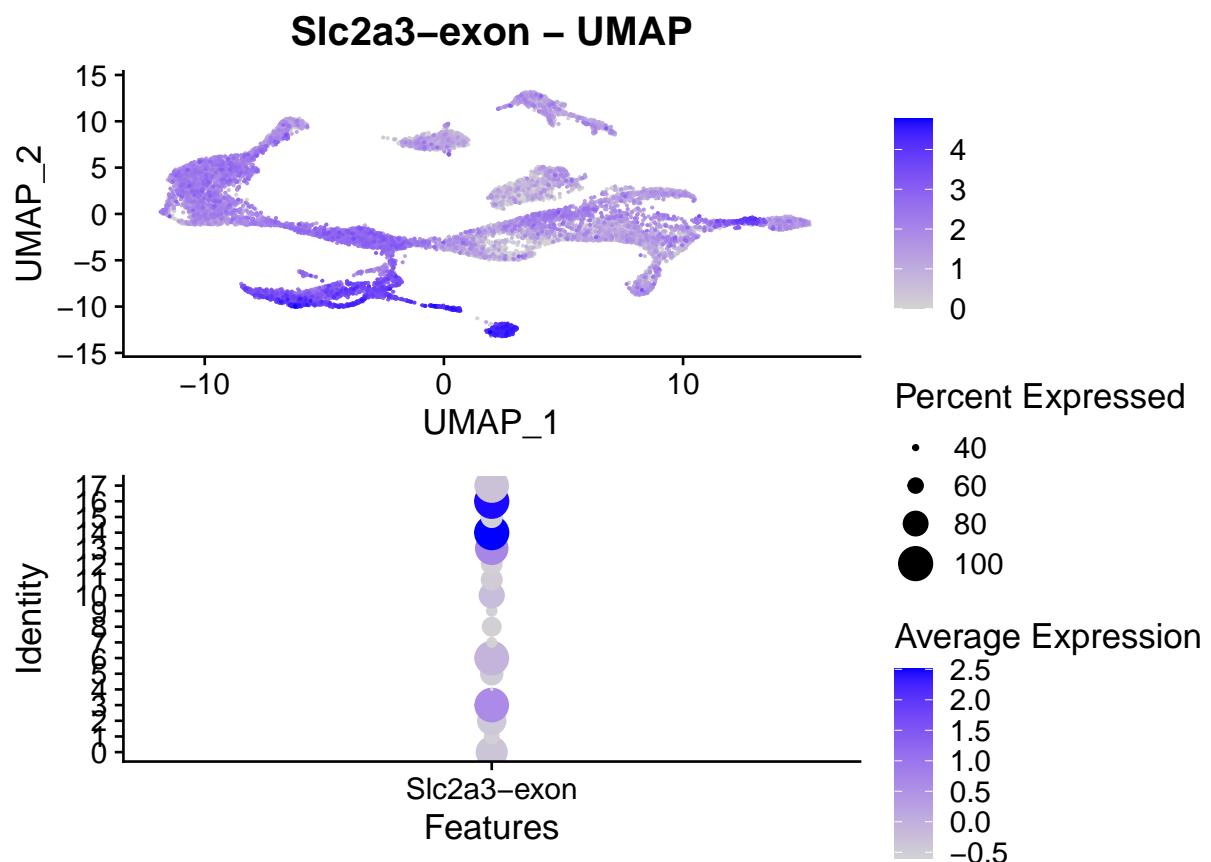
```
##  
## Ttn-exon: hoogste in cluster 15 (278.44), laagste in cluster 17 (0.04)
```



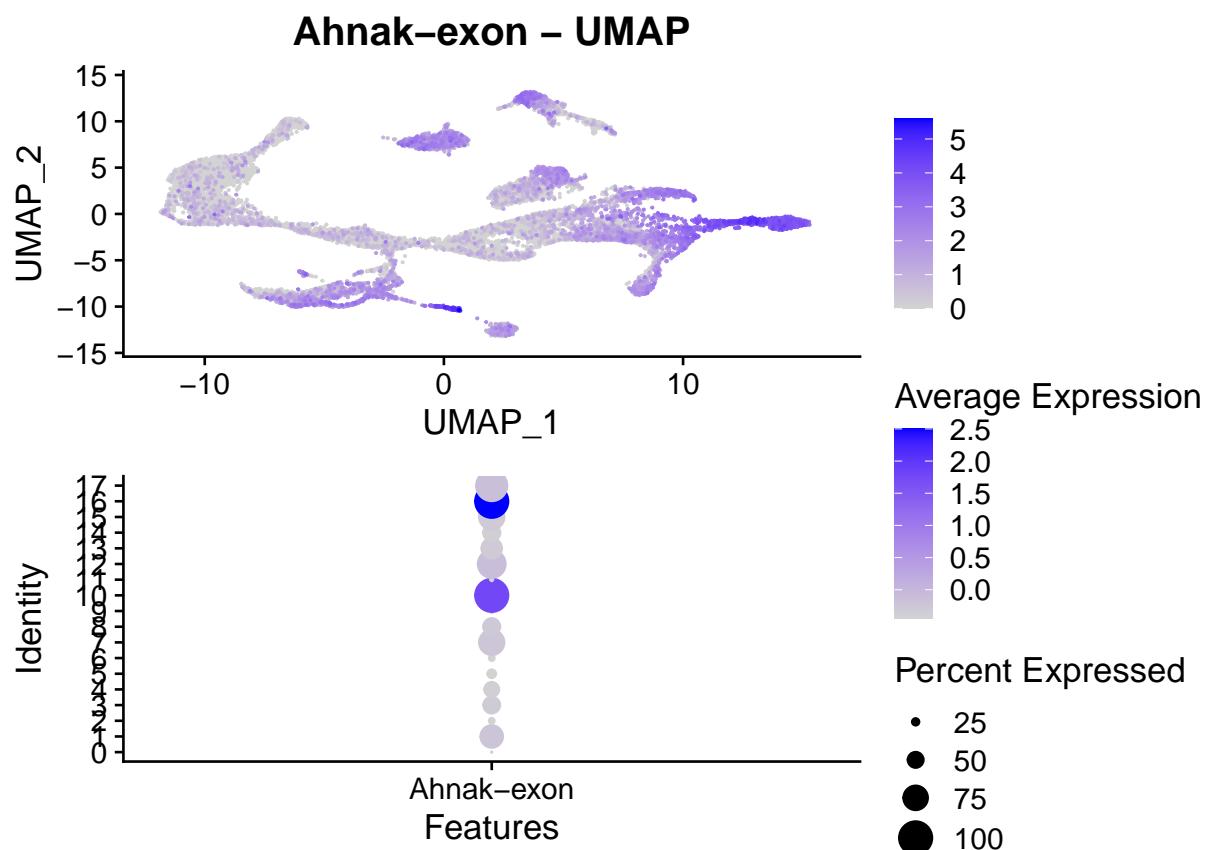
```
##  
## Hbb-bh1-exon: hoogste in cluster 7 (102.41), laagste in cluster 13 (0.26)
```



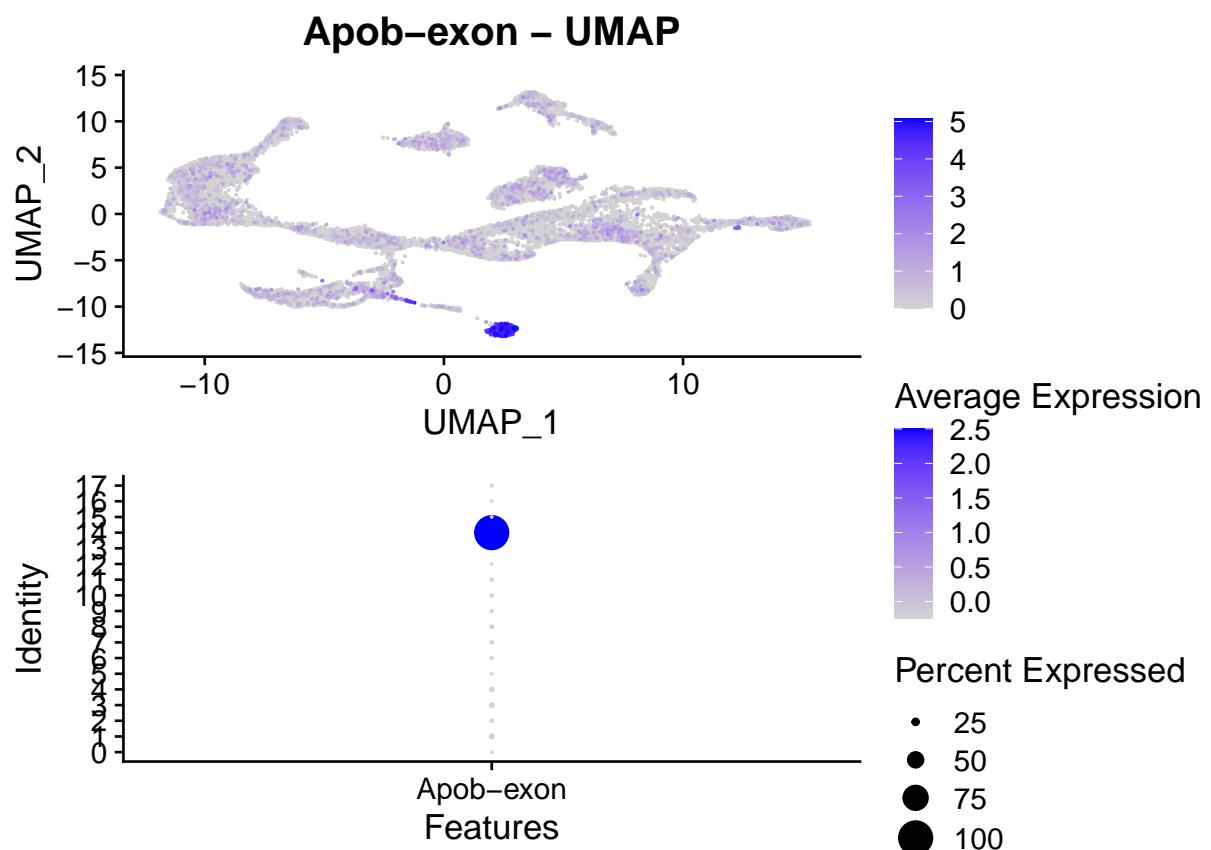
```
##  
## Slc2a3-exon: hoogste in cluster 14 (58.36), laagste in cluster 4 (0.55)
```



```
##
## Ahnak-exon: hoogste in cluster 16 (68.13), laagste in cluster 0 (0.11)
```



```
##  
## Apob-exon: hoogste in cluster 14 (70.34), laagste in cluster 5 (0.10)
```



Dee conclusie: Uit deze resultaten is gekozen voor cluster 14 en 4 op basis van gen Slc2a3-exon. Het exon gen heeft in 2 clusters expressie waardoor deze te vergelijken zijn. Dit is een gen wat gelinked wordt aan tumor ontwikkeling, en het gen is zeer onstabiel en gevoelig voor recombinante veranderingen.