

Seurat_filtering.Rmd

Anne Brussaard

2025-05-12

Om te kijken naar de effecten van filtering op een UMAP worden verschillende manieren van filtering met elkaar vergeleken.

Stap 1: Packages laden. De packages dplyr, ggplot2, pheatmap, tidyr, RColorBrewer, ggrepel, Seurat, Matrix, patchwork en here worden geladen. Deze packages worden met de library functie toegepast in analyse.

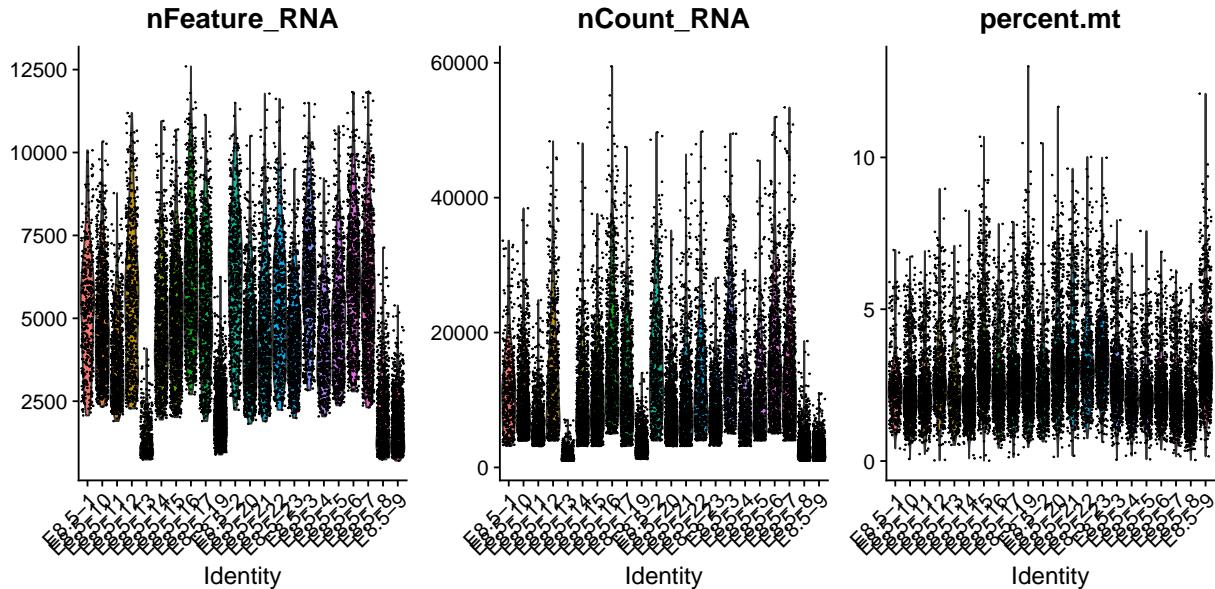
Stap 2: Data wordt geladen. De data wordt als csv bestand ingeladen en als object bewaard. De data is afkomstig van muis embryo's op dag E8.5.

Stap 3: MTX files worden geladen. Het object wordt samen met de matrix tabel ingeladen.

Stap 4: Seurat object maken. Met de packages seurat wordt een object gemaakt waardoor de data gebruikt kan worden voor preprocessing en data analyse.

Het percentage mitochondriale expressie wordt handmatig toegevoegd aan het seurat object omdat deze nodig is voor de kwaliteits check.

Stap 5: Kwaliteit check met visualisatie in violin plot van nFeature_RNA (unieke genen per cel), nCount_RNA (totaal aantal moleculen), en percent.mt (mitochondriale expressie). Dit wordt geplot in een violin plot zodat kan worden gekeken hoe de data is verdeeld. Een laag aantal unieke genen of moleculen kan namelijk wijzen op lege of slechte kwaliteit van de cellen. Een te hoog aantal kan juist wijzen op dubbele metingen. Veder is het mitochondriaal percentage belangrijk omdat een hoog percentage hiervan wijst op dode of slechte kwaliteit cellen.



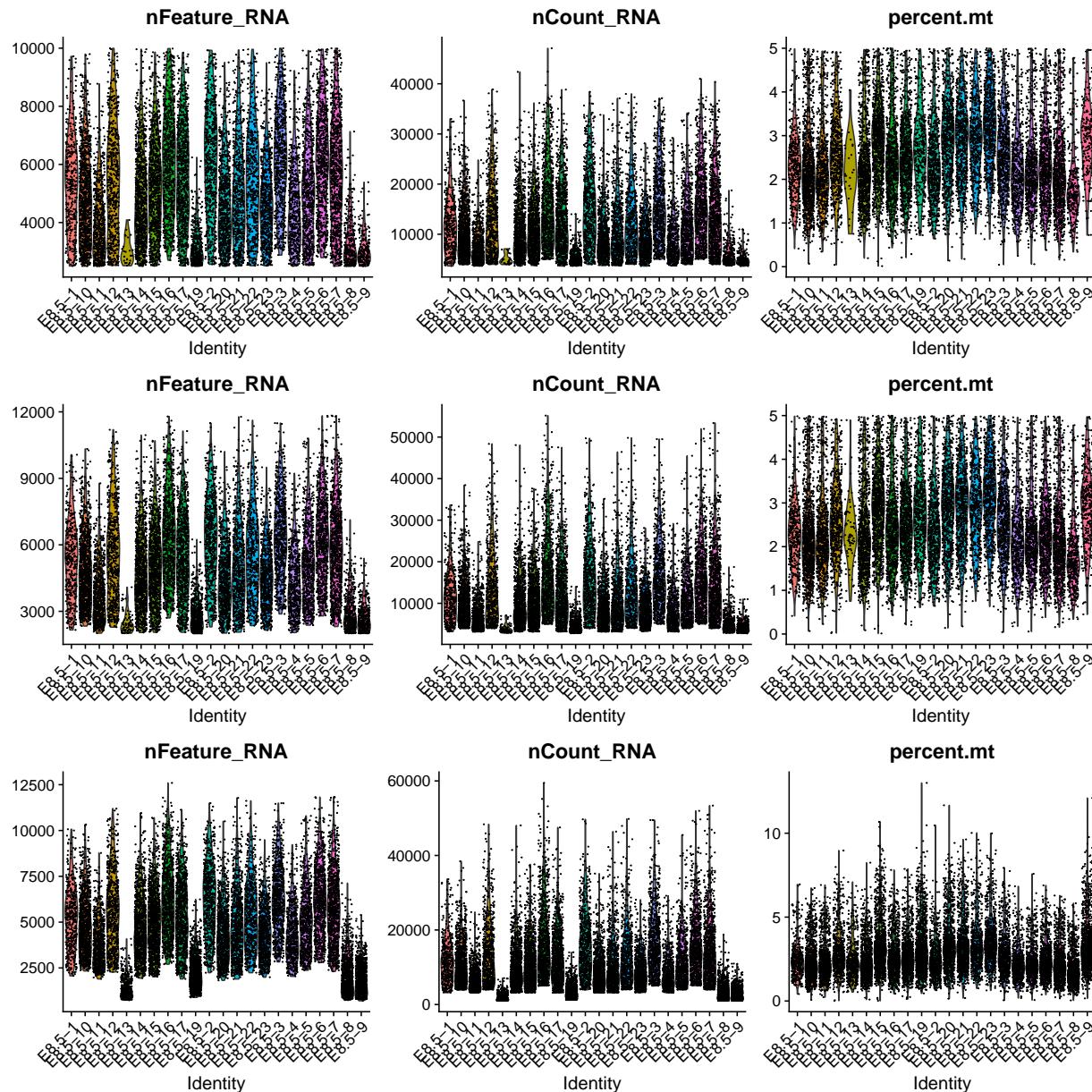
In deze violin plots wordt weergeven hoe de data verdeeld is.

Om te bepalen welke data er meegenomen wordt in verdere analyse worden er verschillende manieren van filteren getest.

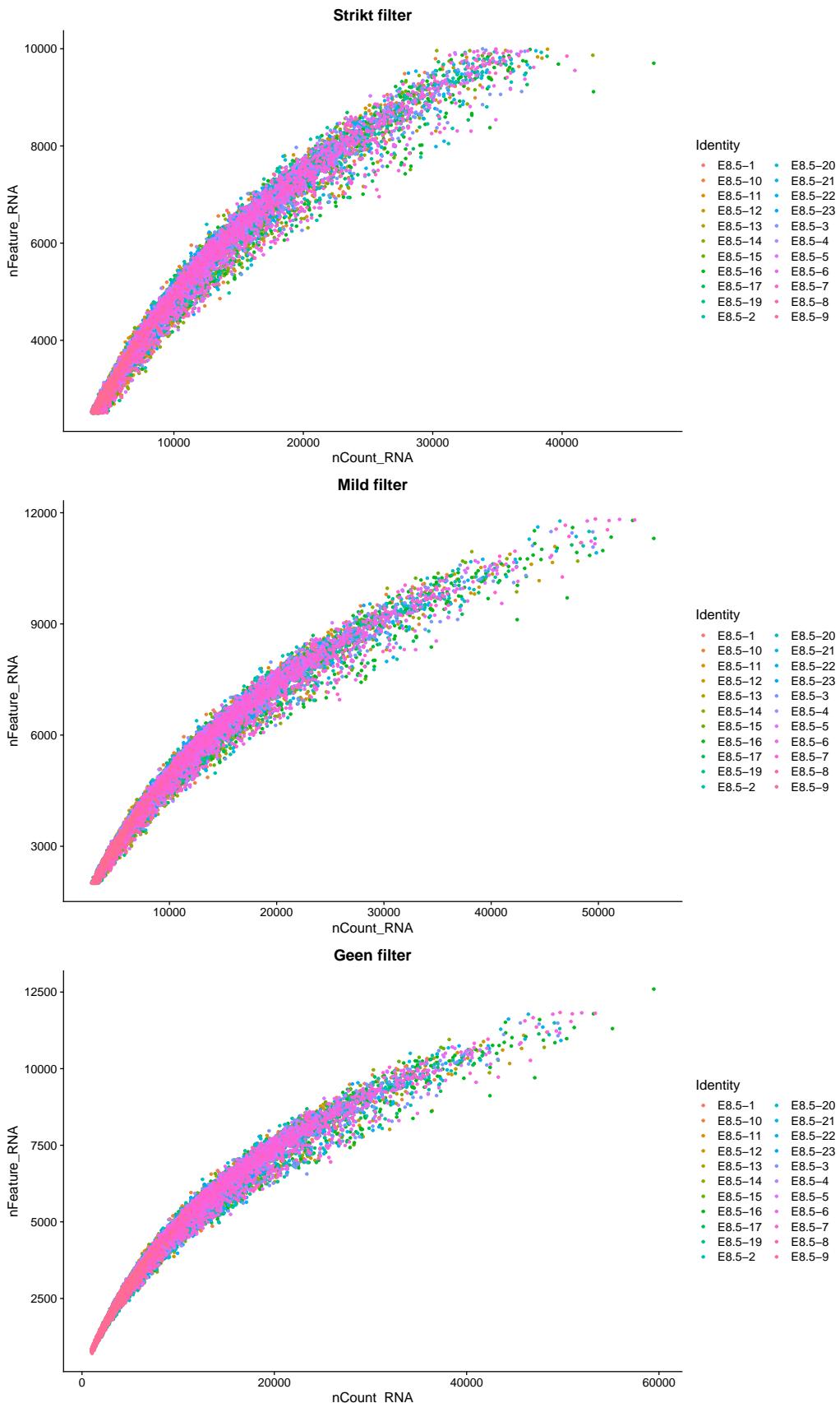
Stap 6. De data kan vervolgens op verschillende manieren worden gefilterd nu is gekozen voor de opties strict, mild en geen filter. Dit is gedaan op basis van de violin plot uit stap 5.

Er is gekozen voor strikt filtering, milde filtering en geen filtering. Dit wordt gedaan te kijken welke filtering het beste is voor de data.

Stap 7. Opnieuw worden de violin plots gemaakt om het effect van de filter stappen te zien op de data.



Stap 8. Ook wordt een feature scatter geplot om te kijken naar de correlatie, en of er uitschieters aanwezig zijn.



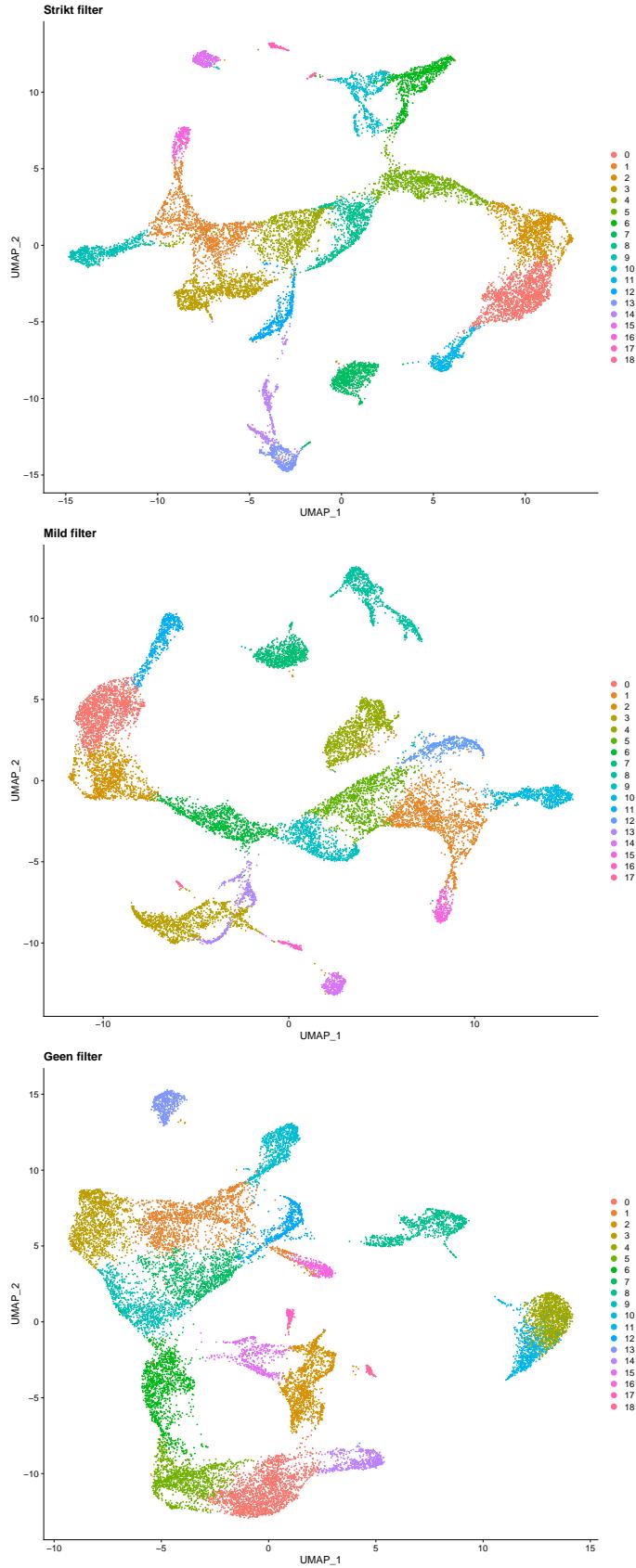
Stap 9. Vervolgens wordt een functie geschreven om de de verder seurat analyse uit te voeren op de verschillende filter mogelijkheden.

Stap 10. Deze functie wordt uitgevoerd op de 3 filter opties.

Stap 11. Van de 3 verschillende filter opties worden UMAPs gemaakt, om het verschil te laten zien.

Stap 12. De UMAPs worden samengevoegd en weergeven.

Vergelijking van filtering effect op UMAP
Strikt, Mild, Geen filtering



Conclusie: Op basis van deze UMAPS is gekozen voor filter optie mild. Dit omdat met deze manier van filteren zo veel mogelijk data wordt meegenomen wat bijdraagt aan de betrouwbaarheid van de verdere analyse. Bij deze optie zijn er ook duidelijke clusters te zien in vergelijking met de geen filter optie. Ook kunnen door de strenge filtering misschien cellen worden weg gefilterd die zorgen voor interessante variatie.

Discussie: Strengere filtering kan wel zorgen voor minder ruis wat kan zorgen voor een duidelijker biologisch verschil. Daarom zou na het ontwikkelen van de analyse pipeline van brie2 is ontwikkeld om zo andere clusters te kunnen vergelijken.