

# 02\_seurat\_data\_integration\_tutorial

2025-11-21

De seurat tutorial genaamd “Introduction to scRNA-seq integration” is uitgevoerd in deze RMarkDown. De dataset die in deze tutorial is gebruikt bestaat uit twee verschillende condities van humane PBMC’s, interferon-gestimuleerde PBMC’s en controle cellen. Het doel van deze tutorial is om de twee datasets samen te voegen om gezamenlijke cel subpopulaties binnen de datasets te kunnen identificeren. Vervolgens kan er onderzocht worden hoe de groepen verschillen tussen de condities.

## Analyse zonder data integratie

De dataset wordt eerst zonder integratie geanalyseerd. De resultaten laten clusters zien die bepaald worden door zowel celtype als stimulatie conditie. Dit levert problemen op tijdens verdere analyse.

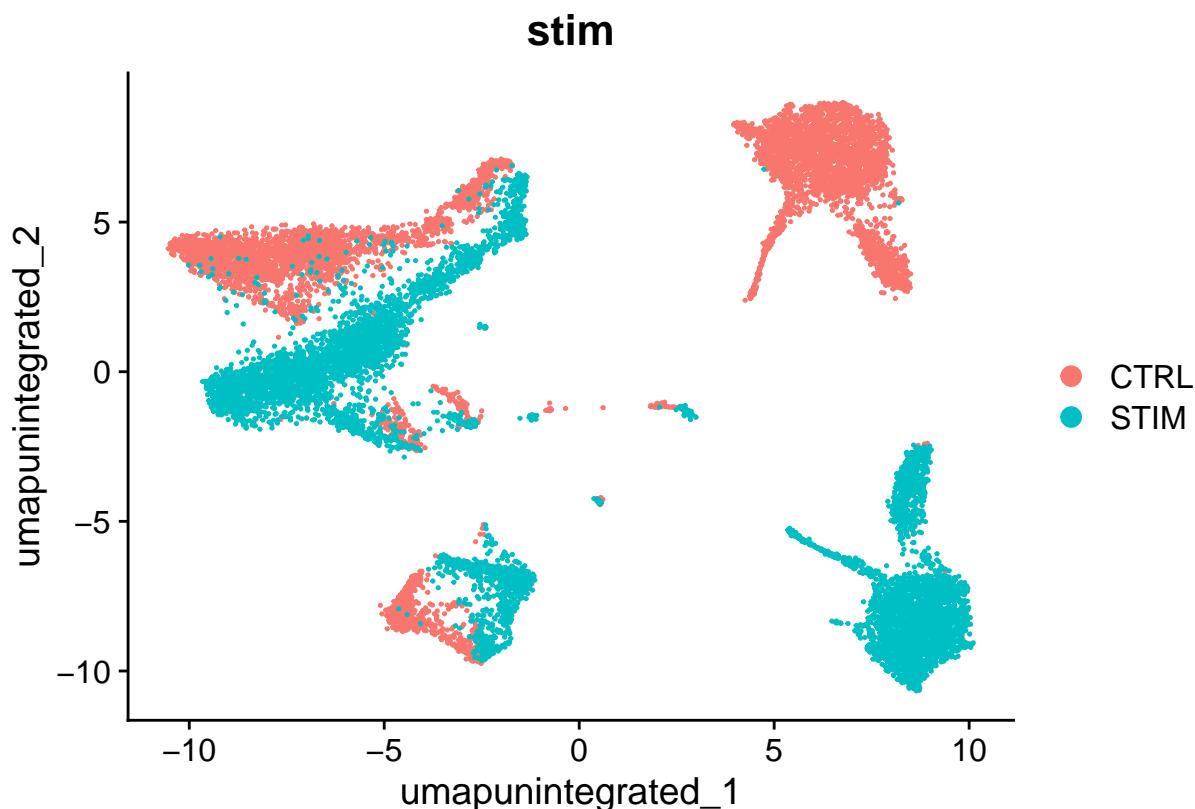


Figure 1: Stim: Elke cel wordt weergegeven als een stipje. Controle is rood en gestimuleerde cellen zijn blauw. Seurat\_clusters: Elke cel wordt weergegeven als een stipje. Er zijn 25 clusters weergegeven met elk een eigen kleur.

## Integratie uitvoeren

Nu worden beide condities samengevoegd waarbij cellen met hetzelfde celtype/subpopulatie gaan clusteren.

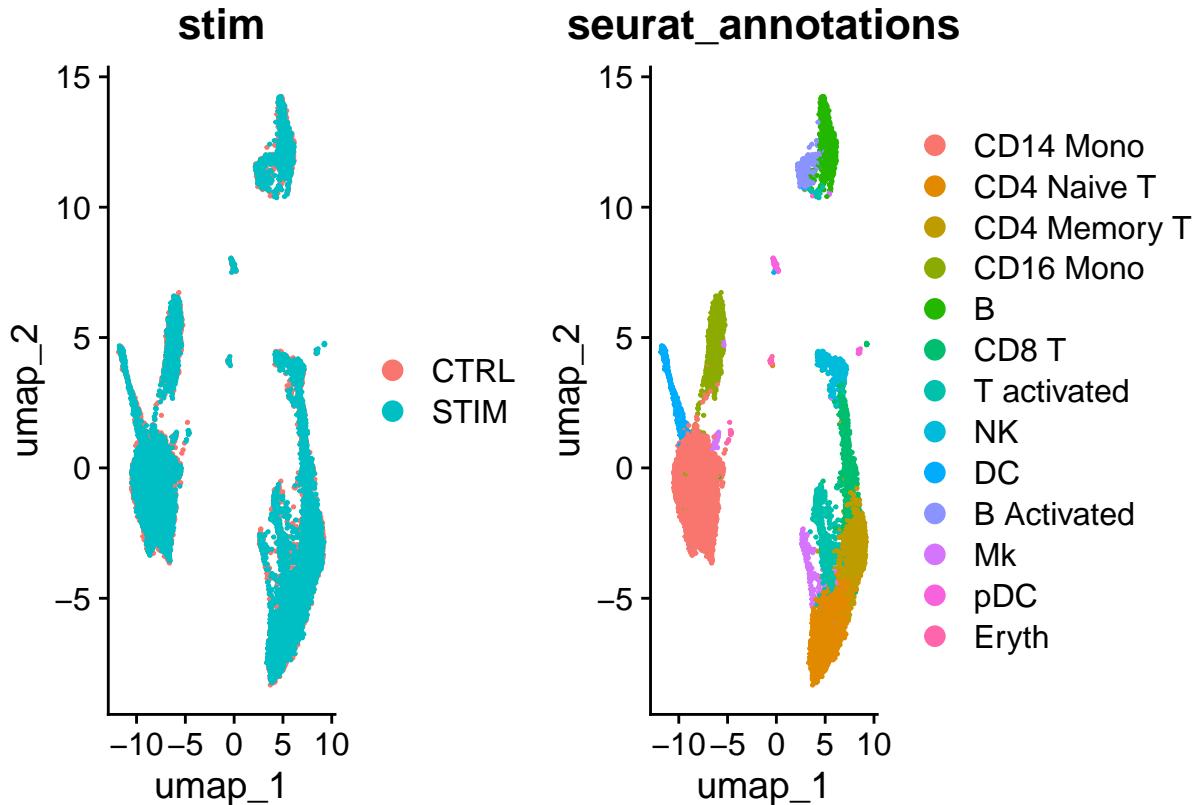


Figure 2: Stim: Elke cel wordt weergegeven als een stipje. Controle is rood en gestimuleerde cellen zijn blauw. Seurat\_annotations: Elke cel wordt weergegeven als een stipje. Er zijn 13 clusters weergegeven met elk een eigen kleuren celtype.

## Identificeer geconserveerde celtypemarkers

Er zijn een aantal celtypemarkers die gebruikt worden om de celtypes te kunnen onderscheiden. De celtypemarkers die in beide groepen aanwezig zijn worden gezocht. Vervolgens worden deze celtypemarkers gebruikt om de verschillende celtypen aan te wijzen binnen de clusters.

## Identificeer differentieel tot expressie gebrachte genen in verschillende omstandigheden

Nadat de gestimuleerde cellen en de controle cellen samengebracht zijn, kunnen er analyses uitgevoerd worden waarbij de twee met elkaar vergeleken kunnen worden. De verschillen die door de stimulatie veroorzaakt worden kunnen dan bekijken worden.

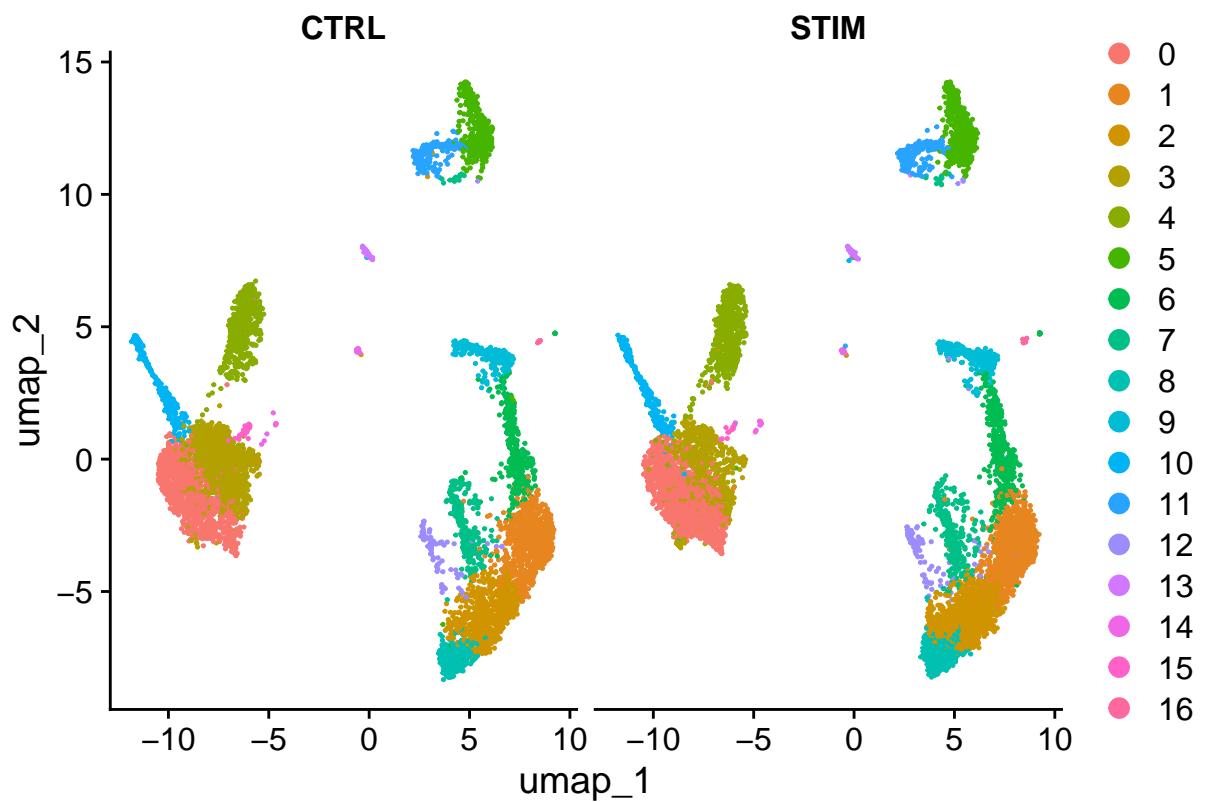


Figure 3: CTRLSTIM: Controle en gestimuleerde cellen worden naast elkaar weergegeven. Elke cel wordt weergegeven als een stipje. Er zijn 16 clusters weergegeven met elk een eigen kleur.

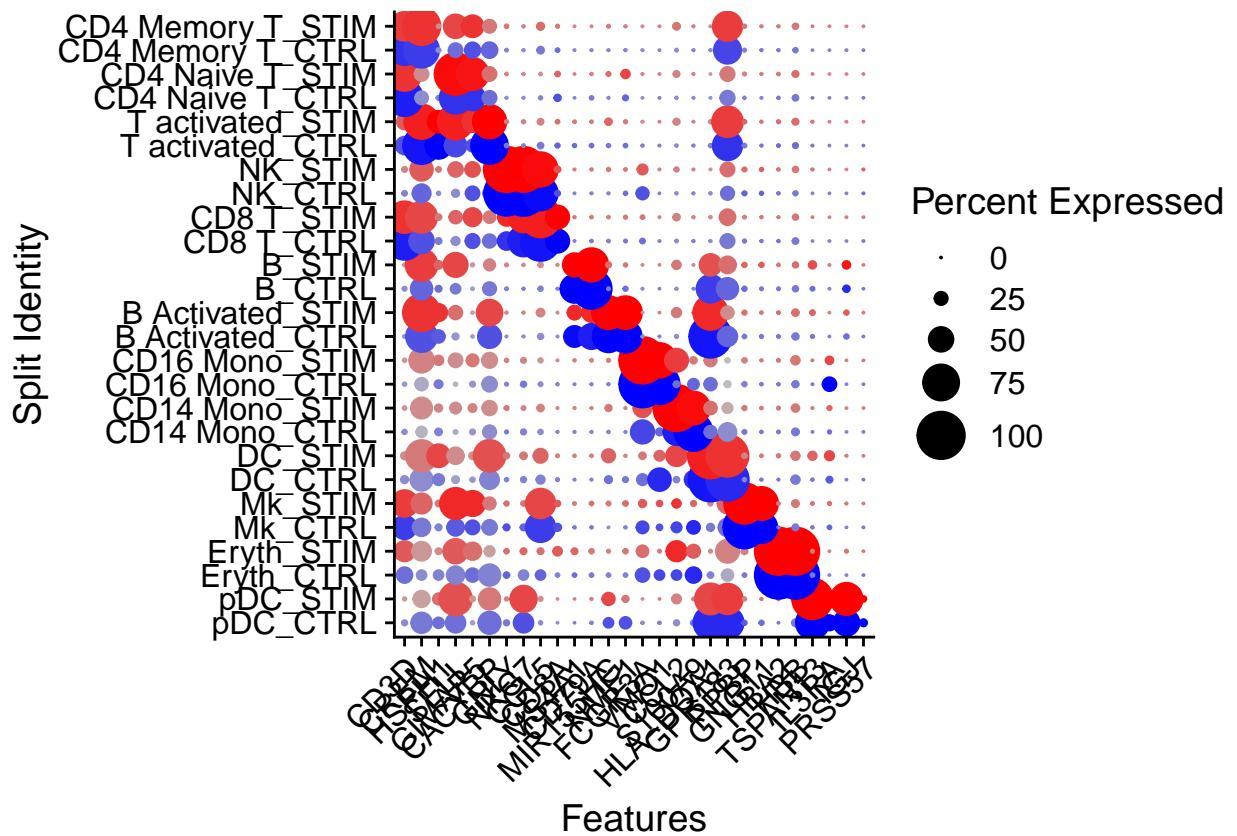


Figure 4: DotPlot: Op de x-as zijn de verschillende features weergegeven. Op de y-as zijn de verschillende celtypes weergegeven, gescheiden in gestimuleerde cellen en de controle cellen. De grootte van de dots geven aan hoeveel procent expressie van de features aanwezig zijn in de celtype, waarbij grotere dots een hoger percentage betekend. Rode dots zijn van de gestimuleerde cellen en blauwe dots zijn van de controlegroep.

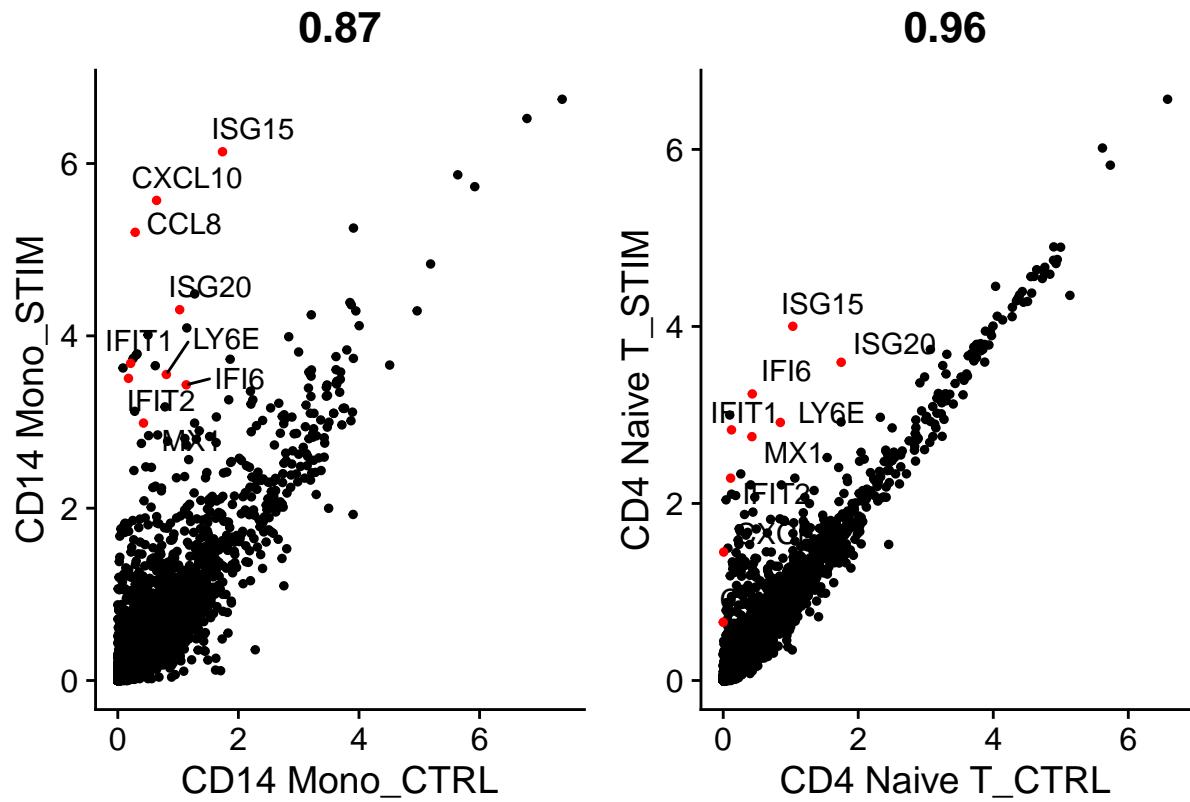
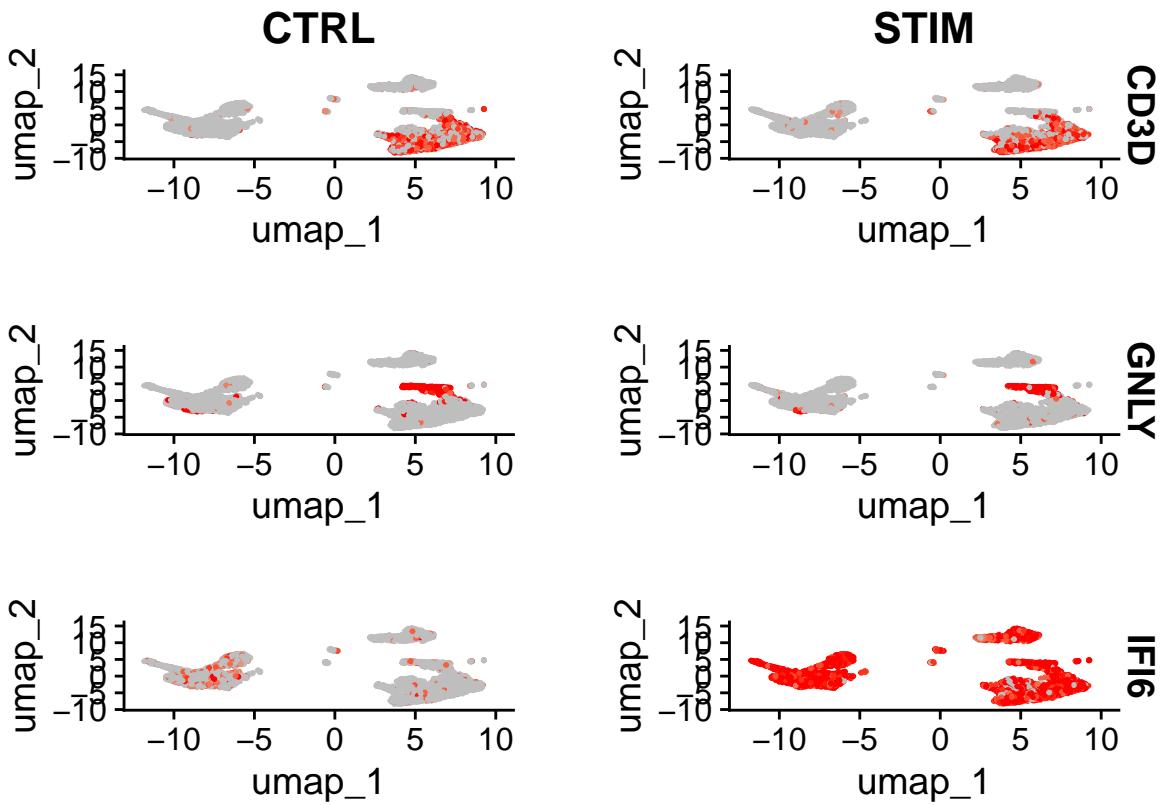
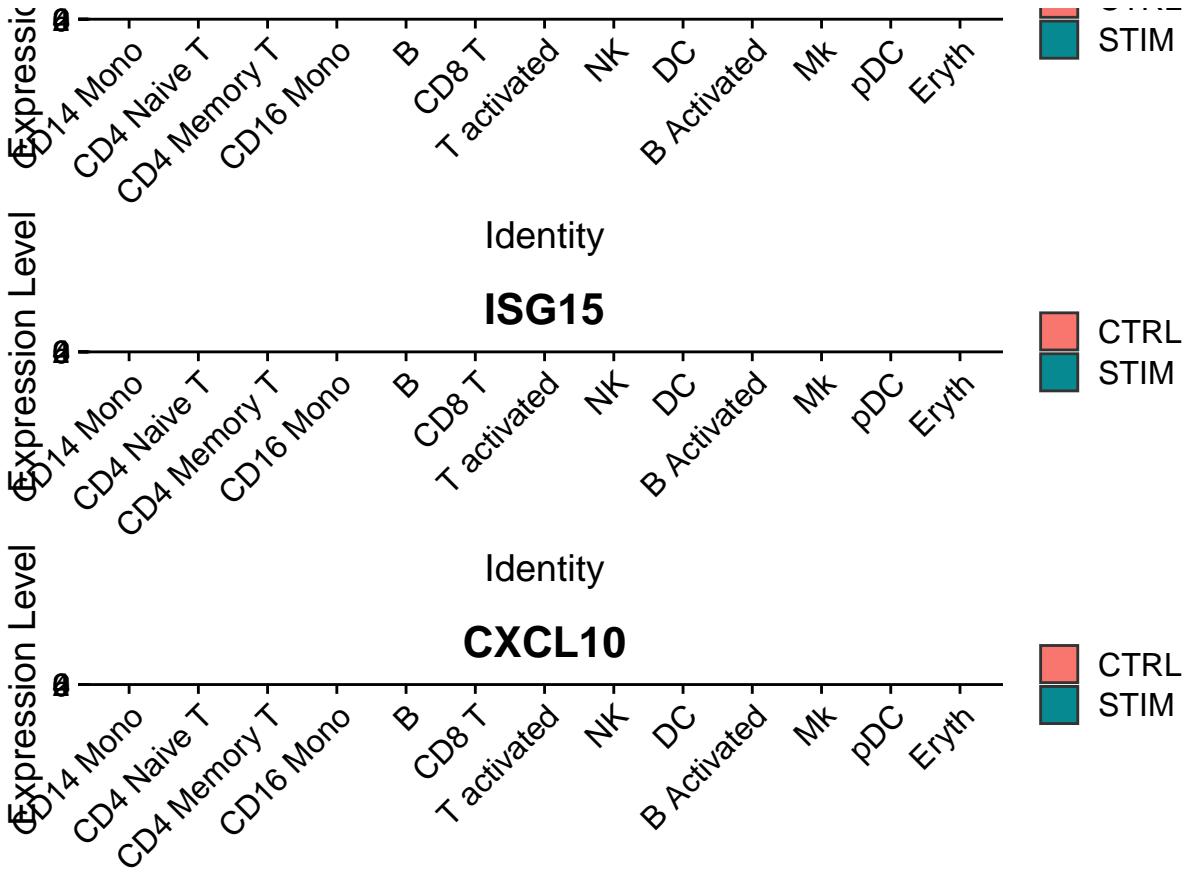


Figure 5: Scatterplots: Links is een scatterplot weergegeven van de genexpressieprofielen van de CD14 monocyten met op de x-as de controle cellen en op de y-as de gestimuleerde cellen. De genen die een sterke respons vertonen op interferonstimulatie zijn rood weergegeven en de naam is erbij gezet. Rechts geldt hetzelfde, maar dan voor de CD4 naïeve T-cellen.





## Integratie uitvoeren met SCTransform-genormaliseerde datasets

