

Transcriptomics: single cell RNAseq

Bioinformatics & Omics

Lernziele

- die Unterschiede zwischen Einzelzell-Sequenzieren und Bulk-Sequenzieren beschreiben: Anwendung, Methode, Analyse
- Technologien zur Erfassung einzelner Zellen listen und die entsprechende Ansätze zueinander vergleichen
- Schritte in Einzelzelle-Analyse motivieren, diese in R ausführen und Ausgaben interpretieren
 - Einzelzell-Transkriptomik-Matrix erstellen
 - Differentielle Expressionsanalyse mit PCA ausführen
 - Clusters mit UMAP und t-SNE bilden um die Kategorien zu identifizieren
- Gene und Pathway Enrichment Analyse durchführen
- Zwecke von räumlichen Einzelzell-Sequenzieren beschreiben

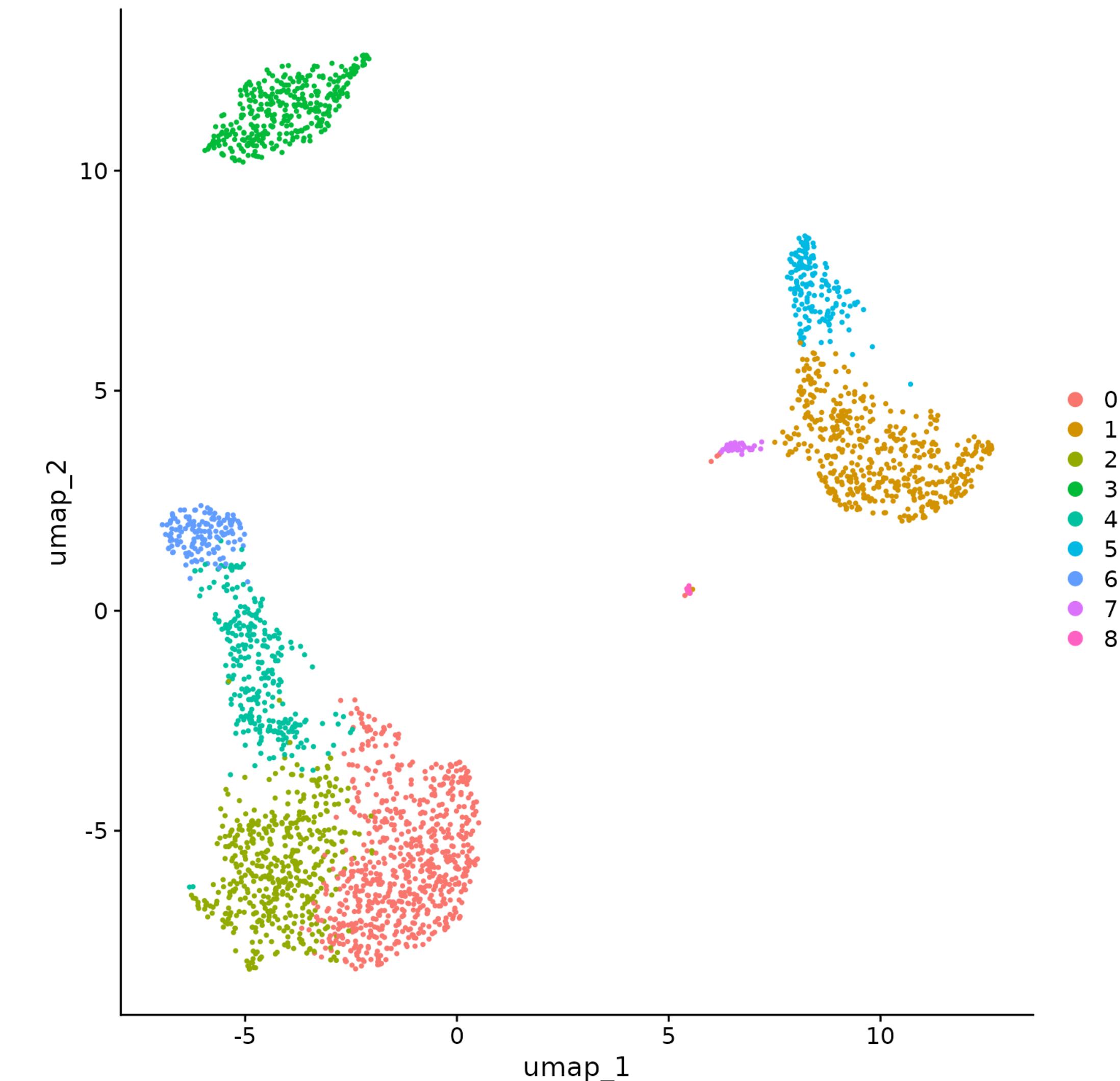
weiter mit Training in R

Für dieses Tutorial analysieren wir einen Datensatz von peripheren mononukleären Blutzellen (PBMC), der von 10X Genomics frei verfügbar ist. Es handelt sich um 2.700 Einzelzellen, die auf dem Illumina NextSeq 500 sequenziert wurden. Die Rohdaten und R-Skript sind online.

```
## 3 x 30 sparse Matrix of class "dgCMatrix"
##
## CD3D  4 . 10 . . 1 2 3 1 . . 2 7 1 . . 1 3 . 2  3 . . . . 3 4 1 5
## TCL1A . . . . . . . 1 . . . . . . . . . . 1 . . . . . . . .
## MS4A1 . 6 . . . . . 1 1 1 . . . . . . . 36 1 2 . . 2 . . .
```

3. Clusterbildung von Zellen

Seurat führt zunächst eine graphbasierte Clusterbildung durch und führt anschliessend eine nichtlineare Dimensionsreduktion (UMAP/tSNE) durch. Dadurch werden ähnliche Zellen anhand ihrer Genexpression gruppiert, um Gruppen von Zellen zu finden, die häufiger miteinander in Verbindung stehen als andere.



Clusterbildung

<https://pair-code.github.io/understanding-umap/>

Zwei Größen:

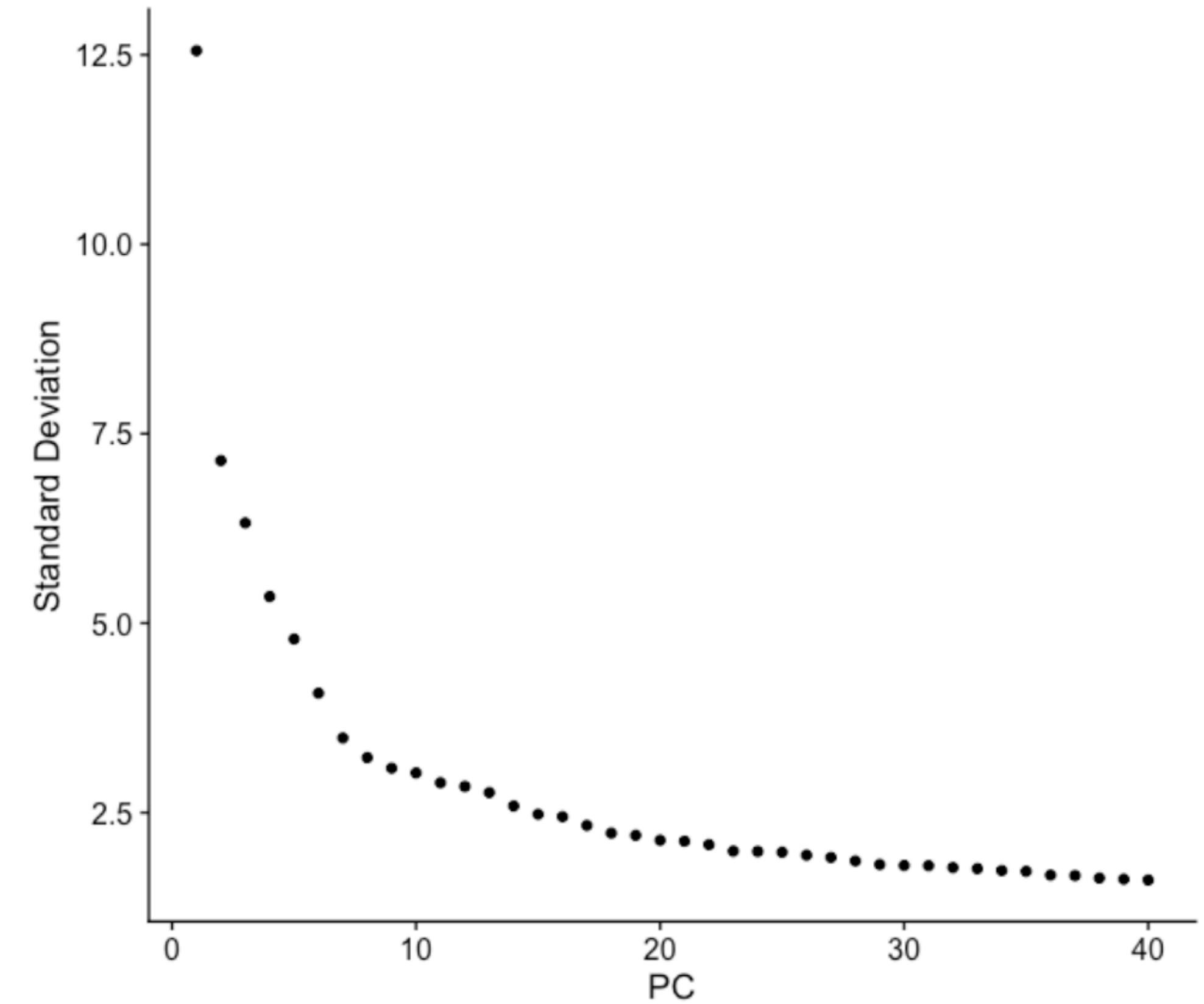
`n_neighbors`: Anzahl von Nachbarn

`min_dist`: minimale Distanz zwischen Nachbarn

Während die meisten Anwendungen von UMAP die Projektion aus hochdimensionalen Daten beinhalten, dient die Projektion aus 3D als nützliche Analogie, um zu verstehen, wie UMAP je nach seinen Parametern globale gegenüber lokalen Strukturen priorisiert.

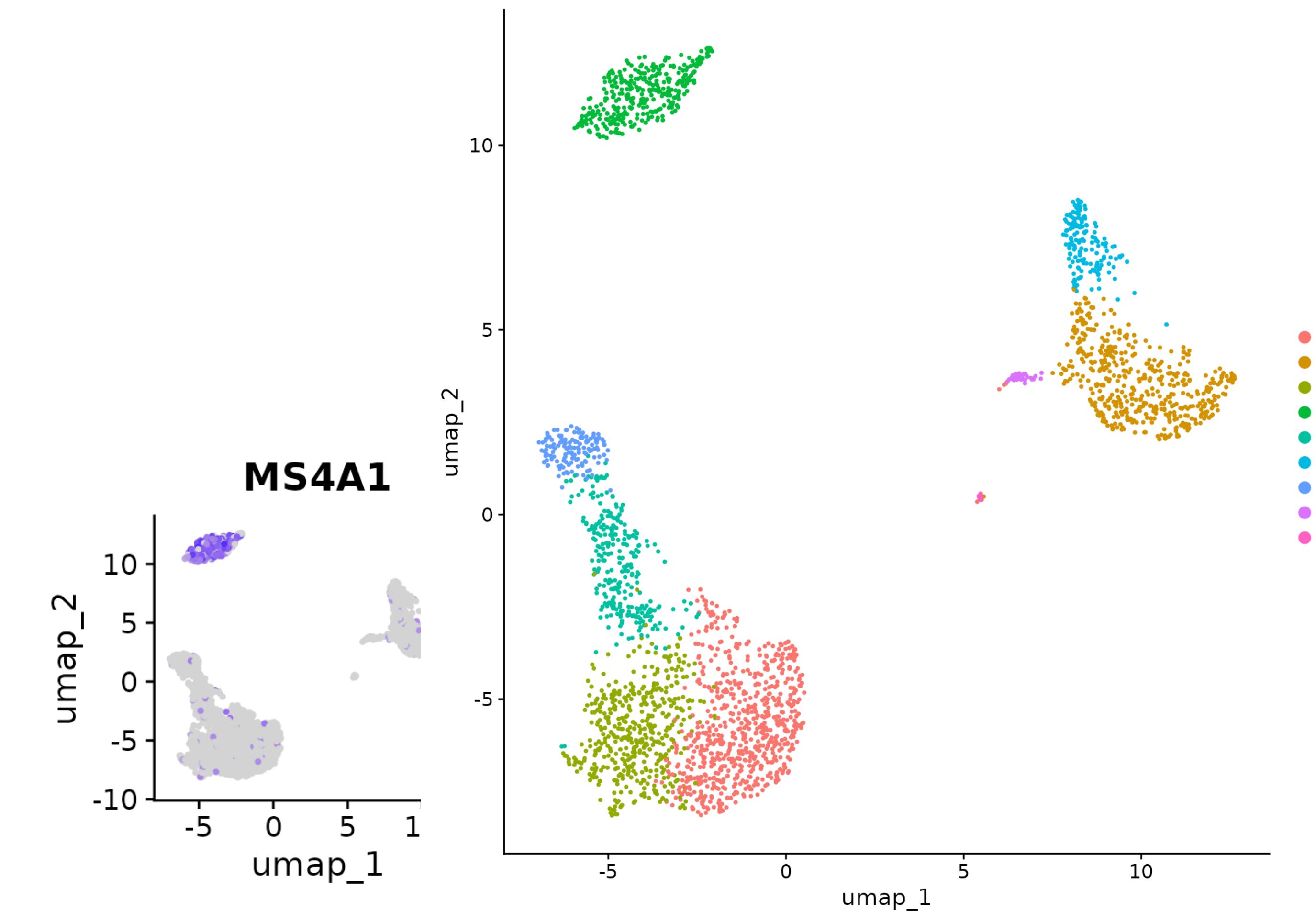
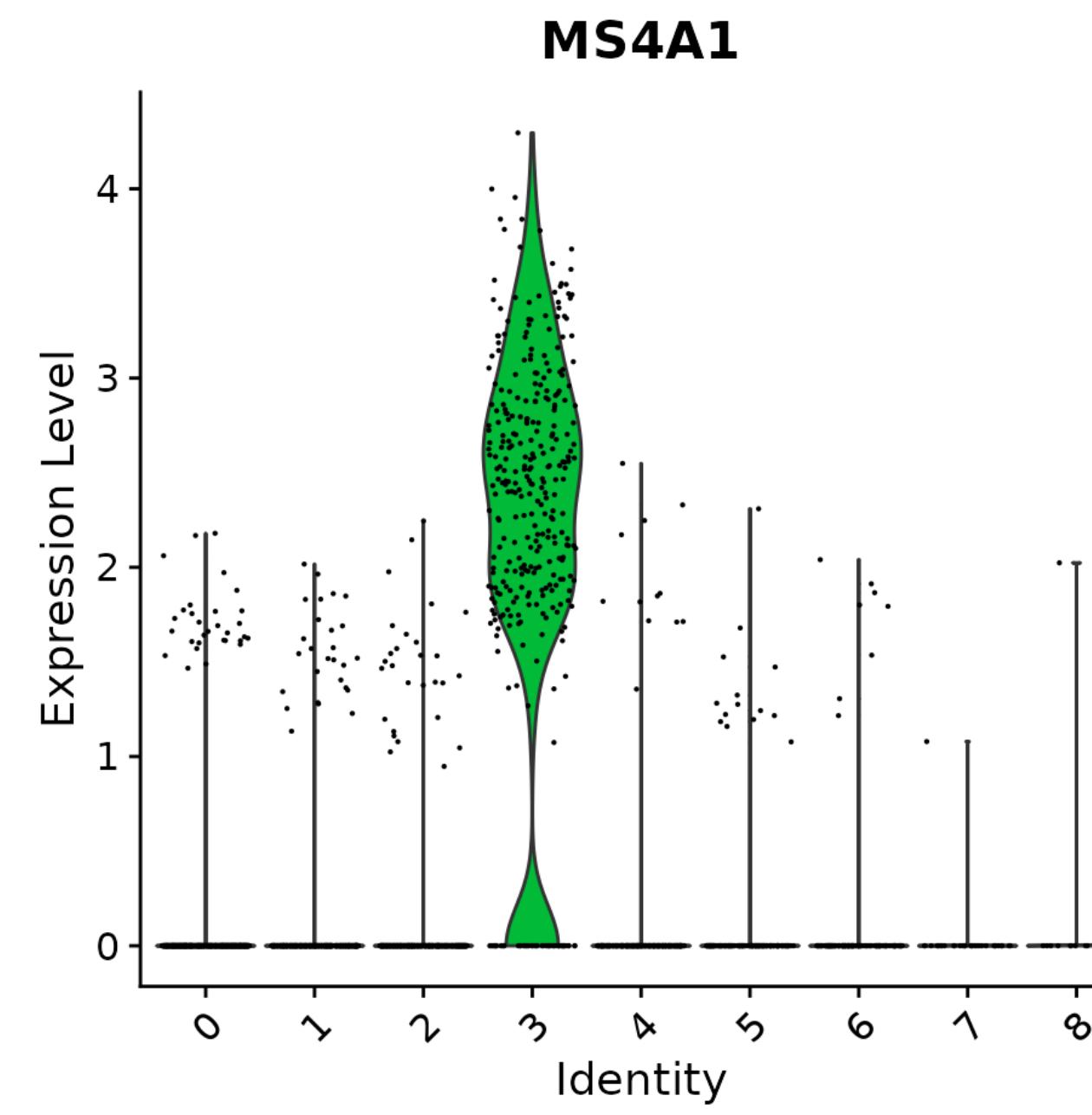
Wie viele Clusters?

Dazu verwenden wir die Darstellung der Hauptkomponenten aus der PCA und deren jeweilige Erklärungskraft. Der Punkt, an dem wir den Knick sehen, ist der Cut-out-Wert. Es ist nicht notwendig, weitere Dimensionen zu berücksichtigen, da diese keinen wesentlichen Beitrag zur Erklärung der Daten mehr leisten.



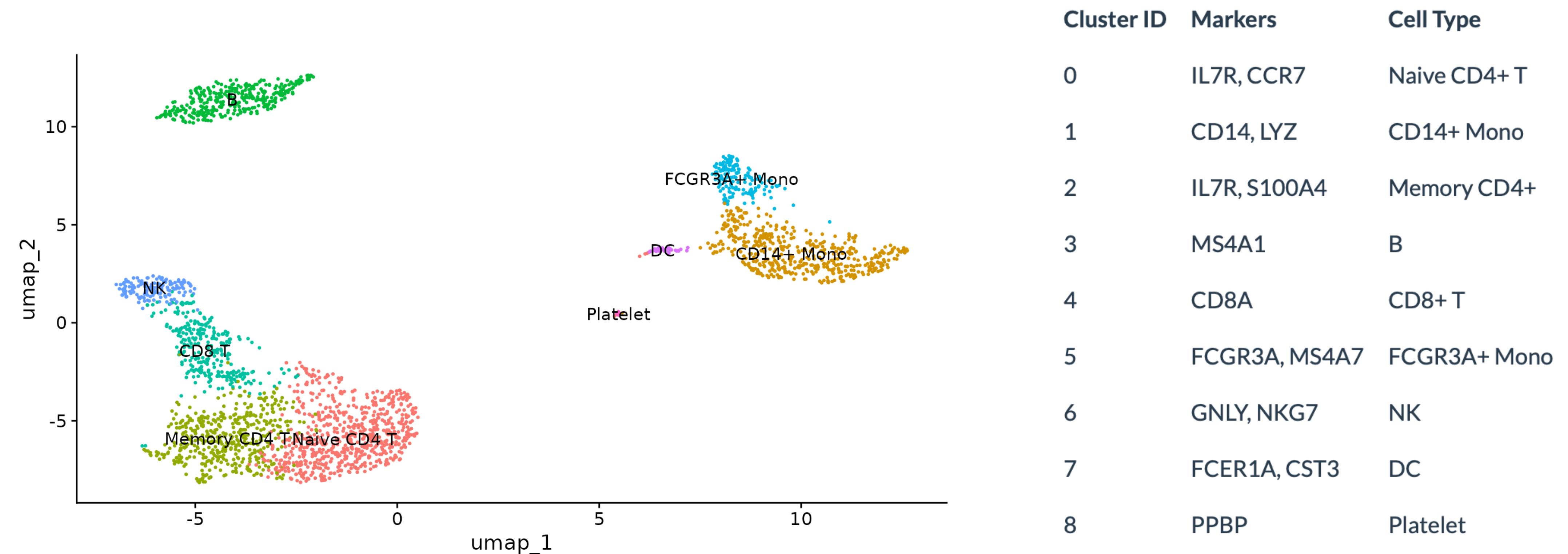
Downstream Analyse

Marker-Gene finden, die Cluster über differentielle Expression definieren

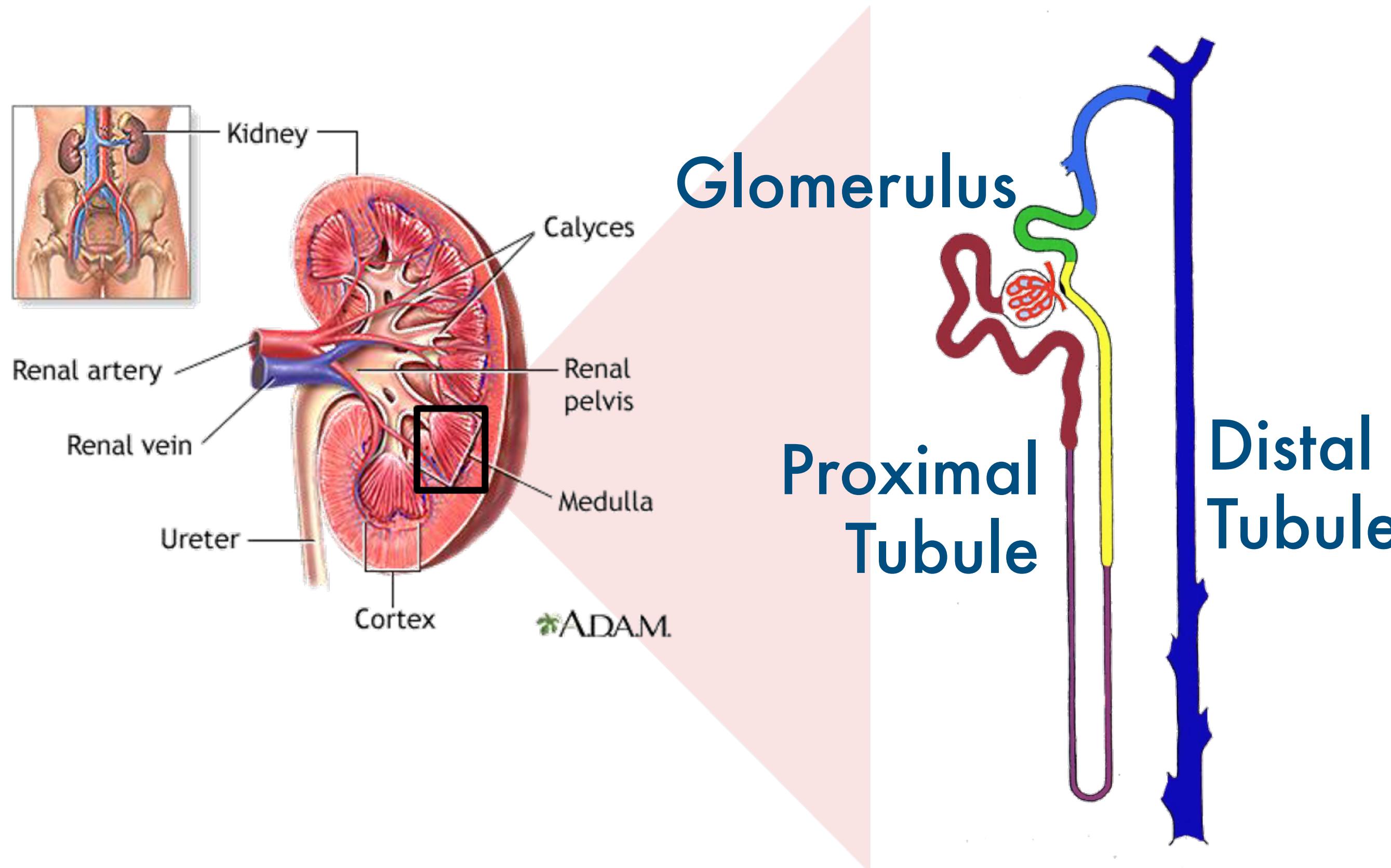


Downstream Analyse

Zuweisung der Zelltyp-Identität zu Clustern auf Grundlage der Literatur: Erstellung einer Expressions-Heatmap für bestimmte Zellen und Merkmale für die wichtigsten Marker.



Zweiter Haus-Datensatz: Niere



310 Zellen sequenziert mit Smart Seq

Marker Gene Analyse

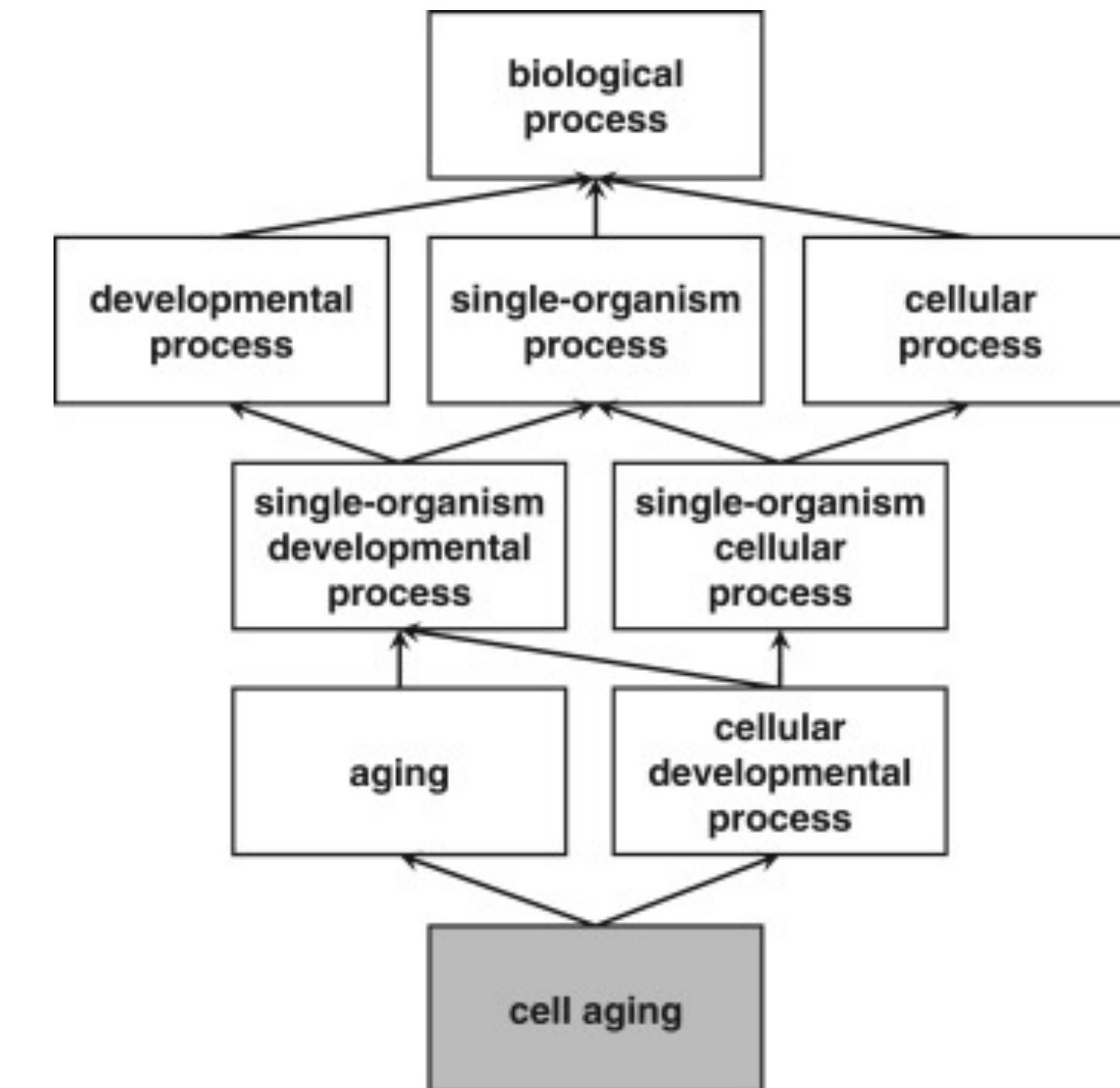
Cluster 0: Aqp2 (Distal Tubule)

Cluster 1: Nol3 (Glomerulus)

Cluster 2: Galnt14 (proximal Tubule)

Gene Ontology und Pathway Enrichment Analyse

Die Genanreicherungsanalyse ist eine rechnergestützte Methode, mit der ermittelt wird, ob eine vordefinierte Gruppe von Genen, beispielsweise solche in einem bestimmten biologischen Signalweg, in einer größeren Genliste überrepräsentiert ist. Sie wird zur Interpretation großer Genexpressionsdatensätze verwendet, indem sie ermittelt, welche biologischen Prozesse oder Signalwege signifikant mit einer bestimmten Erkrankung, einem bestimmten Phänotyp oder einem bestimmten Ansprechen auf eine Behandlung assoziiert sind, und eine lange Liste von Genen zu einer kleineren, übersichtlicheren Liste von Signalwegen zusammenfasst.



Gene Ontology und Pathway Enrichment Analyse

Cluster 0: Aqp2 (Distal Tubule)

Cluster 1: Nol3 (Glomerulus)

Cluster 2: Galnt14 (proximal Tubule)

Bile acid metabolism in the kidney occurs in the **proximal tubules and collecting ducts**, where bile acids are filtered and then reabsorbed to be conserved and returned to circulation. The **proximal tubule** contains the apical sodium-dependent bile acid transporter (ASBT) for reabsorption, while both the proximal and distal parts of the tubules have the farnesoid X receptor (FXR) and other transporters that regulate bile acid levels and play a role in lipid and water metabolism. 

Räumliche Transkriptomik

Das ist eine Technik, die die Genexpression bestimmten Stellen innerhalb eines Gewebes zuordnet und so ein umfassenderes Verständnis der Organisation und Interaktion von Zellen ermöglicht. Sie kombiniert die Hochdurchsatz-Genexpressionsprofilierung mit Positionsinformationen und ermöglicht es Forschern so, zu sehen, wie verschiedene Gene an bestimmten Stellen innerhalb eines Gewebes in seinem natürlichen Zustand exprimiert werden. Dies ist ein leistungsstarkes Werkzeug für die Forschung in Bereichen wie Krebs und Entwicklungsbiologie.

Kombiniere: Single Cell und Spatial Transcriptomics

Die erste weit verbreitete Methode wurde von Ståhl et al. in einer bahnbrechenden Veröffentlichung im Jahr 2016 in Science beschrieben, in der der Begriff „räumliche Transkriptomik“ geprägt wurde. Im Jahr 2018 erwarb 10X Genomics Spatial Transcriptomics als Grundlage für die 10X Visium-Plattform.

