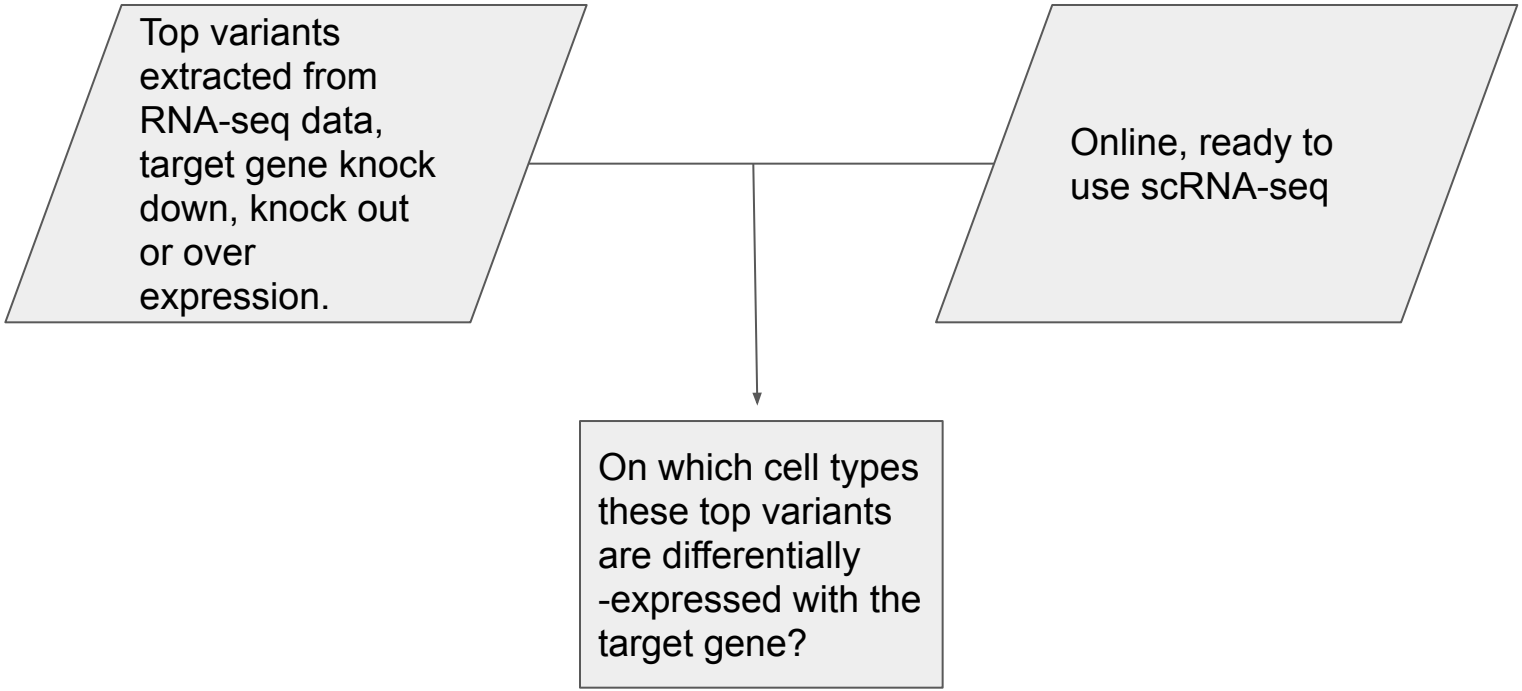
Gene Expression Deconvolution

Maral Dicle Maral
Görkem Durmaz
Ezgi Taş

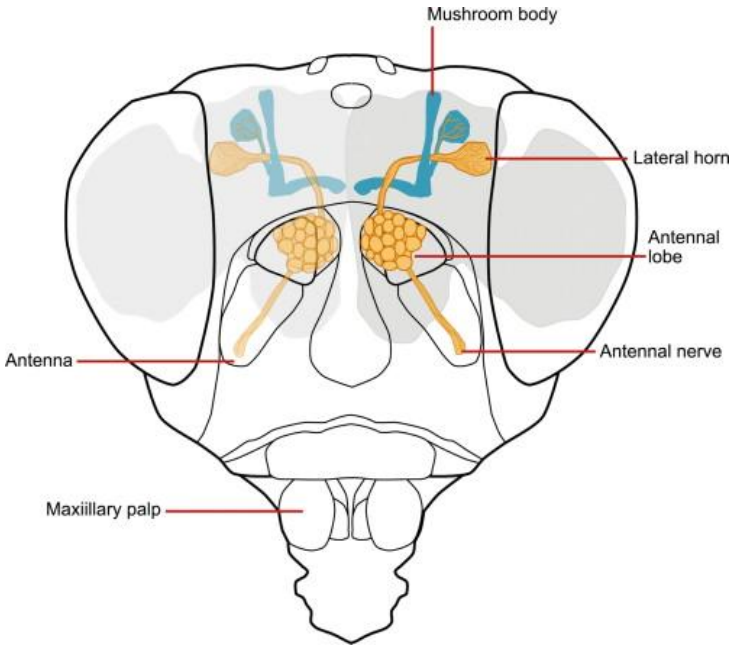
Gene-Expression Deconvolution/Decomposition



Transcriptome Analysis of *IROC* Genes in Fruitfly Olfactory System

- *Role of iroC* transcription factors in olfactory system (*Araucan, caupolican and mirror)*
- Bulk RNA sequencing for *iroC* mutant and wild-type organisms.

In which cell types these genes are differentially expressed?



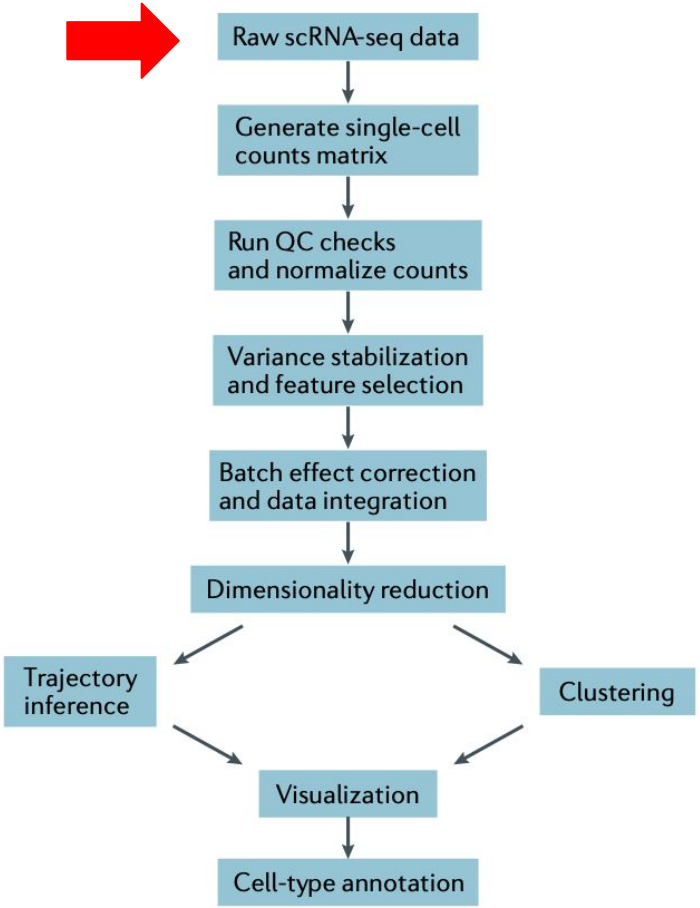
Differentially Expressed Genes

OLFACTORY RECEPTOR	MEAN COUNT
OR47A	1236,71
OR43B	1474.46
OR33C	251.53
OR47B	2087.87
OR85A	787,40
OR82A	399.01

OLFACTORY RECEPTOR	MEAN COUNT
ORCO	54059.51
OR71A	352.57
OR9A	1556.74
OR59B	4225.04
OR67A	905.89

Differential expression analysis done by Ibrahim Ihsan Taskiran, Arzu Celik's Neuroscience Laboratory

Workflow of Single Cell Seq Data Analysis

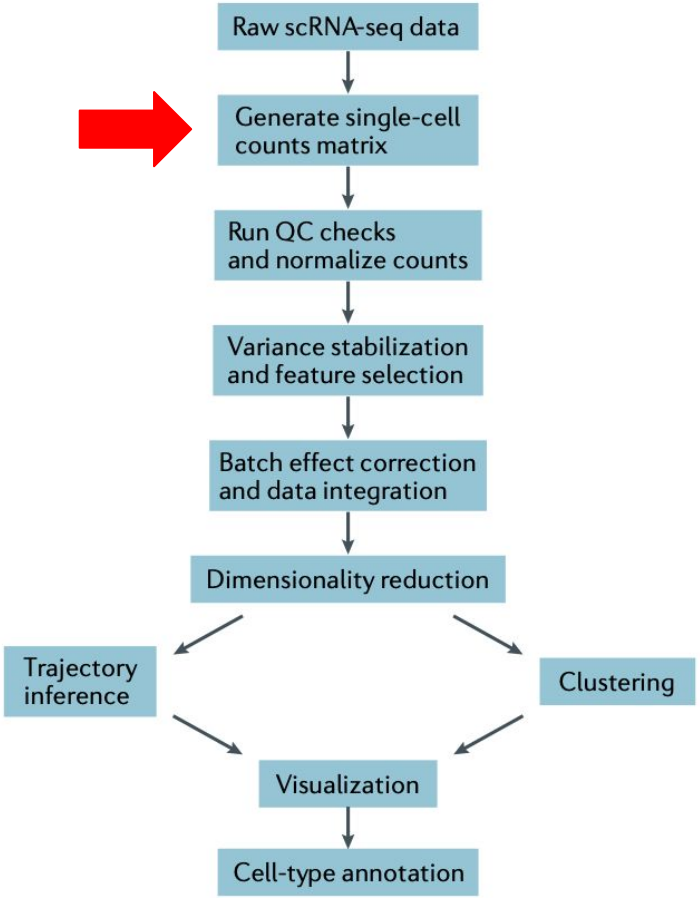


Data Pre-processing

Raw Input

- RNA Seq Read consisting of complementary DNA (cDNA)
- Cell Barcode
- Unique Molecular Index (UMI)

Workflow of Single Cell Seq Data Analysis

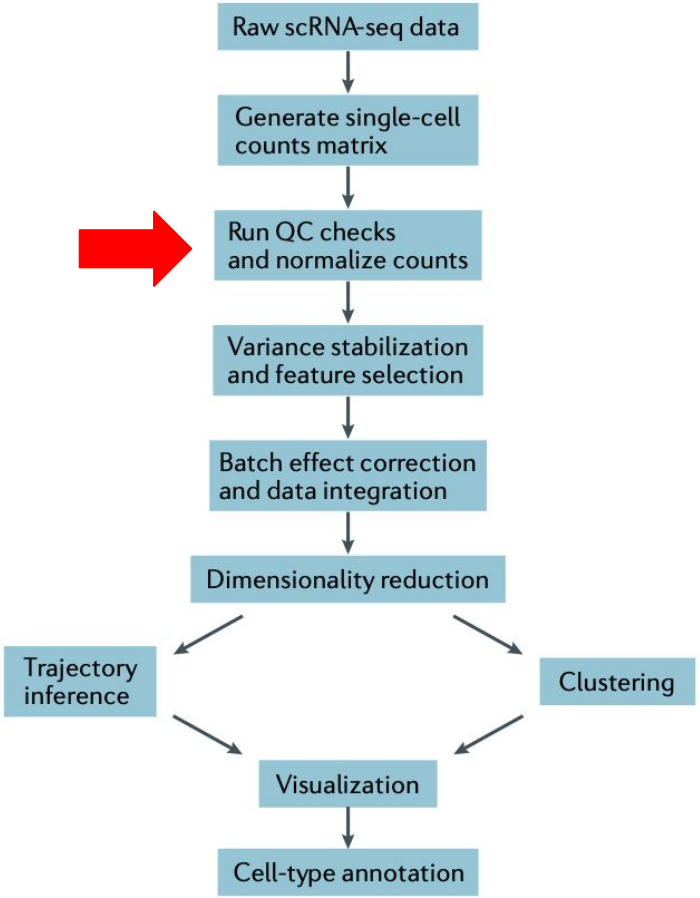


Generating a Gene Expression Matrix

- Alignment of cDNA sequences to a reference genome
- Assigning reads to cell of origin

- CellRanger , dropEst, Kallisto - BUStools, StarSolo

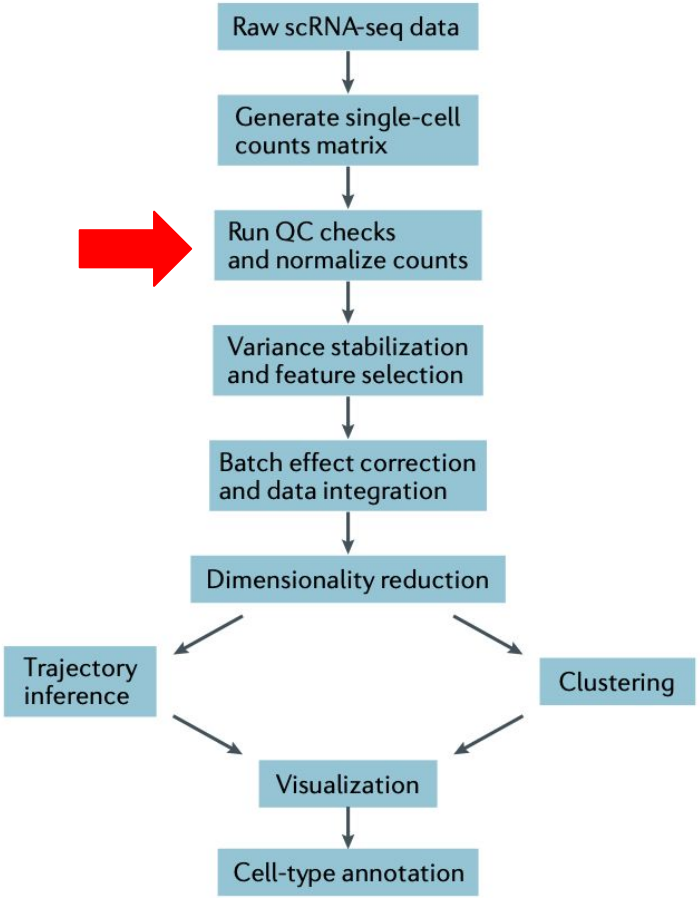
Workflow of Single Cell Seq Data Analysis



Quality Control and Doublet Detection

- Doublets
- Biases & Noise
- Seurat, SCANPY, Scater

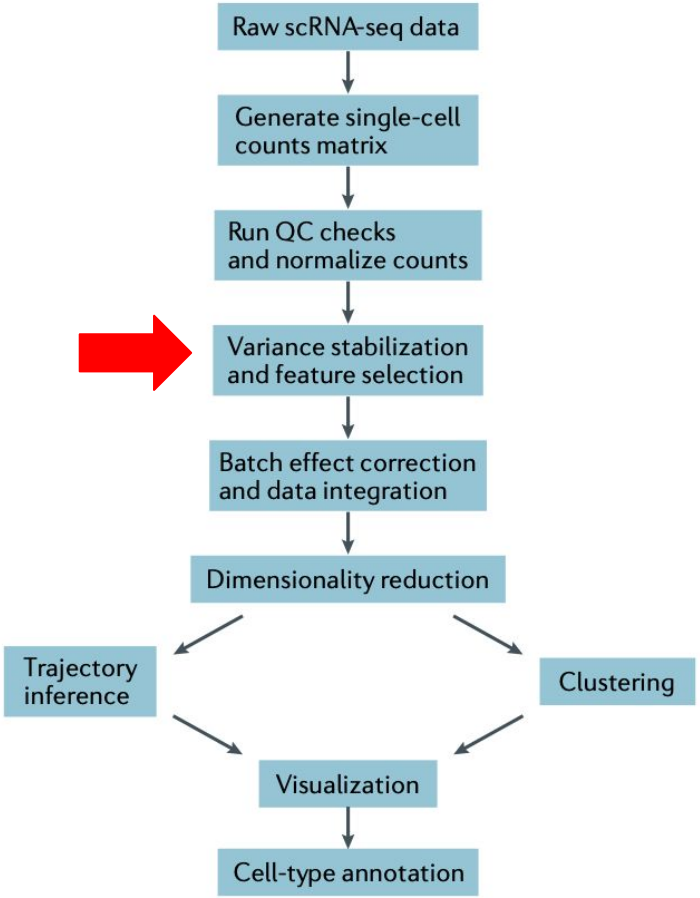
Workflow of Single Cell Seq Data Analysis



Data Normalization

- Normalization is therefore crucial to revealing the true biological heterogeneity of a dataset.

Workflow of Single Cell Seq Data Analysis

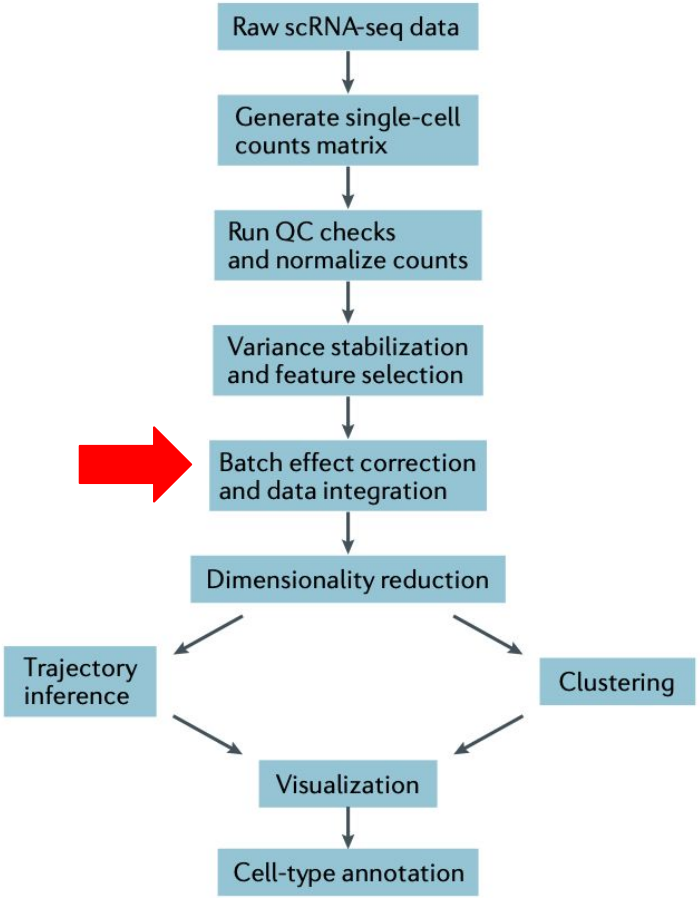


Variance Stabilization

- Enhances the contribution of biologically relevant genes.
- Log-transformation of normalized counts

- Seurat, Pagoda2, SCANPY, ZINB-Wave

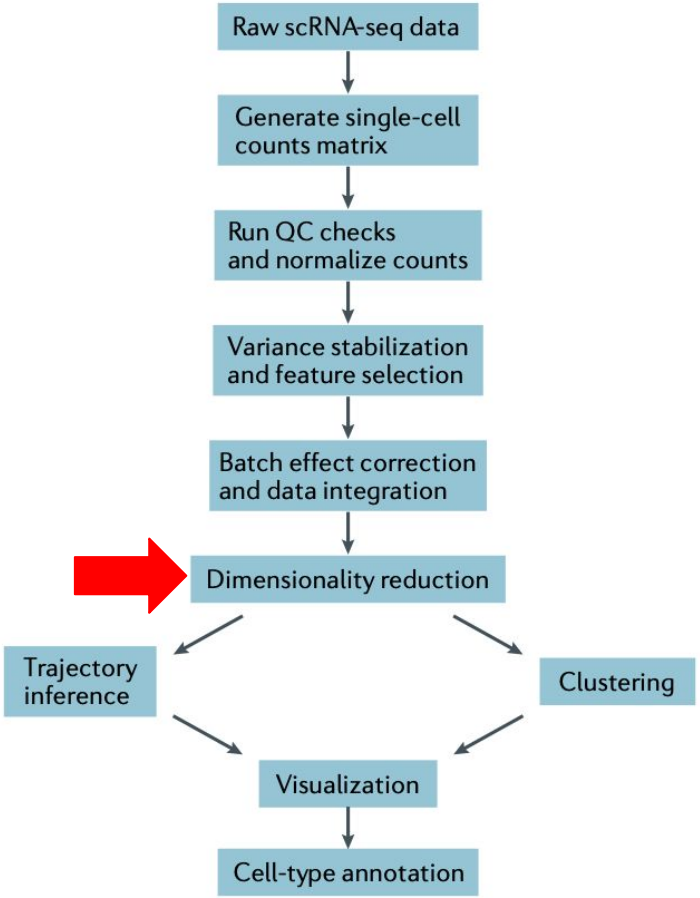
Workflow of Single Cell Seq Data Analysis



Batch Effects and Data Integration

- Integration can help us detect rare cell phenotypes.
- Regression model to eliminate batch differences
- Mutual Nearest Neighbors (MNNs)

Workflow of Single Cell Seq Data Analysis

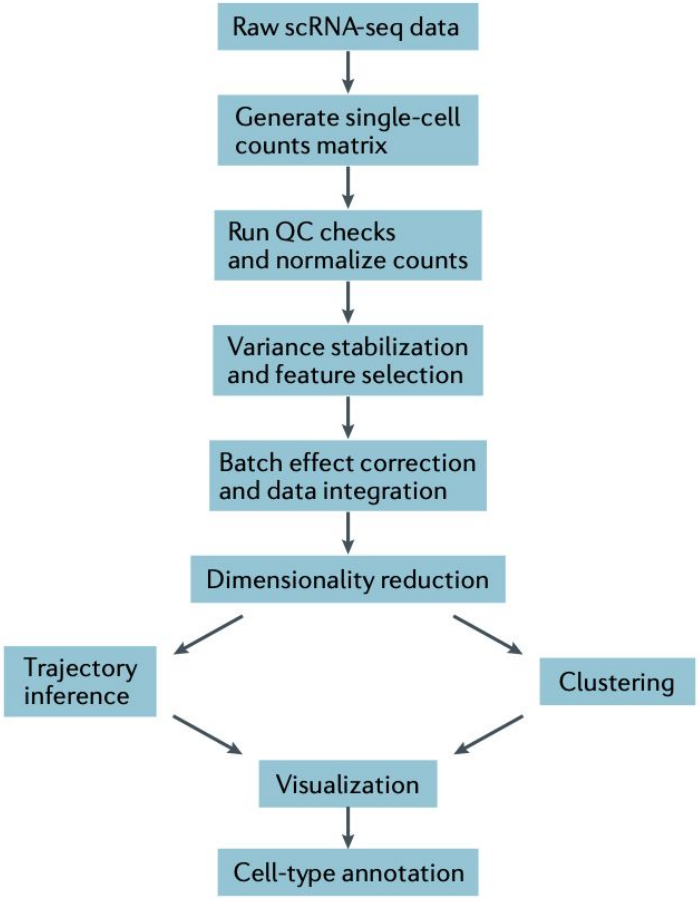


Downstream Analyses

Dimensionality Reduction and Imputation

- A more compact and interpretable data is produced.
- PCA & SCVI

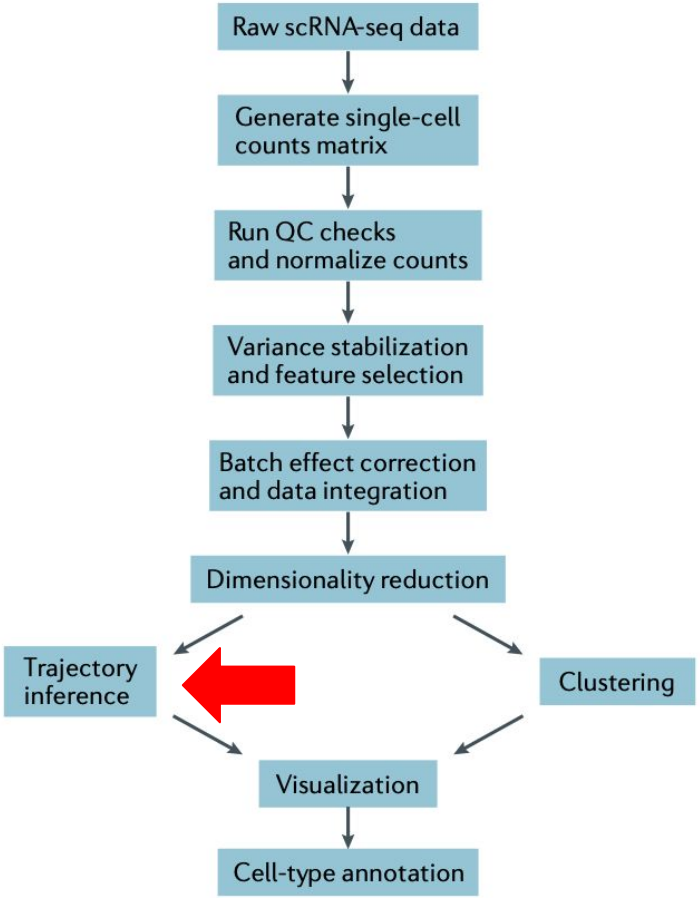
Workflow of Single Cell Seq Data Analysis



Clustering

- The goal of clustering is to identify subpopulations.
- Most popular types of clustering are hierarchal, graph-based and k-means clustering.
- Graph-based clustering is the most popular and best performing algorithm.
  - Louvain and Leiden algorithms

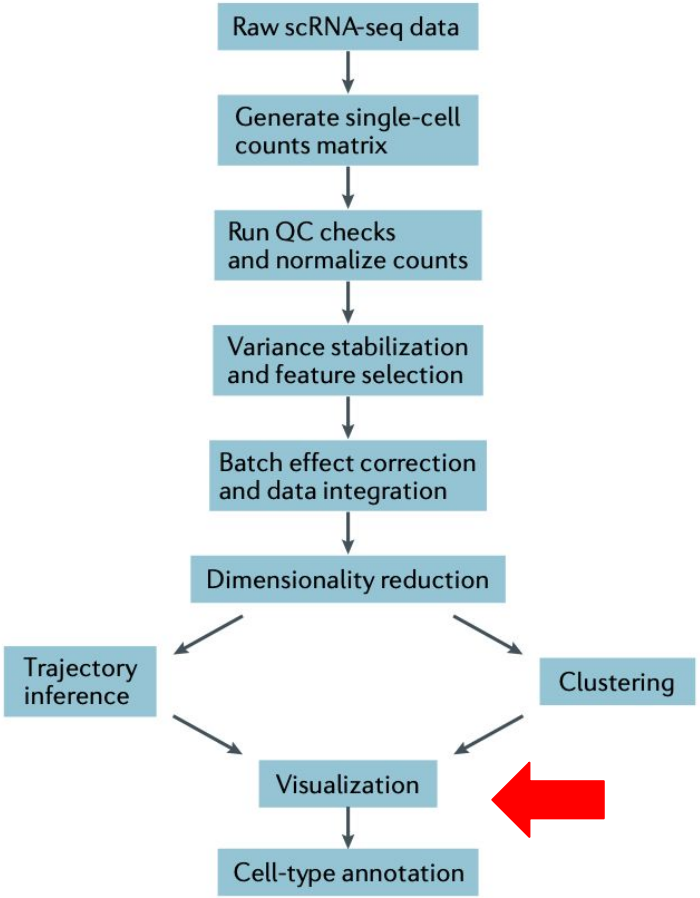
Workflow of Single Cell Seq Data Analysis



Trajectory Inference

- A continuum of gene expression patterns in development

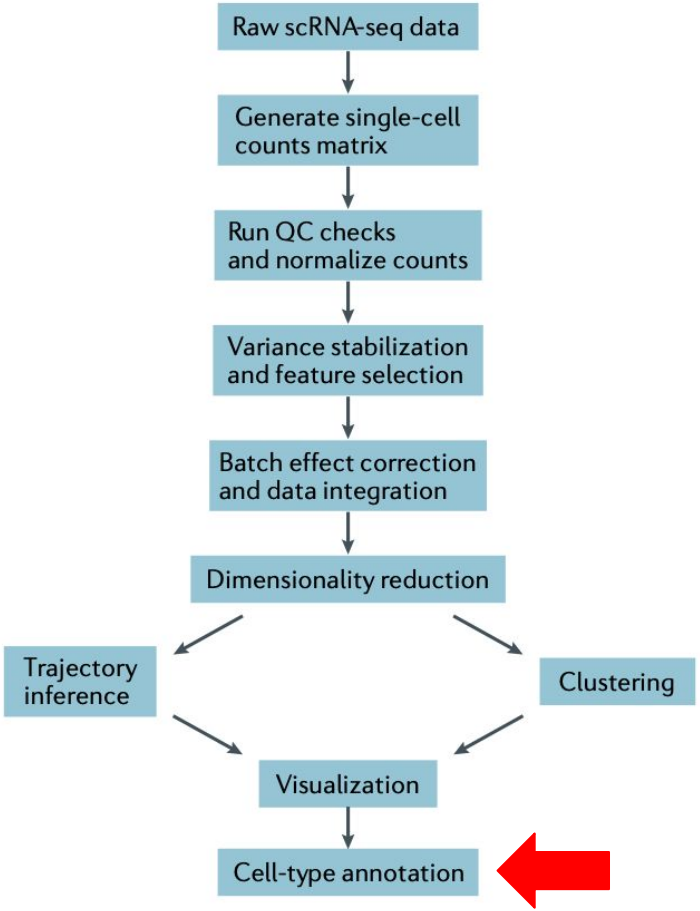
Workflow of Single Cell Seq Data Analysis



Visualization

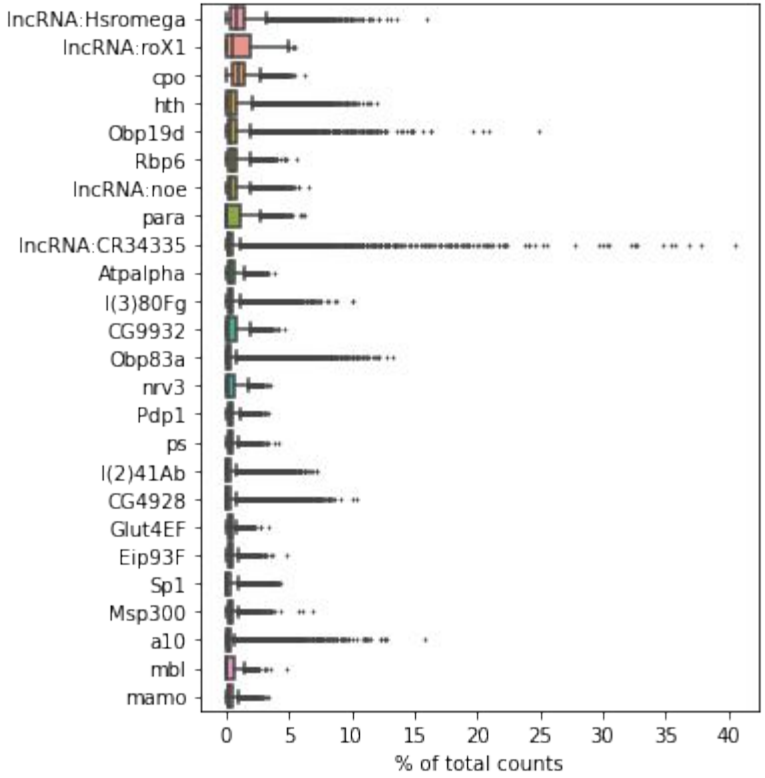
- 2D or 3D scatterplot of cells

Workflow of Single Cell Seq Data Analysis

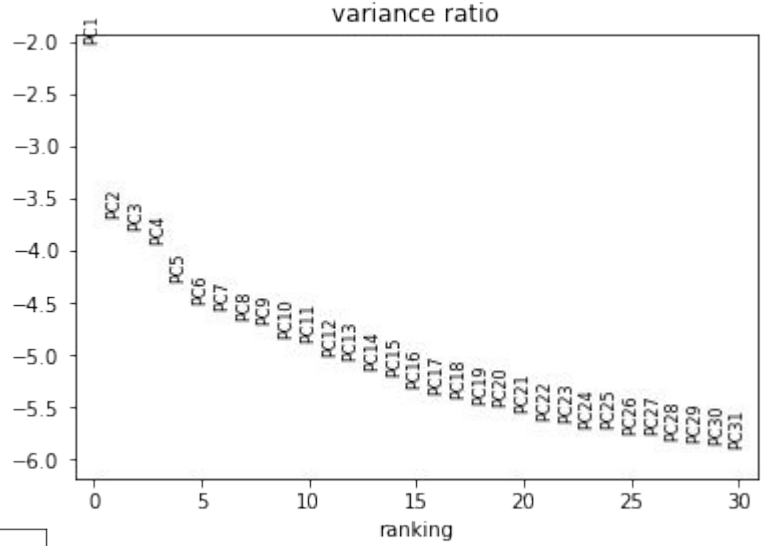


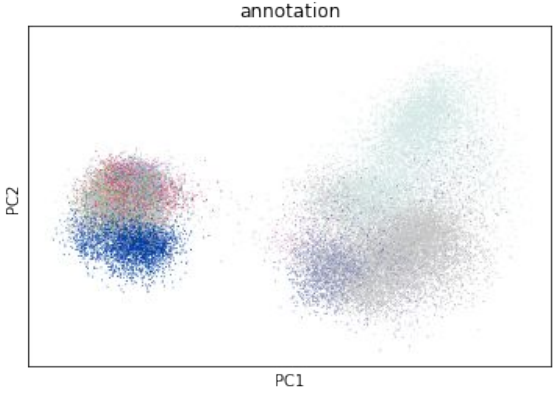
Cell-type Annotation

Preprocessing: Highly expressed genes in all cells



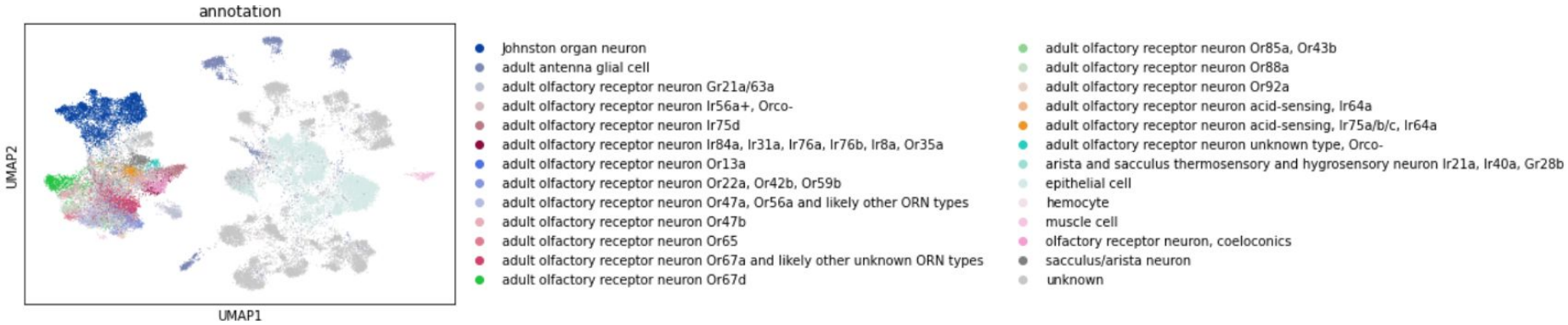
Preprocessing: PCA





- Johnston organ neuron
- adult antenna glial cell
- adult olfactory receptor neuron Gr21a/63a
- adult olfactory receptor neuron Ir56a+, Orco-
- adult olfactory receptor neuron Ir75d
- adult olfactory receptor neuron Ir84a, Ir31a, Ir76a, Ir76b, Ir8a, Or35a
- adult olfactory receptor neuron Or13a
- adult olfactory receptor neuron Or22a, Or42b, Or59b
- adult olfactory receptor neuron Or47a, Or56a and likely other ORN types
- adult olfactory receptor neuron Or47b
- adult olfactory receptor neuron Or65
- adult olfactory receptor neuron Or67a and likely other unknown ORN types
- adult olfactory receptor neuron Or67d
- adult olfactory receptor neuron Or85a, Or43b
- adult olfactory receptor neuron Or88a
- adult olfactory receptor neuron Or92a
- adult olfactory receptor neuron acid-sensing, Ir64a
- adult olfactory receptor neuron acid-sensing, Ir75a/b/c, Ir64a
- adult olfactory receptor neuron unknown type, Orco-
- arista and sacculus thermosensory and hygrosensory neuron Ir21a, Ir40a, Gr28b
- epithelial cell
- hemocyte
- muscle cell
- olfactory receptor neuron, coeloconics
- sacculus/arista neuron
- unknown

Results: Human annotation off cell types



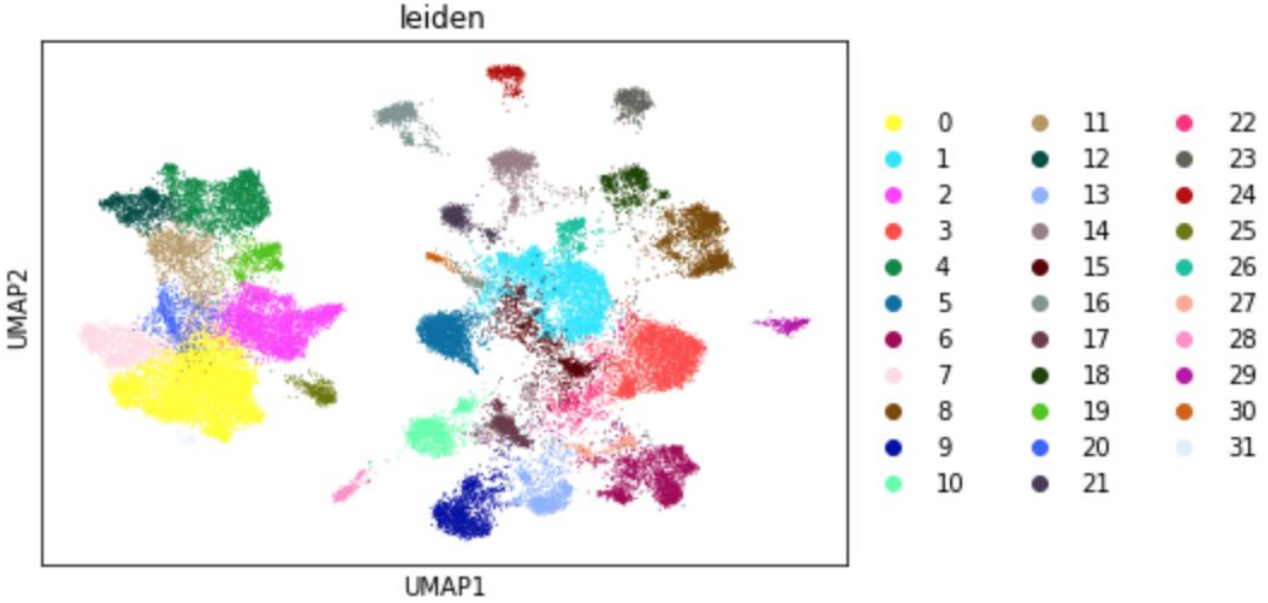
Differentially Expressed Genes

OLFACTORY RECEPTOR	MEAN COUNT
OR47A	1236,71
OR43B	1474.46
OR33C	251.53
OR47B	2087.87
OR85A	787,40
OR82A	399.01

OLFACTORY RECEPTOR	MEAN COUNT
ORCO	54059.51
OR71A	352.57
OR9A	1556.74
OR59B	4225.04
OR67A	905.89

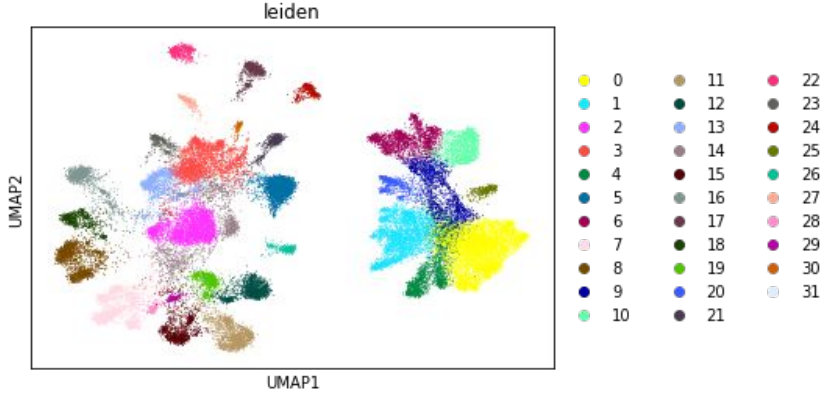
Differential expression analysis done by Ibrahim Ihsan Taskiran, Arzu Celik's Neuroscience Laboratory

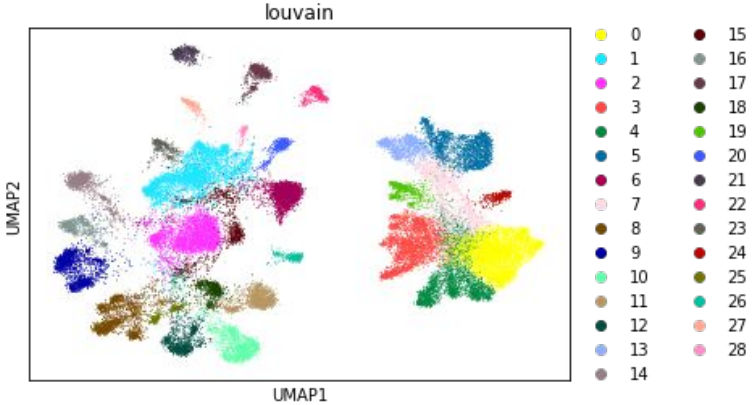
Results: Clustering with leiden algorithm

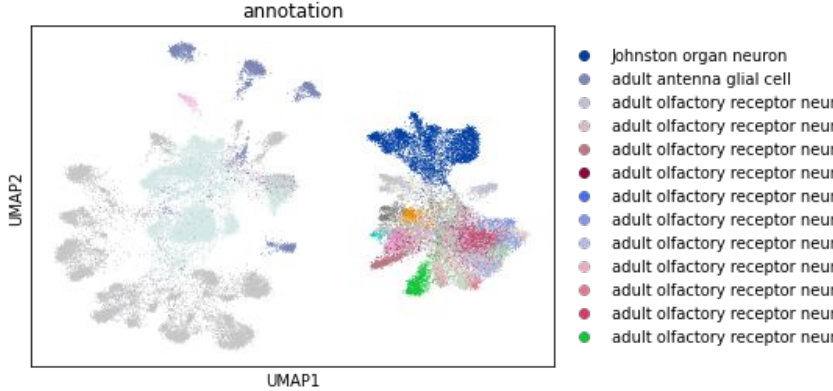


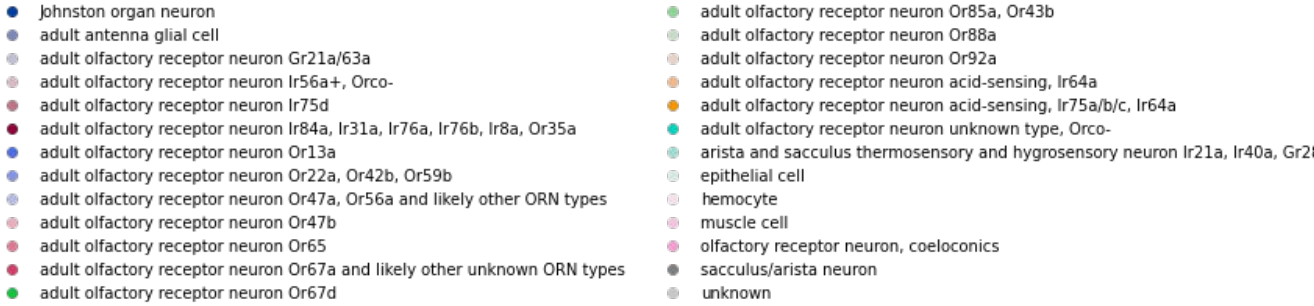
Results: Clustering with different parameters

- Default metrics are used. distance metric is euclidean and number of neighbors is 15. Number of principal components are 0.



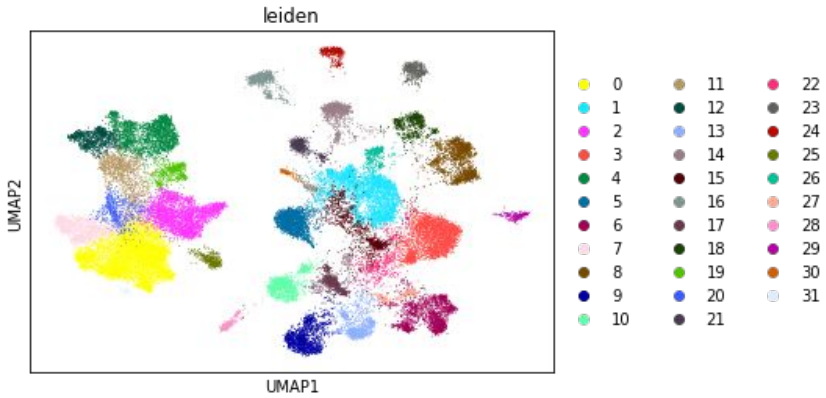


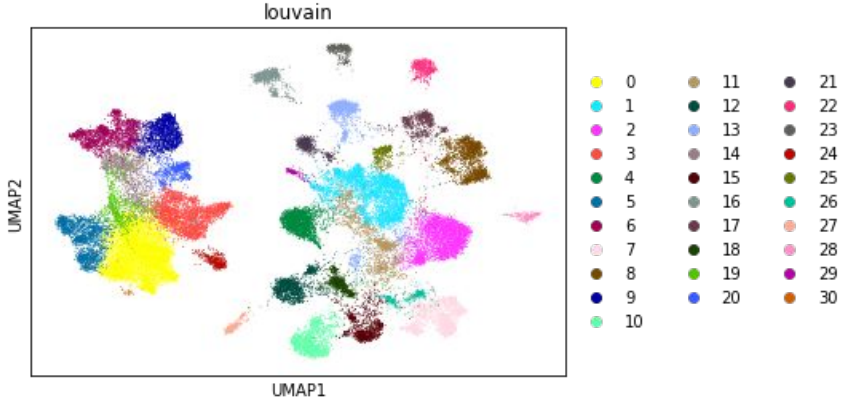


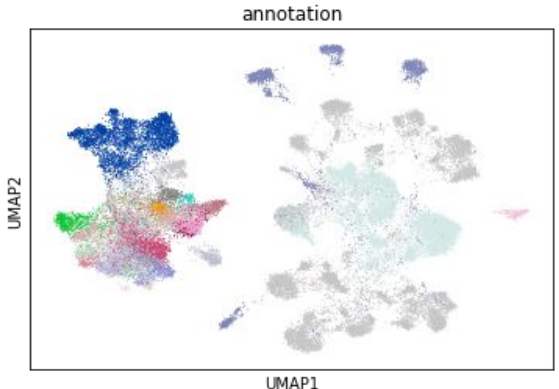


Results: Clustering with different parameters

- Number of neighbors is decreased to 10, number of PCs set to be 31 and metric used is euclidean distance.



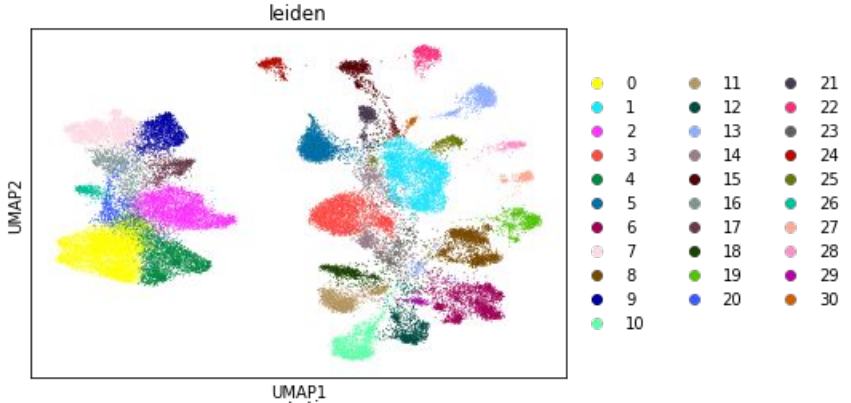


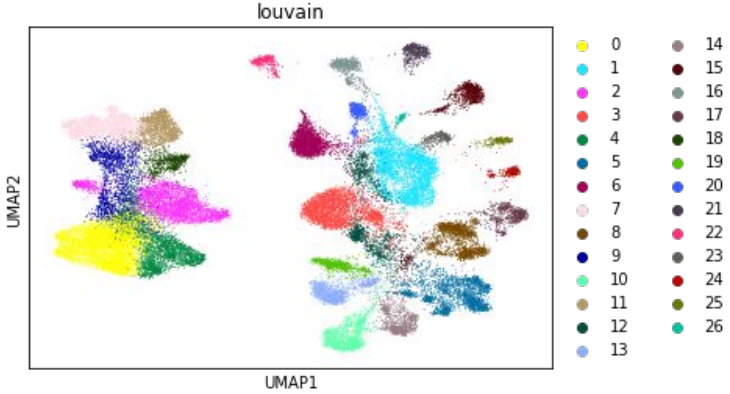


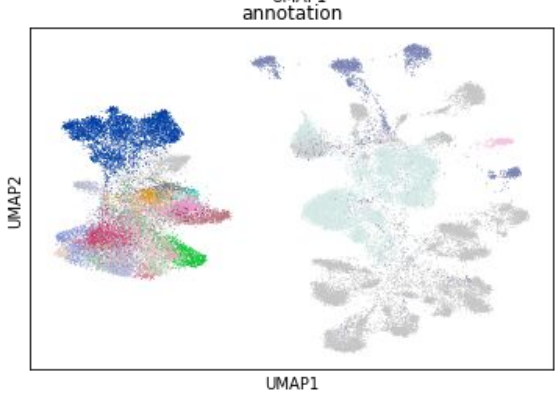
Johnston organ neuron
adult antenna glial cell
adult olfactory receptor neuron Gr21a/63a
adult olfactory receptor neuron Ir56a+, Orco-
adult olfactory receptor neuron Ir75d
adult olfactory receptor neuron Ir84a, Ir31a, Ir76a, Ir76b, Ir8a, Or35a
adult olfactory receptor neuron Or13a
adult olfactory receptor neuron Or22a, Or42b, Or59b
adult olfactory receptor neuron Or47a, Or56a and likely other ORN types
adult olfactory receptor neuron Or47b
adult olfactory receptor neuron Or65
adult olfactory receptor neuron Or67a and likely other unknown ORN types
adult olfactory receptor neuron Or67d
adult olfactory receptor neuron Or85a, Or43b
adult olfactory receptor neuron Or88a
adult olfactory receptor neuron Or92a
adult olfactory receptor neuron acid-sensing, Ir64a
adult olfactory receptor neuron acid-sensing, Ir75a/b/c, Ir64a
adult olfactory receptor neuron unknown type, Orco-
arista and sacculus thermosensory and hygrosensory neuron Ir21a, Ir40a, Gr28
epithelial cell
hemocyte
muscle cell
olfactory receptor neuron, coeloconics
sacculus/arista neuron
unknown

Results: Clustering with different parameters

- Number of neighbors is 15, number of PCs set to be 31 and metric used is euclidean distance.

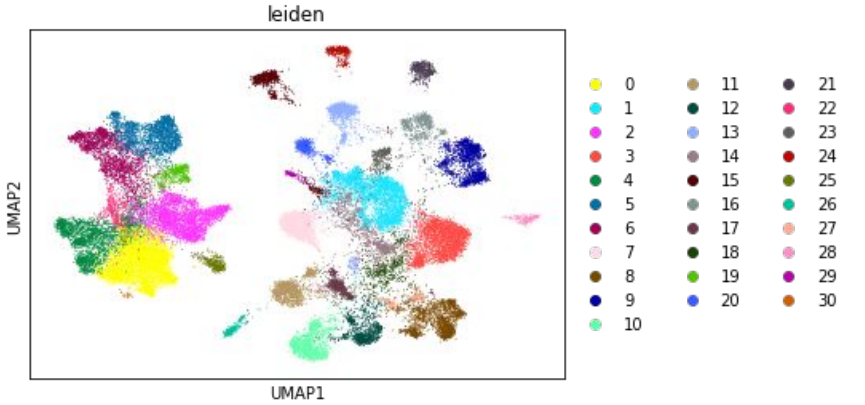


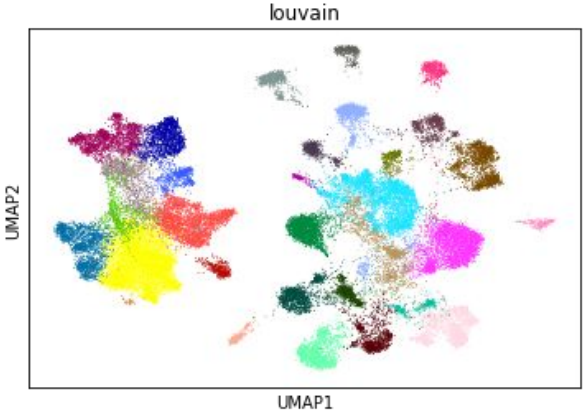


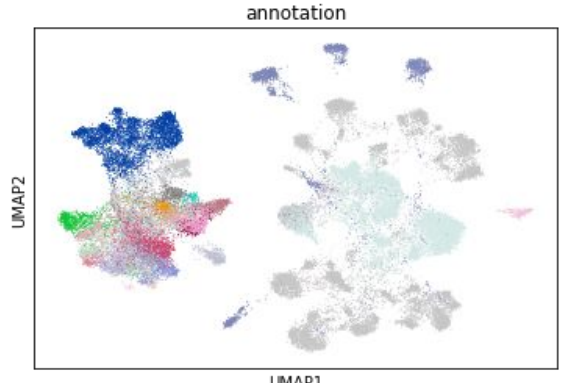


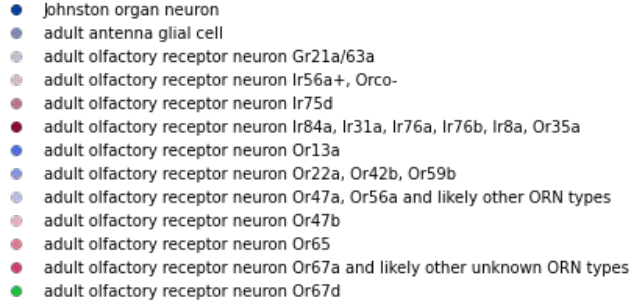
Results: Clustering with different parameters

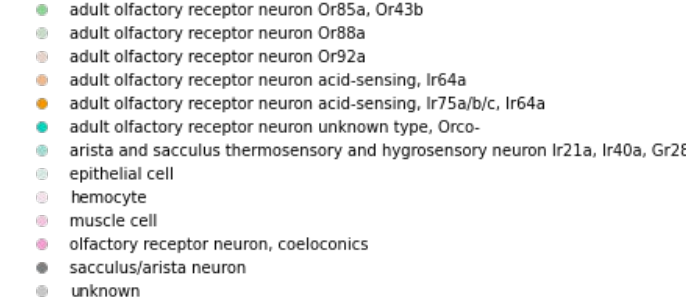
- Number of neighbors is 15, number of PCs set to be 31 and metric used is correlation distance.



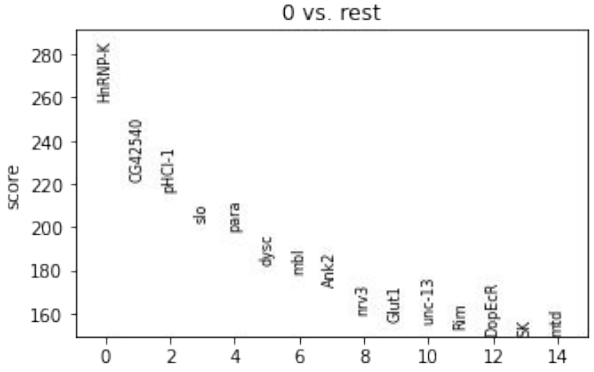


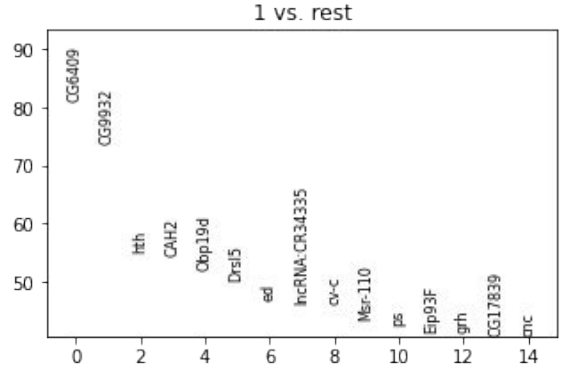


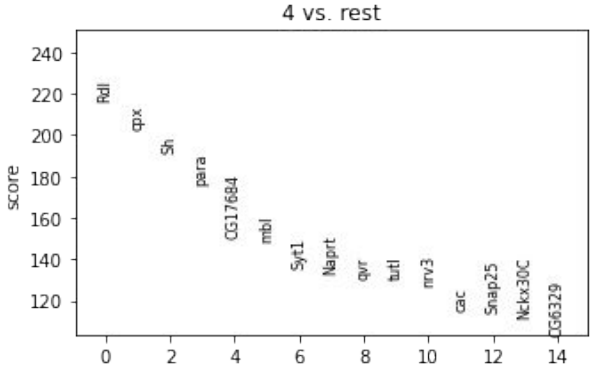


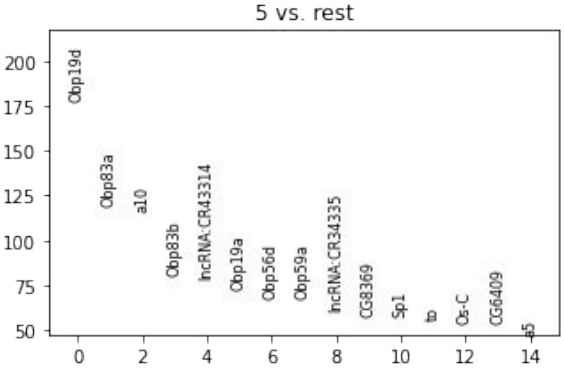


Results: Marker Genes

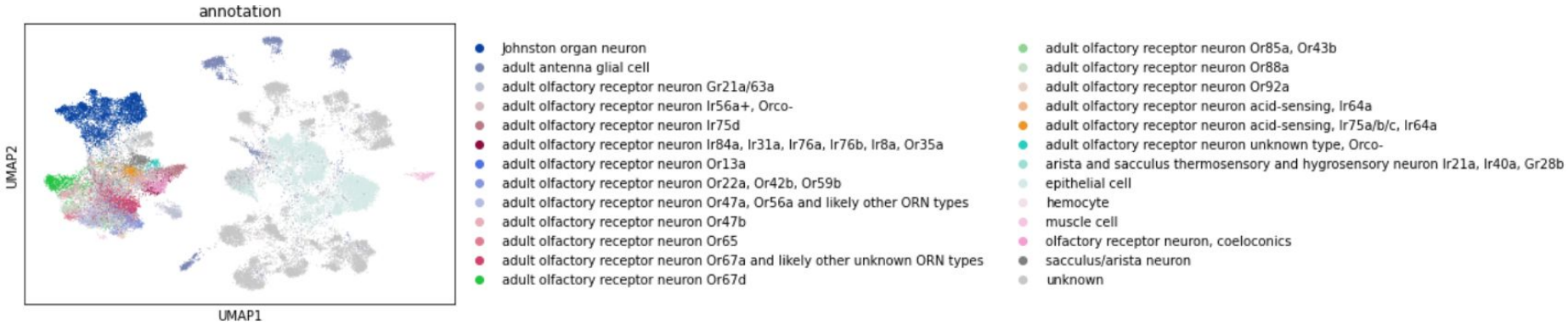




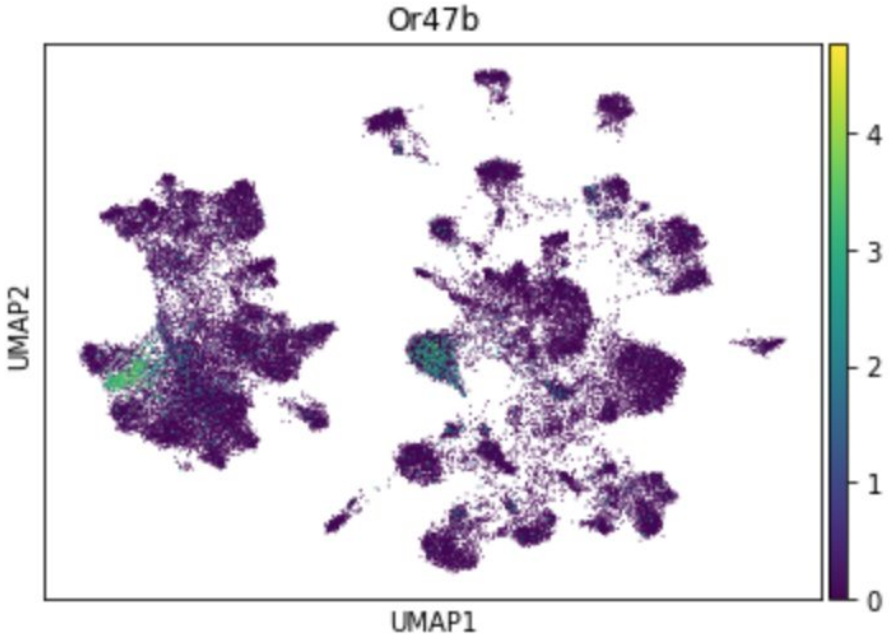




Results: Human annotation off cell types



Results: Expression pattern of Or47b gene across all the clusters



Differentially Expressed Genes

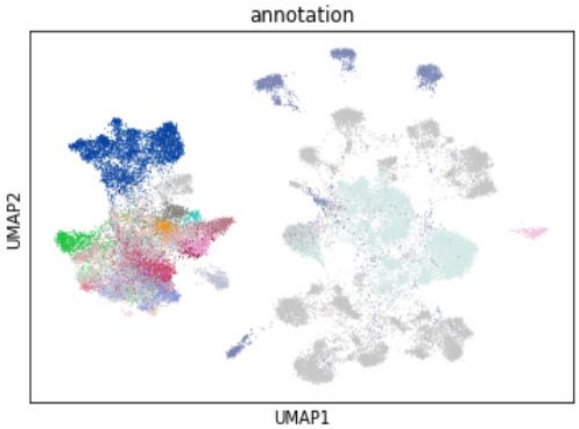
OLFACTORY RECEPTOR	MEAN COUNT
OR47A	1236,71
OR43B	1474.46
OR33C	251.53
OR47B	2087.87
OR85A	787,40
OR82A	399.01

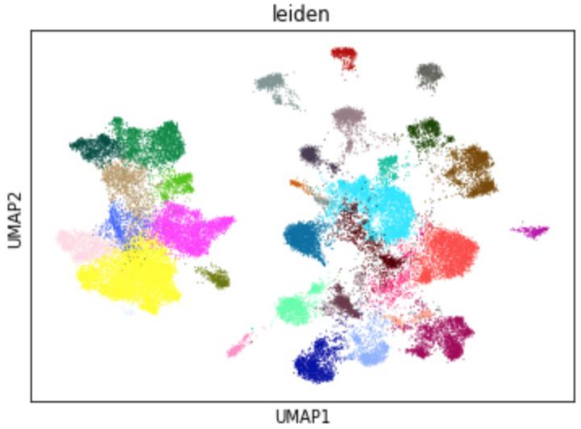
OLFACTORY RECEPTOR	MEAN COUNT
ORCO	54059.51
OR71A	352.57
OR9A	1556.74
OR59B	4225.04
OR67A	905.89

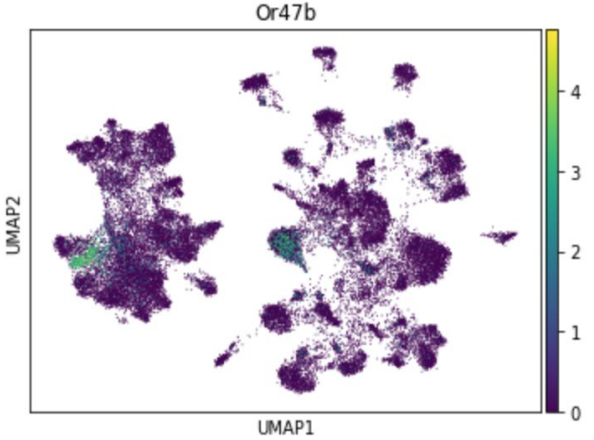
Differential expression analysis done by Ibrahim Ihsan Taskiran, Arzu Celik's Neuroscience Laboratory

Results: Comparison of Or47b expression with clusters

Leiden Algorithm, number of neighbors=10, metric used is euclidean distance.







Results: Number of various cell types in the clusters

	Cluster 0	Cluster 1	Cluster 2
adult olfactory receptor neuron Or22a, Or42b, Or59b	649	0	0
adult olfactory receptor neuron Or47a, Or56a and likely other ORN types	554	0	2
adult olfactory receptor neuron Or47b	8	0	0
adult olfactory receptor neuron Or67a and likely other unknown ORN types	1181	0	22
adult olfactory receptor neuron Or85a, Or43b	161	0	0

Conclusion and Future Work

- We might try **different approaches for cell type annotation** (ML approaches or different human annotations for the cluster marker genes)

- **Subclustering** the interested clusters

- **Adjusted Rand Index** to quantify cluster similarity

- **Bootstrapping** to improve confidence in clusters

References

[1]Y. Wu and K. Zhang, "Tools for the analysis of high-dimensional single-cell RNA sequencing data", *Nature Reviews Nephrology*, vol. 16, no. 7, pp. 408-421, 2020. Available: 10.1038/s41581-020-0262-0 [Accessed 16 January 2022].

[2]B. Hwang, J. Lee and D. Bang, "Single-cell RNA sequencing technologies and bioinformatics pipelines", *Experimental & Molecular Medicine*, vol. 50, no. 8, pp. 1-14, 2018. Available: 10.1038/s12276-018-0071-8 [Accessed 16 January 2022].

[3]H. Li et al., "Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly", 2021. Available: 10.1101/2021.07.04.451050 [Accessed 16 January 2022].

[4]V. Traag, L. Waltman and N. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities", *Scientific Reports*, vol. 9, no. 1, 2019. Available: 10.1038/s41598-019-41695-z [Accessed 16 January 2022].

THANK YOU!