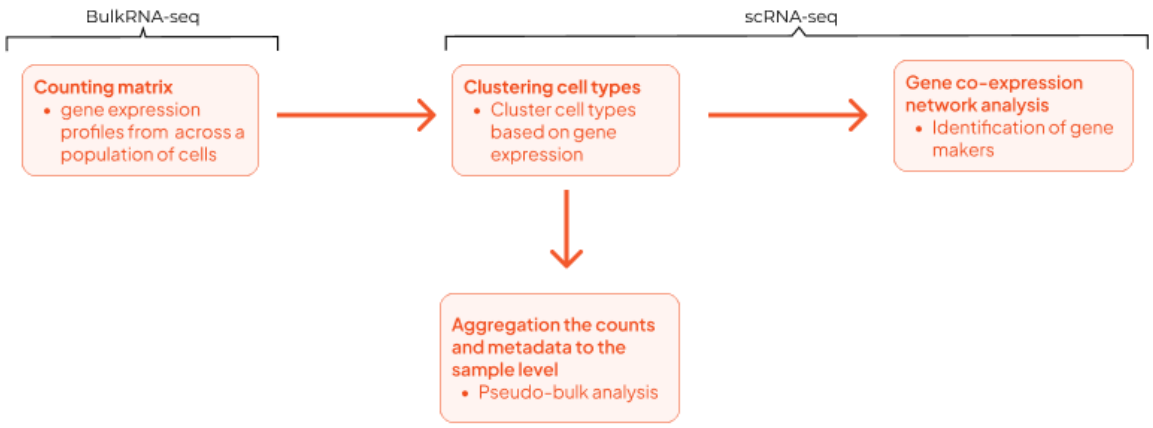
Future research

Background:

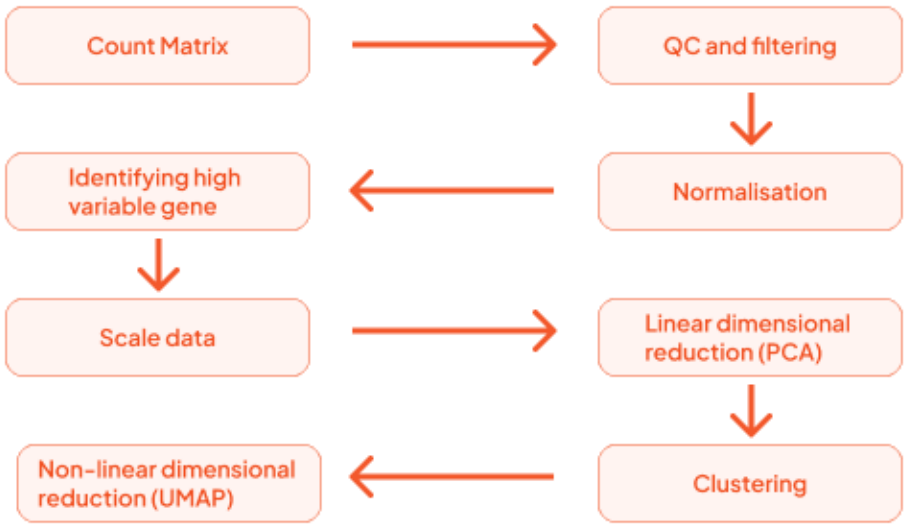
Alzheimer's disease (AD) is the most common neurodegenerative disorder. This neurodegenerative disorder is involved in different cell types in the brain such as neurons, microglia, astrocytes, oligodendrocytes, pericytes. However, transcriptome profiles between these cell types associated with neurodegeneration remain incompletely understood. Bulk RNA-seq and single cell RNA-seq are now available to examine transcriptome profiles of these cell types. While bulk RNA-seq provides insight into the transcriptome which results in an average result of gene expression profiles from across a population of cells, single cell RNA-seq analyse transcriptome of individual cells to profile entire population for clustering cell types and identifying cell markers.

Methodology:

RNA-sequencing of cell types (neurons, microglia, astrocytes, oligodendrocytes, pericytes) from patients with AD and from controls need to be prepared. Analysis of RNA-seq data is comprised of four sub-models: counting matrix, clustering cell types and gene co-expression network analysis, including aggregating the counts and metadata to the sample level. Diagram of the analysis of RNA-seq data is as follows.



BulkRNA-seq	**scRNA-seq**
1)Goal: Analysing gene expression of all cells 2)QC: The overly present based on the measurement: K-mers 3)Normalisation: use a method to make the comparing meaningful 4)Analysis: Identifying biomarker/expression network	1)Goal: Analyse gene expression of single cells 2)QC: Look at the gene/cell/percentage of mitochondria 3)Normalisation: use the method to make the comparing meaningful 4)Analysis: Clustering

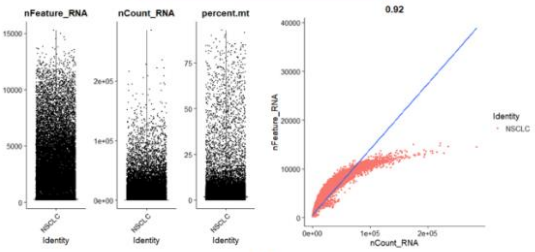
Workflow FASTQ segments-> Cleansing segments-> Mapping segment with Ref. (genome or transcriptome) -> get segment from Ref. segments based on the FASTQ segment -> assemble Ref. segments -> read gene transcript at exon level -> downstream analysis or expression analysis (BAM file). **Output:** Count matrix, Pseudo-bulk analysis	**Workflow** FASTQ-> method of **cell ranger** (transform from FASTQ to barcode matrix) -> Downstream analysis. **Output:** group cell cluster -> gene marker
Downstream analysis:  	

Clustering cell type (Example)

Count Matrix

```
> nsclc.seurat.obj #
An object of class Seurat
32978 features across 71880 samples within 1 assay
Active assay: RNA (32978 features, 0 variable features)
```

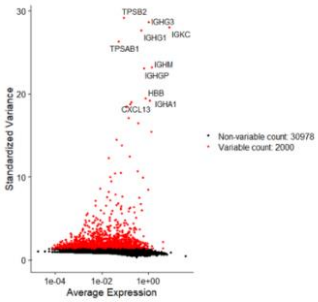
QC and filtering



Normalisation

```
nsclc.seurat.obj <- NormalizeData(nsclc.seurat.obj)
str(nsclc.seurat.obj)
```

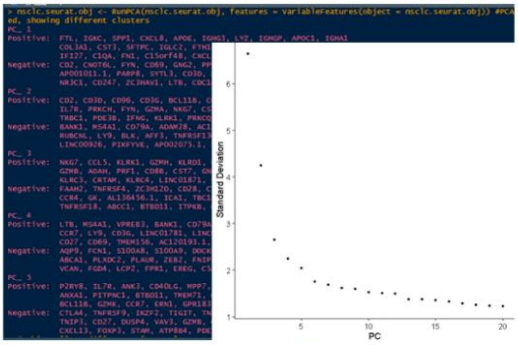
Identifying high variable gene



Scale data

```
all.genes <- rownames(nsclc.seurat.obj)
nsclc.seurat.obj <- ScaleData(nsclc.seurat.obj, features = all.genes)
str(nsclc.seurat.obj)
```

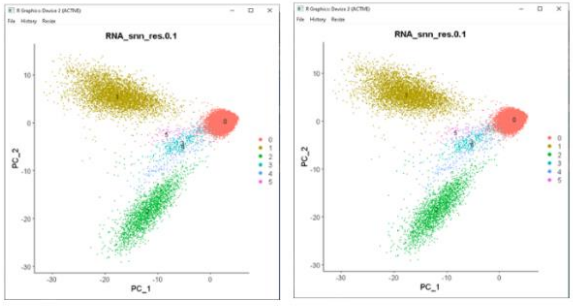
Linear dimensional reduction (PCA)



Clustering

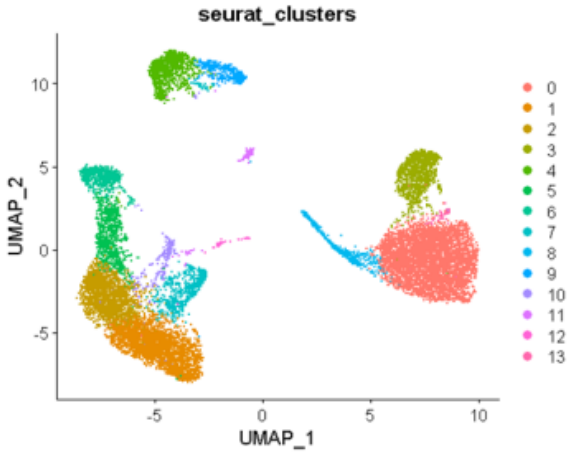
```
# 7. Clustering ------------
nsclc.seurat.obj <- FindNeighbors(nsclc.seurat.obj, dims = 1:15)
```

Non-linear dimensional reduction (UMAP)

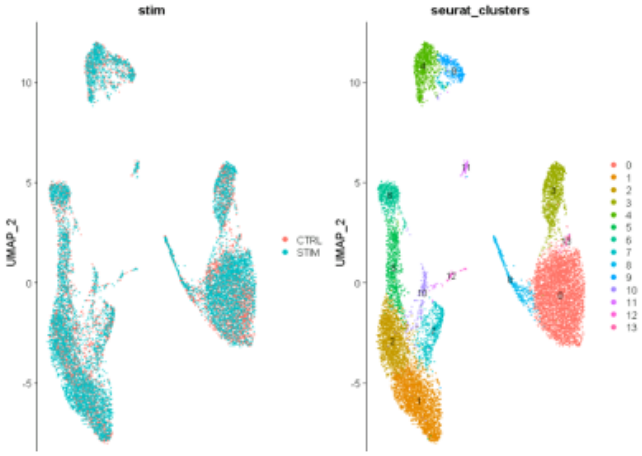


Identifying gene markers (Example)

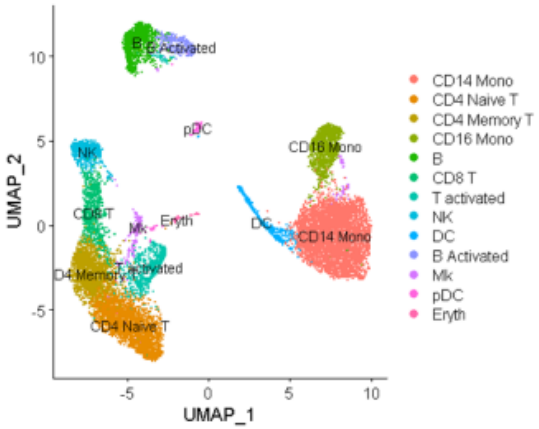
Visualise clusters by UMAP



Show two condition: stimulate , control



Label clusters



Label clusters from stimulate condition and control condition

Split and plot based on conditions and genes