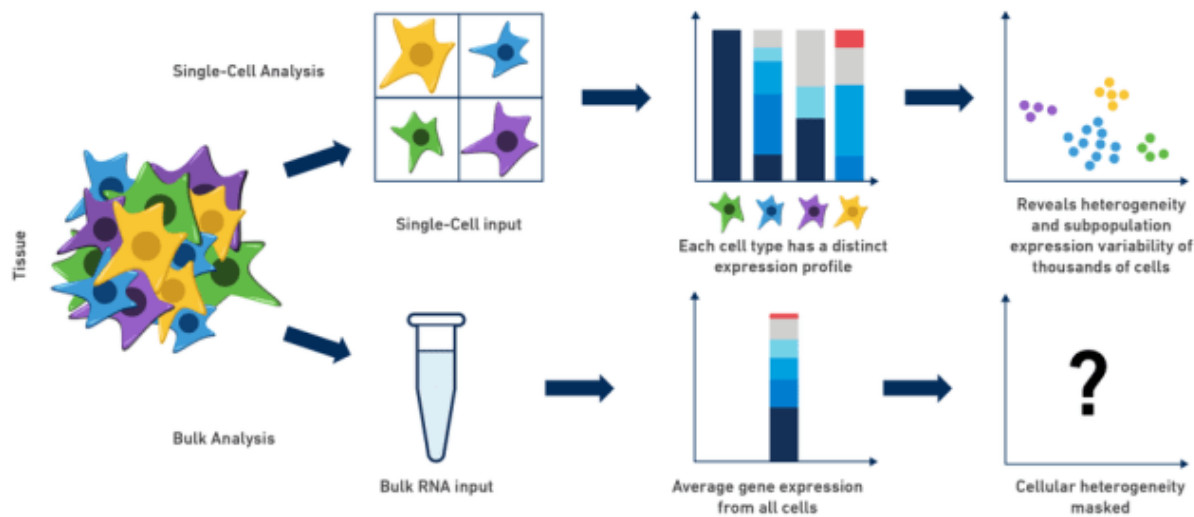


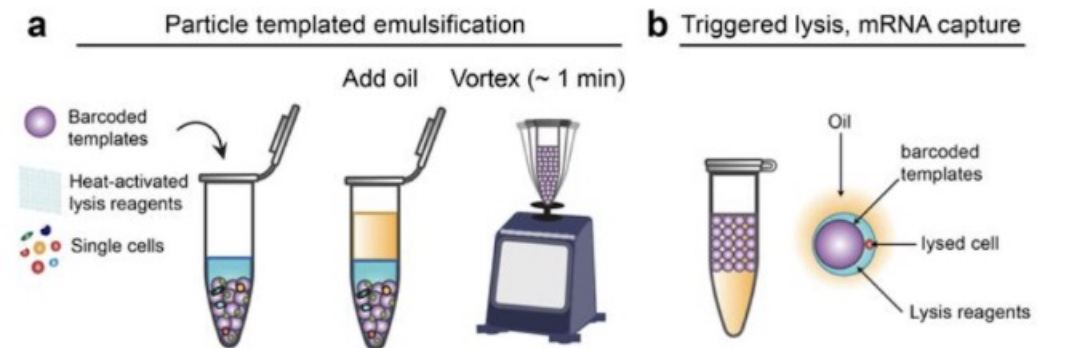
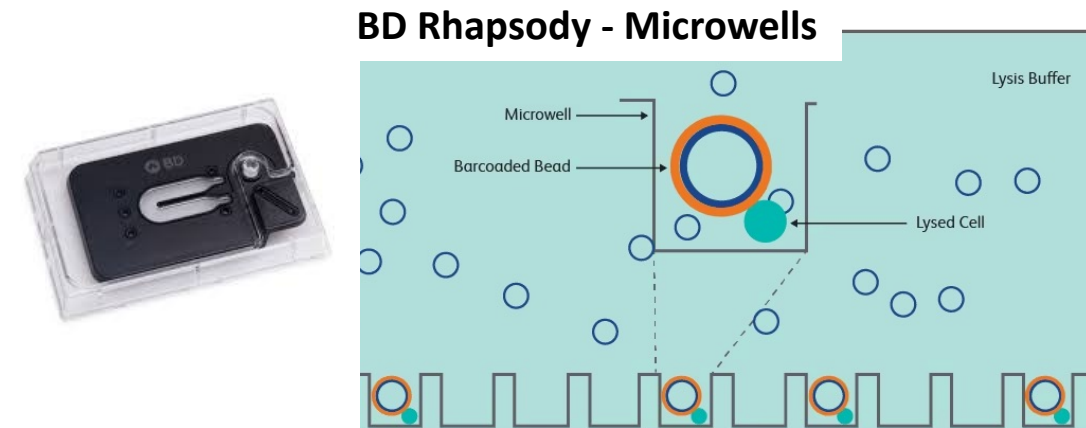
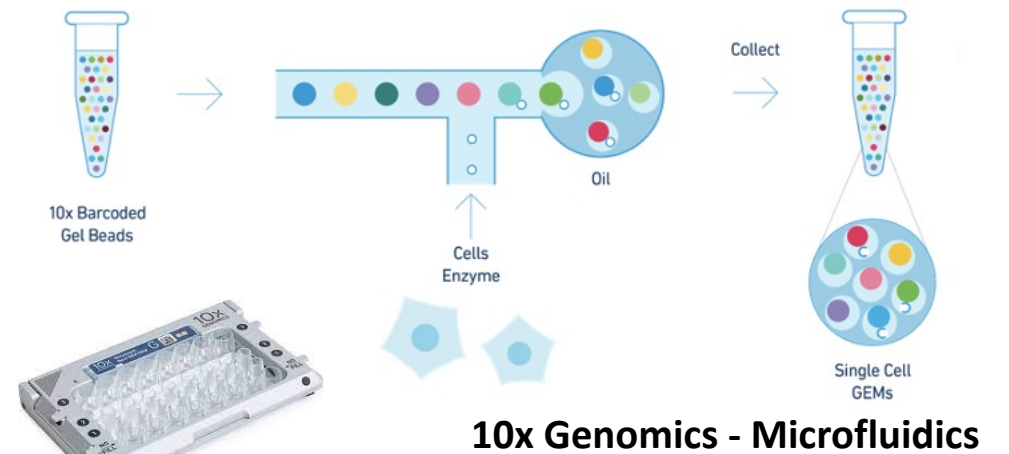
# Introduction to single cell RNA-Seq for trajectory analysis

# Single cell RNA-Seq

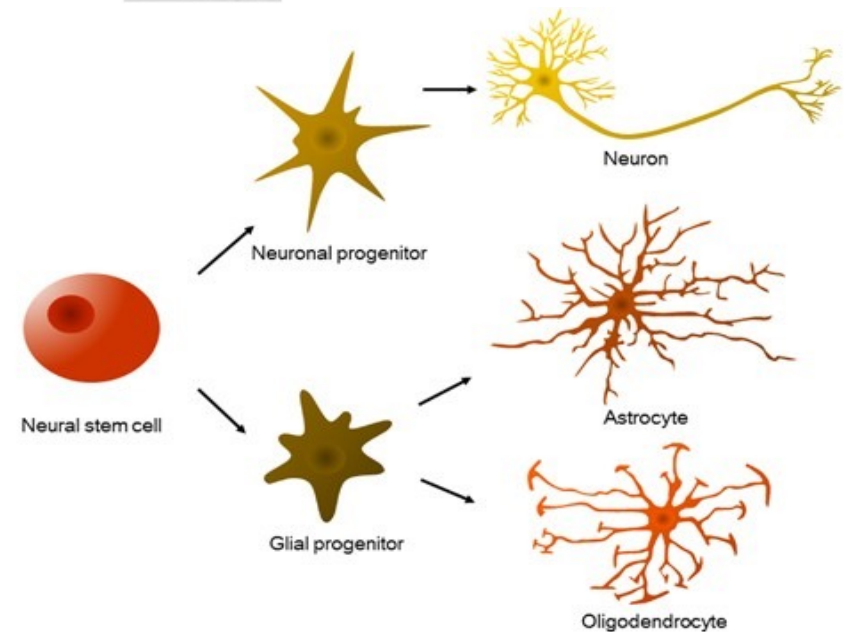
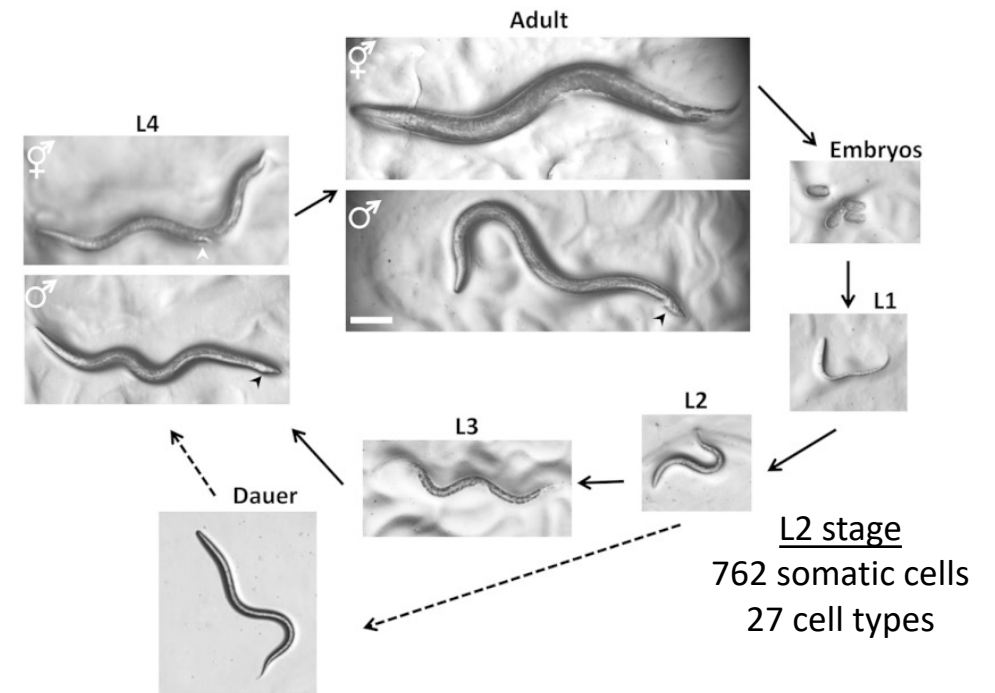


## Highlights

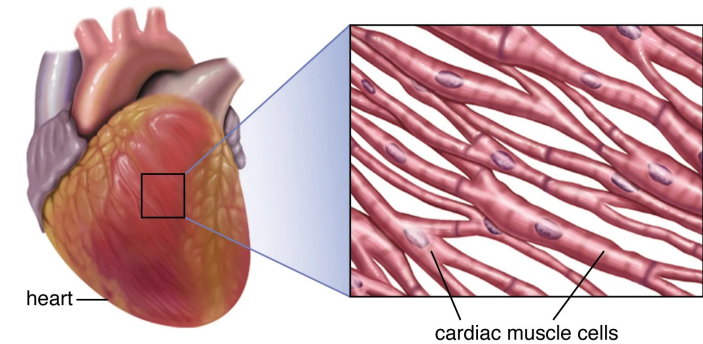
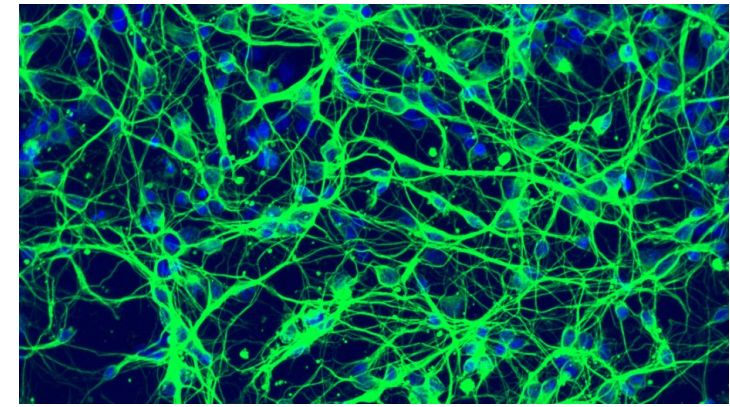
- Acquisition of transcriptome profile for single cell
- 1000 – 10,000 cells per condition; up to million of cells in a single experiment
- Cells can be grouped and defined based on resulting transcriptome profile, follow by group-based analyses, e.g., differential expression analysis
- Multiple technologies, varying in
  - Throughput
  - Sequencing / detection format & requirement
- Compartmentalization



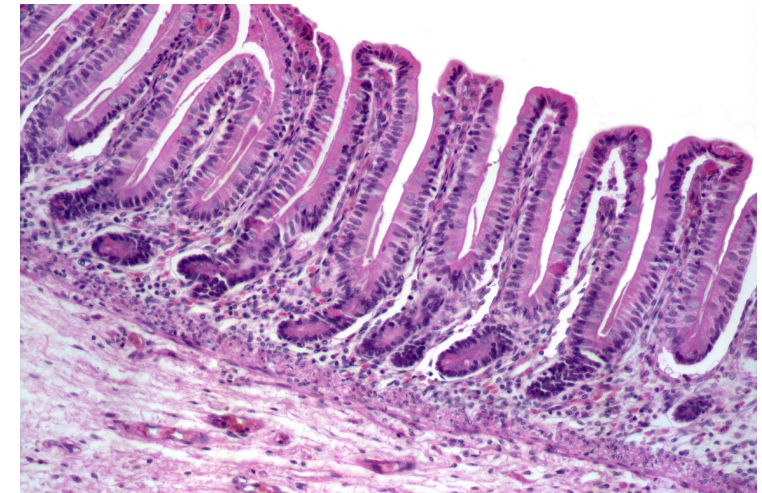
- Suitable for
  - High complexity input sample
    - Organs, whole organism, tissue, organoids, etc.
    - Unformal / Asynchronized cellular response / change
  - Morphologically indistinguishable/ambiguous, or lack of markers
    - (Immuno-)Histology or flow cytometry
    - Cell type / state that are not previously defined
    - Non-model organism
  - Rare cells / cell types
    - E.g, 1-2 rare neurons in L2 stage *C. elegans*
  - Related cell types
    - Cell differentiation
      - Neuronal cell
      - Hematopoietic Stem cell
    - Activation
      - Adaptive immune response



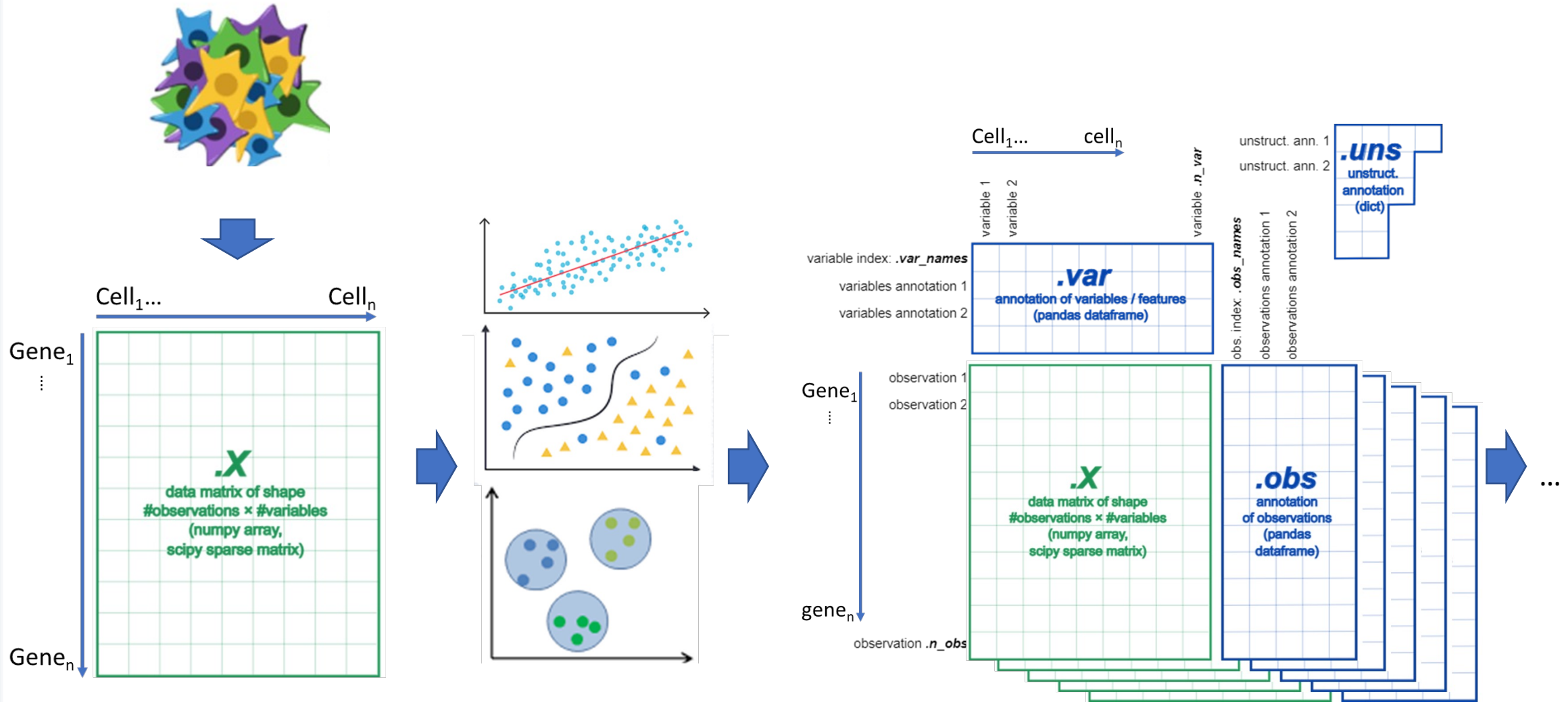
- Limitation
- Costly
  - Instrument, reagents & sequencing
- Incomplete transcriptome
  - Only mRNA transcripts that are highly expressed
  - Non-coding or small RNAs are omitted
  - Lack of isoform / splice variants information
  - Partial vs full transcript
- mRNA transcript level does not always correlate with protein level
  - Expression of known protein markers may not reflect in transcript data
- Single cell suspension required
  - Dead cell, damaged cell, nuclei only, cell fragment, incomplete / biased dissociation
  - Single nuclei as compromise for “difficult tissue”
- Complexity of data analyses / interpretation
  - Quality control, e.g., live/dead, fragment, multiple, etc
  - Cell classification, e.g., Unknown, subtype, non-model organism, etc
- Lack of spatial correlation



© Encyclopædia Britannica, Inc.



# Data structure



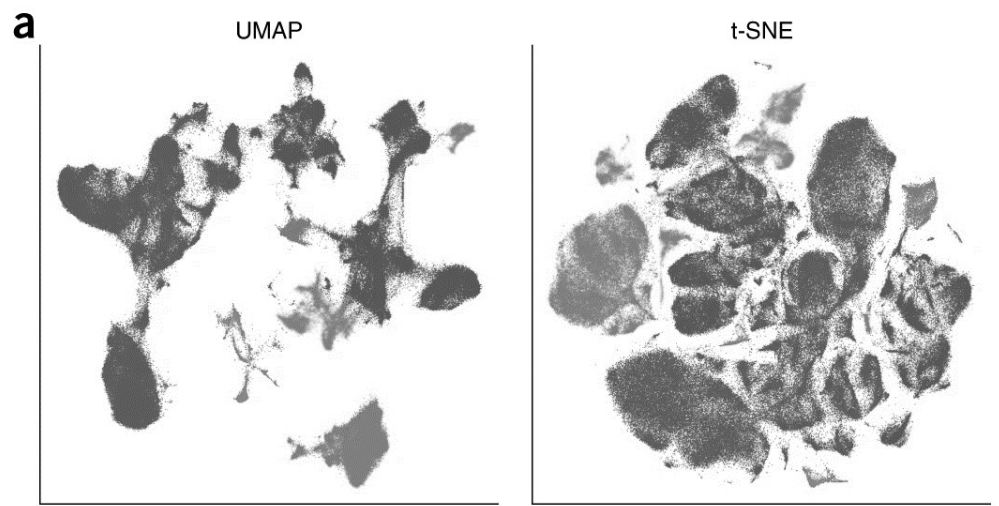
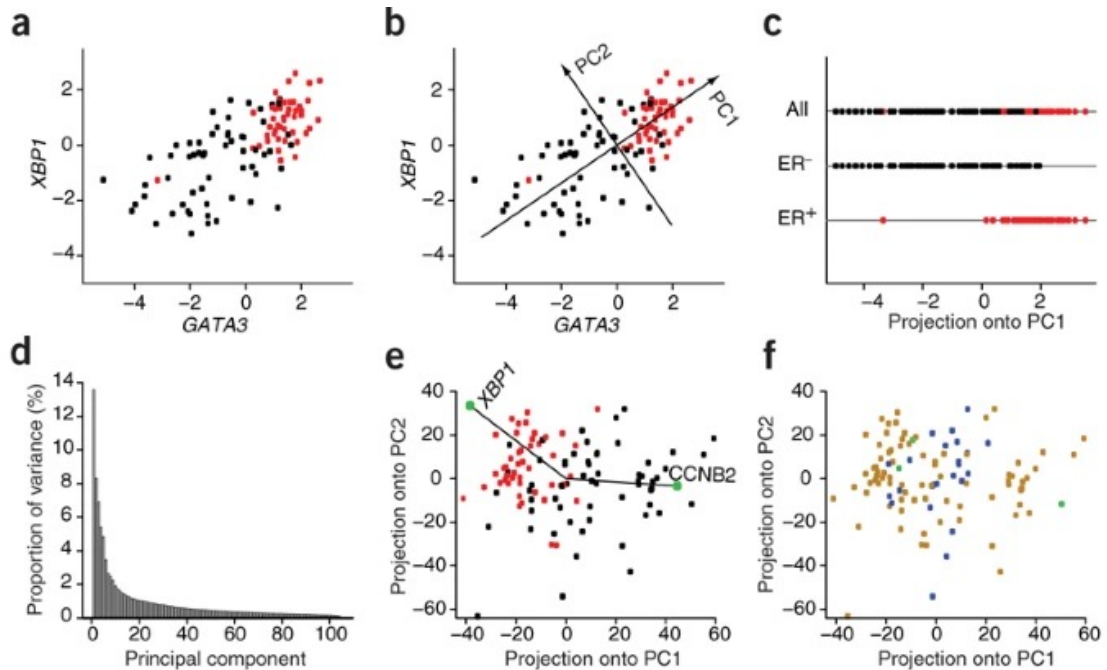
Raw count matrix

Filtering  
Normalization  
Dimension reduction  
Clustering  
Cell Assignment

Annotated count matrix

- Organs, sex, time-point, treatment
- Cluster, cell types, state, coordinate

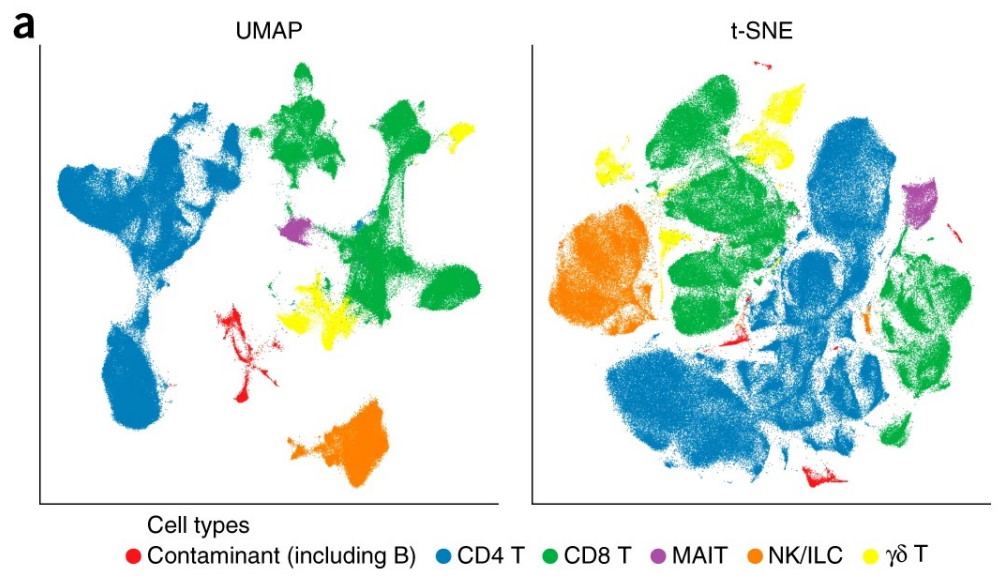
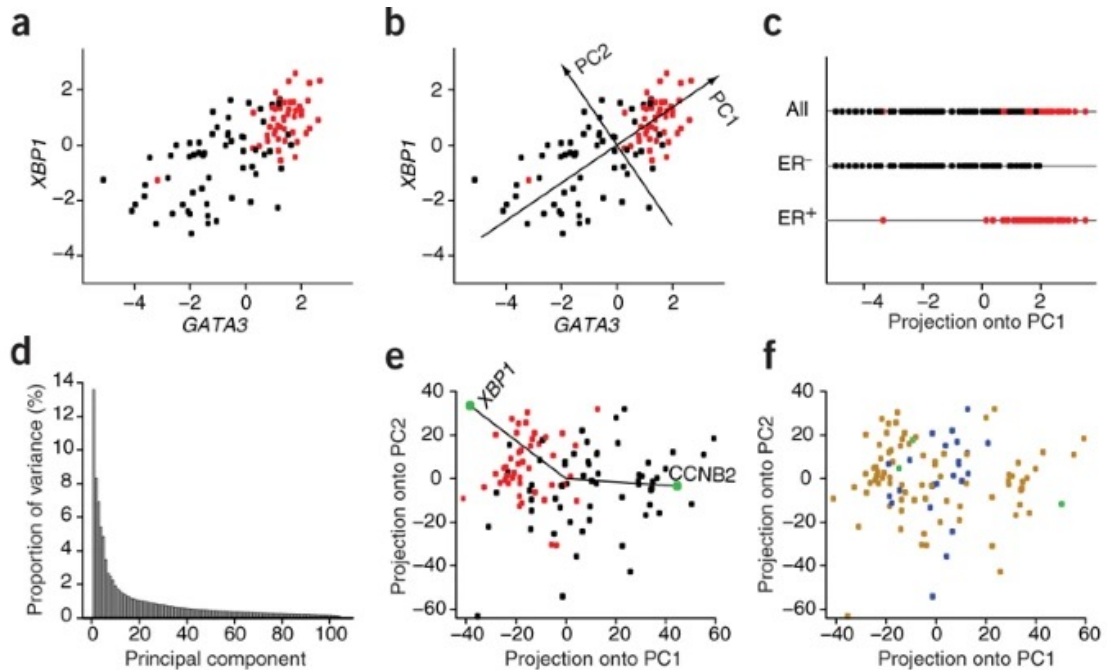
# Dimension reduction, clustering, cell assignment



Uniform Manifold Approximation and Projection & t-distributed stochastic neighbor embedding

<https://www.nature.com/articles/nbt0308-303>  
<https://www.nature.com/articles/nbt.4314>  
<https://www.nature.com/articles/s41598-020-59827-1>

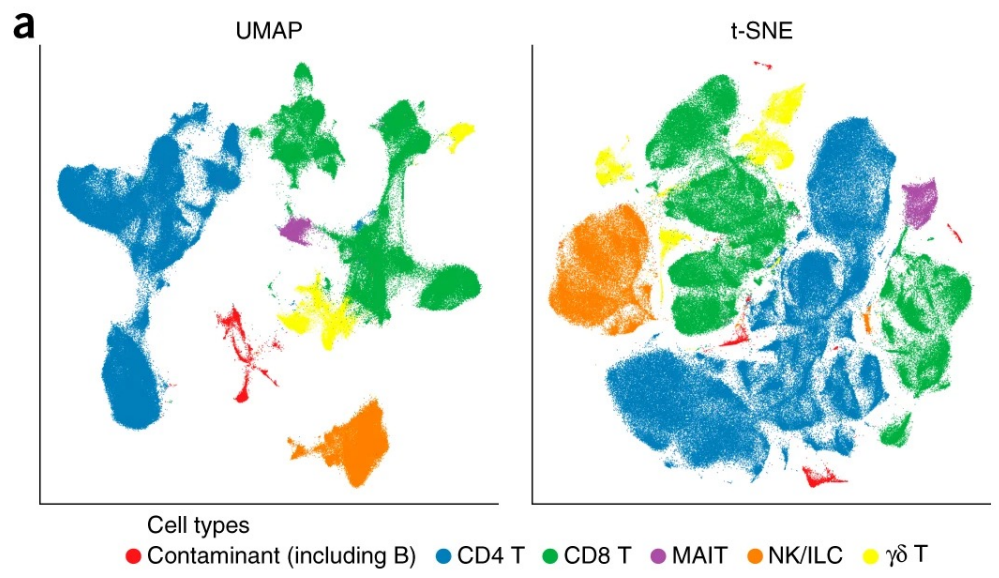
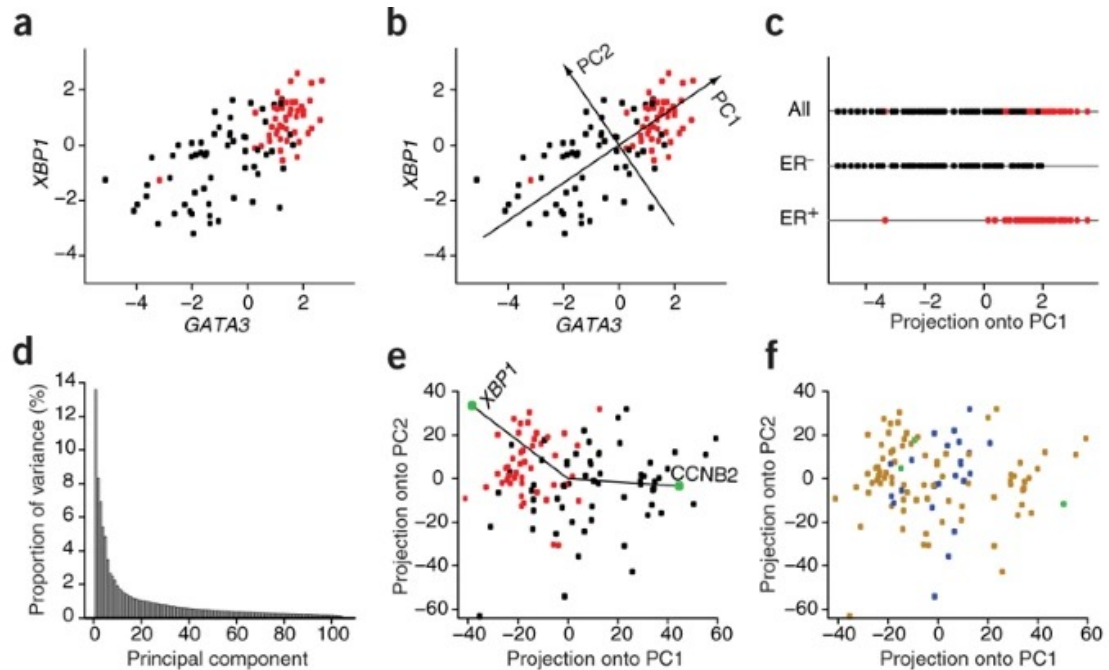
# Dimension reduction, clustering, cell assignment



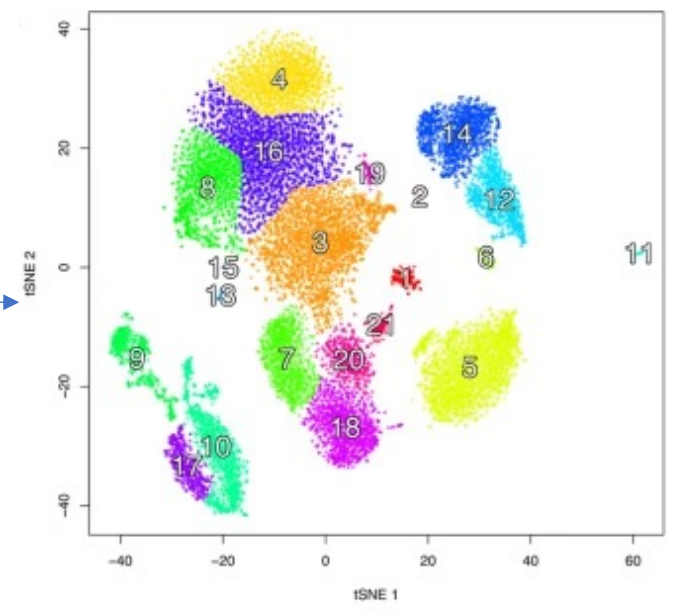
Uniform Manifold Approximation and Projection & t-distributed stochastic neighbor embedding

<https://www.nature.com/articles/nbt0308-303>  
<https://www.nature.com/articles/nbt.4314>  
<https://www.nature.com/articles/s41598-020-59827-1>

# Dimension reduction, clustering, cell assignment



K-nearest neighbor (KNN) graph

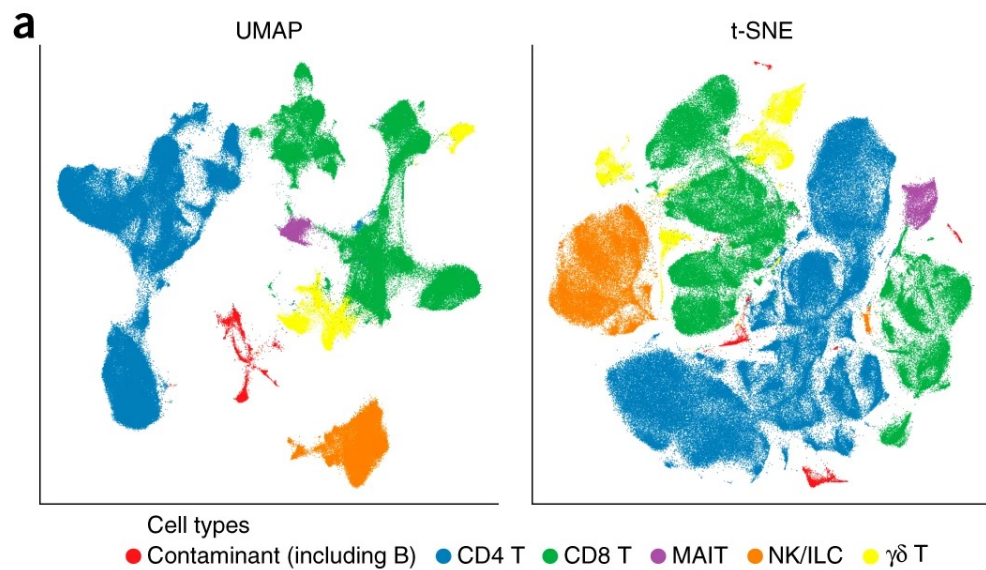
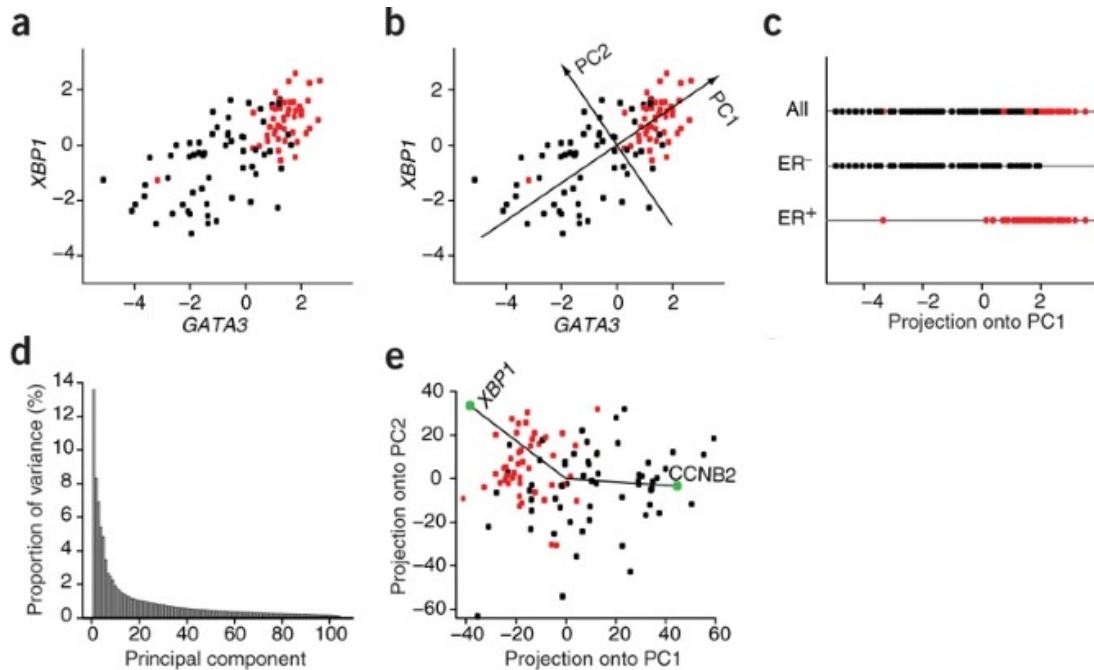


Uniform Manifold Approximation and Projection & t-distributed stochastic neighbor embedding

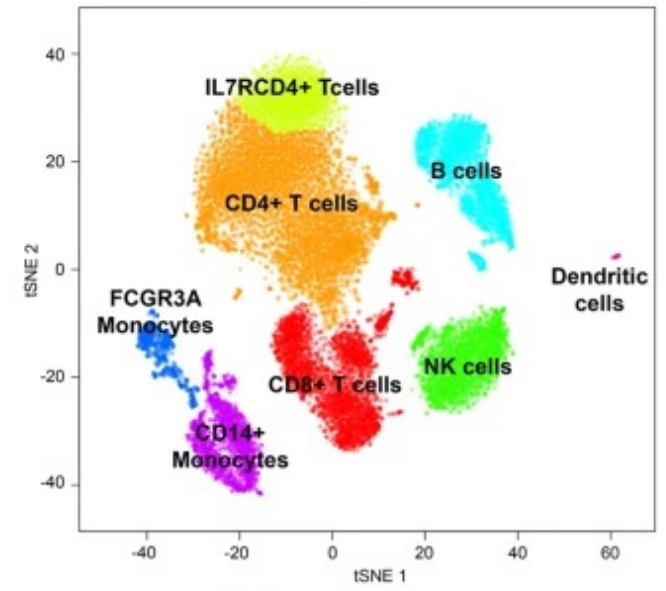
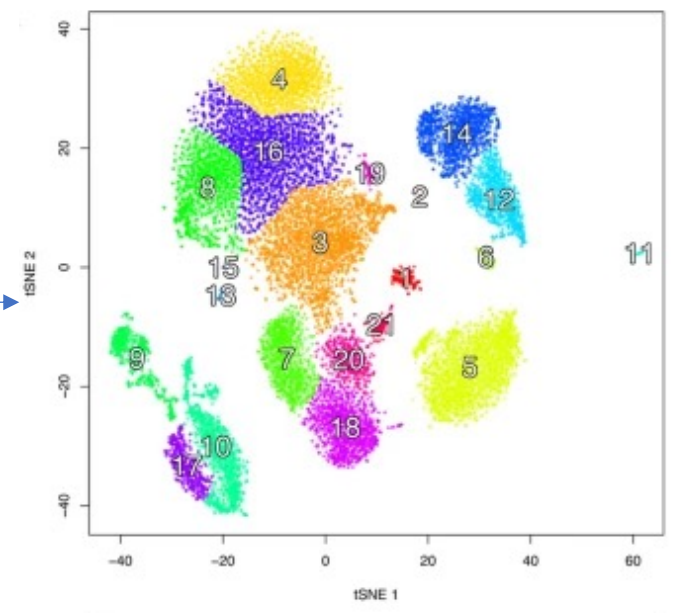
<https://www.nature.com/articles/nbt0308-303>  
<https://www.nature.com/articles/nbt.4314>  
<https://www.nature.com/articles/s41598-020-59827-1>



# Dimension reduction, clustering, cell assignment



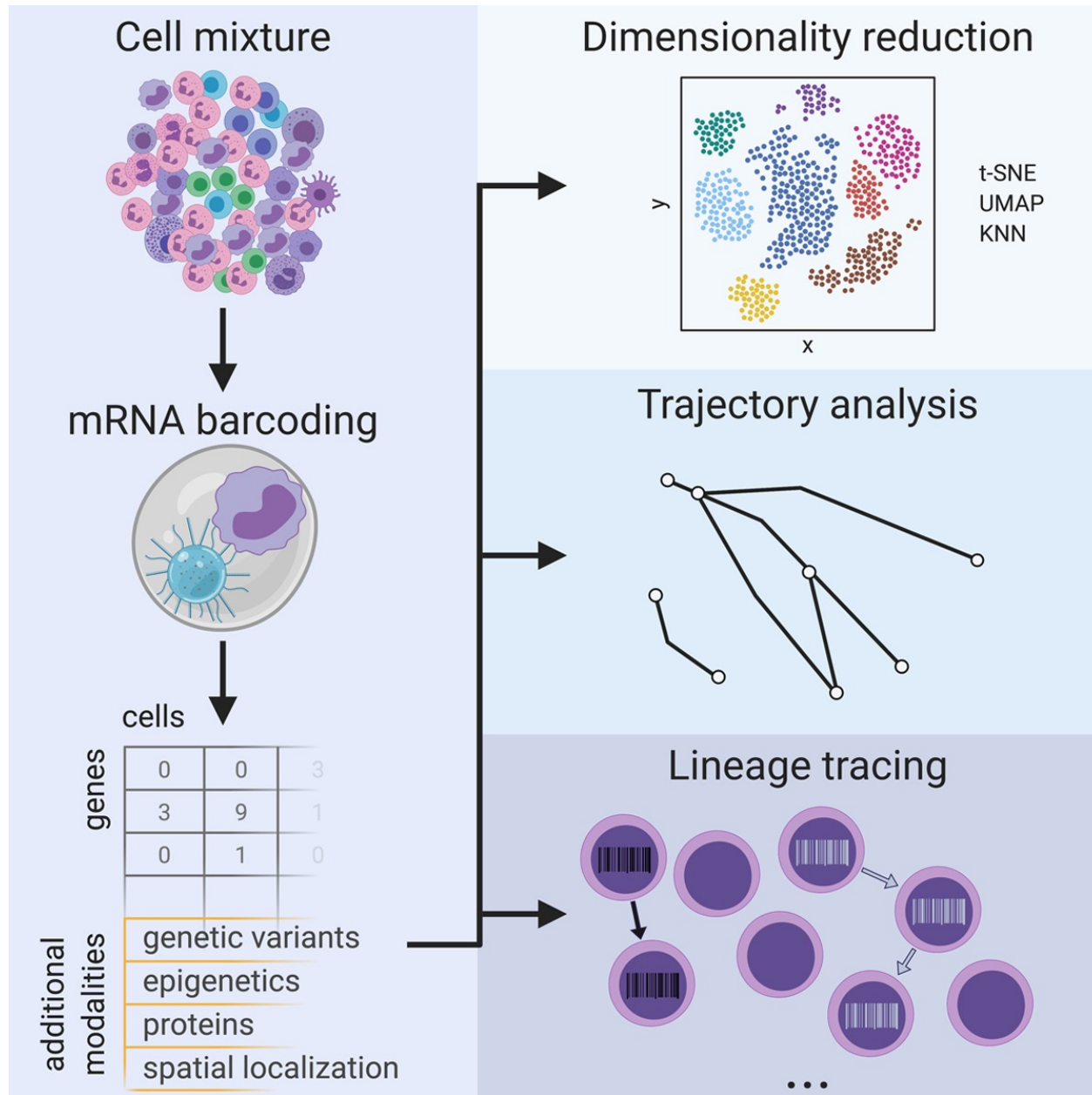
K-nearest neighbor (KNN) graph



Uniform Manifold Approximation and Projection & t-distributed stochastic neighbor embedding

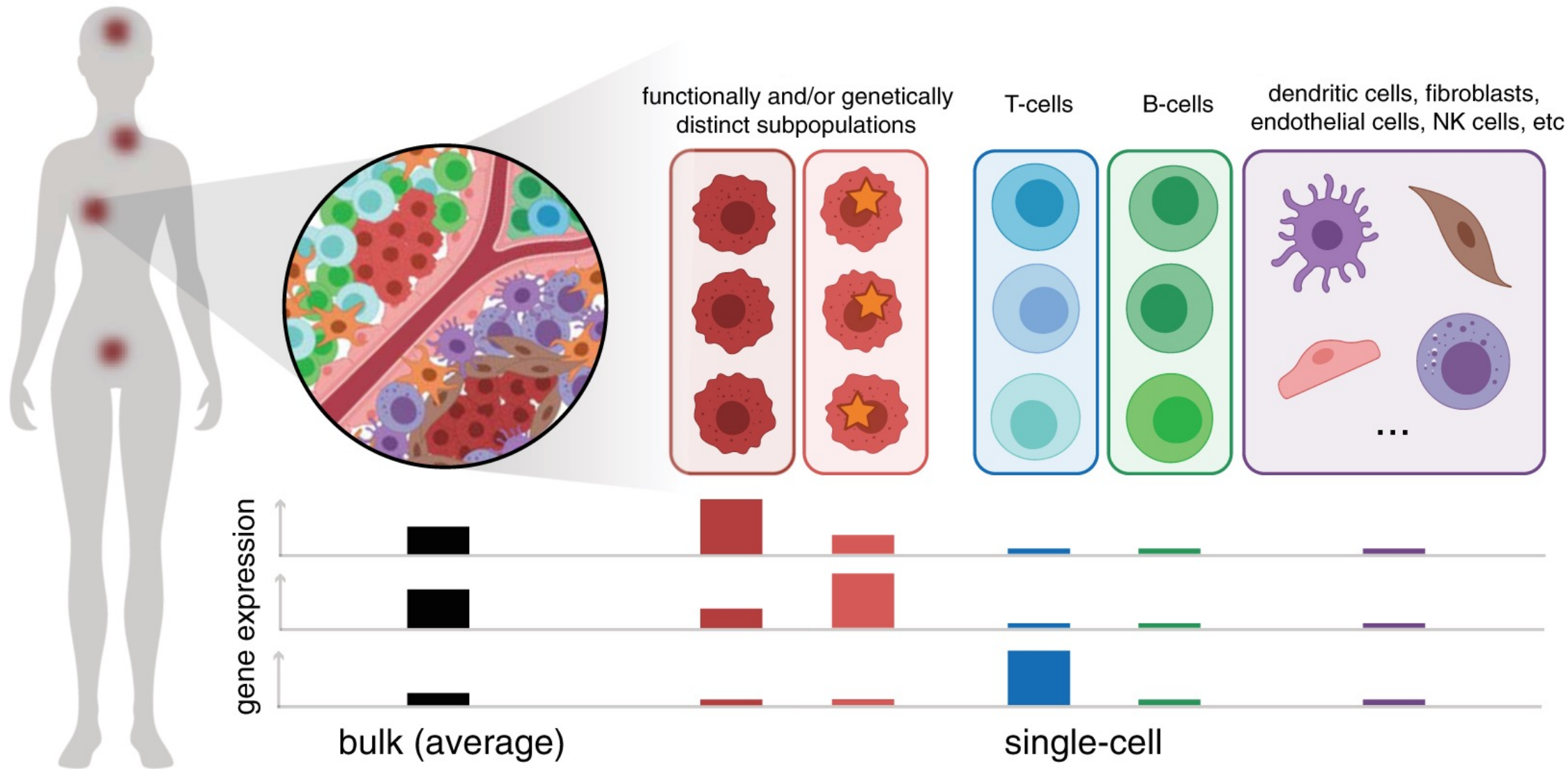
<https://www.nature.com/articles/nbt0308-303>  
<https://www.nature.com/articles/nbt.4314>  
<https://www.nature.com/articles/s41598-020-59827-1>

# What's next?

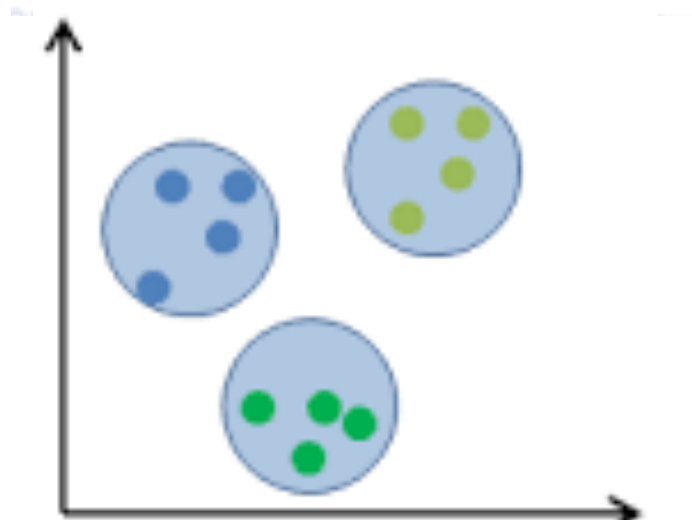
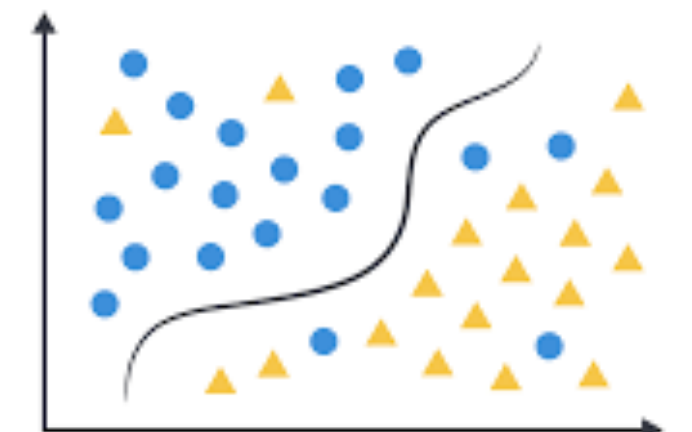
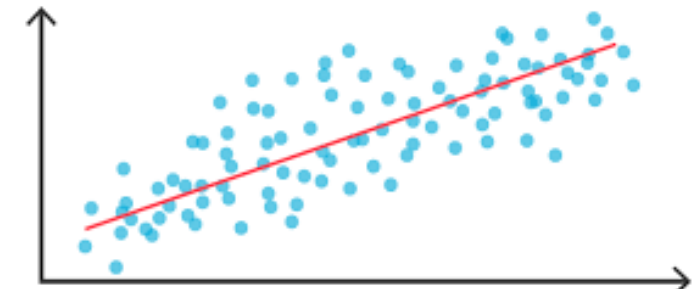
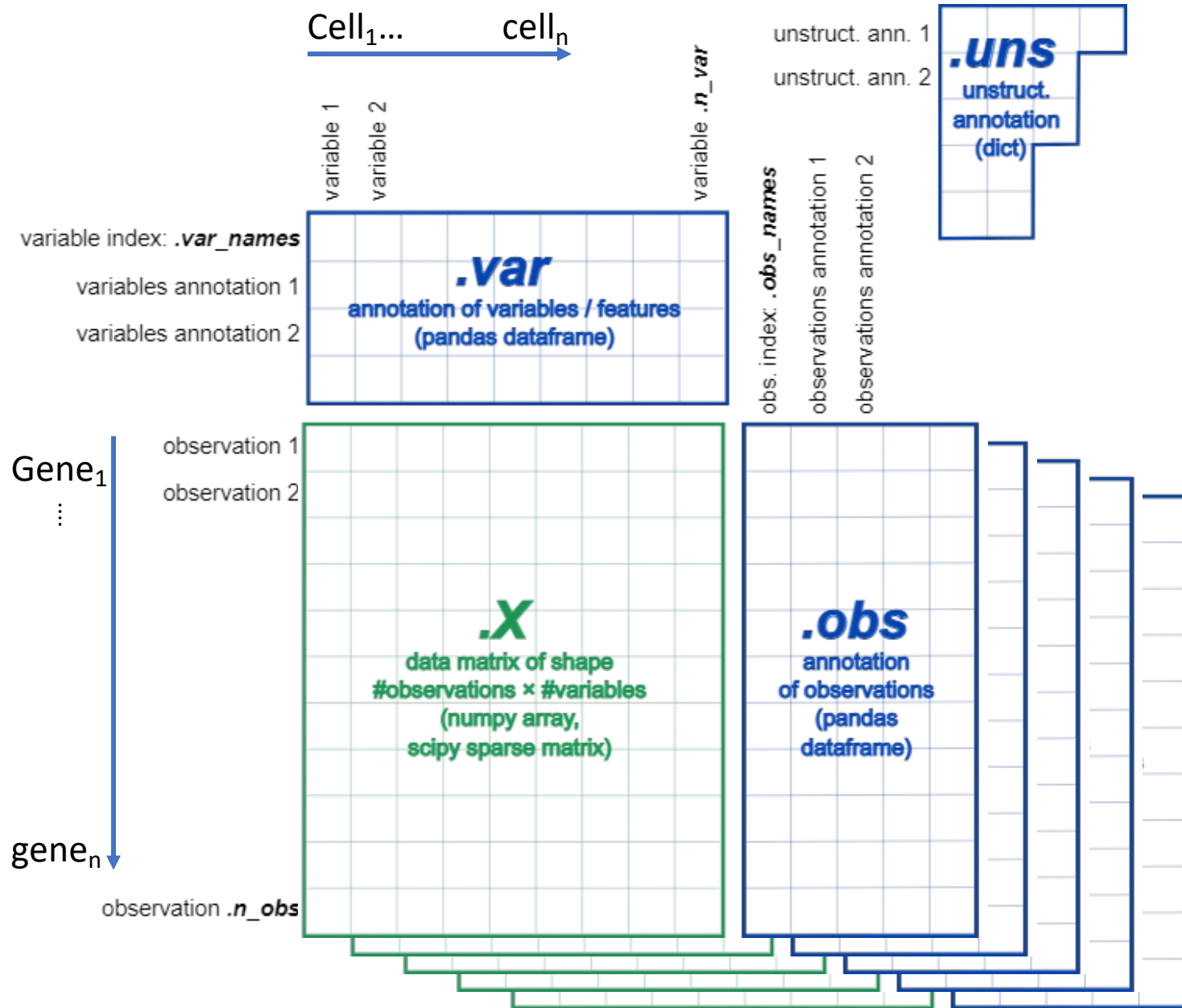


Extra slides

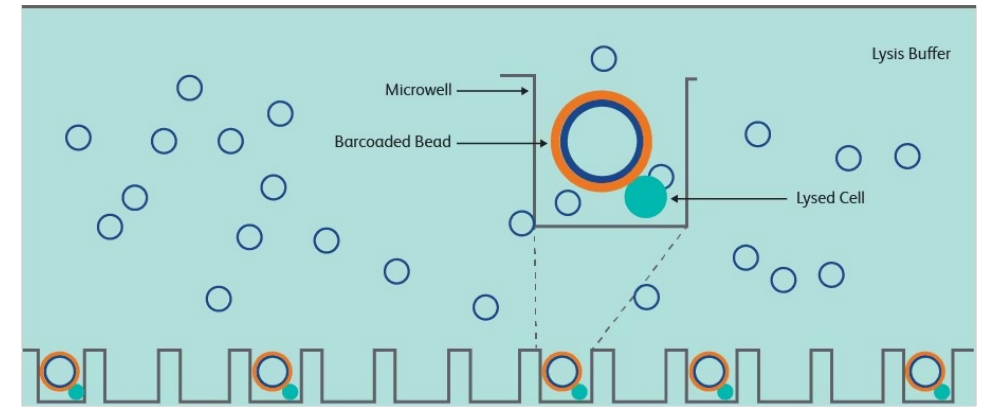
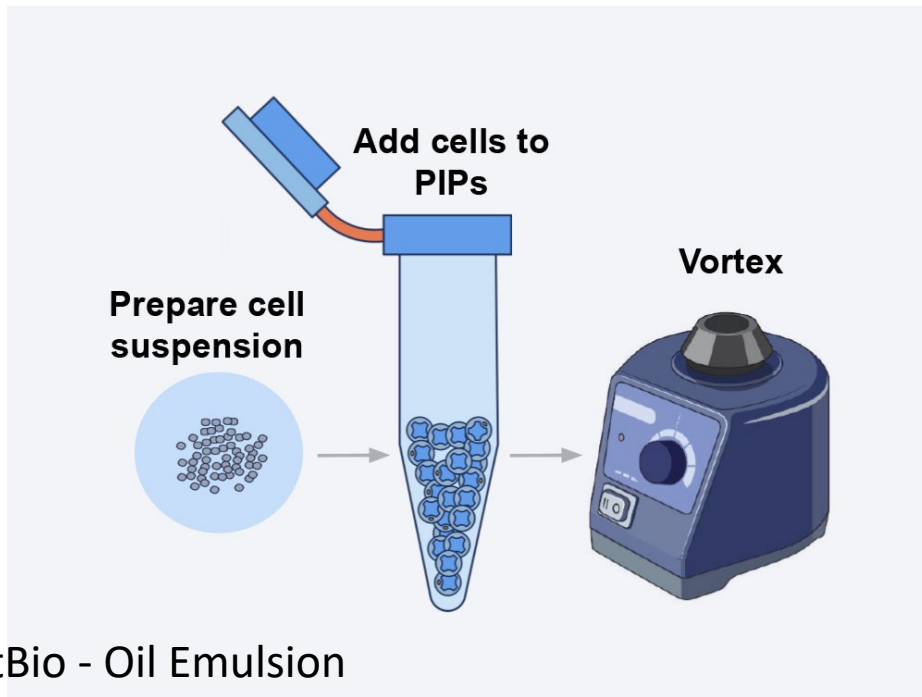
# Single cell RNA-Seq



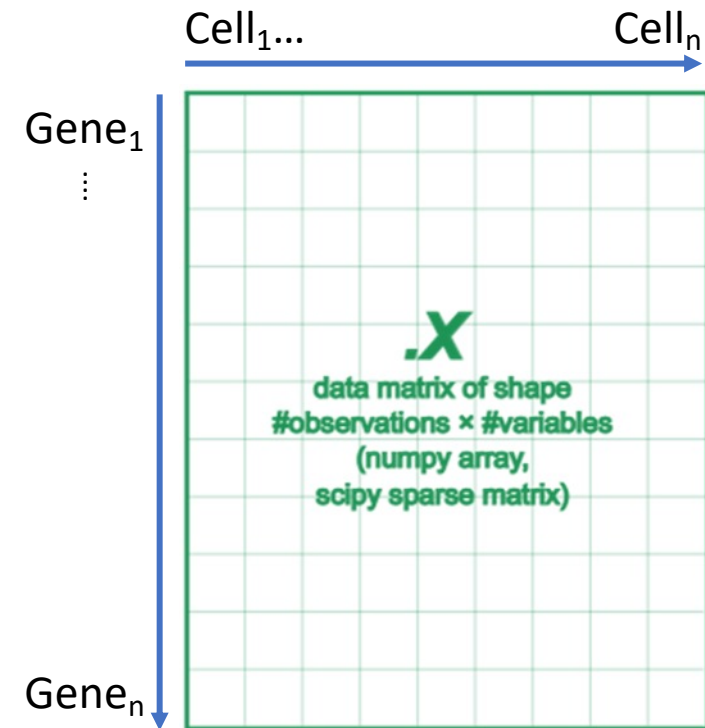
# Data structure



# Single cell RNA-Seq via Compartmentalization



BD Rhapsody - Microwells



# Single cell RNASeq

- A power technology that allows to acquire transcriptome profile within a single cell
- Typically, 1000 – 10,000 cells per conditions, up to million of cells in a single experiment
- Cells can be grouped and defined based on transcriptome profile, follow by group specific differential expression analysis
- Multiple technologies, varying in
  - Throughput
  - Sequencing / detection format & requirement
- High complexity input sample
  - Organs, whole organism, tissue
- Morphologically identical similar or lack of markers
  - (Immuno-)Histology or flow cytometry
  - Cell type / state that are not previously defined
  - Non-model organism
- Rare cells
- Costly
- Incomplete
  - Only genes that are highly expressed
  - mRNA only
    - Non-coding or small RNA omitted
  - Lack of isoform / splice variants information
- mRNA transcript level does not always correlates with protein level
  - Know protein marker may not reflect in transcript data