

Lecture 9: Single-Cell RNA-seq

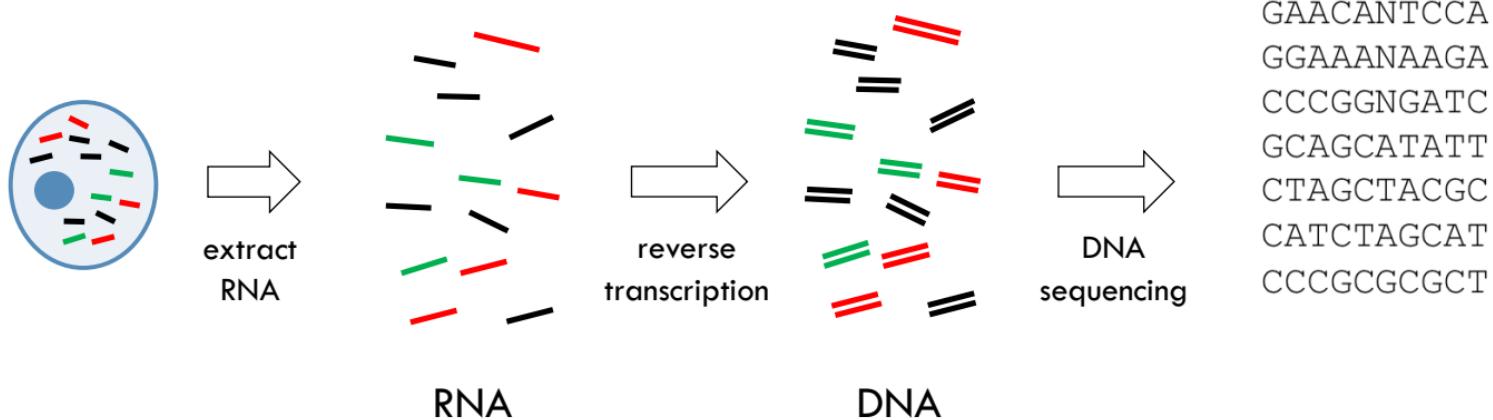


ECE 365

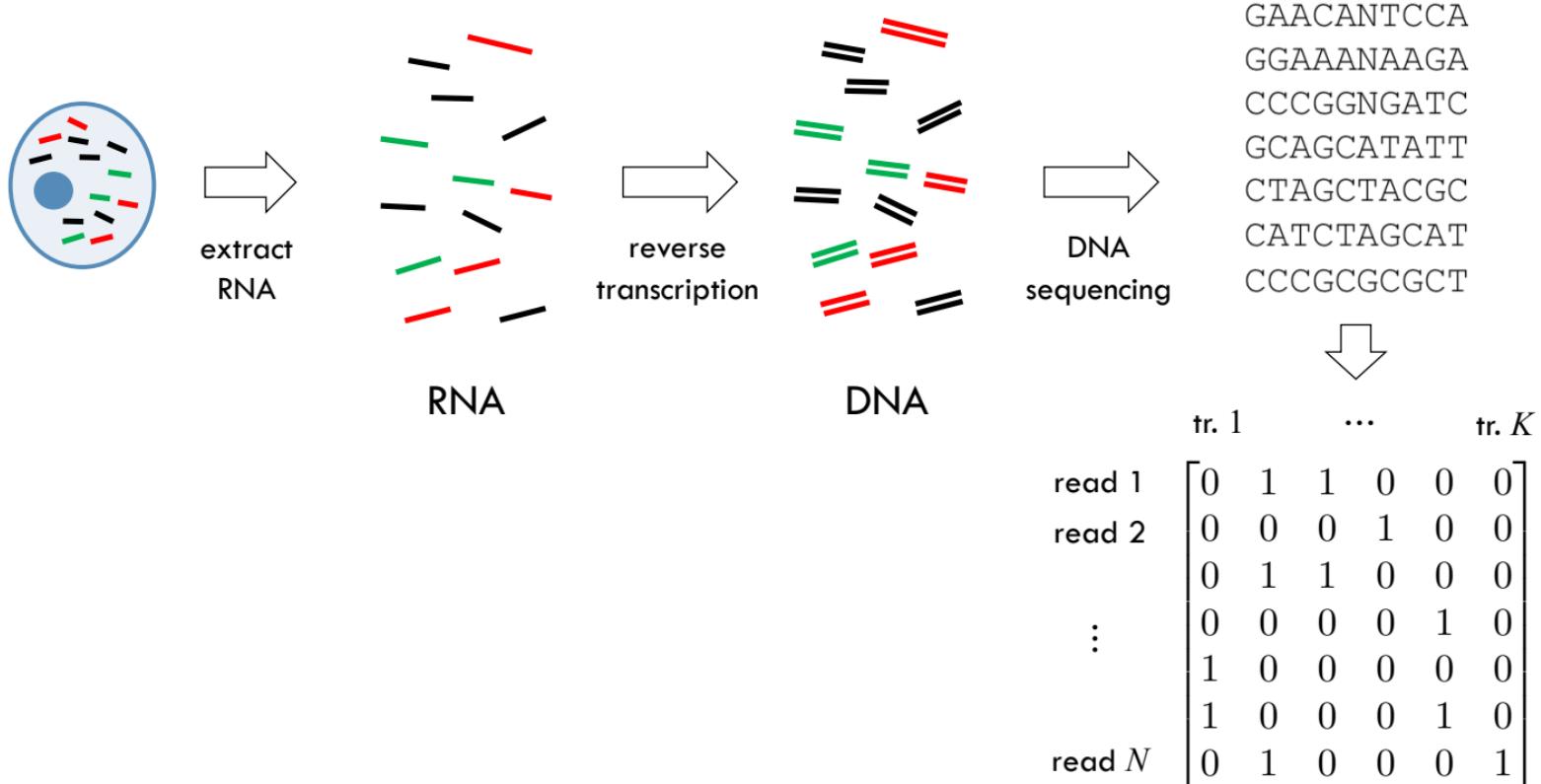
Announcements:

- Lab 4 (RNA-seq) due on April 2
- Quiz 2 on Thursday
 - ▣ Includes all Genomics material (with emphasis on the material since GWAS)

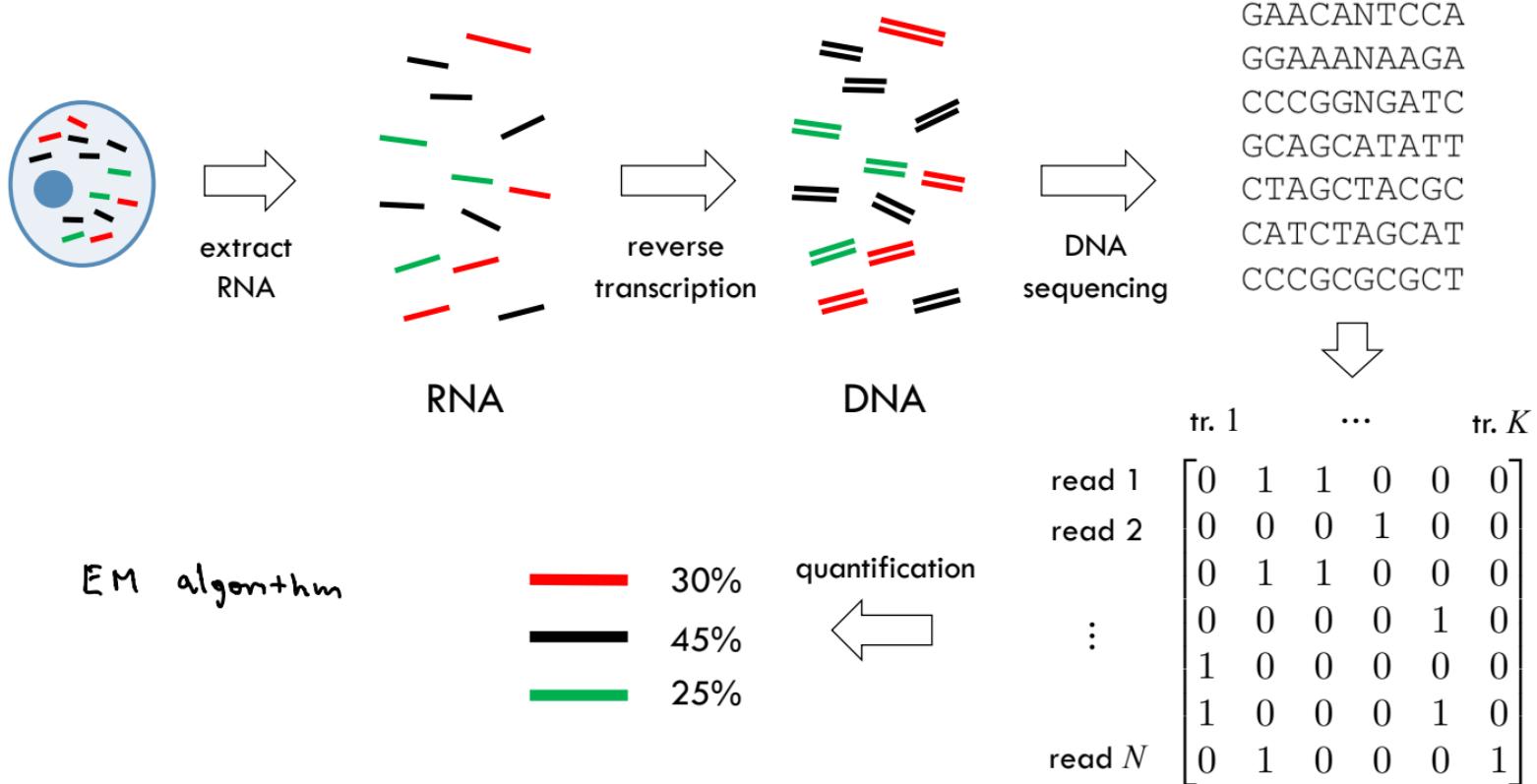
From previous lecture: RNA quantification



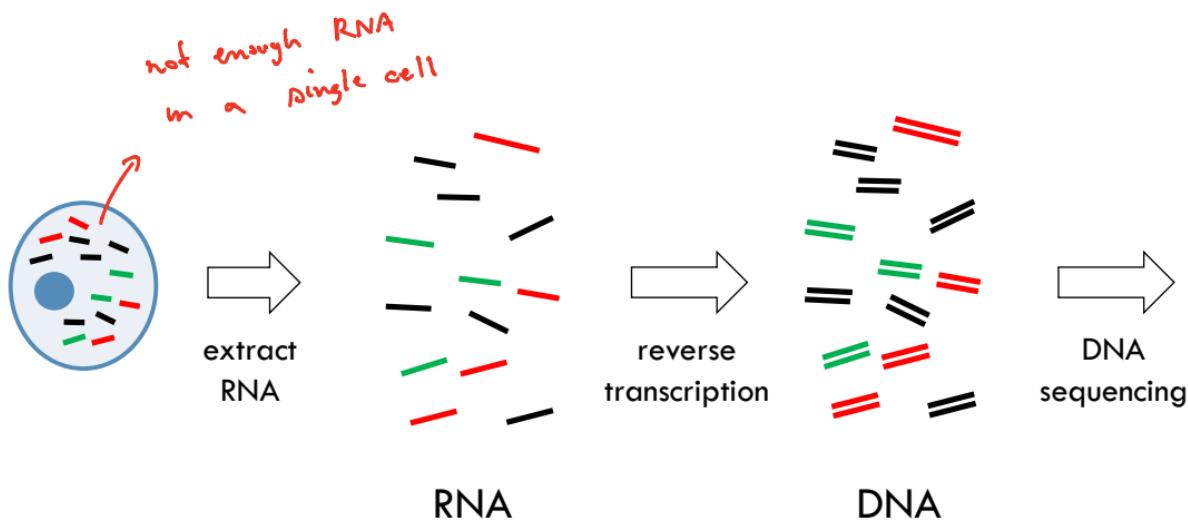
From previous lecture: RNA quantification



From previous lecture: RNA quantification



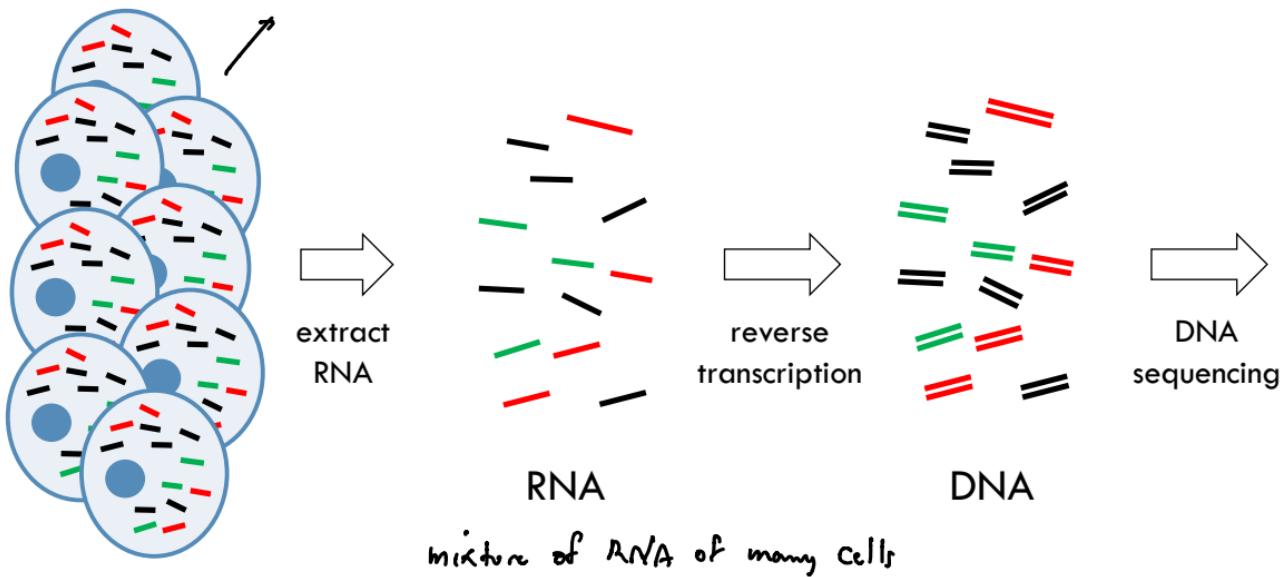
RNA-seq



CCTTCNCACT
TTTTTNCAGA
GAACANTCCA
GGAAANAAGA
CCCGGNGATC
GCAGCATATT
CTAGCTACGC
CATCTAGCAT
CCCGCGCGCT
TCACACATCT
CATGCTAGCT

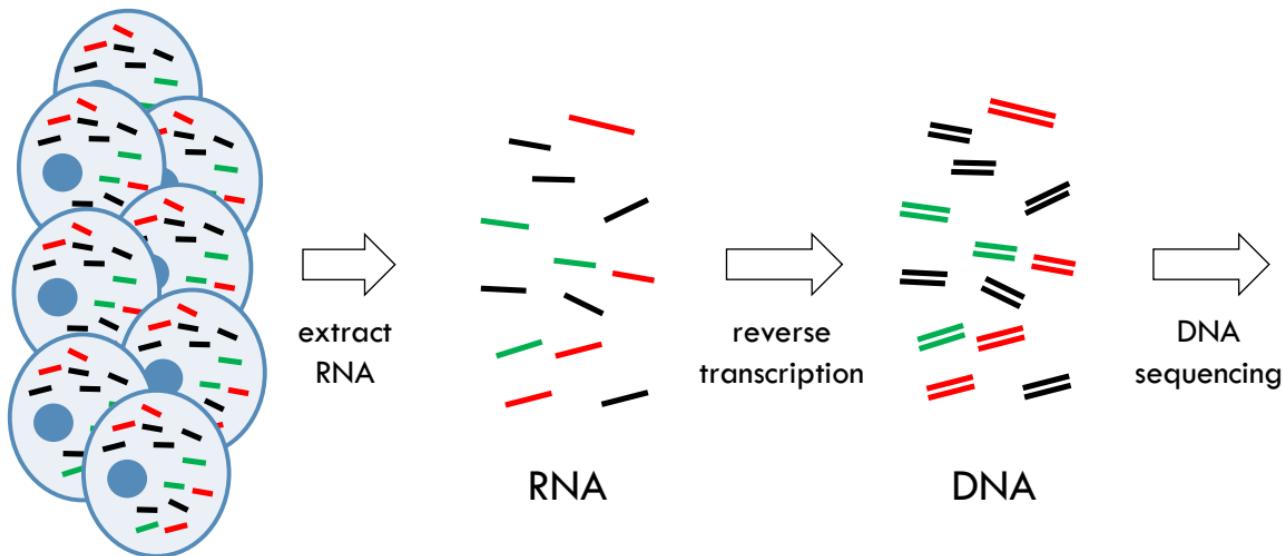
Bulk RNA-seq

cells from same tissue : brain, liver, blood



CCTTCNCACT
TTTTTNCAGA
GAACANTCCA
GGAAANAAGA
CCCGGNGATC
GCAGCATATT
CTAGCTACGC
CATCTAGCAT
CCCGCGCGCT
TCACACATCT
CATGCTAGCT

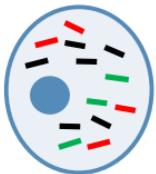
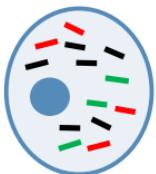
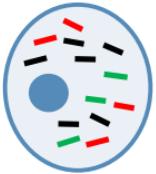
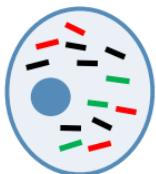
Bulk RNA-seq



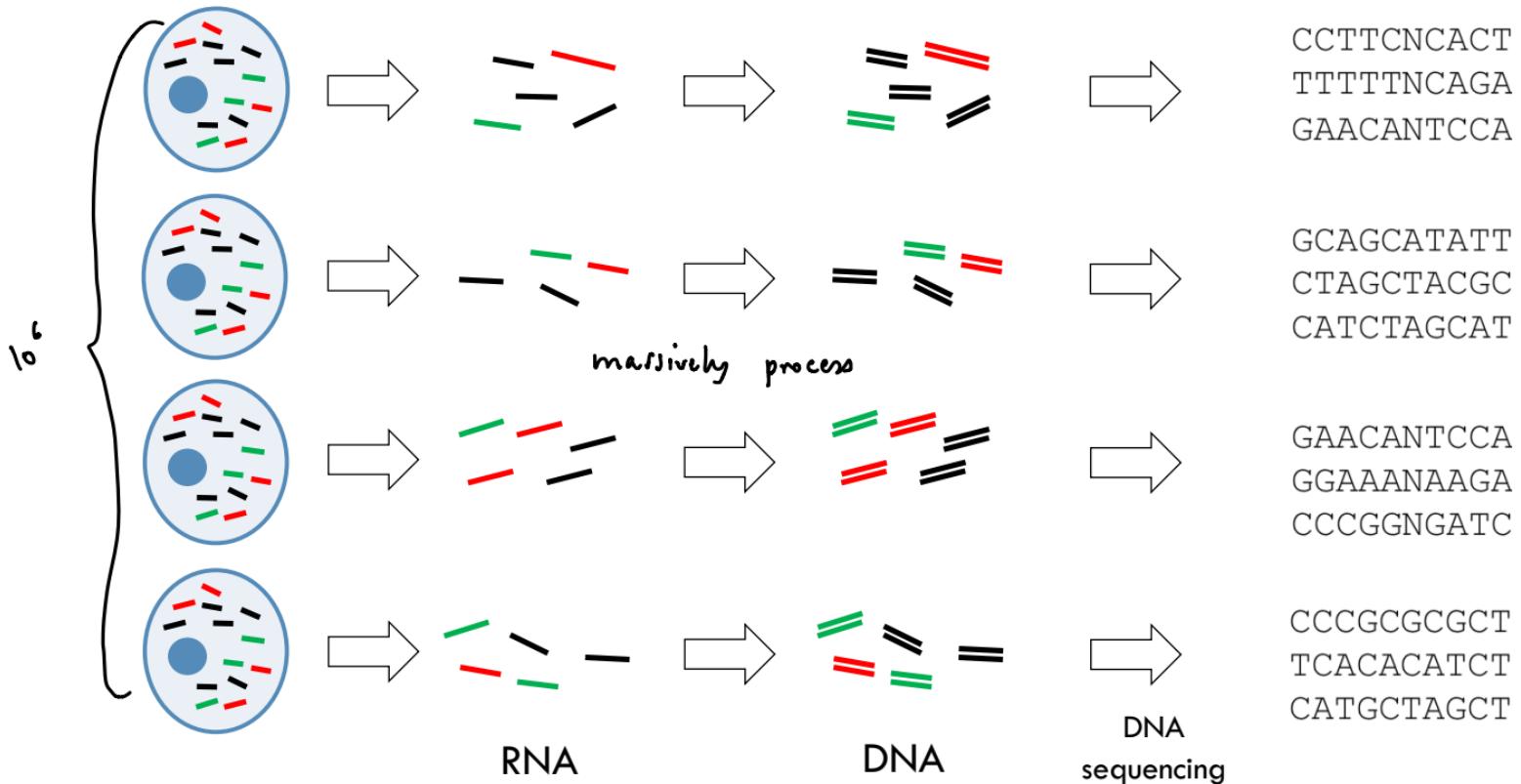
CCTTCNCACT
TTTTTNCAGA
GAACANTCCA
GGAAANAAGA
CCCGGNGATC
GCAGCATATT
CTAGCTACGC
CATCTAGCAT
CCCGCGCGCT
TCACACATCT
CATGCTAGCT

What if there are differences between individual cells?

New technology: Single-Cell RNA-seq

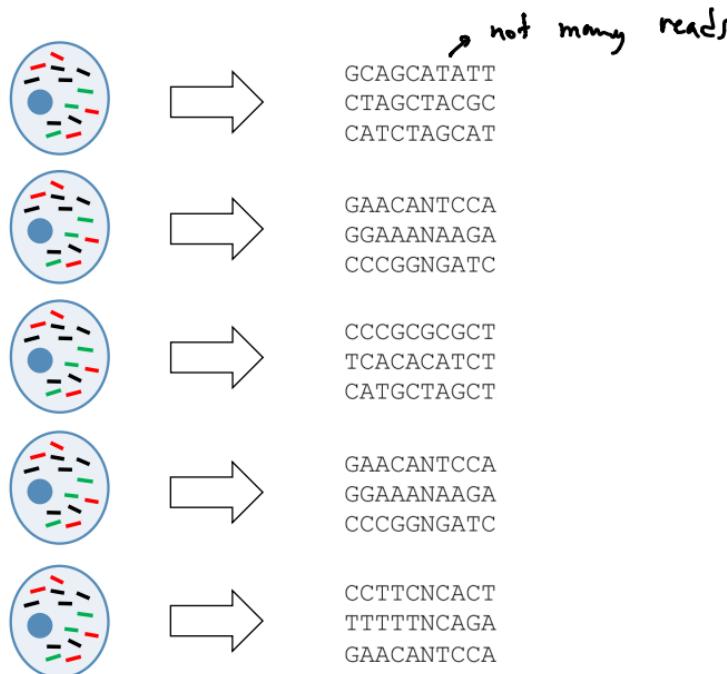


New technology: Single-Cell RNA-seq



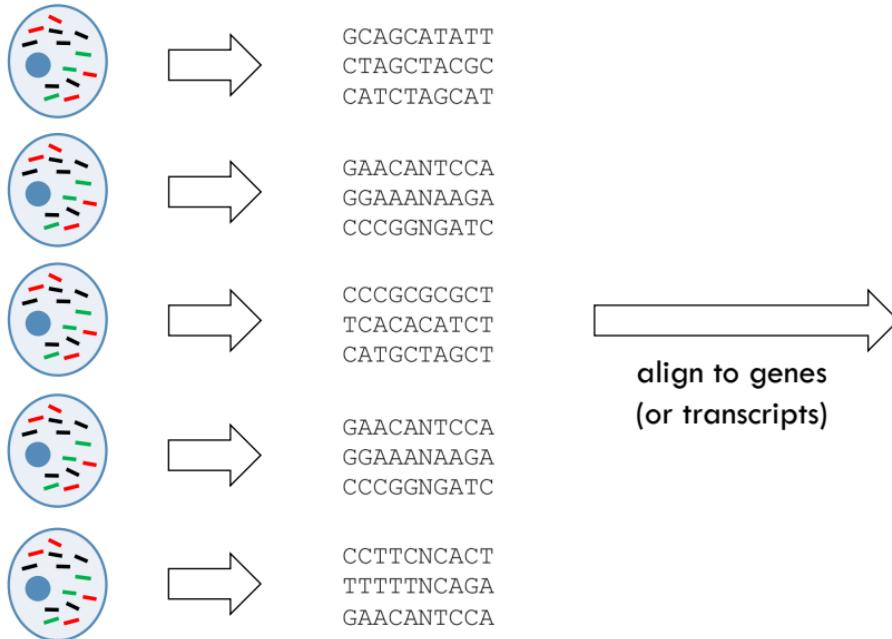
New technology: Single-Cell RNA-seq

- We obtain reads for each cell



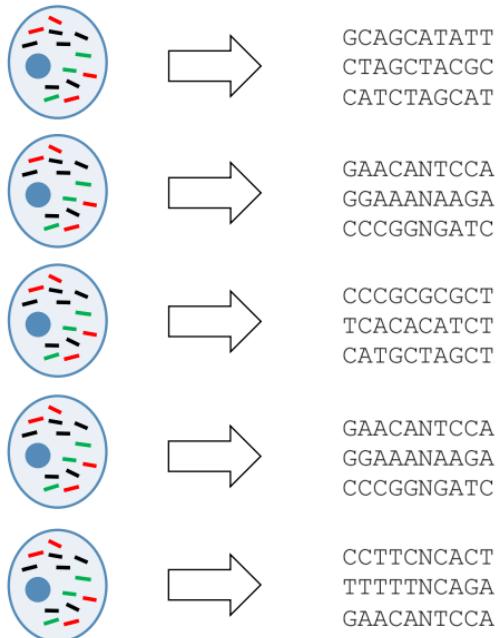
New technology: Single-Cell RNA-seq

- We obtain reads for each cell



New technology: Single-Cell RNA-seq

- We obtain reads for each cell



align to genes
(or ~~transcripts~~)
not enough read data

| | gene 1 | ... | gene m | | | |
|----------|--------|-----|----------|---|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| cell 3 | 0 | 2 | 4 | 0 | 0 | 0 |
| cell 4 | 0 | 0 | 0 | 0 | 9 | 0 |
| cell 5 | 1 | 0 | 0 | 0 | 0 | 0 |
| ⋮ | 7 | 0 | 0 | 0 | 0 | 0 |
| cell n | 0 | 6 | 1 | 0 | 0 | 1 |

read count

Single-Cell RNA-seq Data Analysis

- Genes x Cells matrix with read counts

| | gene 1 | ... | gene m | | |
|--------|--------|-----|--------|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 |
| | 0 | 2 | 4 | 0 | 0 |
| : | 0 | 0 | 0 | 0 | 9 |
| | 1 | 0 | 0 | 0 | 0 |
| | 7 | 0 | 0 | 0 | 8 |
| cell n | 0 | 6 | 1 | 0 | 0 |
| | | | | | 1 |

Annotations:

- The first three rows (cell 1, cell 2, and cell 3) are highlighted with red boxes.
- A red arrow points from the text "similar" to the second row (cell 2).
- A red arrow points from the text "different" to the third row (cell 3).

Single-Cell RNA-seq Data Analysis

- Genes x Cells matrix with read counts

| | gene 1 | ... | | gene <i>m</i> | |
|---------------|--------|-----|---|---------------|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 |
| | 0 | 2 | 4 | 0 | 0 |
| : | 0 | 0 | 0 | 0 | 9 |
| | 1 | 0 | 0 | 0 | 0 |
| | 7 | 0 | 0 | 0 | 8 |
| cell <i>n</i> | 0 | 6 | 1 | 0 | 0 |
| | | | | | 1 |

- Are there different types of cells in the mixture?

Single-Cell RNA-seq Data Analysis

- Genes x Cells matrix with read counts

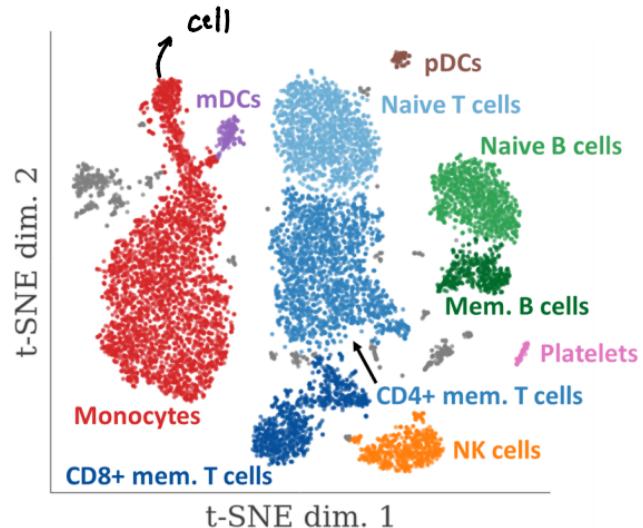
| | 20,000 | | | | | |
|--------|--------|---|-----|--------|---|---|
| | gene 1 | | ... | gene m | | |
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| | 0 | 2 | 4 | 0 | 0 | 0 |
| : | 0 | 0 | 0 | 0 | 9 | 0 |
| | 1 | 0 | 0 | 0 | 0 | 0 |
| | 7 | 0 | 0 | 0 | 8 | 0 |
| cell n | 0 | 6 | 1 | 0 | 0 | 1 |

- Are there different types of cells in the mixture?
- How can we visualize this high-dimensional data?

Single-Cell RNA-seq Data Analysis

- Genes x Cells matrix with read counts

| | gene 1 | ... | | gene <i>m</i> | |
|---------------|--------|-----|---|---------------|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 |
| | 0 | 2 | 4 | 0 | 0 |
| : | 0 | 0 | 0 | 0 | 9 |
| | 1 | 0 | 0 | 0 | 0 |
| | 7 | 0 | 0 | 0 | 8 |
| cell <i>n</i> | 0 | 6 | 1 | 0 | 0 |
| | | | | | 1 |



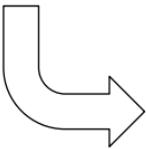
- Are there different types of cells in the mixture?
- How can we visualize this high-dimensional data?

Single-Cell RNA-seq Data Analysis

| | gene 1 | ... | | gene <i>m</i> | |
|---------------|--------|-----|---|---------------|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 |
| ⋮ | 0 | 2 | 4 | 0 | 0 |
| ⋮ | 0 | 0 | 0 | 9 | 0 |
| ⋮ | 1 | 0 | 0 | 0 | 0 |
| ⋮ | 7 | 0 | 0 | 0 | 8 |
| cell <i>n</i> | 0 | 6 | 1 | 0 | 0 |
| | | | | | 1 |

Single-Cell RNA-seq Data Analysis

| | gene 1 | ... | gene <i>m</i> | | | |
|---------------|--------|-----|---------------|---|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| ⋮ | 0 | 2 | 4 | 0 | 0 | 0 |
| ⋮ | 0 | 0 | 0 | 0 | 9 | 0 |
| ⋮ | 1 | 0 | 0 | 0 | 0 | 0 |
| ⋮ | 7 | 0 | 0 | 0 | 8 | 0 |
| cell <i>n</i> | 0 | 6 | 1 | 0 | 0 | 1 |



$$\ln(1 + x)$$

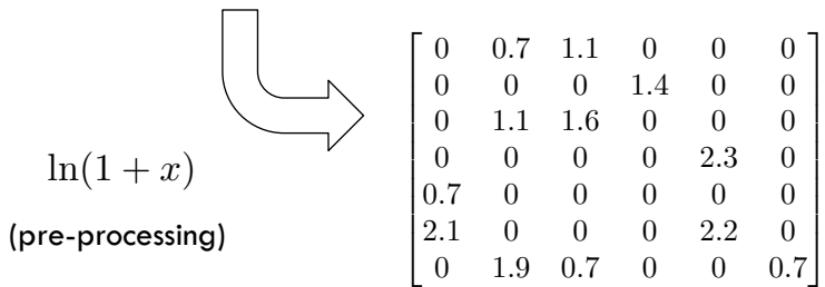
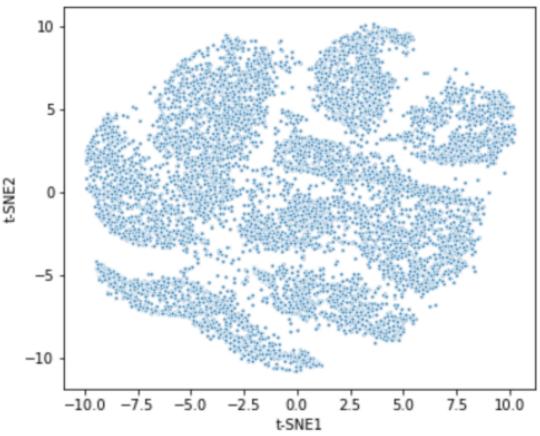
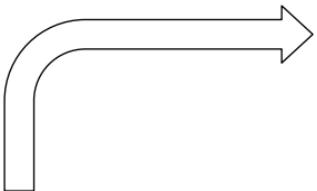
(pre-processing)

$$\begin{bmatrix} 0 & 0.7 & 1.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.4 & 0 & 0 \\ 0 & 1.1 & 1.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.3 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 \\ 2.1 & 0 & 0 & 0 & 2.2 & 0 \\ 0 & 1.9 & 0.7 & 0 & 0 & 0.7 \end{bmatrix}$$

Single-Cell RNA-seq Data Analysis

| | gene 1 | ... | gene <i>m</i> | | | |
|--------|--------|-----|---------------|---|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| ⋮ | 0 | 2 | 4 | 0 | 0 | 0 |
| ⋮ | 0 | 0 | 0 | 0 | 9 | 0 |
| ⋮ | 1 | 0 | 0 | 0 | 0 | 0 |
| ⋮ | 7 | 0 | 0 | 0 | 8 | 0 |
| ⋮ | 0 | 6 | 1 | 0 | 0 | 1 |

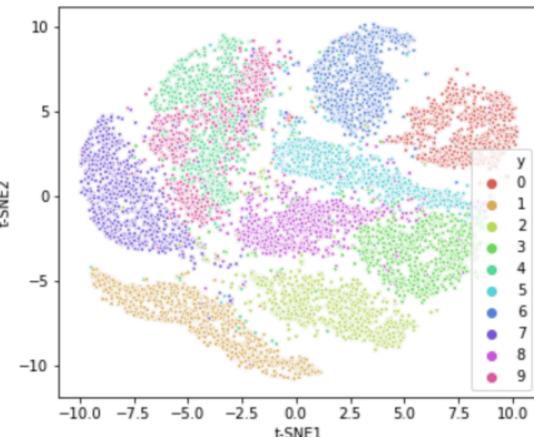
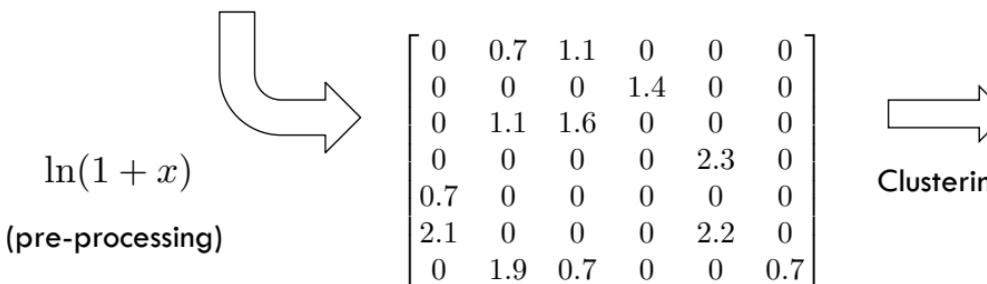
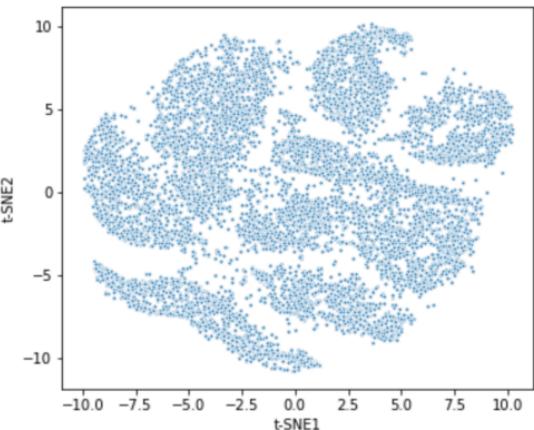
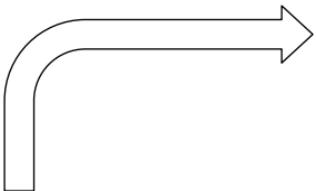
Dimensionality reduction
and visualization



Single-Cell RNA-seq Data Analysis

| | gene 1 | ... | gene <i>m</i> | | | |
|--------|--------|-----|---------------|---|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| ⋮ | 0 | 2 | 4 | 0 | 0 | 0 |
| ⋮ | 0 | 0 | 0 | 0 | 9 | 0 |
| ⋮ | 1 | 0 | 0 | 0 | 0 | 0 |
| ⋮ | 7 | 0 | 0 | 0 | 8 | 0 |
| ⋮ | 0 | 6 | 1 | 0 | 0 | 1 |

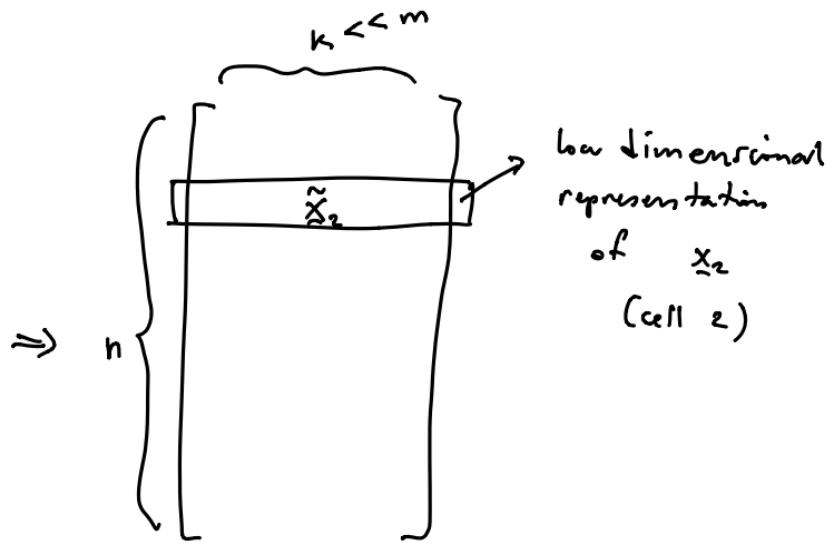
Dimensionality reduction
and visualization



Dimensionality reduction

$m = 20,000$

| | gene 1 | ... | gene m | | | |
|----------|--------|-----|----------|---|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| : | 0 | 2 | 4 | 0 | 0 | 0 |
| | 0 | 0 | 0 | 0 | 9 | 0 |
| | 1 | 0 | 0 | 0 | 0 | 0 |
| | 7 | 0 | 0 | 0 | 8 | 0 |
| cell n | 0 | 6 | 1 | 0 | 0 | 1 |



One approach: PCA

Linear transformation : $\tilde{\underline{x}}_i = W \underline{x}_i$

Dimensionality reduction via PCA

Data matrix \mathbf{X} ($n \times m$)

Perform SVD on \mathbf{X} : $\mathbf{X} = \mathbf{U}_{n \times n} \mathbf{S}_{n \times m} \mathbf{V}^T_{m \times m}$

use this to
get transformation \mathbf{W}

$$\mathbf{V}^T_{m \times m}$$

$$\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots & \sigma_m \end{bmatrix}$$

singular values

$$\sigma_1 > \sigma_2 > \dots > \sigma_m$$

$$\mathbf{V}^T = \left[\begin{array}{c} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_m \end{array} \right] \xrightarrow{\text{pick top } k \text{ rows}} \left[\begin{array}{c} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k \end{array} \right] = \mathbf{W}_{k \times m}$$

now we can find $\tilde{\mathbf{x}}_i = \mathbf{W} \mathbf{x}_i$

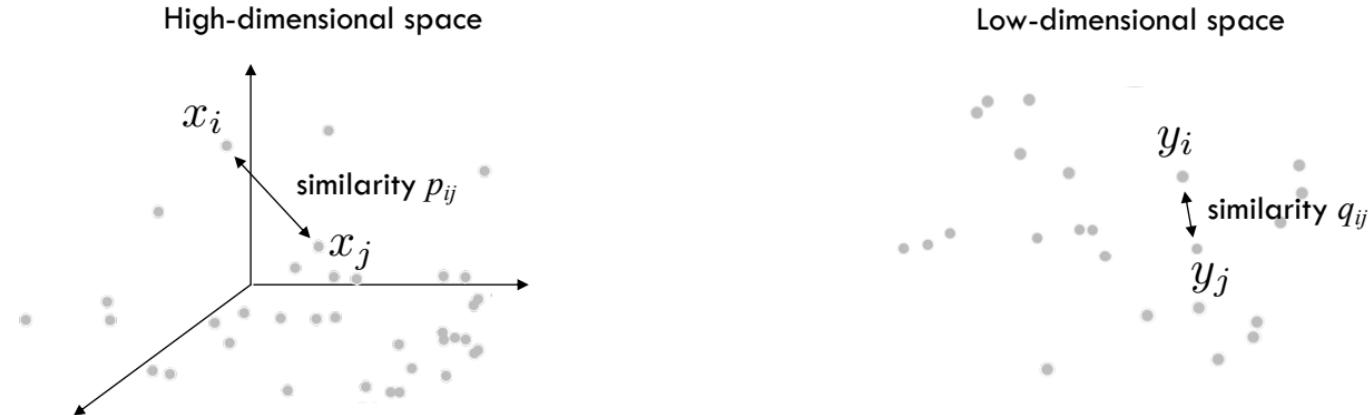
$$\tilde{\mathbf{X}}^T = \mathbf{W} \mathbf{X}^T$$

Dimensionality reduction via PCA

Let's look at an example!

Non-linear dimensionality reduction?

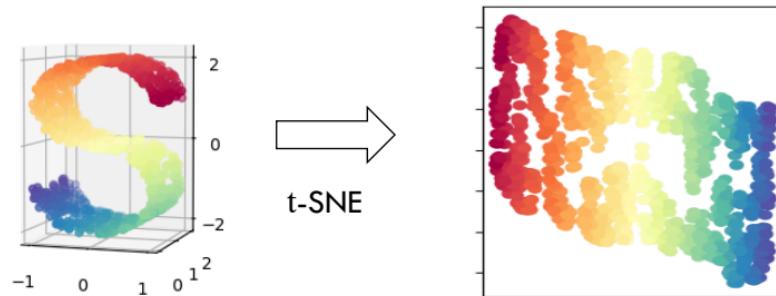
- *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE)
 - Popular *non-linear* dimensionality reduction method
 - High-level idea:



- Use gradient descent to move y_i 's to minimize $D(P\|Q) = \sum_{i\neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$

Non-linear dimensionality reduction via t-SNE

- Pros:
 - Can handle data that with non-linear underlying structures



Non-linear dimensionality reduction via t-SNE

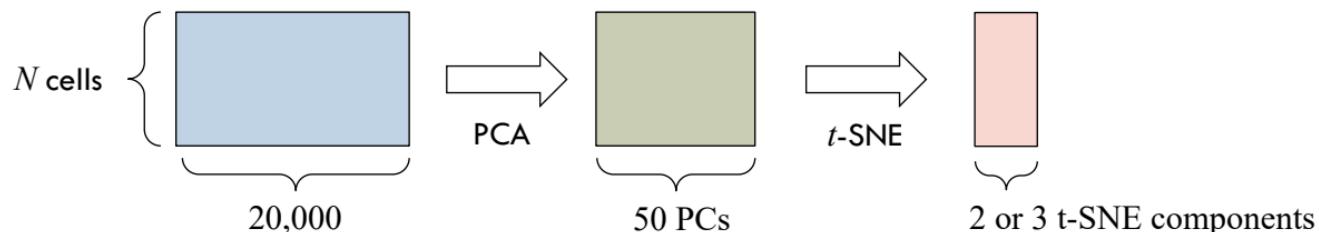
- Pros:

- Can handle data that with non-linear underlying structures
- Best approach to visualize single-cell RNA-seq data

- Cons:

- Mapping is not interpretable
- Random initialization (not always the same result)
- Computationally expensive

- Recommended approach:



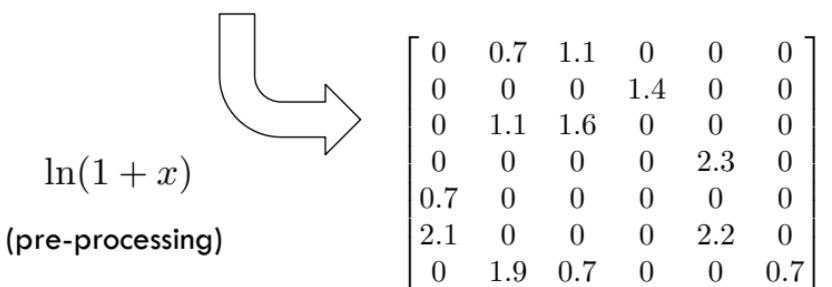
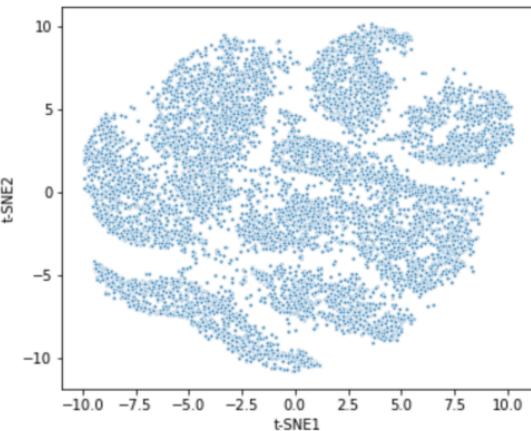
Non-linear dimensionality reduction via t-SNE

- Pros:
 - Can handle data that with non-linear underlying structures
 - Best approach to visualize single-cell RNA-seq data
- Cons:
 - Mapping is not interpretable
 - Random initialization (not always the same result)
 - Computationally expensive
- Let's look at the MNIST example again

Single-Cell RNA-seq Data Analysis

| | gene 1 | ... | gene <i>m</i> | | | |
|--------|--------|-----|---------------|---|---|---|
| cell 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| cell 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| ⋮ | 0 | 2 | 4 | 0 | 0 | 0 |
| ⋮ | 0 | 0 | 0 | 0 | 9 | 0 |
| ⋮ | 1 | 0 | 0 | 0 | 0 | 0 |
| ⋮ | 7 | 0 | 0 | 0 | 8 | 0 |
| ⋮ | 0 | 6 | 1 | 0 | 0 | 1 |

Dimensionality reduction
and visualization



Clustering

