



From latent spaces to living systems: introductory notes on representing single cell biology **across scales**

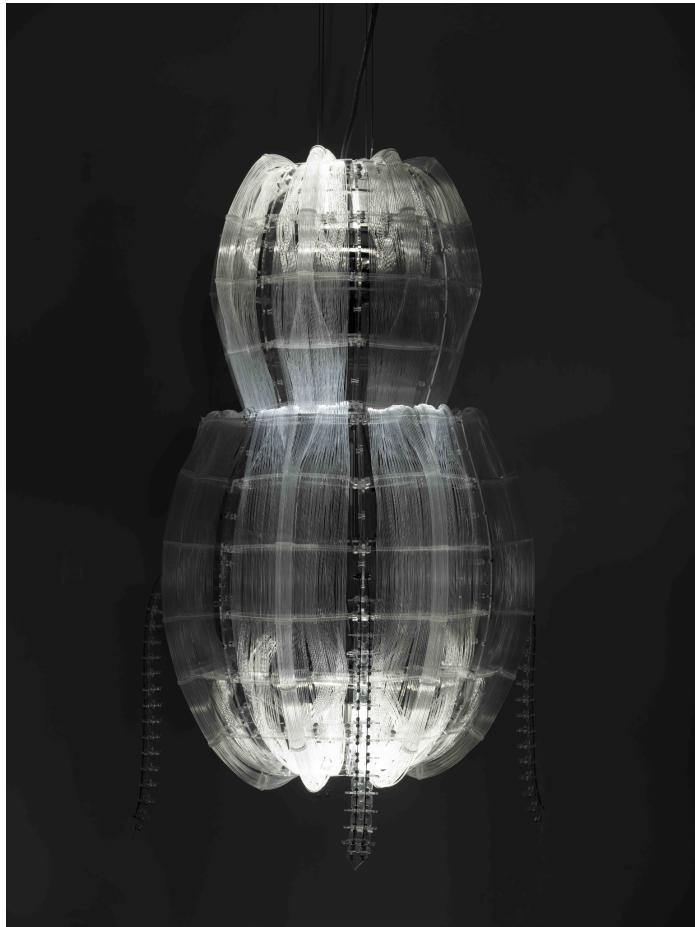
Bianca Dumitrascu
Columbia University
MLSS, Arequipa, Peru, 2025

From latent spaces to living systems

“Scientific truth should be presented in different forms, and should be regarded as equally scientific, whether it appears in the robust form and the vivid coloring of a physical illustration, or in the tenuity and paleness of a symbolic expression.”

James Clerk Maxwell

Disclaimer for the budding ML researcher: A latent space isn't a world – it's a useful scaffold



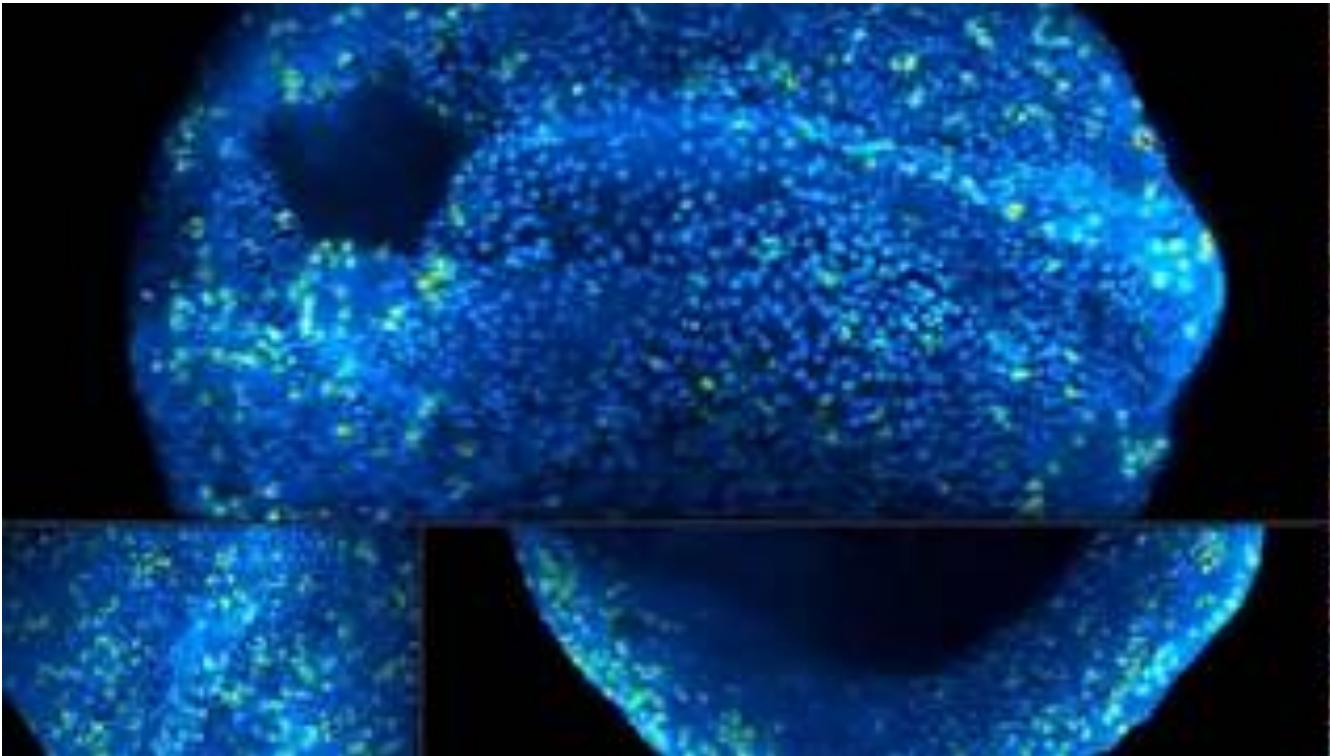
Works by Anicka Yi. **Left:** *Nested Lung* (2023–24). Courtesy Leeum Museum of Art. **Bottom right:** *Each Branch Of Coral Holds Up The Light Of The Moon* (2024) (detail). Courtesy Leeum Museum of Art. **Top right:** *Biologizing the Machine (Tentacular Trouble)*, Venice Biennale 2019

Morphogenesis as entry point into Computational Biology



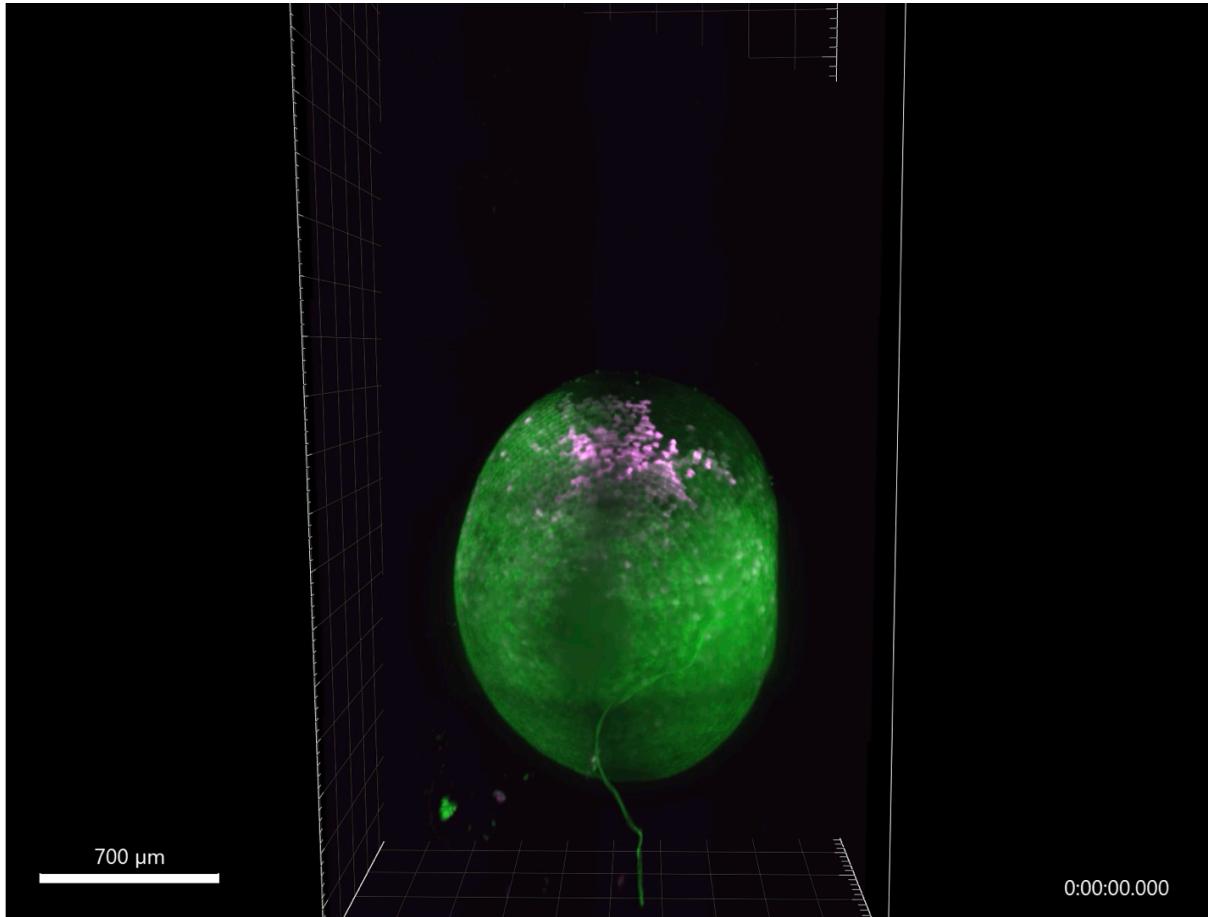
Becoming, Jan van IJken, time-lapse of a salamander growing from one cell

Morphogenesis as entry point into Computational Biology



Kate McDole

Morphogenesis as entry point into Computational Biology



Credit: Jakub Sedzinski

Morphogenesis as entry point into Computational Biology

*What tools do we need to understand a **spatio-temporal** process like **morphogenesis**?*

Morphogenesis as entry point into Computational Biology

*What tools do we need to understand a **spatio-temporal** process like **morphogenesis**?*

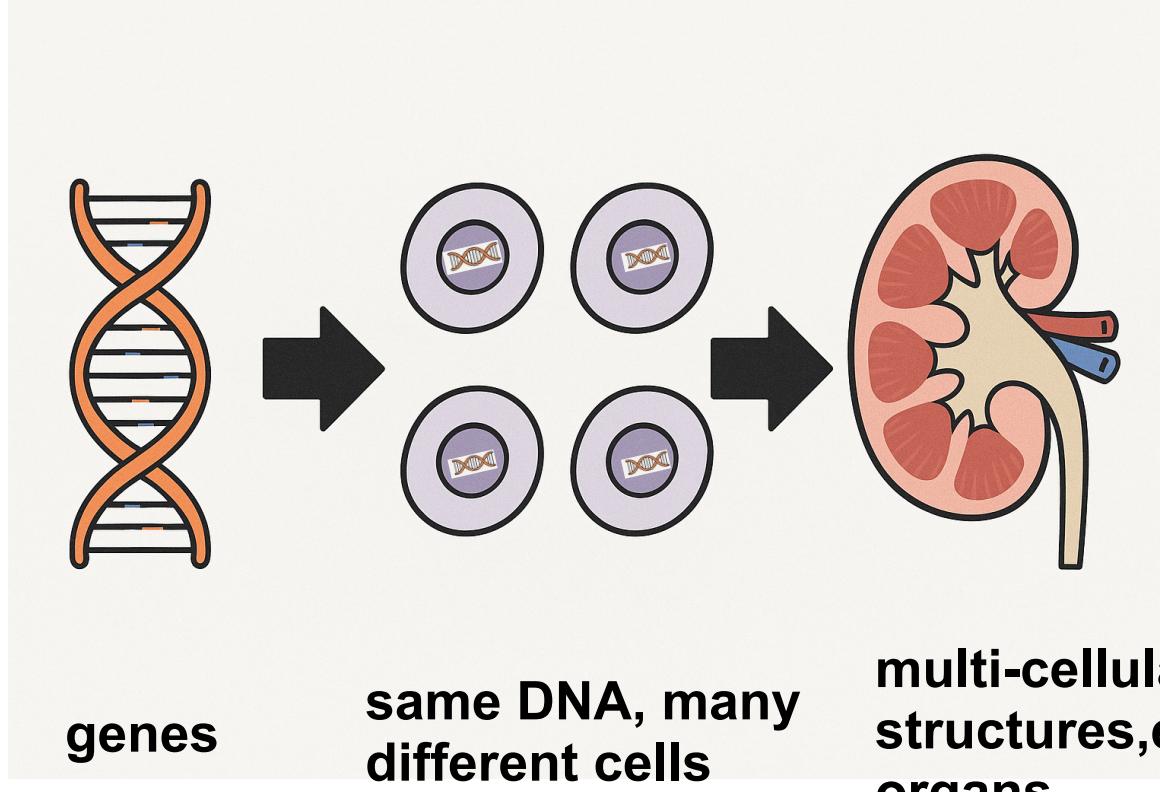
- **Define and identify** the main players
→ **Lecture 1 + 2:** Cells and Latent Spaces

Morphogenesis as entry point into Computational Biology

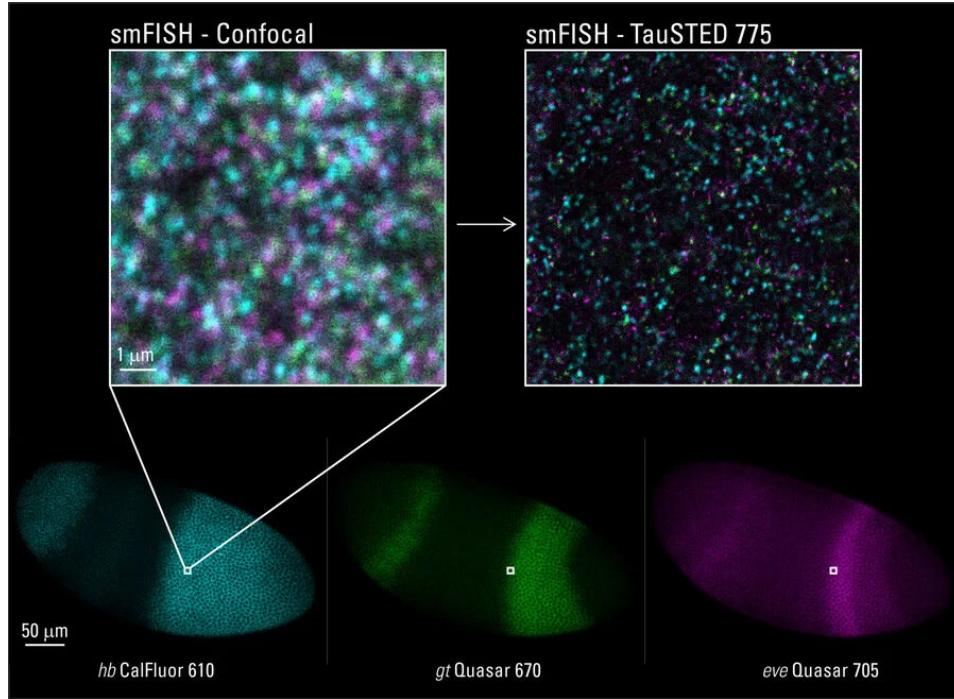
*What tools do we need to understand a **spatio-temporal** process like **morphogenesis**?*

- Define and identify the main players
→ Lecture 1+2: Cells and Latent Spaces
- Map their spatial organization and context
→ Lecture 3+4: Tissue geography
 - Model physical constraints

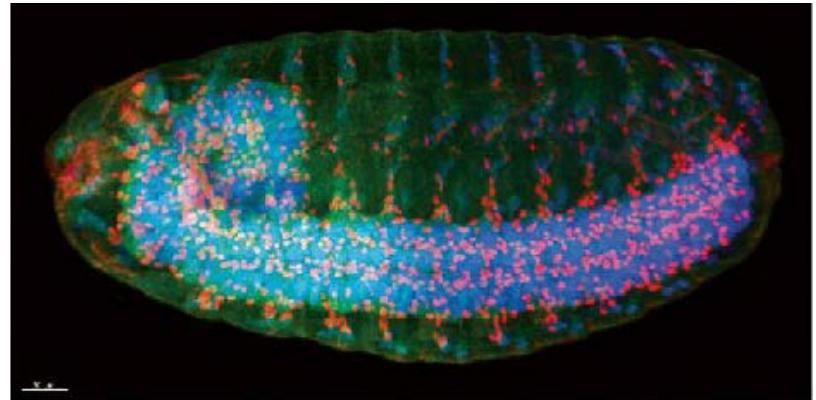
In search of the right unit of representation and the right scale



In search of the right unit of representation and the right scale



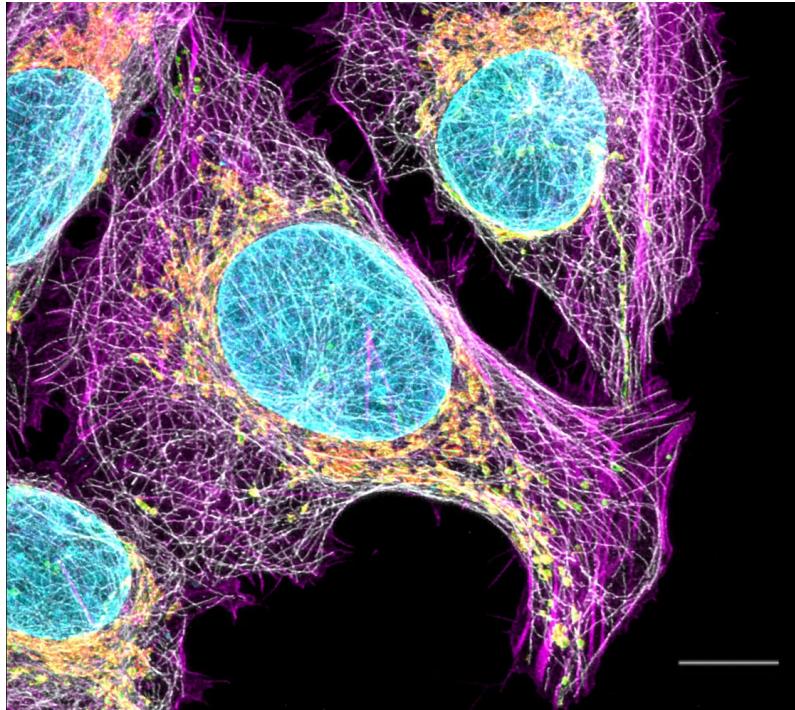
genes?



Single molecule in-situ hybridization (smFISH) of RNA in Drosophila embryo whole-mount preparation. Genes: **hb** (cyan), **gt** (green), **eve** (magenta). Image from <https://www.leica-microsystems.com>

Drosophila (fruit fly) embryo showing Schizo, a protein involved in neural development. Photo: The American Journal of Human Genetics/H. Chung/the Bellen lab.

In search of the right unit of representation and the right scale



genes? cells?

Maximum projection of a z-stack from fixed U2OS cells (human osteosarcoma cell line) labelled with 5 colors, including Tubulin (grey), Actin (magenta), Membranes (cyan). scale bar 10 μ m. Image <https://www.leica-microsystems.com/>

Lecture 1+2: Identity and Representation

Functional unit in these lectures: Cells

- Modalities: **gene expression**, chromatin accessibility
- Concepts: latent variables
 - (foundations) shallow methods (**L1**)
 - + deep, genAI models (**L2**)
- **L1** Techniques: dimensionality reduction, inductive bias, self-supervised learning; PCA, pPCA, FA etc.

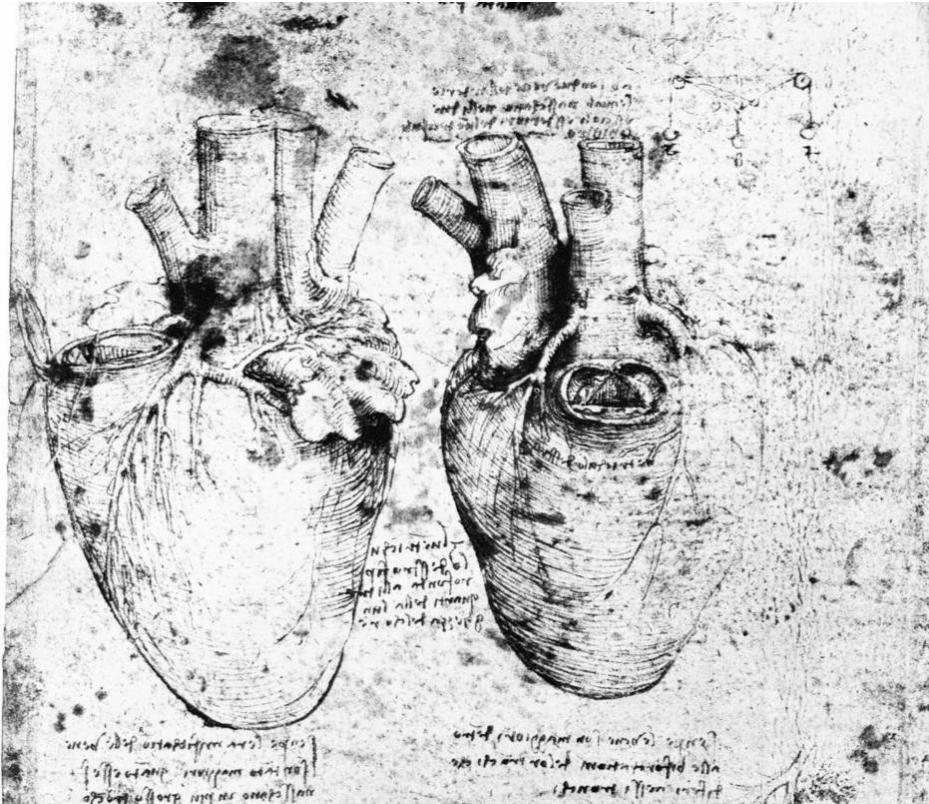
Representing Data with Purpose: seeing is believing a first and very literal student - teacher interaction



Rembrandt's The Anatomy Lesson of Dr. Nicolaes Tulp, 1632.

Fine Art Images/Heritage Images/Getty Images

Representing Data: Zooming In



Anatomical drawings of the heart and its blood vessels

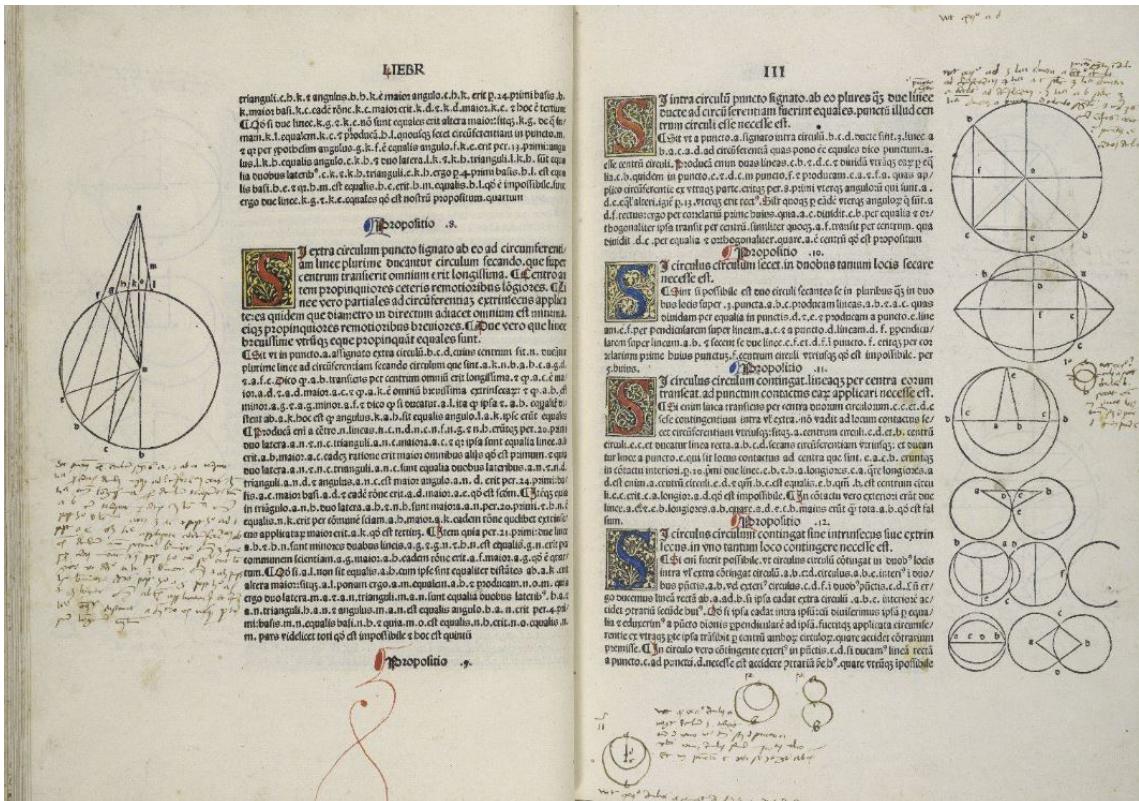
Leonardo da Vinci, c 1513

Representing Form: Zooming Out



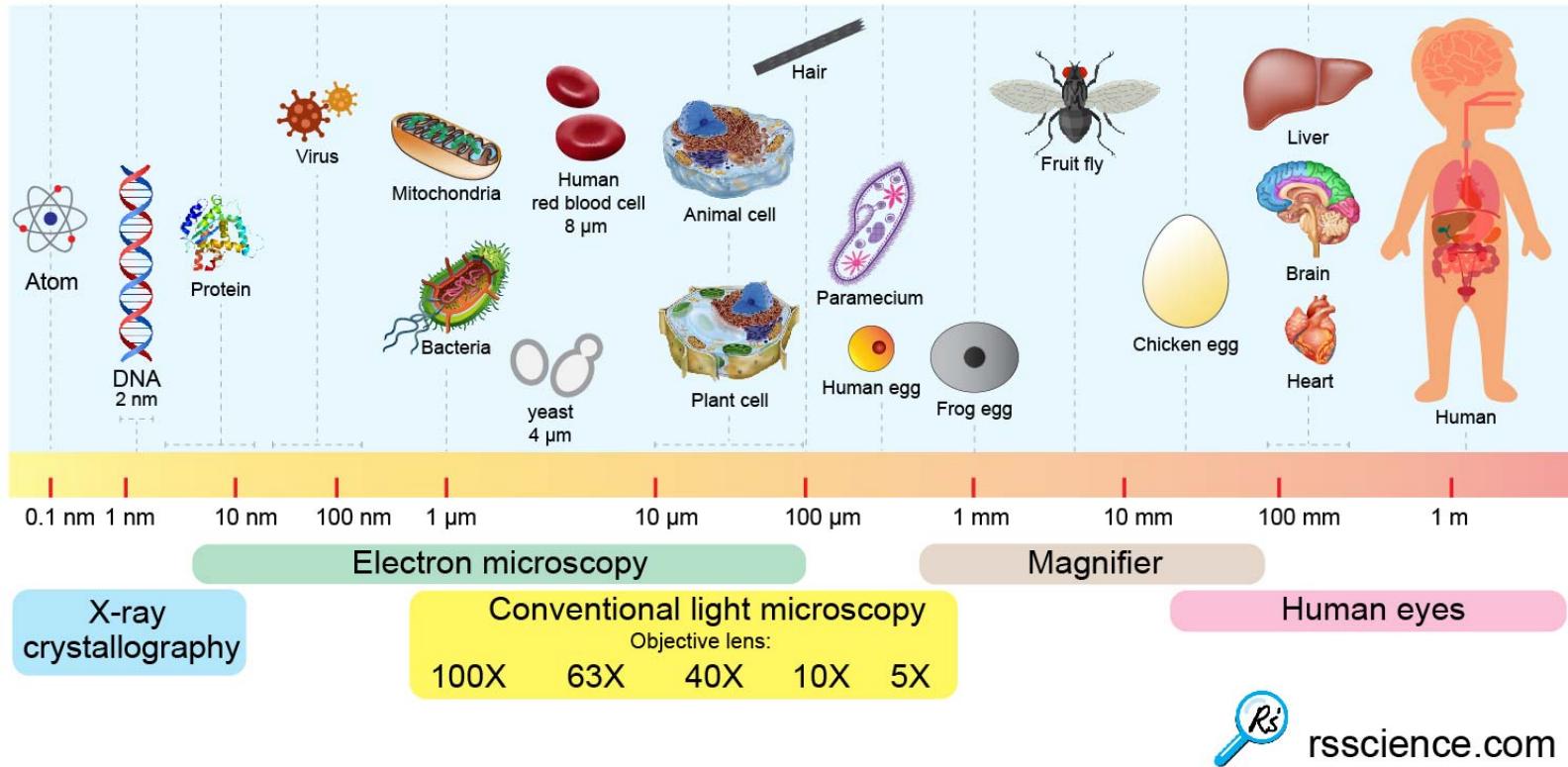
Young Hare, Albrecht Durer, 1502

Representing Geometry



Euclid's Elements, c 330 BC, Printed in Venice, 1482.

Observing cells: zooming way in



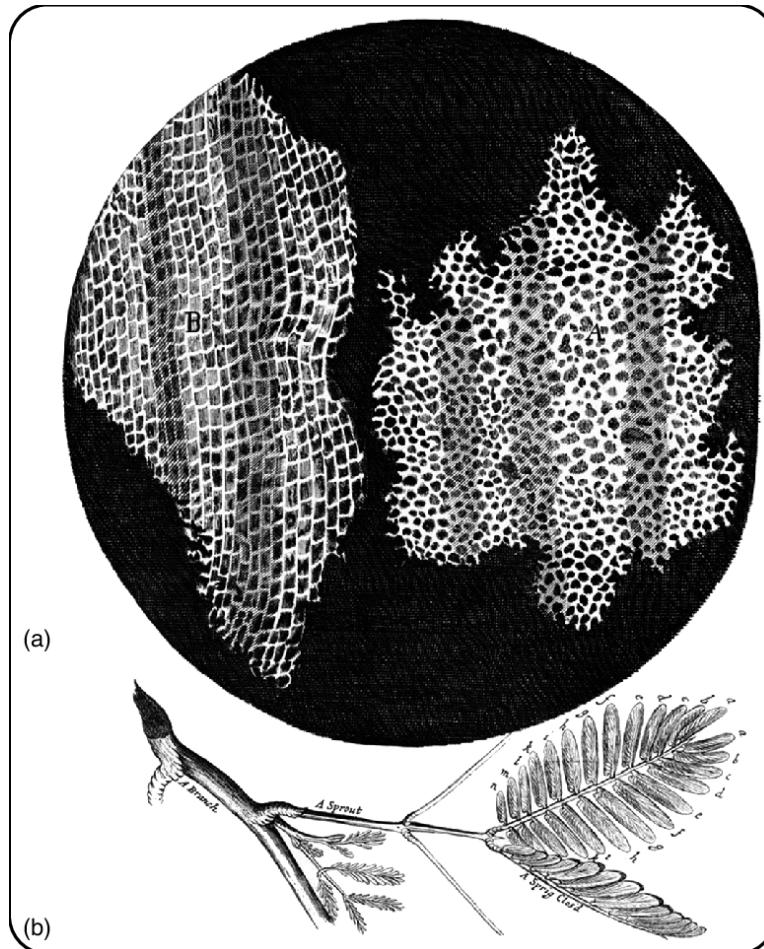
*A cell is to a meter, as a coin
is to the Eiffel tower

Observing the Cell: a Short History

Ingredients:

- Experimental: **Measure new things**
 - Molecular Recordings (Sequencing)
 - Imaging and visualization (Microscopy)
- Conceptual: **Imagine and define** what new things to measure
 - The idea of a gene (Mendel)
- Community: curate, illustrate, communicate, pass on

Observing the Cell: a Short History



drawing of cork cells, Robert Hooke, 1665,
using the first optical microscope

1. 1590 – Invention of the Early Microscope

Zacharias Janssen and his son Hans create a tube with multiple lenses, a forerunner to both the **compound microscope** and **telescope**, sparking the field of microscopy.

2. 1665 – First Use of the Term ‘Cells’

Robert Hooke publishes *Micrographia* and introduces the term “cells,” marking the beginning of **cell biology** as a scientific discipline.

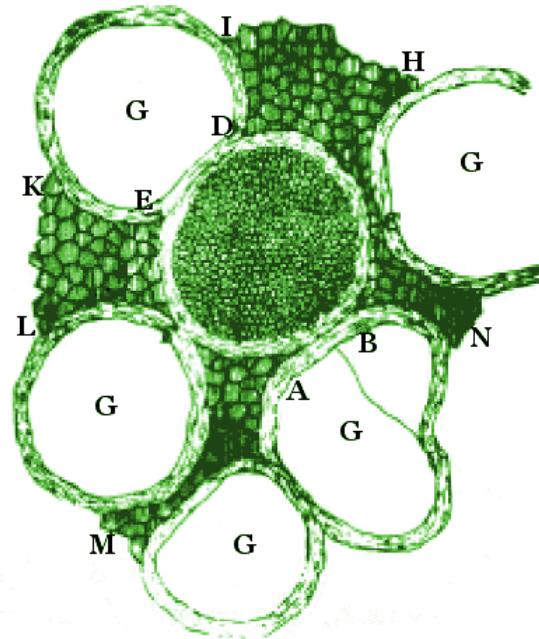
3. 1676 – First Observation of Living Cells

Antonie van Leeuwenhoek observes bacteria and protozoa using his single-lens microscope, expanding the biological world visible to science.

4. 2014 – Nobel Prize for Super-Resolution Microscopy

Awarded to Eric Betzig, Stefan Hell, and William Moerner, this prize recognizes breakthroughs in **super-resolved fluorescence microscopy**, which overcame the optical resolution limit and transformed cellular imaging.

Observing the Cell: from Structure to Context



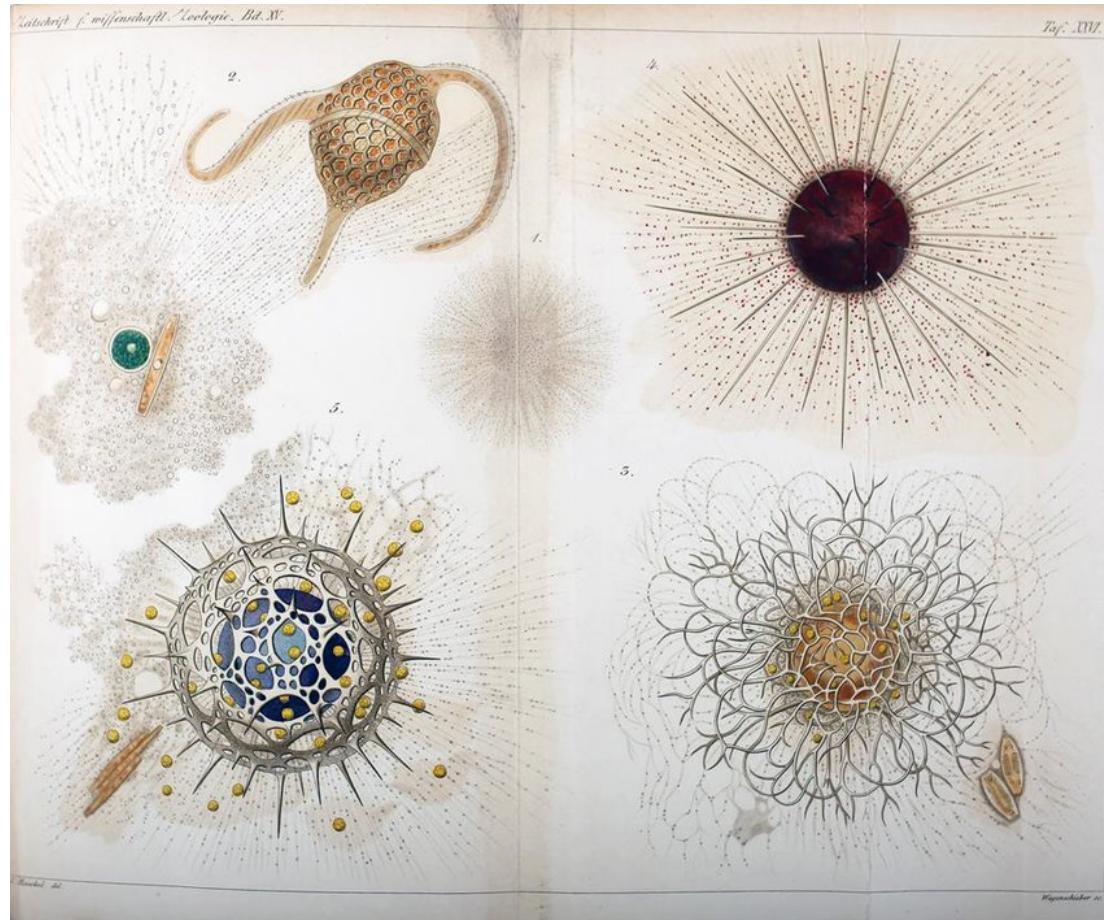
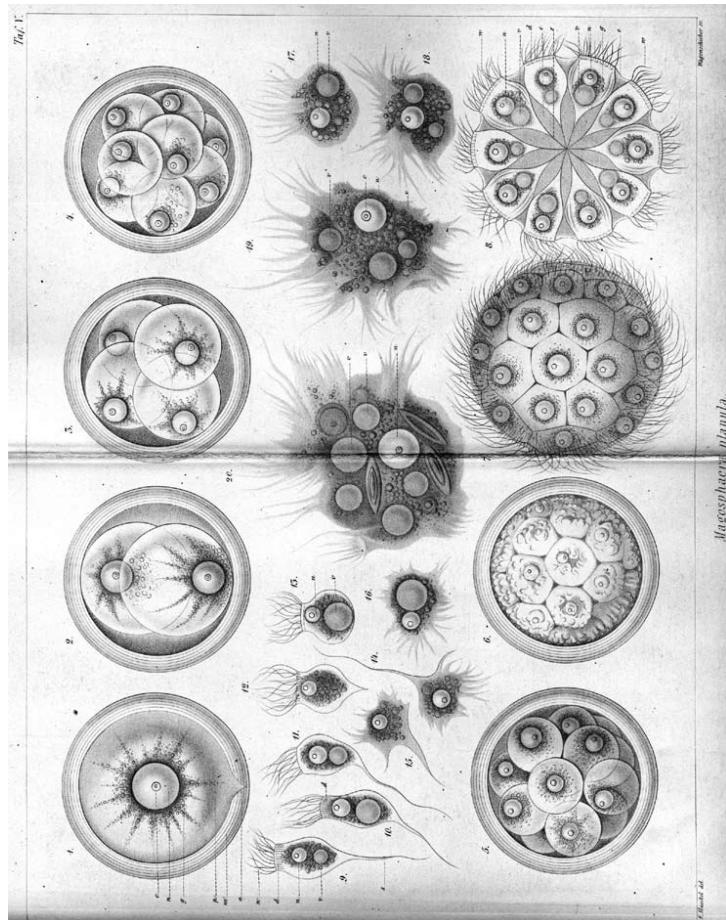
1676 – First Observation of Living Cells

Antonie van Leeuwenhoek observes bacteria and protozoa using his single-lens microscope, expanding the biological world visible to science.

“...I observed certain animalcules, within whole bodies I saw so quick a motion as to exceed belief; they were about the size of a large grain of sand, and their bodies being transparent, that the internal motion could plainly be seen. Among other things, I saw in the body of one of these animalcules a bright and round corpuscle, placed near the head, and in which a very wonderful swift motion was to be seen, consisting of an alternate extension and contraction.”

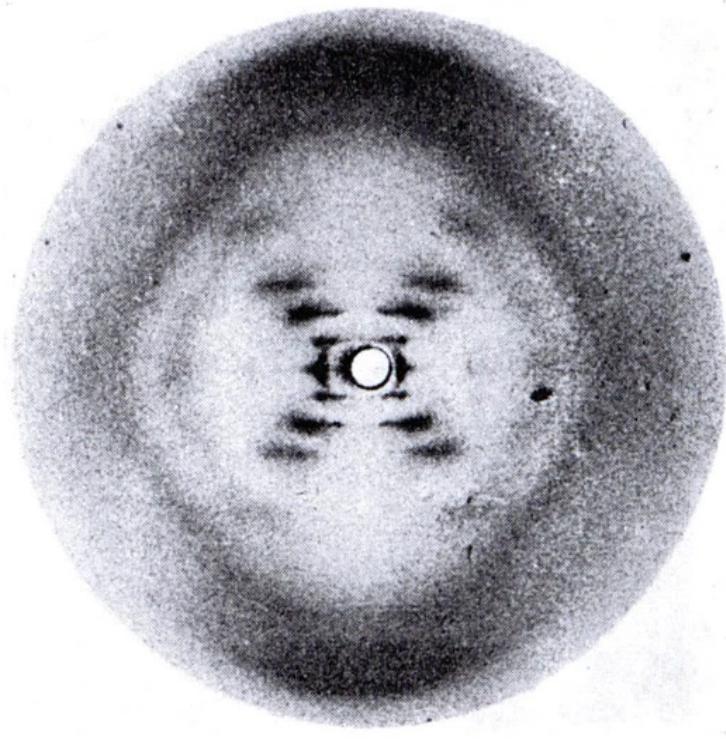
Antonie van Leeuwenhoek
Letter to Anthonio Magliabechi of Florence

Observing the Cell: Qualitative Beginnings

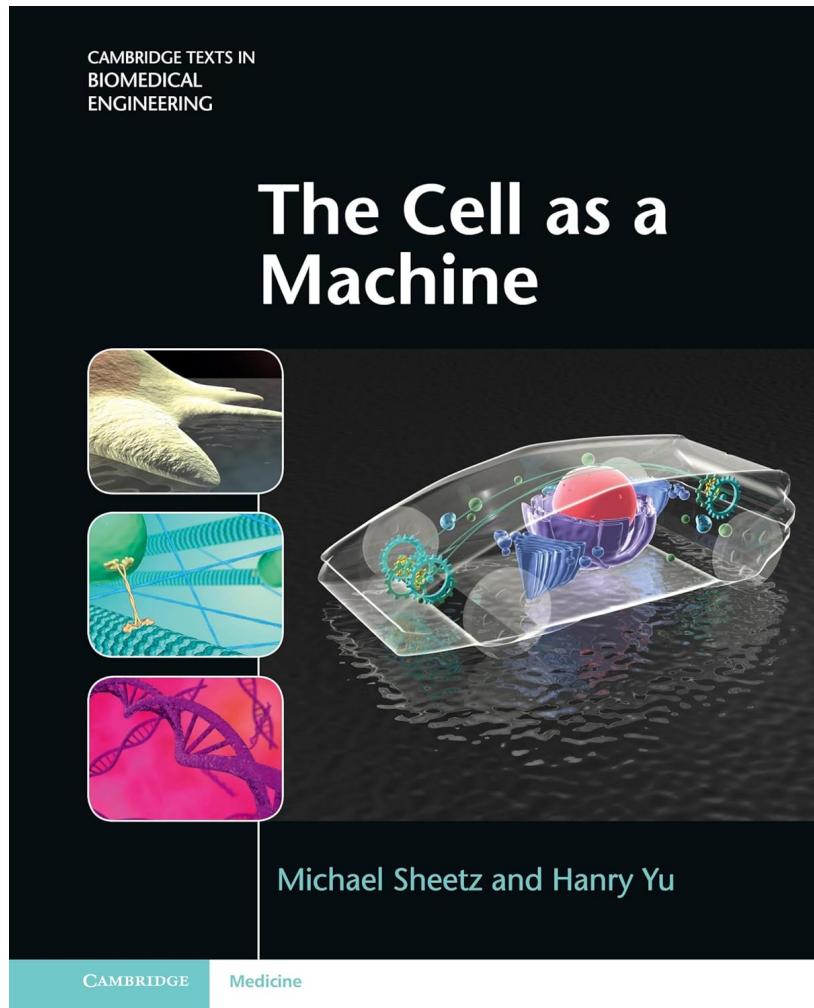


drawings by Ernst Haeckel, cca. 1870

Representing the Cell: quantitative follow-up



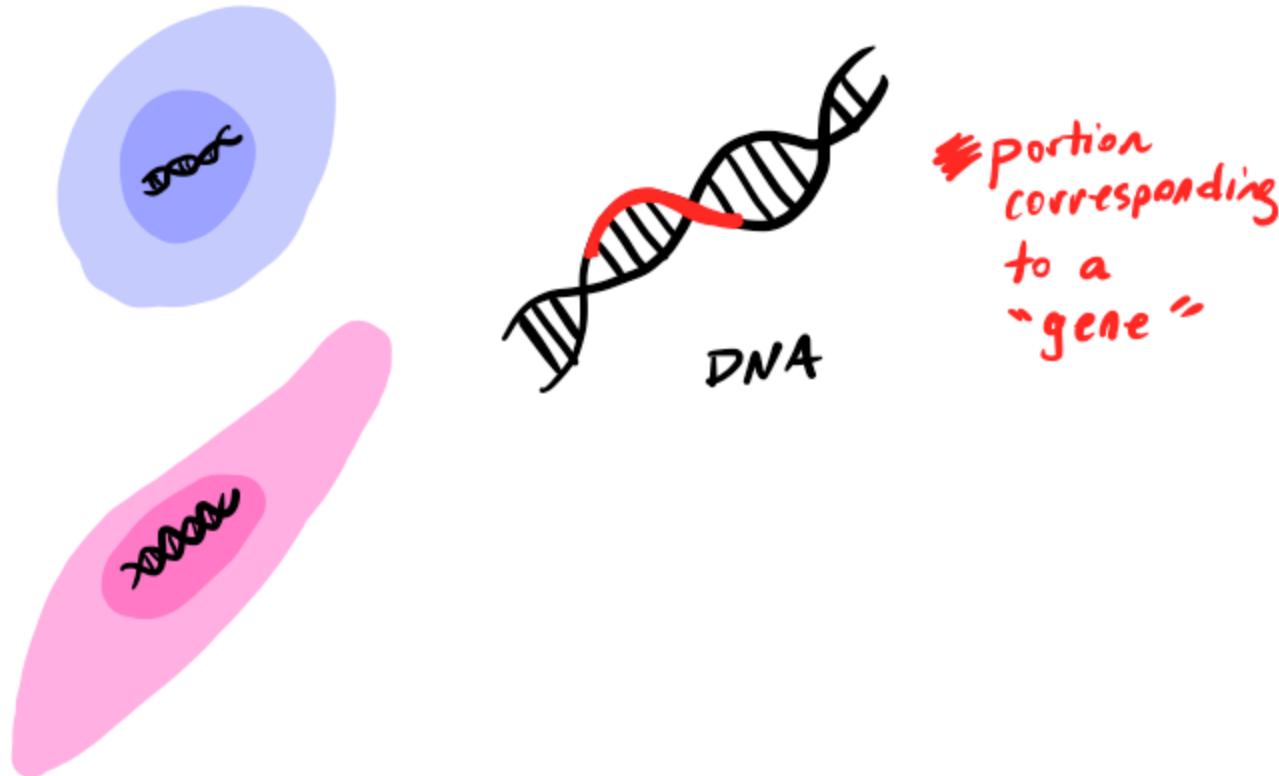
1953 - Franklin's X-ray diagram of the B form of sodium thymonucleate (DNA) fibres, published in Nature on 25 April 1953



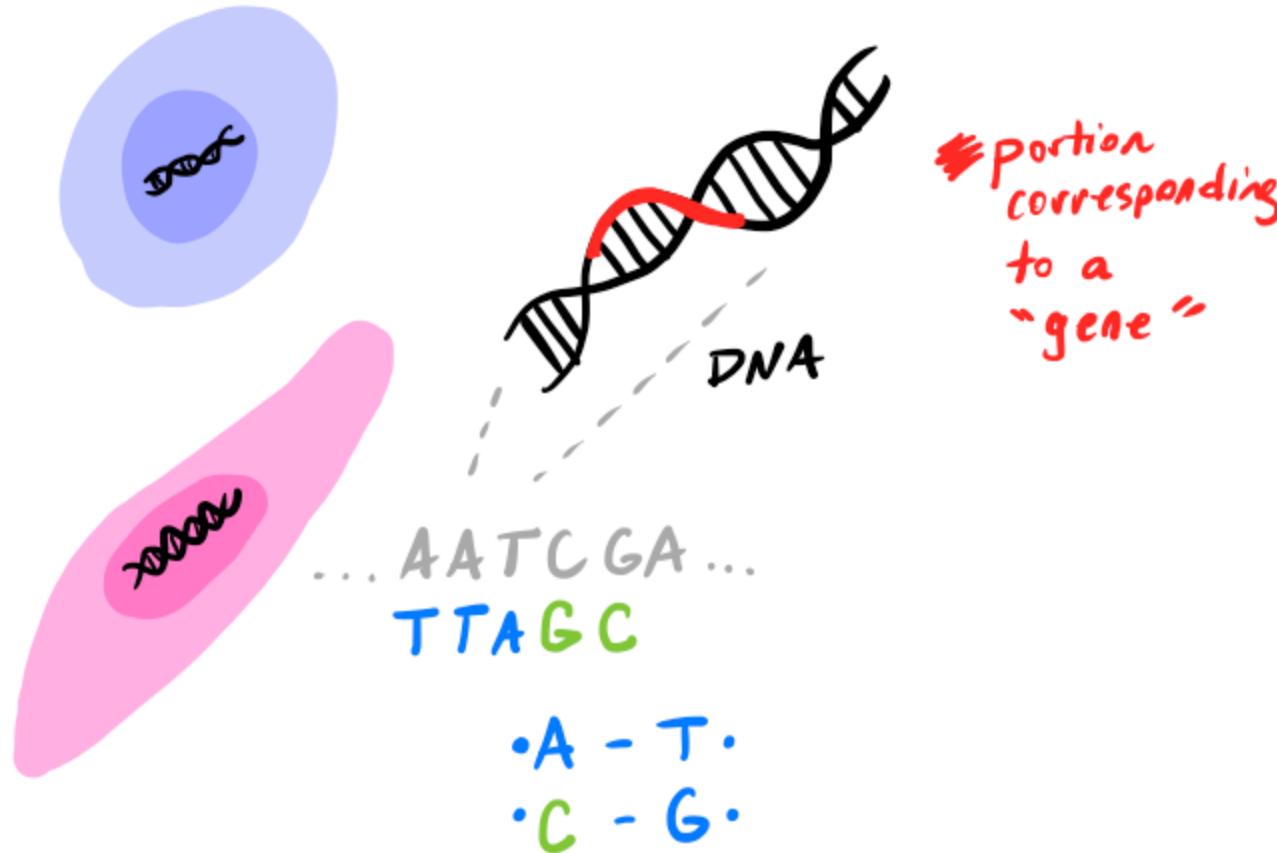
Representing the Cell: focus on genes



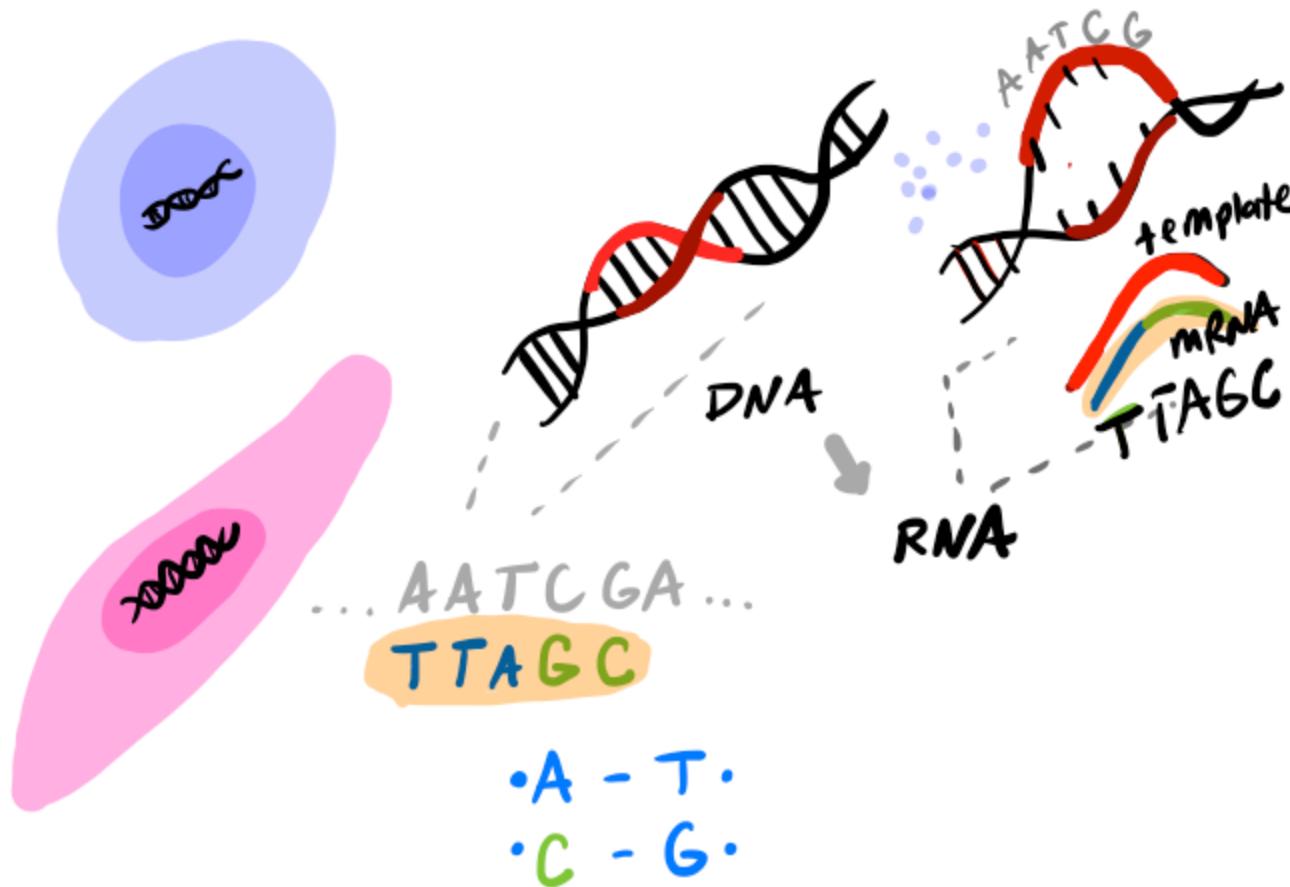
Representing the Cell: focus on genes



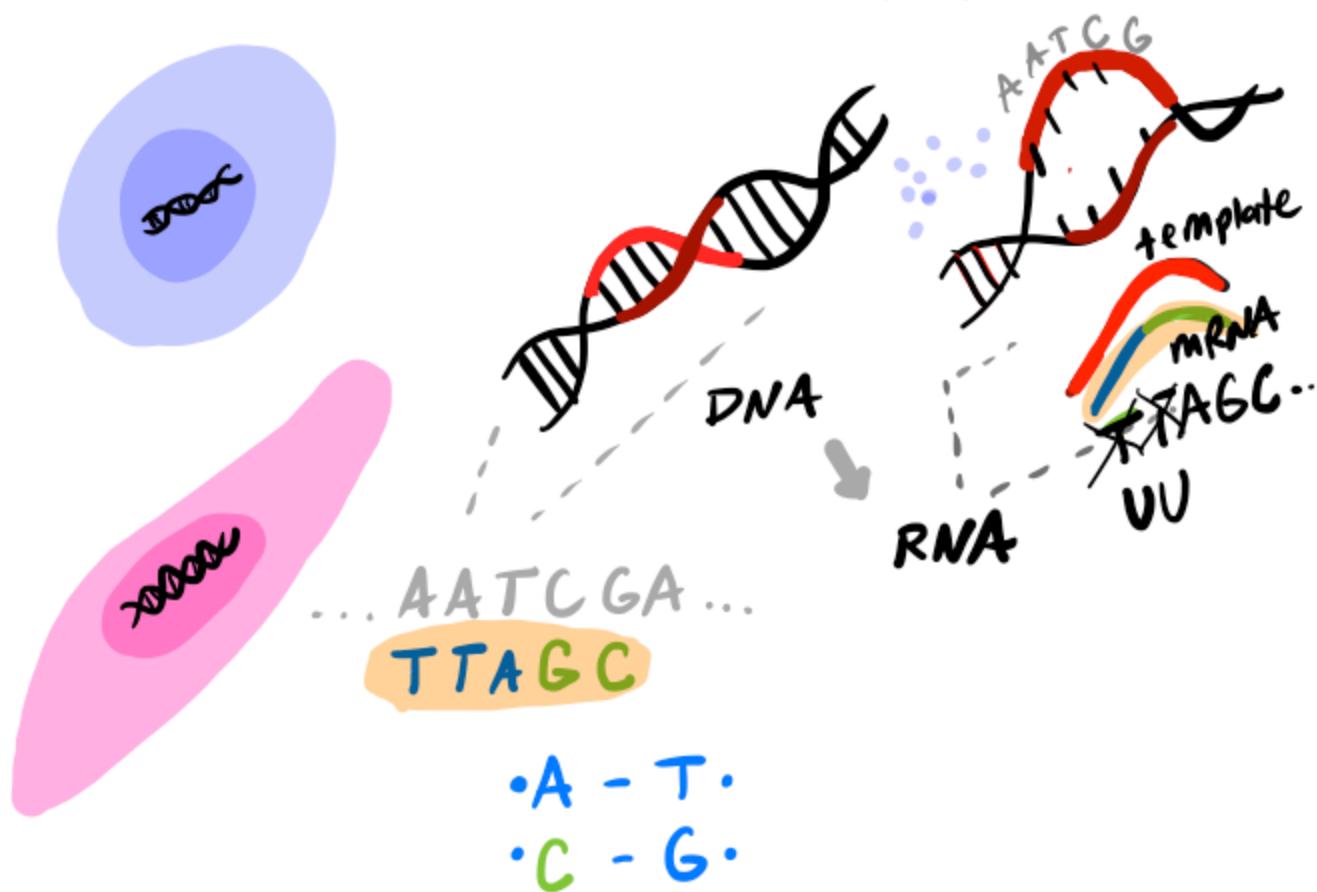
Representing the Cell: focus on genes



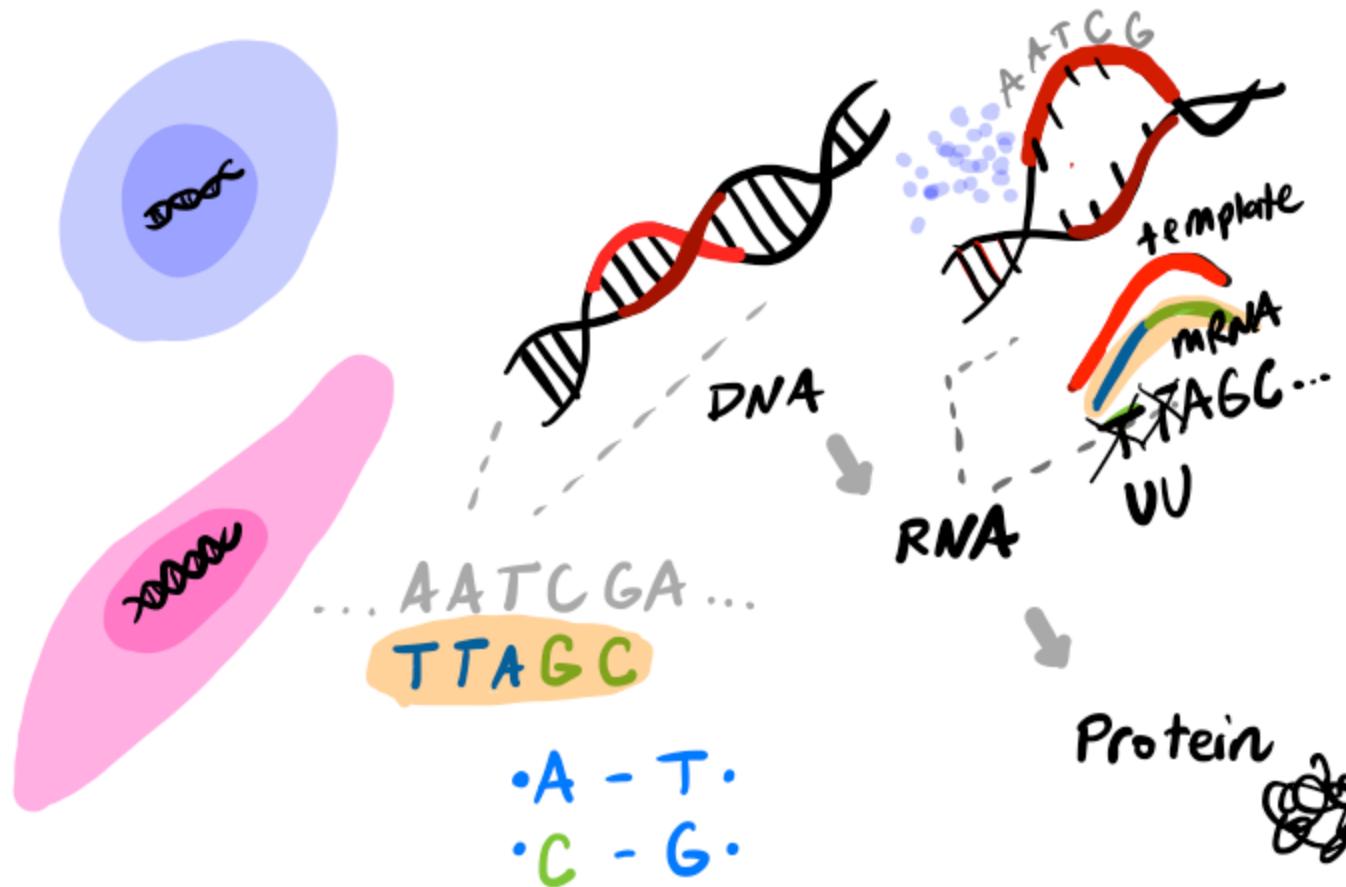
Representing the Cell: transcription



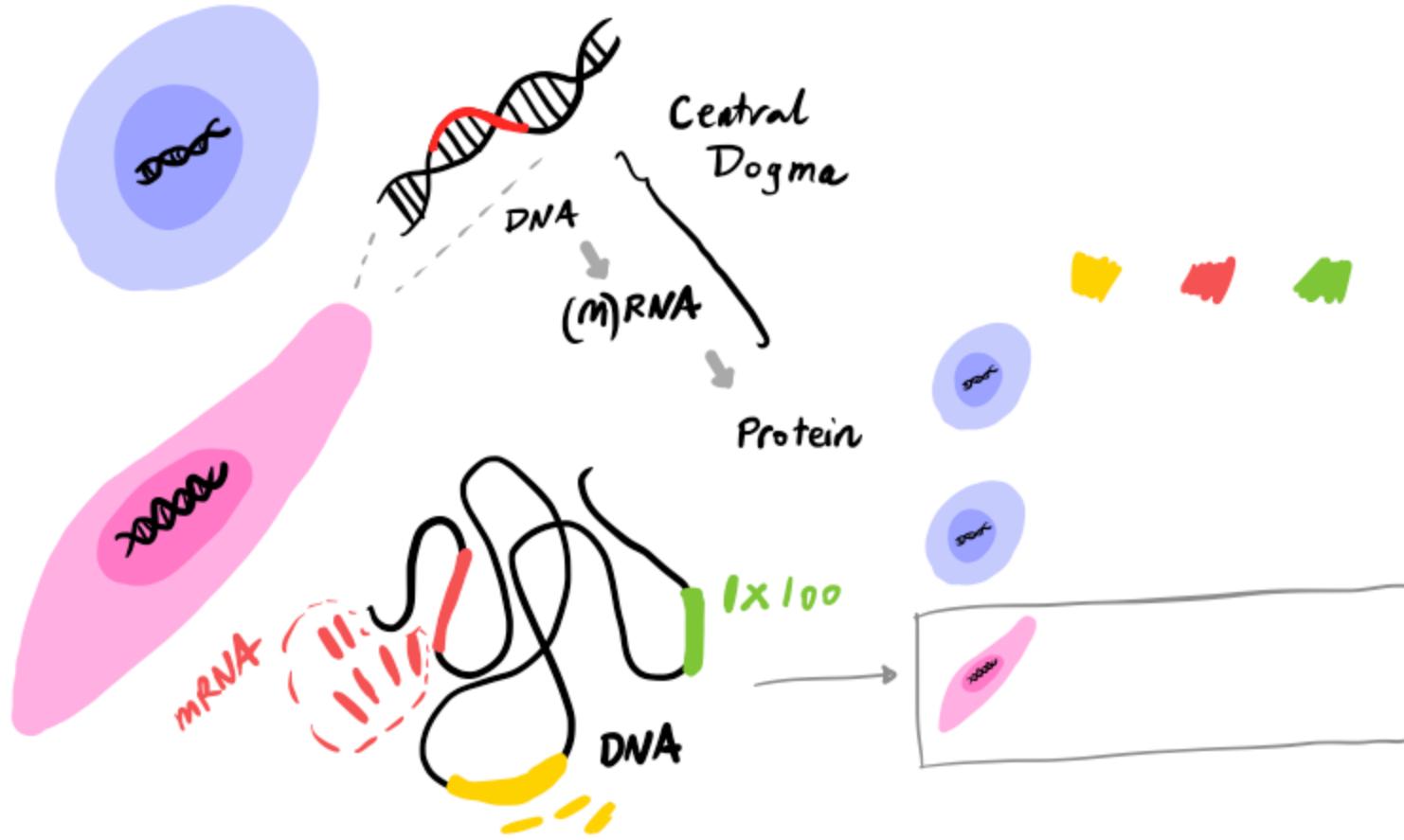
Representing the Cell: transcription



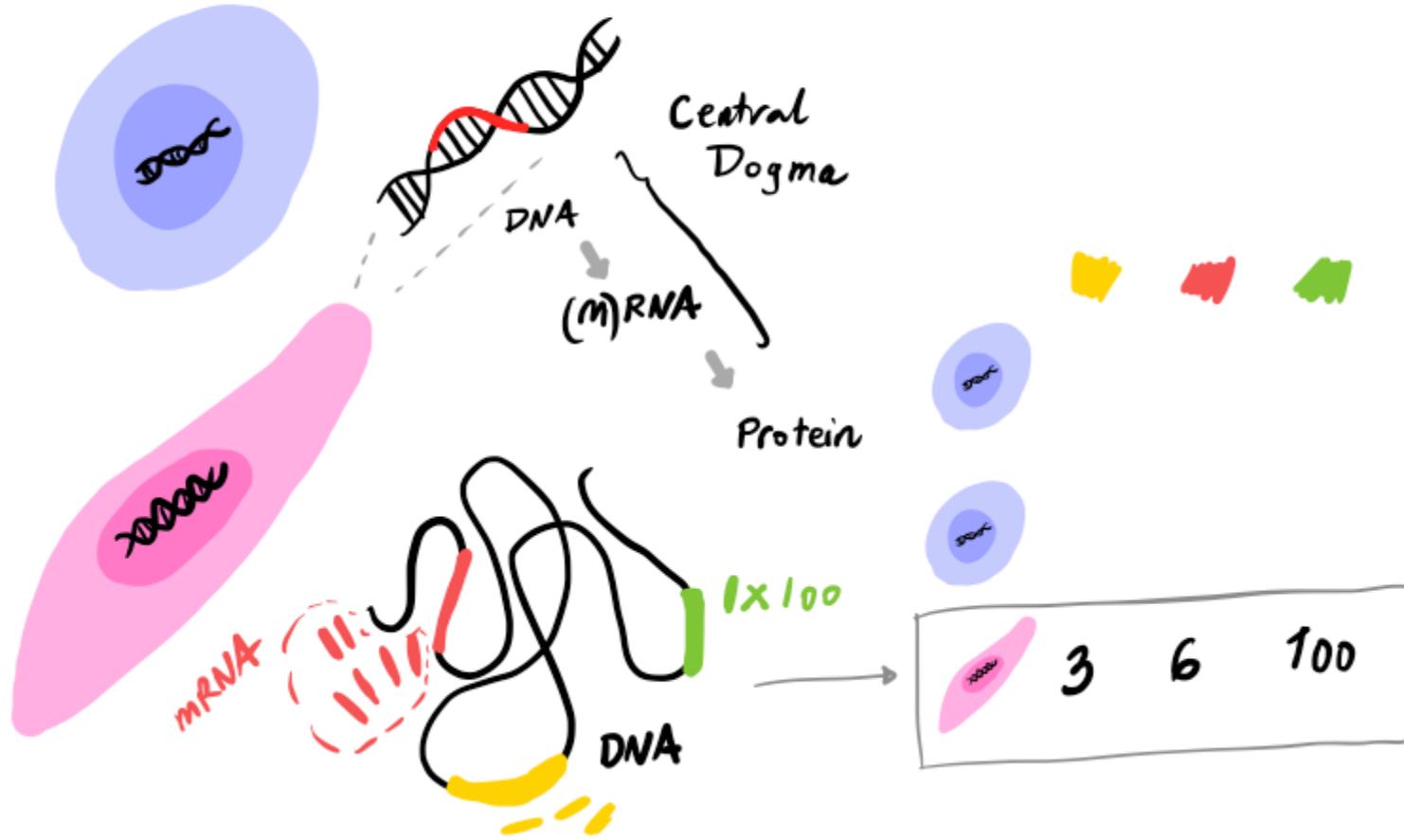
Representing the Cell: translation



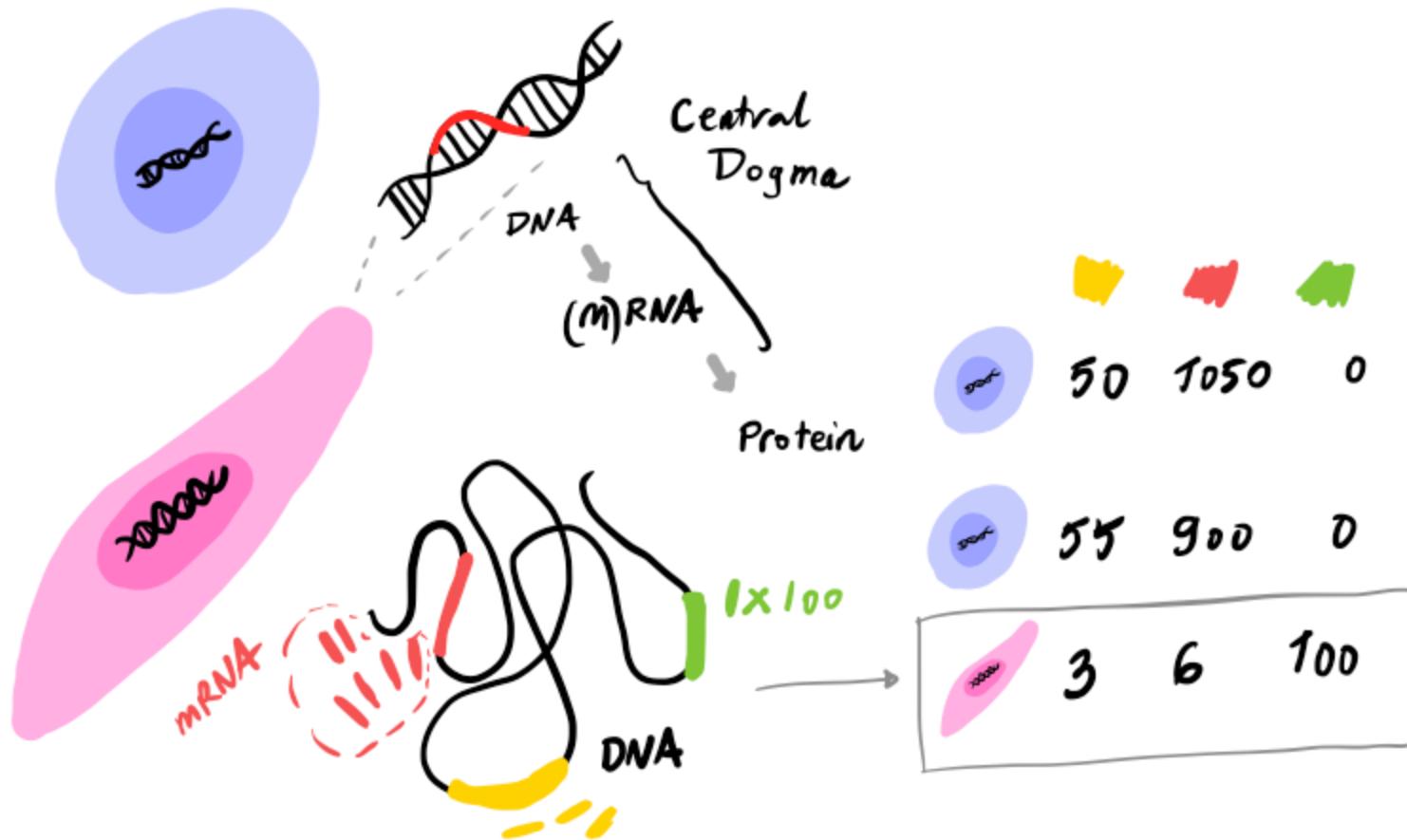
Representing the Cell: transcripts as activity proxies



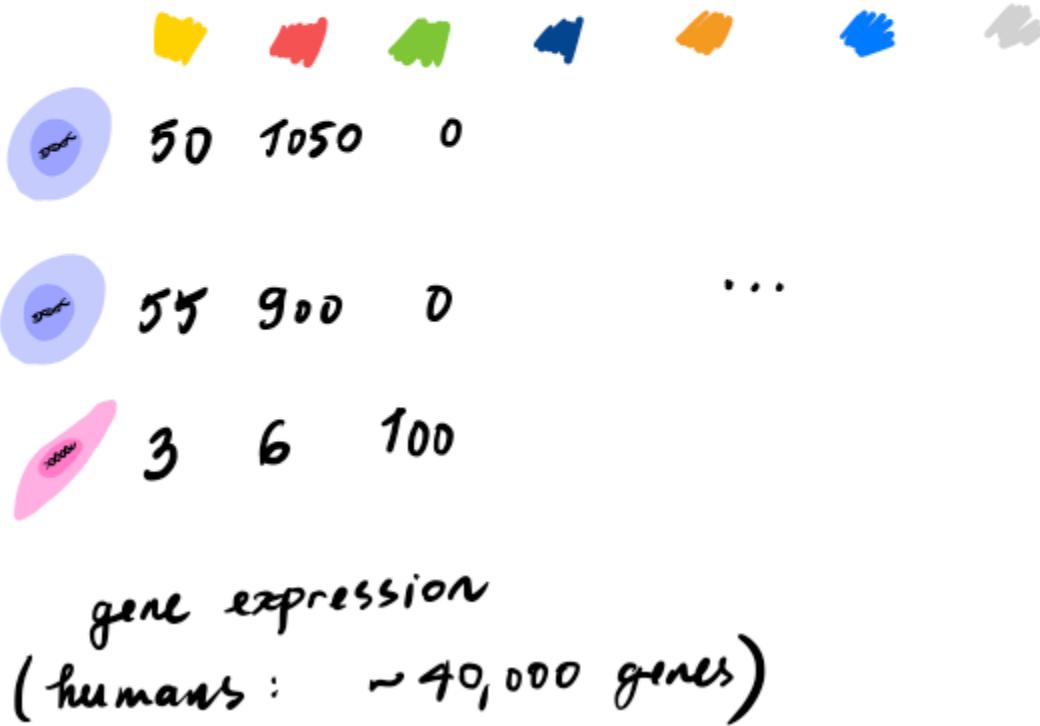
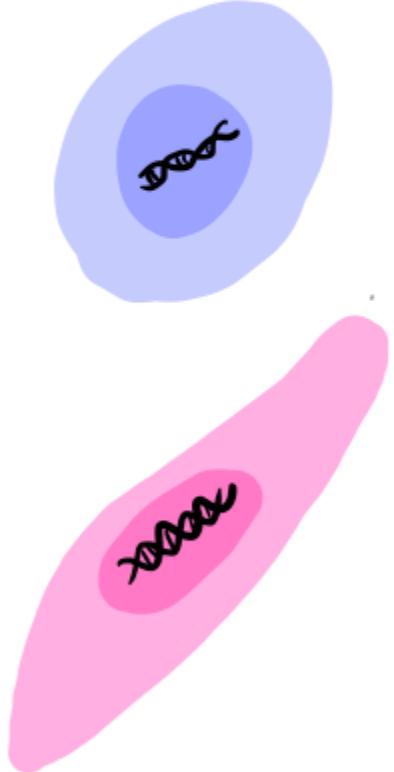
Representing the Cell: transcripts as activity proxies



Representing the Cell: transcripts that count



Representing the Cell: transcripts that count



Representing the Cell: quantitative follow-up

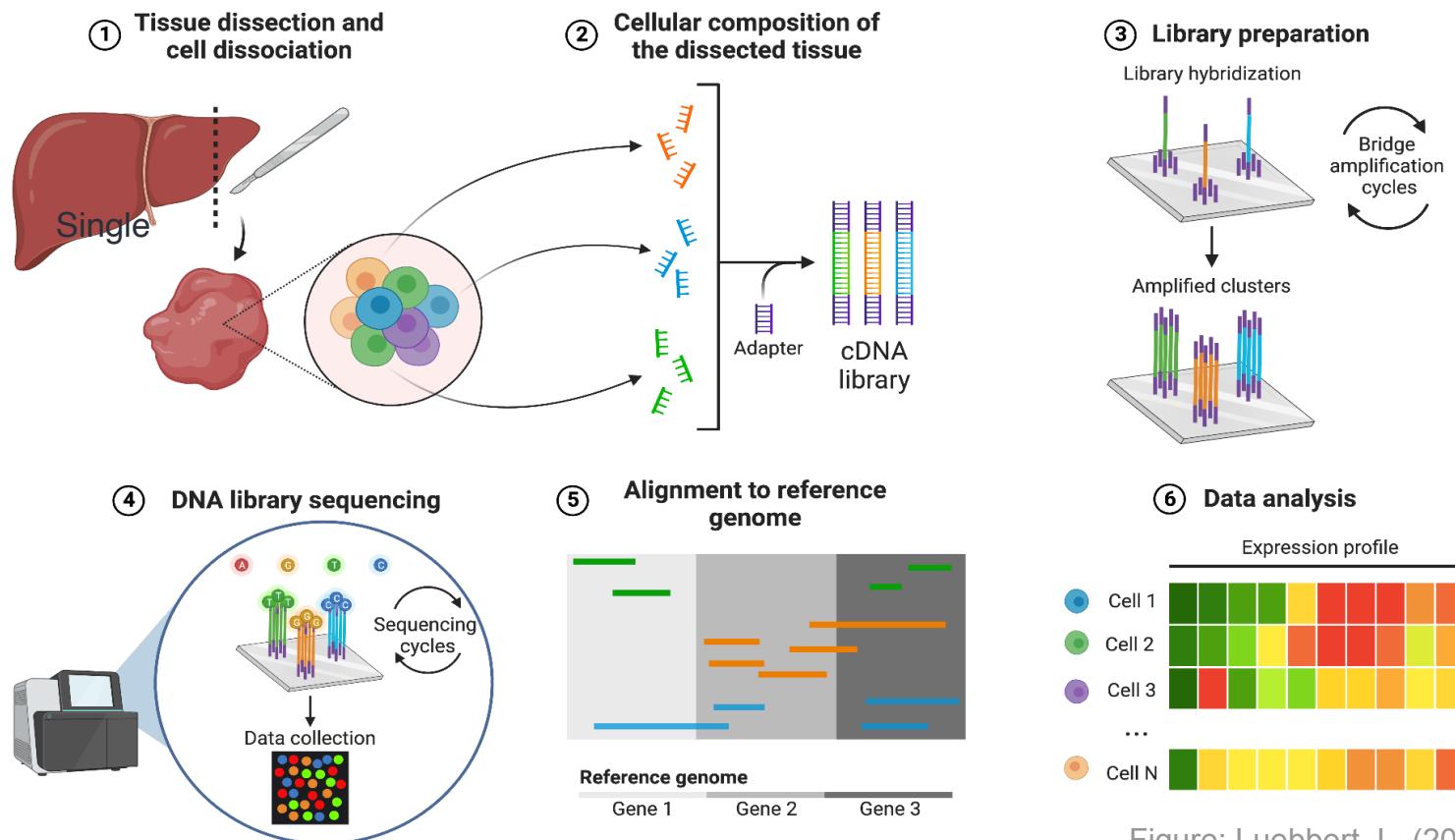


Figure: Luebbert, L. (2025)
BioRender.

A great course about sequencing & where data comes from: <https://www.singlecellcourse.org/introduction-to-single-cell-rna-seq.html>

Representing the Cell: sequencing on steroids

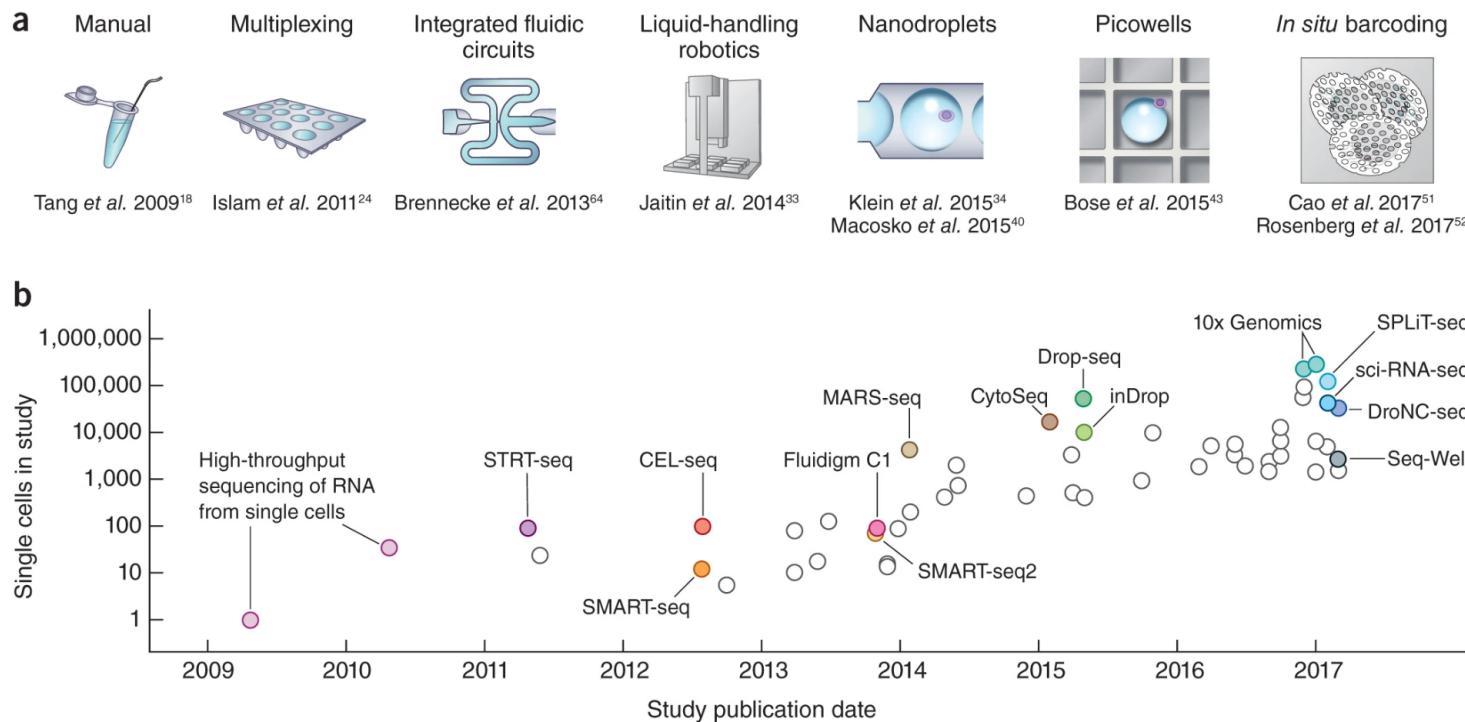
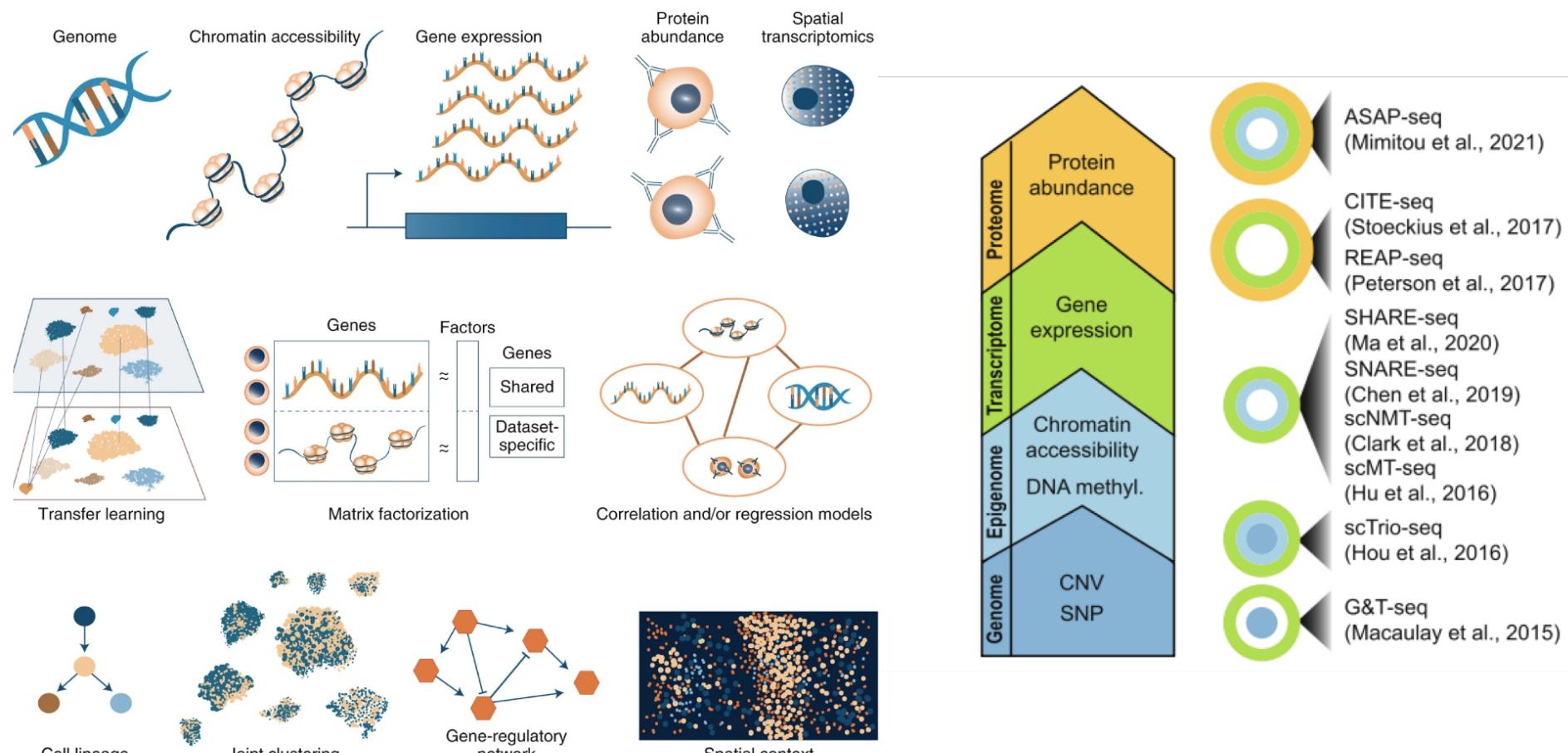


Figure adapted from : Svensson, V., Vento-Tormo, R. & Teichmann, S. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc 13, 599–604 (2018).

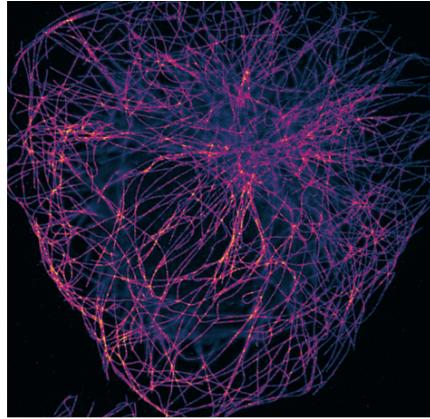
There is lots of data beyond gene expression



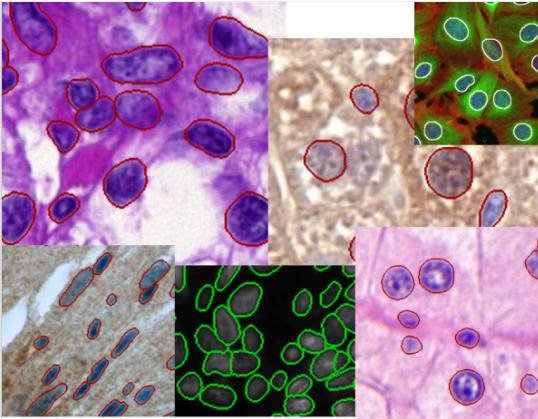
Yang, Huang, Liu, 2021

Efremova & Teichmann, 2020

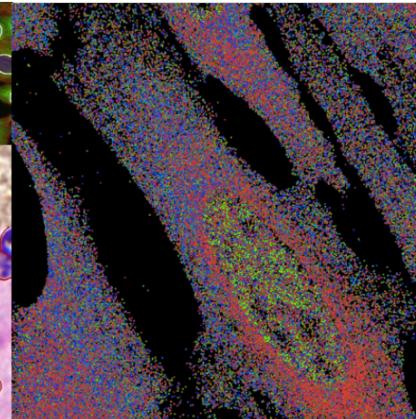
There is lots of data beyond gene expression



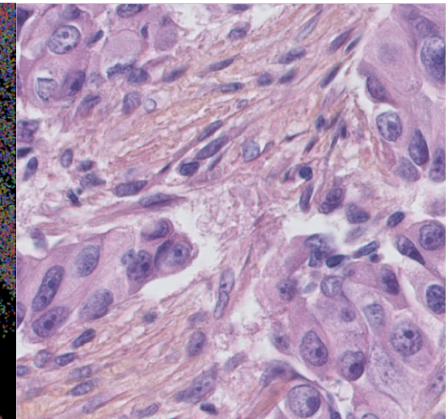
srm.epfl.ch - Tubulin-COS7-



Hollandi et al., Cell Systems, 2020



the Xiaowei Zhuang Laboratory

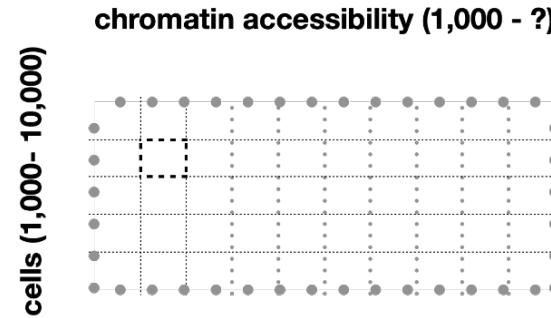
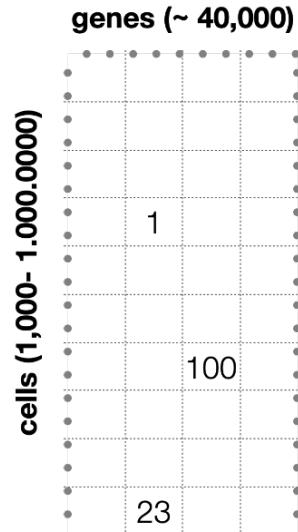


zenodo.org/record/1175282#.YUtlNGZKhJ

Challenges:

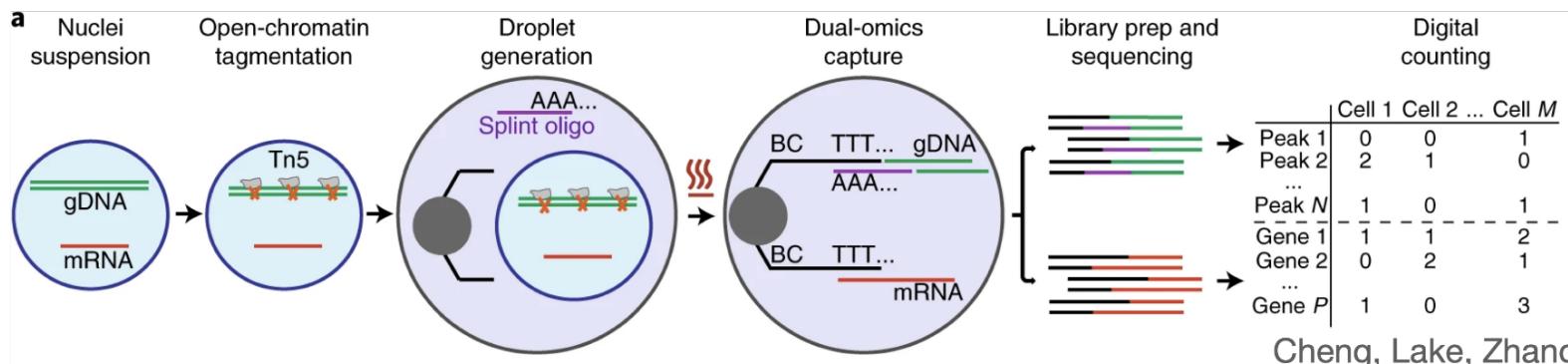
- segmentation
- single object vs multiple object modeling
- feature extraction

The data may come from different experimental conditions and have different sizes and shapes



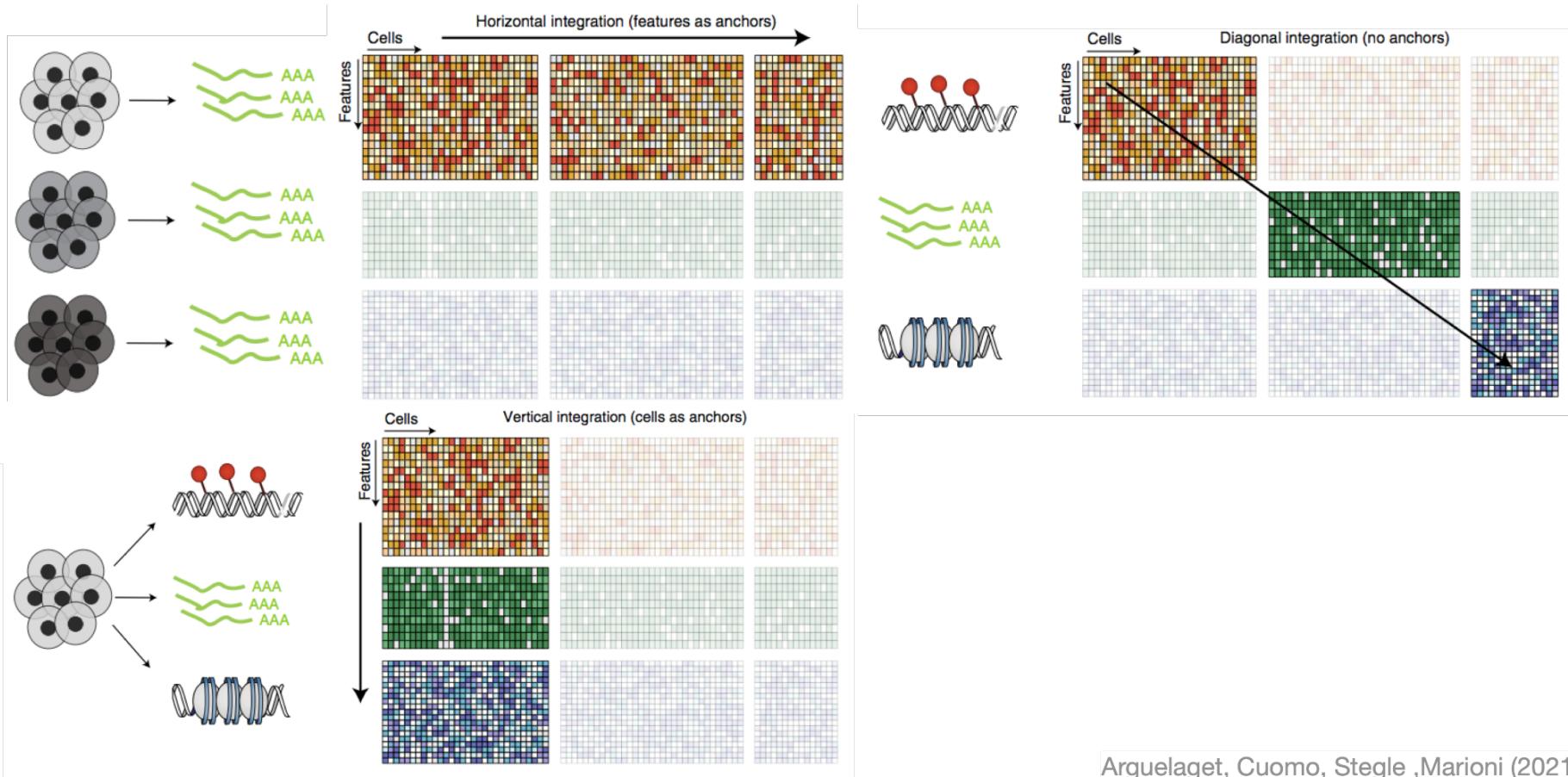
Challenges:

- sparsity
- scaling
- low level processing
- quality control
- normalization



What do we do with all these data?

first, find common coordinate frameworks

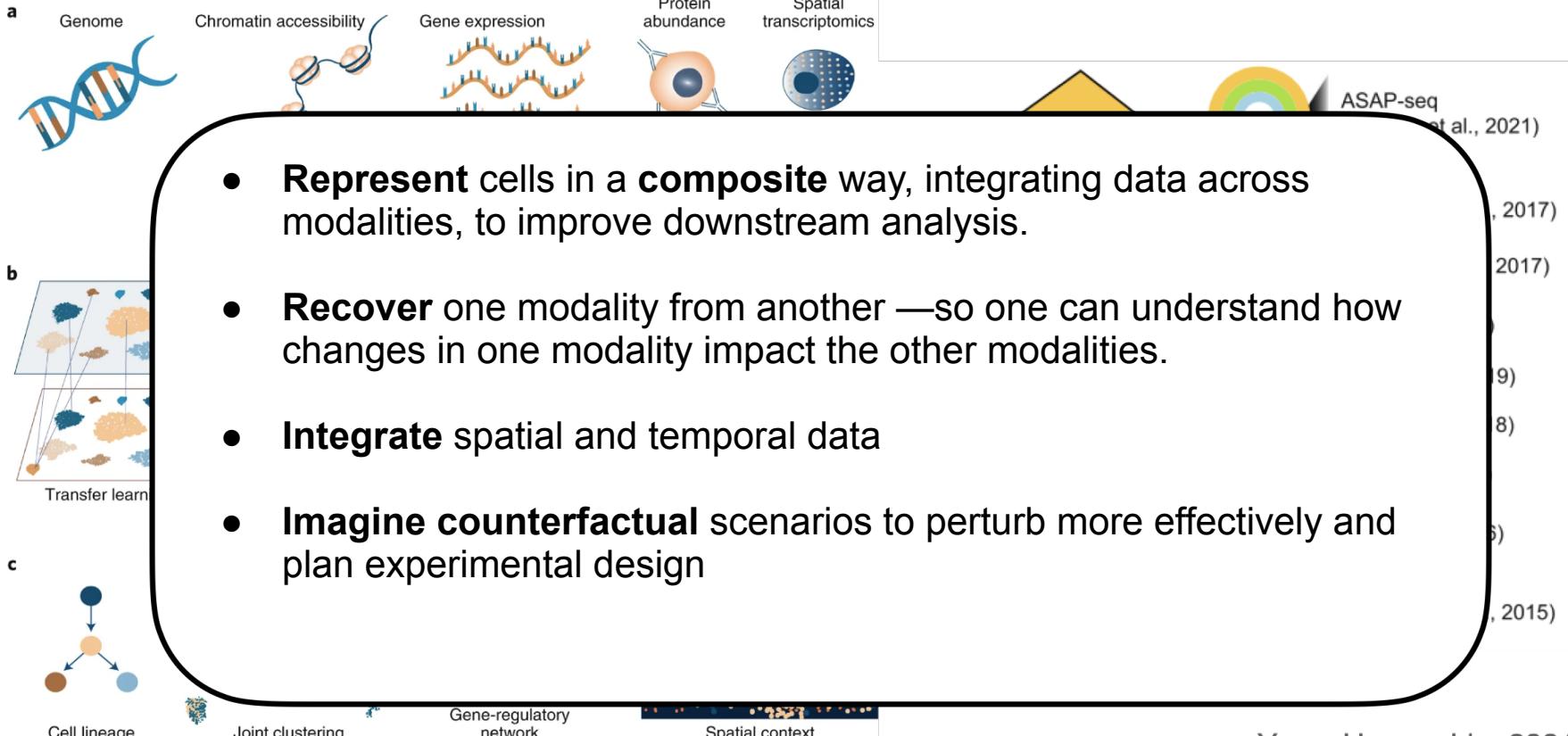


Arguelaguet, Cuomo, Stegle ,Marioni (2021)

What do we do with all these data?

then:

- **Represent** cells in a **composite** way, integrating data across modalities, to improve downstream analysis.
- **Recover** one modality from another —so one can understand how changes in one modality impact the other modalities.
- **Integrate** spatial and temporal data
- **Imagine counterfactual** scenarios to perturb more effectively and plan experimental design



What do we do with all these data?

Task: Dimensionality Reduction

Concept: Latent Variables

Given gene expression profiles for many cells

- **compress** these high-dimensional vectors into a set of **latent variables** that still **explain** what makes one cell different from another
- **cluster (how many cell types? are cell types discrete or continuous?)**

Dimensionality reduction: if cell types were discrete



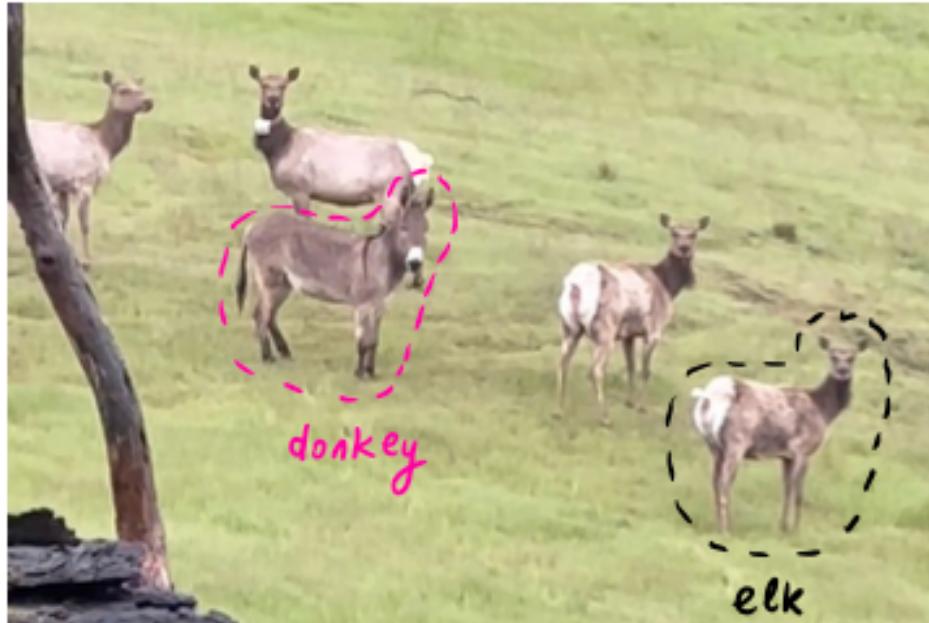
Source : instagram

“lost donkey found living best life
among heard of elks”

Dimensionality reduction

Donkey = cell A

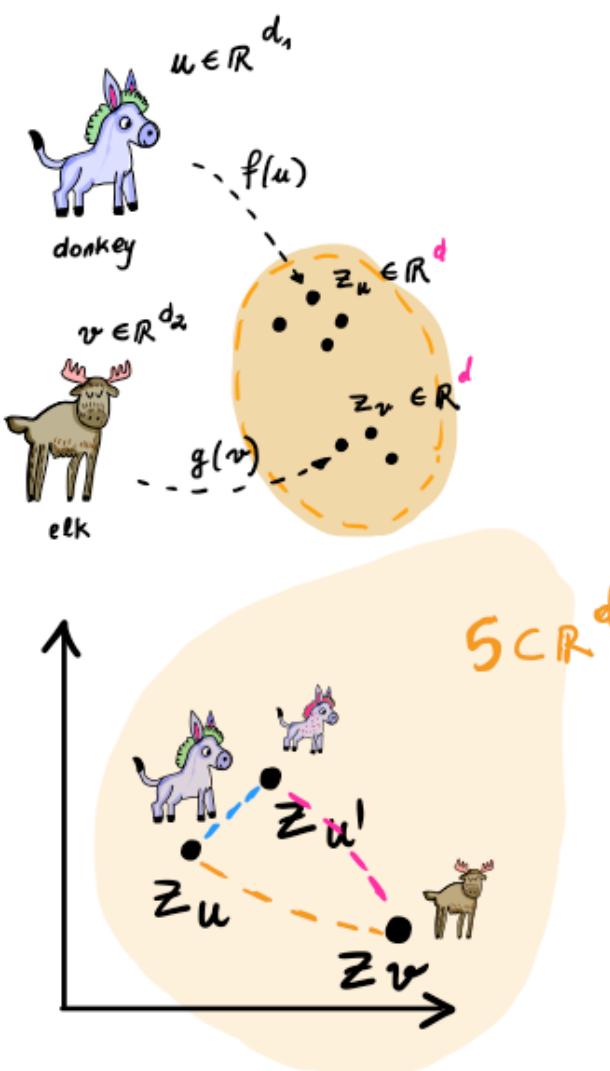
Elk = cell B



Source : instagram

“lost donkey found living best life
among heard of elks”

Dimensionality reduction

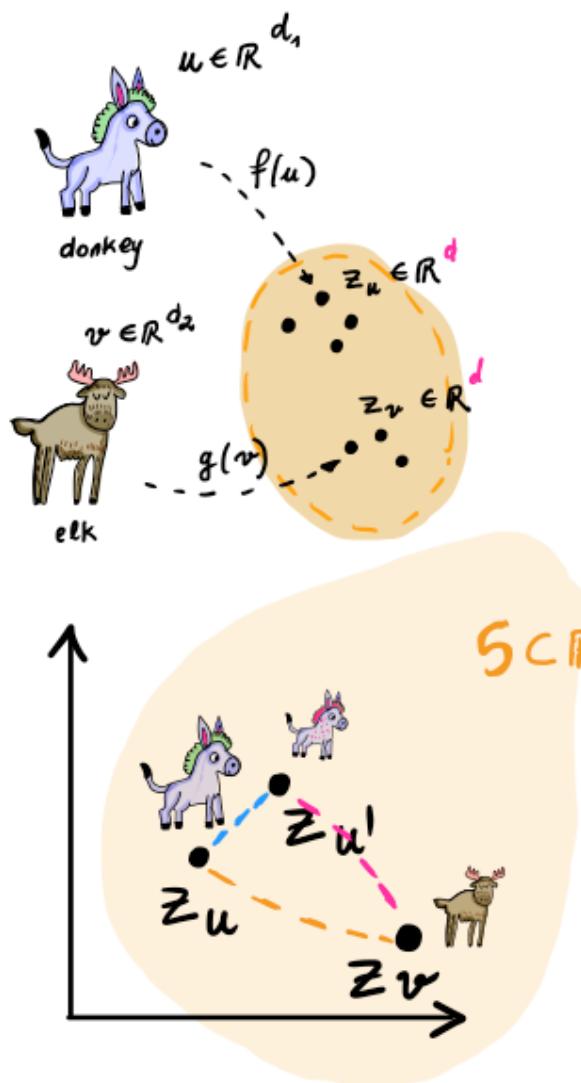


Distances (metric)

For all $z, z' \in S \subseteq \mathbb{R}^d$

- $d(z, z') \geq 0$
- $d(z, z) = 0$
- $d(z, z') = d(z', z)$ (symmetry)
- $d(z_u, z_v) \leq d(z_u, z_{u'}) + d(z_{u'}, z_v)$
(triangle inequality)

Dimensionality reduction



Distances (metric)

For all $z, z' \in \mathcal{S} \subseteq \mathbb{R}^d$

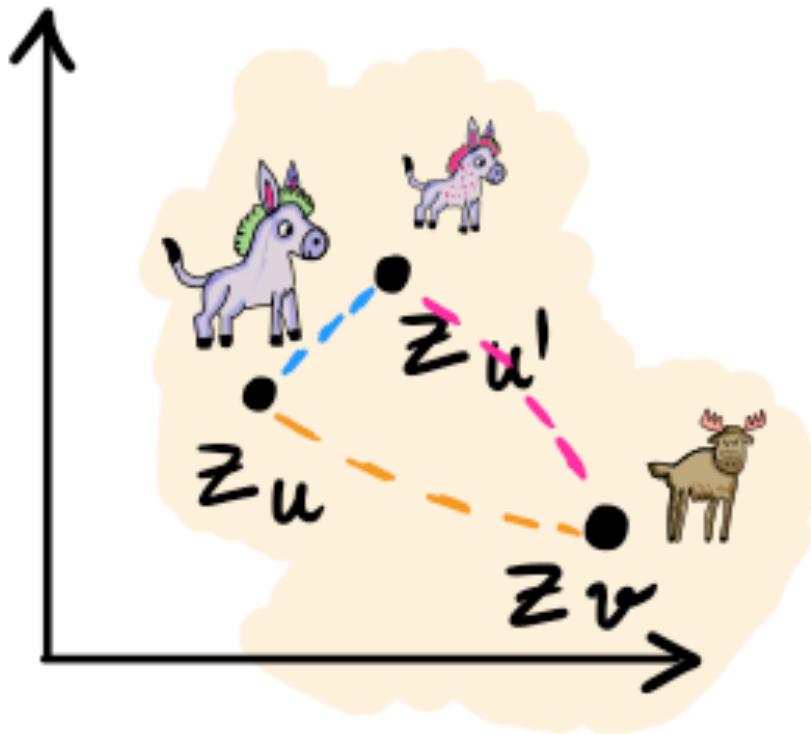
- $d(z, z') \geq 0$
- $d(z, z) = 0$
- $d(z, z') = d(z', z)$ (symmetry)
- $d(z_u, z_v) \leq d(z_u, z_{u'}) + d(z_{u'}, z_v)$
(triangle inequality)

Example :

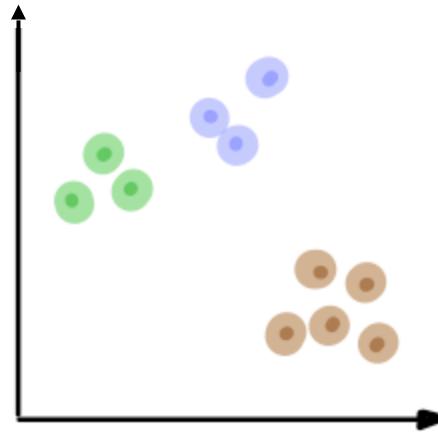
$$z \in \mathbb{R}^2, z' \in \mathbb{R}^2$$

$$\begin{aligned} d(z, z') &\triangleq \|z - z'\|_2 = \\ &= \sqrt{(z_1 - z'_1)^2 + (z_2 - z'_2)^2} \end{aligned}$$

Dimensionality reduction

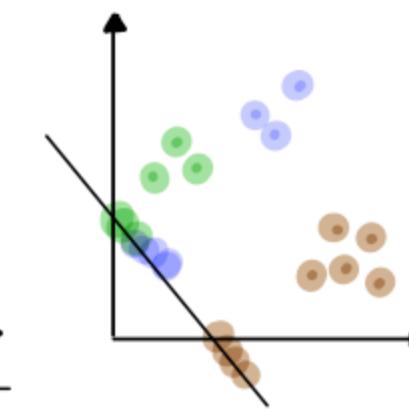
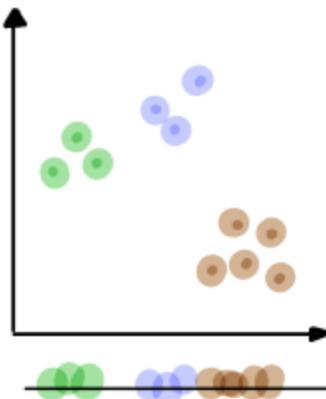
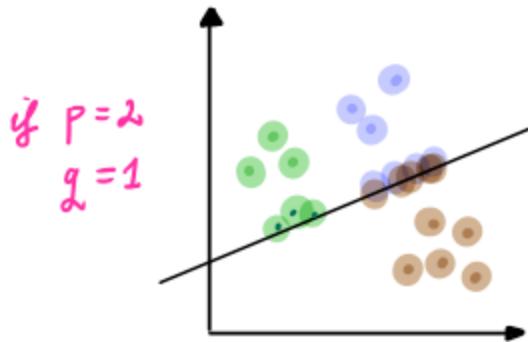


Data visualization with t-SNE

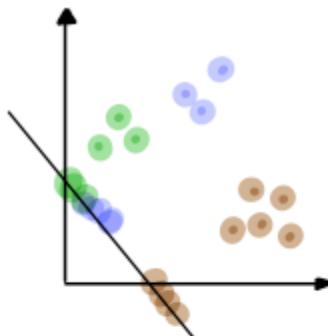


(input) (original) Data : $X \in \mathbb{R}^{n \times p}$
(output) Embedding : $Z \in \mathbb{R}^{n \times q}$
 $x_i \in \mathbb{R}^p, z_i \in \mathbb{R}^q$

(goal) : find z such that the distance between z 's is representative of that between observations X



Data visualization with t-SNE



(input) (original) Data : $X \in \mathbb{R}^{n \times P}$
(output) Embedding : $Z \in \mathbb{R}^{n \times 2}$
 $x_i \in \mathbb{R}^P$, $z_i \in \mathbb{R}^2$

Approach

- ① Define two quantities

$$P_{ij} = \frac{P_{j|i} + P_{i|j}}{2n}$$

with

$$P_{ij}, Q_{ij}$$

$$P_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}$$

$$Q_{ij} = \frac{(1 + \|z_i - z_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|z_k - z_l\|^2)^{-1}}$$

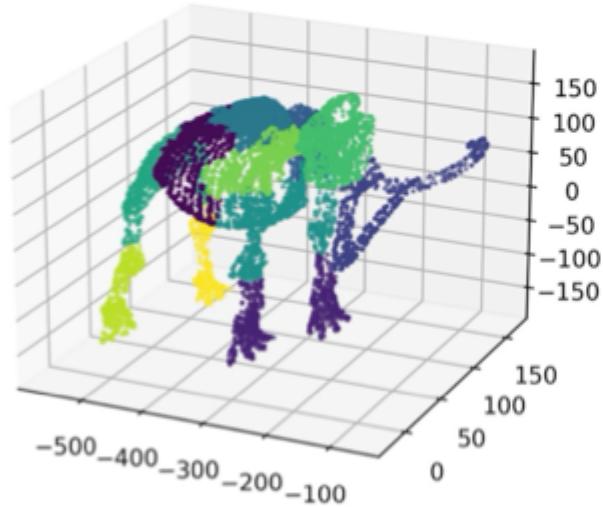
highlight parameters to estimate

- ② Minimize the loss function using gradient descent

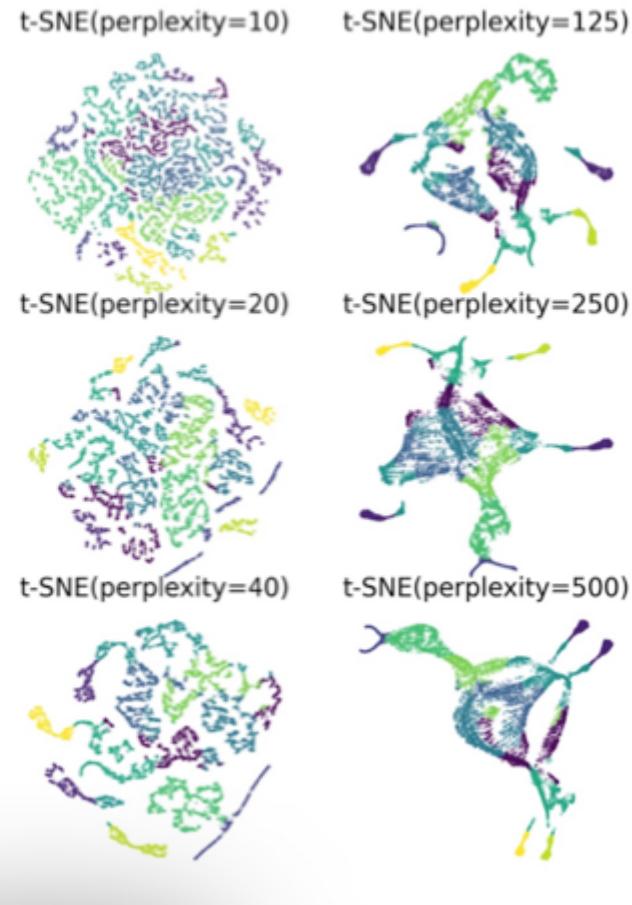
$$\sum_i \sum_j P_{ij} \log \frac{P_{ij}}{Q_{ij}} = C(z)$$

Data visualization with t-SNE

Original Mammoth

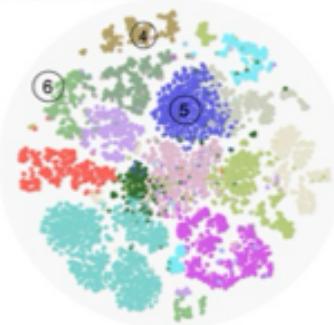


adapted from Wang, Huang et al.
(JMLR, 2021)
Understanding how dim. reduction
"tools work..."

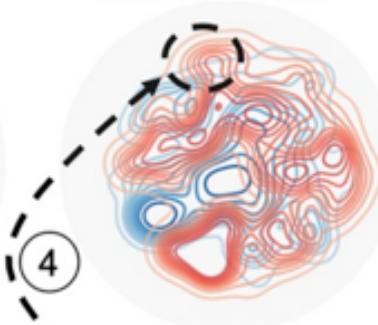


Data visualization with t-SNE

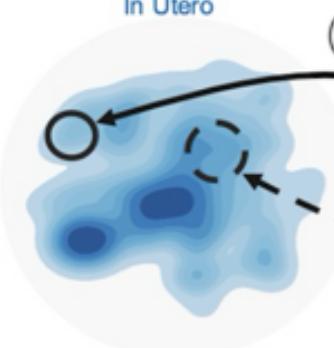
t-SNE perplexity: 5



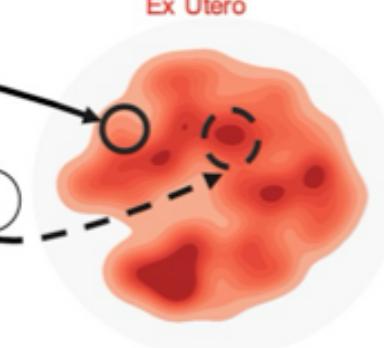
Ex Utero & In Utero



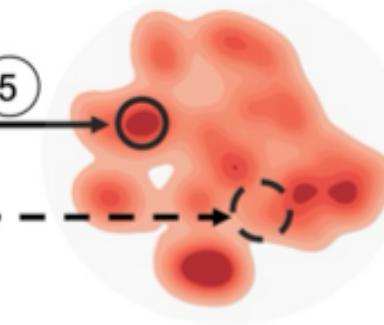
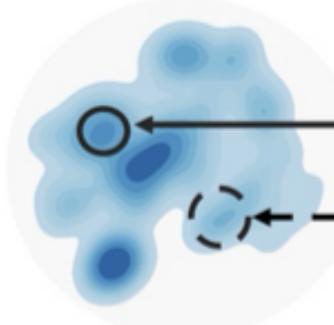
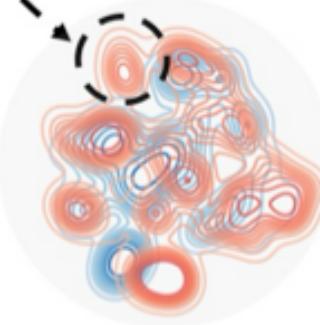
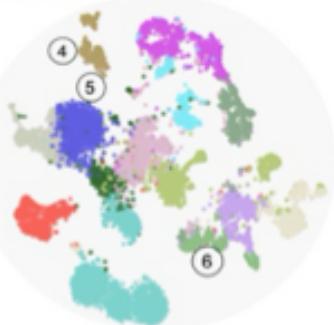
In Utero



Ex Utero



t-SNE perplexity: 50



Cell Types

● Neural Tube ⑥

● Amnion

● Extra-Embryonic Ectoderm

● Placodes ① ④

● Extra-Embryonic Endoderm

● Endothelial

● Foregut Mid-hindgut ② ⑤

● Pharyngeal Mesoderm ②

● Blood

● Mid-hindbrain

● Somitic Mesoderm ③

● Presomitic/Mixed Mesoderm

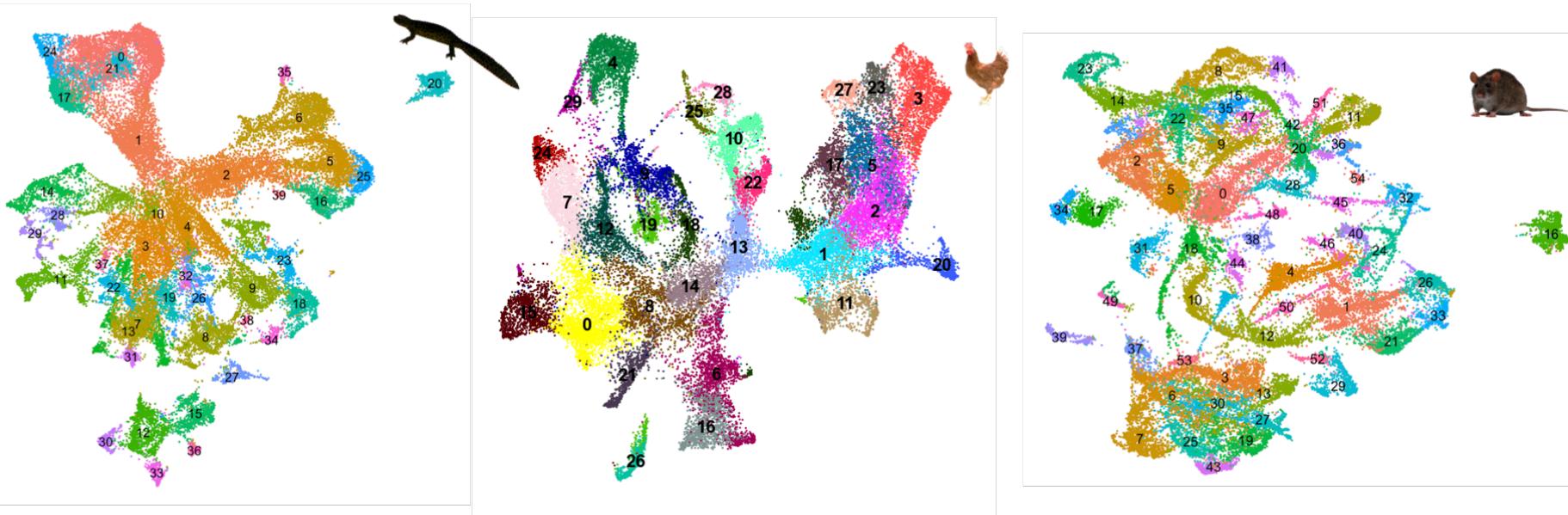
● Cardiac

● Extra-Embryonic Mesoderm

The specious art of single-cell genomics

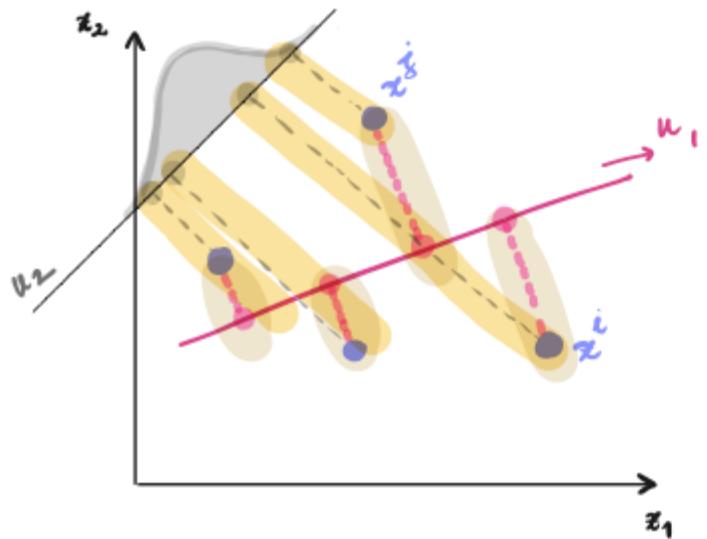
Tara Chari, Lior Pachter

Some desiderata: represent geometry across conditions, find common coordinate frameworks



Zaremba et al, 2025 Schmidt et al., 2023

Toolkit: Principal Component Analysis (PCA)



2 formulations

I. Maximum Variance
(Hotelling, 1933)

II. Minimum Error
(Pearson, 1901)

Toolkit: PCA

Maximum Variance
(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

Toolkit: PCA

Maximum Variance
(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the data

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

if $M=1$
we want to find u
 $u \in \mathbb{R}^D$

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the data

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

if $M=1$
we want to find u
 $u \in \mathbb{R}^D$

Projection : $u^T x_n$

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the data

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the data

if $M = 1$
we want to find $u \in \mathbb{R}^D$
Projection : $u^T x_n$
(scalar value)

Mean : $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

if $M=1$
we want to find u
 $u \in \mathbb{R}^D$

Projection : $u^T x_n$
(scalar value)

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the projected data.

Mean : $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$ Covariance : $S = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

if $M=1$
we want to find u
 $u \in \mathbb{R}^D$

Projection : $u^T x_n$
(scalar value)

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the projected data.

Mean : $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$ Covariance : $S = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T$

Variance of the projection : $\frac{1}{N} \sum_{n=1}^N (u^T x_n - u^T \bar{x})^2$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

Data $\{x_n\}_{n=1}^N$, $x_n \in \mathbb{R}^D$

Goal : find a projection P into a space of dimension $M < D$ that maximizes the variance of the data

$$\text{Variance of the projection} : \frac{1}{N} \sum_{n=1}^N (u^T x_n - u^T \bar{x})^2$$

$$\text{Goal : Find } \underset{u \in \mathbb{R}^D}{\operatorname{argmax}} \frac{1}{N} \sum_{n=1}^N (u^T x_n - u^T \bar{x})^2$$

Toolkit: PCA

Maximum Variance
(Hotelling, 1933)

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N (u^T x_n - u^T \bar{x})^2 &= \frac{1}{N} \sum_{n=1}^N u^T (x_n - \bar{x})^2 u \\ &= u^T \left[\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2 \right] u \end{aligned}$$

Toolkit: PCA

Maximum Variance
(Hotelling, 1933)

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N (u^T x_n - u^T \bar{x})^2 &= \frac{1}{N} \sum_{n=1}^N u^T (x_n - \bar{x})^2 u \\ &= u^T \left[\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2 \right] u \end{aligned}$$

S
(data covariance)

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$\underset{\mathbf{u} \in \mathbb{R}^D}{\operatorname{argmax}} \mathbf{u}^T \mathbf{S} \mathbf{u}$$

subject to

$$\mathbf{u}^T \mathbf{u} = 1$$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$\underset{\mathbf{u} \in \mathbb{R}^D}{\operatorname{argmax}} \mathbf{u}^T \mathbf{S} \mathbf{u}$$

subject to

$$\mathbf{u}^T \mathbf{u} = 1$$

Optimize $\mathbf{u}^T \mathbf{S} \mathbf{u} + \lambda(1 - \mathbf{u}^T \mathbf{u})$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$\underset{\mathbf{u} \in \mathbb{R}^D}{\operatorname{argmax}} \mathbf{u}^T \mathbf{S} \mathbf{u}$$

subject to

$$\mathbf{u}^T \mathbf{u} = 1$$

Optimize $\mathbf{u}^T \mathbf{S} \mathbf{u} + \lambda(1 - \mathbf{u}^T \mathbf{u}) = f(\mathbf{u})$

$$\frac{\partial f(\mathbf{u})}{\partial \mathbf{u}} = 0$$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$\underset{\mathbf{u} \in \mathbb{R}^D}{\operatorname{argmax}} \mathbf{u}^T \mathbf{S} \mathbf{u}$$

subject to

$$\mathbf{u}^* : \mathbf{S} \mathbf{u}^* = \lambda \mathbf{u}^*$$

$$\mathbf{u}^T \mathbf{u} = 1$$

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$\underset{\mathbf{u} \in \mathbb{R}^D}{\operatorname{argmax}} \mathbf{u}^T \mathbf{S} \mathbf{u}$$

subject to

$$\mathbf{u}^* : \mathbf{S} \mathbf{u}^* = \lambda \mathbf{u}^* \quad \mathbf{u}^T \mathbf{u} = 1$$

- \mathbf{u} is an eigenvector of \mathbf{S}
- $\mathbf{u}^T \mathbf{S} \mathbf{u}^* = \lambda$ (variance)
- λ has to be the highest eigenvalue

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$x_1, x_2, \dots, x_N \in \mathbb{R}^D \xrightarrow{\text{projection}} u^T x_n \in \mathbb{R}$$

$$\underset{u \in \mathbb{R}^D}{\operatorname{argmax}} \quad u^T S u$$

\uparrow
 $D \times D$

(u the eigenvector
with the highest
eigenvalue)

Toolkit: PCA

Maximum Variance

(Hotelling, 1933)

$$\underset{u \in \mathbb{R}^D}{\operatorname{argmax}} \frac{u^T S u}{\|u\|^2}$$

(u the eigenvector
with the highest
eigenvalue)

$$x_1, x_2, \dots, x_N \in \mathbb{R}^D \Rightarrow \underset{1 \times D}{u^T} \underset{D \times 1}{x_n} \in \mathbb{R}$$

for $u \in \mathbb{R}^{D \times M}$, $M \geq 2$
how do we choose u ?

Toolkit: dimensionality reduction + probabilistic modeling

PCA: great,
but no probabilistic model of the data

Generative modelling



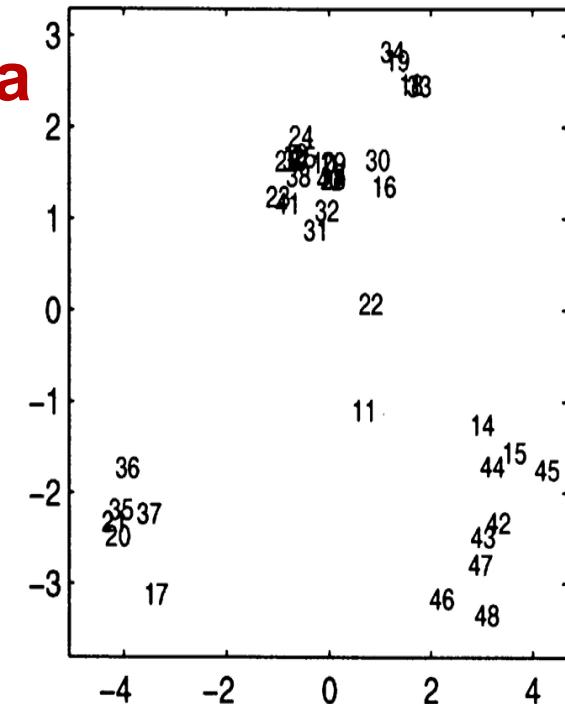
lower dimensional, **latent** embedding

$$z \sim p_\lambda(z)$$

observations (e.g., gene expression)

$$x \sim p_\theta(x|z)$$

Examples (linear): probabilistic PCA, Factor Analysis



pPCA projection of 18-dimensional Tobamovirus data (the counts of amino acid residues per molecule of coat protein)*

*probabilistic PCA (Tipping and Bishop, 1999)
PCA (Hotelling, 1933)

Toolkit: probabilistic PCA



$$z \sim N(0, I_d)$$

$$x = Wz + \varepsilon$$

with $W \in \mathbb{R}^{D \times d}$

$$\varepsilon \sim N(0, \sigma^2 I_d)$$

*probabilistic PCA (Tipping and Bishop, 1999)
PCA (Hotelling, 1933)

Toolkit: probabilistic PCA

$$\begin{array}{c} z \in \mathbb{R}^d \\ \downarrow \\ x \in \mathbb{R}^D \end{array}$$
$$z \sim N(0, I_d)$$
$$x = Wz + \varepsilon$$
$$\text{with } W \in \mathbb{R}^{D \times d}$$
$$\varepsilon \sim N(0, \sigma^2 I_D)$$

Then $x \sim N(0, W^T W + \sigma^2 I_D)$

Toolkit: probabilistic PCA

$$z \in \mathbb{R}^d \quad z \sim N(0, I_d)$$

$$x = Wz + \varepsilon$$

$$\text{with } W \in \mathbb{R}^{D \times d}$$

$$\varepsilon \sim N(0, \sigma^2 I_D)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_D)$$

Objective:

Maximize the likelihood with respect to the parameters W and σ^2

Toolkit: probabilistic PCA

$$\begin{array}{c} z \in \mathbb{R}^d \\ \downarrow \\ x \in \mathbb{R}^D \end{array}$$

$z \sim N(0, I_d)$

$x = Wz + \varepsilon$

with $W \in \mathbb{R}^{D \times d}$

$\varepsilon \sim N(0, \sigma^2 I_D)$

Then $x \sim N(0, W^T W + \sigma^2 I_D)$

Objective:

Maximize the likelihood with respect to the parameters W and σ^2

$$\underset{W, \sigma^2}{\operatorname{argmax}} \quad p(X | W, \sigma^2)$$

Toolkit: probabilistic PCA

$$\begin{array}{c} z \in \mathbb{R}^d \\ \downarrow \\ x \in \mathbb{R}^D \end{array}$$
$$z \sim N(0, I_d)$$
$$x = Wz + \varepsilon$$
$$\text{with } W \in \mathbb{R}^{D \times d}$$

$$\varepsilon \sim N(0, \sigma^2 I_D)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_D)$$

$$\text{Log likelihood: } \mathcal{L} = -\frac{N}{2} \left\{ d \ln(2\pi) + \ln|C| + \text{tr}(C^{-1} S) \right\}$$

$$\text{where } S = \frac{1}{N} \sum_{n=1}^N x_n x_n^T$$

$$C = W W^T + \sigma^2 I_D$$

Toolkit: probabilistic PCA

$$z \in \mathbb{R}^d$$

$$z \sim N(0, I_d)$$

$$x \in \mathbb{R}^D$$

$$x = Wz + \varepsilon$$

$$\text{with } W \in \mathbb{R}^{D \times d}$$

$$\varepsilon \sim N(0, \sigma^2 I_D)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_D)$$

Objective:

Maximize the likelihood with respect to the parameters W and σ^2

$$\underset{W, \sigma^2}{\operatorname{argmax}} \quad p(X | W, \sigma^2)$$

Tipping, Bishop, 1999

Roweis, 1998

$$\hat{W}_{ML} = U \left(I - \hat{\sigma}_{ML}^2 I_d \right)^{1/2} R$$
$$\hat{\sigma}_{ML}^2 = \frac{1}{D-d} \sum_{i=d+1}^D \lambda_i$$

Toolkit: probabilistic PCA

$$\begin{array}{l}
 z \in \mathbb{R}^d \\
 \downarrow \\
 x \in \mathbb{R}^D
 \end{array}
 \quad
 \begin{array}{l}
 z \sim N(0, I_d) \\
 x = Wz + \varepsilon \\
 \text{with } W \in \mathbb{R}^{D \times d} \\
 \varepsilon \sim N(0, \sigma^2 I_D)
 \end{array}$$

Then $x \sim N(0, W^T W + \sigma^2 I_D)$

Objective: Maximize the likelihood with respect to the parameters W and σ^2

$$\underset{W, \sigma^2}{\operatorname{argmax}} \quad p(X | W, \sigma^2)$$

Tipping, Bishop, 1999

Roweis, 1998

$U \rightarrow$ first d eigenvectors of

$$S = \frac{1}{N} \sum_{n=1}^N x_n x_n^T$$

$$\lambda_1 \geq \dots \geq \lambda_D > 0$$

$$\Lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_d \end{pmatrix}$$

$$\hat{W}_{ML} = U \left(\Lambda - \hat{\sigma}_{ML}^2 I_d \right)^{1/2} R$$

$$\hat{\sigma}_{ML}^2 = \frac{1}{D-d} \sum_{i=d+1}^D \lambda_i$$

"variance lost due to the projection"

Toolkit: probabilistic PCA

$$z \in \mathbb{R}^d$$

$$z \sim N(0, I_d)$$

$$x \in \mathbb{R}^D$$

$$x = Wz + \varepsilon$$

with $W \in \mathbb{R}^{D \times d}$

$$\varepsilon \sim N(0, \sigma^2 I_D)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_D)$$

$$\hat{W}_{ML} = U \left(I - \hat{\sigma}_{ML}^2 I_d \right)^{1/2} R$$

$$\hat{\sigma}_{ML} = \frac{1}{D-d} \sum_{i=d+1}^D \lambda_i$$

$U \rightarrow$ first d eigen vectors of

$$S = \frac{1}{N} \sum_{n=1}^N x_n x_n^T$$

$$\lambda_1 \geq \dots \geq \lambda_D > 0$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_d \end{pmatrix}$$

Not immediately obvious how
to connect probabilistic PCA to PCA!

Toolkit: probabilistic PCA

$$z \in \mathbb{R}^d$$

$$z \sim N(0, I_d)$$

$$x \in \mathbb{R}^D$$

$$x = Wz + \varepsilon$$

with $W \in \mathbb{R}^{D \times d}$

$$\varepsilon \sim N(0, \sigma^2 I_D)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_D)$$

Let's look at
 $z|x$!

$$\hat{W}_{ML} = U \left(I - \hat{\sigma}_{ML}^2 I_d \right)^{1/2} R$$

$$\hat{\sigma}_{ML} = \frac{1}{D-d} \sum_{i=d+1}^D \lambda_i$$

$U \rightarrow$ first d
eigen vectors of

with $S = \frac{1}{N} \sum_{n=1}^N x_n x_n^T$

$$\lambda_1 \geq \dots \geq \lambda_d > 0$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_d \end{pmatrix}$$

Toolkit: probabilistic PCA

$$z \in \mathbb{R}^d$$

$$z \sim N(0, I_d)$$

$$x = Wz + \varepsilon$$

$$\text{with } W \in \mathbb{R}^{D \times d}$$

$$\varepsilon \sim N(0, \sigma^2 I_d)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_d)$$

Let's look at

$$z | x !$$

Using Bayes' rule $z|x \sim N(M^{-1}W^T x, \sigma^2 M^{-1})$

$$\text{with } M = W^T W + \sigma^2 I_d$$

$$\hat{W}_{ML} = U \left(\Lambda - \hat{\sigma}_{ML}^2 I_d \right)^{1/2} R$$

$$\hat{\sigma}_{ML} = \frac{1}{D-d} \sum_{i=d+1}^D \lambda_i$$

$U \rightarrow$ first d eigen vectors of

with $S = \frac{1}{N} \sum_{n=1}^N x_n x_n^T$

$$\lambda_1 \geq \dots \geq \lambda_D > 0$$

$$\Lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_d \end{pmatrix}$$

Toolkit: probabilistic PCA

$$\begin{matrix} z \in \mathbb{R}^d \\ \downarrow \\ x \in \mathbb{R}^D \end{matrix}$$

$$z \sim N(0, I_d)$$

$$x = Wz + \varepsilon$$

$$\text{with } W \in \mathbb{R}^{D \times d}$$

$$\varepsilon \sim N(0, \sigma^2 I_d)$$

$$\text{Then } x \sim N(0, W^T W + \sigma^2 I_d)$$

Let's look at

$$z|x!$$

$$\text{Using Bayes' rule } z|x \sim N(M^{-1}W^T x, \sigma^2 M^{-1})$$

$$\text{with } M = W^T W + \sigma^2 I_d$$

$$\text{If } \sigma \rightarrow 0, \text{ then } E(z|x) = (W_{ML}^T W_{ML})^{-1} W_{ML}^T x$$

this is the orthogonal proj in latent space (PCA)

$$\hat{W}_{ML} = U \left(I - \hat{\sigma}_{ML}^2 I_d \right)^{1/2} R$$

$$\hat{\sigma}_{ML} = \frac{1}{D-d} \sum_{i=d+1}^D \lambda_i$$

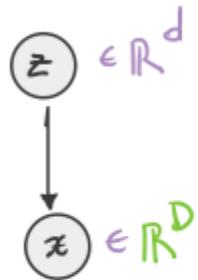
$U \rightarrow$ first d eigen vectors of

$$\text{with } S = \frac{1}{N} \sum_{n=1}^N x_n x_n^T$$

$$\lambda_1 \geq \dots \geq \lambda_d > 0$$

$$\Lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_d \end{pmatrix}$$

Toolkit: probabilistic PCA



$$z \sim N(0, I_d)$$

$$x = Wz + \varepsilon$$

with $W \in \mathbb{R}^{D \times d}$

$$\varepsilon \sim N(0, \sigma^2 I_d)$$

Good things:

- handles noise
- is generative
so we can do EM
- + sampling to
help w. missing data

Toolkit: probabilistic PCA

$$\begin{array}{ccc} z \in \mathbb{R}^d & z \sim N(0, I_d) & \\ \downarrow & & \\ x \in \mathbb{R}^D & x = Wz + \varepsilon & \\ & \text{with } W \in \mathbb{R}^{D \times d} & \\ & \varepsilon \sim N(0, \sigma^2 I_D) & \end{array}$$

Good things:

- × handles noise
- × is generative
so we can do EM
+ sampling to
help w. missing data

Opportunities

- × real data might not be gaussian
- × the latent → data map might not be linear
- × why should all the features be equally noisy?
- × what is the right dimension d

Probabilistic PCA + Inductive Bias

ZIFA

Pierson, E & Yau, C (2015)

$$\begin{aligned} z &\sim N(0, I) && \times \text{ has lots} \\ &&& \text{of zeroes} \\ x|z &\sim Wz + \epsilon && \times \text{ sparsity seems} \\ &&& \text{stochastic} \end{aligned}$$

?

PPCA

Data

Probabilistic PCA + Inductive Bias

ZIFA

Prierson, E & Yau, C (2015)

$$z \sim N(0, I)$$

✗ has lots
of zeroes

solution:
model the
zeroes
accordingly!

$$x|z \sim Wz + \epsilon$$

✗ sparsity seems
stochastic

P PCA

Data

Probabilistic PCA + Inductive Bias

ZIFT

Pierson, E & Yau, C (2015)

$$z \sim N(0, I)$$

$$x|z \sim Wz + \epsilon$$

PPCA

not sparse

x has lots
of zeroes

x sparsity seems
stochastic

Data

sparse

$$z \sim N(0, I)$$

$$x|z \sim Wz + \epsilon$$

$$x = (x_{ij})_{i=1,n}^{j=1,B}$$

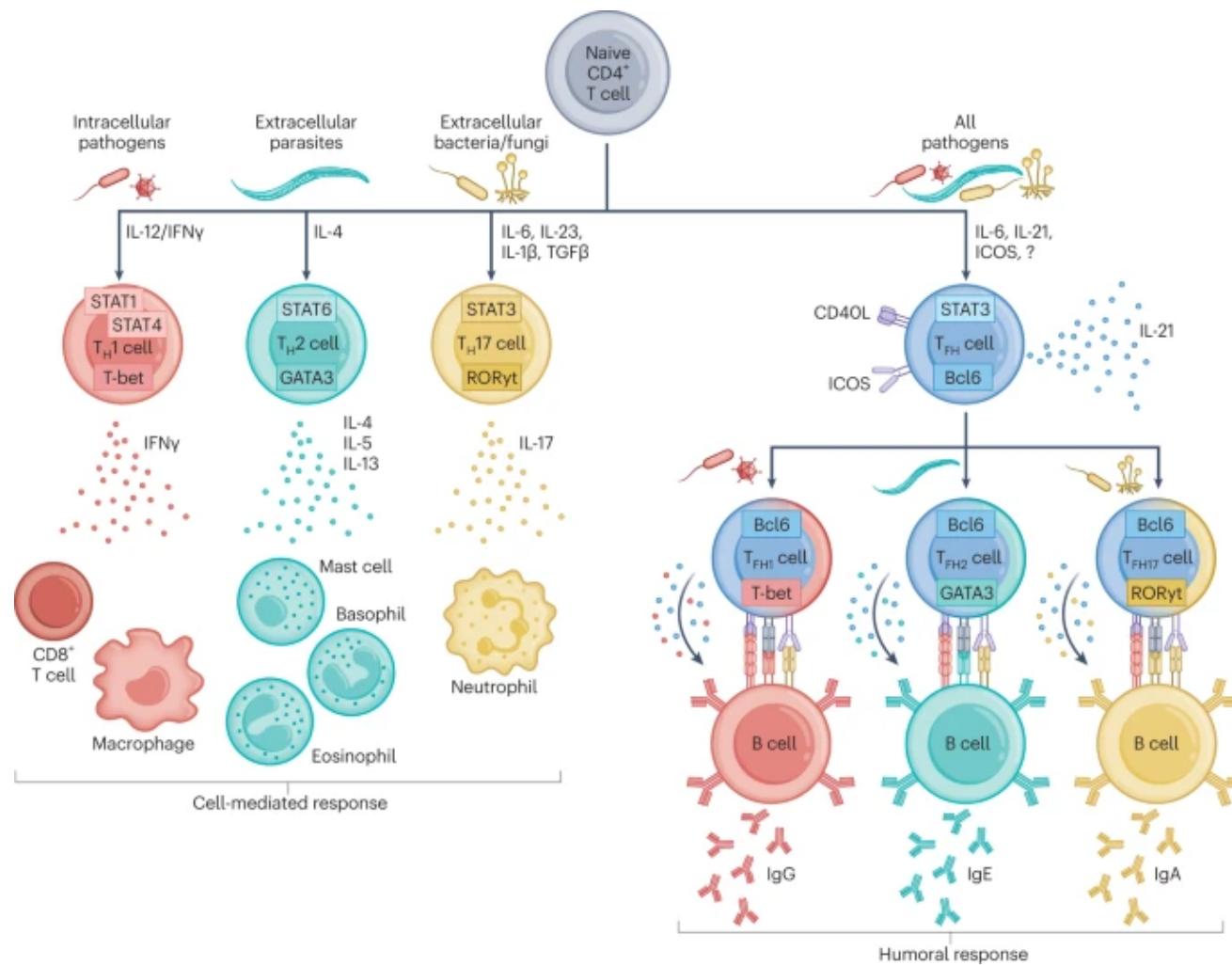
$$\rho_0 = \exp(-\lambda x_{ij}^2)$$

$$h_{ij}|x_{ij} \sim \text{Bern}(\rho_0)$$

$$y_{ij} = (1 - h_{ij}) x_{ij}$$

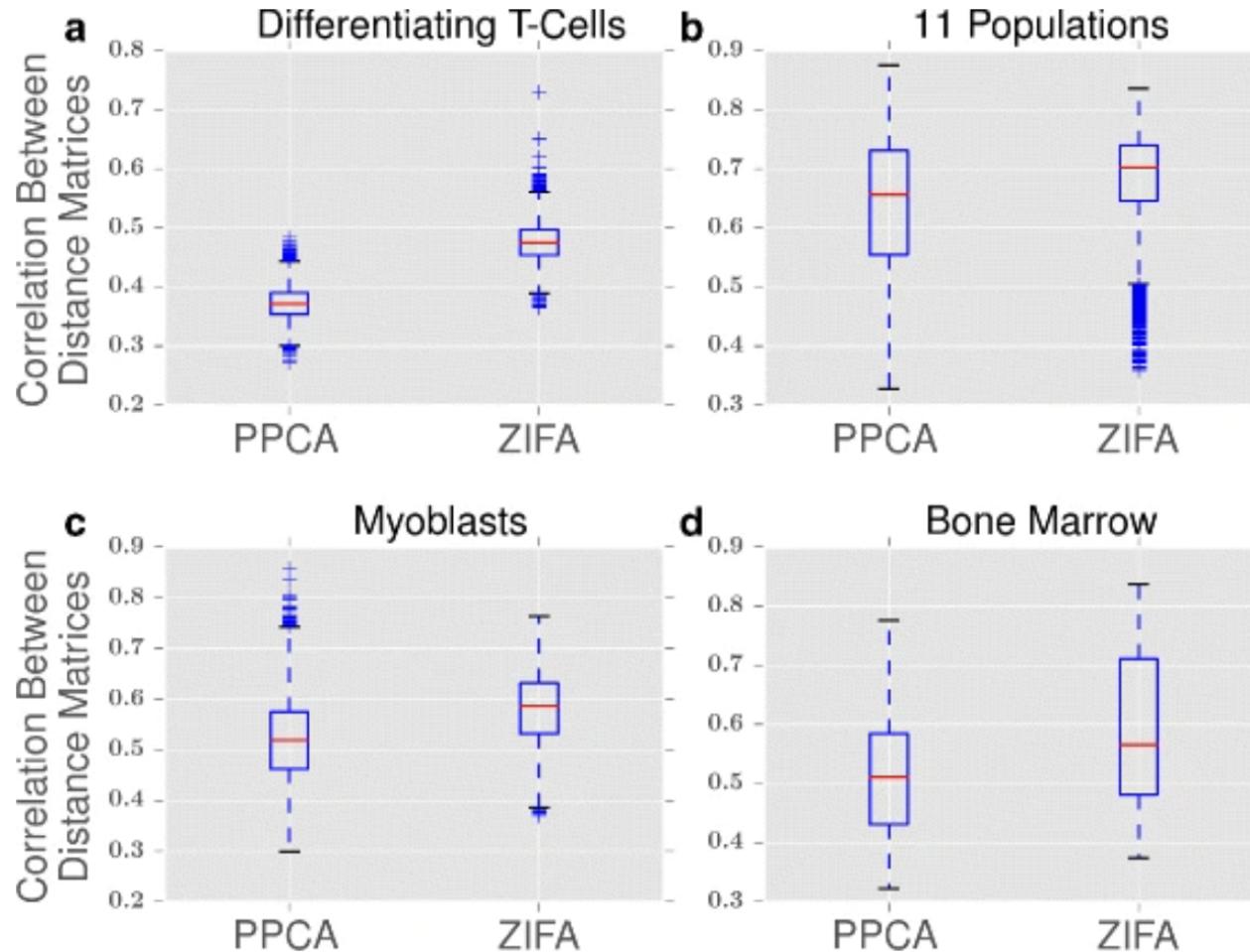
sparse

If you have ground truth, then you can tell if you are doing well



From: Künzli, M., Masopust, D. CD4⁺ T cell memory. Nat Immunol 24, 903–914 (2023)

If you have ground truth, then you can tell if you are really doing well



From: ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis

Toolkit: probabilistic PCA

$$\begin{array}{ccc} z \in \mathbb{R}^d & z \sim N(0, I_d) \\ \downarrow & \\ x \in \mathbb{R}^D & x = Wz + \varepsilon \\ & \text{with } W \in \mathbb{R}^{D \times d} \\ & \varepsilon \sim N(0, \sigma^2 I_D) \end{array}$$

Good things:

- handles noise
- is generative
so we can do EM
+ sampling to
help w. missing data

Opportunities

- real data might not be gaussian
 - the latent → data map might not be linear
 - why should all the features be equally noisy?
 - what is the right dimension d
- RNA-seq is count data
so Binomial or Poisson is more fitting
- gene regulatory networks, SeVi
- factor analysis
- always a problem

Toolkit: probabilistic PCA

$$z \in \mathbb{R}^d$$

$$z \sim N(0, I_d)$$

$$x \in \mathbb{R}^D$$

$$x = Wz + \varepsilon$$

$$\text{with } W \in \mathbb{R}^{D \times d}$$

$$\varepsilon \sim N(0, \sigma^2 I_D)$$

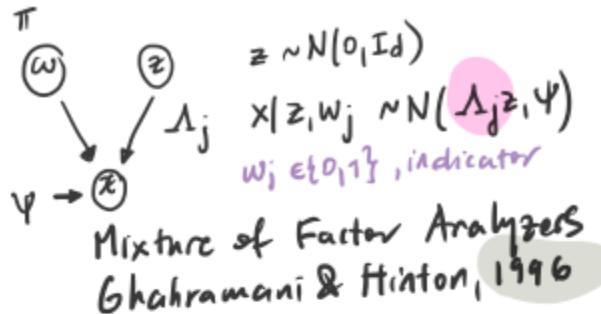


$$\varepsilon \sim N(0, \Psi)$$

$$\Psi = \begin{pmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots & \sigma_D^2 \end{pmatrix}$$

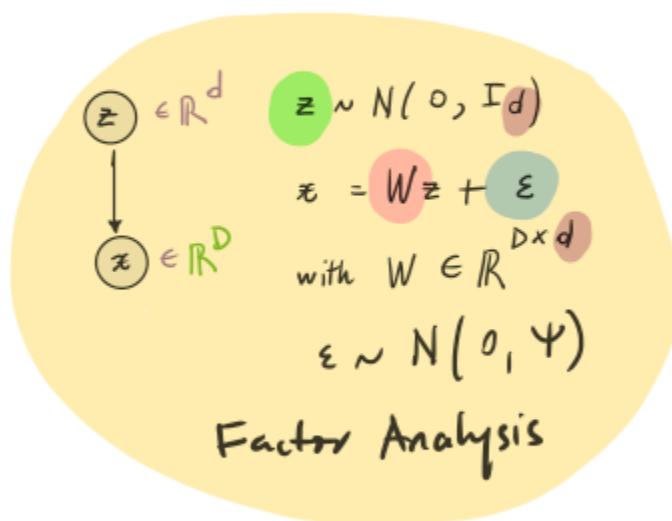
Factor Analysis!

A family of Factor Analysis methods for all that you seq

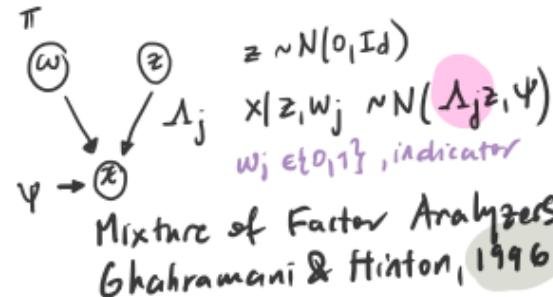


$\psi_i^{-1} \sim Ga(\alpha, \beta)$
 $z_{ik} \sim N(0, \sigma_{ik}^2)$
Sparse Factor Analysis
Engelhardt & Stephens, 2010

Priors on W
spike & slab
West 2003
Rockova &
George 2015
+ many

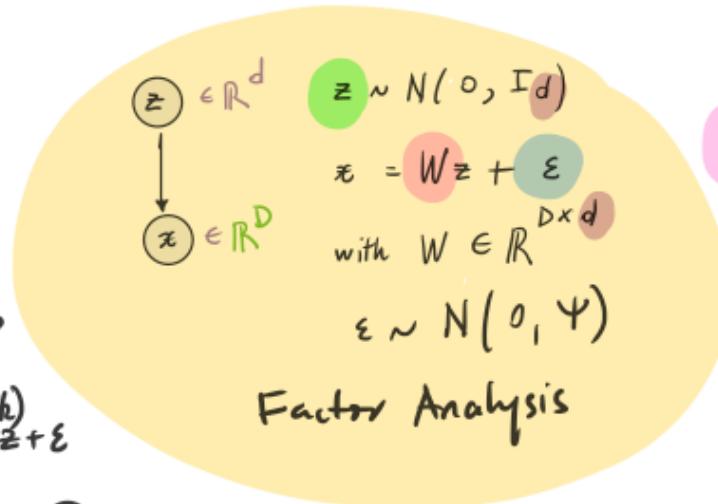
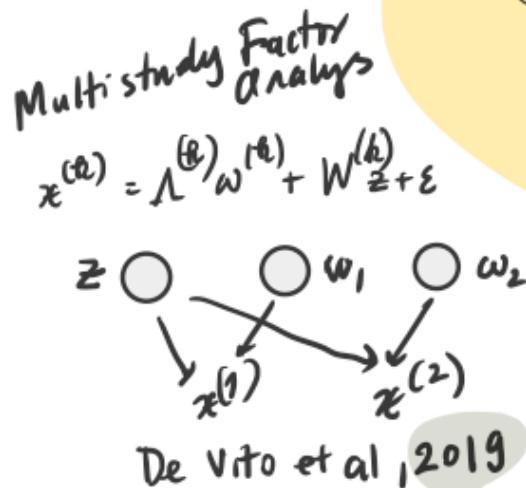


A family of Factor Analysis methods for all that you seq



$\psi_i^{-1} \sim Ga(\alpha, \beta)$
 $z_{ik} \sim N(0, \sigma_{ik}^2)$
Sparse Factor Analysis
Engelhardt & Stephens, 2010

Priors on W
spike & slab
West 2003
Rockova &
George 2015
+ many



many applications
≥ 2020

↑
usually
a Gaussian
Process Prior
(more about
this Lecture 3)

↑
spatial
extensions
Toussaint &
Engelhardt 2022
dynamic/
temporal
extensions

Next time

- **Going deep and nonlinear and what we hope to gain from that**
- **Learning from multiple conditions**

Practical toolkit: variational autoencoders, self-supervised learning



From latent spaces to living systems: Lecture 2

What do we gain from nonlinearity?

Bianca Dumitrascu
Columbia University
MLSS, Arequipa, Peru, 2025