



From latent spaces to living systems: Lecture 2

**What do we gain from nonlinearity?
How do we deal with conditions and
multiple modalities?**

Bianca Dumitrascu

Columbia University

MLSS, Arequipa, Peru, 2025

More on cells



Wassily Kadinsky, Color Study

Fun fact: Wound healing and beer



Around 2100 BC, a clay tablet described washing wounds with beer and hot water

All animals heal, some better than others

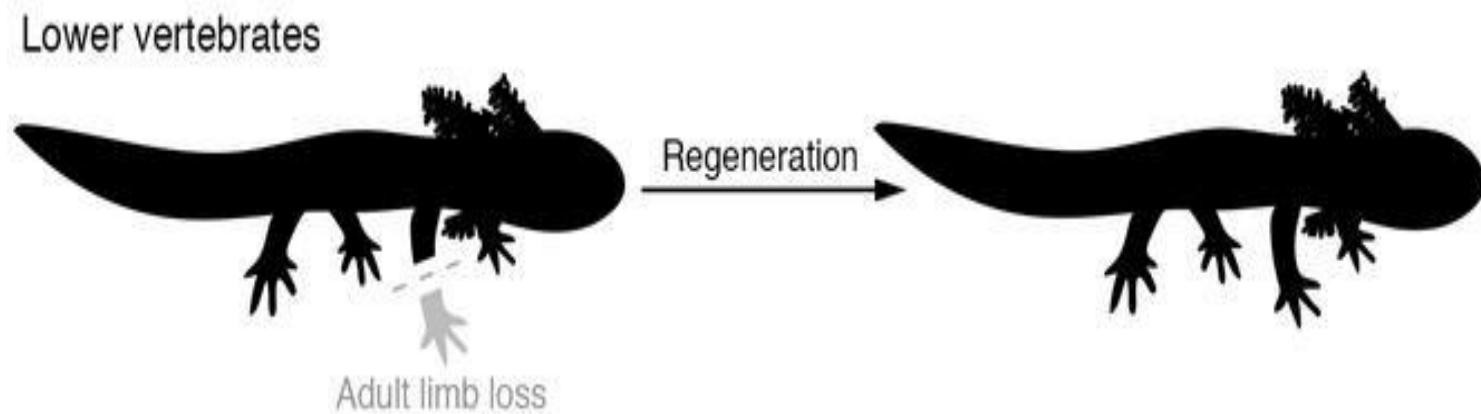
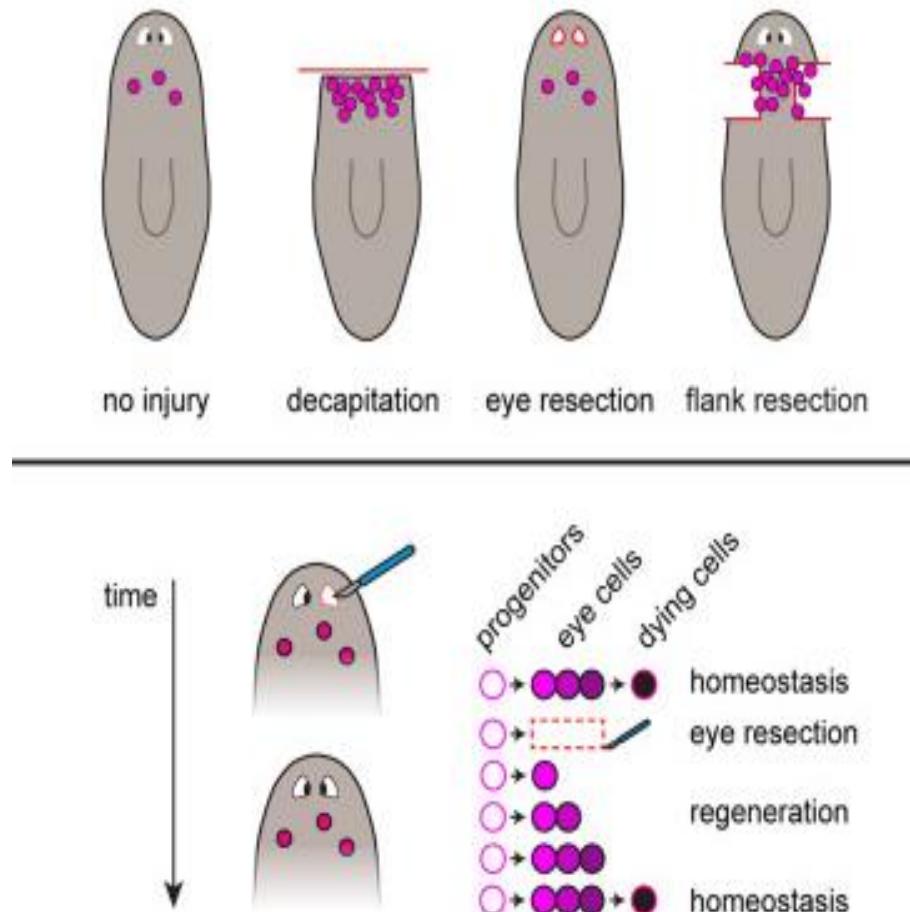


Figure adapted from: The balancing act of the liver: tissue regeneration versus fibrosis. Cordero-Espinoza & Huch. *J Clin Invest*. 2018;128(1):85-96. Bianca Dumitrescu, Machine Learning for Computational Biology, MESSI, 2023.

Some regenerate even their eyes

Some even regenerate: do not try this at home!



Graphical abstract from: LoCascio, Samuel A., Sylvain W. Lapan, and Peter W. Reddien. "Eye absence does not regulate planarian stem cells during eye regeneration." *Developmental Cell* 40, no. 4 (2017): 381-391.

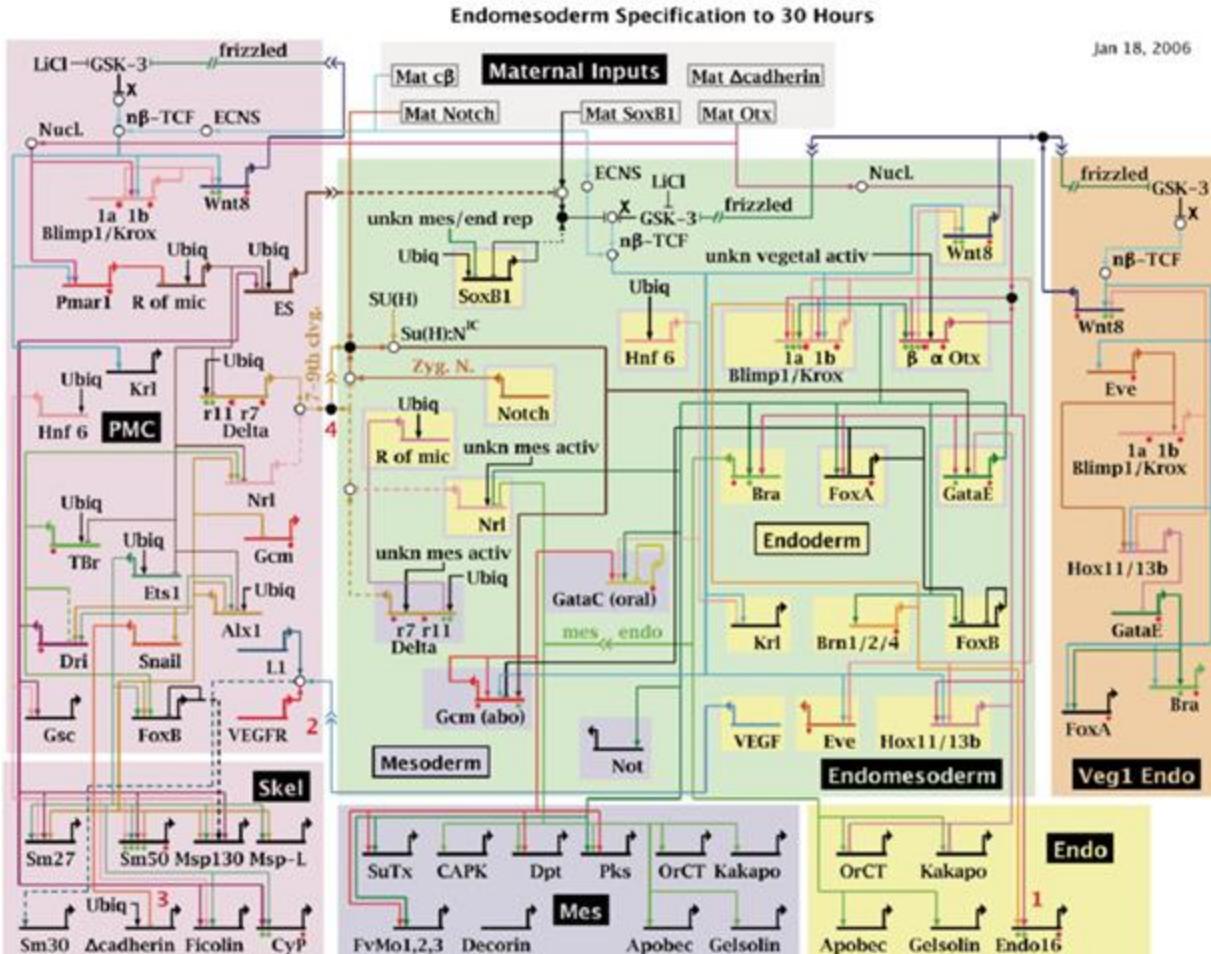
From last time: a glimpse on cell identity

Gene expression data: up to millions of points/ cells (rows; depending on question or system) and up to 40K genes (columns; depending on technology or system), multiple conditions



Dream data: non destructive, spatio-temporal gene expression and other modalities across many perturbations (genomic or environmental)

It turns out cells can compute and behave in complex ways, you could even say nonlinear



Ubiqu = ubiquitous; Mat = maternal; activ = activator; rep = repressor;
 unkn = unknown; Nucl. = nuclearization; χ = β -catenin source;
 $n\beta$ -TCF = nuclearized $n\beta$ -catenin-Tcf1; ES = early signal;
 ECNS = early cytoplasmic nuclearization system; Zyg. N. = zygotic Notch

Copyright © 2001–2006 Hamid Bolouri and Eric Davidson

Nonlinear gene interactions drive cell decision making

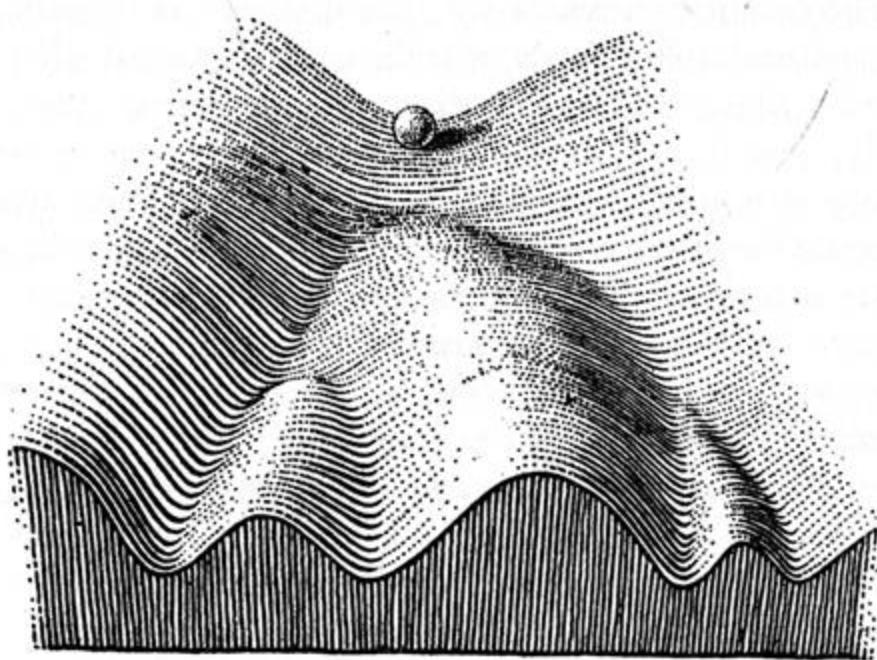


FIGURE 4

Part of an Epigenetic Landscape. The path followed by the ball, as it rolls down towards the spectator, corresponds to the developmental history of a particular part of the egg. There is first an alternative, towards the right or the left. Along the former path, a second alternative is offered; along the path to the left, the main channel continues leftwards, but there is an alternative path which, however, can only be reached over a threshold.

From Waddington, C. H. *The Strategy of the Genes* (Geo Allen & Unwin, London, 1957).

Nonlinear gene interactions drive cell decision making

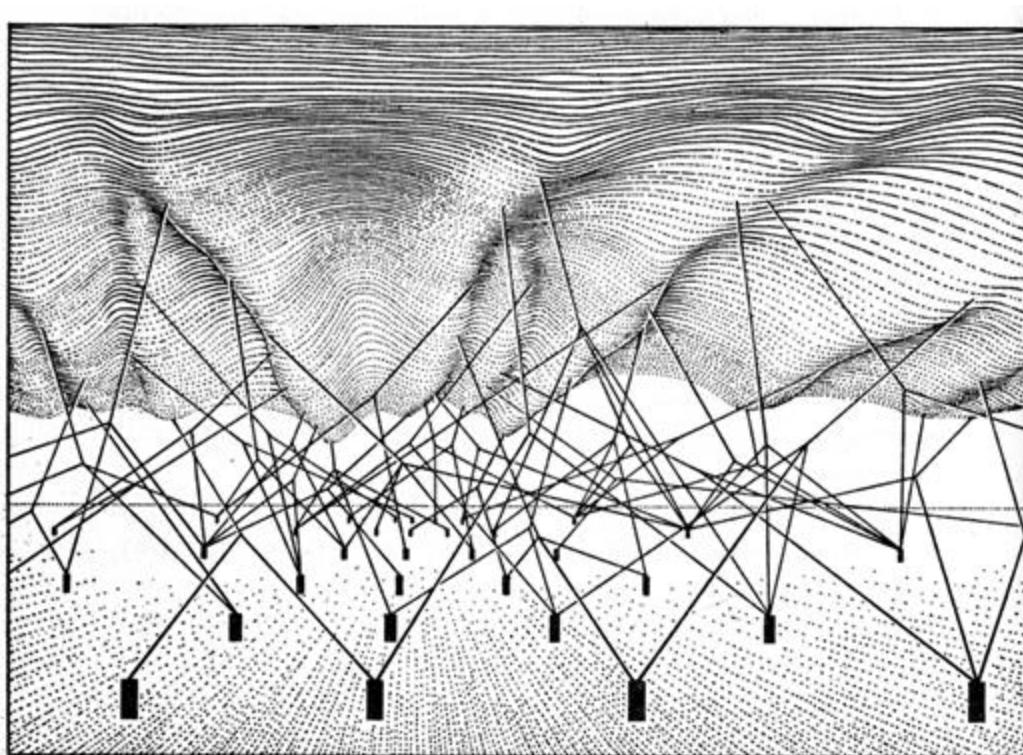
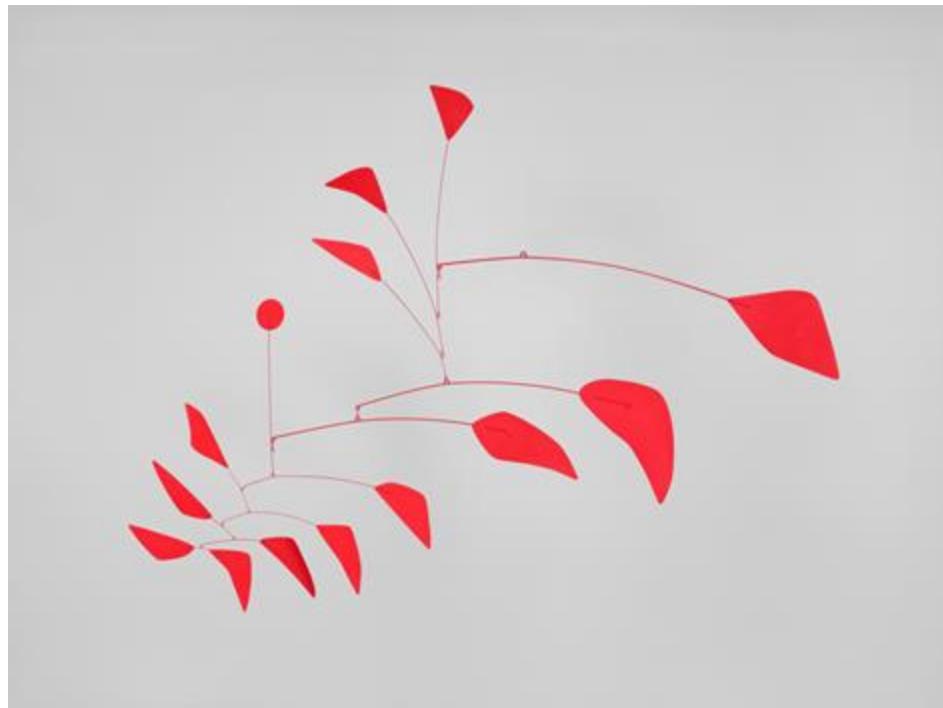


FIGURE 5

The complex system of interactions underlying the epigenetic landscape. The pegs in the ground represent genes; the strings leading from them the chemical tendencies which the genes produce. The modelling of the epigenetic landscape, which slopes down from above one's head towards the distance, is controlled by the pull of these numerous guy-ropes which are ultimately anchored to the genes.

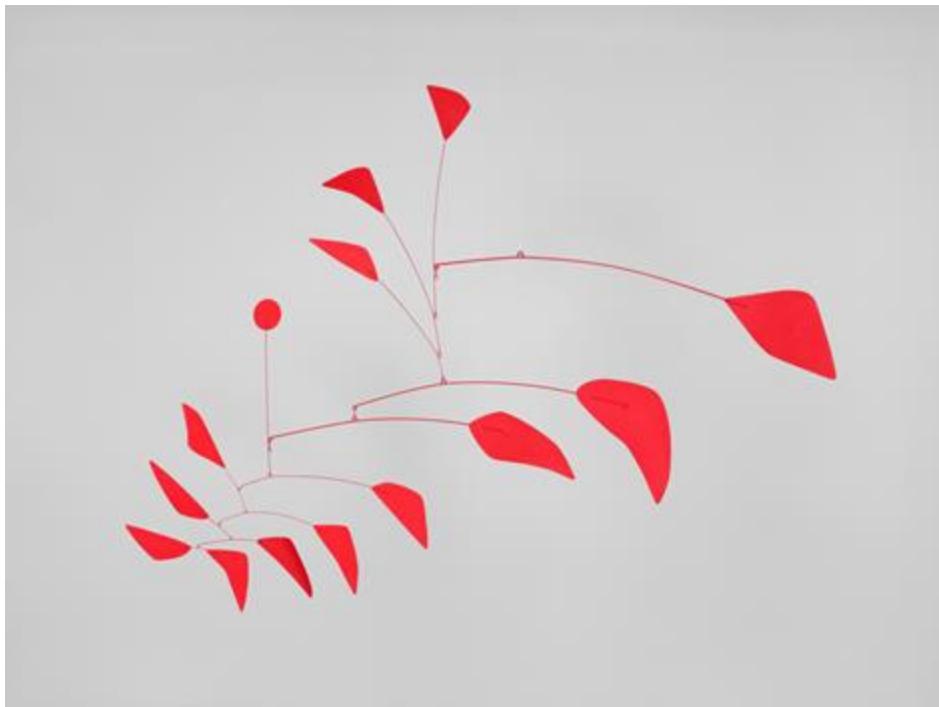
From Waddington, C. H. *The Strategy of the Genes* (Geo Allen & Unwin, London, 1957).

Nonlinear gene interactions drive cell decision making

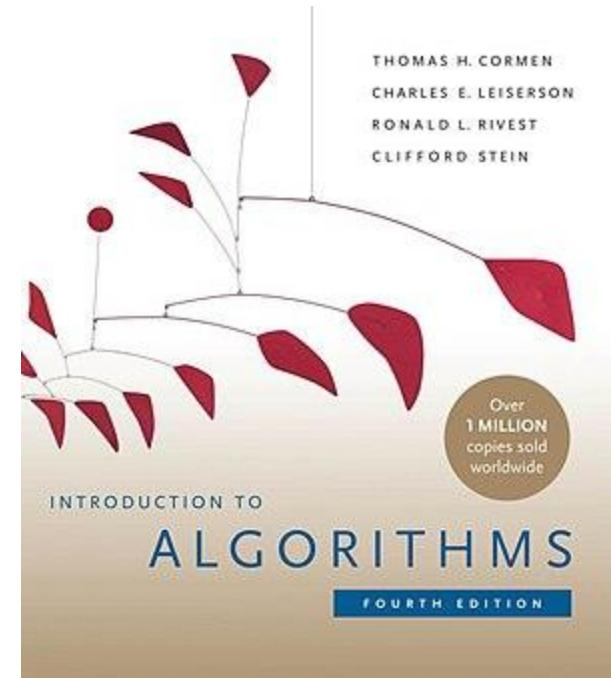


Alexander Calder

Nonlinear gene interactions drive cell decision making



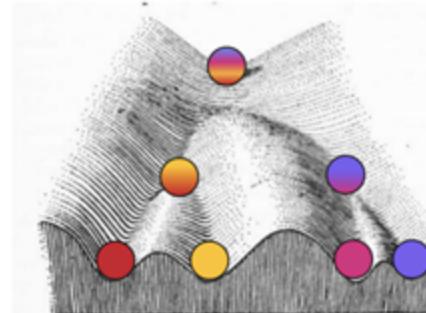
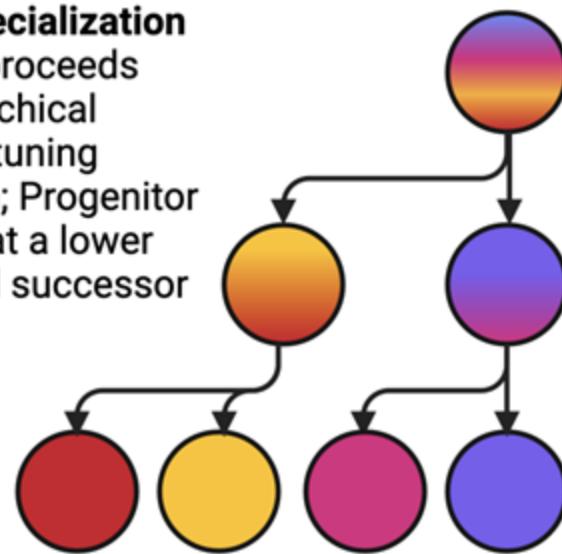
Alexander Calder



Nonlinear gene interactions drive cell decision making

Hierarchical specialization

Differentiation proceeds through a hierarchical process of fine-tuning cellular function; Progenitor states express at a lower level associated successor expression programs



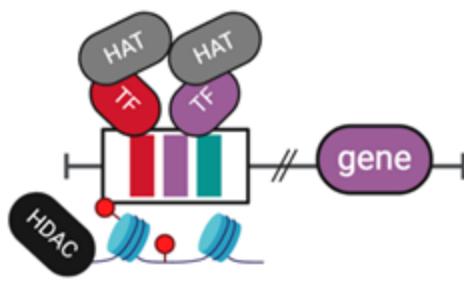
Adapted from “Hierarchical cell identities emerge from animal gene regulatory mechanisms”, Grishechkin, Abhirup Mukherjee, Omer Karin (2024).

v

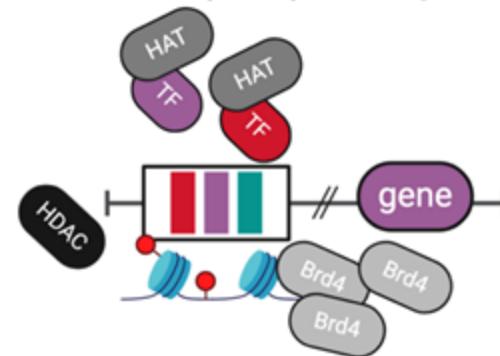
Bianca Dumitrascu, Machine Learning for Computational Biology, MLSS, 2025

Nonlinear gene interactions drive cell decision making

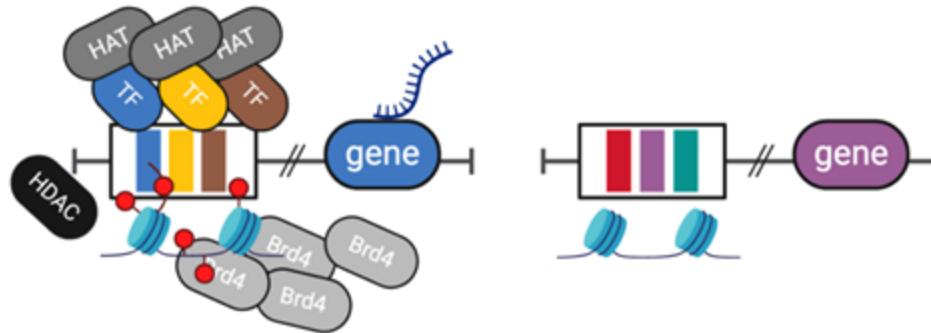
Transcription factor binding recruits HATs ("writers") to specific enhancers



Brd4 proteins ("readers") are recruited by acetylation signal

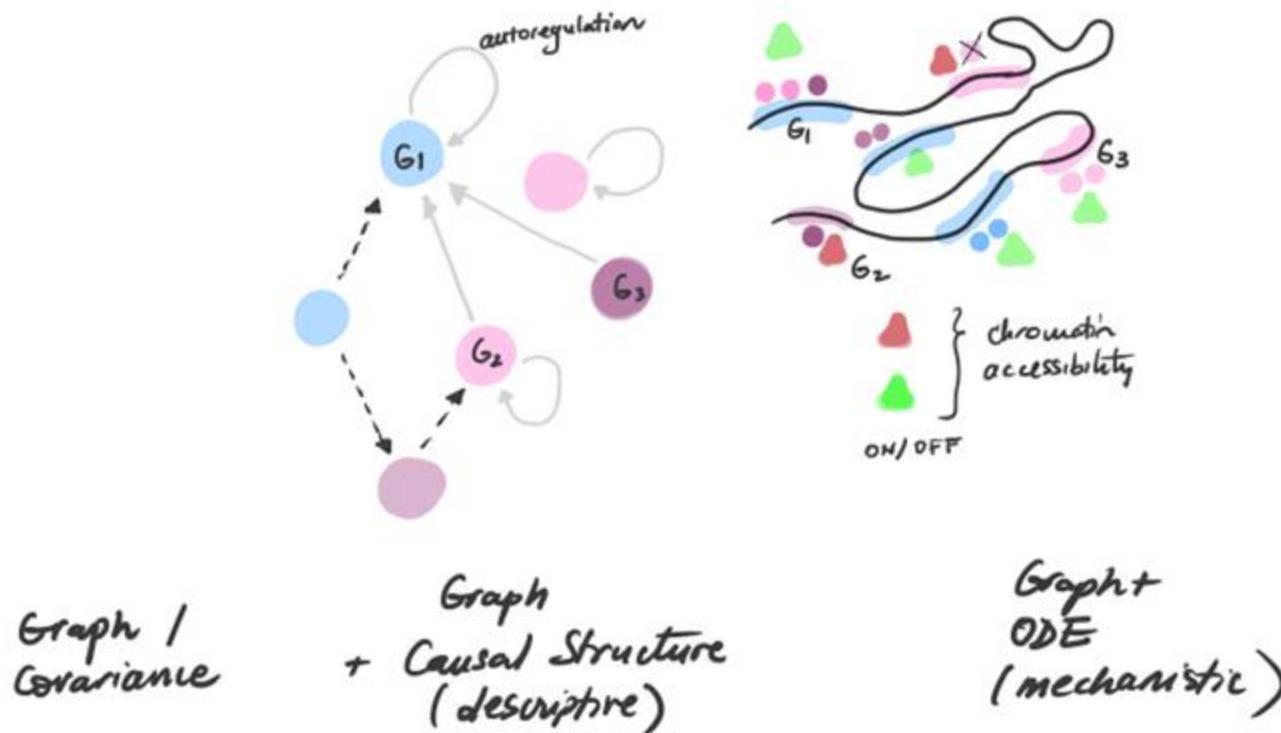


Genomic distribution of reader proteins, which control transcriptional elongation, determines gene expression



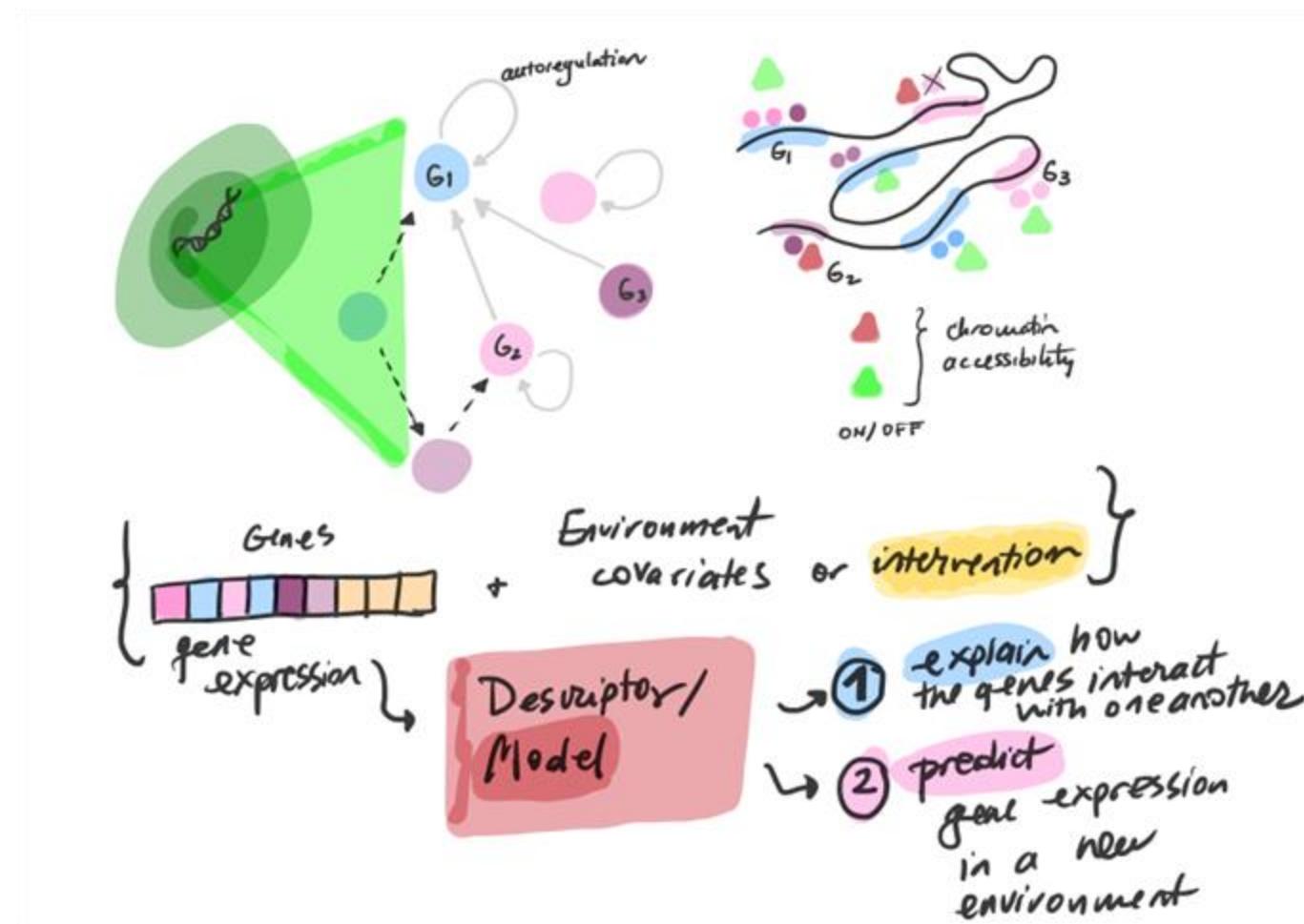
Adapted from "Hierarchical cell identities emerge from animal gene regulatory mechanisms", Grishechkin, Abhirup Mukherjee, Omer Karin (2024).

The case for nonlinear dimensionality reduction



- Linear, additive assumptions like those in PCA, NMF or Factor Analysis can find axis of variation or individual contributors, but can obfuscate how modalities interact with one another
- Biology is capable of generating complex responses, it's unlikely these behaviors can be **encoded** with only linear assumptions

The case for nonlinear dimensionality reduction



Building block: the variational autoencoder

Data x

Goal model $p_{\theta}(x)$

Approach assume a latent z
and a generative model

$$p_{\theta}(x, z) = p_{\theta}(x|z) p(z)$$

Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." 20 Dec. 2013,

Data x

Goal model $P_\theta(x)$

Approach assume a latent z
and a generative model

$$P_\theta(x, z) = P_\theta(x|z) P_\theta(z)$$

With maximum likelihood

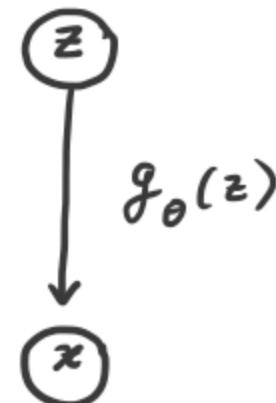
$$P_\theta(x) = \int P_\theta(z) P_\theta(x|z) dz$$



Data x

Goal model $p_{\theta}(x)$

Approach assume a latent z
and a generative model



$$p_{\theta}(x, z) = p_{\theta}(x|z) p_{\theta}(z)$$

With maximum likelihood

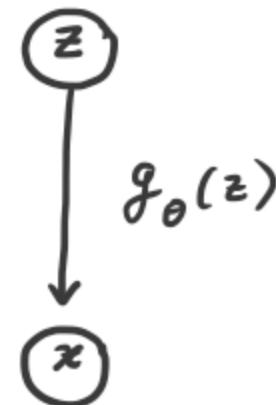
$$p_{\theta}(x) = \int p_{\theta}(z) \underbrace{p_{\theta}(x|z)}_{\text{decoder}} dz \leftarrow \text{typically intractable}$$

hence $p_{\theta}(z|x) = \frac{p_{\theta}(x,z)}{p_{\theta}(x)}$ \leftarrow also typically intractable

Data x

Goal model $p_{\theta}(x)$

Approach assume a latent z
and a generative model

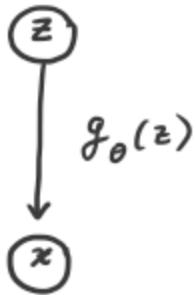


$$p_{\theta}(x, z) = p_{\theta}(x|z) p_{\theta}(z)$$

With maximum likelihood

$$p_{\theta}(x) = \int p_{\theta}(z) \underbrace{p_{\theta}(x|z)}_{\text{decoder}} dz \leftarrow \text{typically intractable}$$

hence $p_{\theta}(z|x) = \frac{p_{\theta}(x,z)}{p_{\theta}(x)}$ \leftarrow also typically intractable

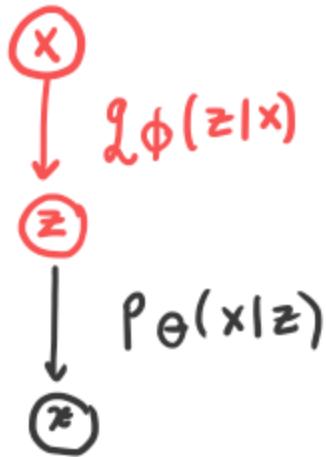


$$p_{\theta}(x) = \int p_{\theta}(z) \underbrace{p_{\theta}(x|z)}_{\text{decoder}} dz \quad \leftarrow \text{typically intractable}$$

$$p_{\theta}(z|x) = \frac{p_{\theta}(x,z)}{p_{\theta}(x)} \quad \leftarrow \text{also typically intractable}$$

Solution: introduce a parametric inference model

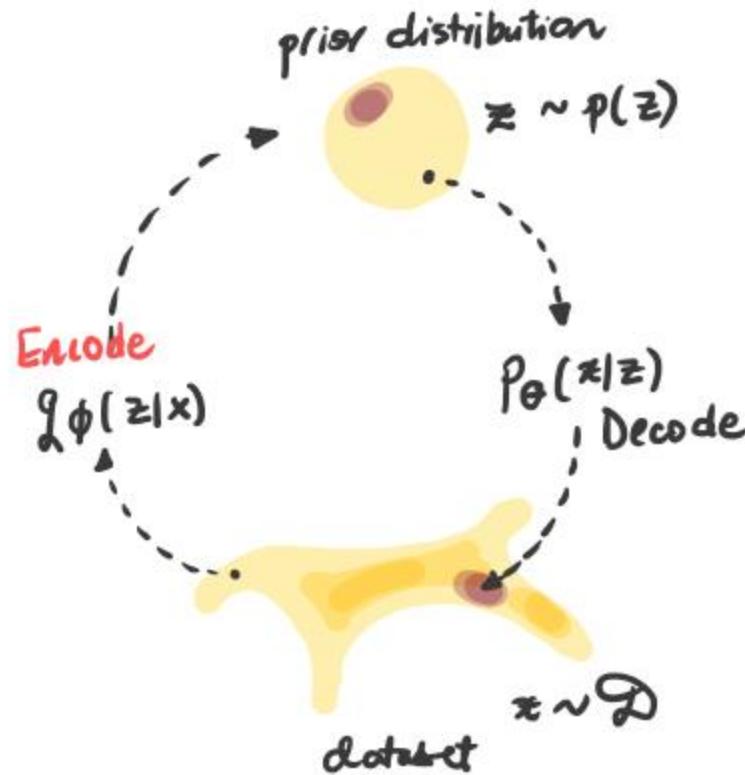
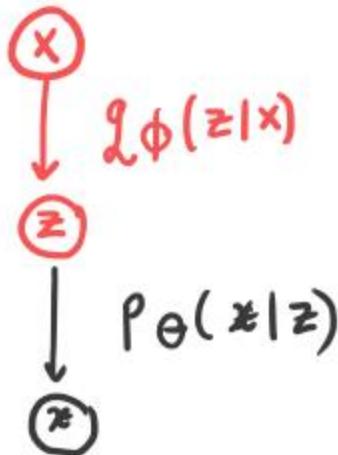
\$q_{\phi}(z|x)\$
 encoder
 recognition model



$p_{\theta}(x|z)$: decoder

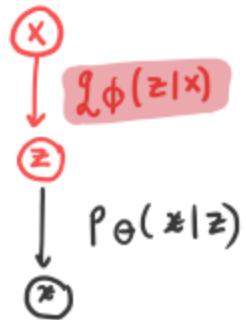
$q_{\phi}(z|x)$: encoder

ϕ : variational parameters



we want that

$$q_\phi(z|x) \approx p_\theta(z|x)$$



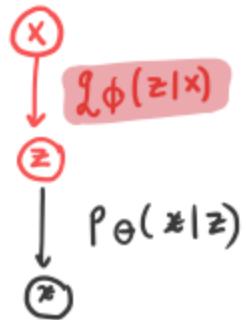
$$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x)] \\ = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{p_\theta(z|x)} \right]$$

$$q_\phi(z|x) \approx p_\theta(z|x)$$

$$= \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{q_\phi(z|x)} \cdot \frac{q_\phi(z|x)}{p_\theta(z|x)} \right]$$

Then

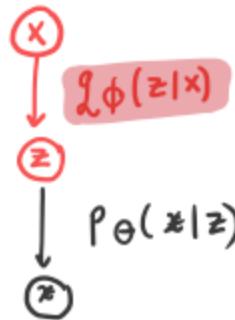
$$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] + \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)} \right]$$



we want that

$$q_\phi(z|x) \approx p_\theta(z|x)$$

$$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x|z)}{q_\phi(z|x)} \right] + \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)} \right]$$



we want that

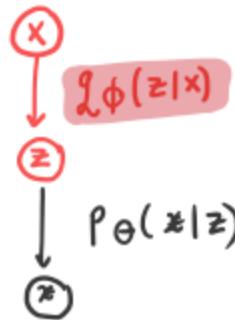
$$q_\phi(z|x) \approx p_\theta(z|x)$$

$$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] + \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)} \right]$$

z
x

≥ 0 (Kullback - Leibler divergence)

$$D_{KL}(q_\phi(z|x) \parallel p_\theta(z|x))$$



we want that

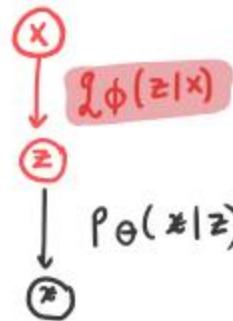
$$q_\phi(z|x) \approx p_\theta(z|x)$$

$$\underbrace{D_{KL}(q_\phi(z|x) || p_\theta(z|x))}_{\geq 0 \text{ (Kullback-Leibler divergence)}}$$

$$\mathcal{L}_{\theta, \phi}(x) \quad \begin{matrix} \text{(Evidence} \\ \text{lower bound} \\ \text{ELBO)} \end{matrix}$$

$$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] + \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)} \right]$$

Hence $\mathcal{L}_{\theta, \phi}(x) = \log p_\theta(x) - KL(\cdot || \cdot) \leq \log p_\theta(x)$



we want that

$$q_{\phi}(z|x) \approx p_{\theta}(z|x)$$

$$\mathcal{L}_{\theta, \phi}(x) \quad \begin{array}{l} \text{(Evidence} \\ \text{lower bound} \\ \text{ELBO)} \end{array}$$

≥ 0 (Kullback-Leibler divergence)

$$D_{KL}(q_{\phi}(z|x) \parallel p_{\theta}(z|x))$$

$$\log p_{\theta}(x) = \mathbb{E}_{q_{\phi}(z|x)} \left[\log \frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \right] + \mathbb{E}_{q_{\phi}(z|x)} \left[\log \frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right]$$

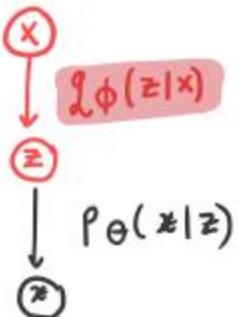
Hence $\mathcal{L}_{\theta, \phi}(x) = \log p_{\theta}(x) - KL(\cdot \parallel \cdot) \leq \log p_{\theta}(x)$

① $D_{KL}(q_{\phi}(z|x) \parallel p_{\theta}(z|x))$ is \star the distance of the approximate posterior from the true posterior

\star the gap between the ELBO and the marginal likelihood

② maximizing $\mathcal{L}_{\theta, \phi}(x)$ means

\star maximize the marginal likelihood $p_{\theta}(x)$ \star minimize the KL: posterior match is better



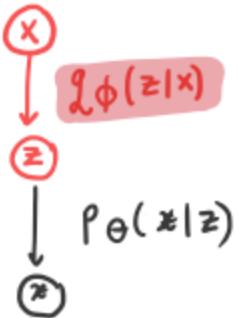
Optimizing $\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{g_{\phi}(z|x)} \left[\log \frac{p_{\theta}(x, z)}{g_{\phi}(z|x)} \right]$
 via Stochastic gradient Descent (SGD)

Given x_1, x_2, \dots, x_n iid.

$$\mathcal{L}_{\theta, \phi}(x) = \sum_{i=1}^n \mathcal{L}_{\theta, \phi}(x_i)$$

Goal : estimate $\nabla_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x_i)$

$$\nabla_{\theta} \mathcal{L}_{\theta, \phi}(x) = \nabla_{\theta} \mathbb{E}_{g_{\phi}(z|x)} \left[\log p_{\theta}(x, z) - \log g_{\phi}(z|x) \right]$$



Optimizing $\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right]$
 via Stochastic gradient Descent (SGD)

Given x_1, x_2, \dots, x_n iid.

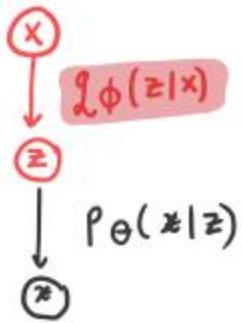
$$\mathcal{L}_{\theta, \phi}(x) = \sum_{i=1}^n \mathcal{L}_{\theta, \phi}(x_i)$$

Goal : estimate $\nabla_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x_i)$

This is the easy part!

$$\begin{aligned} \nabla_{\theta} \mathcal{L}_{\theta, \phi}(x) &= \nabla_{\theta} \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x, z) - \log q_\phi(z|x)] \\ &= \mathbb{E}_{q_\phi(z|x)} [\nabla_{\theta} \log p_\theta(x, z) - \nabla_{\theta} \log q_\phi(z|x)] \end{aligned}$$

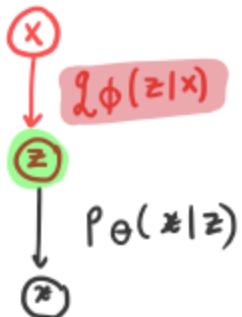
can get samples $\tilde{z} \sim q_\phi(z|x)$
and evaluate



Optimizing $\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right]$
 via Stochastic gradient Descent (SGD)

Goal : estimate $\nabla_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x_i)$

$$\begin{aligned}\nabla_{\phi} \mathcal{L}_{\theta, \phi}(x) &= \nabla_{\phi} \mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x, z) - \log q_\phi(z|x) \right] \\ &\neq \mathbb{E}_{q_\phi(z|x)} \left[\nabla_{\phi} (\overbrace{\quad}^{\text{---}}) \right]\end{aligned}$$

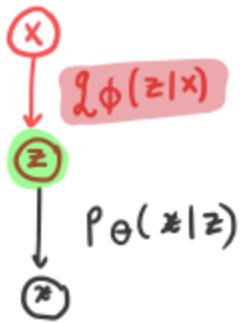


Optimizing $\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_{\phi}(z|x)} \left[\log \frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \right]$
 via Stochastic gradient Descent (SGD)

Goal : estimate $\nabla_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x_i)$

$$\begin{aligned} \nabla_{\phi} \mathcal{L}_{\theta, \phi}(x) &= \nabla_{\phi} \mathbb{E}_{q_{\phi}(z|x)} \left[\log p_{\theta}(x, z) - \log q_{\phi}(z|x) \right] \\ &\neq \mathbb{E}_{q_{\phi}(z|x)} \left[\nabla_{\phi} (\cdot) \right] \end{aligned}$$

Solution : reparametrization trick (if z is continuous)



Goal : estimate $\nabla_{\theta, \phi} \mathcal{L}_{\theta, \phi}(x_i)$

$$\nabla_{\phi} \mathcal{L}_{\theta, \phi}(x) = \nabla_{\phi} \mathbb{E}_{g_{\phi}(z|x)} [\log p_{\theta}(x, z) - \log g_{\phi}(z|x)]$$

Solution : reparametrization trick (if z is continuous)

write $z = g(\epsilon, \phi, x)$ s.t.

- $\textcircled{*}$ g is invertible
- $\textcircled{*}$ ϵ 'noise' $p(\epsilon)$ independent of x or ϕ

Then :

$$\begin{aligned} \nabla_{\phi} \mathbb{E}_{g_{\phi}(z|x)} [f(z)] &= \nabla_{\phi} \mathbb{E}_{p(\epsilon)} [f(z)] \\ &= \mathbb{E}_{p(\epsilon)} [\nabla_{\phi} f(z)] \leftarrow \text{Great, can use MC again} \end{aligned}$$

Common Setup



$$q_{\phi}(z|x)$$

with $q_{\phi}(z|x) = N(z, \mu, \text{diag}(\sigma^2))$

where

$$\mu, \log \sigma = \text{Neural Net}_{\phi}(x)$$

$$q_{\phi}(z|x) = \prod_i q_{\phi}(z_i|x)$$

$$= \prod_i N(z_i; \mu_i, \sigma_i^2)$$

& with reparametrization

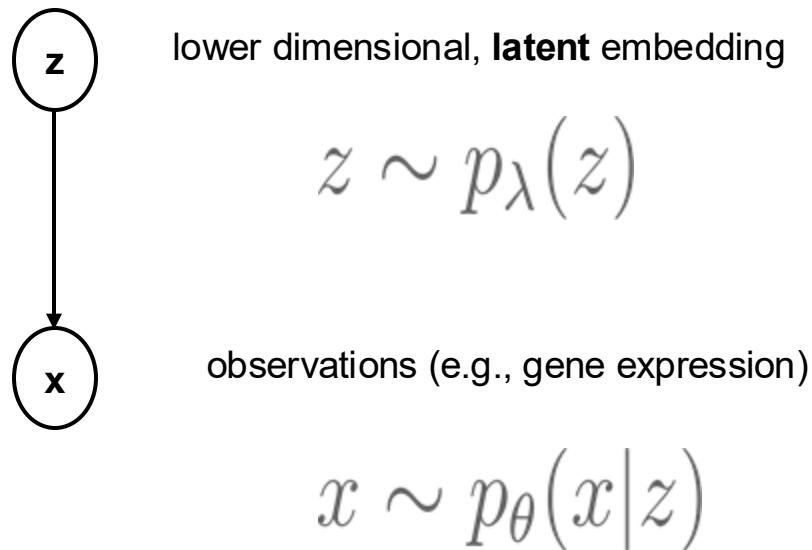
$$\varepsilon \sim N(0, I)$$

$$z = \mu + \sigma \odot \varepsilon$$

From VAE to scVI

* What is special about RNA-seq?

Generative modelling

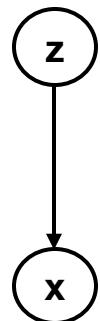


Examples (linear): probabilistic PCA, Factor Analysis

From VAE to scVI

* What is special about RNA-seq?

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

Incorporating some domain knowledge:

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

$$\tilde{x}|z \sim N(Az + \mu, W)$$

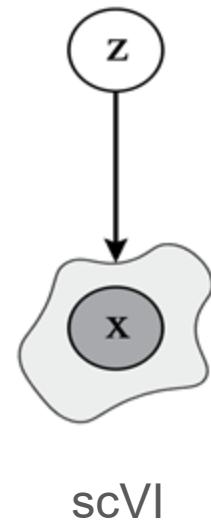
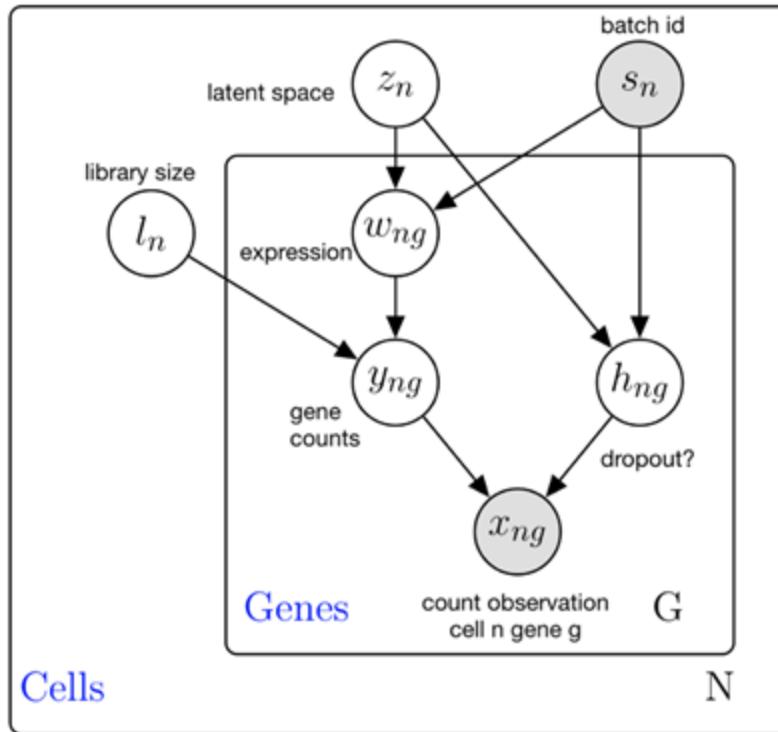
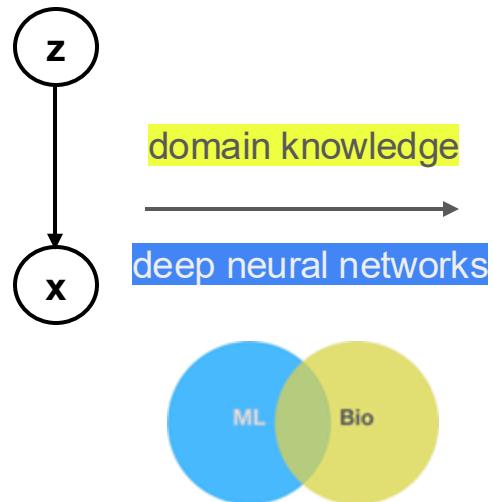
$$h|\tilde{x} \sim Bernoulli(f(\tilde{x}))$$

$$x|\tilde{x}, h \sim (1 - h)\tilde{x}$$

Motivation: a) single cell data is sparse; b) a linear map between latent z and x can aid interpretability

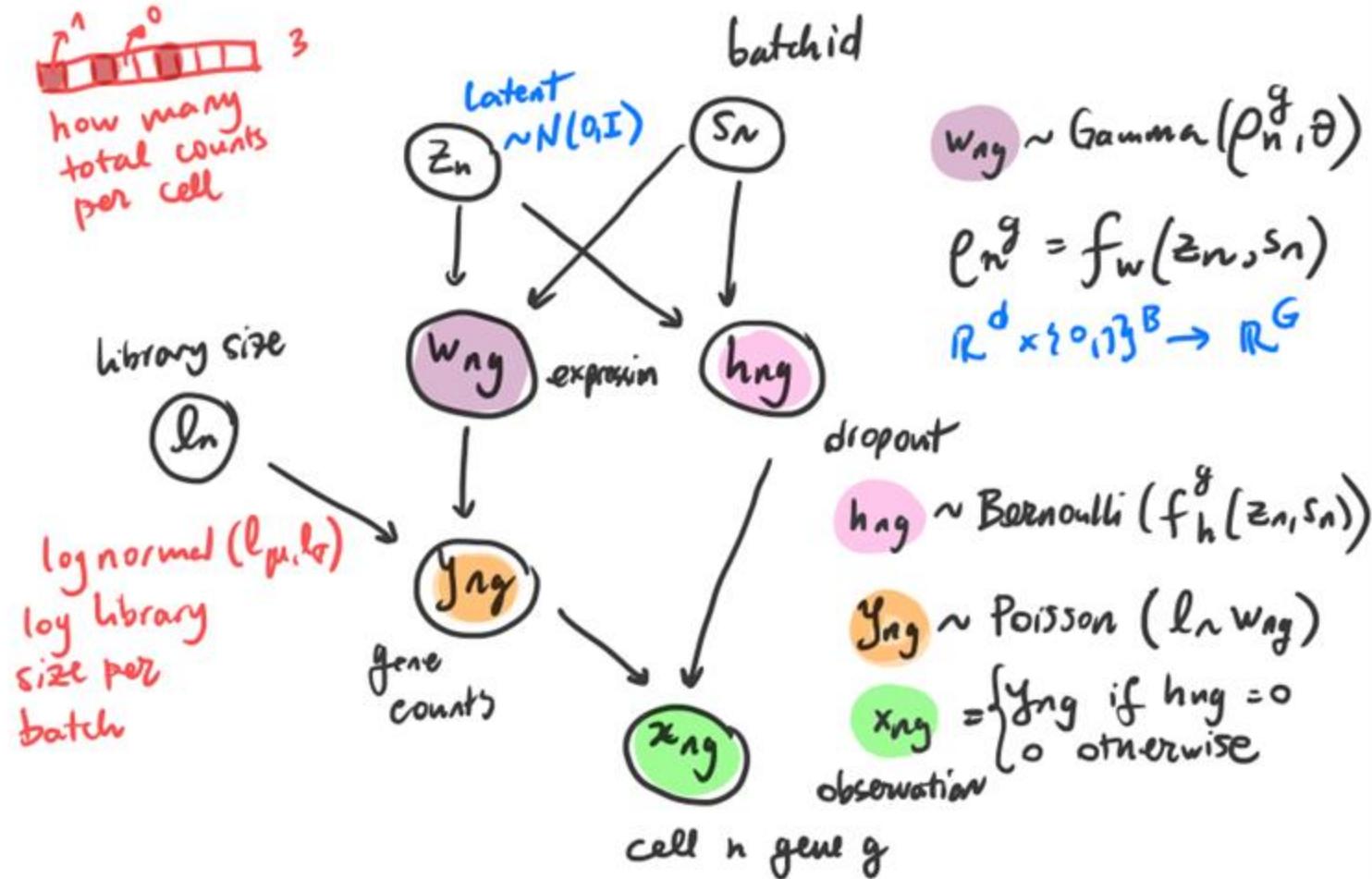
Pierson, E., & Yau, C. (2015). ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome biology*, 16, 1-10.

From VAE to scVI

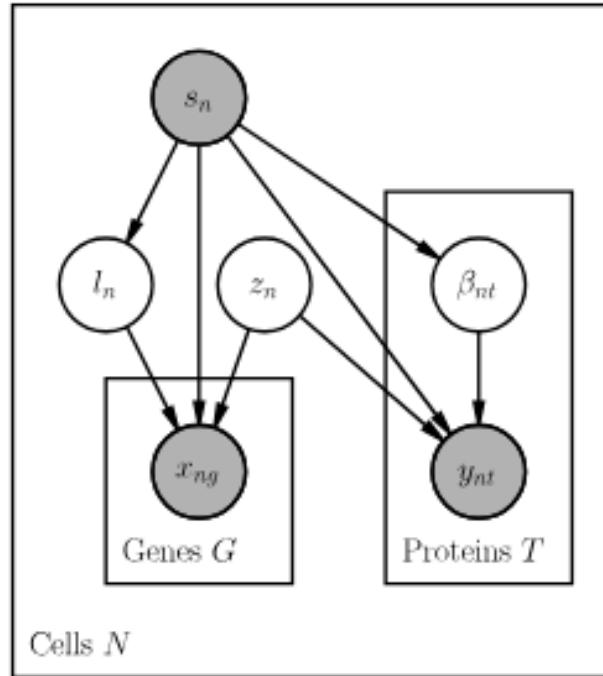


Lopez R., Regier J., Cole M.B. et al. 2018

From VAE to scVI



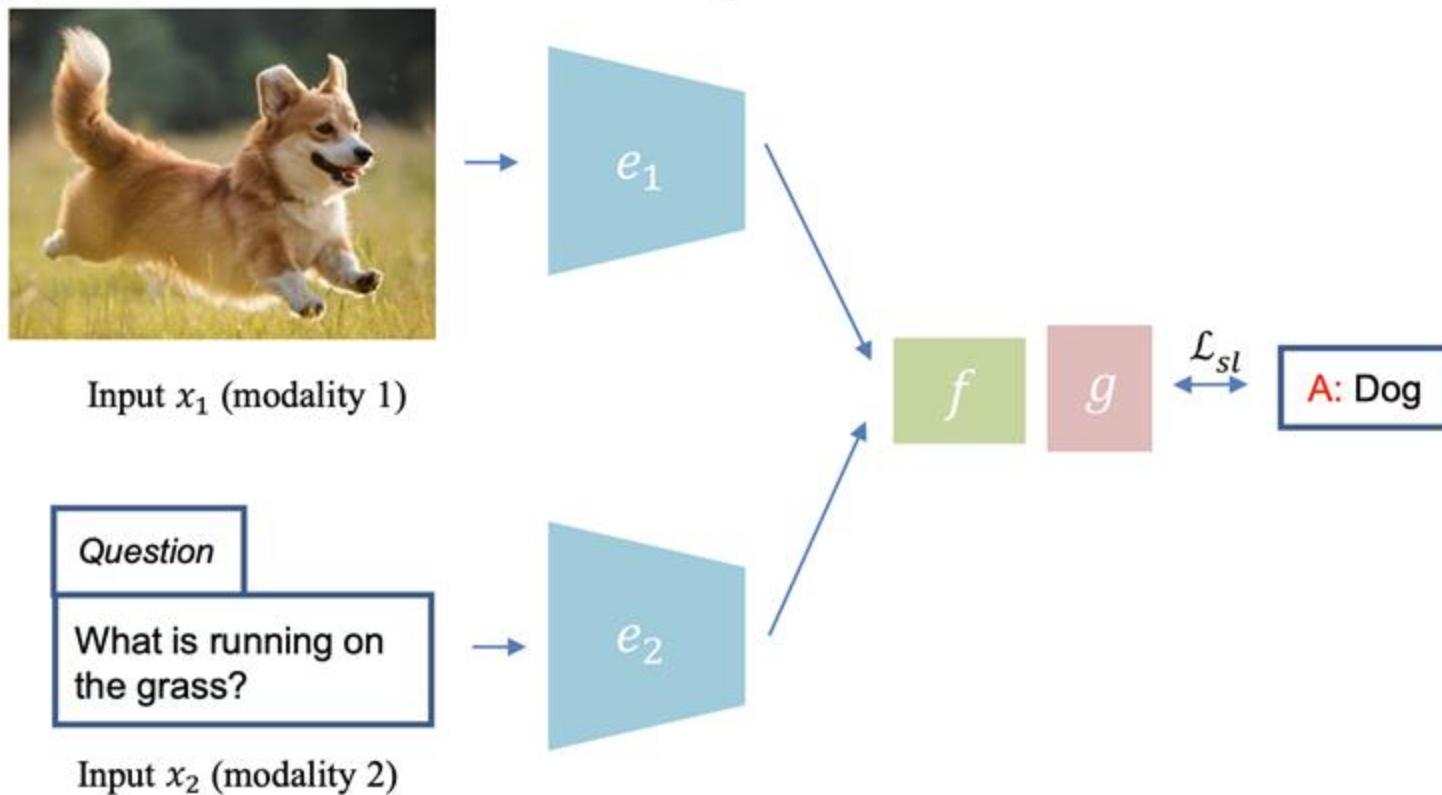
From scVI (uni-modal) to totalVI (multi-modal)



Gayoso*, Zoë Steier* et al (2021), Joint probabilistic modeling of single-cell multi-omic data with totalVI

Multi-modal data you say?

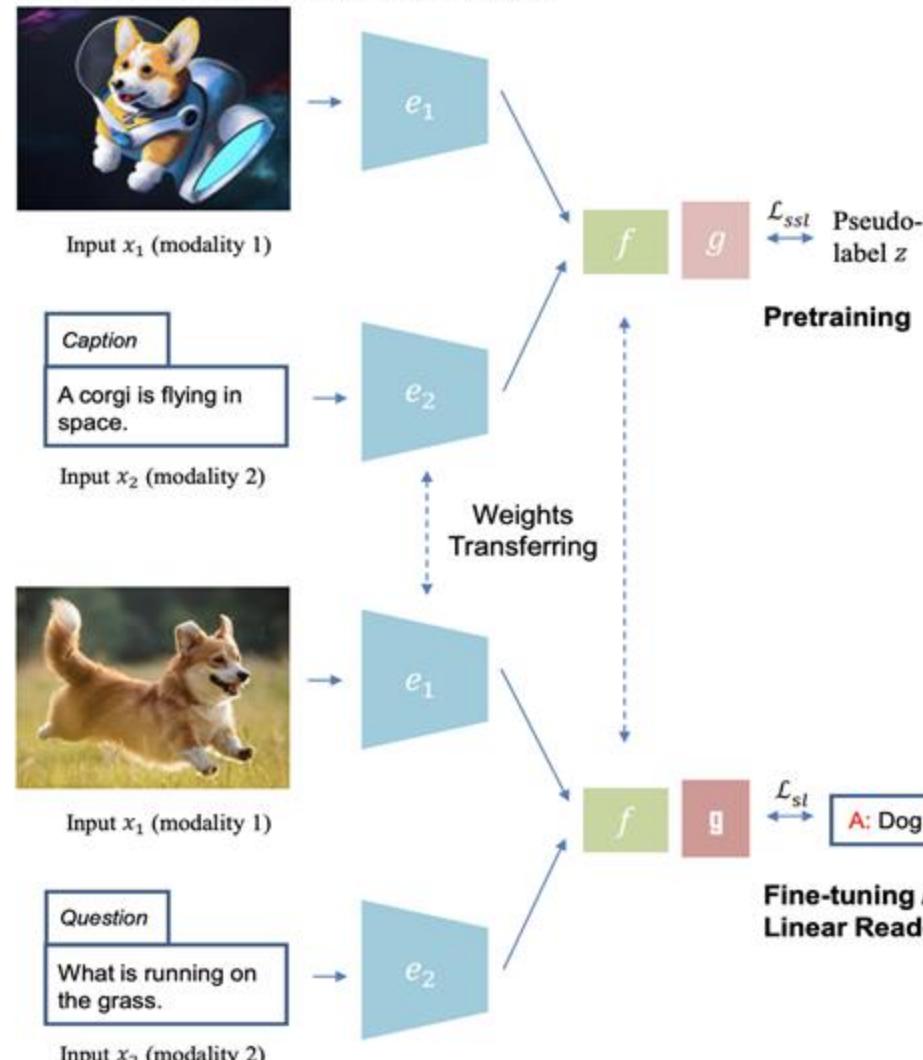
(a) Supervised Multimodal Learning



Zong et al, Self-Supervised Multimodal Learning: A Survey
Yongshuo(2024)

Multi-modal data you say?

(b) Self-Supervised Multimodal Learning



Zong et al, Self-Supervised Multimodal Learning:
A Survey Yongshuo(2024)

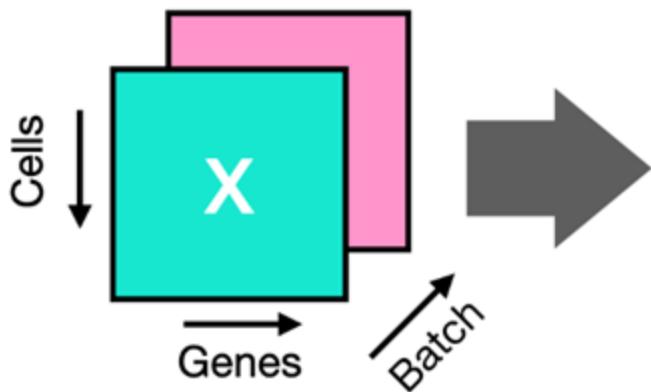
Bianca Dumitrascu, Machine Learning for Computational Biology, MLSS, 2025

scVI: flexible, specialized, performs well at cell typing

2

Nineteen generic and specialized self-supervised learning methods

Input Data



1

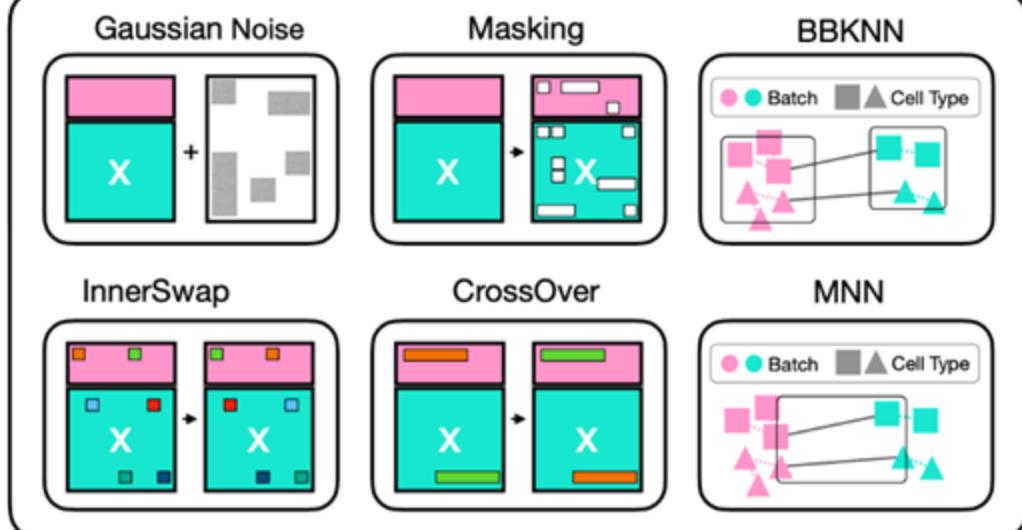
Nine single-cell genomics datasets
(seven uni-modal and two multi-modal)

Representation Learning

SSL Methods

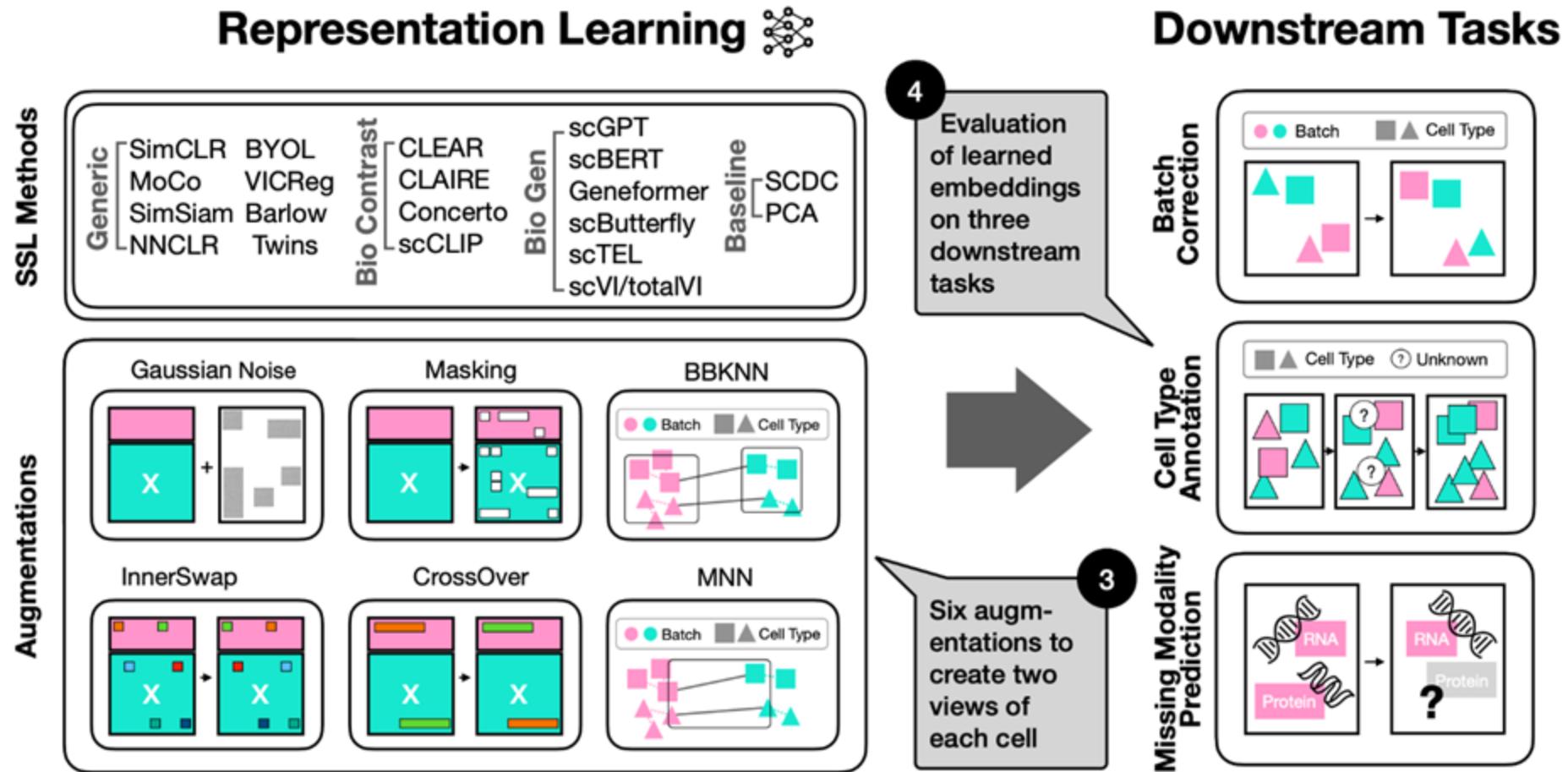
Generic	SimCLR	BYOL	Bio Contrast	CLEAR	Bio Gen	scGPT	Baseline
MoCo		VICReg		CLARE		scBERT	SCDC
SimSiam		Barlow		Concerto		Geneformer	PCA
NNCLR		Twins		scCLIP		scButterfly	
						scTEL	
						scVI/totalVI	

Augmentations



Ovcharenko et al, scSSL-Bench: Benchmarking Self-Supervised Learning for Single-Cell Data(2025)

scVI: flexible, specialized, performs well at cell typing



Ovcharenko et al, scSSL-Bench: Benchmarking Self-Supervised Learning for Single-Cell Data(2025)

scVI is okay at cell typing, but totalVI is not as good at integrating data from multiple-modalities

Method	Multi-Modal data						Single Modality Datasets								
	PBMC-M			BMMC			PBMC			Pancreas			Immune Cell Atlas		
	Bio	Batch	Total	Bio	Batch	Total	Bio	Batch	Total	Bio	Batch	Total	Bio	Batch	Total
scGPT (zero-shot)	—	—	—	—	—	—	0.440 ± 0.010	0.469 ± 0.017	0.451 ± 0.013	0.473 ± 0.001	0.168 ± 0.003	0.351 ± 0.002	0.380 ± 0.012	0.516 ± 0.014	0.435 ± 0.008
scGPT (finetuned)	—	—	—	—	—	—	0.940 ± 0.012	0.514 ± 0.011	0.770 ± 0.011	0.873 ± 0.003	0.345 ± 0.047	0.662 ± 0.021	0.979 ± 0.017	0.485 ± 0.019	0.781 ± 0.017
Geneformer (finetuned)	—	—	—	—	—	—	0.024 ± 0.000	0.462 ± 0.000	0.199 ± 0.000	0.004 ± 0.000	0.437 ± 0.000	0.177 ± 0.000	0.013 ± 0.000	0.265 ± 0.000	0.114 ± 0.000
scButterfly	0.702 ± 0.002	0.391 ± 0.003	0.577 ± 0.002	0.781 ± 0.000	0.297 ± 0.004	0.587 ± 0.002	—	—	—	—	—	—	—	—	—
scTEL	0.089 ± 0.002	0.800 ± 0.004	0.373 ± 0.001	0.000 ± 0.006	0.706 ± 0.005	0.282 ± 0.005	—	—	—	—	—	—	—	—	—
totalVI / scVI	0.702 ± 0.002	0.305 ± 0.002	0.543 ± 0.001	0.755 ± 0.002	0.272 ± 0.001	0.562 ± 0.002	0.918 ± 0.015	0.871 ± 0.000	0.899 ± 0.009	0.805 ± 0.002	0.511 ± 0.007	0.688 ± 0.001	0.862 ± 0.033	0.593 ± 0.013	0.754 ± 0.024

Ovcharenko et al, scSSL-Bench: Benchmarking Self-Supervised Learning for Single-Cell Data(2025)

scVI is okay at cell typing, but totalVI is not as good at integrating data from multiple-modalities

Inductive Bias Methods

Method	PBMC-M			BMMC		
	Bio	Batch	Total	Bio	Batch	Total
scGPT (zero-shot)	—	—	—	—	—	—
scGPT (finetuned)	—	—	—	—	—	—
Geneformer (finetuned)	—	—	—	—	—	—
scButterfly	0.702 ± 0.002	0.391 ± 0.003	0.577 ± 0.002	0.781 ± 0.000	0.297 ± 0.004	0.587 ± 0.002
scTEL	0.089 ± 0.002	0.800 ± 0.004	0.373 ± 0.001	0.000 ± 0.006	0.706 ± 0.005	0.282 ± 0.005
totalVI / scVI	0.702 ± 0.002	0.305 ± 0.002	0.543 ± 0.001	0.755 ± 0.002	0.272 ± 0.001	0.562 ± 0.002

Generic SSL Methods

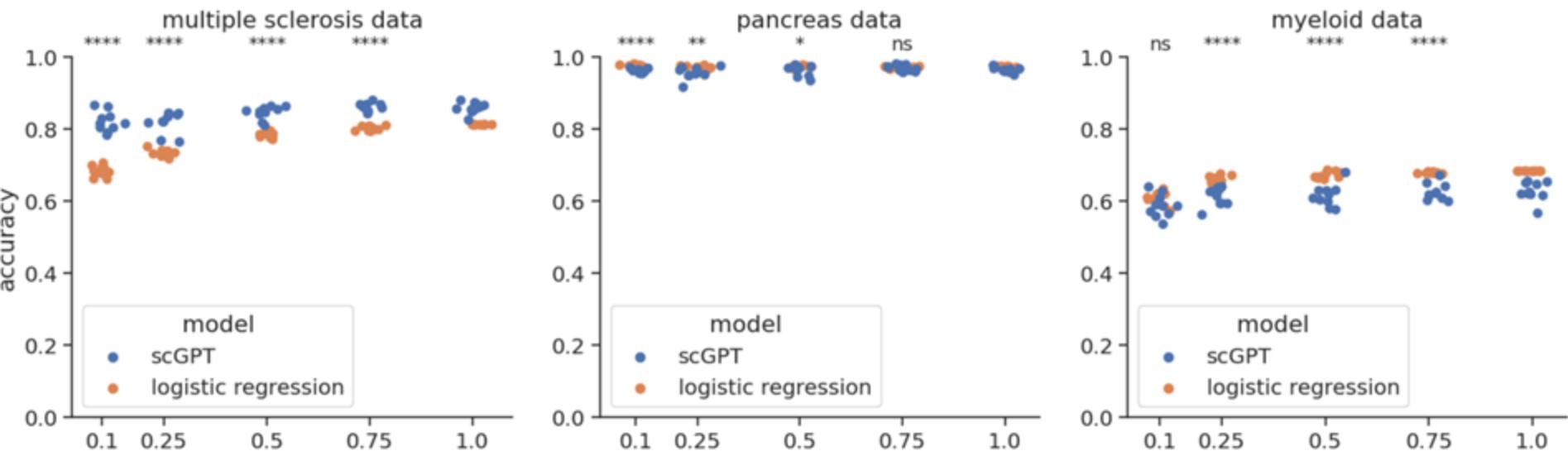
Method	PBMC-M			BMMC		
	Bio	Batch	Total	Bio	Batch	Total
SimCLR	0.877 ± 0.020	0.434 ± 0.001	0.700 ± 0.012	0.877 ± 0.025	0.601 ± 0.002	0.767 ± 0.016
MoCo	0.786 ± 0.005	0.581 ± 0.016	0.704 ± 0.003	0.647 ± 0.048	0.819 ± 0.024	0.716 ± 0.038
SimSiam	0.903 ± 0.057	0.455 ± 0.029	0.724 ± 0.046	0.753 ± 0.007	0.571 ± 0.002	0.680 ± 0.005
NNCLR	0.877 ± 0.033	0.534 ± 0.004	0.740 ± 0.018	0.819 ± 0.021	0.580 ± 0.008	0.723 ± 0.016
BYOL	0.928 ± 0.065	0.493 ± 0.016	0.754 ± 0.033	0.742 ± 0.043	0.693 ± 0.016	0.722 ± 0.019
VICReg	0.814 ± 0.039	0.405 ± 0.026	0.651 ± 0.013	0.832 ± 0.051	0.656 ± 0.009	0.761 ± 0.027
Barlow Twins	0.902 ± 0.048	0.430 ± 0.014	0.713 ± 0.034	0.859 ± 0.018	0.612 ± 0.011	0.760 ± 0.006

“For uni-modal data (PBMC, Pancreas, and Immune Cell Atlas), the specialized encoder-decoder method scVI, the domain-specific SSL method CLAIRE, and a foundation model scGPT outperform other methods. For the multi-modal datasets PBMC-M and BMMC, generic methods achieve higher scores.”

Ovcharenko et al, scSSL-Bench: Benchmarking Self-Supervised Learning for Single-Cell Data(2025)

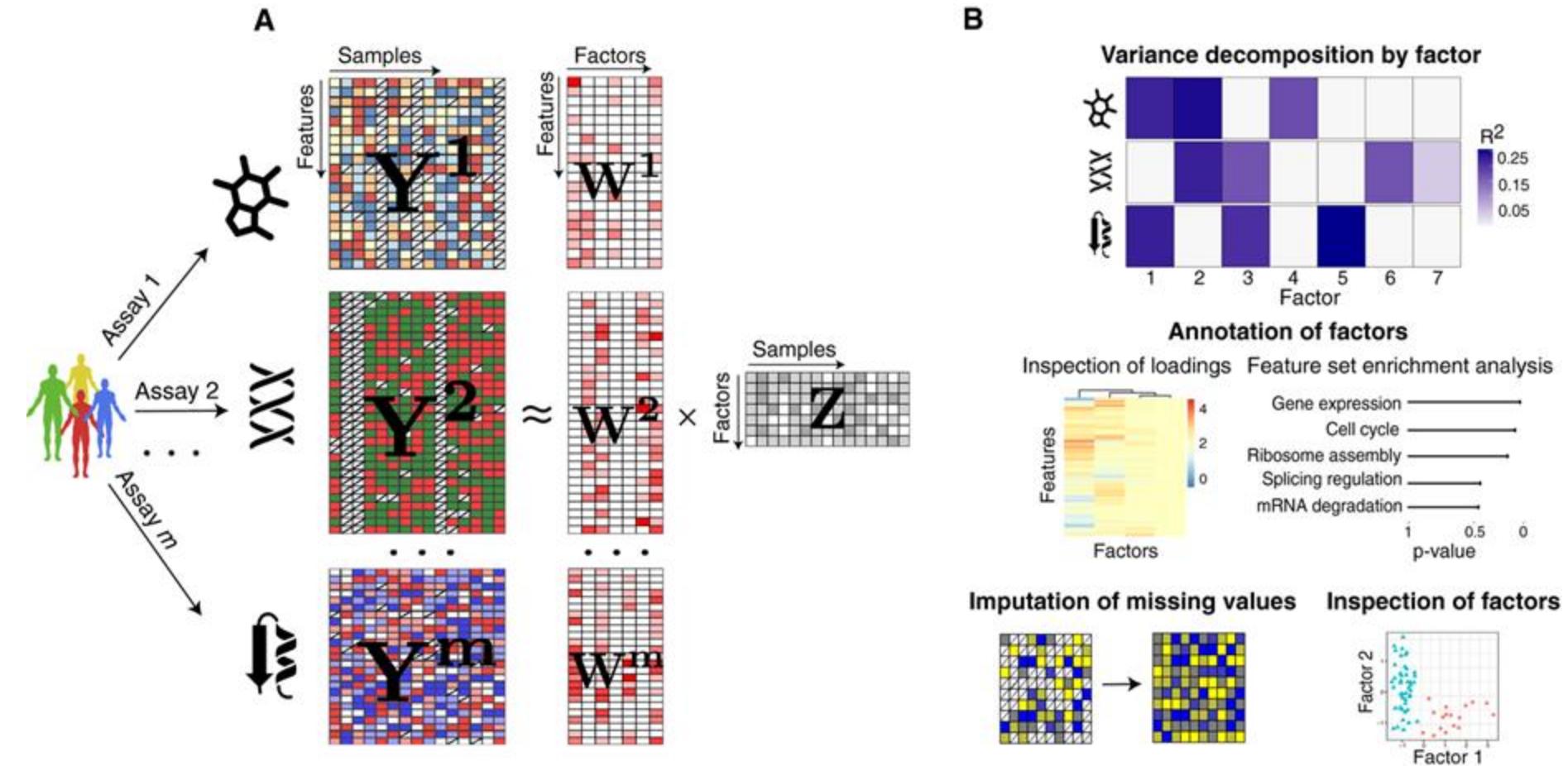
Bianca Dumitrascu, Machine Learning for Computational Biology, MLSS, 2025

Benchmarking studies have started showing simpler baselines too



From: A Deep Dive into Single-Cell RNA Sequencing Foundation Models (Boiarsky et al, 2024)

Benchmarking studies have started showing simpler baselines too, but some are still missing



Benchmarking studies have started showing simpler baselines too, but some are still missing

MoFA +

Argelaguet et al. 2020

: a multi-study
variant

Input: a matrix for each group or data modality

Factor
Matrix

(G)

Dg

sample #

(M)

Dm

feature #

Model: $Y_{gm} = Z_g^T W_m + \epsilon_{gm}$

$N_g \times K \quad K \times D_m$

Argelaguet, Velten et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018

Benchmarking studies have started showing simpler baselines too, but some are still missing

MoFA +

Argelaguet et al. 2020

: a multi-study
variant

Input: a matrix for each group or data modality

Factor Matrix

(G)

D_g
sample #

(M)

D_m
feature #

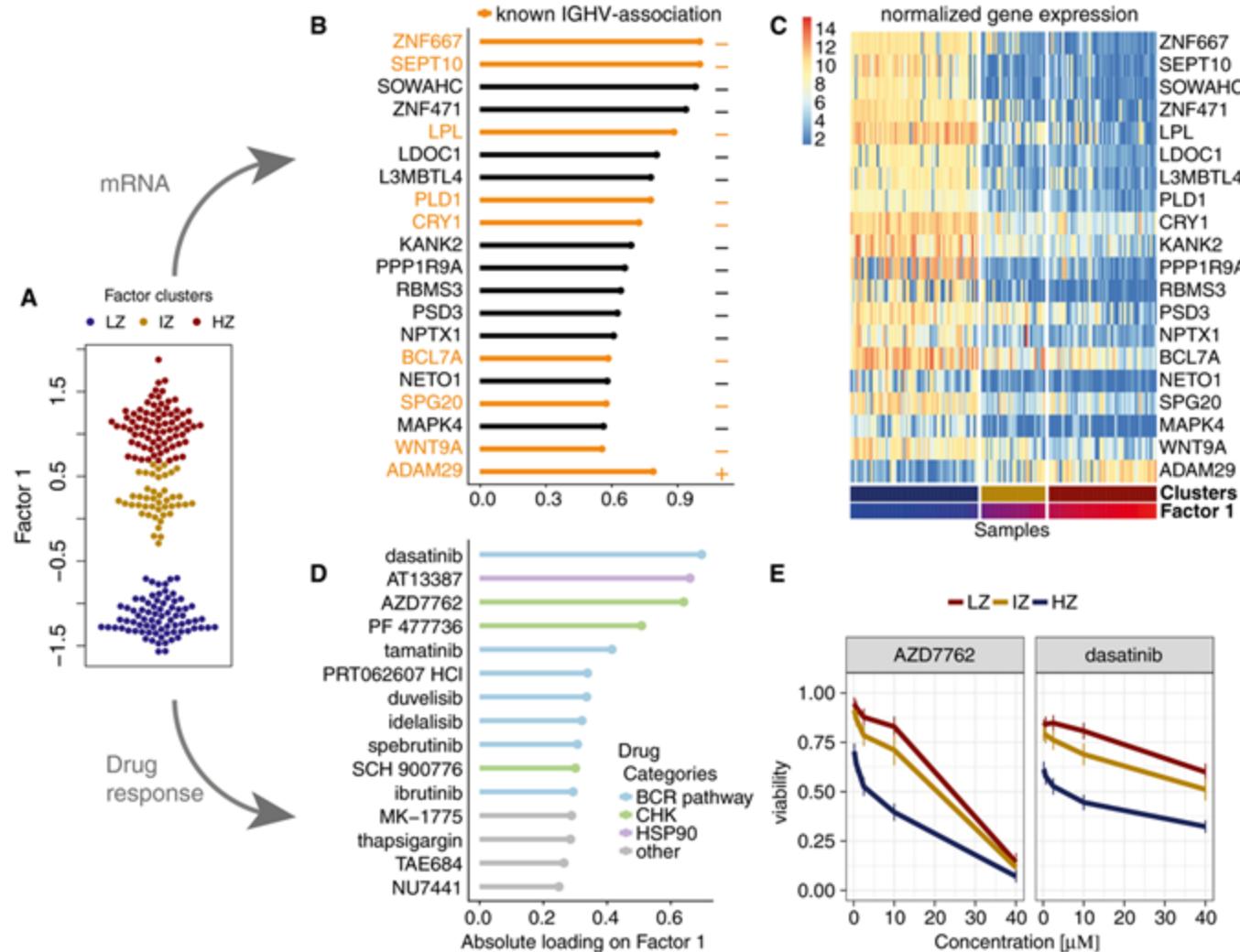
Model: $Y_{gm} = Z_g^T W_m + \epsilon_{gm}$

N_g × K K × D_m

Also $Y_{gm} = f(Z_g^T W_m + \epsilon_{gm})$
s.t. Y_{gm} : Gaussian, Bernoulli,
Poisson

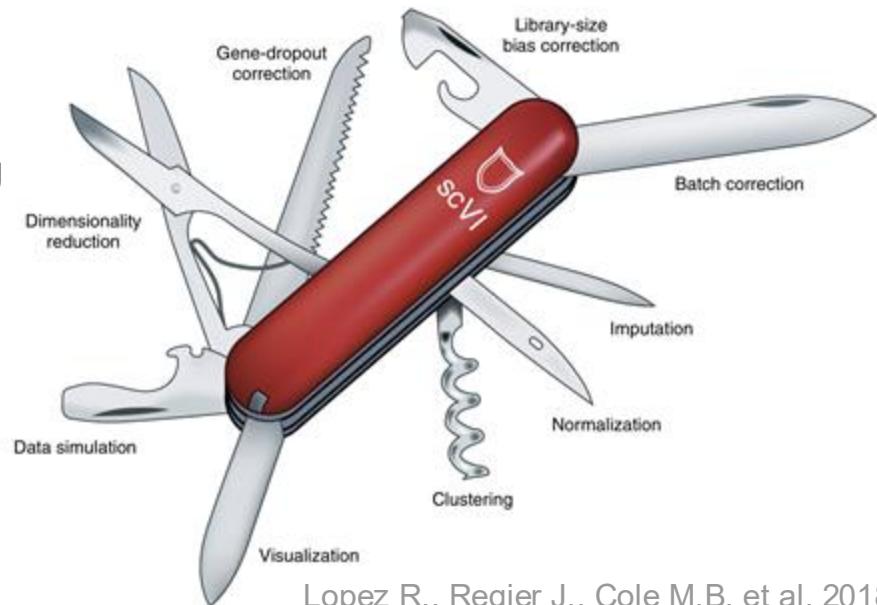
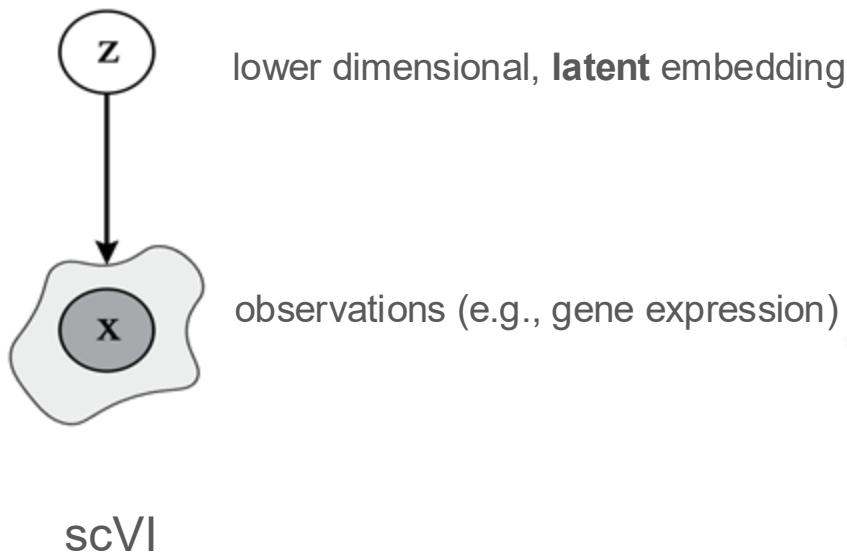
Argelaguet, Velten et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018

Benchmarking studies have started showing simpler baselines too, but some are still missing



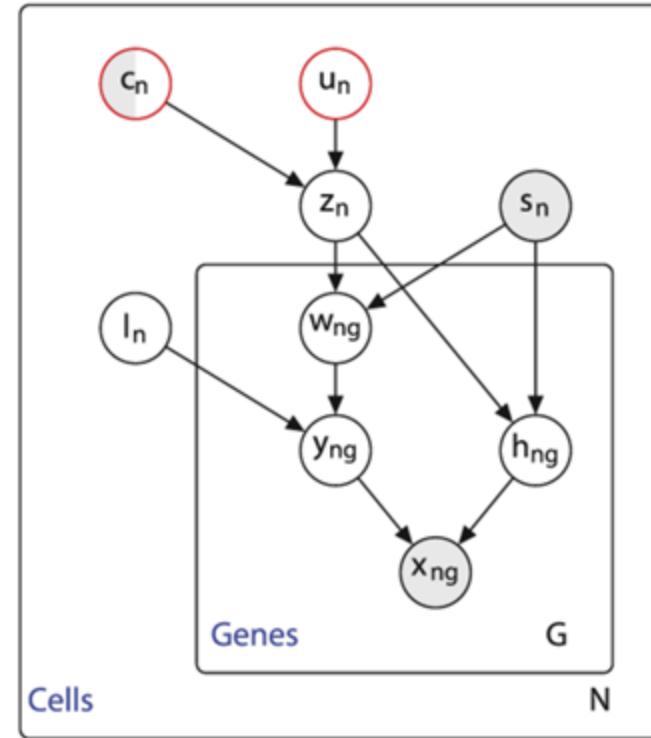
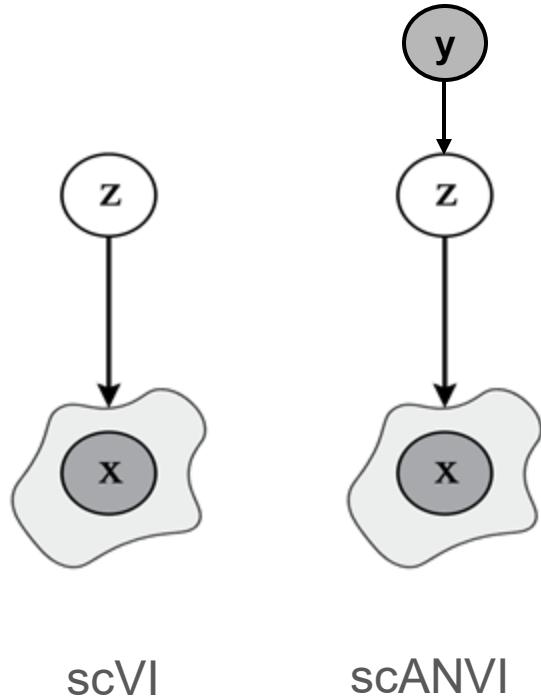
Argelaguet, Velten et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018

VAE/ scVI a scaffold for many tasks



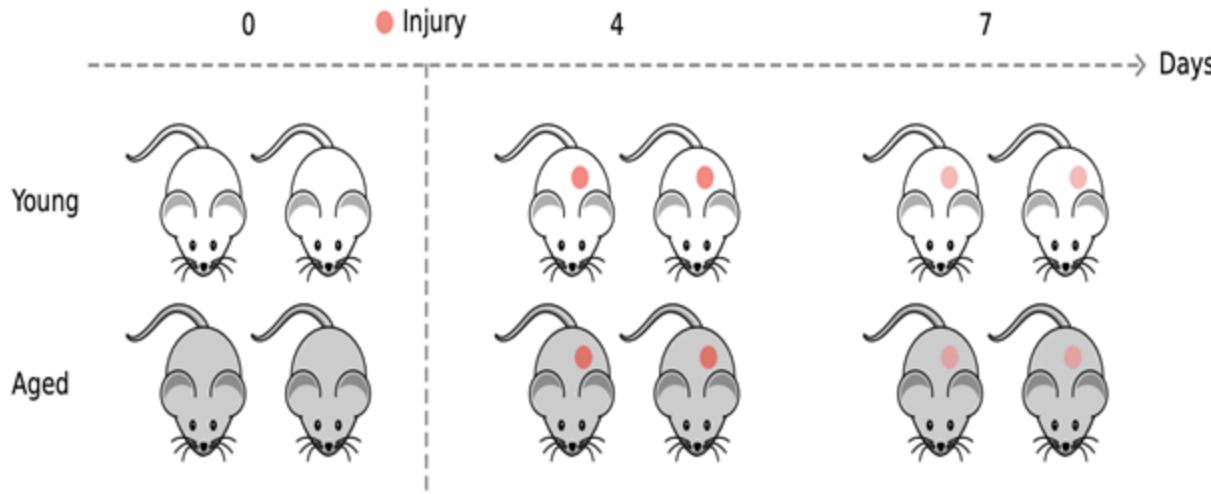
Lopez R., Regier J., Cole M.B. et al. 2018
Grønbech et al. 2020
Pierson E. & Yau C. 2015
Risso D. et al. 2018

How well does scVI deal with conditional data?



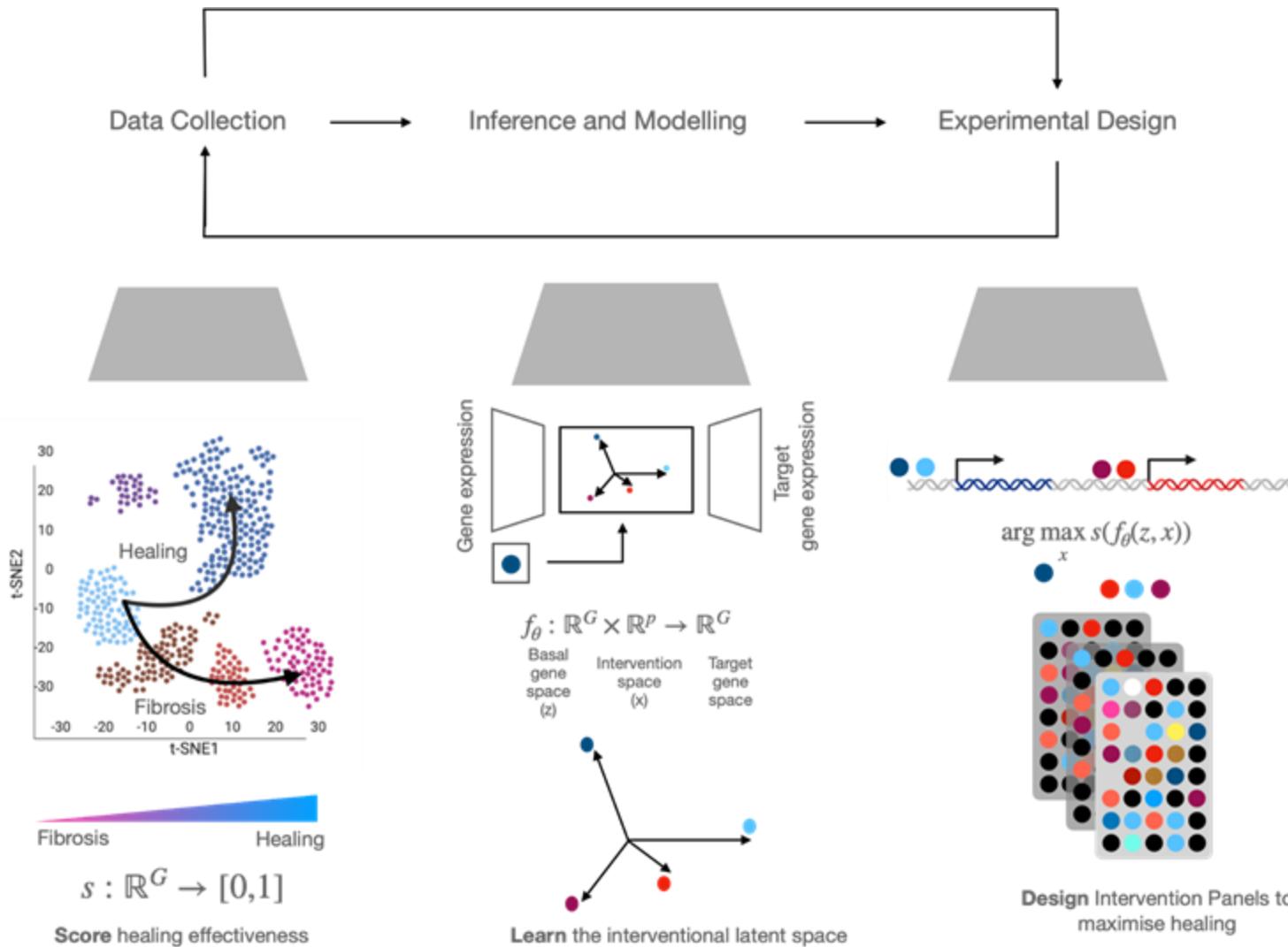
scANVI plate diagram, shaded elements are observed

Example conditional data & challenges

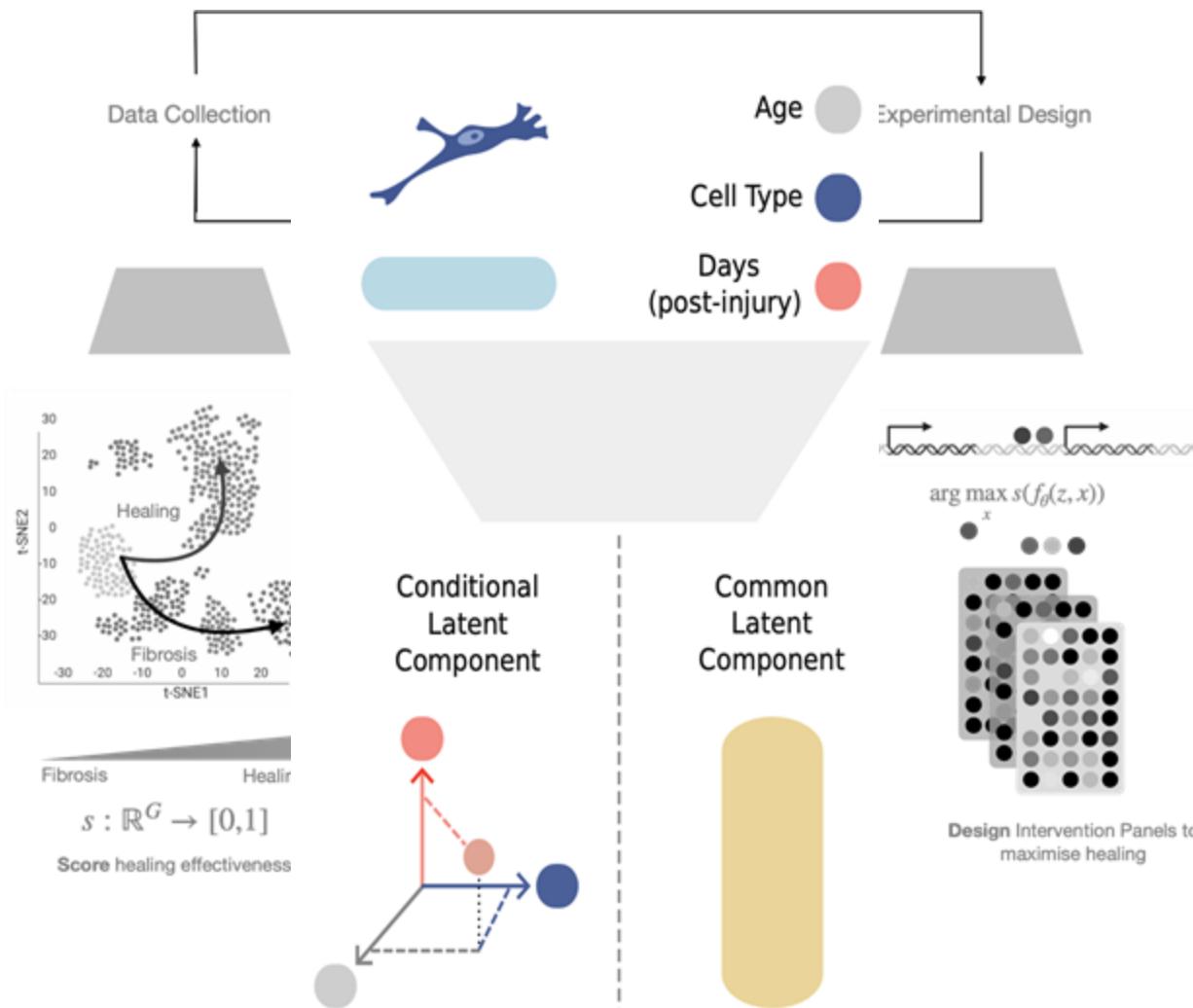


- Experimental challenges:
 - Difficult to observe every phase of healing in real-time.
 - Destructive sampling limits longitudinal insights.
- Computational challenges:
 - **Disentangle** condition **specific** signals from signals **common** across conditions! Aging and disease conditions add complexity, but how different can we expect the transcriptomics signals to be during healing?
 - **Interpret** the latent space by connecting with transcriptomic wide observations!

An idealized lab-in-the-loop workflow for accelerating wound healing



A first goal: representations across multiple attributes



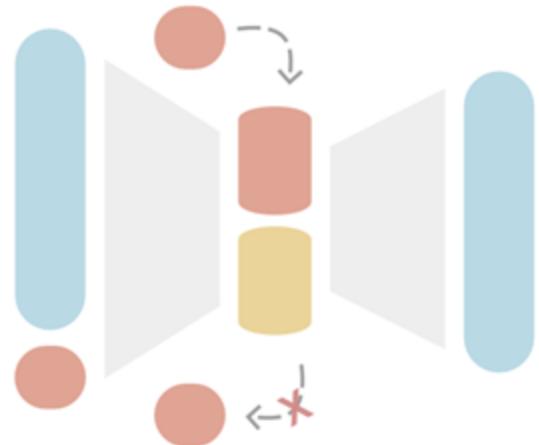
Joint and specific representations across multiple attributes



scVI



scANVI

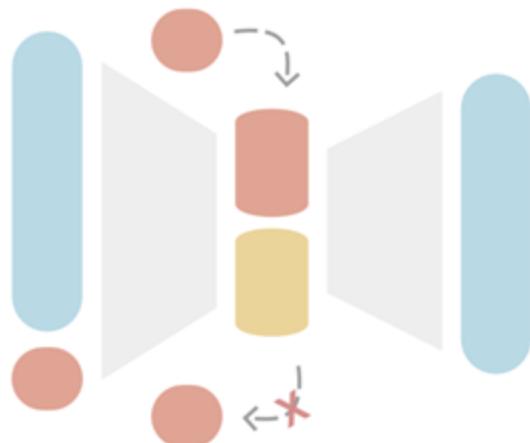


Patches*

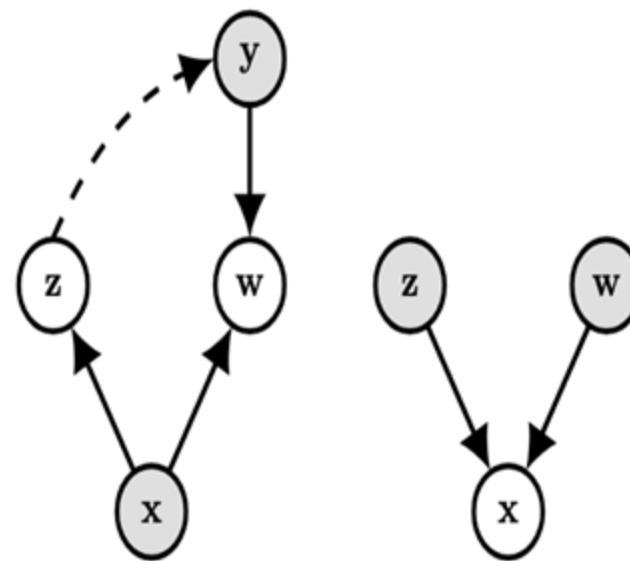
Bianca Dumitrascu, M

*Beker, O et al. Patches: A Representation Learning framework for Wound Healing (2024)

Joint and specific representations across multiple attributes

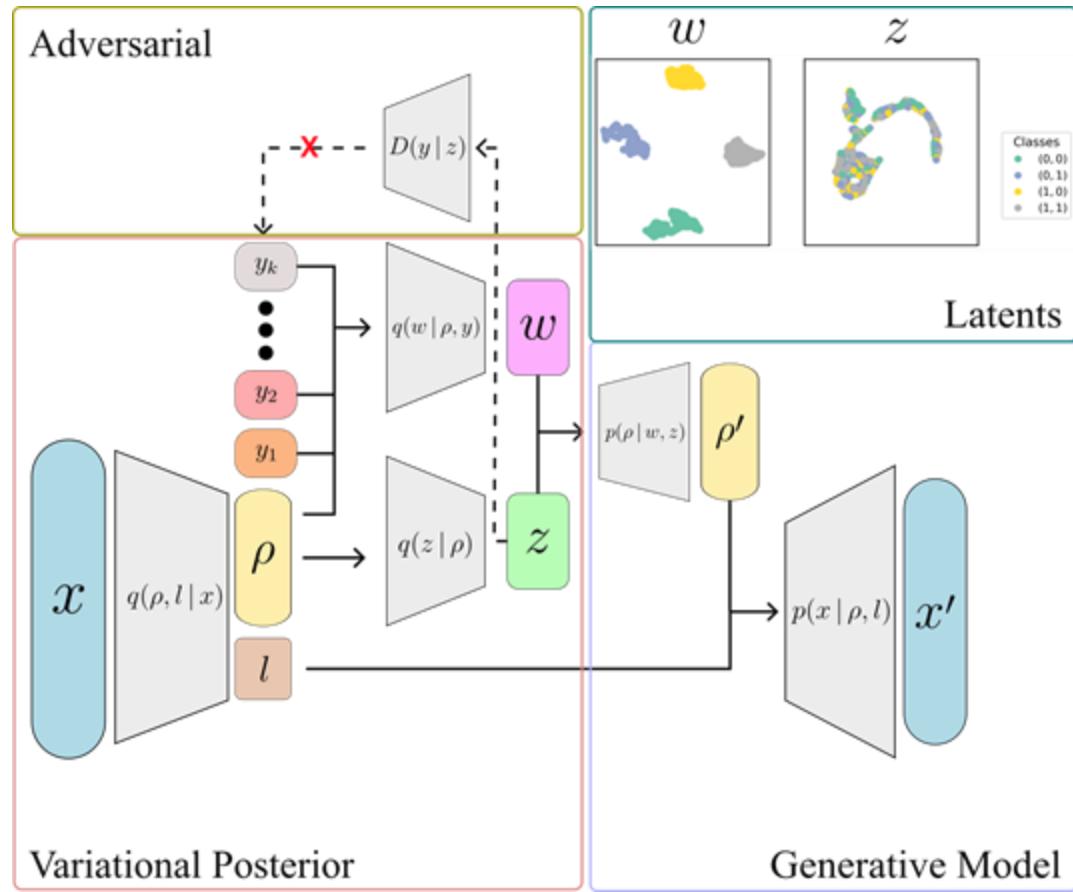
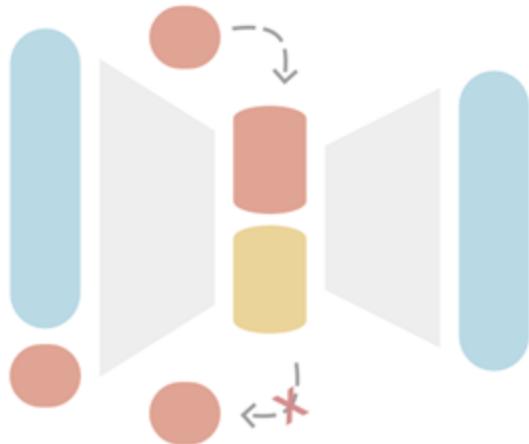


Patches simplified



Latent space of Common Space VAE

Joint and specific representations across multiple attributes



Patches simplified

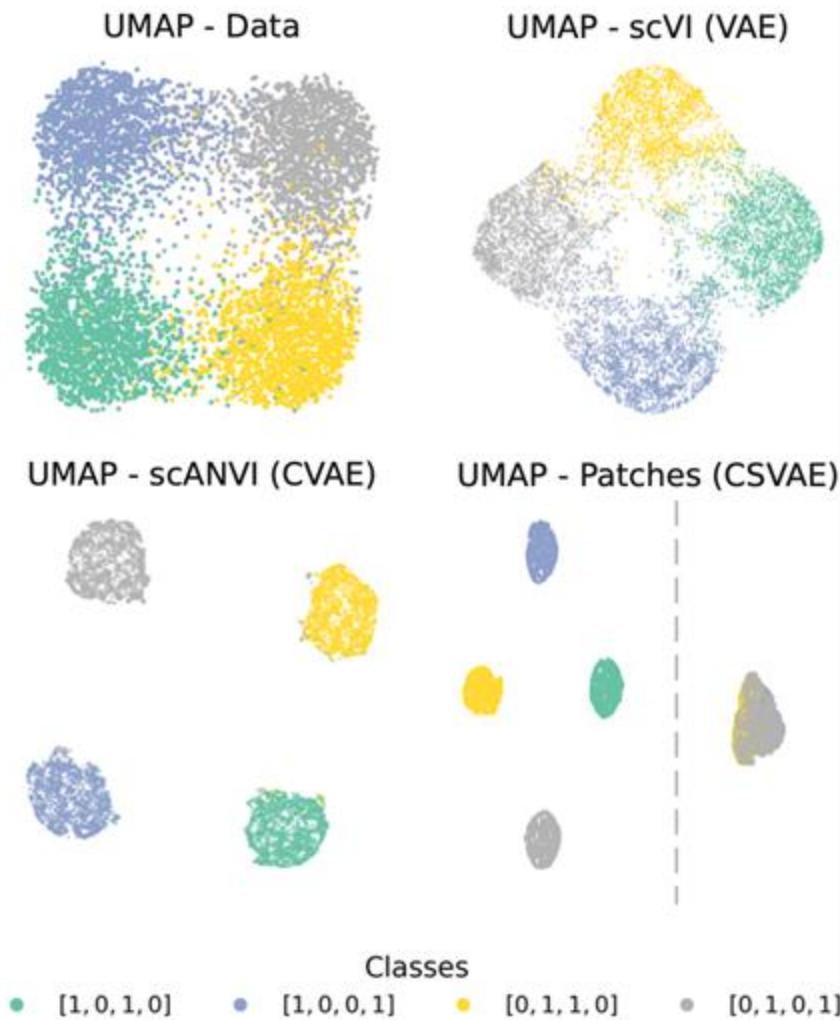
Patches in detail

Synthetic Example 1

Patches learns **shared** and **distinct** representations in a simple, linear model

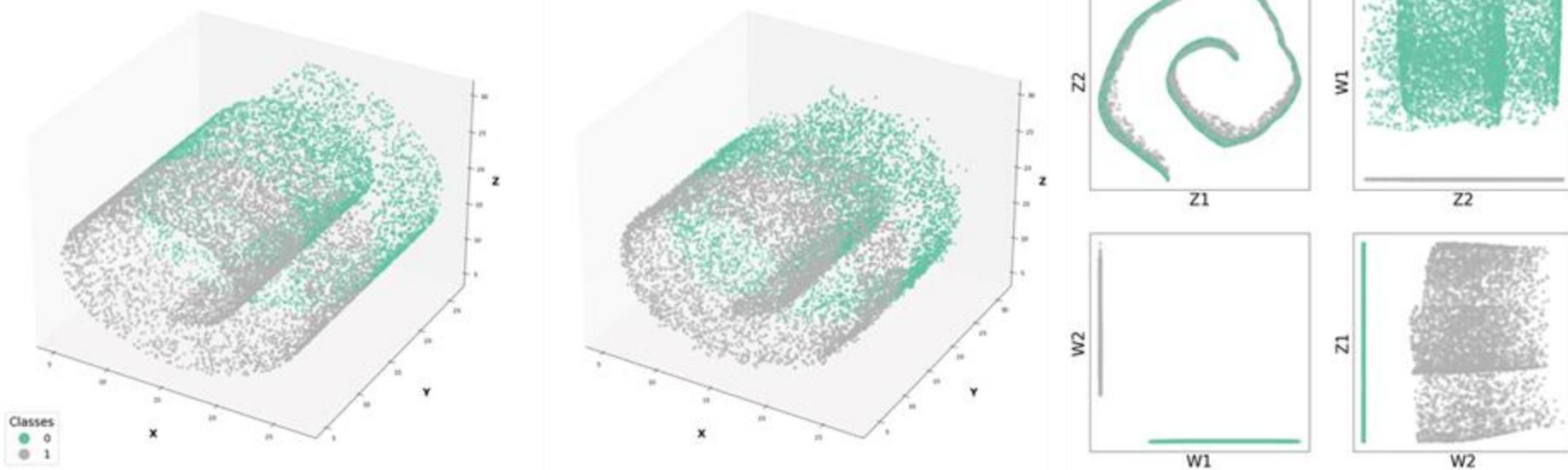
Generative model:

- $z \sim \text{Multivariate Normal}(0, I)$
- $w \sim \text{Mixture of Gaussians}$ with distinct means corresponding to one of 4 labels
- a linear identity decoder



Synthetic Example 2: Swiss Roll

Patches disentangled representations in nonlinear data

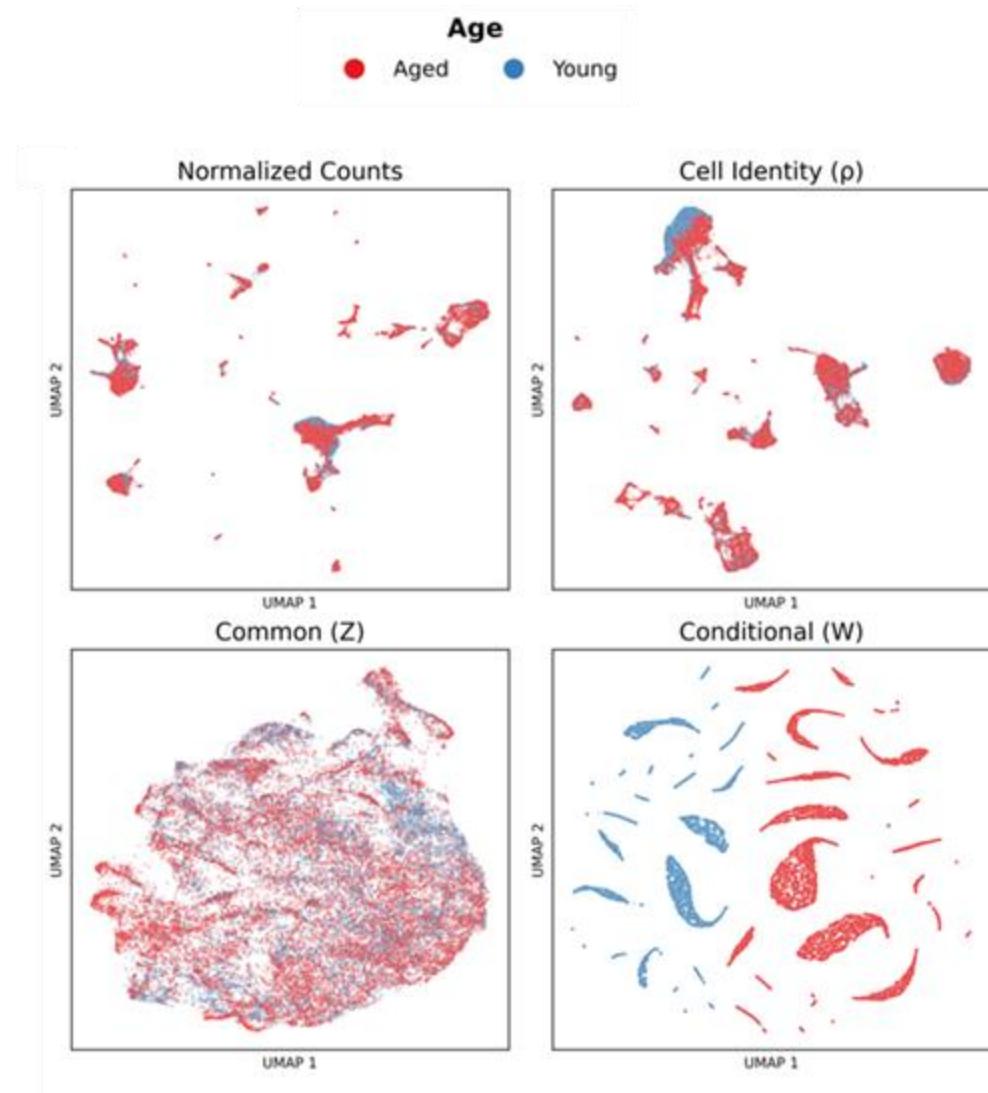


Wound Healing Example

Patches learns **age** specific representations

Data: Vu et al. (2022), young (7-wk) and old (88-wk) mice across 3 timepoints (number of cells, n = 26,966)

Provided attributes: Number of days post wounding (time), age & cell type

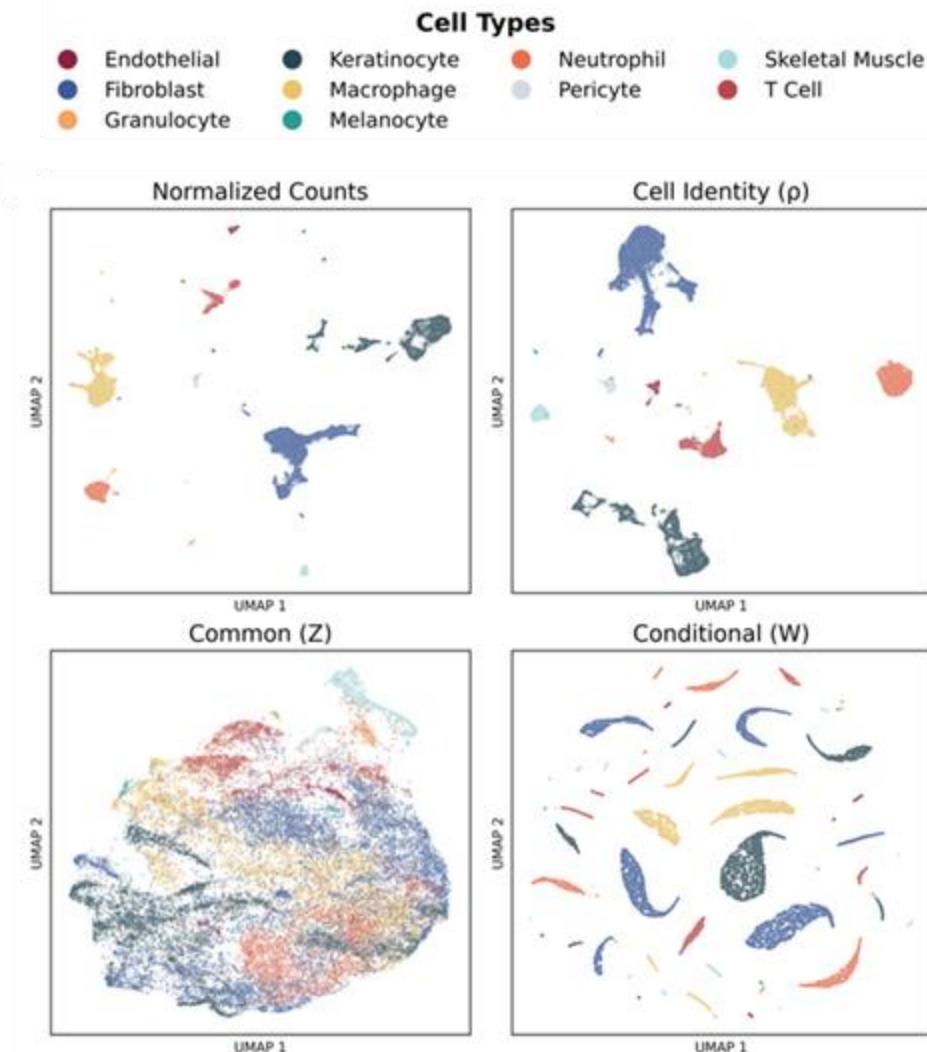


Wound Healing Example

Patches learns **cell type** specific representations in wound healing data

Data: Vu et al. (2022), young (7-wk) and old (88-wk) mice across 3 timepoints (number of cells, n = 26,966)

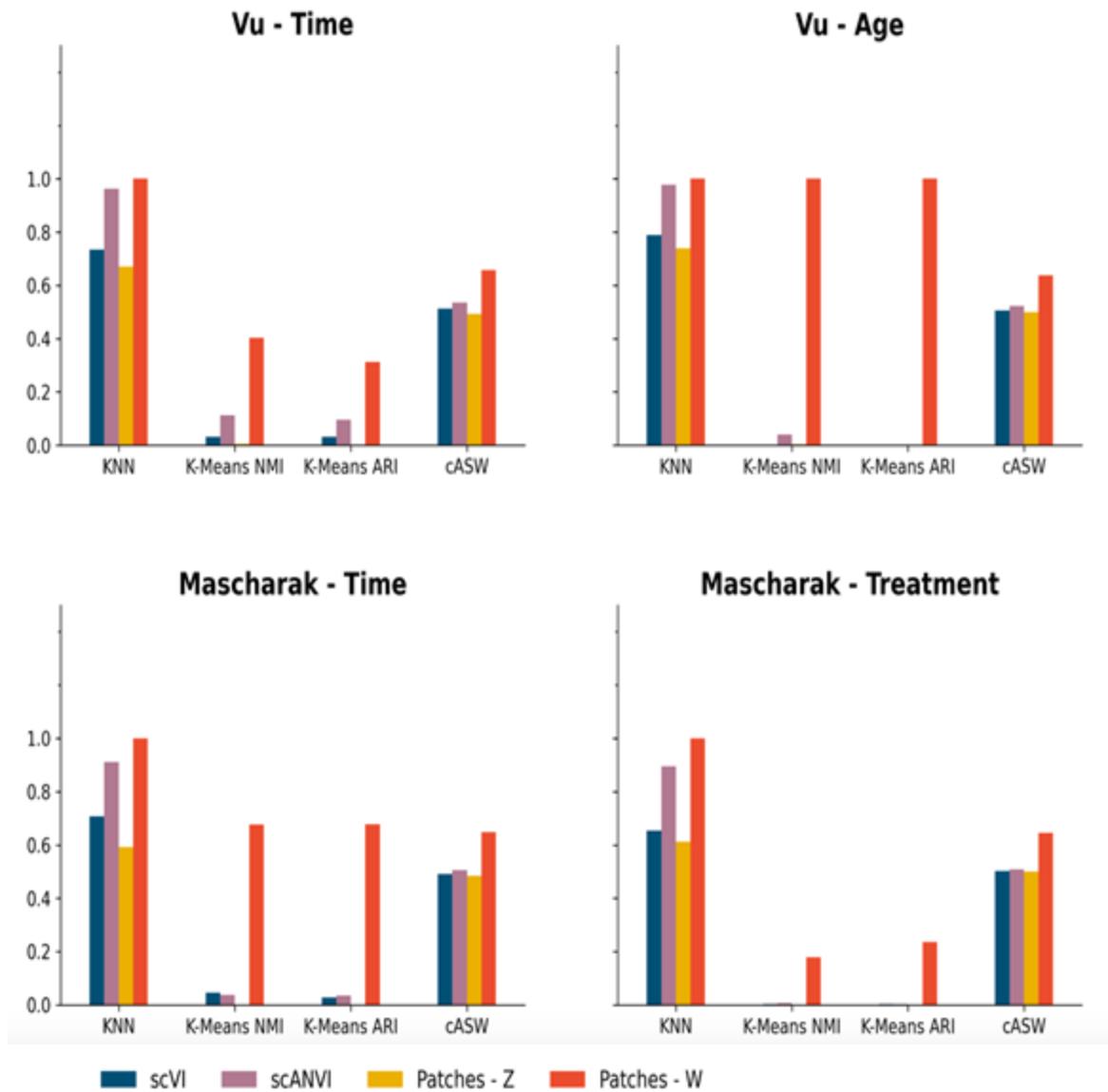
Provided attributes: Number of days post wounding (time), age & cell type



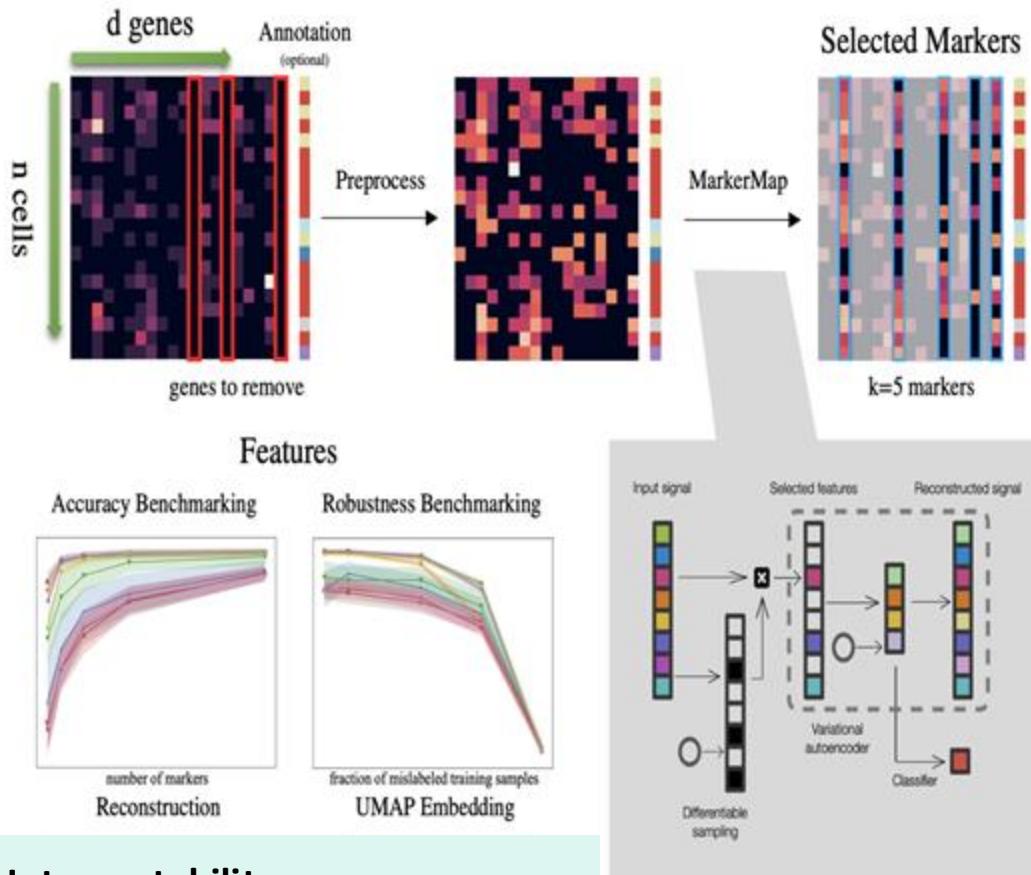
Benchmarking of multi-condition learning

Latent space
separability scores

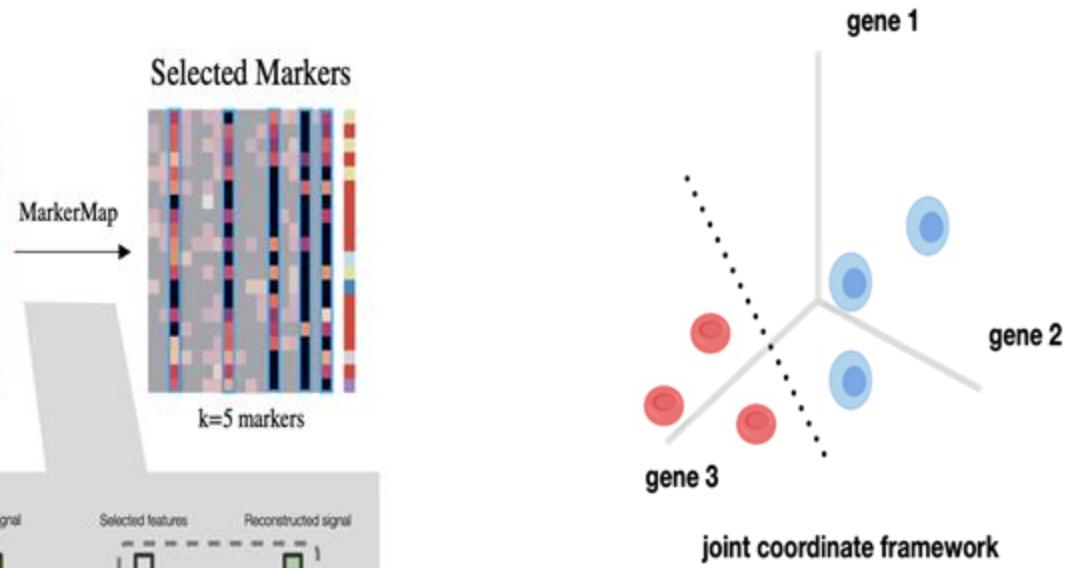
Aged
vs Young



Can the representations lead to biological interpretation?

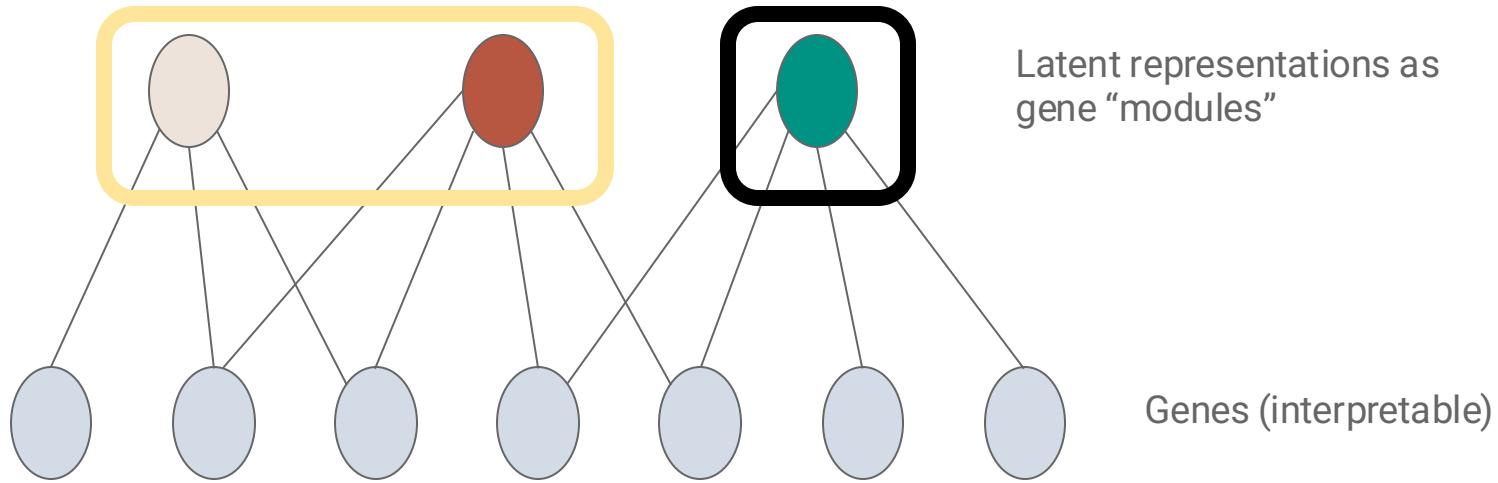


Interpretability:
which **parts of the input**, maximize or
define a quantity of interest?



Gregory, Wilson, et al. "MarkerMap: nonlinear marker selection for single-cell studies." *npj Systems Biology and Applications* 10.1 (2024): 17.
Bianca Dumitrescu, Machine Learning for Computational Biology, MLSS, 2025

Can the representations lead to biological interpretation?



Interpretable input:

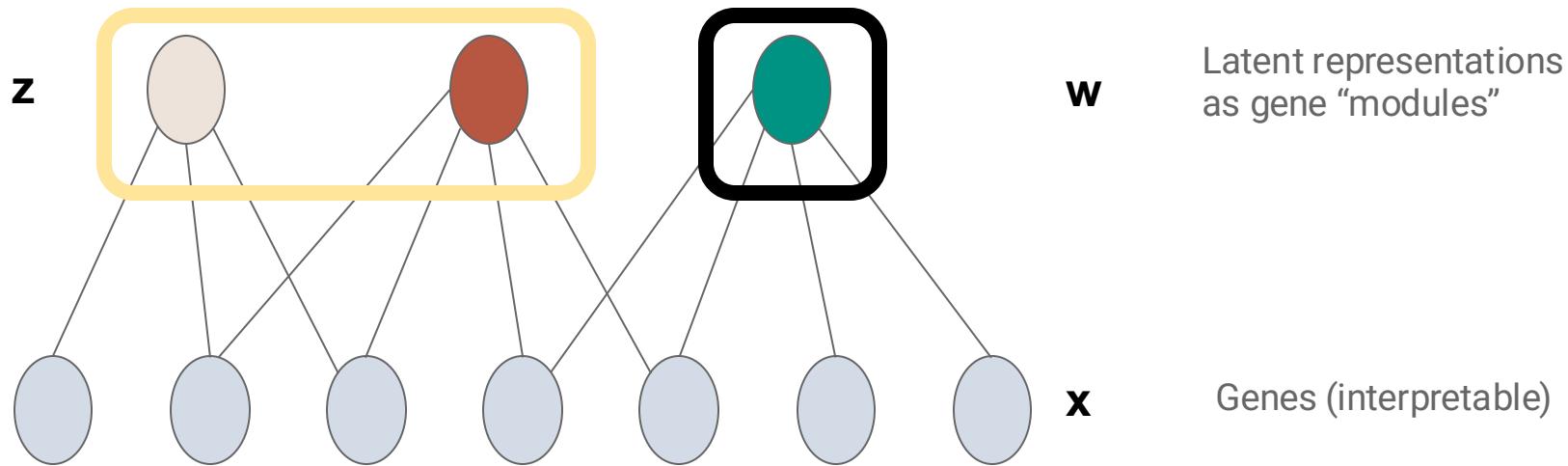
which **parts of the input**, maximize or define a quantity of interest?

vs

Interpretable latent space:

Can we relate a learnt latent representation to bits that are already interpretable?

Can the representations lead to biological interpretation?



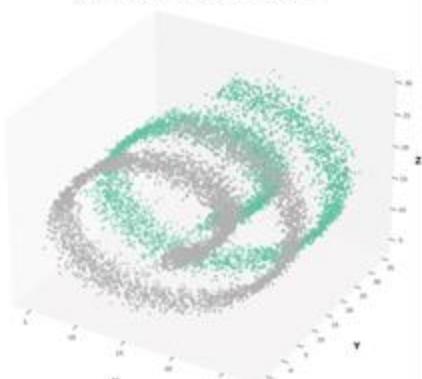
$$\text{Decoder}(x|z, w) = \lambda_z z + \lambda_w w$$

Interpretable latent space:

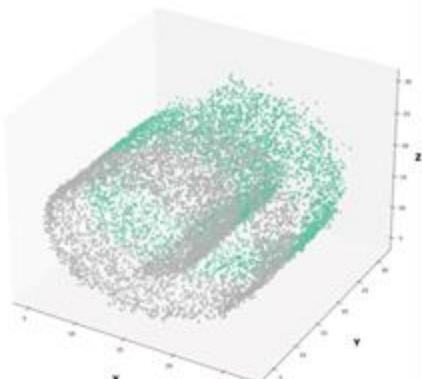
Can we relate a learnt latent representation to bits that are already interpretable?

Linear-decoder Patches identifies axis of variation

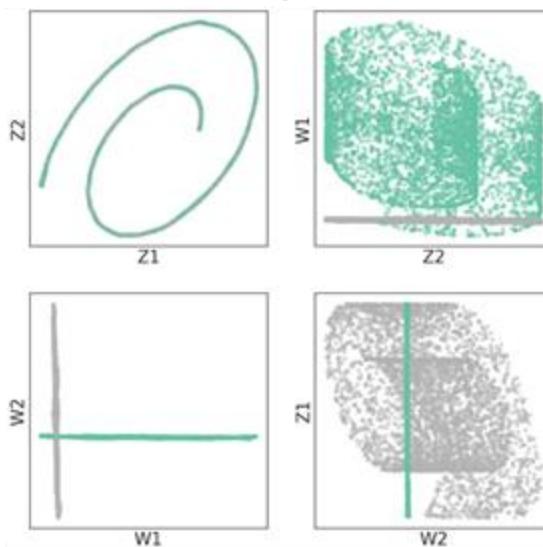
Swiss Roll - Linear Decoder



Swiss Roll - Standard Decoder



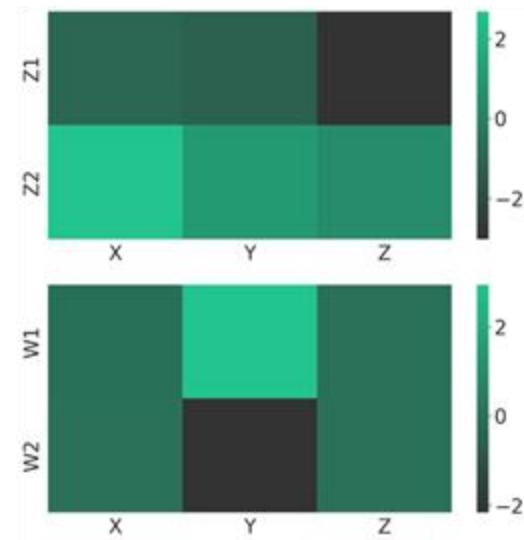
Latent Representations



pro: interpretability

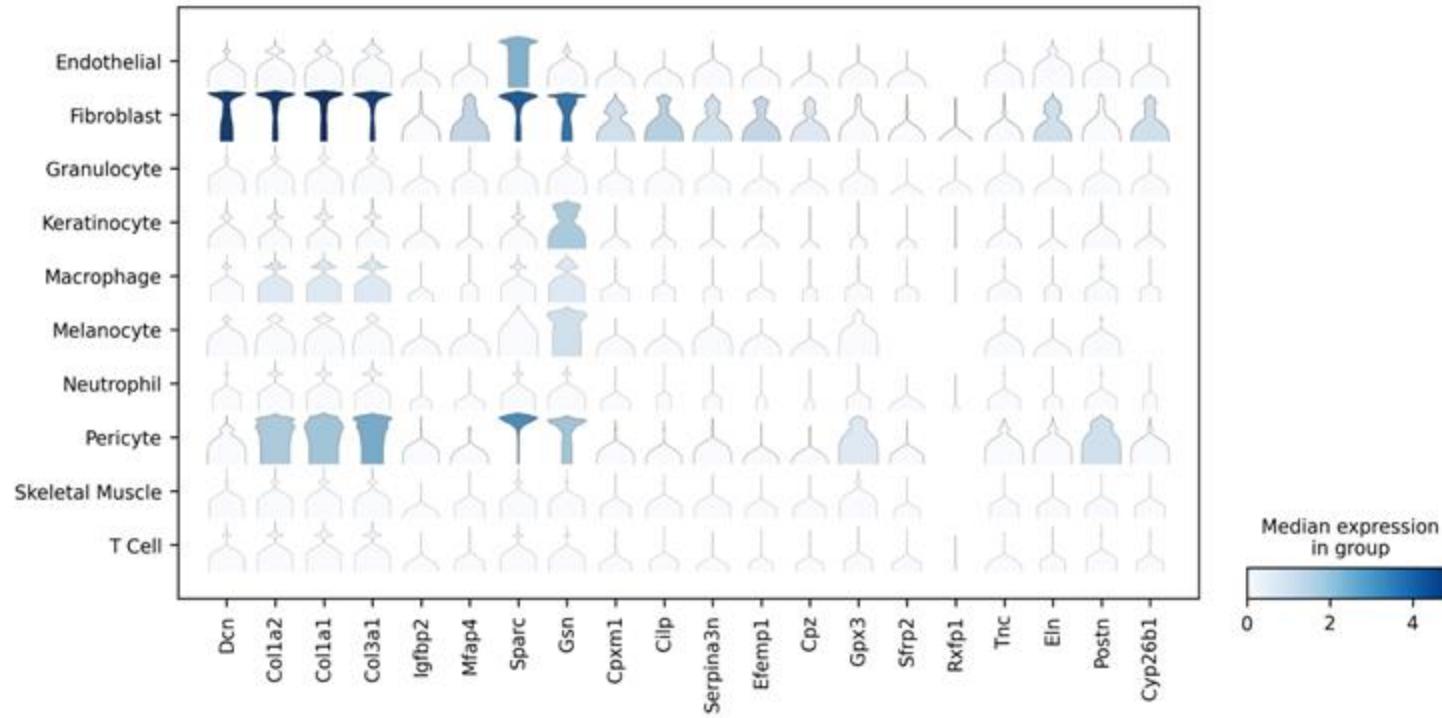
con: worst reconstruction

Coefficients

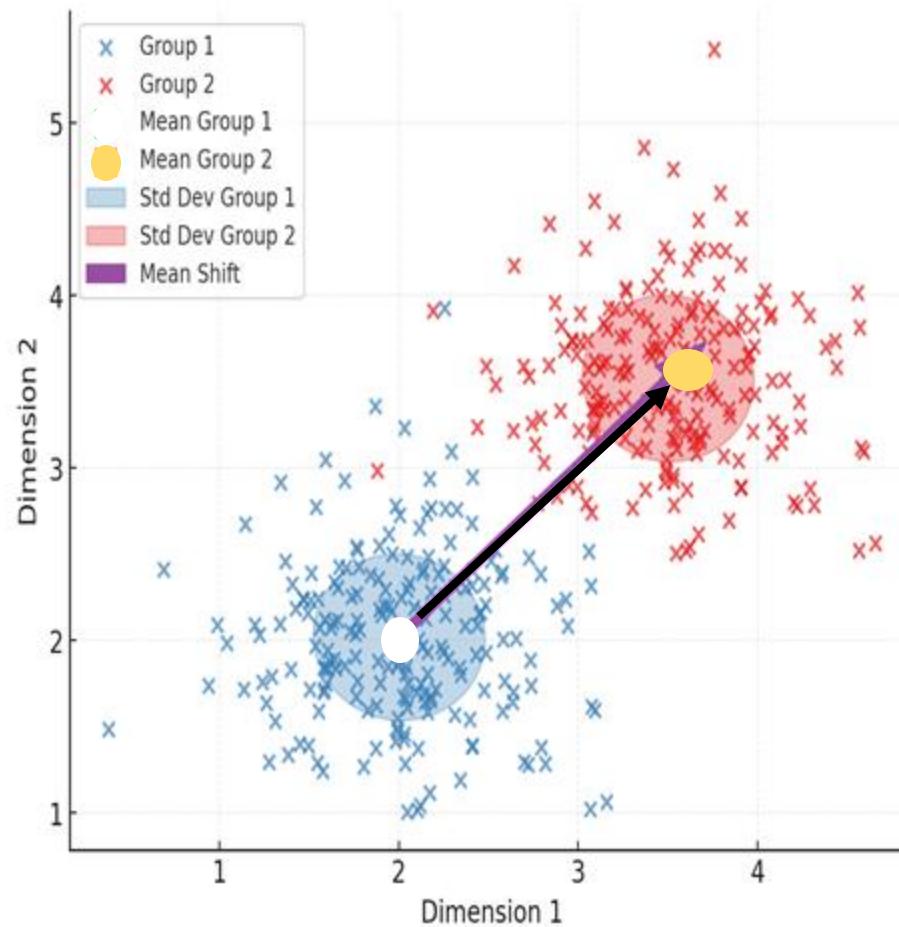


Linear-decoder Patches finds cellular markers

Fibroblast - Prop: 31.85%



Can I design counterfactuals across conditions?

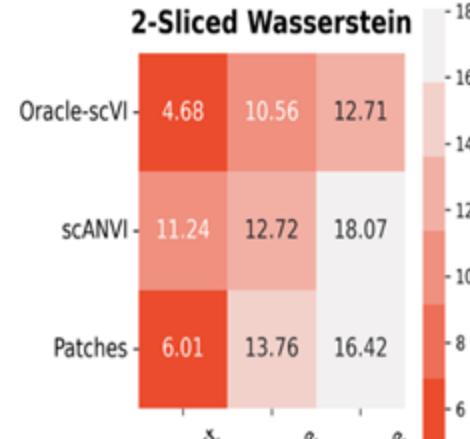
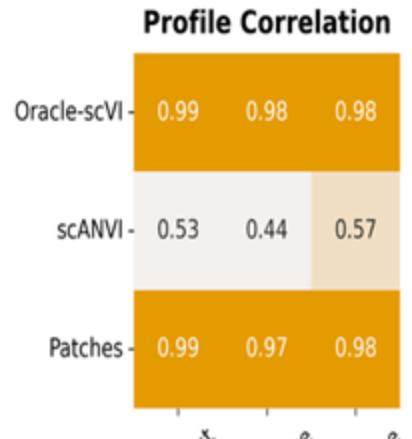
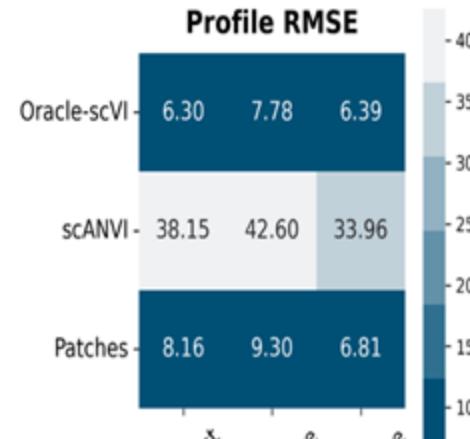


Oracle: knows the value of the mean shift

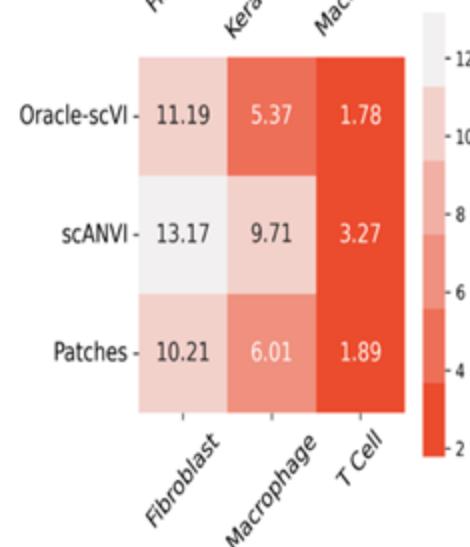
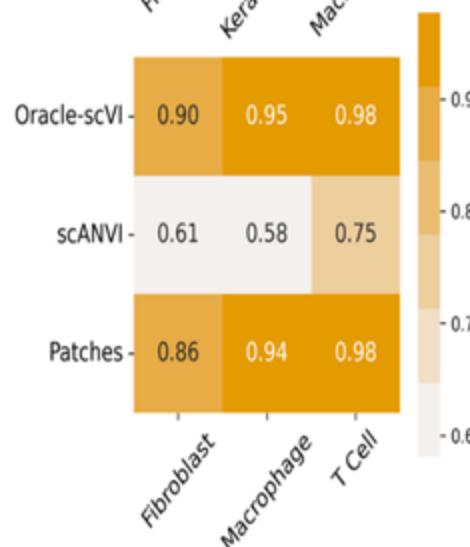
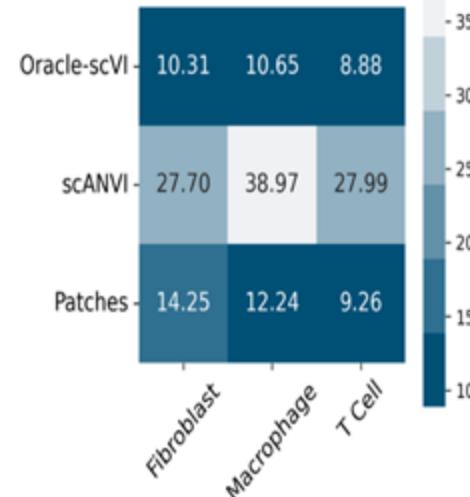


Can I design counterfactuals across conditions?

Vu et al. (2022)



Mascharka et al. (2022)



Reconstruction of transferred temporal profiles (aging the sample)



IN THE CITY OF NEW YORK

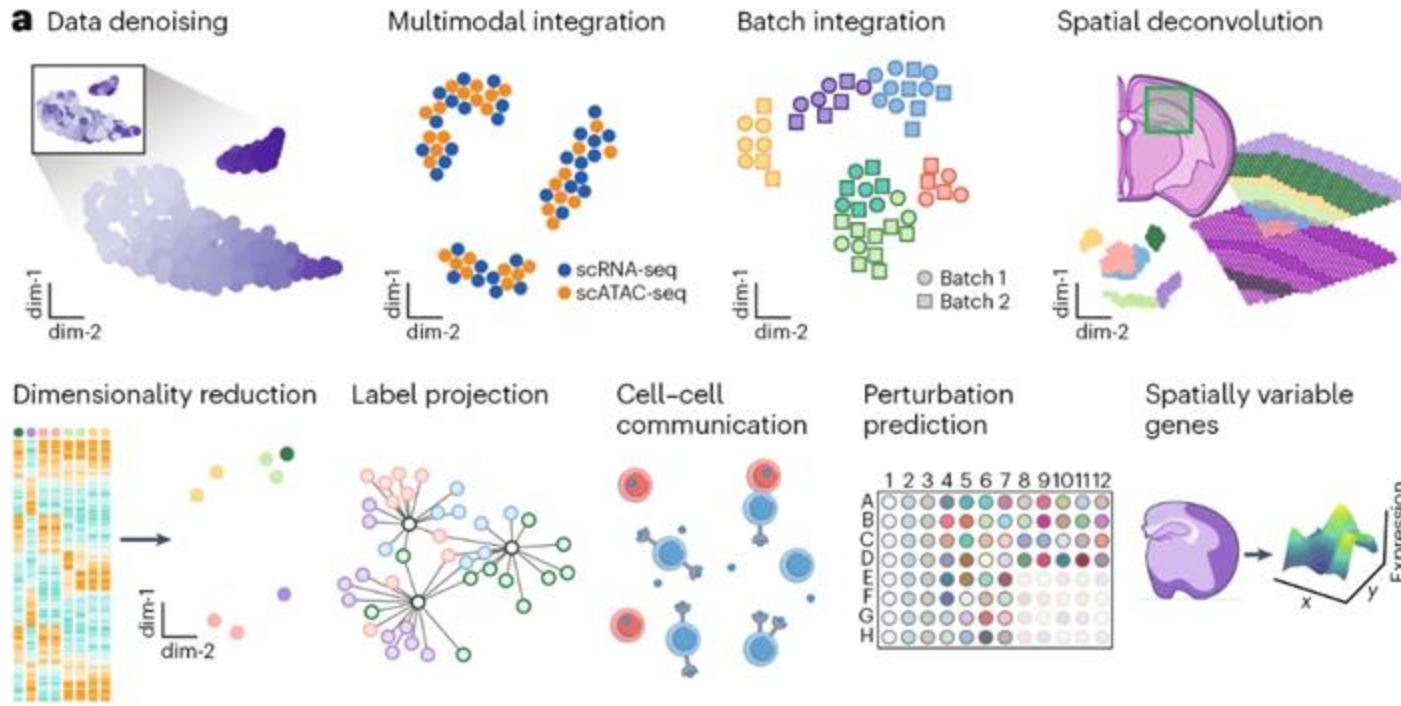
Dianica Dumitrascu, Machine Learning for Computational Biology, MLE50, 2025

Toolkit: scX, going forward

- * **the good: modular, open source**
- * **the bad: ~1000x algorithms and growing**

Toolkit: scX

* part of the solution: community benchmarks



From Luecken, M.D., Gigante, S., Burkhardt, D.B. et al.
Defining and benchmarking open problems in single-cell
analysis. *Nat Biotechnol* 43, 1035–1040 (2025)

What do we do with all these models?

Task: Interpretability or what do the models tell us about real biology?

Given a latent representation of cells (e.g., from a VAE or factor model), how can we assign *biologically meaningful* interpretations to each dimension?

Can we link these dimensions to gene modules, regulatory factors, or known biological pathways?

How do we **balance predictive performance with human insight?**

What do we do with all these models?

Task: Interpretability or what do the models tell us about real biology?

Given a latent representation of cells:

- Can I “name” any of the latent variables? Does this naming mean anything to a computational biologist?
- Can I intervene on these latent variables experimentally? If not, how do I validate information about them?

What do we do with all this data?

Task/Concept: Disentanglement or can I imagine counterfactuals, alternatives so I plan better

Given a latent representation of cells (e.g., from a VAE or factor model), how can we assign *biologically meaningful* interpretations to each dimension?

Can we link these dimensions to gene modules, regulatory factors, or known biological pathways?

How do we **balance predictive performance with human insight?**

Next time

- **From cells to spatial contexts, from spatial contexts to spatio-temporal data**
- **Practical toolkit: spatial statistics, morphology quantification, graph neural networks**



From latent spaces to living systems: Lecture 3 & 4

Dynamic tissue geography