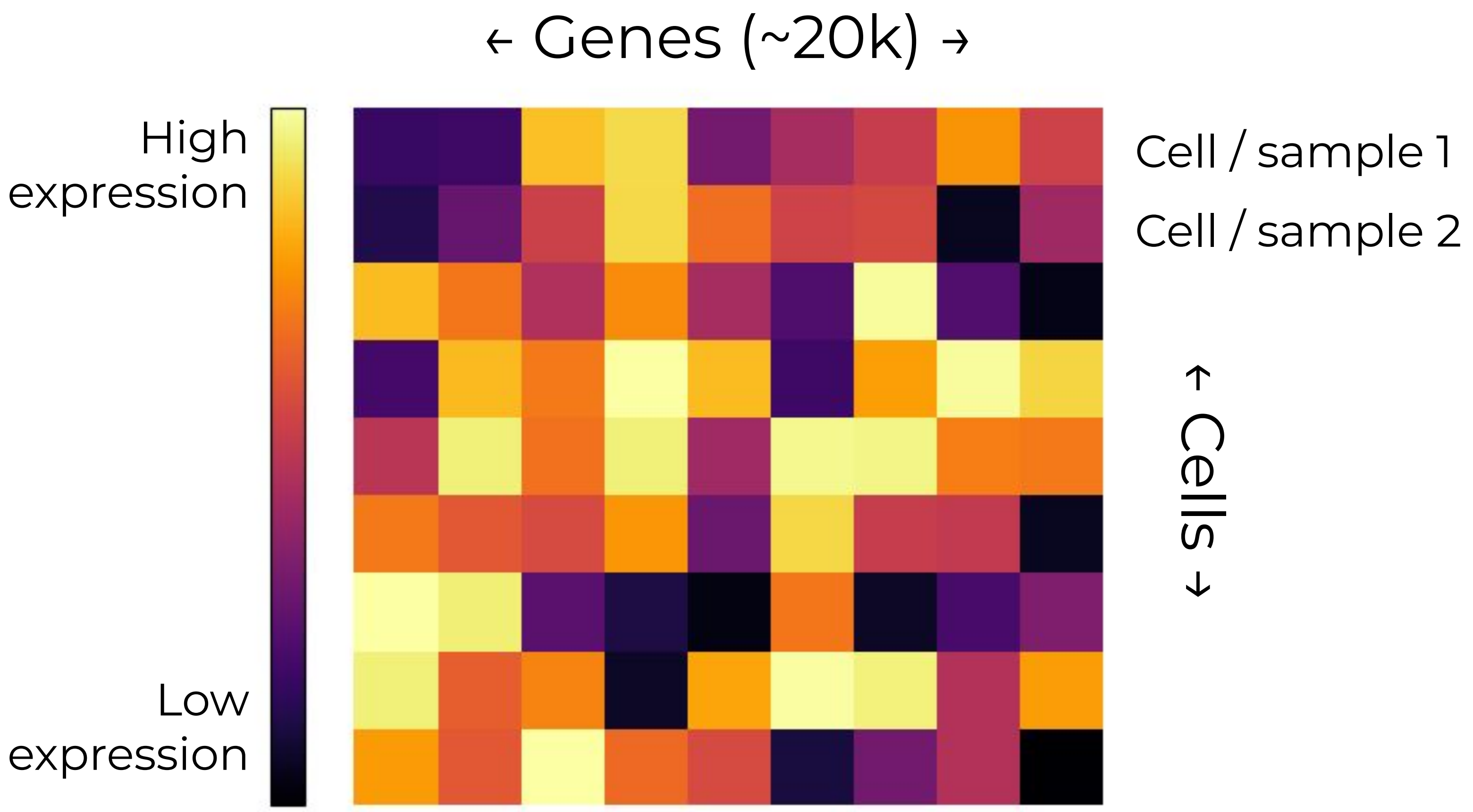
Contrastive Mixture of Posteriors for Counterfactual Inference, Data Integration and Fairness

Adam Foster[1] Árpi Vezér[2] Craig A Glastonbury[2 3]
Páidí Creed[2] Sam Abubudeh[2] Aaron Sim[2]

[1]Microsoft Research, Cambridge. Work completed at BenevolentAI and University of Oxford. [2]BenevolentAI, London. [3]Human Technopole, Milan, Italy.

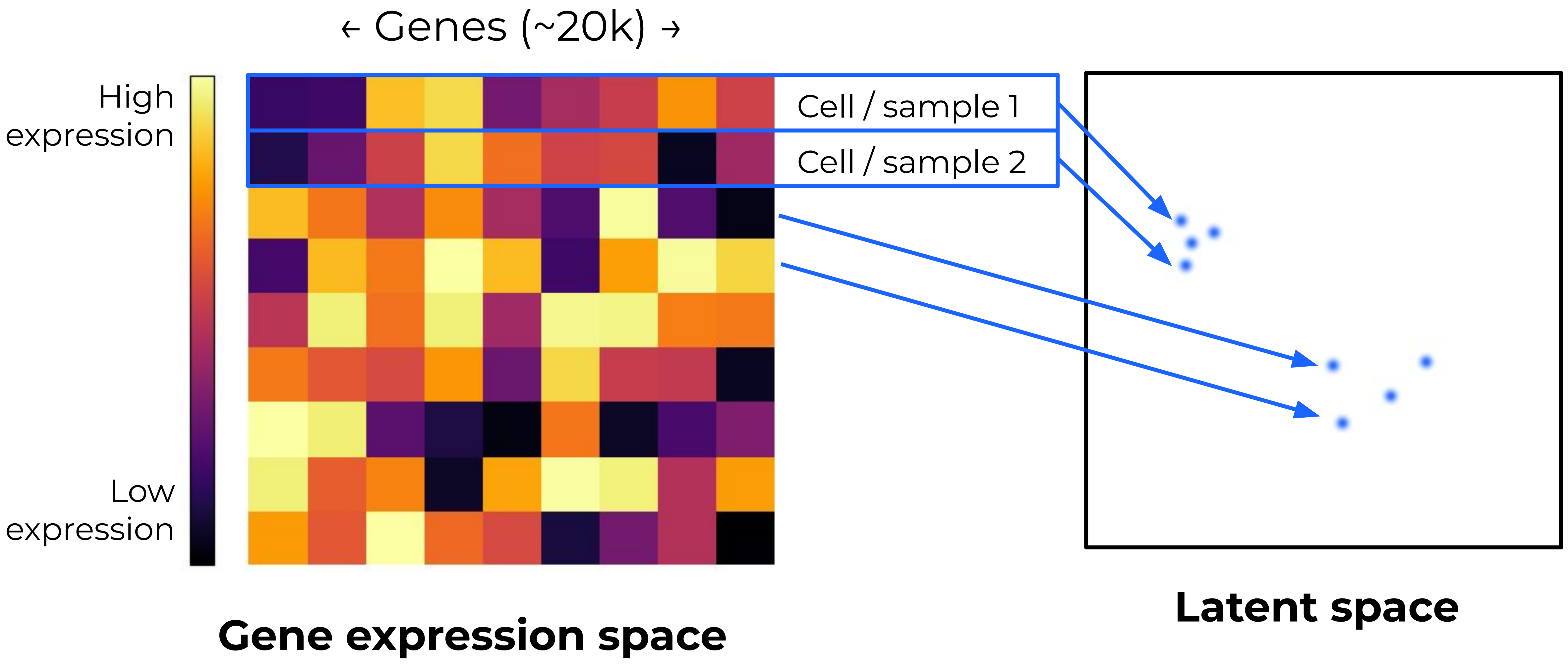
Benevolent AI

‘Omics data

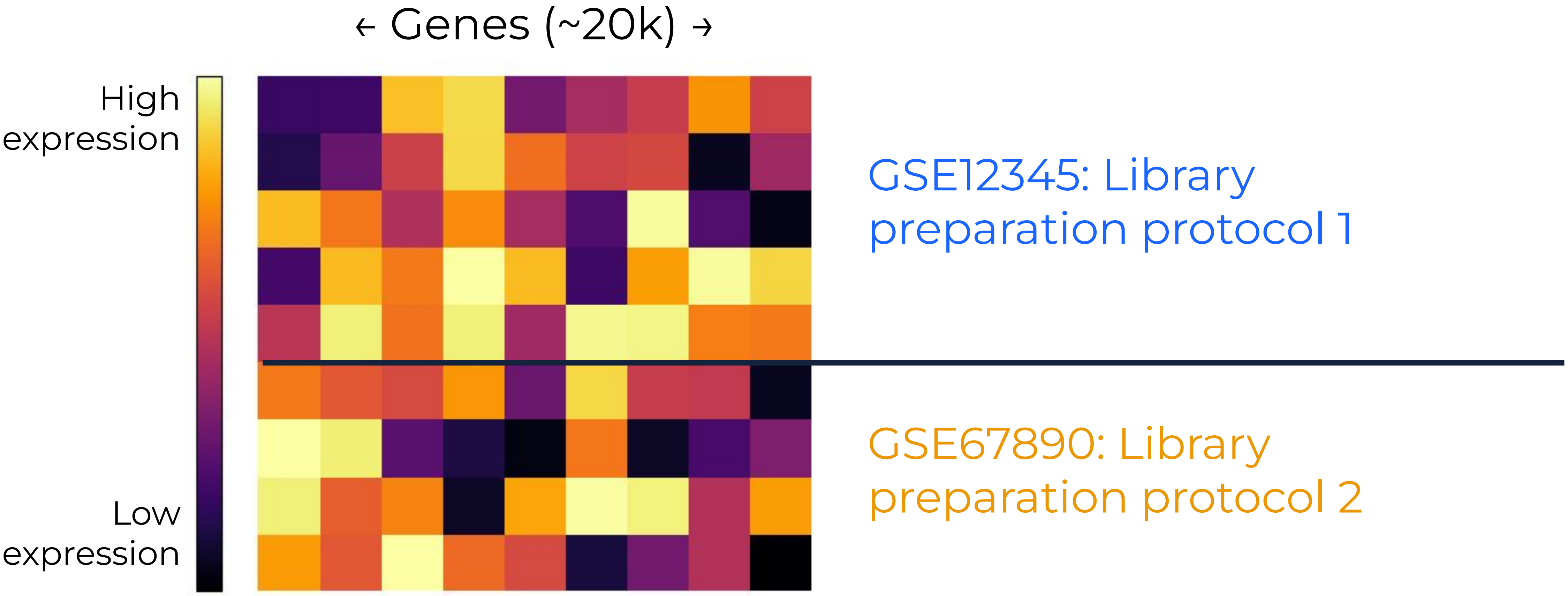


Particularly interested in **transcriptomics** and **single-cell RNA-seq**

Representation learning for ‘omics data

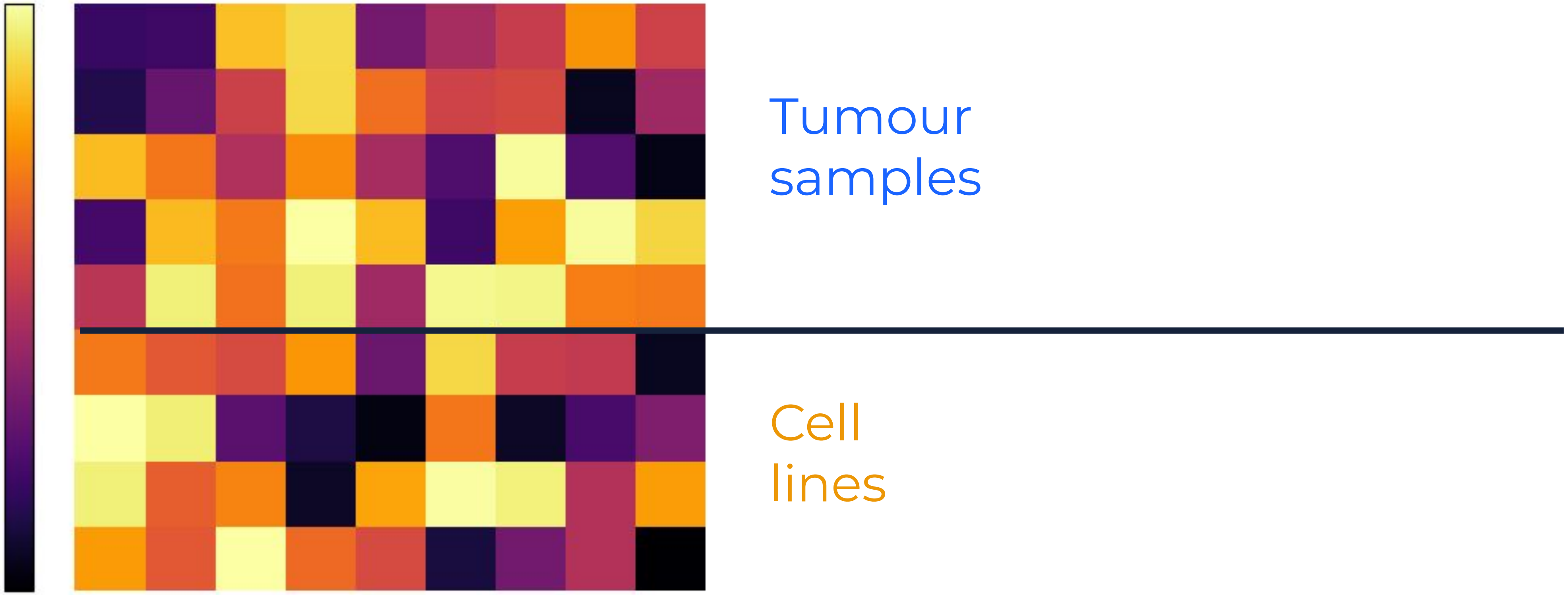


Challenge 1: data integration & batch correction



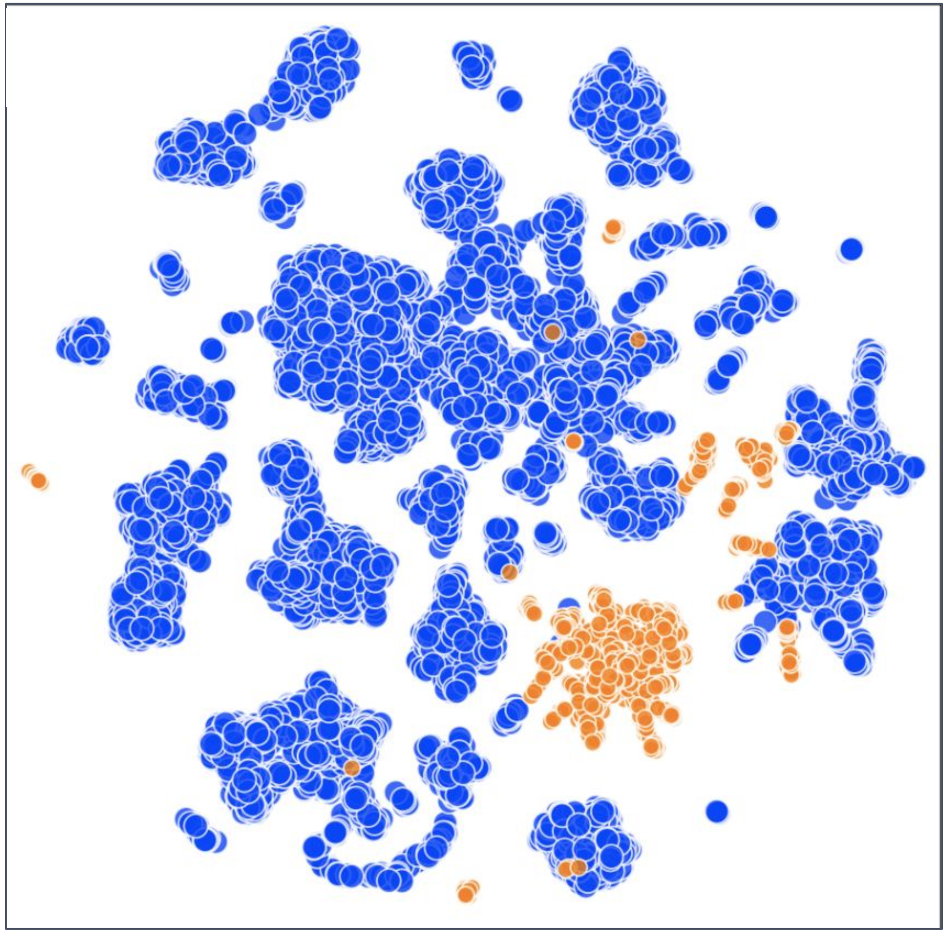
As in Korsunsky et al., 2019
Harmony paper

Challenge 1: data integration & batch correction



As in Warren et al., 2021
Celligner paper

Challenge 1: data integration & batch correction



● Tumour samples (TCGA)

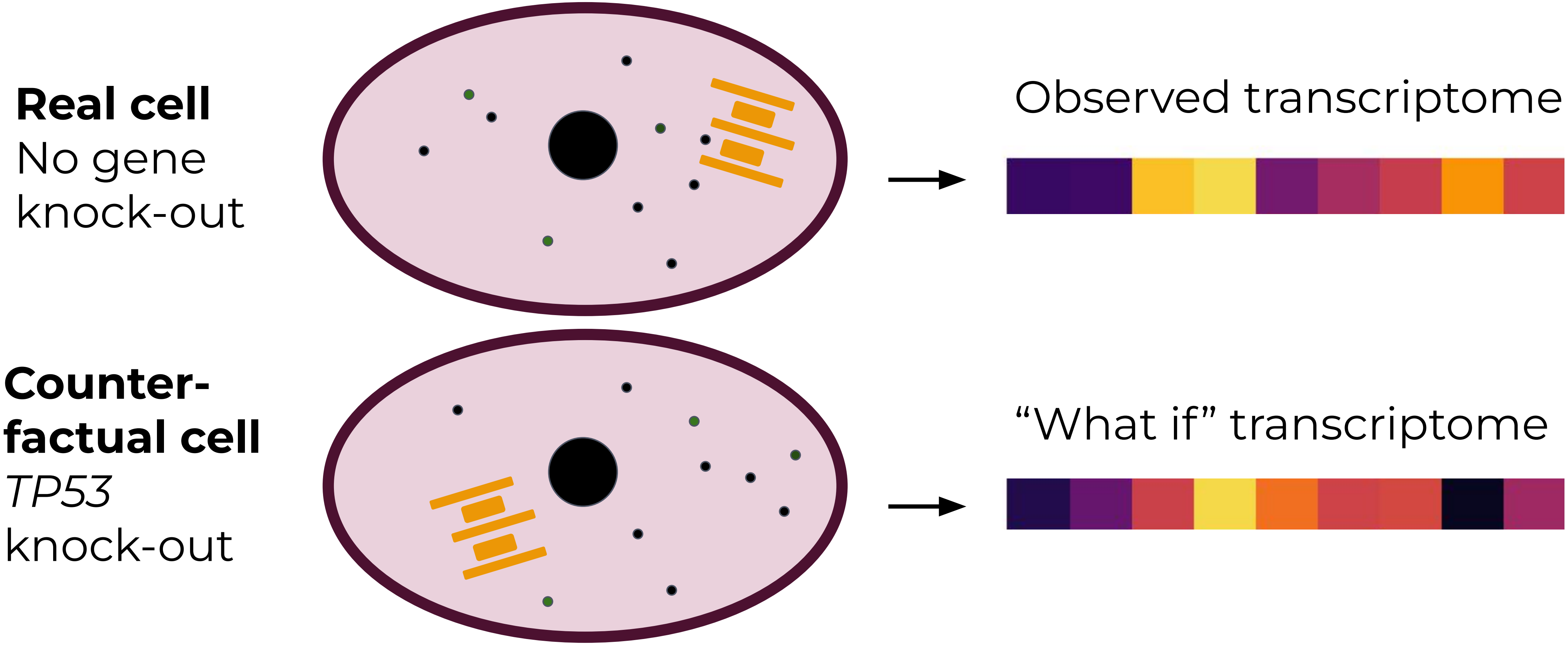
● Cell lines (CCLE)

Datasets contain **comparable cell populations**, but there is **unwanted global variation**

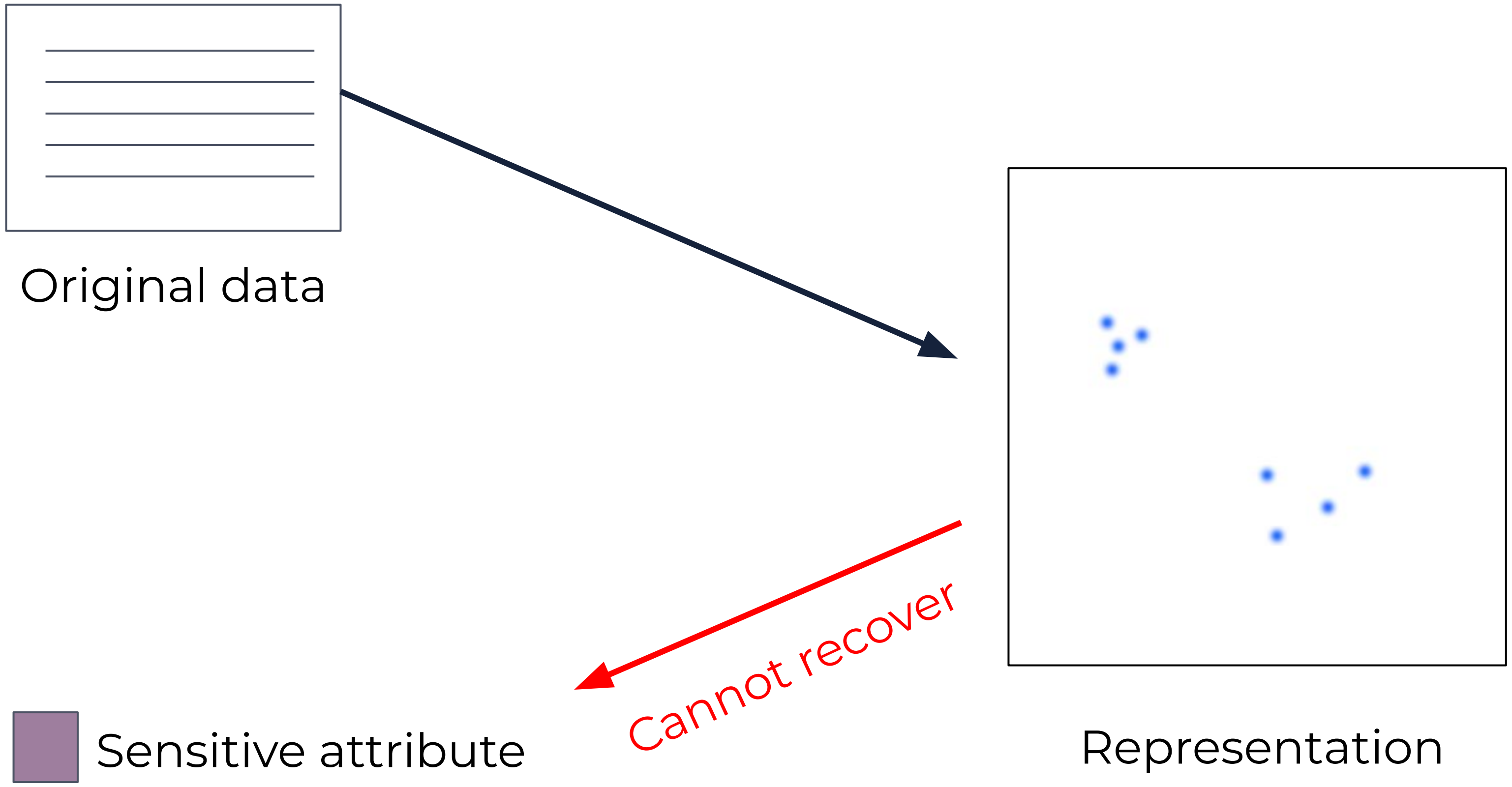
Aim: subtract out the tumour / cell line global variation

*The Cancer Genome Atlas

Challenge 2: counterfactual predictions of effects of interventions

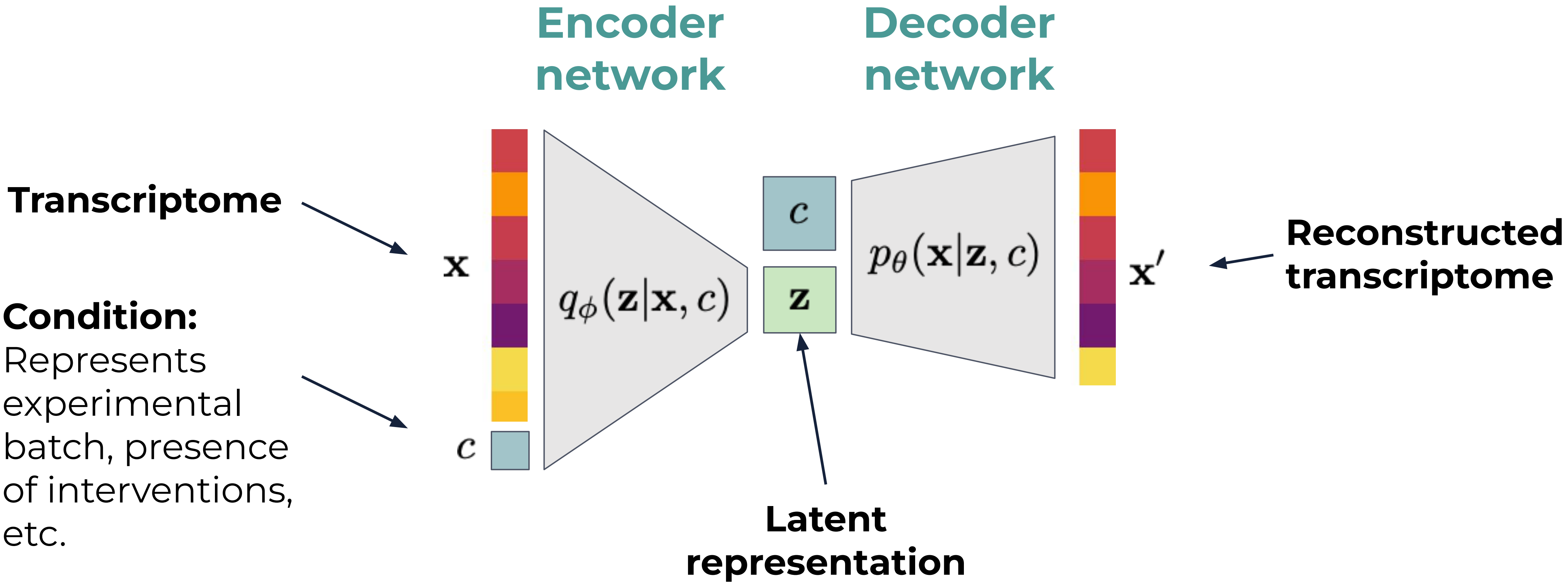


Challenge 3: learning fair representations

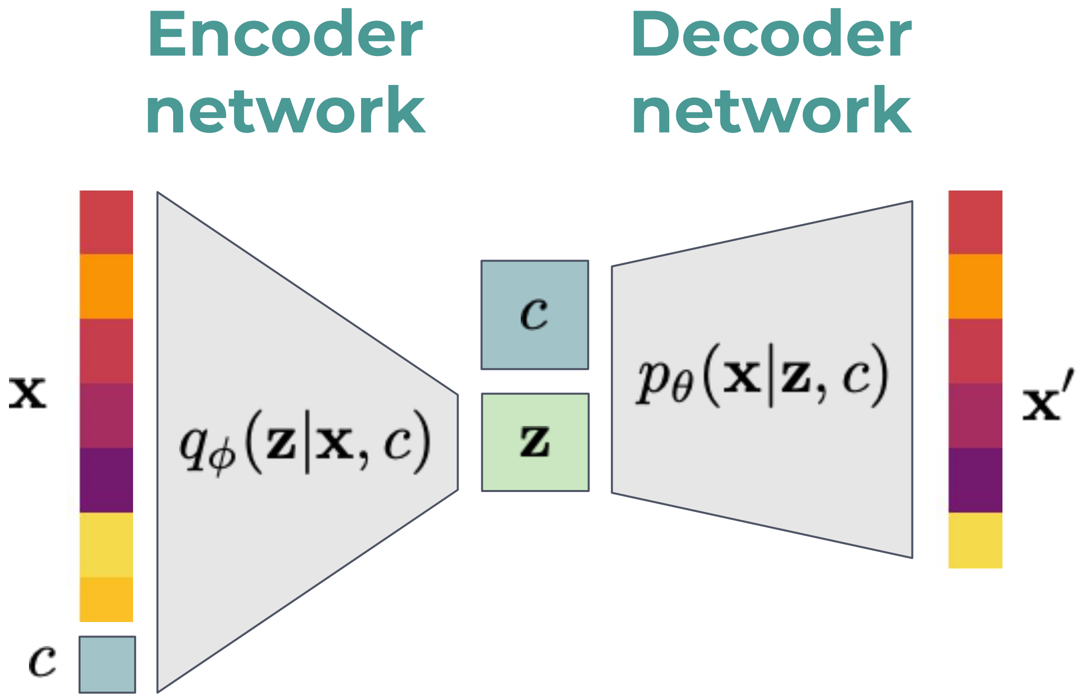


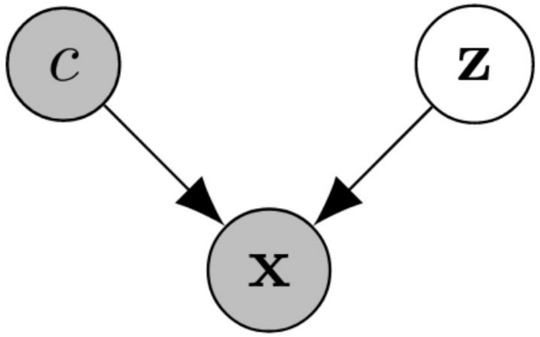
Conditional Variational AutoEncoders

Conditional Variational Autoencoder (CVAE)



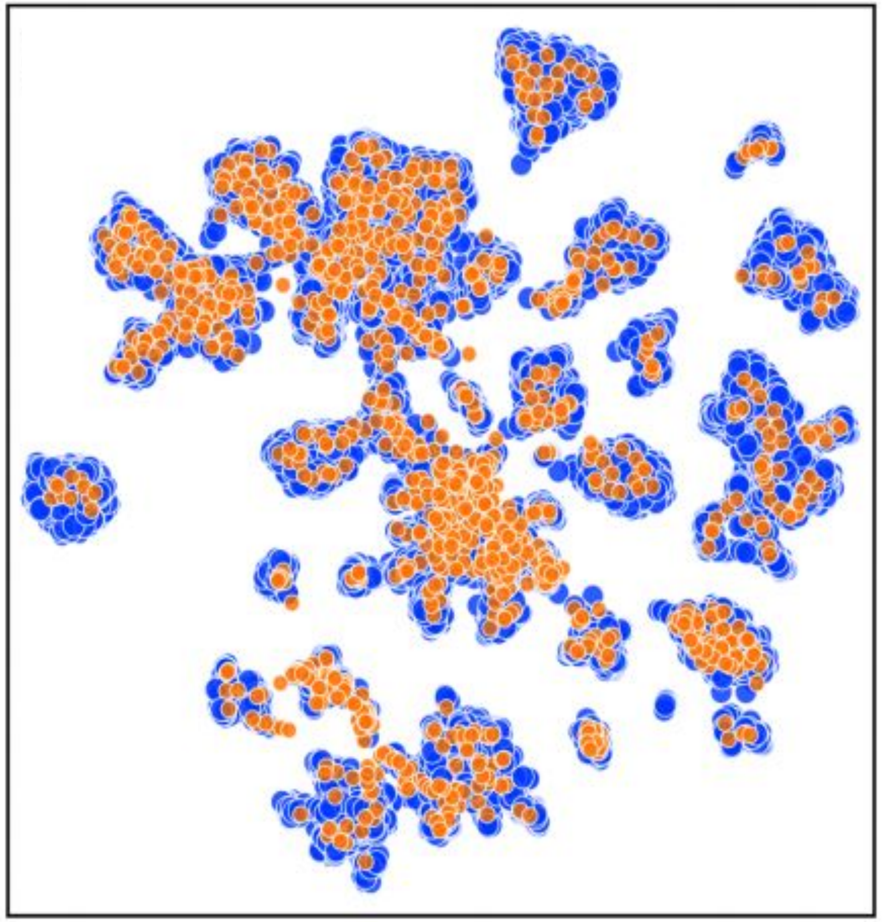
Conditional Variational Autoencoder (CVAE)



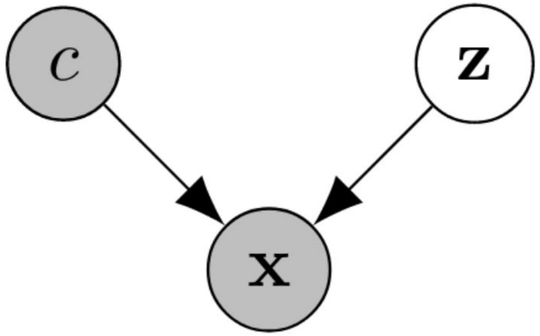


z and c are modelled as independent

The importance of latent space alignment

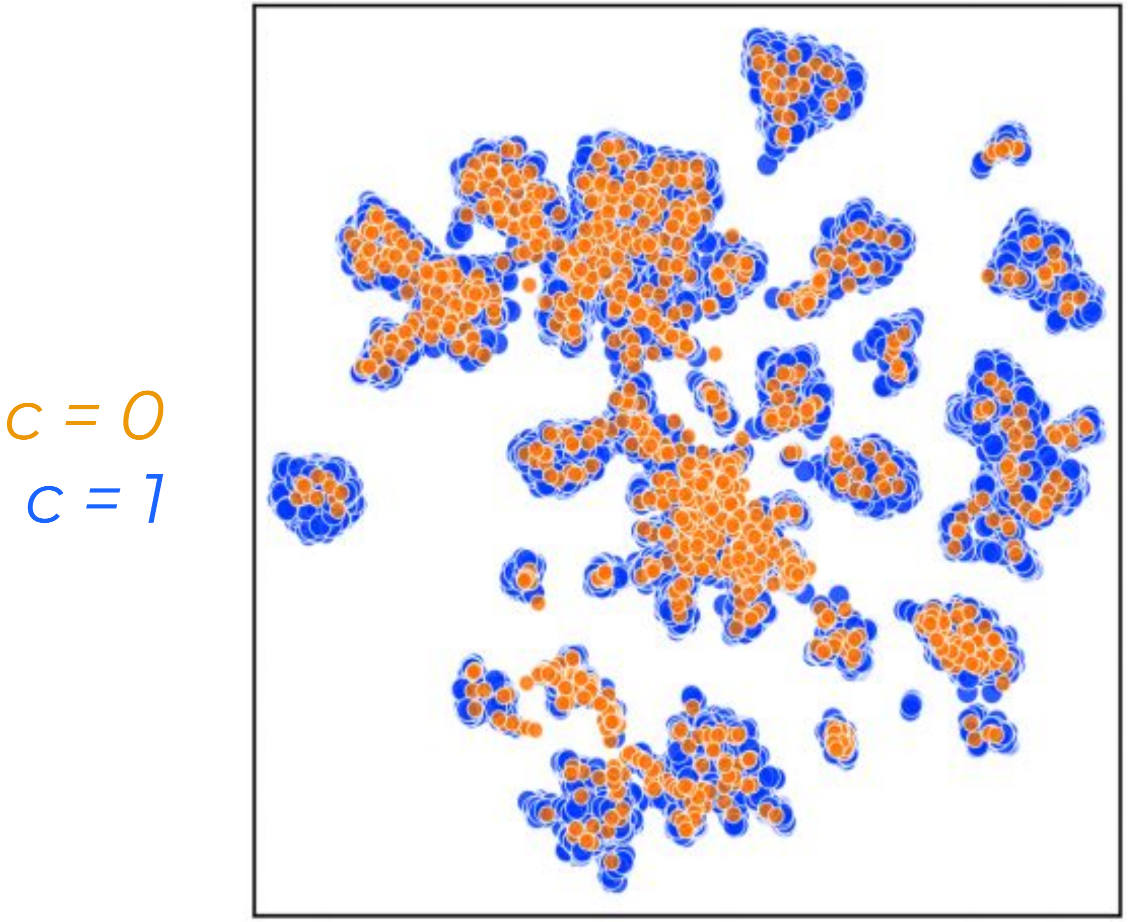


z plot



z and c are modelled as independent

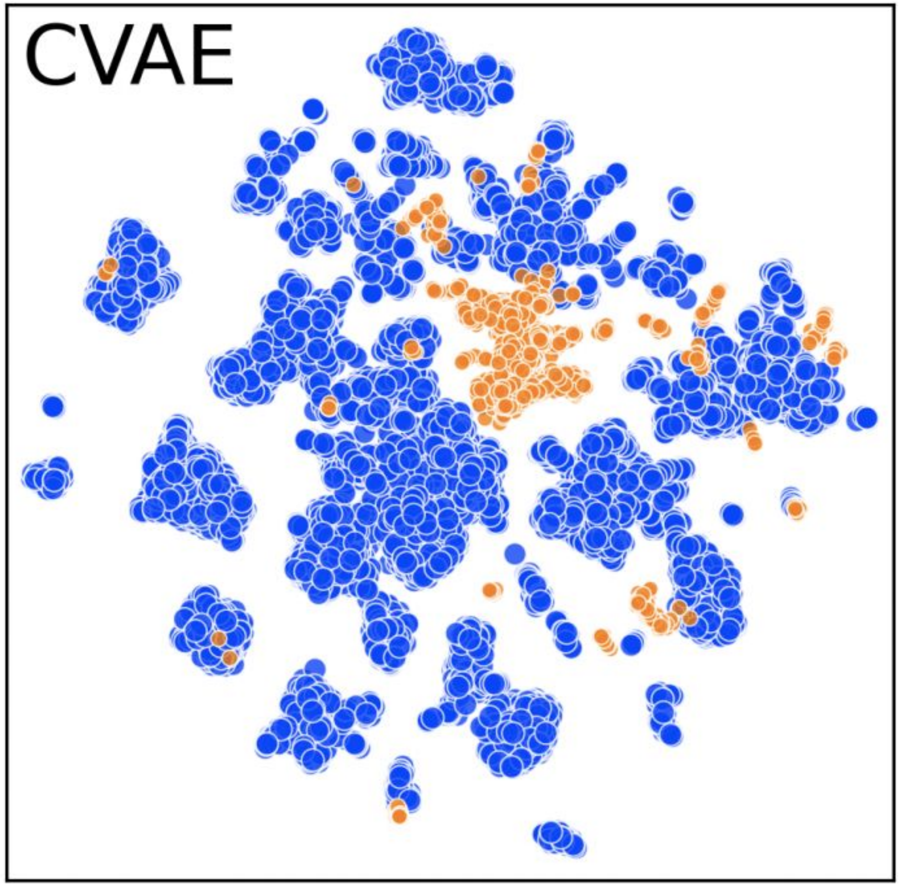
The importance of latent space alignment



z plot

This alignment directly gives us **data integration** in latent space

However...



- The model is under no obligation to use the condition label c
- Training standard CVAE on Tumour / Cell lines does not lead to good alignment

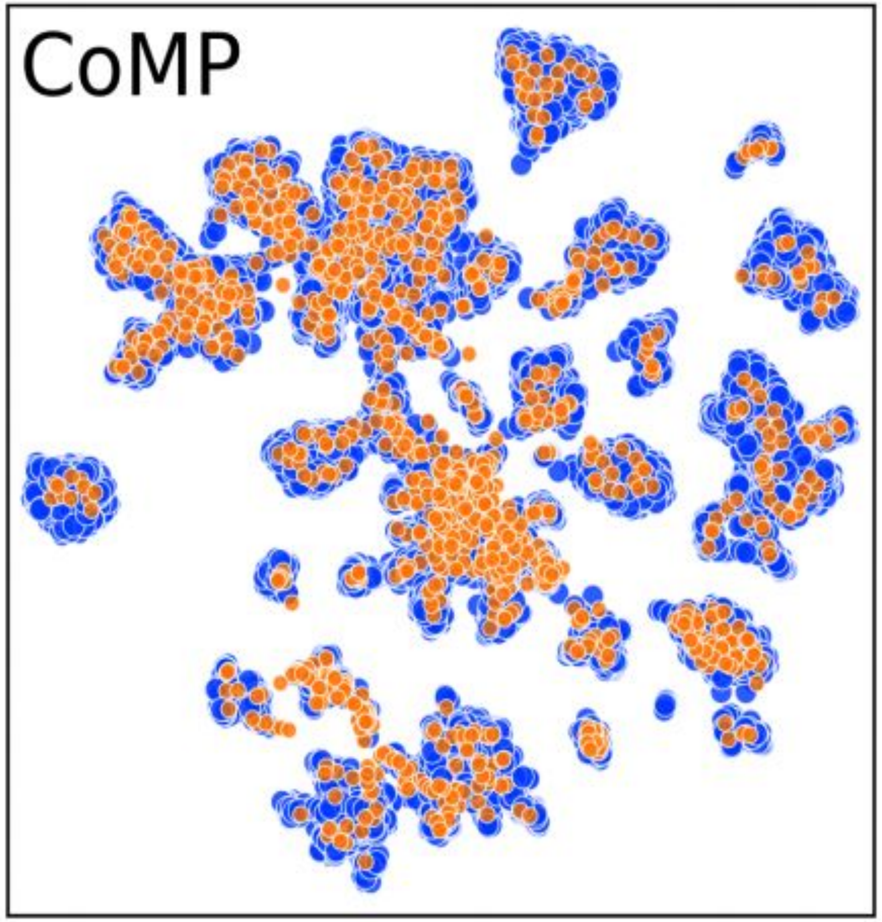
Contrastive Mixture of Posteriors Misalignment Penalty

Aim

Train a CVAE such that

$$\mathbf{z} \perp\!\!\!\perp c$$

holds

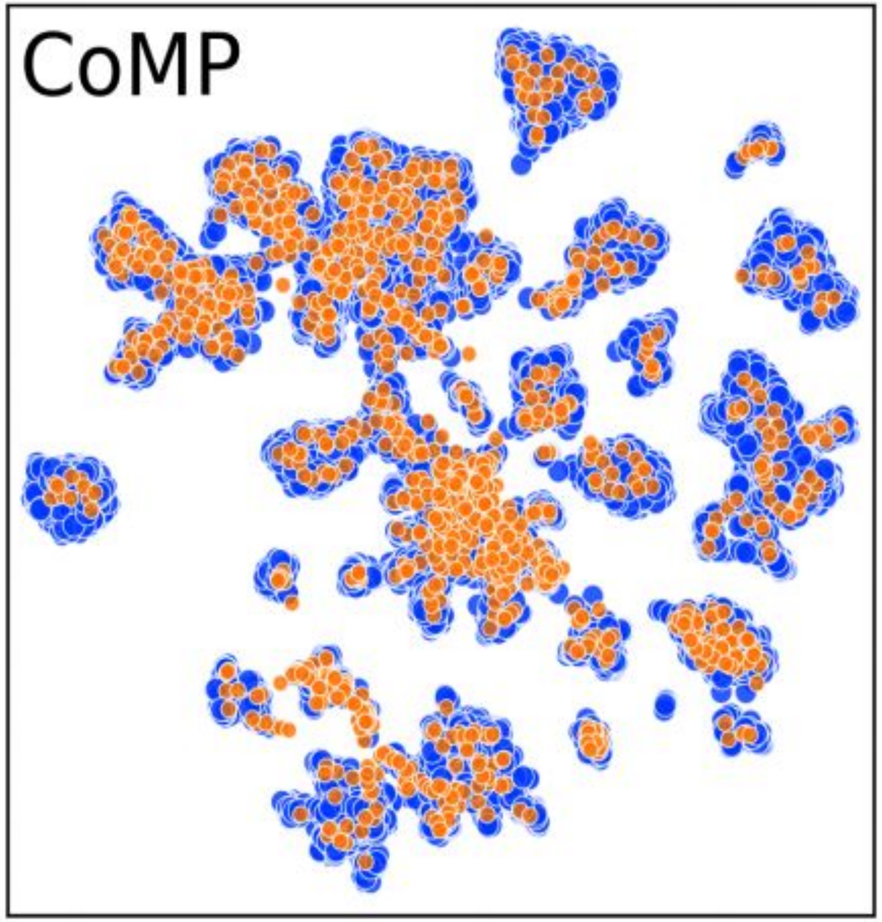


Aim (equivalent form)

$$\textit{for every } c \in \mathcal{C}, \mathrm{KL}(q(\mathbf{z}|c) \| q(\mathbf{z}|\neg c)) = 0$$

Marginal distribution of ***z*** for cells in condition c

Marginal distribution of ***z*** for cells in any other condition



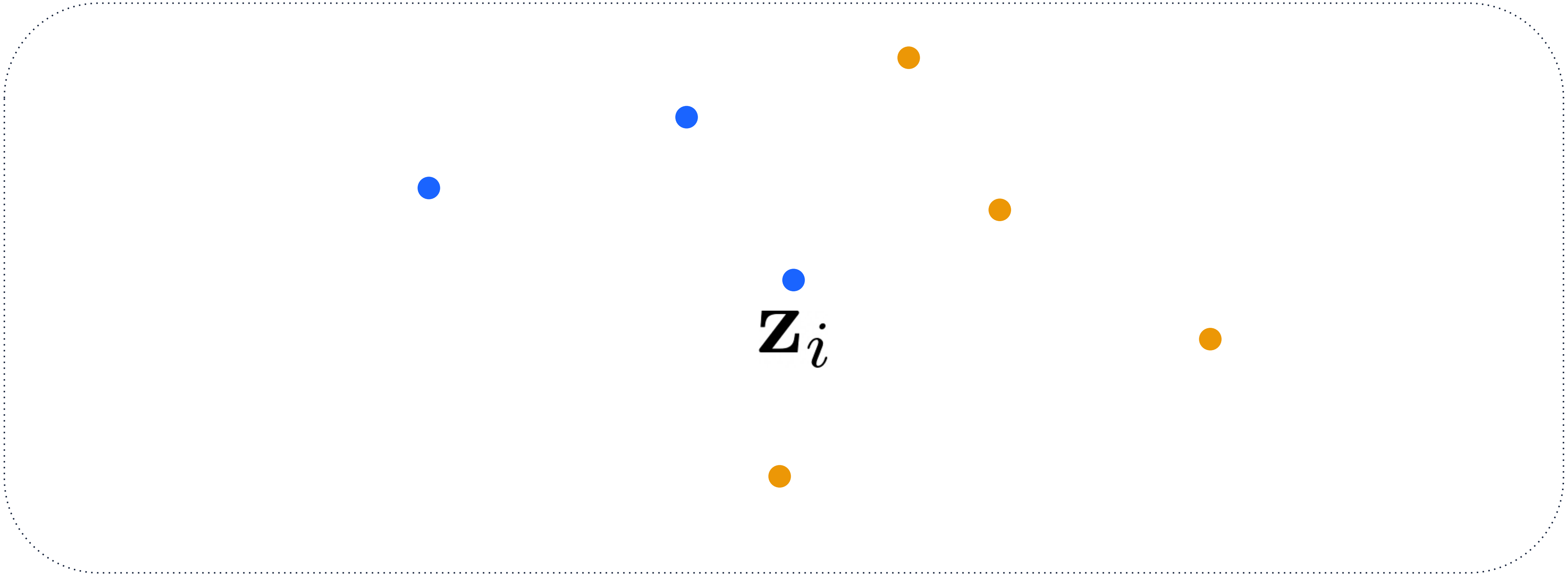
CoMP: Contrastive Mixture of Posteriors

$$\text{CoMP penalty} = \frac{1}{B}\sum_{i=1}^{B}\log\left(\frac{1}{|I_{c_i}|}\sum_{j\in I_{c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_i)\right) - \log\left(\frac{1}{|I_{\neg c_i}|}\sum_{j\in I_{\neg c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_j)\right).$$

CoMP: Contrastive Mixture of Posteriors

$$\text{CoMP penalty} = \frac{1}{B}\sum_{i=1}^{B} \log\left(\frac{1}{|I_{c_i}|}\sum_{j \in I_{c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_i)\right) - \log\left(\frac{1}{|I_{\neg c_i}|}\sum_{j \in I_{\neg c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_j)\right).$$

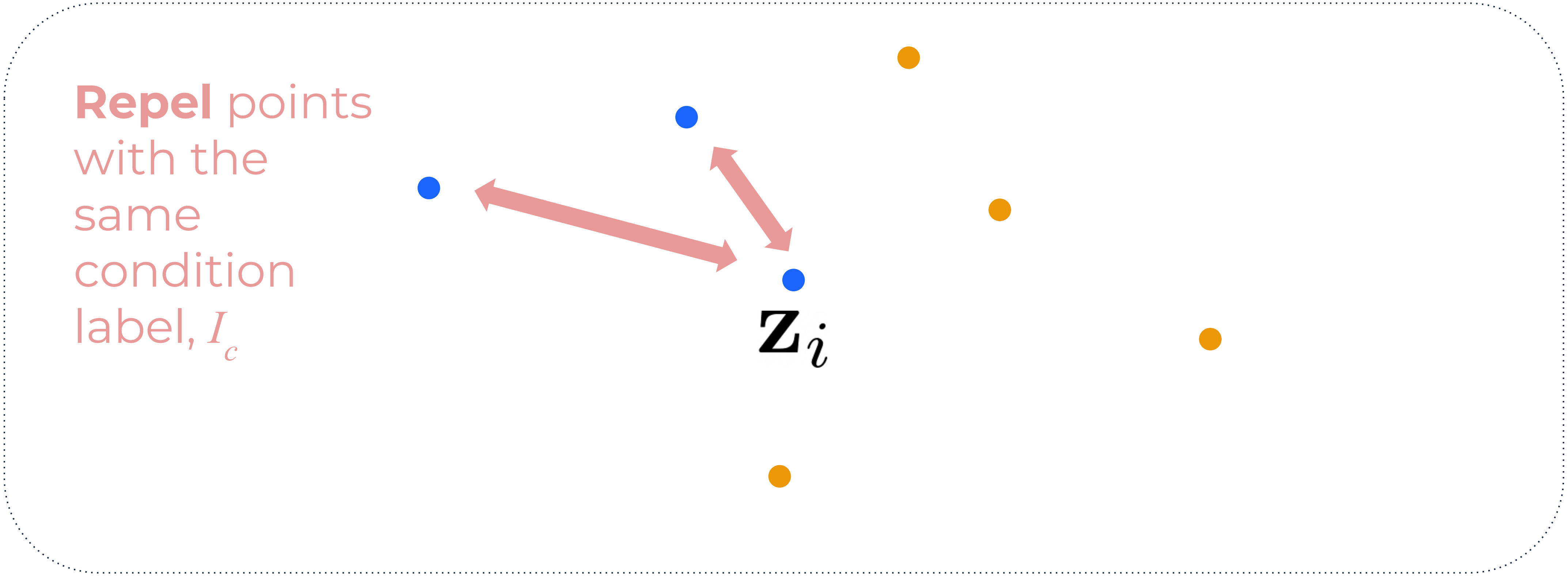
Representations of one training batch (size *B*)



CoMP: Contrastive Mixture of Posteriors

$$\text{CoMP penalty} = \frac{1}{B}\sum_{i=1}^{B} \log\left(\frac{1}{|I_{c_i}|}\sum_{j\in I_{c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_i)\right) - \log\left(\frac{1}{|I_{\neg c_i}|}\sum_{j\in I_{\neg c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_j)\right).$$

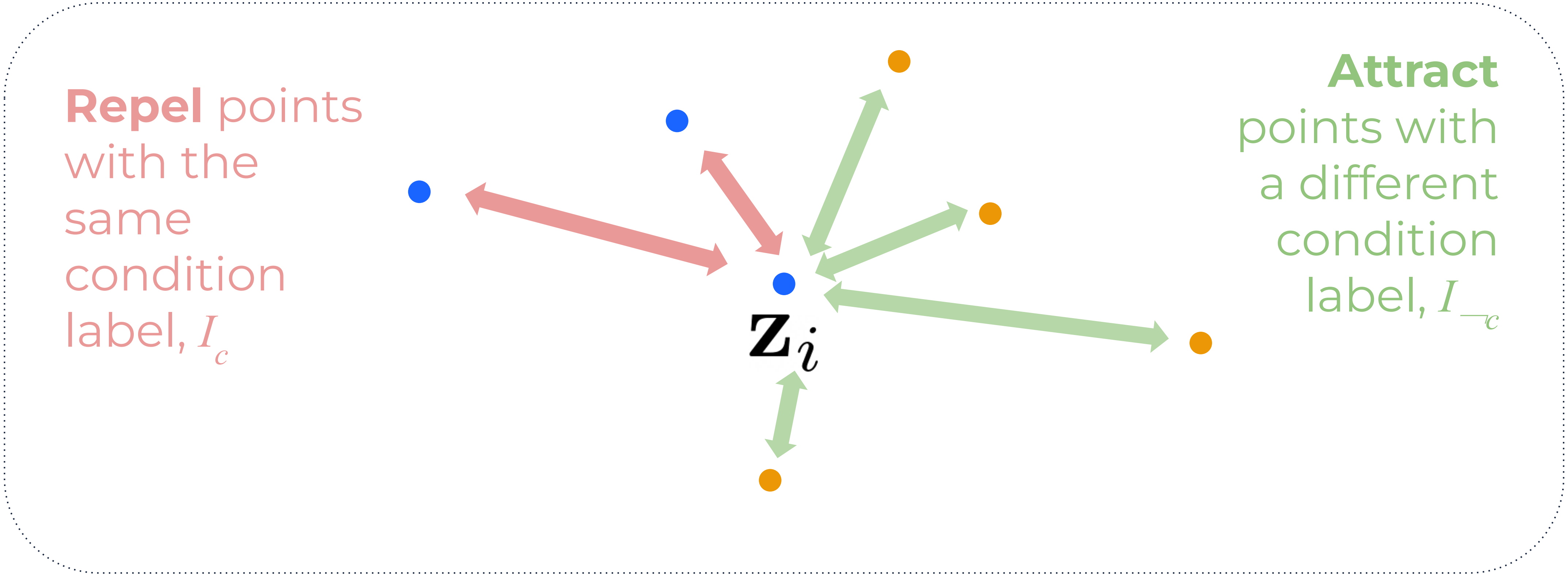
Representations of one training batch (size *B*)



CoMP: Contrastive Mixture of Posteriors

$$\text{CoMP penalty} = \frac{1}{B}\sum_{i=1}^{B} \log\left(\frac{1}{|I_{c_i}|}\sum_{j \in I_{c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_i)\right) - \log\left(\frac{1}{|I_{\neg c_i}|}\sum_{j \in I_{\neg c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_j)\right).$$

Representations of one training batch (size *B*)



CoMP: Contrastive Mixture of Posteriors

Theorem 1. *The CoMP misalignment penalty satisfies*

$$\mathbb{E}_{\prod_{i=1}^{B} p(\mathbf{x}_i, c_i) q(\mathbf{z}_i|\mathbf{x}_i, c_i)} \left[\frac{1}{B} \sum_{i=1}^{B} \log\left(\frac{1}{|I_{c_i}|} \sum_{j \in I_{c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_i) \right) - \log\left(\frac{1}{|I_{\neg c_i}|} \sum_{j \in I_{\neg c_i}} q(\mathbf{z}_i|\mathbf{x}_j, c_j) \right) \right]$$

$$\geq \sum_{c \in \mathcal{C}} p(c) \, \mathrm{KL} \left[q(\mathbf{z}|c) || q(\mathbf{z}|\neg\, c) \right]$$

CoMP: Contrastive Mixture of Posteriors

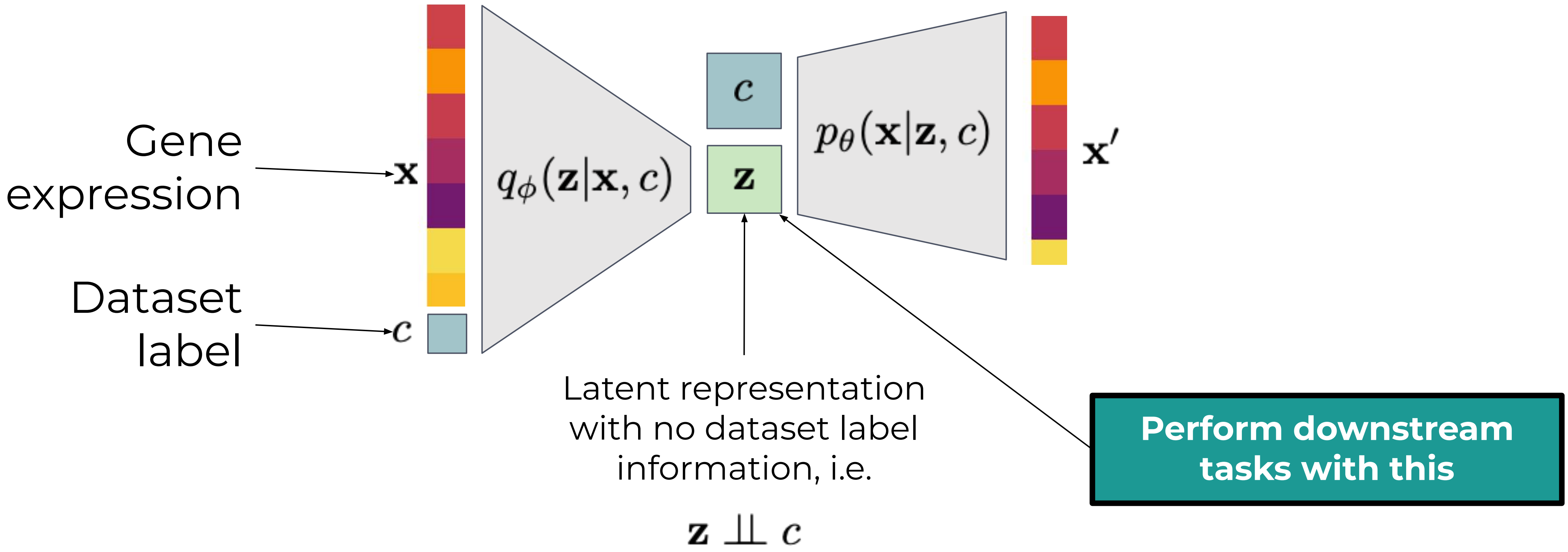
Our training objective to max is a penalised ELBO

$$\mathrm{ELBO} - \gamma(\mathrm{CoMP\ \ penalty})$$

Applying CoMP

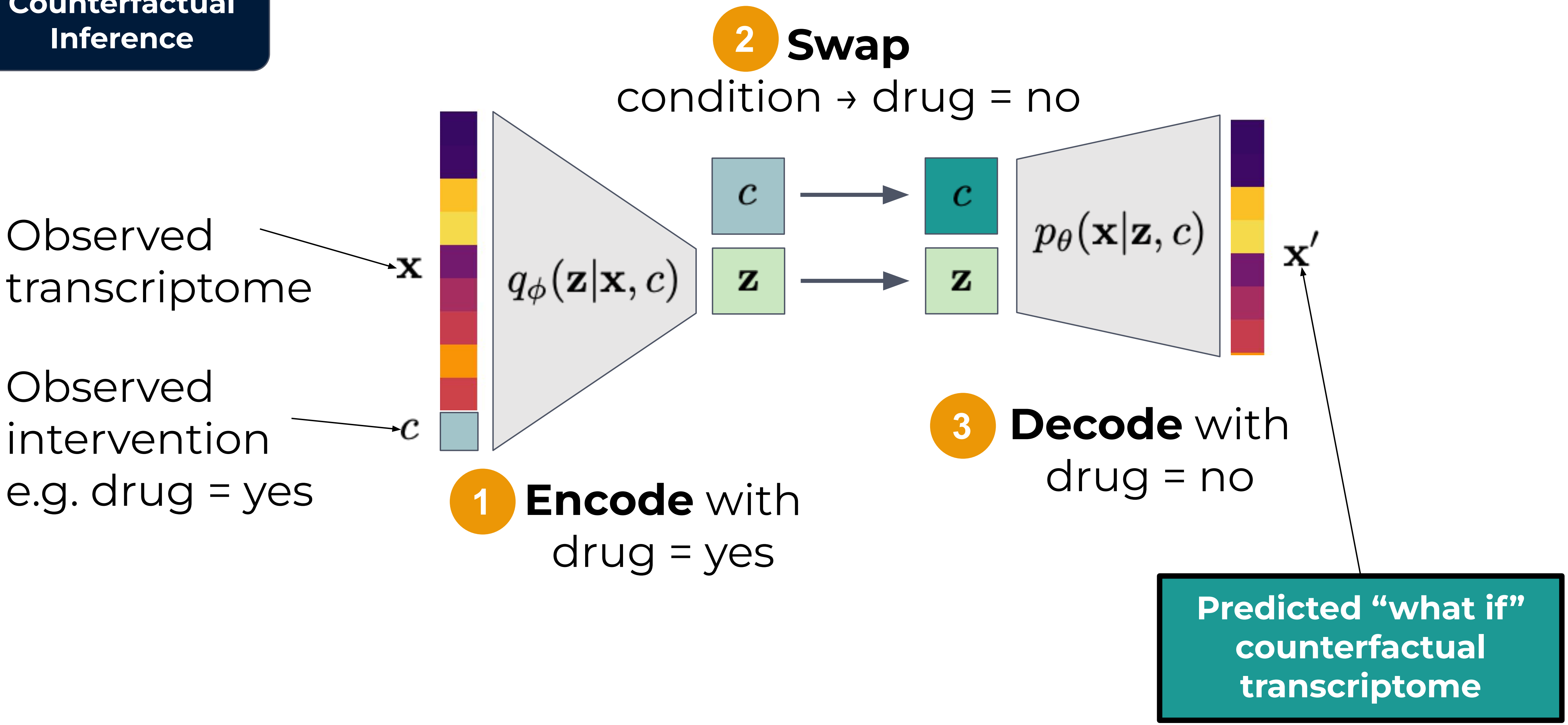
How do we use the CoMP CVAE model?

1 **Data Integration**



How do we use the CoMP CVAE model?

2 **Counterfactual Inference**



Theory

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

Counterfactual identifiability

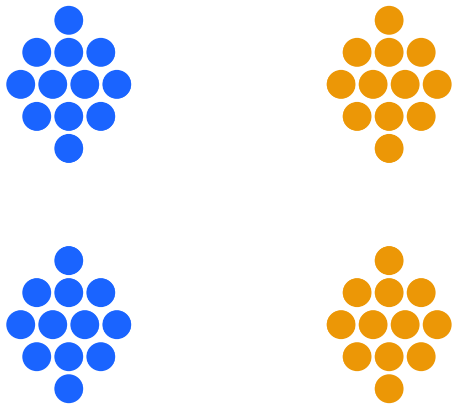
Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.

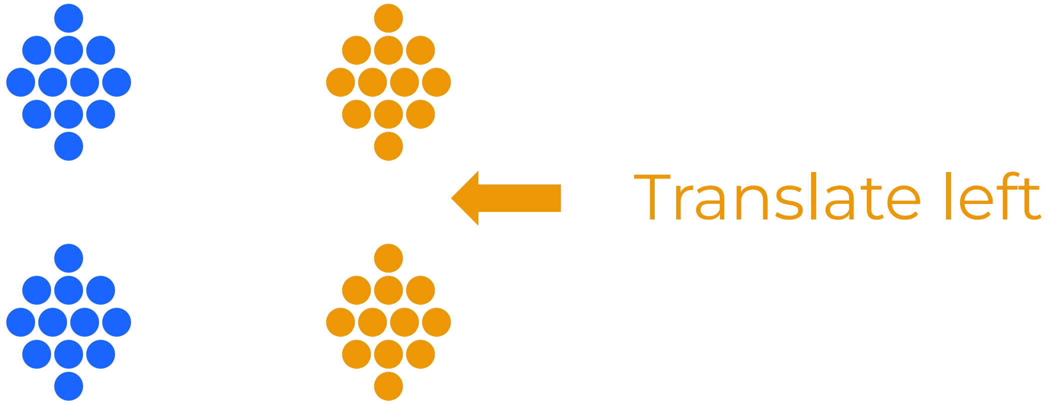


Latent space with **z**
and *c* not independent

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.

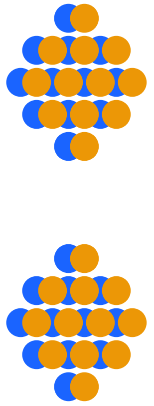


Latent space with **z**
and c not independent

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.

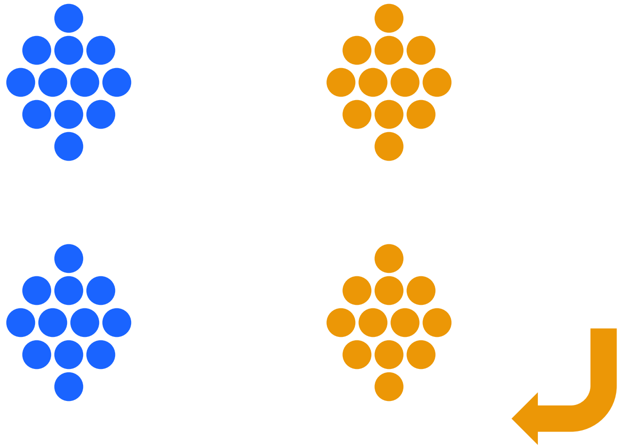


Latent space with **z** and c independent

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.



Rotate 180°

Latent space with **z** and c not independent

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

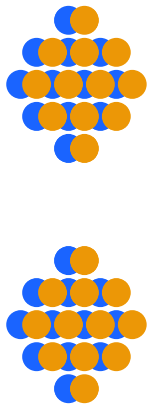
No.

Latent space with **z** and c not independent

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.

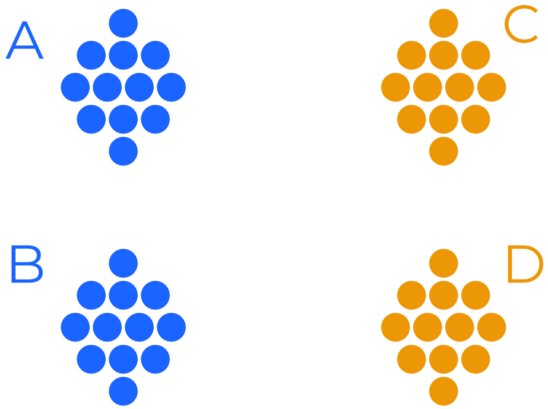


Latent space with **z** and c independent

Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No.

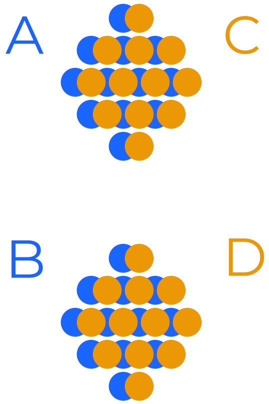


Latent space with **z** and *c* not independent

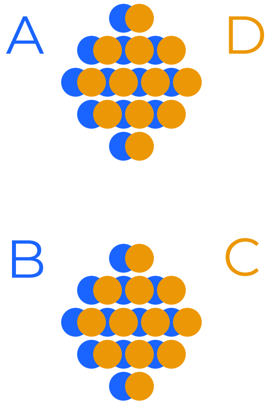
Counterfactual identifiability

Question: Is $\mathbf{z} \perp\!\!\!\perp c$ sufficient to find counterfactuals, assuming the model is correct?

No. There is no way to tell which of these models is the right one.



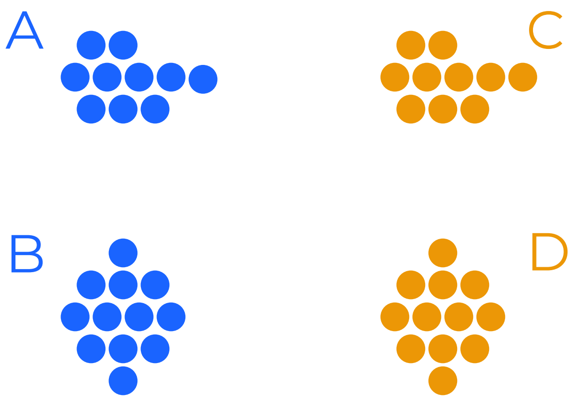
Translation



Rotation

Counterfactual identifiability

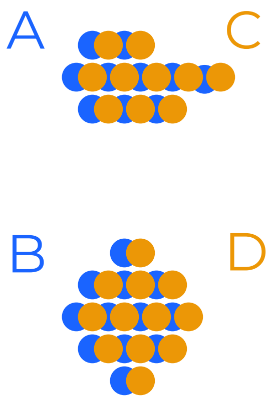
Theorem (inexact statement): If we have a linear decoder and we assume that there is no exact, linear symmetry of the true latent distribution, then counterfactuals are identifiable.



Latent space with **z** and *c* not independent

Counterfactual identifiability

Theorem (inexact statement): If we have a linear decoder and we assume that there is no exact, linear symmetry of the true latent distribution, then counterfactuals are identifiable.



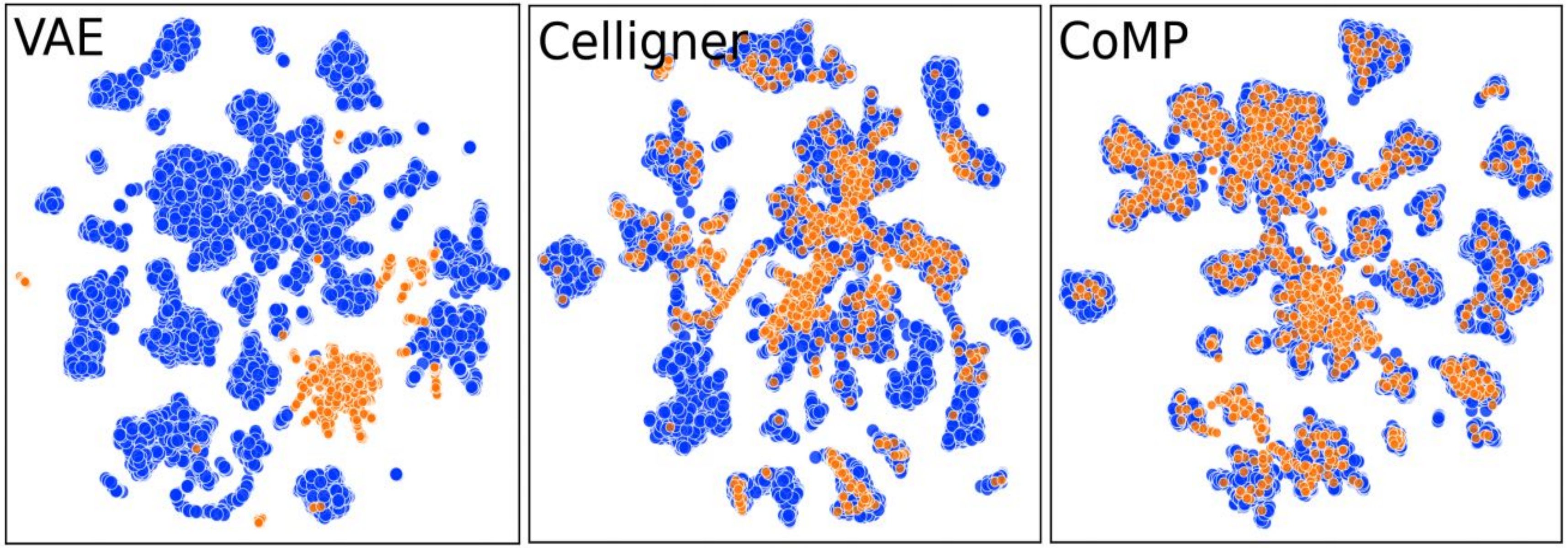
Unique latent space with **z** and c independent

In cell biology, exact symmetries are rarely seen in practice.

Results

1. Powerful data integration tool

Uncorrected | **Partially aligned** | **Our method**



- Tumour samples (TCGA - Largest 15 cancer types)
- Cell lines (CCLE)

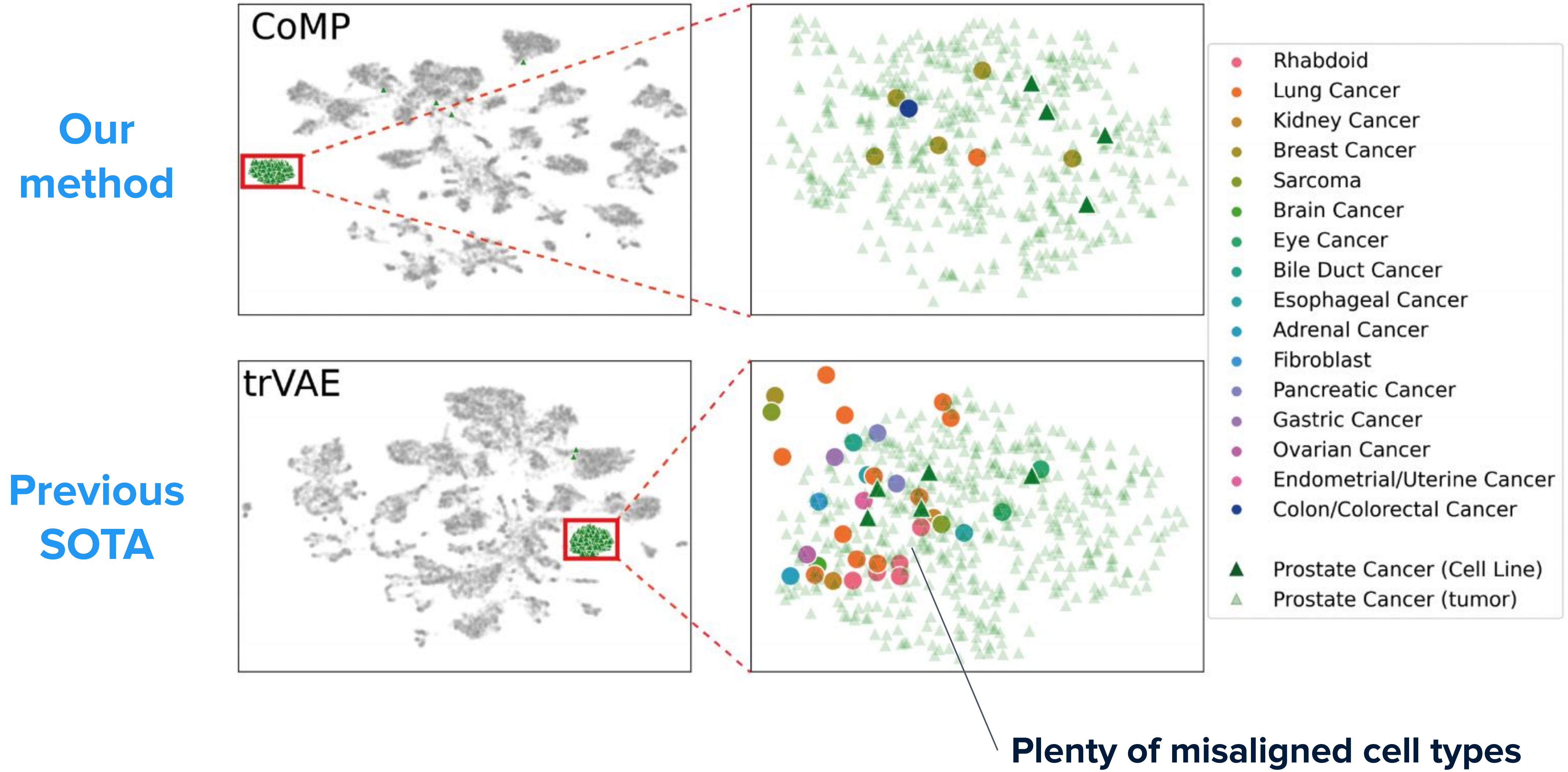
1. Powerful data integration tool

	s	kBET	$\tilde{s}$	m-kBET
VAE	0.658	0.974	0.803	0.581
CVAE	0.554	0.931	0.684	0.571
VFAE	0.168	0.258	0.198	0.188
trVAE	0.096	0.163	0.138	0.123
Celligner	0.082	0.525	0.568	0.226
CoMP	**0.023**	**0.160**	**0.094**	**0.101**

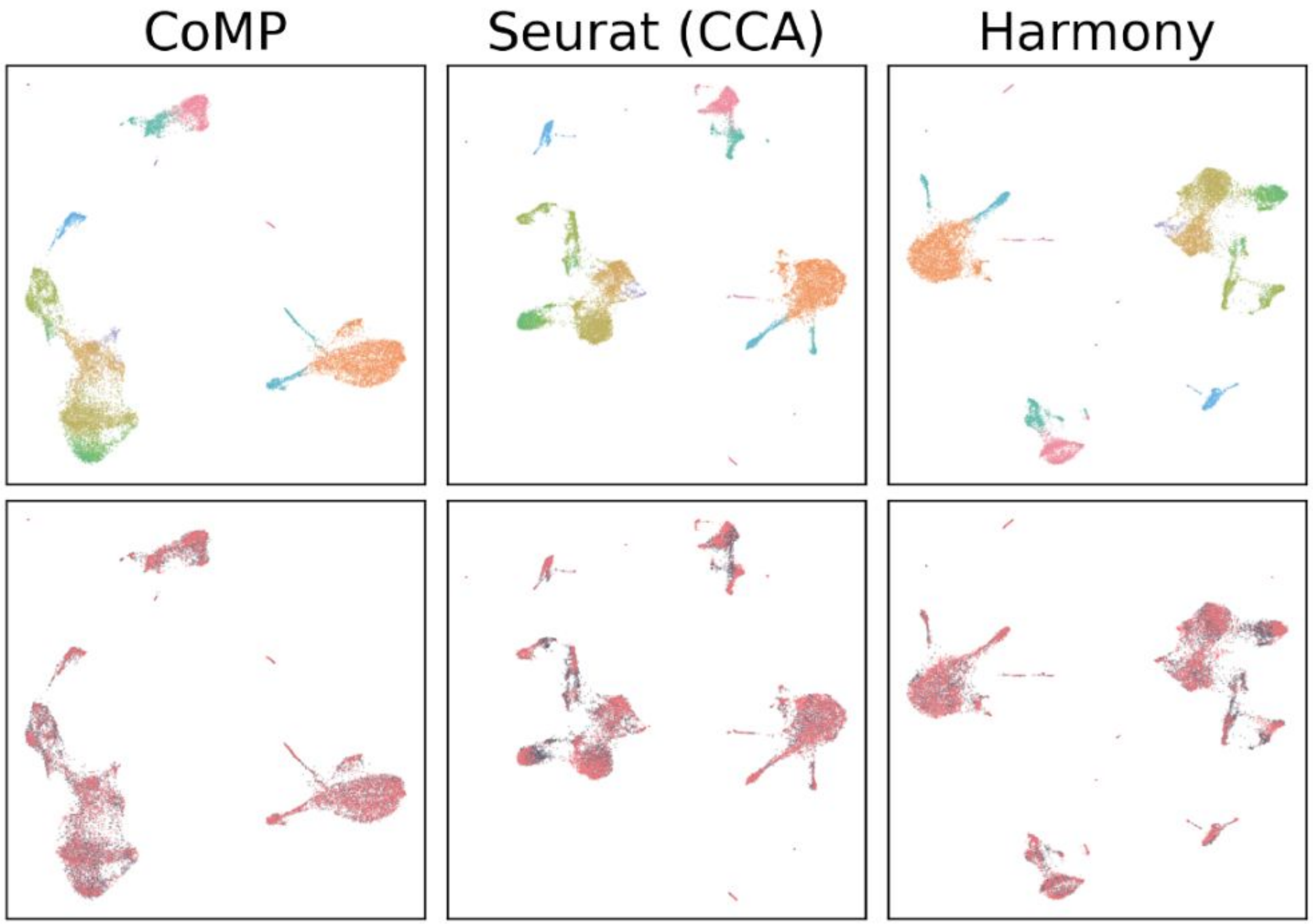
Best global alignment

Best alignment of correctly matching cell types

2. Fewer mis-alignments

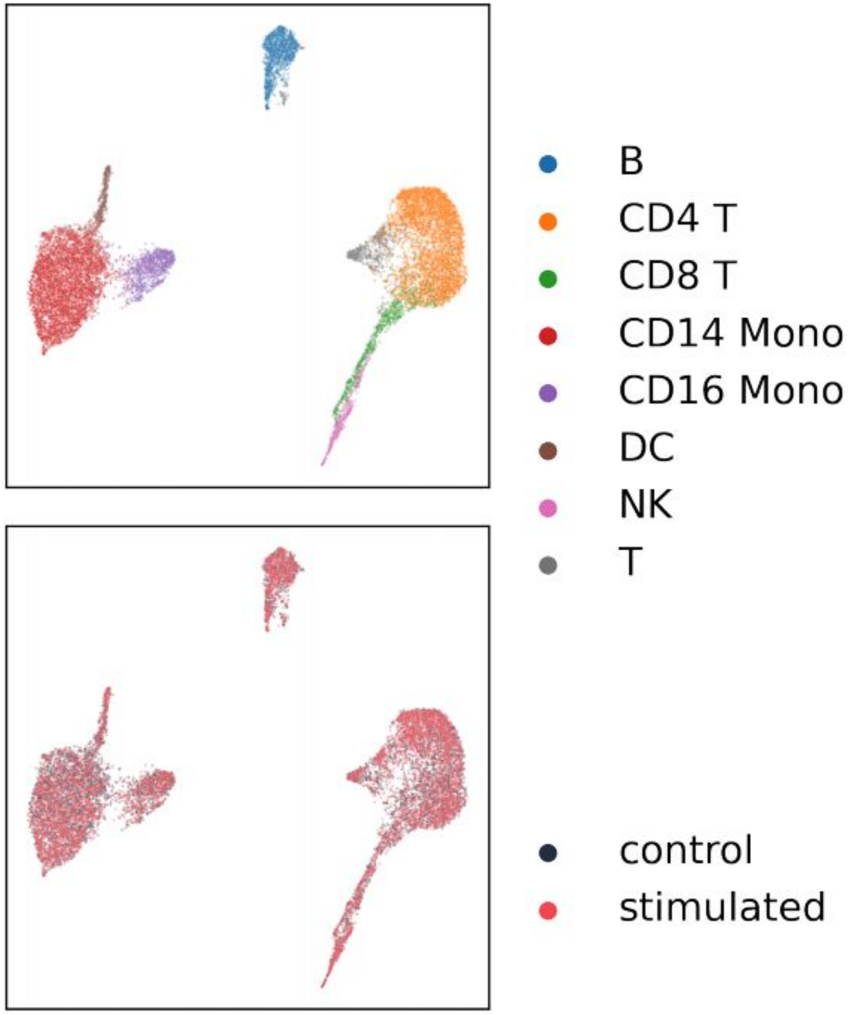


3. Better than Seurat, Harmony for scRNA-seq data integration tasks

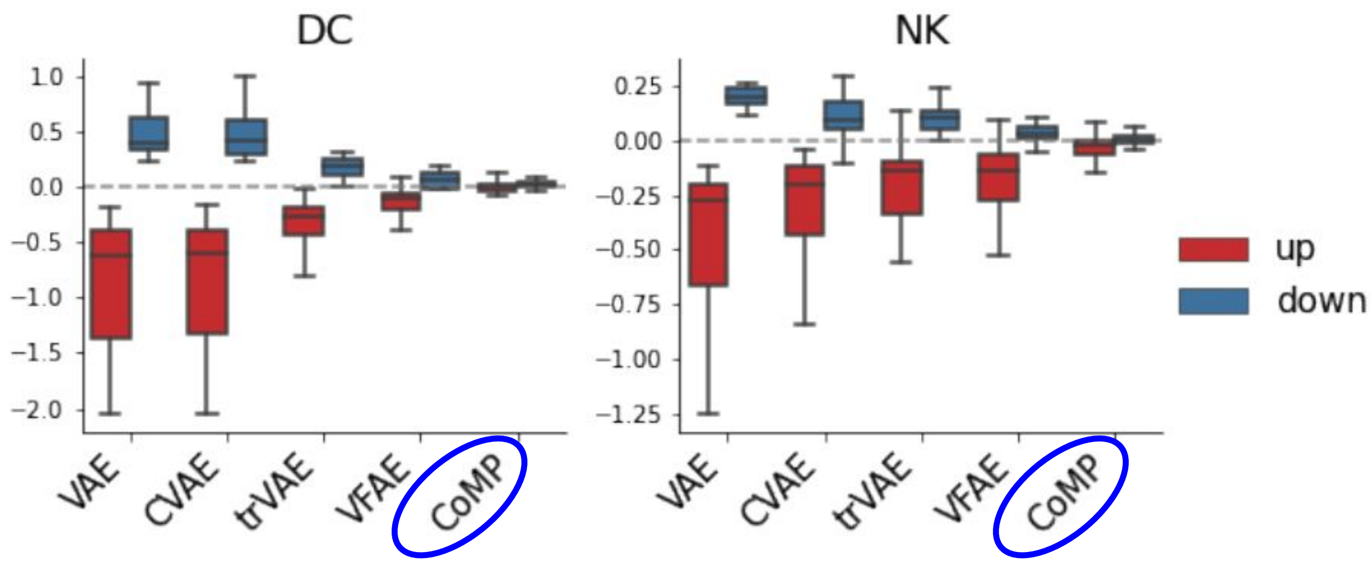


	s	kBET	$\tilde{s}$	m-kBET
Seurat CCA	0.0176	0.436	0.022	0.356
Harmony	0.0158	0.318	0.013	0.245
CoMP	**0.0004**	**0.164**	**0.0011**	**0.120**

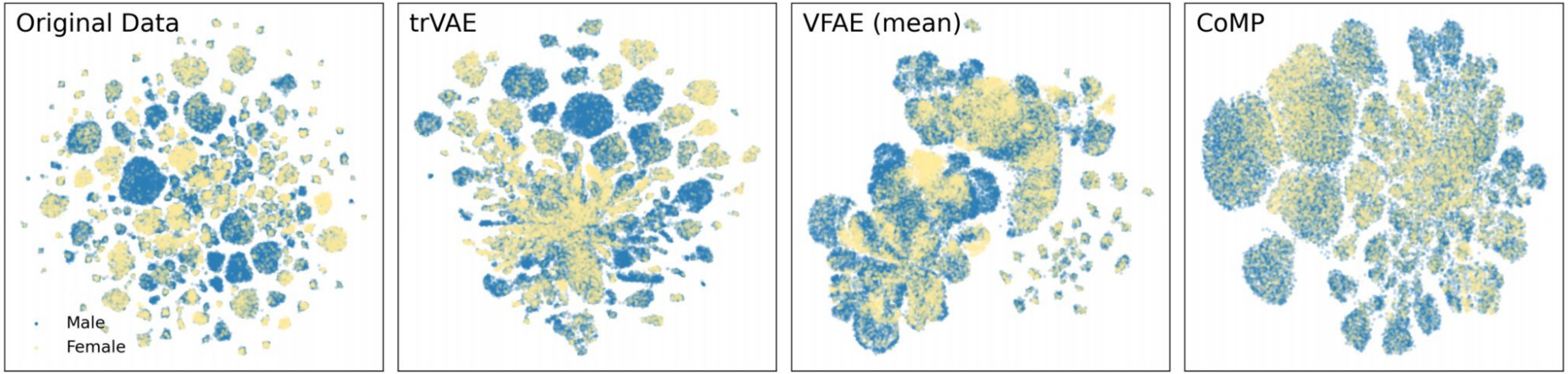
4. Counterfactual inference to predict effects of drug



Counterfactual inference under IFN-beta stimulation



5. Fair but informative representations



	Gender Acc.	Income Acc.	$s_{k,c}$	kBET$_{k,\alpha}$
Original data	0.796	**0.849**	0.067	0.786
VAE	0.764	0.812	0.054	0.748
CVAE	0.778	0.819	0.054	0.724
VFAE (sampled) [18]	0.680	0.815	-	-
VFAE (mean)	0.789	0.805	0.046	0.571
trVAE	0.698	0.808	0.066	0.731
CoMP (ours)	**0.679**	0.805	**0.011**	**0.451**

Thank you!

Árpi Vezér
Craig A Glastonbury
Páidí Creed
Sam Abubudeh
Aaron Sim

Our code is available `github.com/BenevolentAI/CoMP`

Find the paper `"Contrastive Mixture of Posteriors"`

BenevolentAI