Supplementaries

Total loss function for multiDGD

$$-\log p(x, z, \phi, \theta) = \sum\nolimits_{\mathrm{mod}} \mathrm{Loss}_{\mathrm{recon}}^{\mathrm{mod}} + \mathrm{Loss}_{\mathrm{rep}}^{\mathrm{basal}} + \sum\nolimits_{cov} \mathrm{Loss}_{\mathrm{rep}}^{\mathrm{cov}} + \mathrm{Loss}_{\mathrm{prior}}$$

The softball priors

The composition of the Dirichlet prior loss is given as follow:

$$\begin{aligned} &\mathrm{Loss}_{\mathrm{prior}}^{\mathrm{basal}} = -\log p(\phi) = -\log p(\mu, w\,, -\log \Sigma) \\ &= -\log(p(\mu)\, p(w)\, p(-\log \Sigma)) \end{aligned}$$

$$\begin{aligned} p(\mu) &= \prod_k p_{\mathrm{Softball}}(\mu_k|\mathrm{scale}, \mathrm{sharpness}) \\ p(w) &= \prod_k \mathrm{Dir}(\mu_k|\alpha) \\ p(-\log \Sigma) &= \prod_k \prod_l \mathcal{N}\left(-\log \Sigma_{k,l}| -\log 0.2 \times \frac{scale}{K}, 1\right) \end{aligned}$$

Figures

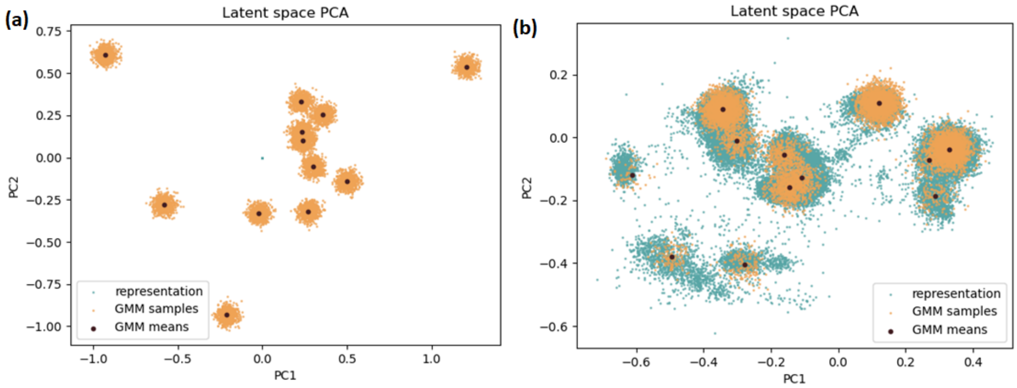

Figure 1: *Latent space visualization. (a) Initialized latent space. (b) Latent space learnt after training.*

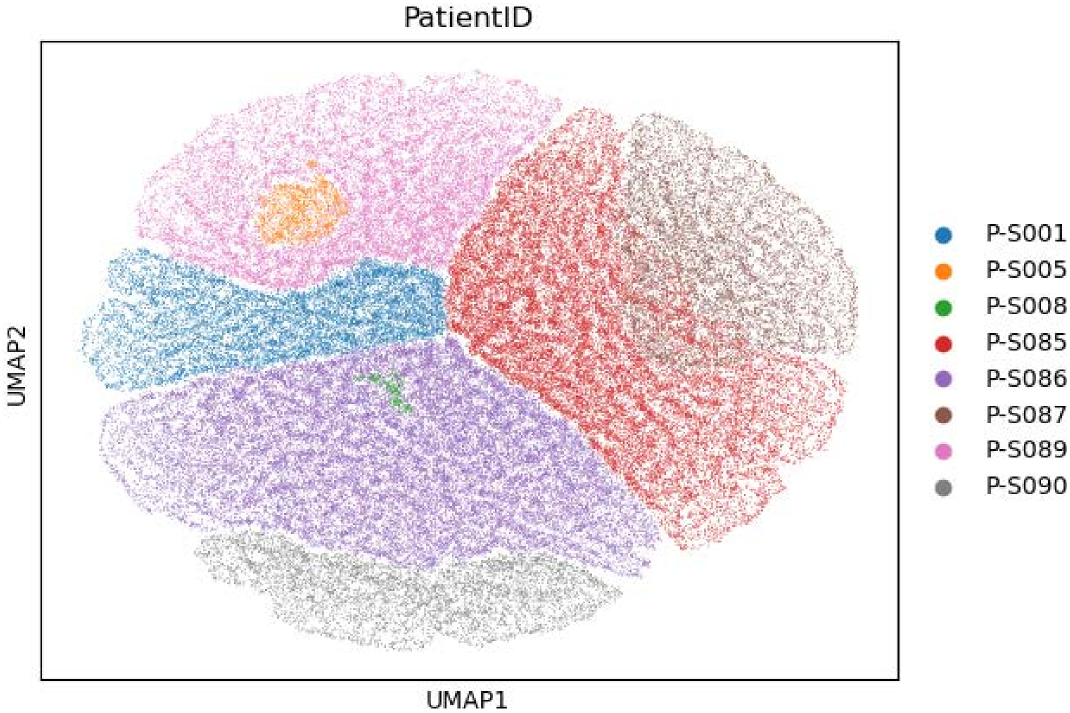

Figure 2: *Latent space learnt for covariate in multiDGD*

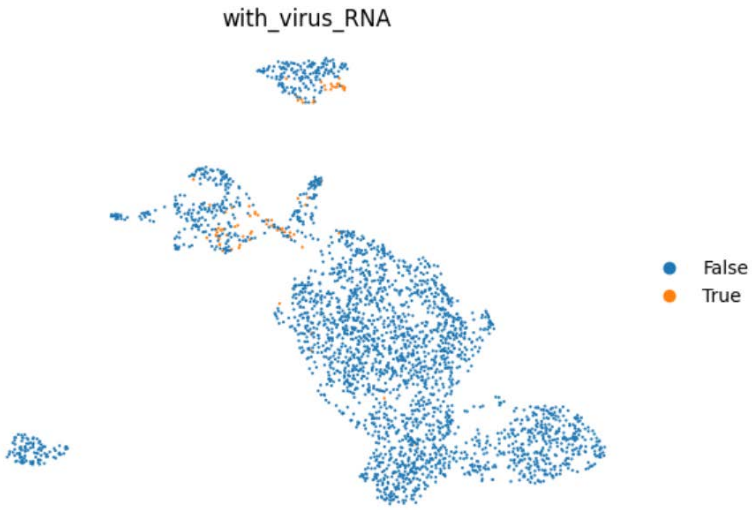

Figure 3: *Testing the model's ability to identify cell's phenotypes like infection status in low-dimensional space*

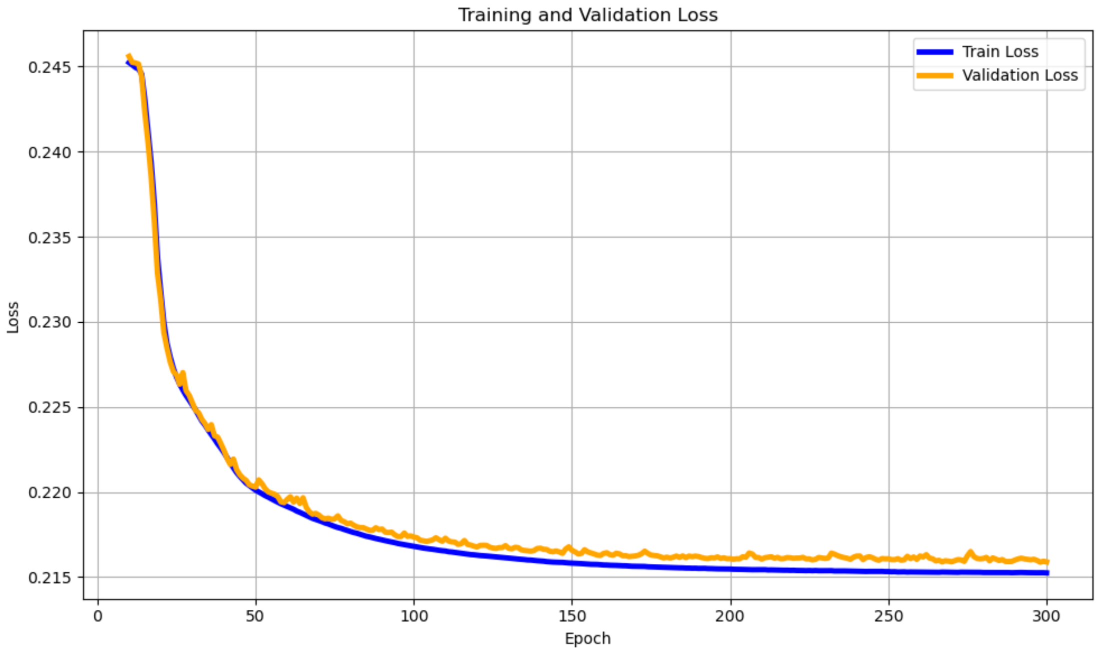

Figure 4: *Training progress of the scDGD model*

Tables

Table 1: DGD Hyperparameters Search

Parameter	Search
Learning rate decay	$1e-4$, $1e-3$
Dirichlet alpha(α)	1,2,5
Hidden layers	20,50,100
Dropout	0.1, 0.2
Standard deviation (GMM)	0.1,0.2,0.5
Batch size	200~800