Identifying Effects of Disease on Single-Cells with Domain-Invariant Generative Modeling

Abdul Moeed
German Cancer Research Center (DKFZ)
Heidelberg, Germany
abdul.moeed@dkfz-heidelberg.de

Martin Rohbeck
German Cancer Research Center (DKFZ)
Heidelberg, Germany
martin.rohbeck@dkfz-heidelberg.de

Kai Ueltzhöffer
European Molecular Biology Laboratory (EMBL)
Heidelberg, Germany
kai.ueltzhoeffer@embl.de

Pavlo Lutsik
German Cancer Research Center (DKFZ)
Heidelberg, Germany
p.lutsik@dkfz-heidelberg.de

Oliver Stegle
German Cancer Research Center (DKFZ)
Heidelberg, Germany
o.stegle@dkfz-heidelberg.de

Marc Jan Bonder
German Cancer Research Center (DKFZ)
Heidelberg, Germany
m.bonder@dkfz-heidelberg.de

Abstract

A core challenge in computational biology is predicting the effects of disease on healthy tissue. From the machine learning perspective, effects of disease and other stimulations on gene expression of single cells can be modeled as a domain shift in a low-dimensional latent space applied to healthy cells. Guided by principles of domain-invariance and compositional models, we present *single-cell Domain Shift Autoencoder (scDSA)*, a deep generative model for disentangling disease-invariant and disease-specific gene programs at single-cell resolution. scDSA uncovers latent factors that are conserved across healthy and disease cell states, and learns how these factors interact with disease. We show that our model i) predicts counterfactual healthy cell-types of diseased cells in unseen patients, ii) captures interpretable representations of disease(s), and iii) learns interaction of disease effects and cell-types. scDSA helps to further our understanding of how diseases perturb healthy tissue on a patient-specific basis therefore enabling advances in personalized healthcare.

1 Introduction

Advancements in RNA sequencing technologies are enabling researchers to measure gene expression levels of thousands of genes in individual cells – giving rise to so-called single-cell RNA sequencing (scRNA-seq) data. As the scale of such data has grown, so has the need to process and extract meaningful information from it. Such high resolution data holds the promise to deepen our knowledge of how certain diseases affect healthy tissue on a cellular level.

Variation in gene expression of cells can be explained by biological processes (cell/tissue-specific functionality) as well as technical artefacts of sequencing (batch effects). Additionally, these sources of variation may be further influenced by disease. While existing machine learning approaches serve adequately to explain latent sources of variation in single-cell gene expression absent of disease

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Table 1: scDSA comparison with related methods.

Method	Count-based	Domain-invariant space	Domain-specific space	OOD predictor
scVI	✓	-	-	-
CPA	✓	-	✓	-
MOFA	✓	✓	-	-
DIVA	-	✓	✓	✓
scDIVA	✓	✓	✓	✓
scDSA	✓	✓	✓	✓

[1, 2, 3], it remains an open challenge to explain how specific diseases impact specific tissue at cellular scale.

Building on existing ideas related to domain-invariant representation learning [4, 5], our approach relies on disentangling disease from healthy signal in cells, by learning invariant representations of cells across healthy and disease states. This allows us to infer disease effect and apply it to healthy cells – subsequently admitting for counterfactual predictions for both healthy and disease cells. To this end, we propose the *single-cell Domain Shift Autoencoder (scDSA)* – a deep generative model which i) learns to remove disease effects from cells in order to learn a domain-invariant latent space, and ii) adds these effects back to the latent space linearly to recover the input. As diseases can affect cell-types differentially, we also model an interaction effect of domain and cell-type in the latent space. To be able to predict counterfactual healthy cell-types of disease cells, we encourage cell-type information to be encoded in the domain-invariant space via an auxillary cell-type prediction task. To achieve domain-invariant representations, we explicitly penalize leakage of domain information in the domain-invariant space through adversarial training. Through various tasks and benchmarks we show that *scDSA* predicts both raw gene expression counts and cell-type labels in unseen disease conditions (out-of-distribution setting), achieving either comparable or superior results to existing methods. Moreover, we analyze disease embeddings learned by *scDSA* and show that they capture meaningful dimensions of diseases.

2 Related work

Due to high-dimensionality of scRNA-seq data, latent variable methods such as factor analysis [6, 3] and variational autoencoders (VAEs) [7, 2] have been popular choices to perform dimensionality reduction. Such procedures not only explain how raw gene expression counts are generated by co-expression patterns but also aid in downstream tasks such as cell-type classification. To account for the multitude of sources of variation in scRNA-seq data – such as batch effects, perturbations and disease – specialized models have been developed on top of the aforementioned techniques (Table 1).

Single-cell variation inference (scVI) [2] is a VAE tailored for scRNA-seq data. Being an amortized version of the standard variational inference approach – it has no mechanism to disentangle sources of variation per se. The compositional perturbation autoencoder (CPA) [8] extends this approach by explicitly modeling domain-specific representations with residual variation pushed in the so-called basal cell state. However, CPA encodes cell-types as covariates instead of encoding cell-type information in the basal state, and as such cannot infer cell-types in OOD samples. Furthermore, it does not model the interaction of cell-types and domains in latent space. MOFA [3] is a popular approach that uses factor analysis to explain shared variation in multi-omic readouts. Domain-specific variation in MOFA is treated as noise thus not modeled explicitly. The domain-invariant variational autoencoder (DIVA) [9] model is a domain generalization method in the machine learning literature. It aims to simultaneously disentangle domain-invariant and domain-specific representations via training a domain-invariant predictor. However, it doesn't enforce independence of domain-invariant and domain-specific information which could leak from one space to the other. To make it amenable for scRNA-seq data, we reimplemented DIVA – in the fashion of scVI – by replacing the Gaussian data likelihood with a Negative Binomial likelihood (typical choice to model gene expression count data [1, 2]), accounting for library size with a learnable parameter and conditioning the encoders and decoder on batch ID to account for batch effects. We henceforth call this version single-cell DIVA (scDIVA). Subsequently, we primarily focus on comparing scDSA to scDIVA.

3 Method

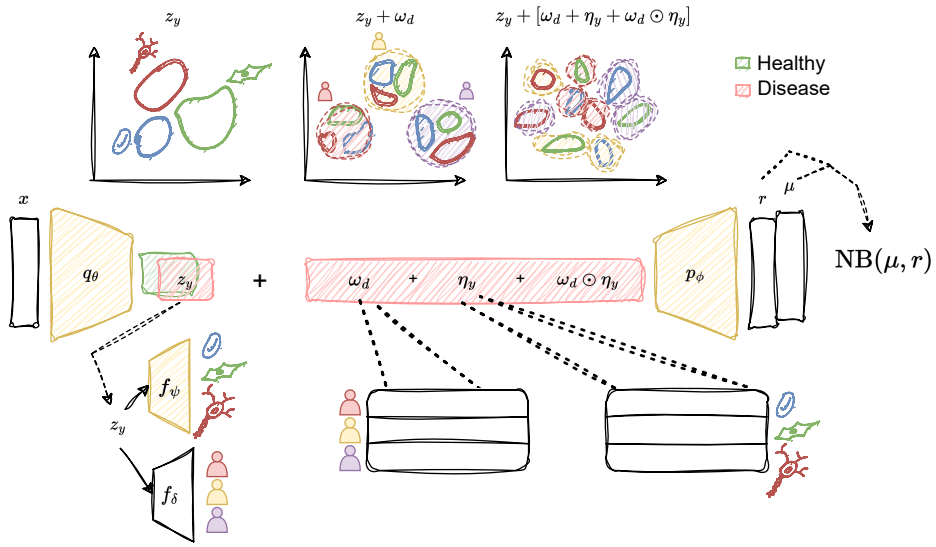


Figure 1: Graphical illustration of the architecture of scDSA. The model learns a domain-invariant representation z_y of the input x, and learns a shift with learnable embeddings ω_d and η_y in case it is perturbed by disease (here depicted as cancer patients). Domain-invariance in z_y is achieved via predictor f_ψ and discriminator f_δ which are trained jointly with the VAE. Top row of UMAP plot sketches how each part of the latent space is expected to look like.

3.1 Model

Given observations in the form $\mathcal{D} = (x^i, y^i, d^i)$, where $x^i \in \mathbb{N}_0^D$ is the vector of expression counts of D genes for sample i, y^i and d^i are discrete cell-type and domain labels, we are interested in learning a domain-invariant representation $z_y^i \in \mathbb{R}^K$ of x^i such that $z_y^i \perp\!\!\!\perp d^i$. In the following we drop index i for readability. We assume that the observed gene counts in a cell are generated by low-dimensional latent processes. For cells that are in the healthy state, these processes correspond to active cell-type specific gene programs z_y. Cells that are affected by disease, are assumed to have a perturbation applied to z_y. This is formalized as follows:

$$z = z_y + \mathbb{1}_{d \neq \text{healthy}} \Big[\omega_d + \eta_y + \omega_d \odot \eta_y \Big], \tag{1}$$

where $z \in \mathbb{R}^K$ represents x in latent space, z_y is the domain-invariant representation, $\mathbb{1}_{d \neq healthy}$ is an indicator function which denotes that the domain shift terms are only added to cells not in the *healthy* domain, $\omega_d \in \mathbb{R}^K$ is the shift vector for disease d, $\eta_y \in \mathbb{R}^K$ is the shift vector for cell-type y and $\odot$ denotes the Hadamard-product of ω_d and η_y. We assume that ω_d captures global shift properties of domain d independent of cell-type y, while η_y captures global shift effects for each cell-type y regardless of domains. As we are dealing with count data, observed gene count vector x follows a Negative Binomial distribution. As scRNA-seq readouts depend on the so-called library size (larger cells correspond to more gene counts), we need to account for this artefact. Similar to [2], it is modeled as a scalar $l \in \mathbb{R}$ sampled from a log normal distribution:

$$l \sim \text{log normal}(l_\mu(z_l), l_\sigma(z_l)), \tag{2}$$

$$x \sim \text{NB}(\mu(z) * l, r(z)), \tag{3}$$

where $\mu(z) * l \in \mathbb{R}^D$ and $r \in \mathbb{R}^D$ are the mean and dispersion parameters respectively (Figure 1).

3.2 Inference

In order to learn the aforementioned latent representation of the data, we use a VAE-based model. We train an encoder $q_\theta(z_y|x)$ and decoder $p_\phi(x|z)$ as inference and generative models respectively.

The decoder outputs parameters of the aforementioned Negative Binomial distribution. To enforce cell-type information being encoded in z_y, we put a conditional prior on the variational posterior, also parametrized as an encoder $p_{\theta_y}(z_y|y)$. The variational lower bound thus becomes:

$$\text{ELBO} = \log p_\phi(x|z) - \beta_y KL(q_{\theta_{z_y}}(z_y|x)||p_{\theta_y}(z_y|y)) - \beta_l KL(q_{\theta_{z_l}}(z_l|x)||p_{\theta_l}(z_l)) \tag{4}$$

The posterior for library scale is inferred via encoder $q_{\theta_{z_l}}(z_l|x)$ with prior $p_{\theta_l}(z_l)$ instantiated as $\mathcal{N}(\mu_l, \sigma_l)$ where μ_l and σ_l are the empirical mean and variance of the sequencing batch. The embeddings ω and η are optimized as free parameters.

To enforce domain invariance in z_y, a GAN-style strategy [10] is employed: we alternate between 1) optimizing a discriminator f_δ which takes as input the domain-invariant representation z_y and predicts domain d, and 2) fixing f_δ, computing cross-entropy loss, subtracting it from the VAE loss and optimizing the VAE. Step 1 encourages f_δ to improve prediction of domains from z_y while step 2 ensures that the VAE is penalized in proportion to accuracy achieved by f_δ. Finally, to further enforce cell-type encoding in z_y, a predictor f_ψ is trained in conjunction with the VAE to predict cell-type y from z_y. The complete objective function for *scDSA* is thus:

$$J = \text{ELBO} + \alpha_y \log f_\psi(y|z_y) - \alpha_d \log f_\delta(d|z_y), \tag{5}$$

where α_y and α_d are regularization parameters for the predictor and discriminator respectively, with positive support. In cases where certain training data is missing cell-type annotations, we propose a semi-supervised approach similar to [9]:

$$z = z_y + \mathbb{1}_{d \neq \text{healthy}} \Big[\sum_y f_\psi(y|z_y)(\omega_d + \eta_y + \omega_d \odot \eta_y) \Big], \tag{6}$$

where we impute cell-type information by pseudo-labeling cells using f_ψ. When training on a semi-supervised batch, only the parameters of the VAE and discriminator are optimized while f_ψ is kept fixed, in contrast to [9] where the predictor is also optimized.

For interpretation of disease embeddings compared to control samples, we initialize the embedding corresponding to control domain as a zero vector and disable gradient updates on this embedding.

4 Results

We evaluate various aspects of model learning via different tasks: 1) transfer of labels to an unseen experiment (referred to as query-to-reference data mapping), 2) inference of leukemic cells from unseen cancer patients, and 3) inference of cells from an unseen perturbation (COVID) while trained on control and Influenza A cells. For comparison, we use a vanilla VAE from the scVI package as baseline. In addition, we use our scDIVA implementation which explicitly tries to disentangle domain-specific variation from shared variation. We observe that scDIVA and scDSA have comparable data fit on novel domains while outperforming VAE. Furthermore, adding the domain-celltype interaction to scDSA further improves generalization (Table 2). For OOD cell-type prediction, we use the trained predictors of scDIVA and scDSA, while for scVI, a neural network is trained on its latent representations. To avoid circularity of cell-type predictions from gene expression, all datasets used in our experiments derive cell-type labels from cell surface protein expression (CITE-seq).

For proof-of-concept, we start with mapping cells from unseen experiments onto cells of known cell-types (query-to-reference mapping). The goal is to learn representations of cells such that they generalize to data generated by unseen experiments. For this, we take healthy human bone marrow (BMMC) and peripheral blood (PBMC) data from the Azimuth annotated reference atlas [11], that includes cells from four experiments. Furthermore, we add healthy samples from [12] where the authors study gene expression variation in patients with Acute Myeloid Leukemia (AML) – more on this in Section 4.1. To evaluate how the model fits cells from unseen experiments, we holdout data from one study during training. As seen in Table 2, scDSA out-performs scVI and has comparable performance to scDIVA on data fit of the holdout set. Additionally, we assess the impact of interaction of domain and cell-types in our model, where we see that adding the interaction term further decreases the negative log-likelihood (NLL) of holdout data.

4.1 OOD prediction of cancer cell states

Next, in order to assess generalization of the model in cells with disease, we use the entire AML dataset [12] which contains CITE-seq derived cell-type labels and clonal information identified by

Table 2: Mean negative log-likelihood on OOD holdout sets (lower is better). Values have been multiplied by a factor of 100 for readability. *scDSA-interaction* refers to model with disease-celltype interaction included.

Task	Holdout	scVI	scDIVA	scDSA	scDSA-interaction
Query-to-reference	Study	31.2 ± 0.4	15.5 ± 0.2	17.3 ± 0.5	16.6 ± 1.3
AML	Patients	62.1 ± 1.0	23.4 ± 0.2	24.1 ± 1.1	25.5 ± 1.7
COVID/IAV	Condition	45.1 ± 0.5	17.6 ± 3.4	19.8 ± 0.3	19.1 ± 0.4

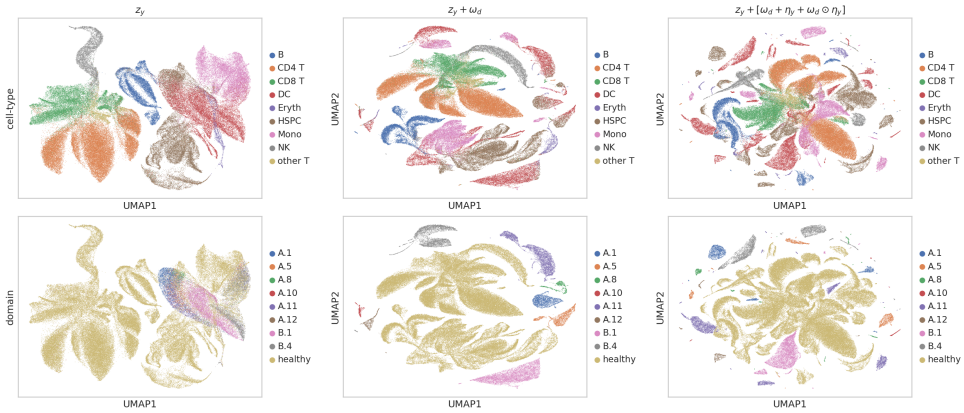


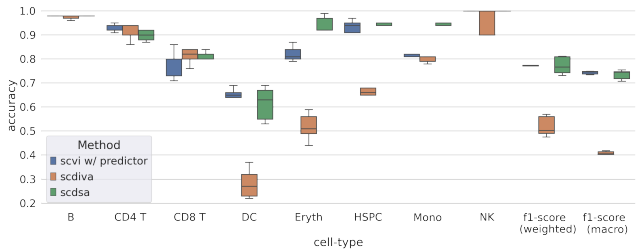
Figure 2: UMAP visualizations of latent spaces learned by scDSA on AML data. Top row: Colored by cell-type, bottom row: Colored by domain. Each column depicts how domain shifts (with and without cell-type interaction) affect latent space. Left-most column: domain-invariant representation z_y, Center column: Domain-shift vectors ω_d added to z_y, Right-most column: Full latent state with interaction of domain and cell-types.

mitochondrial mutations. The dataset comprises of cells from two healthy donors, as well as a mixture of healthy and leukemic cells from 13 AML patients. Data is split into three sets: training (8 leukemic + control donors), validation (3 patients) and holdout (2 patients). Healthy cells, regardless of patient/donor, are assigned a single domain ('*healthy*'). AML cancer clones are highly heterogeneous among patients; each patient is therefore treated as a separate domain. The latent space of the model consists of 64 latent dimensions.

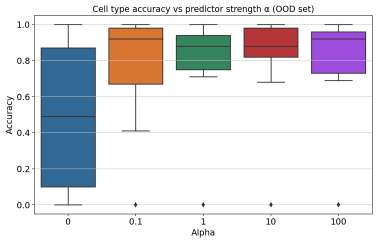
For qualitative assessment of the learned latent space of scDSA, we plot UMAP visualizations of the training data as seen in Fig 2. As can be obvserved, predominantly cell-type information is encoded in the domain-invariant space z_y, regardless of the domain, as we aimed with this model. Next, we add the clone/disease embeddings ω to each cell (column 2), where we see that patient-specific clusters are formed. Finally, we add the interaction of clone and cell-types to z_y to observe that clusters are now both clone- and cell type-specific.

To quantify how well scDSA generalizes to unseen clones, we infer counterfactual healthy cell-types of leukemic cells for samples of holdout patients. Due to imbalance of cell type abundance, the results are stratified by cell types and reported in Fig 3a. As we can see, scDSA has an overall lead over competing methods, while also being robust across cell types. In contrast, scDIVA is not able to generalize to unseen clones. We hypothesize that this is due to scDIVA not explicitly penalizing domain information flow into the domain-invariant space.

As an ablation for whether it helps to jointly train f_ψ as opposed to VAE having no supervision signal for encoding cell-type information in z_y, we compare effects on OOD prediction of different values of predictor strength α_y (Fig 3b). As we can see, having no predictor severely reduces OOD accuracy of cell-type prediction. Interestingly, increasing α_y exponentially has modest returns on this dataset.



(a) Cell-type prediction accuracy of leukemic cells from holdout patients.



(b) Effect of predictor strength α_y on average OOD accuracy.

Figure 3: Summary of prediction accuracy on Acute Myeloid Leukemia (AML) datasets

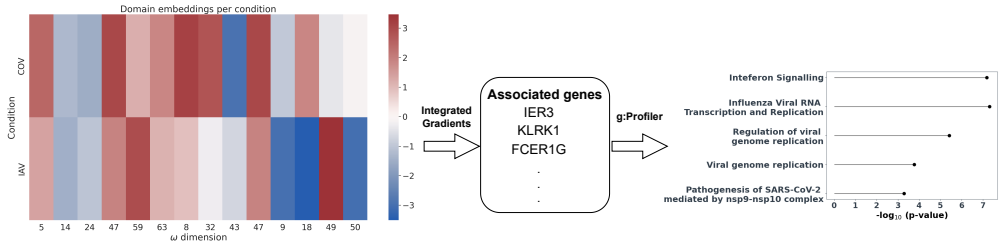


Figure 4: Relevant latent dimensions of disease conditions – SARS-CoV-2 (COV), Influenza A virus (IAV) – and top-n genes associated with each. Selected genes were queried with g:Profiler for known pathways they are involved in.

4.2 Interpretation of disease embeddings

To interpret disease embeddings learned by scDSA, we use an scRNA-seq dataset with healthy controls along with cells stimulated with Influenza A virus (IAV) and SARS-CoV-2 (COV) [13]. Our goal here is to identify latent factors in disease embeddings ω that are either highly active or inactive compared to control samples.

We train scDSA with 64 latent dimensions and three domains (not-stimulated (NS), IAV, COV). As mentioned in Section 3, the control embedding is set to a null vector. We then identify latent dimensions of ω_{COV} and ω_{IAV} which have high positive or negative entries (Fig 4). Some dimensions have high/low values for both diseases (e.g. dimensions 5, 14, 24) while others are relevant for only a single condition (e.g. dimension 50). To find out which genes are highly associated with each of these latent dimensions, we run Integrated Gradients [14]. Finally, to interpret which signalling pathways these genes might play a role in, we use g:Profiler [15] to search for biological enrichments of genes important in the latent factors. Controlling for background genes expressed in the dataset, we find that the query genes are indeed involved in specific COVID-19 and IAV-related mechanisms, as well as more general viral pathways such as 'viral genome replication' (Fig 4).

5 Discussion

In this work, we propose scDSA – an approach to disentangle and identify disease-induced effects on gene expression of single-cells. scDSA uses a compositional latent space where first, disease effects are removed via adversarial training, and then linearly added back to an otherwise healthy representation of a cell. It also models how specific interaction of a cell-type and disease might effect this perturbation. We evaluate scDSA on various tasks: mapping a holdout dataset onto a reference, learning clone-invariant and clone-specific states of AML cells as well as modeling cellular responses to SARS-CoV-2 and Influenza A virus. While primarily focused on scRNA-seq data, we note that scDSA can more generally be adapted to other omics data with minor modifications.

One major drawback of scDSA pertains to the complexity of model training. As scDSA is composed of many separate components (VAE, discriminator, predictor) with each having its own set hyperpa-

rameters, it might be a challenge to find the optimal set for certain tasks. Although in our experiments minimal hyperparameter tuning was conducted, as most hyperparameters were kept fixed for all tasks, in theory the choice of such variables can drastically alter model convergence rates.

Future work will focus on further understanding differential transformation of cell-types by known mechanisms of certain diseases. In the case of cancer, for example, the domain shift embedding ω could be modeled as a function of cellular copy number variation (CNV), mutational burden or DNA methylation status. Another promising direction is incorporating single-cell spatial readouts. With true single-cell spatial transcriptomic technologies such as Xenium on the rise, scDSA could be modified to encode spatial dependencies of cells via Graph Neural Networks [16]. This is especially useful in diseases where spatial architecture of cell systems plays an important role, as has been demonstrated for breast cancer [17].

Acknowledgments and Disclosure of Funding

We would like to thank the Helmholtz Information and Data Science School for Health (HIDSS4Health) in addition to the German Cancer Research Center (DKFZ) for providing funding for this project.

References

[1] Gökcen Eraslan, Lukas M Simon, Maria Mircea, Nikola S Mueller, and Fabian J Theis. Single-cell rna-seq denoising using a deep count autoencoder. *Nature communications*, 10(1):390, 2019.

[2] Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

[3] Ricard Argelaguet, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C Marioni, and Oliver Stegle. Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome biology*, 21(1):1–17, 2020.

[4] Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-supervised learning with data augmentations provably isolates content from style, 2022.

[5] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

[6] Florian Buettner, Naruemon Pratanwanich, Davis J McCarthy, John C Marioni, and Oliver Stegle. f-sclvm: scalable and versatile factor analysis for single-cell rna-seq. *Genome biology*, 18:1–13, 2017.

[7] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[8] Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, Leon Hetzel, Yuge Ji, Ignacio L Ibarra, Sanjay R Srivatsan, Mohsen Naghipourfar, Riza M Daza, Beth Martin, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology*, page e11517, 2023.

[9] Maximilian Ilse, Jakub M Tomczak, Christos Louizos, and Max Welling. Diva: Domain invariant variational autoencoders. In *Medical Imaging with Deep Learning*, pages 322–348. PMLR, 2020.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[11] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.

[12] Sergi Beneyto-Calabuig, Anne Kathrin Merbach, Jonas-Alexander Kniffka, Magdalena Antes, Chelsea Szu-Tu, Christian Rohde, Alexander Waclawiczek, Patrick Stelmach, Sarah Gräßle, Philip Pervan, et al. Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. *Cell Stem Cell*, 30(5):706–721, 2023.

[13] Yann Aquino, Aurélie Bisiaux, Zhi Li, Mary O'Neill, Javier Mendoza-Revilla, Sarah Hélène Merkling, Gaspard Kerner, Milena Hasan, Valentina Libri, Vincent Bondet, et al. Dissecting human population variation in single-cell responses to sars-cov-2. *Nature*, pages 1–9, 2023.

[14] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[15] Uku Raudvere, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, and Jaak Vilo. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research*, 47(W1):W191–W198, 2019.

[16] David S Fischer, Anna C Schaar, and Fabian J Theis. Modeling intercellular communication in tissues using spatial graphs of cells. *Nature Biotechnology*, 41(3):332–336, 2023.

[17] Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, et al. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature*, 611(7936):594–602, 2022.

[18] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19:1–5, 2018.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Appendix

Data Pre-processing

1000, 1000, 3000 genes were selected for query-to-reference mapping, AML prediction and COVID/IAV tasks respectively using scanpy's [18] `sc.pp.highly_variable_genes()` function. Label matching for Azimuth and AML datasets were done manually. For AML task, cells with uncertain cancer/healthy label were dropped.

Model Architectures

All weights are initialized with `torch.nn.init.xavier_uniform_()`.

Encoders

Table 3: Architecture of encoders used in scDSA and scDIVA

Layer	Details (type, activation)
1	Linear(n-genes, 600), ReLU
2	Linear(600, 200), ReLU
3.1	Linear(200, 64), -
3.2	Linear(200, 64), Softplus

Decoders

Table 4: Architecture of decoders used in scDSA and scDIVA. 3.2 refers to dispersion parameter of NB distribution which is shared across cells, therefore specified as a learnable tensor of size *n-genes*.

Layer	Details (type, activation)
1	Linear(64, 400), SELU
2	Linear(400, 800), ReLU
3.1	Linear(800, n-genes), Softmax
3.2	Parameter(n-genes), -

Classifiers

The same architecture is used for predictors and discriminator. *output-dim* corresponds to *n-domains* for discriminator and to *n-celltypes* for predictors.

Hyperparameters

Adam [19] optimizer is used for all parameter learning. To account for scale imbalance in loss terms, we run a warmup epoch without optimization to evaluate normalization constants for each loss component. Each loss term (VAE, predictor, discriminator) is then normalized by corresponding constants during training. Each model is trained for `max_epochs=200` with early stopping enabled.

Table 5: Architecture of classifiers used in scDSA and scDIVA

Layer	Details (type, activation)
1	Linear(64, 28), ReLU
2	Linear(28, 16), ReLU
3	Linear(16, output-dim), -

Table 6: Hyperparameters for scDSA training runs across tasks

Hyperparameter	Value
α_y	10
α_d	1
β_y	1
β_l	1
learning-rate$_{\text{VAE}}$	1e-3
learning-rate f_ψ	1e-2
learning-rate f_δ	1e-2

Table 7: Hyperparameters for scDIVA training runs across tasks.

Hyperparameter	Value
α_y	4800
α_d	2000
β_y	1
β_d	1
β_l	1
learning-rate$_{\text{VAE}}$	1e-3
learning-rate f_y	1e-3
learning-rate f_d	1e-3