UNSUPERVISED DEEP DISENTANGLED REPRESENTATION OF SINGLE-CELL OMICS WITH DRVI

Amir Ali Moinfar
School of Computing, Information and Technology, Technical University of Munich, Garching, Germany.
Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany.

Fabian J. Theis *
School of Computing, Information and Technology, Technical University of Munich, Garching, Germany.
Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany.
TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany.

ABSTRACT

Single-cell genomics allows for the unbiased exploration of cellular heterogeneity. Representation learning methods summarize high-dimensional single-cell data into a manageable latent space in a typically nonlinear fashion, allowing cross-sample integration or generative modeling. However, these methods often produce entangled representations, limiting interpretability and downstream analyses. Existing disentanglement methods instead either require supervised information or impose sparsity and linearity, which may not capture the complexity of biological data. We, therefore, introduce Disentangled Representation Variational Inference (DRVI), an unsupervised deep generative model that learns nonlinear, disentangled representations of single-cell omics. This is achieved by combining recently introduced additive decoders with nonlinear pooling, for which we theoretically prove disentanglement under reasonable assumptions. We validate DRVI's disentanglement capabilities across diverse relevant biological problems, from development to perturbational studies and cell atlases, decomposing, for example, the Human Lung Cell Atlas into meaningful, interpretable latent dimensions. Moreover, we demonstrate that if applied to batch integration, DRVI's integration quality does not suffer from the disentanglement constraints and instead is on par with entangled integration methods. With its disentangled latent space, DRVI is inherently interpretable and facilitates the identification of rare cell types, provides novel insights into cellular heterogeneity beyond traditional cell types, and highlights developmental stages.

1 INTRODUCTION

Single-cell RNA-seq technologies are widely used for profiling transcriptomes or other omics levels at single-cell resolution in a high-throughput fashion Svensson et al. (2018); Angerer et al. (2017). Consequently, many tools and methods have emerged to facilitate the analysis of these datasets Zappia & Theis (2021). Among these, generative models have effectively summarized data within a latent space, particularly when dealing with multiple batches Lopez et al. (2018); Lotfollahi et al. (2019), outperforming more traditional methods in large-scale benchmarks and competitions for integration and cross-modality inference Lance et al. (2022).

Despite their widespread use, generative models commonly produce entangled latent representations, where multiple unrelated biological processes are intertwined within a single dimension Kunes et al. (2023). This limits the interpretability of the latent space, restricting its direct application for downstream tasks Svensson et al. (2020); DeTomaso & Yosef (2021). As a result, the latent space is often only used as an intermediate, and its utility is limited to neighborhood graphs constructed on top of the cellular representations. Specifically, common techniques like Uniform manifold approximation and projection (UMAP) McInnes et al. (2018); Becht et al. (2018) for visualization and

*Correspondence to: fabian.theis@helmholtz-muenchen.de

Leiden clustering Traag et al. (2019) for cell type identification exclusively rely on these neighborhood graphs. While clustering methods are essential for extracting biological variations from the high-dimensional single-cell omics data Kiselev et al. (2019), they have limitations. They cannot capture continuous processes or identify shared programs among multiple cell types as a single concept and often identify the most dominant signals, such as cell types, which makes it challenging to uncover secondary signals. Disentanglement effectively addresses these limitations by definition.

Disentanglement refers to the ability of a model to separate and represent distinct underlying processes in individual latent dimensions. Several disentangled latent space models have been proposed for single-cell omics data. However, they either rely on linear assumptions Argelaguet et al. (2020); Hackenberg et al. (2024), auxiliary data, such as variations of interest Lopez et al. (2022); Bereket & Karaletsos (2023); Hediyeh-zadeh et al. (2024), or both Lotfollahi et al. (2023); Gut et al. (2021); Nazaret et al. (2023); Kunes et al. (2023) to enforce disentanglement. In practice, for new discoveries, supervised data is absent or, if present, frequently biased. And, linear models fail to capture the complexity of biological data, particularly in scenarios involving multiple batches. These limitations motivate the development of unsupervised nonlinear disentangled representation learning. While there are a few unsupervised and unconstrained disentanglement models Lotfollahi et al. (2021); Yu & Welch (2021), we will experimentally show none of these unconstrained models demonstrate satisfactory disentanglement performance, which may explain the fact that they see little practical use.

To address these challenges, we leverage and extend the recently proposed framework of additive decoders Lachapelle et al. (2023), which enables latent variable identification in the context of image generation. We present Disentangled Representation Variational Inference (DRVI), an unsupervised generative model capable of learning non-linear interpretable disentangled latent representations from single-cell count data. DRVI is built upon additive decoders followed by a non-linear pooling function. We theoretically prove that this architecture achieves disentanglement under specific assumptions: additivity of independent underlying processes in count space and the existence of markers for each process. Empirical evaluations across a diverse range of biological applications demonstrate DRVI's capabilities in learning disentangled latent representations, outperforming existing methods in terms of disentanglement without reducing integration quality. DRVI facilitates the identification of biological states by representing cell types across individual latent factors in cell atlases, enables exploration of cellular heterogeneity by identifying biologically meaningful variations beyond cell types, captures developmental stages in developmental datasets, identifies coherent groups of effective perturbations in perturbational studies. Additionally, DRVI allows users to selectively highlight or exclude latent factors associated with specific known variations, offering flexible post-hoc analysis capabilities.

2 Methods

2.1 DRVI: Disentangled Representation Variational Inference

DRVI is an unsupervised deep generative model that employs log-sum-exp (LSE) pooled additive decoders to disentangle latent factors. Additive decoder architecture has been shown to enforce disentanglement under specific assumptions Lachapelle et al. (2023). However, without additional modifications, it does not achieve disentangled latent representations for single-cell omics data. To address this, we combine additive decoders with variational autoencoders and incorporate multiple modifications that achieve both theoretical and practical disentanglement of latent factors. Most importantly, we introduced LSE aggregation, which, from a mechanistic point of view, assumes that biological processes add up in count space (Supplementary Figure 1).

2.2 The Generative model

Assume we have an observed count matrix $\boldsymbol{X} \in \mathbb{N}^{N \times M}$, where N is the number of cells and M is the number of genes. Optionally, we can consider a covariate vector $\boldsymbol{C} \in \mathcal{R}^{N \times R}$, where R is the dimensionality of the covariates for each cell. For example, in batch correction frameworks, $\boldsymbol{C}_i$ could be a one-hot vector representing the batch identity if modeled as in scVI Lopez et al. (2018), or an embedding vector if modeled as in scPoli De Donno et al. (2023) (DRVI accepts both

conventions). Then, the goal is to discover latent factors $\boldsymbol{Z} \in \mathcal{R}^{N \times K}$ where K is the latent space dimension.

DRVI accepts Normal, Poisson, and Negative Binomial distributions as noise models. When using the Normal noise model, the observed data (x) should be log-transformed. We primarily parametrize the distributions in log space to ensure a correct and clear formulation within our framework. This approach also emphasizes that all other methods calculate their parameters in log space and transform them back to count space in the last step.

The generative process is defined by the pooling of multiple decoders that map latent space to parameters of the noise distribution. Specifically:

$$Z \sim Normal(0, I) \tag{1}$$

$$\log \mu = f(Z, C) = \sigma(\sum_{1 \leq k \leq K} \psi \circ f^{(k)}(Z_k, C)) \tag{2}$$

$$X_g \sim Dist(\log \mu_g + \log L, \log \theta_g) \tag{3}$$

Where f is the decoder, $f^{(k)}$ denotes the additive decoder subunits (referred to as basis functions), σ and ψ form pooling function, $Dist$ models the noise distribution, θ_g is a parameter indicating gene-wise dispersion or variance, L corresponds to the random variable indicating library size of the cell, and Z_k denotes kth dimension of Z.

Fitting the generative process given observed data involves maximizing the log-likelihood function $\log \mathbb{P}_\theta(X|C) = \sum_n \log \mathbb{P}_\theta(\boldsymbol{X}_n|\boldsymbol{C}_n)$ where $\boldsymbol{X}$ is the matrix of observations, C indicates covariates matrix (such as technical batch), and θ represents all parameters in the model. Although the log-likelihood decomposes into sample log-likelihoods, maximizing each term requires integration over all Z values, which is intractable. We employ Variational Bayes, which constructs an approximate posterior probability distribution function $q_\phi(Z|X, C)$ parameterized by ϕ Kingma & Welling (2013). This distribution is typically formulated as a simple multivariate Gaussian distribution.

Similar to other frameworks, the ELBO is optimized by constructing an autoencoder where $\mathbb{P}_\theta(X|Z, C)$, $q_\phi(Z|X, C)$, and $\mathbb{P}(Z)$ correspond to the decoder, the encoder, and the prior distribution, respectively Lopez et al. (2018); De Donno et al. (2023); Lotfollahi et al. (2020). Stochastic gradient descent is used to optimize the negative ELBO, and the reparametrization trick is employed to sample from the approximate posterior q Kingma & Welling (2013); Kingma & Ba (2014).

2.3 Additive Decoders, Pooling, and Disentanglement

We build on the theoretical foundations of additive decoders Lachapelle et al. (2023) to establish disentanglement criteria within DRVI. Our primary methodological contribution is the introduction of an additional pooling function to additive decoders.

Sufficient Nonlinearity of defined generative process(adapted from pure additive decoders Lachapelle et al. (2023)): DRVI's decoder is sufficiently nonlinear (main requirement of disentanglement) if, for all z in the latent space, the following matrix has linearly independent columns:

$$W_f(z) := \left[Df^{(k)}(z_k) \times \psi'\left(f^{(k)}(z_k)\right) \middle| D^2 f^{(k)}(z_k) \times \psi'\left(f^{(k)}(z_k)\right) + (Df^{(k)}(z_k))^2 \times \psi''\left(f^{(k)}(z_k)\right) \right]_{1 \leq k \leq K} \tag{4}$$

Where $Df^{(k)}(z_k)$ and $D^2 f^{(k)}(z_k)$ denote first-order and second-order derivatives of $f^{(k)}$ at point z_k, respectively. The functions ψ' and ψ'' are the first-order and second-order derivatives of ψ. Note that $Df^{(k)}(z_k)$ and $D^2 f^{(k)}(z_k)$ are $M \times 1$ matrices that collectively form an $M \times 2K$ matrix $W_f(z)$.

2.4 LSE Pooling

Consider the case where $\sigma = \log$ and $\psi = \exp$. Then $f(z) = \log \sum_{1 \leq k \leq K} \exp\left(f^{(k)}(z_k)\right)$.

As we have $\psi'(x) = \psi''(x) = \exp(x)$ for first-order and second-order derivates. Sufficient nonlinearity in this case reduces to:

$$W_f^{LSE}(z) := \left[Df^{(k)}(z_k) \times \exp\left(f^{(k)}(z_k)\right) \middle| \left((Df^{(k)}(z_k))^2 + D^2 f^{(k)}(z_k)\right) \times \exp\left(f^{(k)}(z_k)\right) \right]_{1 \leq k \leq K}. \tag{5}$$

Consider the GT processes underlying single-cell omics data, although we have no access to them. Assume each underlying process has a distinct set of marker genes, and each set has at least two marker genes with different log-fold-change (LFC) rates upon process activation (genes are non-identical in log-transformed space). In Appendix A.1, we show how to form a block diagonal matrix with these marker genes and prove sufficient nonlinearity of matrix $W_f^{LSE}(z)$ in almost every point of the domain when basis functions are piecewise linear. Although the proof requires the basis functions to be piecewise linear, we observed that this constraint can be relaxed in practice, and RELU activation functions can be replaced by smooth activation functions such as ELU and Softplus. As stated before, we use the ELU activation function that makes the decoder smoother and has almost linear behavior in positive and negative enough areas of its domain.

An important aspect of LSE pooling is its mechanistic interpretation. As described before, the decoder of DRVI (as in most other methods) models the mean parameters of the output distributions in the log-transformed space. This to the assumption that "activities induced by different factors (independent biological processes) sum up in the count space." Formalization of the interpretability of DRVI is available in Appendix A.2.

3 METRICS

Since DRVI aims to bridge the gap between disentangled linear models and flexible non-disentangled methods, we evaluate its performance from two perspectives: disentanglement and integration. We assess the integration quality of multi-batch datasets based on the scIB metrics framework. Details on these metrics can be found at Luecken et al. (2022b). As in the original study, the total score is derived by averaging biological conservation and batch integration scores with weights of 0.6 and 0.4, respectively.

Disentanglement refers to the existence of a one-to-one correspondence between the GT processes and the learned latent factors. Given a one-to-one correspondence, one can use an appropriate similarity function to measure how much each ground truth process is similar to the corresponding latent factor. Accordingly, we measure disentanglement by finding the one-to-one correspondence that maximizes the average pairwise similarities. This category of metrics is known as the Latent Matching Score (LMS). Since the number of GT processes is unknown to the models, the number of GT processes may differ from the number of latent dimensions. Therefore, we define the LMS metrics by finding the best non-overlapping matching between the GT processes and the latent space dimensions Lachapelle et al. (2023):

$$\text{LMS-}F_{sim}(P, Z) = \frac{1}{|P|} \max_{\pi \in \Pi(P,Z)} \left[\sum_{p_l, z_k \in \pi} F_{sim}(p_l, z_k) \right], \tag{6}$$

where F_{sim} can be any similarity function comparing learned factors with respect to GT processes, P indicates the set of vectors corresponding to all ground-truth processes, Z is the set of latent dimensions, and Π is the set of all injective correspondences from P to Z.

In the following, we define three similarity functions based on Spearman correlation, Mutual Information, and Nearest Neighbor Consistency.

Scaled Mutual Information (SMI): We use Mutual Information as a nonlinear metric to compare GT processes and latent space dimensions. Since mutual information is bounded by the entropy of the GT process, we scale it so that the maximum value of +1 becomes achievable.

$$SMI(p_l, z_k) = \frac{\text{MI}(p_l, z_k)}{\text{H}(p_l)}, \tag{7}$$

where MI is the mutual information function and H is the entropy function. SMI is also known by other names such as proficiency or uncertainty coefficient White et al. (2004).

Same-Process Neighbors (SPN): Consider a process with only two "on" and "off" states. Then, we would like to see the cells in the "on" state to form a group in a latent space. Accordingly, we measure the non-random probability for which a cell and its right neighbor are in the "on" state at the same time. So we have:

$$SPN(p, z_k) = \max\left(0, \frac{\mathbb{P}\big(p_j = 1 \big| p_i = 1 \wedge i, j \in \text{NG}(z_k)\big) - \mathbb{P}_{p_i \in p}\big(p_i = 1\big)}{1 - \mathbb{P}_{p_i \in p}\big(p_i = 1\big)}\right), \tag{8}$$

where $NG(z_k)$ indicates the set of all neighboring cells (no cell in between, not necessarily nearest neighbors) in dimension k of the latent space. The function $SPN(p_l, z_k)$ measures the frequency of two neighboring cells being active simultaneously minus the probability of such an outcome when cells are shuffled randomly. This metric is also adjusted to reach +1 as its maximum.

Absolute Spearman Correlations (ASC): This metric is defined based on the absolute Spearman correlation between a GT process and a latent space dimension as

$$ASC(p_l, z_k) = \big|Corr_s(p_l, z_k)\big|, \tag{9}$$

where $Corr_s$ indicates the Spearman correlation function. As Spearman correlation of discrete GT with respect to the continuous values in the latent space will be substantially small, the other two metrics are more suitable in our case.

We use $LMS - SMI$ as the default metric to evaluate the disentanglement of a trained model with respect to known biological processes. While the SMI and ASC metrics accept discrete and continuous processes, the SPN metric is designed specifically for on/off processes.

In addition to the LMS metrics, we use two other secondary metric families to better understand the model's behavior. To determine if a GT process is represented in at least one factor, we find the most similar dimension in the derived latent space for each GT process and then average these similarities across all GT processes. This is called the Most Similar Averaging Score (MSAS). We use the Most Similar Gap Score (MSGS), which builds on MSAS by adding a term that penalizes the presentation of a GT process in multiple dimensions. Formally, we have:

$$\text{MSAS-}F_{sim}(P, Z) = \frac{1}{|P|} \sum_{p_l \in P} \max_{z_k \in Z} F_{sim}(p_l, z_k) \tag{10}$$

$$\text{MSGS-}F_{sim}(P, Z) = \frac{1}{|P|} \sum_{p_l \in P} \max_{z_k \in Z} \left[F_{sim}(p_l, z_k) - \max_{z_{k'} \neq z_k \in Z} F_{sim}(p_l, z_k)\right], \tag{11}$$

where F_{sim} can be any of the SMI, SPN, or ASC functions, and P and Z are defined as before. When using the SMI similarity function, MSGS reduces to the Mutual Information Gap (MIG) metric as defined in Ref Chen et al. (2018).

We primarily use the LMS metric family for evaluation and examine the MSAS and MSGS metric families to further evaluate model behavior. This choice is because MSAS and MSGS metrics do not penalize the presence of multiple irrelevant biological signals within a single dimension.

4 RESULTS

DRVI OUTPERFORMS PREVIOUS DISENTANGLEMENT APPROACHES WHILE MAINTAINING INTEGRATION QUALITY

Learning disentangled latent representation aims to identify biological processes in different axes of the latent space. We refer to all (often unknown) biological processes that are represented in the data and have contributed to its formation as ground truth (GT) processes. From the GT processes to the observed single-cell omics data, we can consider the biology describing cells as a stochastic GT generative function. We call a latent space to be disentangled when latent space axes correspond to the GT processes (see 'Definitions and Notations' in Methods). Accordingly, we can measure the disentanglement by finding and assessing the best matching between GT processes and latent space

dimensions. However, the GT processes and the GT generative function are often unreachable. Here, we propose to address this conundrum by assessing the disentanglement using supervised information such as annotated cell types, applied perturbations, or basic known biological processes such as cell cycle (Figure 1a). Disentanglement is therefore measured by matching latent variables and proxy variables, technically called latent matching score (LMS) in the disentanglement literature Lachapelle et al. (2023) - by default, we measure the matching quality based on the Scaled Mutual Information (SMI, see 'Metrics' in Methods).

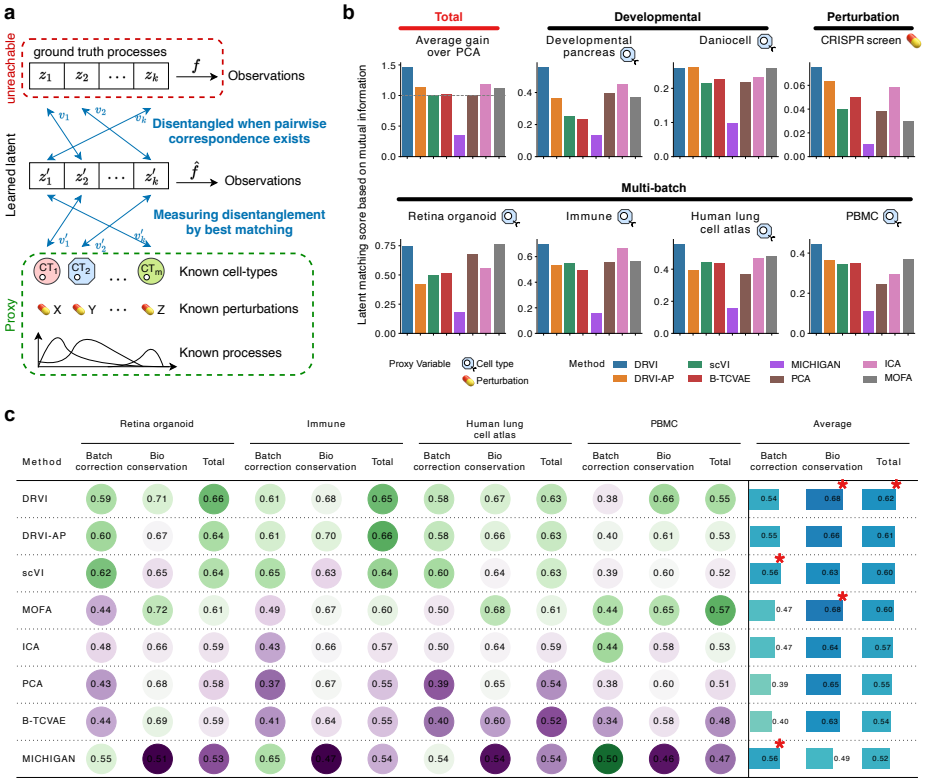


Figure 1: **Benchmarking DRVI in terms of disentanglement and integration quality.** **a,** Evaluation of disentanglement requires access to ground truth biological processes, which are often unreachable. We use cell type annotations, perturbations, and known processes as proxies when available. **b,** Disentanglement was evaluated using latent matching scores, calculated by comparing the alignment of latent dimensions to proxy variables using mutual information. These scores were obtained from benchmarked models across diverse datasets. The proxy variable that serves as the ground truth for each dataset is specified. The total score is determined by averaging the improvements relative to PCA (top left). DRVI achieves the best result. **c,** scIB Integration scores for each individual multi-batch dataset. The last three columns indicate the average scores, with the best method in each category marked with a star. DRVI achieves the best results in unsupervised integration.

To systematically assess how well DRVI learns disentangled latent representations, we benchmarked DRVI against six unsupervised methods covering linear (PCA, ICA, and MOFA) and nonlinear (scVI, β-TCVAE, and MICHIGAN) representation assumptions. Among these methods, ICA, MOFA, β-TCVAE, and MICHIGAN are designed to provide disentangled latent representations. In addition we added DRVI with average pooling (referred to as DRVI-AP) to highlight the importance of the LSE pooling. To cover a wide range of scenarios, we evaluated the mentioned methods on two atlas-scale multi-batch data (HLCA Sikkema et al. (2023) and PBMC Stephenson et al. (2021)), a small multi-batch data (immune Luecken et al. (2022b)), a dataset consisting of primary and organoid cells (retina organoid Cowan et al. (2020)), one small-scale developmental data (de-

velopmental pancreas Bastidas-Ponce et al. (2019)), one large-scale developmental data (Daniocell Sur et al. (2023)), and one genetic perturbation dataset (CRISPR screen Norman et al. (2019)).

When evaluating DRVI on individual datasets, DRVI consistently outperforms or matches the best benchmarked methods in providing disentangled latent representations (Figure 1b).

To address the inherent variability across different datasets, we normalized each metric by the corresponding PCA score before aggregation. The resulting metric is referred to as the gain over PCA (see 'Metrics' in Methods). The total score was calculated by averaging the gain over PCA across all datasets (Figure 1b, top left). We find that also in the overall metric, DRVI outperforms other benchmarked models, achieving a 23.9% improvement over ICA, the next best-performing model.

We explored two additional similarity functions, absolute Spearman correlations (ASC) and same-process neighbors (SPN), to assess the benchmarked methods using the LMS criteria (Supplementary Figure 2.). For completeness, we also provide the results of two additional disentanglement metric families (Supplementary Figure 3): Most Similar Averaging Score (MSAS) and Most Similar Gap Score (MSGS).

Current unsupervised methods such as ICA and MOFA enforce disentanglement through strongly limiting model constraints like linearity and sparsity, restricting their ability to represent rich, multi-batch data. To assess the applicability of disentangled models on complex data, we assess integration quality alongside disentanglement. We consider the datasets containing multiple batches and assessed the biological signal conservation and batch correction quality based on the scIB metrics Luecken et al. (2022b). Considering the average performance across all datasets, DRVI achieved the highest total integration score among all benchmarked methods (Figure 1c).

DRVI ALLOWS ATLAS-LEVEL IDENTIFICATION OF BIOLOGICAL PROCESSES

To demonstrate the capability of DRVI in disentangling latent factors and identifying biological signals in complex data, we applied DRVI to the Human Lung Cell Atlas (HLCA) Sikkema et al. (2023). We use the curated core of this atlas (Figure 2a). consisting of 166 samples from 14 datasets and including a total of 585K cells with curated annotations across 61 distinct cell identities in the finest level annotations.

We identified 31 latent dimensions corresponding to cell types in the finest-level annotations with a favorable one-to-one correspondence to cell identities (Figure 2b top and Supplementary Figure A.4). Given that the disentanglement metrics already assess the identification quality of the cell-type indicator dimensions, we will not focus on evaluating this aspect. However, it is noteworthy that DRVI effectively identifies most of the challenging cell types highlighted by the HLCA study, including Migratory DCs (identified by DR 39+), AT0 cells (identified by DR 34+), hillock-like epithelial cells (identified by DR 26-), and pre-TB secretory through transitional club-AT2 cell (identified by DR 32+), Interestingly pre-TB secretory cells are also described in the HLCA study only through marker genes corresponding to transitional club-AT2 cells Sikkema et al. (2023). This is in full alignment with DRVI's latent geometry where AT0 and pre-TB secretory are together identified by DR 34+. that were previously identifiable only after multiple iterations of clustering Sikkema et al. (2023), or clustering at very high resolutions with hundreds of clusters (Supplementary Figure 7). Given that DRVI can capture these cell types in just 52 non-vanished dimensions, we expect it to significantly facilitate the systematic discovery and seed annotation of novel and rare cell types.

To demonstrate the validity of the identified variations beyond cell types, we followed the interpretability pipeline and found the non-linear gene programs related to each non-cell-type dimension (Supplementary Figures 5 and 6). As a result, we identified relevant biological descriptions for 22 of the 25 dimensions expected to represent biological processes (Figure 1c). DRVI successfully identified variations beyond cell types. Here, we highlighted some interesting examples of dimensions (Figure 2d-h). DR 64- identifying IFI27+ Alveolar Macrophages as an even finer cellular state Li et al. (2022); Aegerter et al. (2022); Bailey et al. (2023). DR 29+ indicates Dissociation stress response, a confounding process that can easily be removed from DRVI latent dimension van den Brink et al. (2017); Denisenko et al. (2020). DR 48+ highlights the upregulation of Metallothionein (MT) family genes, indicating a response to metal ions. A cell subset characterized by the expression of CXCL9, CXCL10, and CXCL11 is highlighted by DR 46+, suggesting CXCR3 chemokine receptor binding process Tokunaga et al. (2018), which is particularly relevant in the context of lung

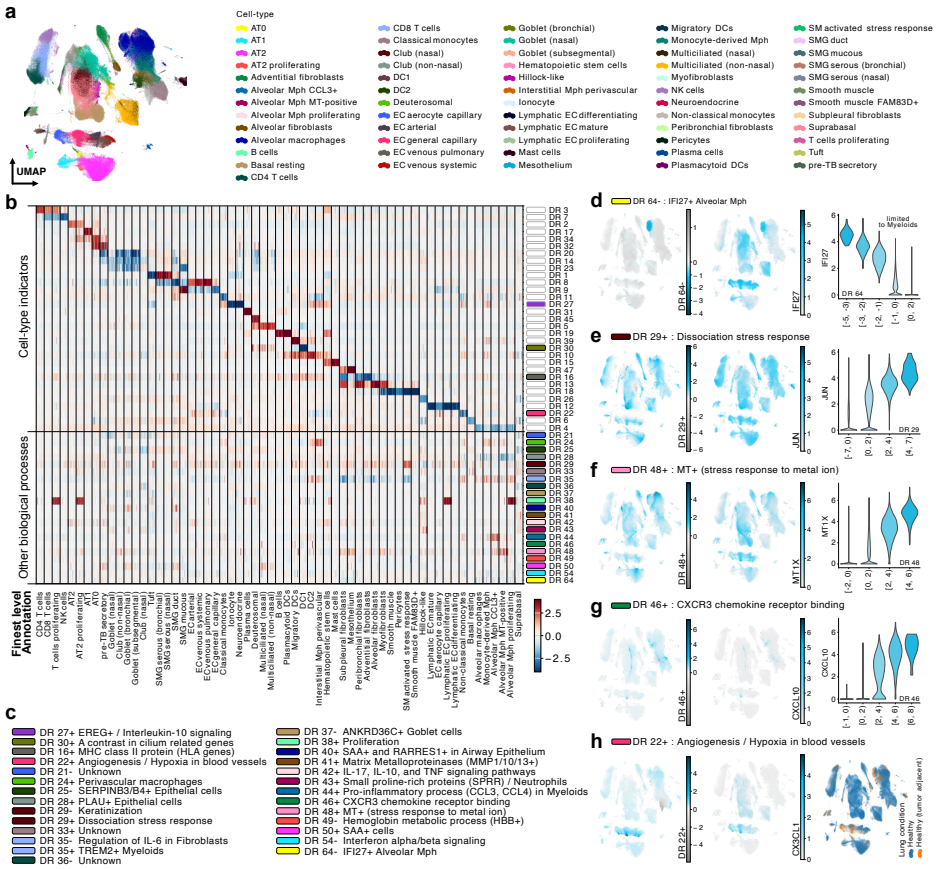


Figure 2: **DRVI uncovers the variations in HLCA beyond cell types. a,** UMAPs of the HLCA integrated by benchmarked methods. Cells are colored by third-level annotations. **b,** DRVI latent dimensions. The heatmap illustrates each dimension's activity in cells vertically grouped by finest level annotations. Cells are subsampled to have an equal number of cells in each cell type. The dimensions are horizontally grouped into two categories: cell-type indicators dimensions (top) and dimensions indicating other biological processes (bottom). Dimensions representing non-cell-type processes are color-coded on the right-hand side of the plot. Vanished dimensions are omitted. For better illustration, some tail values are clipped. **c,** For each dimension indicating a non-cell-type biological process, a short description is provided. **d-g,** Demonstration of five example dimensions. Left: Activity of the example dimensions on the UMAP. Middle: The expression of a relevant gene on the UMAP. Right: Violin plots indicating the expression levels of relevant genes versus the activity of dimensions. **h,** The activity of DR 22+ and one of the identified genes in the nonlinear gene program. This dimension indicates angiogenesis and hypoxia in endothelial cells and is highly active in some tumor-adjacent cells. Mph, Macrophage.

diseases Kameda et al. (2020); Callahan et al. (2021). Finally, as an interesting example, DR 22+ is primarily active in endothelial cells, related to genes such as CX3CL1, SERPINE1, AKAP12, GRP4, and FSTL3 all known to be upregulated under hypoxic conditions Korbecki et al. (2020); Azimi et al. (2017); Finger et al. (2015); Qi et al. (2021); Biron-Shental et al. (2008), a common phenomenon observed in tumor-adjacent environments Chen et al. (2023). The increased abundance of tumor-adjacent endothelial cells in the higher end of DR 22+ supports this interpretation.

Altogether, DRVI enables the disentanglement of biological processes within the HLCA, a large-scale cell atlas. This facilitates the identification of rare cell types and the underlying biological processes.

MEANINGFULNESS STATEMENT

This work addresses entanglement as the main issue of the interpretability in generative models. We proposed "disentangled representation" as a meaningful representation capable of simultaneously modeling discrete states and continuous processes, which is more general than clustering methods producing mutually exclusive sets. Accordingly, we proposed a systematic framework to evaluate disentanglement in single-cell omics. We proposed DRVI, a generative model that solves the disentanglement problem. DRVI demonstrates superior performance in disentanglement and integration. Moreover, it effectively identifies rare cell types and biological processes far beyond cell-types in the Human Lung Cell Atlas (HLCA) as an atlas scale example.

ACKNOWLEDGMENTS

We thank Malte D. Luecken and Lisa Sikkema for providing constructive feedback on this work and for their expertise in HLCA analysis, Sara Jimenes and Dominik Klein for their valuable insights and discussions on the pancreas data, Leander Dony for a fruitful discussion on the required number of latent dimensions, Lisa Sikkema, Alessandro Palma, Mojtaba Bahrami, and Anastasia Litinetskaya for their constructive criticism and valuable feedback on the manuscript, Mohammad Lotfollahi for discussions on related works, Philipp Weiler for suggestions on the choice of developmental datasets, Daniel Strobl for suggestions on enrichment analysis, and Michaela Müller for a discussion on evaluation framework. Finally, we thank all the members of the Theislab for their insightful discussions and support. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the Leibniz Prize and project 458958943 (grant number 5010338). A.A.M. is a member of the ELLIS PhD Program of the European Laboratory for Learning and Intelligent Systems (ELLIS) Society. F.J.T. acknowledges support from the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI (grant number ZT-I-PF-5-01) and from the Leibniz Prize from DFG.

REFERENCES

Helena Aegerter, Bart N Lambrecht, and Claudia V Jakubzick. Biology of lung macrophages in health and disease. *Immunity*, 55(9):1564–1580, September 2022.

Philipp Angerer, Lukas Simon, Sophie Tritschler, F Alexander Wolf, David Fischer, and Fabian J Theis. Single cells make big data: New challenges and opportunities in transcriptomics. *Curr. Opin. Syst. Biol.*, 4:85–91, August 2017.

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 Biol.*, 21(1):111, May 2020.

Iman Azimi, Rosalie M Petersen, Erik W Thompson, Sarah J Roberts-Thomson, and Gregory R Monteith. Hypoxia-induced reactive oxygen species mediate N-cadherin and SERPINE1 expression, EGFR signalling and motility in MDA-MB-468 breast cancer cells. *Sci. Rep.*, 7(1):15140, November 2017.

Joseph I Bailey, Connor H Puritz, Karolina J Senkow, Nikolay S Markov, Estefani Diaz, Emmy Jonasson, Zhan Yu, Suchitra Swaminathan, Ziyan Lu, Samuel Fenske, Rogan A Grant, Hiam Abdala-Valencia, Ruben J Mylvaganam, Janet Miller, R Ian Cumming, Robert M Tighe, Kymberly M Gowdy, Ravi Kalhan, Manu Jain, Ankit Bharat, Chitaru Kurihara, Ruben San Jose Estepar, Raul San Jose Estepar, George R Washko, Ali Shilatifard, Jacob I Sznajder, Karen M Ridge, G R Scott Budinger, Rosemary Braun, Alexander V Misharin, and Marc A Sala. Expansion of profibrotic monocyte-derived alveolar macrophages in patients with persistent respiratory symptoms and radiographic abnormalities after COVID-19. July 2023.

Aimée Bastidas-Ponce, Sophie Tritschler, Leander Dony, Katharina Scheibner, Marta Tarquis-Medina, Ciro Salinno, Silvia Schirge, Ingo Burtscher, Anika Böttcher, Fabian J Theis, Heiko Lickert, and Mostafa Bakhti. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. *Development*, 146(12), June 2019.

Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W H Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W Newell. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.*, 37(1):38–44, December 2018.

Michael Bereket and Theofanis Karaletsos. Modelling cellular perturbations with the sparse additive mechanism shift variational autoencoder. November 2023.

Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.*, 38(12): 1408–1414, December 2020.

T Biron-Shental, W T Schaiff, E Rimon, T L Shim, D M Nelson, and Y Sadovsky. Hypoxia enhances the expression of follistatin-like 3 in term human trophoblasts. *Placenta*, 29(1):51–57, January 2008.

Triet M Bui, Hannah L Wiesolek, and Ronen Sumagin. ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis. *J. Leukoc. Biol.*, 108(3):787–799, September 2020.

Victoria Callahan, Seth Hawks, Matthew A Crawford, Caitlin W Lehman, Holly A Morrison, Hannah M Ivester, Ivan Akhrymuk, Niloufar Boghdeh, Rafaela Flor, Carla V Finkielstein, Irving Coy Allen, James Weger-Lucarelli, Nisha Duggal, Molly A Hughes, and Kylene Kehn-Hall. The Pro-Inflammatory chemokines CXCL9, CXCL10 and CXCL11 are upregulated following SARS-CoV-2 infection in an AKT-Dependent manner. *Viruses*, 13(6), June 2021.

T Chen, Xuechen Li, R Grosse, and D Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Adv. Neural Inf. Process. Syst.*, abs/1802.04942, February 2018.

Zhou Chen, Fangfang Han, Yan Du, Huaqing Shi, and Wence Zhou. Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions. *Signal Transduct. Target. Ther.*, 8 (1):70, February 2023.

Cameron S Cowan, Magdalena Renner, Martina De Gennaro, Brigitte Gross-Scherf, David Goldblum, Yanyan Hou, Martin Munz, Tiago M Rodrigues, Jacek Krol, Tamas Szikra, Rachel Cuttat, Annick Waldt, Panagiotis Papasaikas, Roland Diggelmann, Claudia P Patino-Alvarez, Patricia Galliker, Stefan E Spirig, Dinko Pavlinic, Nadine Gerber-Hollbach, Sven Schuierer, Aldin Srdanovic, Marton Balogh, Riccardo Panero, Akos Kusnyerik, Arnold Szabo, Michael B Stadler, Selim Orgül, Simone Picelli, Pascal W Hasler, Andreas Hierlemann, Hendrik P N Scholl, Guglielmo Roma, Florian Nigsch, and Botond Roska. Cell types of the human retina and its organoids at Single-Cell resolution. *Cell*, 182(6):1623–1640.e34, September 2020.

Creators Kristian K Ullrich1 Show affiliations 1. Max Planck Institute for Evolutionary Biology. zebrafish GSE223922 scRNA data set objects.

Carlo De Donno, Soroor Hediyeh-Zadeh, Amir Ali Moinfar, Marco Wagenstetter, Luke Zappia, Mohammad Lotfollahi, and Fabian J Theis. Population-level integration of single-cell datasets enables multi-scale analysis across samples. *Nat. Methods*, 20(11):1683–1692, November 2023.

Elena Denisenko, Belinda B Guo, Matthew Jones, Rui Hou, Leanne de Kock, Timo Lassmann, Daniel Poppe, Olivier Clément, Rebecca K Simmons, Ryan Lister, and Alistair R R Forrest. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.*, 21(1):130, June 2020.

David DeTomaso and Nir Yosef. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.*, 12(5):446–456.e9, May 2021.

Elizabeth C Finger, Laura Castellini, Erinn B Rankin, Marta Vilalta, Adam J Krieg, Dadi Jiang, Alice Banh, Wayne Zundel, Marianne Broome Powell, and Amato J Giaccia. Hypoxic induction of AKAP12 variant 2 shifts PKA-mediated protein phosphorylation to enhance migration and metastasis of melanoma cells. *Proc. Natl. Acad. Sci. U. S. A.*, 112(14):4441–4446, April 2015.

Gilles Gut, Stefan G Stark, Gunnar Rätsch, and Natalie R Davidson. pmVAE: Learning interpretable Single-Cell representations with pathway modules. January 2021.

Maren Hackenberg, Niklas Brunn, Tanja Vogel, and Harald Binder. Infusing structural assumptions into dimension reduction for single-cell RNA sequencing data to identify small gene sets. February 2024.

Soroor Hediyeh-zadeh, Tom Fischer, and Fabian J Theis. Disentanglement via mechanism sparsity by replaying realizations of the past. March 2024.

Lukas Heumos, Yuge Ji, Lilly May, Tessa Green, Xinyue Zhang, Xichen Wu, Johannes Ostner, Stefan Peidli, Antonia Schumacher, Karin Hrovatin, Michaela Müller, Faye Chong, Gregor Sturm, Alejandro Tejada, Emma Dann, Mingze Dong, Mojtaba Bahrami, Ilan Gold, Sergei Rybakov, Altana Namsaraeva, Amir Moinfar, Zihe Zheng, Eljas Roellin, Isra Mekki, Chris Sander, Mohammad Lotfollahi, Herbert B Schiller, and Fabian J Theis. Pertpy: an end-to-end framework for perturbation analysis. *bioRxiv*, pp. 2024.08.04.606516, August 2024.

Blanca E Himes, Xiaofeng Jiang, Peter Wagner, Ruoxi Hu, Qiyu Wang, Barbara Klanderman, Reid M Whitaker, Qingling Duan, Jessica Lasky-Su, Christina Nikolos, William Jester, Martin Johnson, Reynold A Panettieri, Jr, Kelan G Tantisira, Scott T Weiss, and Quan Lu. RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. *PLoS One*, 9(6):e99625, June 2014.

A Mcgarry Houghton. Matrix metalloproteinases in destructive lung disease. *Matrix Biol.*, 44-46: 167–174, May 2015.

Karin Hrovatin, Amir Ali Moinfar, Luke Zappia, Alejandro Tejada Lapuerta, Ben Lengerich, Manolis Kellis, and Fabian J Theis. Integrating single-cell RNA-seq datasets with substantial batch effects. *bioRxiv*, February 2024.

Justyna Jarczak, Ewa M Kościuczuk, Paweł Lisowski, Nina Strzałkowska, Artur Jóźwik, Jarosław Horbańczuk, Józef Krzyżewski, Lech Zwierzchowski, and Emilia Bagnicka. Defensins: natural component of human innate immunity. *Hum. Immunol.*, 74(9):1069–1079, September 2013.

Masami Kameda, Mitsuo Otsuka, Hirofumi Chiba, Koji Kuronuma, Takehiro Hasegawa, Hiroki Takahashi, and Hiroki Takahashi. CXCL9, CXCL10, and CXCL11; biomarkers of pulmonary inflammation associated with autoimmunity in patients with collagen vascular diseases-associated interstitial lung disease and interstitial pneumonia with autoimmune features. *PLoS One*, 15(11): e0241719, November 2020.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv [cs.LG]*, December 2014.

Diederik P Kingma and Max Welling. Auto-Encoding variational bayes. December 2013.

Vladimir Yu Kiselev, Tallulah S Andrews, and Martin Hemberg. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.*, 20(5):273–282, May 2019.

Jan Korbecki, Donata Simińska, Klaudyna Kojder, Szymon Grochans, Izabela Gutowska, Dariusz Chlubek, and Irena Baranowska-Bosiacka. Fractalkine/CX3CL1 in neoplastic processes. *Int. J. Mol. Sci.*, 21(10):3723, May 2020.

Russell Z Kunes, Thomas Walle, Max Land, Tal Nawy, and Dana Pe'er. Supervised discovery of interpretable gene programs from single-cell data. *Nat. Biotechnol.*, September 2023.

Sébastien Lachapelle, Divyat Mahajan, Ioannis Mitliagkas, and Simon Lacoste-Julien. Additive decoders for latent variables identification and cartesian-product extrapolation. July 2023.

Christopher Lance, Malte D Luecken, Daniel B Burkhardt, Robrecht Cannoodt, Pia Rautenstrauch, Anna Laddach, Aidyn Ubingazhibov, Zhi-Jie Cao, Kaiwen Deng, Sumeer Khan, Qiao Liu, Nikolay Russkikh, Gleb Ryazantsev, Uwe Ohler, NeurIPS 2021 Multimodal data integration competition participants, Angela Oliveira Pisco, Jonathan Bloom, Smita Krishnaswamy, and Fabian J Theis. Multimodal single cell data integration challenge: Results and lessons learned. In *NeurIPS 2021 Competitions and Demonstrations Track*, pp. 162–176. PMLR, July 2022.

Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B Schiller, Dana Pe'er, and Fabian J Theis. CellRank for directed single-cell fate mapping. *Nat. Methods*, 19(2):159–170, February 2022.

Pureun Haneul Lee, Byeong Gon Kim, Sun Hye Lee, June Hyuck Lee, Sung Woo Park, Do Jin Kim, Choon Sik Park, George D Leikauf, and An Soo Jang. Alteration in claudin-4 contributes to airway inflammation and responsiveness in asthma. *Allergy Asthma Immunol. Res.*, 10(1):25–33, January 2018.

Xin Li, Fred W Kolling, Daniel Aridgides, Diane Mellinger, Alix Ashare, and Claudia V Jakubzick. ScRNA-seq expression of IFI27 and APOC2 identifies four alveolar macrophage superclusters in healthy BALF. *Life Sci Alliance*, 5(11), November 2022.

Wenli Liu and Griffin P Rodgers. Olfactomedin 4 expression and functions in innate immunity, inflammation, and cancer. *Cancer Metastasis Rev.*, 35(2):201–212, June 2016.

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

Romain Lopez, Nataša Tagasovska, Stephen Ra, Kyunghyn Cho, Jonathan K Pritchard, and Aviv Regev. Learning causal representations of single cells via sparse mechanism shift modeling. November 2022.

Mohammad Lotfollahi, F Alexander Wolf, and Fabian J Theis. scgen predicts single-cell perturbation responses. *Nat. Methods*, 16(8):715–721, August 2019.

Mohammad Lotfollahi, Mohsen Naghipourfar, Fabian J Theis, and F Alexander Wolf. Conditional out-of-distribution generation for unpaired data using transfer VAE. *Bioinformatics*, 36(Suppl_2): i610–i617, December 2020.

Mohammad Lotfollahi, Leander Dony, Harshita Agarwala, and Fabian Theis. Out-of-distribution prediction with disentangled representations for single-cell RNA sequencing data. *bioRxiv*, pp. 2021.09.01.458535, September 2021.

Mohammad Lotfollahi, Sergei Rybakov, Karin Hrovatin, Soroor Hediyeh-Zadeh, Carlos Talavera-López, Alexander V Misharin, and Fabian J Theis. Biologically informed deep learning to query gene programs in single-cell atlases. *Nat. Cell Biol.*, 25(2):337–350, February 2023.

Malte Luecken, Maren Buttner, Anna Danese, Marta Interlandi, Michaela Müller, Daniel Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, Fabian Theis, and Kridsadakorn Chaichoompu. Benchmarking atlas-level data integration in single-cell genomics - integration task datasets, January 2022a.

Malte D Luecken, Daniel Bernard Burkhardt, Robrecht Cannoodt, Christopher Lance, Aditi Agrawal, Hananeh Aliee, Ann T Chen, Louise Deconinck, Angela M Detweiler, Alejandro A Granados, Shelly Huynh, Laura Isacco, Yang Joon Kim, Dominik Klein, Bony de Kumar, Sunil Kuppasani, Heiko Lickert, Aaron McGeever, Honey Mekonen, Joaquin Caceres Melgarejo, Maurizio Morri, Michaela Müller, Norma Neff, Sheryl Paul, Bastian Rieck, Kaylie Schneider, Scott Steelman, Michael Sterr, Daniel J Treacy, Alexander Tong, Alexandra-Chloe Villani, Guilin Wang, Jia Yan, Ce Zhang, Angela Oliveira Pisco, Smita Krishnaswamy, Fabian J Theis, and Jonathan M Bloom. A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. August 2021.

Malte D Luecken, M Büttner, K Chaichoompu, A Danese, M Interlandi, M F Mueller, D C Strobl, L Zappia, M Dugas, M Colomé-Tatché, and Fabian J Theis. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods*, 19(1):41–50, January 2022b.

Nadzeya Marozkina, Laura Smith, Yi Zhao, Joe Zein, James F Chmiel, Jeeho Kim, Janna Kiselar, Michael D Davis, Rebekah S Cunningham, Scott H Randell, and Benjamin Gaston. Somatic cell hemoglobin modulates nitrogen oxide metabolism in the human airway epithelium. *Sci. Rep.*, 11 (1):15498, July 2021.

Isidoro Martínez, Luis Lombardía, Blanca García-Barreno, Orlando Domínguez, and José A Melero. Distinct gene subsets are induced at different time points after human respiratory syncytial virus infection of A549 cells. *J. Gen. Virol.*, 88(Pt 2):570–581, February 2007.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv [stat.ML]*, February 2018.

Yoshikuni Mimata, Akihisa Kamataki, Shinya Oikawa, Kenya Murakami, Miwa Uzuki, Tadashi Shimamura, and Takashi Sawai. Interleukin-6 upregulates expression of ADAMTS-4 in fibroblast-like synoviocytes from patients with rheumatoid arthritis: IL-6 upregulates ADAMTS-4 in RA-FLS. *Int. J. Rheum. Dis.*, 15(1):36–44, February 2012.

Rodrigo Nalio Ramos, Yoann Missolo-Koussou, Yohan Gerber-Ferder, Christian P Bromley, Mattia Bugatti, Nicolas Gonzalo Núñez, Jimena Tosello Boari, Wilfrid Richer, Laurie Menger, Jordan Denizeau, Christine Sedlik, Pamela Caudana, Fiorella Kotsias, Leticia L Niborski, Sophie Viel, Mylène Bohec, Sonia Lameiras, Sylvain Baulande, Laëtitia Lesage, André Nicolas, Didier Meseure, Anne Vincent-Salomon, Fabien Reyal, Charles-Antoine Dutertre, Florent Ginhoux, Lene Vimeux, Emmanuel Donnadieu, Bénédicte Buttard, Jérôme Galon, Santiago Zelenay, William Vermi, Pierre Guermonprez, Eliane Piaggio, and Julie Helft. Tissue-resident FOLR2+ macrophages associate with CD8+ T cell infiltration in human breast cancer. *Cell*, 185(7):1189–1207.e25, March 2022.

Achille Nazaret, Joy Linyue Fan, Vincent-Philippe Lavallée, Andrew E Cornish, Vaidotas Kiseliovas, Ignas Masilionis, Jaeyoung Chun, Robert L Bowman, Shira E Eisman, James Wang, Lingting Shi, Ross L Levine, Linas Mazutis, David Blei, Dana Pe'er, and Elham Azizi. Deep generative model deciphers derailed trajectories in acute myeloid leukemia. November 2023.

Thomas M Norman, Max A Horlbeck, Joseph M Replogle, Alex Y Ge, Albert Xu, Marco Jost, Luke A Gilbert, and Jonathan S Weissman. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science*, 365(6455):786–793, August 2019.

Yoshiki Ohta, Yosuke Sasaki, Mitsuji Saito, Miki Kushima, Masafumi Takimoto, Akira Shiokawa, and Hidekazu Ota. Claudin-4 as a marker for distinguishing malignant mesothelioma from lung carcinoma and serous adenocarcinoma. *Int. J. Surg. Pathol.*, 21(5):493–501, October 2013.

Haining Qi, Cuiyun Yao, Jianhong Xing, and Ying Qin. Hypoxia-induced GPR4 suppresses trophoblast cell migration and proliferation through the MAPK signaling pathway. *Reprod. Toxicol.*, 99:1–8, January 2021.

Sandra Ruiz García, Marie Deprez, Kevin Lebrigand, Amélie Cavard, Agnès Paquet, Marie-Jeanne Arguel, Virginie Magnone, Marin Truchi, Ignacio Caballero, Sylvie Leroy, Charles-Hugo Marquette, Brice Marcet, Pascal Barbry, and Laure-Emmanuelle Zaragosi. Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development*, 146(20):dev.177428, October 2019.

Lisa Sikkema, Ciro Ramírez-Suástegui, Daniel C Strobl, Tessa E Gillett, Luke Zappia, Elo Madissoon, Nikolay S Markov, Laure-Emmanuelle Zaragosi, Yuge Ji, Meshal Ansari, Marie-Jeanne Arguel, Leonie Apperloo, Martin Banchero, Christophe Bécavin, Marijn Berg, Evgeny Chichelnitskiy, Mei-I Chung, Antoine Collin, Aurore C A Gay, Janine Gote-Schniering, Baharak Hooshiar Kashani, Kemal Inecik, Manu Jain, Theodore S Kapellos, Tessa M Kole, Sylvie Leroy, Christoph H Mayr, Amanda J Oliver, Michael von Papen, Lance Peter, Chase J Taylor, Thomas Walzthoeni, Chuan Xu, Linh T Bui, Carlo De Donno, Leander Dony, Alen Faiz, Minzhe Guo, Austin J Gutierrez, Lukas Heumos, Ni Huang, Ignacio L Ibarra, Nathan D Jackson, Preetish Kadur Lakshminarasimha Murthy, Mohammad Lotfollahi, Tracy Tabib, Carlos Talavera-López, Kyle J Travaglini, Anna Wilbrey-Clark, Kaylee B Worlock, Masahiro Yoshida, Lung Biological Network Consortium, Maarten van den Berge, Yohan Bossé, Tushar J Desai, Oliver Eickelberg, Naftali Kaminski, Mark A Krasnow, Robert Lafyatis, Marko Z Nikolic, Joseph E Powell, Jayaraj Rajagopal, Mauricio Rojas, Orit Rozenblatt-Rosen, Max A Seibold, Dean Sheppard, Douglas P Shepherd, Don D Sin, Wim Timens, Alexander M Tsankov, Jeffrey Whitsett, Yan Xu, Nicholas E Banovich, Pascal Barbry, Thu Elizabeth Duong, Christine S Falk, Kerstin B Meyer, Jonathan A Kropski, Dana Pe'er, Herbert B Schiller, Purushothama Rao Tata, Joachim L Schultze, Sara A Teichmann, Alexander V Misharin, Martijn C Nawijn, Malte D Luecken, and Fabian J Theis. An integrated cell atlas of the lung in health and disease. *Nat. Med.*, 29(6):1563–1577, June 2023.

Caroline L Sokol and Andrew D Luster. The chemokine system in innate immunity. *Cold Spring Harb. Perspect. Biol.*, 7(5), January 2015.

Emily Stephenson, Gary Reynolds, Rachel A Botting, Fernando J Calero-Nieto, Michael D Morgan, Zewen Kelvin Tuong, Karsten Bach, Waradon Sungnak, Kaylee B Worlock, Masahiro Yoshida,

Natsuhiko Kumasaka, Katarzyna Kania, Justin Engelbert, Bayanne Olabi, Jarmila Stremenova Spegarova, Nicola K Wilson, Nicole Mende, Laura Jardine, Louis C S Gardner, Issac Goh, Dave Horsfall, Jim McGrath, Simone Webb, Michael W Mather, Rik G H Lindeboom, Emma Dann, Ni Huang, Krzysztof Polanski, Elena Prigmore, Florian Gothe, Jonathan Scott, Rebecca P Payne, Kenneth F Baker, Aidan T Hanrath, Ina C D Schim van der Loeff, Andrew S Barr, Amada Sanchez-Gonzalez, Laura Bergamaschi, Federica Mescia, Josephine L Barnes, Eliz Kilich, Angus de Wilton, Anita Saigal, Aarash Saleh, Sam M Janes, Claire M Smith, Nusayhah Gopee, Caroline Wilson, Paul Coupland, Jonathan M Coxhead, Vladimir Yu Kiselev, Stijn van Dongen, Jaume Bacardit, Hamish W King, Cambridge Institute of Therapeutic Immunology and Infectious Disease-National Institute of Health Research (CITIID-NIHR) COVID-19 BioResource Collaboration, Anthony J Rostron, A John Simpson, Sophie Hambleton, Elisa Laurenti, Paul A Lyons, Kerstin B Meyer, Marko Z Nikolić, Christopher J A Duncan, Kenneth G C Smith, Sarah A Teichmann, Menna R Clatworthy, John C Marioni, Berthold Göttgens, and Muzlifah Haniffa. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.*, 27(5):904–916, May 2021.

Yu Sun, Namratha Sheshadri, and Wei-Xing Zong. SERPINB3 and B4: From biochemistry to biology. *Semin. Cell Dev. Biol.*, 62:170–177, February 2017.

Abhinav Sur, Yiqun Wang, Paulina Capar, Gennady Margolin, Morgan Kathleen Prochaska, and Jeffrey A Farrell. Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. *Dev. Cell*, 58(24):3028–3047.e12, December 2023.

Valentine Svensson, Roser Vento-Tormo, and Sarah A Teichmann. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.*, 13(4):599–604, April 2018.

Valentine Svensson, Adam Gayoso, Nir Yosef, and Lior Pachter. Interpretable factor models of single-cell RNA-seq via variational autoencoders. *Bioinformatics*, 36(11):3418–3421, June 2020.

Ryuma Tokunaga, Wu Zhang, Madiha Naseem, Alberto Puccini, Martin D Berger, Shivani Soni, Michelle McSkane, Hideo Baba, and Heinz-Josef Lenz. CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - a target for novel cancer therapy. *Cancer Treat. Rev.*, 63:40–47, February 2018.

V A Traag, L Waltman, and N J van Eck. From louvain to leiden: guaranteeing well-connected communities. *Sci. Rep.*, 9(1):5233, March 2019.

Susanne C van den Brink, Fanny Sage, Ábel Vértesy, Bastiaan Spanjaard, Josi Peterson-Maduro, Chloé S Baron, Catherine Robin, and Alexander van Oudenaarden. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods*, 14(10):935–936, September 2017.

Philipp Weiler, Marius Lange, Michal Klein, Dana Pe'er, and Fabian Theis. CellRank 2: unified fate mapping in multiview single-cell data. *Nat. Methods*, June 2024.

J V White, S Steingold, and C Fournelle. Performance metrics for group-detection algorithms. *Proceedings of Interface*, 2004.

Kangyun Wu, Derek E Byers, Xiaohua Jin, Eugene Agapov, Jennifer Alexander-Brett, Anand C Patel, Marina Cella, Susan Gilfilan, Marco Colonna, Daniel L Kober, Tom J Brett, and Michael J Holtzman. TREM-2 promotes macrophage survival and lung disease after respiratory viral infection. *J. Exp. Med.*, 212(5):681–697, May 2015.

Donghai Xiong, Yian Wang, and Ming You. A gene expression signature of TREM2hi macrophages and $\gamma\delta$ T cells predicts immunotherapy response. *Nat. Commun.*, 11(1):5084, October 2020a.

Donghai Xiong, Yian Wang, and Ming You. A gene expression signature of TREM2hi macrophages and $\gamma\delta$ T cells predicts immunotherapy response. *Nat. Commun.*, 11(1):5084, October 2020b.

Hanli Xu, Shuye Lin, Ziyun Zhou, Duoduo Li, Xiting Zhang, Muhan Yu, Ruoyi Zhao, Yiheng Wang, Junru Qian, Xinyi Li, Bohan Li, Chuhan Wei, Keqiang Chen, Teizo Yoshimura, Ji Ming Wang, and Jiaqiang Huang. New genetic and epigenetic insights into the chemokine system: the latest discoveries aiding progression toward precision medicine. *Cell. Mol. Immunol.*, 20(7):739–776, July 2023.

Chao Yang, Jason R Siebert, Robert Burns, Zachary J Gerbec, Benedetta Bonacci, Amy Rymaszewski, Mary Rau, Matthew J Riese, Sridhar Rao, Karen-Sue Carlson, John M Routes, James W Verbsky, Monica S Thakar, and Subramaniam Malarkannan. Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. *Nat. Commun.*, 10(1):3931, September 2019.

Masahiro Yoshida, Kaylee B Worlock, Ni Huang, Rik G H Lindeboom, Colin R Butler, Natsuhiko Kumasaka, Cecilia Dominguez Conde, Lira Mamanova, Liam Bolt, Laura Richardson, Krzysztof Polanski, Elo Madissoon, Josephine L Barnes, Jessica Allen-Hyttinen, Eliz Kilich, Brendan C Jones, Angus de Wilton, Anna Wilbrey-Clark, Waradon Sungnak, J Patrick Pett, Juliane Weller, Elena Prigmore, Henry Yung, Puja Mehta, Aarash Saleh, Anita Saigal, Vivian Chu, Jonathan M Cohen, Clare Cane, Aikaterini Iordanidou, Soichi Shibuya, Ann-Kathrin Reuschl, Iván T Herczeg, A Christine Argento, Richard G Wunderink, Sean B Smith, Taylor A Poor, Catherine A Gao, Jane E Dematte, NU SCRIPT Study Investigators, Gary Reynolds, Muzlifah Haniffa, Georgina S Bowyer, Matthew Coates, Menna R Clatworthy, Fernando J Calero-Nieto, Berthold Göttgens, Christopher O'Callaghan, Neil J Sebire, Clare Jolly, Paolo De Coppi, Claire M Smith, Alexander V Misharin, Sam M Janes, Sarah A Teichmann, Marko Z Nikolić, and Kerstin B Meyer. Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature*, 602(7896):321–327, February 2022.

Hengshi Yu and Joshua D Welch. MichiGAN: sampling from disentangled representations of single-cell data using generative adversarial networks. *Genome Biol.*, 22(1):158, May 2021.

Luke Zappia and Fabian J Theis. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol.*, 22(1):301, October 2021.

Hui Zhang, Alvin T Kho, Qing Wu, Andrew J Halayko, Karen Limbert Rempel, Robert P Chase, Neil B Sweezey, Scott T Weiss, and Feige Kaplan. CRISPLD2 (LGL1) inhibits proinflammatory mediators in human fetal, adult, and COPD lung fibroblasts and epithelial cells. *Physiol Rep*, 4 (17), September 2016.

Ziyue Zhou, Xiaoxiang Zhou, Xu Jiang, Bo Yang, Xin Lu, Yunyun Fei, Lidan Zhao, Hua Chen, Li Zhang, Xiaoyan Si, Naixin Liang, Yadong Wang, Dan Yang, Yezi Peng, Yiying Yang, Zhuoran Yao, Yangzhige He, Xunyao Wu, Wen Zhang, Min Wang, Huaxia Yang, and Xuan Zhang. Single-cell profiling identifies IL1Bhi macrophages associated with inflammation in PD-1 inhibitor-induced inflammatory arthritis. *Nat. Commun.*, 15(1):2107, March 2024.

A Appendix

A.1 Sufficient nonlinearity for LSE pooling

Assume that, in single-cell omics data, each underlying process has a particular set of at least two marker genes with different log-fold-change (LFC) rates that are present upon process activation. In addition, assume these underlying processes form the basis functions $f^{(k)}$ in the ground truth additive generative process, and the basis functions are linear almost everywhere in the domain. Then, the matrix $W_f^{LSE}(z)$ has full-rank columns for almost every point z in the domain of the generative process.

Proof. Consider the assumed marker genes. We show that the limitation of the $W_f^{LSE}(z)$ to these marker genes has full-rank columns. To this end, we sort the marker genes to reach a matrix with rectangular blocks of 1 and zeros everywhere else. Let's call the marker genes indicator matrix T.

Then we have:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{12}$$

$$W_f^{LSE}(z) = \begin{bmatrix} A_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots & B_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ A_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots & B_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ A_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots & B_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & A_{4,2} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & B_{4,2} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & A_{4,3} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & B_{4,3} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A_{6,3} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & B_{6,3} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A_{6,4} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & B_{6,4} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A_{6,5} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & B_{6,5} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & A_{6,4} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & B_{6,4} & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{13}$$

Where $A_{i,j} = Df^{(j)}(z_i) \times \exp\left(f^{(j)}(z_i)\right)$ and $B_{i,j} = ((Df^{(j)}(z_i))^2 + D^2 f^{(j)}(z_i)) \times \exp\left(f^{(j)}(z_i)\right)$. By reordering the columns of $W_f^{LSE}(z)$, we will reach a block sparse matrix as below.

$$W_f^{LSE}(z) = \begin{bmatrix} A_{1,1} & B_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ A_{2,1} & B_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ A_{3,1} & B_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A_{4,2} & B_{4,2} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A_{4,3} & B_{4,3} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & A_{6,3} & B_{6,3} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & A_{6,4} & B_{6,4} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & A_{6,5} & B_{6,5} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & A_{6,4} & B_{6,4} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{14}$$

Having independent columns in such a matrix is equivalent to having independent columns in each block. Since each block is of the form

$$\begin{bmatrix} A_{i,j} & B_{i,j} \\ A_{i+1,j} & B_{i+1,j} \\ A_{i+2,j} & B_{i+2,j} \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} Df^{(j)}(z_i) & (Df^{(j)}(z_i))^2 + D^2 f^{(j)}(z_i) \\ Df^{(j)}(z_{i+1}) & (Df^{(j)}(z_{i+1}))^2 + D^2 f^{(j)}(z_{i+1}) \\ Df^{(j)}(z_{i+2}) & (Df^{(j)}(z_{i+2}))^2 + D^2 f^{(j)}(z_{i+2}) \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \exp\left(f^{(j)}(z_i)\right) & 0 \\ 0 & \exp\left(f^{(j)}(z_i)\right) \end{bmatrix} \tag{15}$$

This is the point we use: (I) the assumption that the basis functions are piecewise linear and (II) the assumption that each process has at least two markers with different log-fold change rates. The first assumption implies $D^2 f^{(j)}(z_i)$ is zero almost everywhere. And the second assumption implies that $Df^{(j)}(z_i)$ and $(Df^{(j)}(z_i))^2$ columns are independent. For the latter, assume that the mentioned columns are dependent on point z. Then there exists constant λ where $Df^{(j)}(z_i) = \lambda(Df^{(j)}(z_i))^2$ for all marker genes i relevant to the process j. Since the derivatives of the basis function are non-zero almost everywhere. This simply implies that $Df^{(j)}(z_i) = 1/\lambda$ or a non-negligible set of points in the domain, which is in contradiction to having two marker genes with different LFC rates (The basis functions are defined in the log-transformed space. So their derivatives indicate the rate of the change of the gene in log-space upon activation of the process). The contradiction completes the proof and implies that all the blocks of the matrix $W_f^{LSE}(z)$ and, consequently, the matrix itself have independent columns and are sufficiently nonlinear. □

A.2 INTERPRETABILITY

Learned latent representations may present known or unknown abstract concepts. Interpreting latent space dimensions in terms of the genes involved helps understand unknown abstract concepts and find explanations for known ones. In factorization models, the loadings (weights of the low-rank matrices) are interpreted as the effect of each gene on each factor. However, in nonlinear models, interpretability is challenging since the dependencies are nonlinear, and the effect of each latent space dimension also depends on the values of other latent dimensions. While this challenge still exists in DRVI, the additive architecture facilitates interpretability as each latent space dimension is decoded separately until the final pooling function.

We use the additive decoder architecture to interpret the latent space dimensions in terms of the genes involved and explain each basis function separately.

$$\log \mu = \log \sum_{1 \leq k \leq K} \exp\left(f^{(k)}(z_k, c)\right) \tag{16}$$

where μ, z_k, $f^{(k)}$, and c are defined as in the average pooling case.

Accordingly, for each gene g and dimensions i of the latent space we have:

$$\mu_g = \sum_{k \in K \wedge k \neq i} \left[\exp\left(f^{(k)}(z_k, c)\right)\right]_g + \left[\exp\left(f^{(i)}(z_i, c)\right)\right]_g \tag{17}$$

Here, the effect of perturbing z_i in count space solely depends on $f^{(i)}(z_i, c)$. Formally:

$$\begin{aligned} \text{effect}^{\text{LSE}}_{\text{counts}}(Z_i, g) = \mathop{\mathbb{E}}_{c \in C} \Bigg[& \max_{z \in [\min(Z_i), \max(Z_i)]} \left[\exp\left(f^{(i)}(z, c)\right)\right]_g - \\ & \min_{z \in [\min(Z_i), \max(Z_i)]} \left[\exp\left(f^{(i)}(z, c)\right)\right]_g \Bigg] \end{aligned} \tag{18}$$

A.3 DATASETS

Below, we provide information about the datasets used in this work

IMMUNE DATASET

The immune dataset consists of 32,484 cells collected from 4 human PBMC studies, including 9 batches and 16 distinct cell types after pre-processing. The data is obtained from Luecken et al. (2022) Luecken et al. (2022a;b). The "Villani" study was excluded due to its non-integer values. The dataset was subset to 2,000 highly-variable genes (HVGs) in a batch-aware manner Luecken et al. (2022b).

THE HLCA

The HLCA consists of a core dataset and an extended version Sikkema et al. (2023). We obtained the curated core part from cellxgene (https://cellxgene.cziscience.com/collections/6f6d381a-7701-4781-935c-db10d30de293), which includes 584,944 cells across 166 samples from 14 datasets. This dataset contains annotations at different levels. We used samples as the batch covariate and the finest-level annotations comprising 61 cell types to benchmark the methods. The dataset was subset to 1,996 HVGs originally used to construct the HLCA reference model available on Zenodo (https://doi.org/10.5281/zenodo.7599104).

CRISPR Screen Dataset

The CRISPR screen dataset, also known as the Norman perturb-seq dataset, comprises 104,339 single-cells of K562 cell line (including control cells) perturbed by 106 single-gene perturbations and 131 combinatorial perturbations Norman et al. (2019). The 238 unique perturbation signatures were used to benchmark methods. We downloaded the pre-processed data using the pertpy package Heumos et al. (2024). The data was limited to the 5,000 HVGs originally provided in the pre-processed data.

PBMC Dataset

The PBMC dataset comprises 647,366 cells across 143 samples from 130 healthy and COVID-19 patients Stephenson et al. (2021). The data was collected at three sites, and we used the site as the batch covariate to account for the highest technical variation. We used the finest-level annotation containing 51 unique cell types for benchmarking the methods. The data was obtained from https://www.covid19cellatlas.org/ and was subset to 2,000 HVGs in a batch-aware manner.

Retina Organoid Dataset

The retina organoid data consists of 19,768 cells from fovea, 34,723 cells from periphery, and 43,857 retina organoid cells (total 98,348 cells) Cowan et al. (2020). The cells come from 41 samples and are annotated into 21 unique cell types. The data is obtained from https://cellxgene.cziscience.com/collections/2f4c738f-e2f3-4553-9db2-0582a38ea4dc and was subset to 2,001 HVGs shared between primary and organoid data Hrovatin et al. (2024).

Developmental Pancreas Dataset

The developmental pancreas dataset consists of 3,696 pancreatic mouse cells during endocrinogenesis at embryonic day 15.5 Bastidas-Ponce et al. (2019). The data was obtained from the scVelo package Bergen et al. (2020). We used scVelo to pre-process and subset the data to the top 2,000 HVGs. The CellRank package provides a subset of this data with finer annotations Lange et al. (2022); Weiler et al. (2024). We transferred these finer annotations to the complete dataset and used the finest annotation comprising 15 unique cell states for benchmarking.

Daniocell Dataset

The Daniocell dataset is a comprehensive transcriptional atlas of early zebrafish development, comprising 489,686 cells from 62 stages during zebrafish embryogenesis Sur et al. (2023). The data covers 20 tissues and 156 identity clusters, which we used to benchmark the data. We obtained the h5ad object from https://zenodo.org/records/8133569 Creators Kristian K Ullrich1 Show affiliations 1. Max Planck Institute for Evolutionary Biology. The pre-processing of this data involved marking data with the annotations provided in the original publication, removing author-identified doublets, and subsetting data to 2,000 HVGs.

NeurIPS 2021 Dataset

The NeurIPS 2021 Dataset contains nearly 120K single-cell paired RNA-ATAC and RNA-ADT (antibody-derived tags) measurements from the human bone marrow of 10 donors Luecken et al. (2021). We used the ATAC modality of this dataset consisting of 69,247 cells from 13 batches covering 22 cell types. The dataset is retrieved from GSE194122 and subset to 14,865 highly variable peaks.

A.4 Identification of the non-cell-type dimensions for the HLCA

Here, we try to identify and explain the non-cell-type dimensions not described in the results section. This is done based on the activity of the dimensions.

DR 64-: As the first example, we showcased a subtle variation within Alveolar Macrophages (AMs - a subset of Myeloid cells), identified by DR 64- (Figure 2d). IFI27 is the only gene significantly associated with this dimension, and is known to define a subcluster of AMs Li et al. (2022); Aegerter

et al. (2022); Bailey et al. (2023). Despite its expression in other cell types, such as blood vessels, DRVI has nonlinearly localized its expression to AMs due to the distinct patterns of co-regulated genes. The violin plot of the IFI27 gene with respect to DR 64 in the Myeloid cells confirms the mentioned relationship.

DR 29+: The activity of DR 29+ spans almost all cell types. This dimension is highly associated with the genes ATF3, EGR1, JUN, FOSB, and FOS as top indicators of the nonlinear gene program, suggesting that it represents single-cell dissociation stress response van den Brink et al. (2017); Denisenko et al. (2020). The violin plot of the JUN gene with respect to DR 29 highlights this relationship (Figure 2e).

DR 48+: The upregulation of Metallothionein (MT) family genes characterizes DR 48+, indicating a response to metal ions. The monotonic relationship between DR 48 and MT1X, visualized in the violin plot, further supports this finding (Figure 2f). While this dimension is active in MT-positive AM cells, DR 48+ is not exclusive to this cell type. Moreover, MT+ and IFI27+ AMs (DR 48+ and DR 64-) are not disjoint biological subtypes, indicating DRVI's superiority over traditional clustering methods in representing parallel overlapping processes.

DR 46+: A cell subset characterized by the expression of CXCL9, CXCL10, and CXCL11 is highlighted by DR 46+. GSEA analysis suggests an association with CXCR3 chemokine receptor binding Tokunaga et al. (2018), which is particularly relevant in the context of lung diseases Kameda et al. (2020); Callahan et al. (2021). This dimension is not specific to any particular cell cluster, indicating a shared response signature.

DR 22+: DR 22+ is primarily active in endothelial cells, with GSEA analysis suggesting an association with angiogenesis. The upregulation of top relevant genes, including CX3CL1, SERPINE1, AKAP12, GRP4, and FSTL3 under hypoxic conditions Korbecki et al. (2020); Azimi et al. (2017); Finger et al. (2015); Qi et al. (2021); Biron-Shental et al. (2008), a common phenomenon observed in tumor-adjacent environments Chen et al. (2023), indicates a potential role in tumor growth. The increased abundance of tumor-adjacent endothelial cells in the higher end of DR 22+ further supports this interpretation. (Figure 2h)

DR 27+: This dimension indicates an inflammatory process in myeloid cells. The primary gene associated with this dimension is EREG, with genes such as IL1B, IL1A, CXCL8, and CXCL3 upregulated in this subpopulation. GSEA analysis suggests an association with the regulation of apoptotic signaling and interleukin-10 signaling. As mentioned earlier, there are other dimensions representing inflammation, each with its own specific gene signature.

DR 30+: This dimension highlights a gradient in ciliated cells. DNAH11, DNAAF1, DNAH12, and DNAH6 are among the top identifiers of this dimension, expressed across ciliated cells, with significantly higher expression observed in the upper gradient of DR 30+. GSEA also suggests cilium for this dimension.

DR 16+: All the genes identified for DR16+ are human leukocyte antigen gene complex (HLA) genes, including HLA-DRA, HLA-DRB1, and HLA-DPA. GSEA suggests that this dimension corresponds to the MHC class II protein complex. Notably, while HLA genes are generally more highly expressed in certain cell types like myeloid cells, this dimension provides a per-cell-type normalized representation, covering similar ranges across relevant cell types, such as myeloid cells and blood vessels.

DR 21-: This dimension highlights a subset of epithelial cells mostly active in nasal goblet cells, identified by KLK7, TMPRSS11B, SPINK7, ECM1, SCEL, and KLK6 as their top relevant genes. We could not identify a biologically meaningful explanation for this dimension.

DR 24+: FOLR2, F13A1, STAB1, CCL13, and SELENOP are the main identifiers of the gene program associated with DR 24+. These genes indicate a perivascular population of macrophages Bailey et al. (2023); Nalio Ramos et al. (2022). High activity of this dimension in Interstitial Mph perivascular cells confirms this relevance.

DR 25-: SERPINB3 and SERPINB4 are the top identified gene for DR 25-, active in epithelial cells. Although the coexpression of these two genes is biologically relevant Sun et al. (2017), we could not find a specific biological explanation for the subpopulation of cells active in this dimension.

DR 28+: The activity of DR 28+ indicates the expression of PLAU, its receptor PLAUR, and an elevated expression of CLDN4 (Claudin-4) in airway epithelium. While PLAU and CLDN4 are separately suggested as markers of lung injury or infection, we could not find a comprehensive description for this dimension Lee et al. (2018); Martínez et al. (2007); Ohta et al. (2013).

DR 29-: This dimension highlights a subset of epithelial cells, including Hillock-like cells. Our interpretability pipeline identified KRT6B, KRT13, KRT16, and KRT6A as key markers, suggesting a keratization process within this dimension. Previous studies (Ruiz García et al. (2019); Yoshida et al. (2022)) have employed a diverse set of KRT genes to characterize epithelial subpopulations. In particular, KRT16 and KRT6A are among the markers of certain Hillock-like cell populations.

DR 33+: EDN1, ERRFI1, DST, and TNFRSF12A are the main indicators of DR 33+. This dimension is more active in basal cells, with a significant upregulation of cells from the Nawijn_2021 dataset. However, a clear technical or biological explanation for this dimension remains open.

DR 35-: This dimension indicates a subset of fibroblast lineage and smooth muscle cells identified by a gene program composed of genes such as CRISPLD2, IL6, and ADAMTS4. IL-6 is a pro-inflammatory cytokine, reported to induce expression of ADAMTS4 in rheumatoid arthritis samples Mimata et al. (2012). Conversely, CRISPLD2 is a known inhibitory regulator of IL-6 Zhang et al. (2016); Himes et al. (2014). The coexpression of these genes suggests a dynamic of IL-6 regulation. The exact role of this process needs further experiments.

DR 35+: C1QA, C1QB, C1QC, and TREM2 are the top indicators of the DR 35+, Suggesting that this dimension represents TREM2 and the C1Q complex in myeloid cells, including TREM2+ macrophages Xiong et al. (2020a); Wu et al. (2015); Xiong et al. (2020b).

DR 36-: LYPD3, CALML3, AQP3, TASCTD2, and LY6D are the genes identifiers of the nonlinear gene program for DR 36-. The role of this dimension is not identified.

DR 37-: ANKRD36C gene is the main driver of the DR 37- expressed only in a subset of goblet cells. Comparison with cell types suggests that this dimension is closely related to the Bronchial subsets of the Goblet cells. However, we could not find any evidence of the validity of such a subcluster in the literature, and this dimension seems to be dataset-specific.

DR 38+: DR 38+, characterized by MKI67, TYMS, and PCLAF expression, is associated with cell-cycle processes as identified by GSEA. This aligns with the cell-type annotations.

DR 40+: This dimension highlights a subset of airway epithelial cells, with RARRES1, CXCL6, SERPINB4, and IL17C as the key drivers of its gene expression program. While GSEA suggested an IL-17 signaling pathway, we found the relationship between these genes and the pathway to be weak. As a result, we were unable to identify a meaningful explanation for this dimension.

DR 41+: This dimension identifies a subset of epithelial cells, with MMP1, STC1, MMP13, and MMP10 as their top relevant genes. Accordingly, this dimension highlights the activity of certain proteins in the Matrix Metalloproteinase family, suggesting an ongoing destructive process Houghton (2015).

DR 42+: With genes such as CCL20, CXCL2, CXCL3, CSF2, CSF3, and ICAM1 among the top associated genes, GSEA identifies IL-17, IL-10, and TNF signaling pathways for DR 42+. This suggests an ongoing inflammation process Bui et al. (2020).

DR 43+: This dimension highlights two subsets of cells. The tiny subset in Myeloid cells is associated with genes such as OLFM4, S100A7, and DEFA3, suggesting a small population of neutrophils Liu & Rodgers (2016); Jarczak et al. (2013). The other subset in airway epithelium is targeted by S100A7, SPRR2D, SPRR1B, SPRR2E, SPRR2F, and SPRR3 in the top relevant genes. GSEA suggests keratinization and formation of the cornified envelope for this dimension. The coexistence of two unrelated cell subsets within this dimension suggests a potential failure in disentanglement, possibly due to insufficient latent dimensions.

DR 44+: This dimension indicates a subset of myeloid cells characterized by but not limited to CCL3, CCL4, CCL3L1, CCL4L2, CXCL3, IL1A, and IL1B genes. CCL3 (MIP-1α) and CCL4 (MIP-1β) and their paralogs CCL3L1 and CCL4L2, are proinflammatory chemokines that attract immune cells Xu et al. (2023). IL-1β is an inflammatory cytokine that activates cells to produce more CCL3 and CCL4 chemokines Sokol & Luster (2015). GSEA describes this dimension by

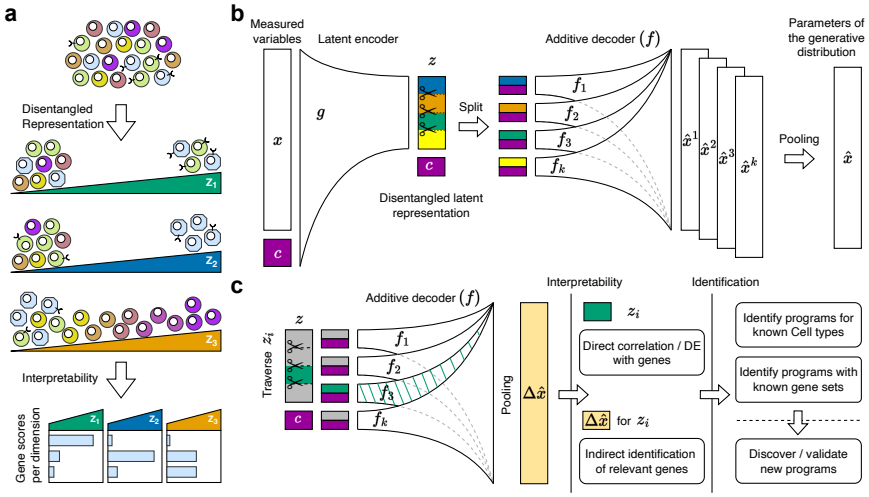
pro-inflammatory process and cytokine activity. This subpopulation of myeloid cells is of biological and clinical interest Zhou et al. (2024). It is worth mentioning that DR 27+ complements this by highlighting another population of myeloid cells undergoing interleukin-1 signaling pathway.

DR 49-: With HBA2, HBA1, and HBB as its top identifiers, GSEA analysis suggests that DR 49- represents the hemoglobin metabolic process. While this process is expected to be most active in erythroid cells that are not present in the HLCA, we observed non-negligible expression of hemoglobin genes in other cell types. The expression of hemoglobin genes in the HLCA could be attributed to technical or biological factors Yang et al. (2019); Marozkina et al. (2021).

Dim DR 50+: SAA1 and SAA2 are the main identifiers of the dim DR 50+. So, we can conclude that this dimension presents the activity of SAA as an inflammatory marker.

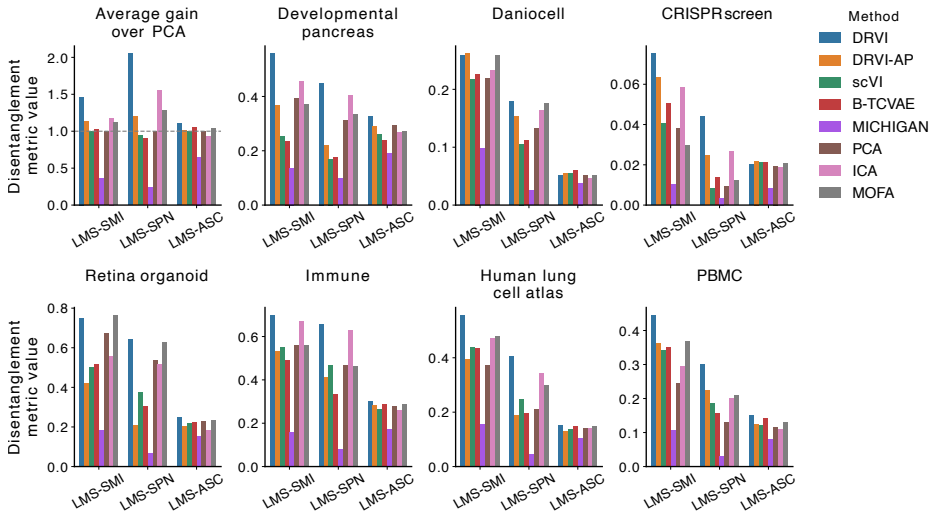
DR 54-: The nonlinear gene program associated with DR 54- includes multiple genes such as ISG15, IFIT1, IFI44L, and IFI6. GSEA describes this dimension by interferon alpha/beta signaling and response to virus pathways.

A.5 VISLUALIZATION OF THE MODEL ARCHITECTURE

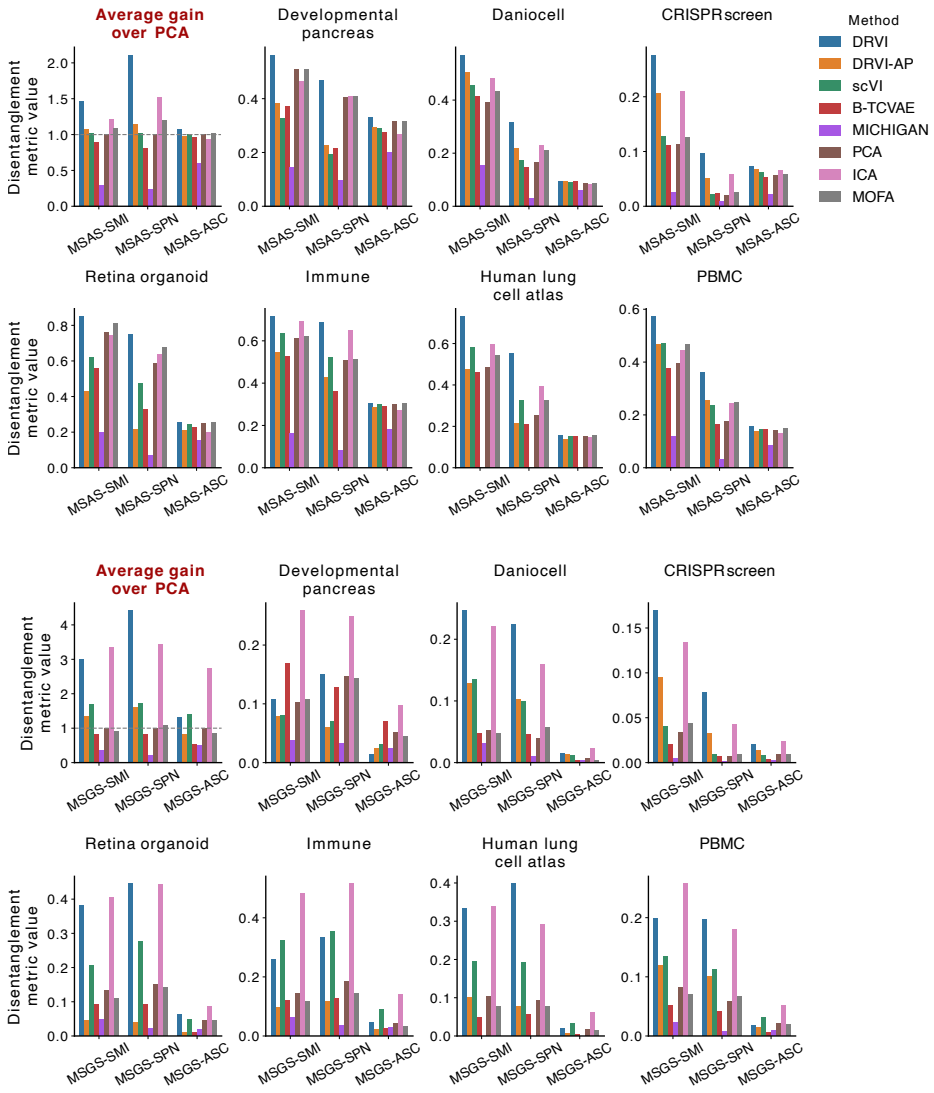


Supplemental Figure 1: **DRVI enables learning interpretable disentangled latent representations. a,** Disentanglement concept. A disentangled latent model should encode concepts such as cell properties, cell types, and developmental stages in distinct dimensions. Each dimension can then be interpreted as the effect of the dimension on the gene expression. **b,** Schematic of the DRVI model. The model consists of a nonlinear encoder and an additive decoder followed by a pooling function that enforces the disentanglement into the encoder and latent factors. **c,** Interpreting disentangled factors. Relevant genes can be discovered by traversing a latent factor or using DE analysis tools. While some factors can be identified using known gene sets or known cell type annotations, others can be validated as potentially novel programs.

A.6 RESULTS FOR ADDITIONAL METRICS

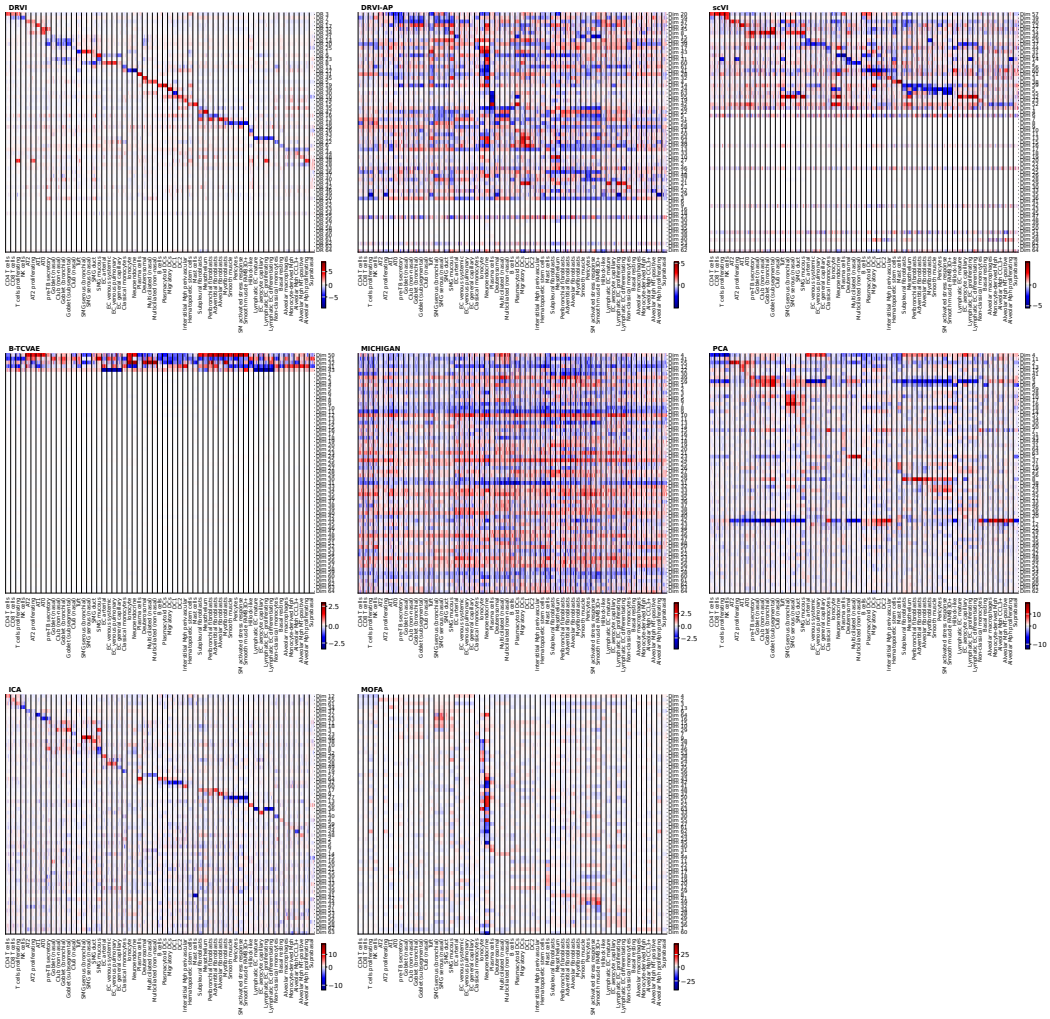


Supplemental Figure 2: **Results of the primary disentanglement metric family.** The results of the LMS metrics, as primary metric family, over different datasets are provided. LMS measures the ability of the model to capture biological processes in individual latent dimensions by matching latent dimensions and ground-truth biological processes. ASC, absolute Spearman correlations; LMS, latent matching score; SMI, scaled mutual information; SPN, same-process neighbors.

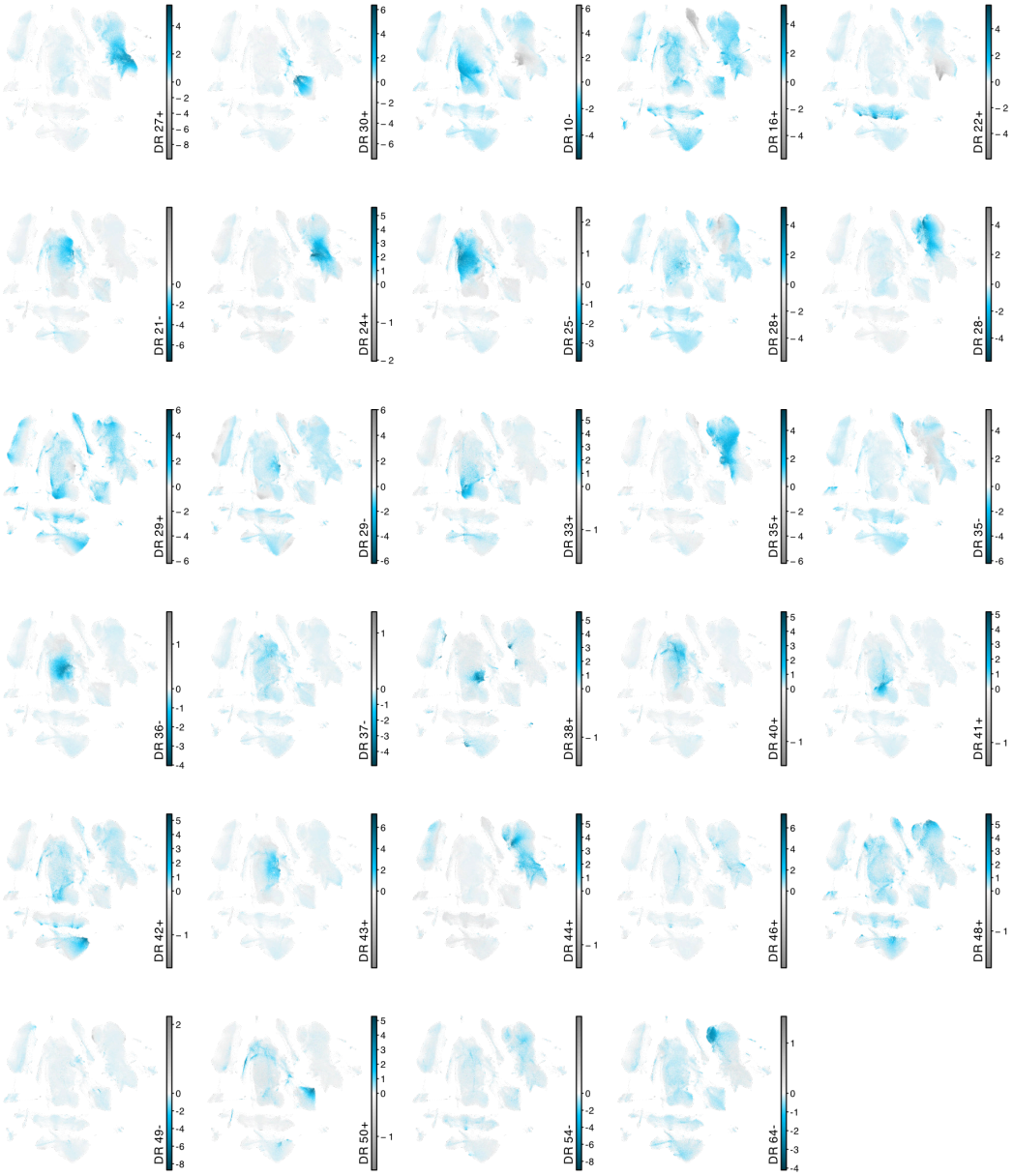


Supplemental Figure 3: **Results of the secondary disentanglement metrics.** The results of the MSAS and MSGS as secondary metrics over different datasets are provided. MSAS shows the capability of the model to capture biological processes in individual latent dimensions regardless of the redundancy. MSGS measures the amount of captured biological information that is specific to a single (probably overlapping) dimension by penalizing redundancy. Unlike LMS as the main metric, neither MSAS nor MSGS penalizes overlapping processes in a single dimension. ASC, absolute Spearman correlations; LMS, latent matching score; MSAS, Most Similar Averaging Score; MSGS, Most Similar Gap Score; SMI, scaled mutual information; SPN, same-process neighbors.

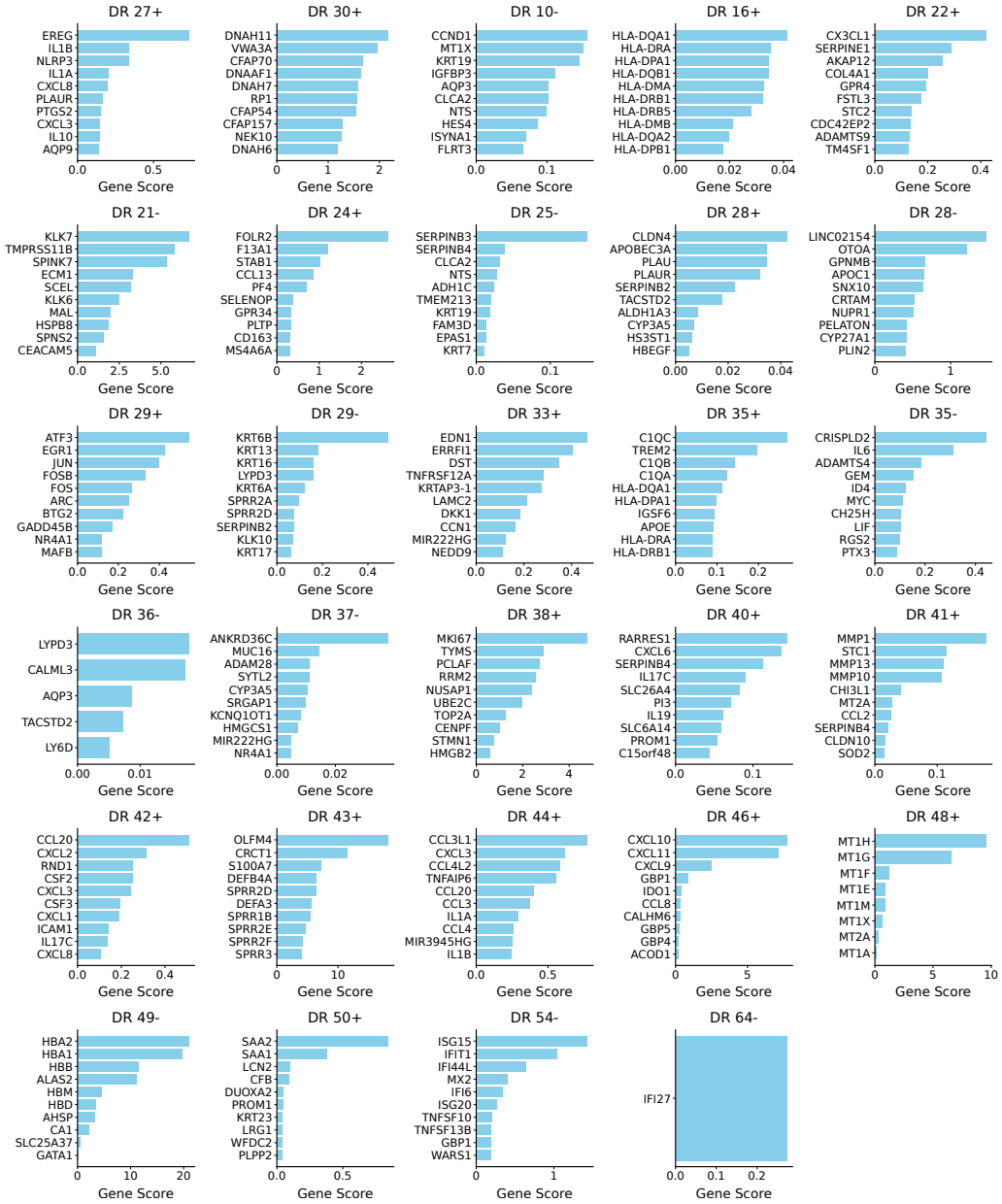
A.7 ADDITIONAL PLOTS FOR HLCA EXAMPLE



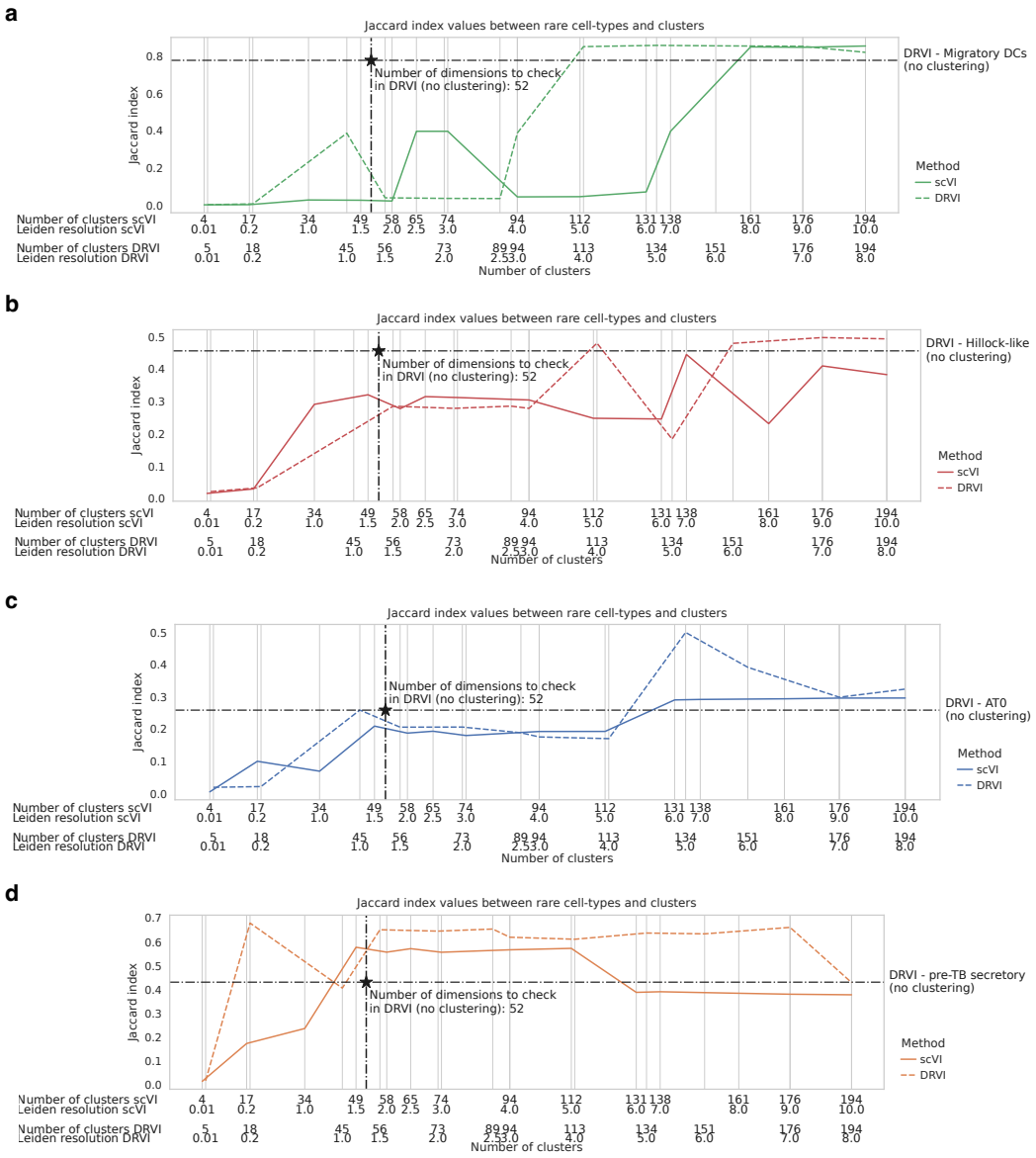
Supplemental Figure 4: **Latent space visualization for benchmarked methods on the HLCA.** For each method, a heatmap visualizing latent representation activity across different cell types is provided. Cells were subsampled to ensure a balanced number of cells from each cell type. Dimensions are reordered using a shared algorithm to maximize the diagonal appearance.



Supplemental Figure 5: **Activity of DRVI factors representing non-cell-type processes in the HLCA on UMAP** For dimensions representing non-cell-type processes, the activity on the UMAP is provided.



Supplemental Figure 6: **Interpretability of DRVI latent factors corresponding to non-cell-type processes in the HLCA.** The top relevant genes and their scores are plotted for each non-cell-type latent factor of DRVI trained on the HLCA.



Supplemental Figure 7: **Rare Cell-Type Identification Performance of Leiden Clustering in the HLCA.** Leiden clustering was applied to DRVI and scVI embeddings at varying resolutions. For each cell type, we calculated the Jaccard index between the cell type and its most relevant cluster at different resolutions. Additionally, we evaluated the performance of DRVI by calculating the Jaccard index between the cell type and the separation achieved using a simple threshold on the most relevant DRVI dimension (marked by star). Migratory DCs were accurately captured by the Leiden algorithm only at the resolution of 8.0 with 161 clusters. AT0 cells were identified with comparable quality to DRVI starting from the resolution of 6.0 with 131 clusters. Hillock-like cells were captured as well as DRVI at the resolution of 7.0 with 138 clusters. Pre-TB secretory cells were captured well at a resolution of 1.5 with 49 clusters. These results demonstrate that identifying at least three rare cell types requires the creation of approximately 130 to 160 clusters using Leiden clustering. Given that DRVI can capture these cell types within its 52 non-vanished dimensions, we suggest DRVI as an efficient valuable tool for seed annotation and discovery of rare cell populations.