Disentangling latent representations of single cell RNA-seq experiments

Jacob C. Kimmel
Calico Life Sciences, LLC
South San Francisco, CA, 94080
jacob@calicolabs.com

Abstract

Single cell RNA sequencing (scRNA-seq) enables transcriptional profiling at the resolution of individual cells. These experiments measure features at the level of transcripts, but biological processes of interest often involve the complex coordination of many individual transcripts. It can therefore be difficult to extract interpretable insights directly from transcript-level cell profiles. Latent representations which capture biological variation in a smaller number of dimensions are therefore useful in interpreting many experiments. Variational autoencoders (VAEs) have emerged as a tool for scRNA-seq denoising and data harmonization, but the correspondence between latent dimensions in these models and generative factors remains unexplored. Here, we explore training VAEs with modifications to the objective function (i.e. β-VAE) to encourage disentanglement and make latent representations of single cell RNA-seq data more interpretable. Using simulated data, we find that VAE latent dimensions correspond more directly to data generative factors when using these modified objective functions. Applied to experimental data of stimulated peripheral blood mononuclear cells, we find better correspondence of latent dimensions to experimental factors and cell identity programs, but impaired performance on cell type clustering.

***K*eywords** single cell RNA-seq · variational autoencoder · VAE · disentangle

Publication Status: This pre-print represents the final output of a preliminary research direction and will not be updated or published in an archival journal. We are happy to discuss future directions we believe to be promising with any interested researchers.

1 Introduction

scRNA-seq experiments can capture many sources of biological variation, including differences between cell identities, responses to perturbations, and developmental programs [1, 2, 3]. The atomic units of an RNA-seq experiment are read counts for individual gene transcripts. However, many processes of interest in biology involve the interaction of many genes in coordinated gene expression programs (GEPs) and may be better represented using a smaller number of dimensions. Many dimensionality reduction methods have been proposed for scRNA-seq data [4, 5, 6, 7], including the recent introduction of variational autoencoder (VAE) based methods [8, 9, 10]. While VAEs have several desirable properties, the latent spaces they learn to encode may be difficult to interpret.

Recent work on VAEs has attempted to encourage "disentangled" latent spaces which may be more interpretable. A disentangled latent space has direct correspondence between dimensions in the latent space and generative factors – parameters of the underlying process that generated the observed data [11]. In the case of single cell RNA-seq data, we may imagine that cell identity, cell cycle state, and the activity of other gene expression programs constitute generative factors. Methods to enforce disentanglement in VAEs largely focus on modifying the objective function [12]. Here, we explore using one of these disentanglement techniques (β-VAE) to encourage disentanglement in VAE latent spaces learned for scRNA-seq data. Leveraging simulated data where ground truth values for generative factors are known, we find that these methods improve the correspondence between dimensions of the latent space and generative factors.

1.1 Variational Autoencoders

Variational autoencoders (VAEs) learn a generative model of observed data $\mathbf{X}$ by taking advantage of a lower dimensional latent space $\mathbf{Z}$. Briefly, VAEs jointly learn to map observations $x \in \mathbf{X}$ to points $z \in \mathbf{Z}$ ("encoding") and to perform the inverse mapping ("decoding"). This two-way mapping is enabled by a flexible encoder $q(z|x)$ and decoder $p(x|z)$, often implemented as neural networks. The encoder posterior $q(z|x)$ is regularized to match a prior distribution on the latent variables, $p(z)$.

VAEs are trained by jointly optimizing (1) the reconstruction error of observations x and (2) divergence of the latent distribution $q(z|x)$ from the prior $p(z)$. The VAE objective is traditionally formulated with two components, each addressing one of these desired properties [13].

$$\mathcal{L}(\theta, \phi, x) = \mathbf{E}_{q_\phi(z|x)}[-\log p_\theta(x|z)] + D_{\text{KL}}(q_\phi(z|x)||p(z))$$

Models of this form have recently been applied to single cell RNA-seq data by multiple groups, yielding effective approaches for gene expression denoising, visualization, and data harmonization [8, 9, 10].

1.2 Encouraging disentanglement in the VAE latent space

A standard choice for a Gaussian prior on the latent space $p(z) = \mathcal{N}(0, \mathbf{I})$ promotes independence among the latent dimensions, since the covariance matrix is diagonal. The β-VAE approach [14] leverages this property to enforce independence between dimensions of the latent distribution and encourage disentanglement. This is achieved by weighting the KL-divergence between the prior and latent posterior in the standard VAE objective by a coefficient $\beta > 1$.

An addition to the β-VAE framework encourages an explicit divergence from the prior distribution that increases during the course of training [15]. This is implemented by penalizing the difference between a "channel capacity" constant C and the KL-divergence. Together these modifications add two additional parameters to the objective function:

$$\begin{aligned}\mathcal{L}(\theta, \phi, x) = \mathbf{E}_{q_\phi(z|x)}&[-\log p_\theta(x|z)] \\ &+ \beta|D_{\text{KL}}(q_\phi(z|x)||p(z)) - C|\end{aligned}$$

2 Experiments

Here, we use the recently introduced Single Cell Variational Inference (scVI) framework[9] as a baseline model, and augment the standard objective function during training by either (1) altering the β parameter and/or (2) increasing the channel capacity parameter C to encourage disentanglement in the latent space. In this parameter space, $\beta = 1$ and $C = 0$ represents a baseline scVI model.

For all experiments, we use $n = 128$ units in both the encoder and decoder layers of scVI. We set the number of latent dimensions to 32 and train for 400 epochs using the Adam optimizer with a learning rate of 10^{-4} followed by 200 epochs with a learning rate of 10^{-5}. For experiments where $C > 0$, we linearly increase the channel capacity from $C = 0.1$ to the maximum value over $20,000$ iterations. We fit models for 10 random starts and average evaluation metrics to account for stochasticity in optimization.

2.1 Simulated Data

We simulate an scRNA-seq experiment using the Splatter statistical framework [16, 17]. We simulate 10,000 cells with 25,000 genes from 5 cell types. Each cell type is defined by expression of a "cell identity" GEP. We also simulate an "activity" GEP which is utilized within 3 of the 5 cell types (Fig. 1A).

This dataset provides a ground truth (1) cell identity, and (2) gene sets associated with each GEP. Formally, we define a GEP as a group of genes that are co-differently expressed. Each gene in a GEP is scaled by a differential expression coefficient D where $\log D \sim \mathcal{N}(2, 1)$ in cells where that GEP is active. To quantify GEP utilization in each cell, we compute a rank-based score in the same manner as AUCell [18] that we refer to as a "GEP score". We consider the scores of these ground truth GEPs to be generative factors in the data which we wish to capture in dimensions of the latent space (Fig. 1B).

2.2 Recovering generative factors in simulated data

To determine if a modification to the VAE objective improves interpretability of the latent space, we require a quantitative metric for interpretability. Quantitative metrics for disentanglement are still an active area of research and no well-defined standard exists [12]. Here, we focus on a simplistic metric to evaluate the correspondence between latent dimensions and ground truth GEPs. For each ground truth GEP, we compute Spearman correlations ρ of the GEP score with each dimension of the latent space. We consider the maximum absolute correlation to reflect the best correspondence between a latent dimension and a GEP. This metric does not reflect overall disentanglement in the latent space, which inherently must consider the uniqueness of correspondence between latent dimensions and generative factors.

We fit scVI models with varying values for β and C. We find that increasing $\beta > 1$ while holding $C = 0$ significantly increases the correlation between ground truth GEPs and latent dimensions at some values ($\beta = 10$, t-test, $q < 0.05$, Benjamini-Hochberg). However at higher values ($\beta = 50$), decreased performance is observed for some ground truth GEPs (Fig. 2A). Increasing the channel capacity to $C = 10$ ameliorates the detrimental effect from larger values of β, although the maximum correlation between some GEPs and a latent dimension is decreased (Fig. 1C).

To determine if highly correlated latent dimensions correspond to specific ground truth GEPs, we visualize the values of the latent dimension most correlated with each ground truth GEP in a UMAP projection. Using the baseline $\beta = 1$, we find that latent dimensions are not specific for a ground truth GEP. Using $\beta = 10$ appears to improve

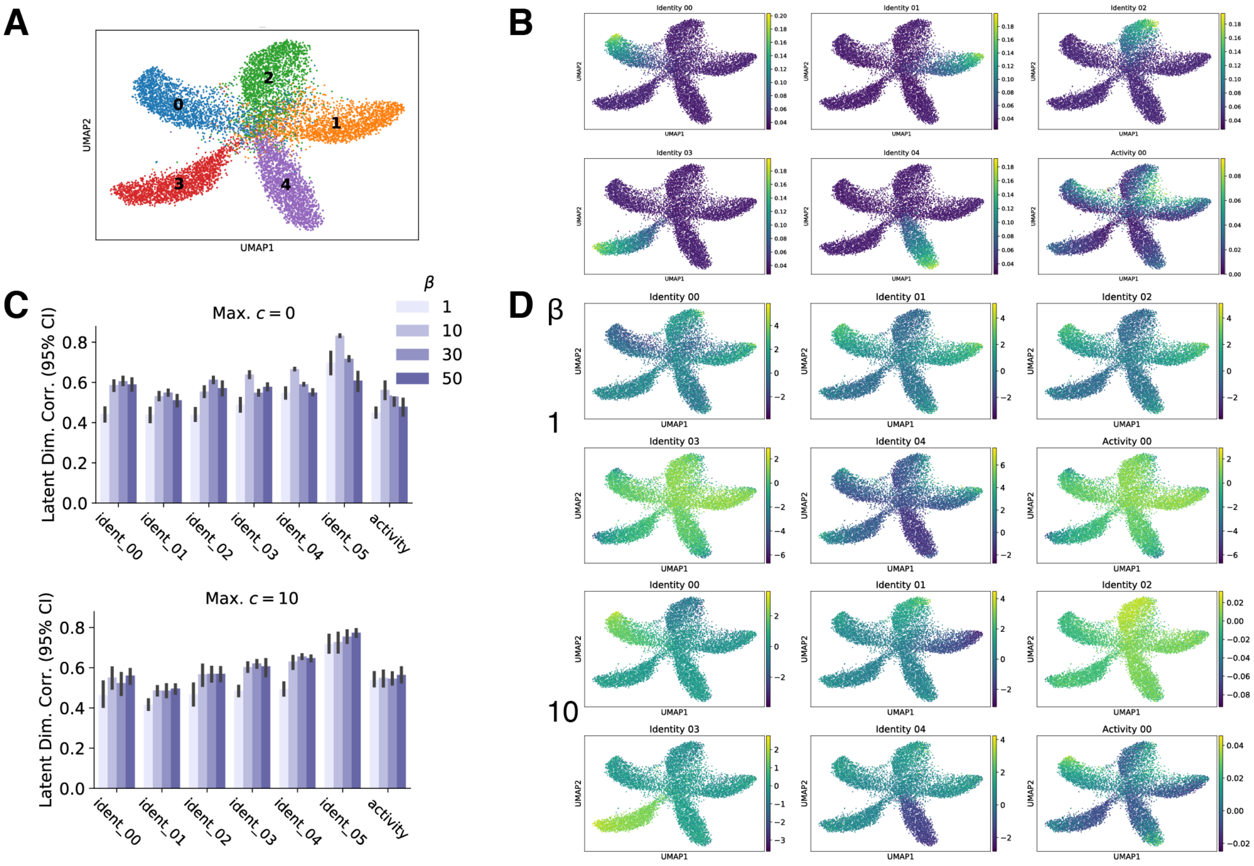

Figure 1: **Increasing β improves correspondence between simulated GEPs and VAE latent dimensions. (A)** UMAP projection of simulated single cell data. Ground truth cell type identities for each cell are overlaid in color. **(B)** Ground truth rank-based GEP score of the 5 cell type identity programs and the "activity" program in each cell. **(C)** Maximum absolute Spearman correlation of GEP scores with dimensions of the latent space as a function of β. **(D)** Values for the latent dimension with the highest correlation with each GEP are presented on a UMAP projection.

the correspondence of latent dimensions to ground truth GEPs. We note that with $\beta = 10$, the most correlated latent dimensions for each cell identity GEP appear to specifically mark that cell identity (Fig. 1D). For the latent space learned with $\beta = 10$, note that the dimensions for identities 1 and 5 are inverse mappings and that the dimension for identity 2 shows a less dramatic correspondence than other GEP:dimension pairs.

2.3 Recovering experimental perturbations and cell identity in PBMCs

To determine if modified VAE objectives can capture cell identity programs and experimental factors, we trained scVI models on experimental data from peripheral blood mononuclear cells (PBMCs) [19]. The data contain 8 unique cell types, each of which is observed before and after stimulation with IFNβ (Fig. 2A). We treat the IFNβ experimental condition as a generative factor we wish to recover. As before, we evaluate the maximum Spearman correlation between latent dimensions and IFNβ condition for a range of β. We find that increasing β leads to modest improvements in this correlation, but even $\beta = 1$ models learn a dimension with strong correlation (Fig. 2B). Visualizing the latent dimensions that correspond best to IFNβ condition confirms that even the $\beta = 1$ models learn a dimension that segregates the experimental conditions (Fig. 2E).

Latent spaces are also used for unsupervised cell type identification by Louvain community detection [20]. We evaluate this method in each latent space using the Adjusted Rand Index (ARI) between the Louvain partition and ground truth cell types. We use a range of resolutions and take the maximum ARI to mimic human adjustment based on visualization. We find that unsupervised clustering efficacy decreases as β increases, suggesting that stronger regularization may be undesirable for some downstream tasks (Fig. 2C).

To determine if we recover cell identity GEPs in each latent space, we fit logistic regression models to classify each cell type based on each individual latent dimension. We fit each regression model to distinguish a single target cell type (i.e. B cells) from all other cell types. We perform

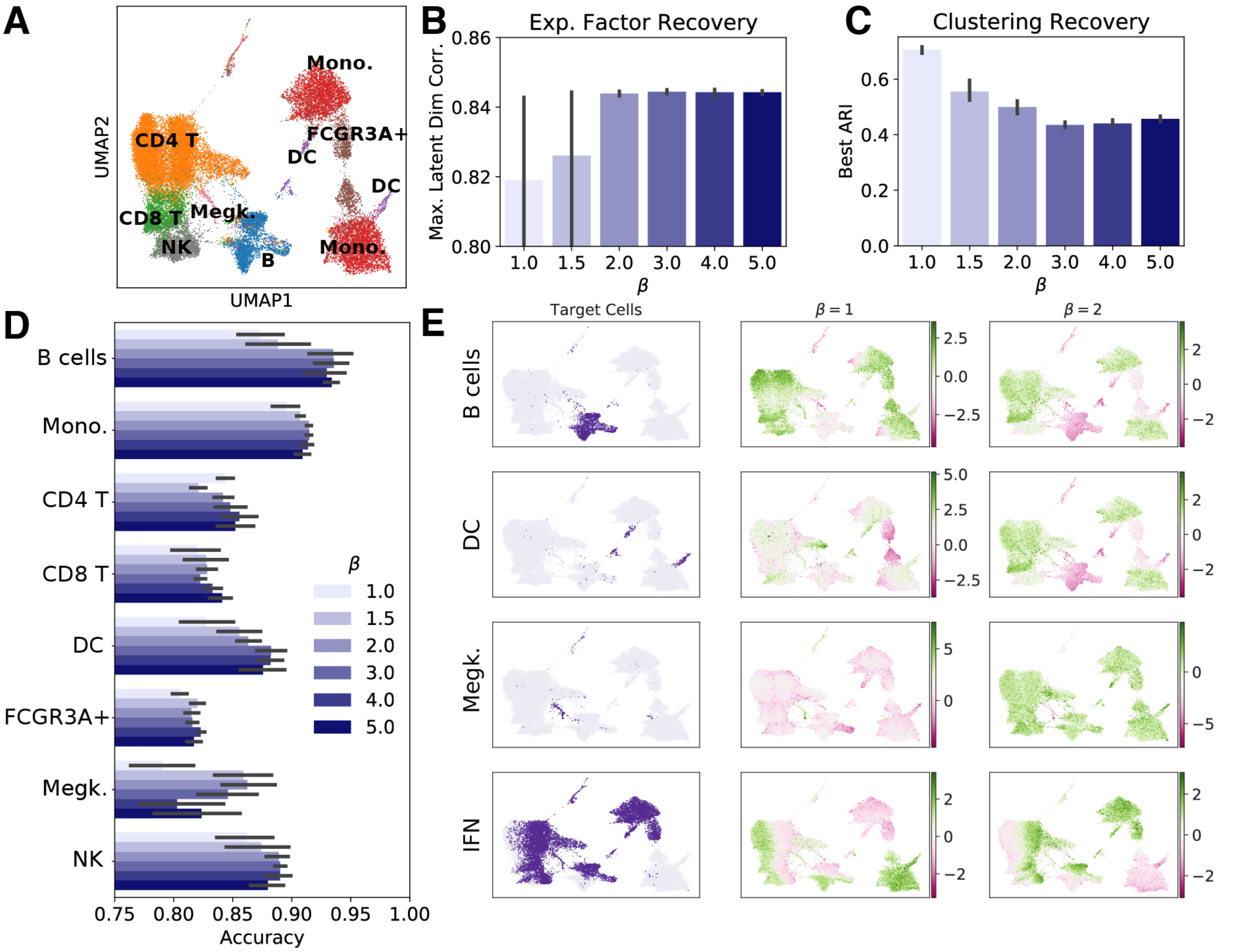

Figure 2: **Increasing β modestly improves experimental factor recovery but impairs cell type clustering. (A)** UMAP projection of PBMC data with cell type labels. **(B)** Correlation of latent dimensions with IFNβ treatment status. **(C)** Adjusted Rand Index for cell type recovering by community detection in latent spaces. **(D)** Max logistic regression cell type classification accuracy (mean 5-fold CV) based on a single latent dimension. **(E)** Values of the latent dimension with the best correspondence to each cell identity program or IFNβ condition (rows) are visualized in UMAP projections. Cells that should be distinguished by each latent dimension are highlighted on the left, and the best dimension from VAEs with $\beta = 1$ (center) or $\beta = 2$ (right) are shown.

class balancing before fitting and report accuracy as the mean of 5-fold cross-validation. Here, we assume that a latent dimension representing a cell identity program will allow for better classification of the corresponding cell type. We consider the dimension with the maximum classification accuracy to have the best cell identity program correspondence.

We find that increasing $\beta > 1$ improves the correspondence of latent dimensions and cell identity programs by this metric for most cell types (Fig. 2D, E). However, counterexamples also exist – we find that correspondence between latent dimensions and the CD4 T cell identity program decreases for some values of $\beta > 1$. Taken together, these results suggest that modifying the VAE objective can improve correspondence between latent dimensions and some generative factors, but may also decrease performance on some downstream tasks like cell type clustering.

3 Conclusions

We find that a modified VAE objective (β-VAE) designed to encourage disentanglement improves the correspondence between VAE latent dimensions and ground truth gene expression programs in simulated data. Applied to experimental data, we observe modest improvements in the correspondence of latent dimensions with experimental conditions and cell identity programs. However, we also find that the performance of cell type clustering decreases in the same conditions.

These results suggest that fitting VAE models to scRNA-seq data with the β-VAE objective may improve the interpretability of latent spaces, but that these modifications may decrease performance on some downstream tasks. We note that the simplistic metrics of correspondence we employ here do not measure disentanglement directly. We believe these results motivate further research on the application of disentanglement methods to single cell RNA-seq models. Multiple groups have proposed alternative formulations of the VAE objective that may ameliorate the detrimental effects of disentanglement methods we observe here. Likewise, additional quantitative metrics of disentanglement have been proposed that may more accurately identify alignment of latent variables with generative factors [21, 22, 23]. Application of these techniques may prove fruitful and remains an exciting direction for future work.

References

[1] Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, John J Trombetta, David A Weitz, Joshua R Sanes, Alex K Shalek, Aviv Regev, and Steven A McCarroll. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5):1202–1214, May 2015.

[2] Allon M Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*, 161(5):1187–1201, May 2015.

[3] Alex K Shalek, Rahul Satija, Joe Shuga, John J Trombetta, Dave Gennert, Diana Lu, Peilin Chen, Rona S Gertner, Jellert T Gaublomme, Nir Yosef, Schraga Schwartz, Brian Fowler, Suzanne Weaver, Jing Wang, Xiaohui Wang, Ruihua Ding, Raktima Raychowdhury, Nir Friedman, Nir Hacohen, Hongkun Park, Andrew P May, and Aviv Regev. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, 510(7505):263–269, June 2014.

[4] 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. *Nature Biotechnology*, 37(1):38–44, December 2018.

[5] Gökcen Eraslan, Lukas M Simon, Maria Mircea, Nikola S Mueller, and Fabian J Theis. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun*, pages 1–14, January 2019.

[6] Christopher Yau. ZIFA: Dimensionality reduction for zero-inflated single cell gene expression analysis. *Genome Biol*, pages 1–10, October 2015.

[7] Philipp Angerer, Laleh Haghverdi, Maren Büttner, Fabian J Theis, Carsten Marr, and Florian Buettner. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*, 32(8):1241–1243, April 2016.

[8] Jiarui Ding, Anne Condon, and Sohrab P Shah. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nat Commun*, 9(1):2002, May 2018.

[9] Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, pages 1–11, November 2018.

[10] Chenling Xu, Romain Lopez, Edouard Mehlman, Jeffrey Regier, Michael I Jordan, and Nir Yosef. Harmonization and Annotation of Single-cell Transcriptomics data with Deep Generative Models. *bioRxiv*, pages 1–46, January 2019.

[11] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: a review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, August 2013.

[12] Michael Tschannen, Olivier Bachem, and Mario Lucic. Recent Advances in Autoencoder-Based Representation Learning. December 2018.

[13] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv*, 2013.

[14] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. *ICLR*, 2017.

[15] Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in β-VAE. *arXiv*, April 2018.

[16] Luke Zappia, Belinda Phipson, and Alicia Oshlack. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol*, 18(1):174, September 2017.

[17] Dylan Kotliar, Adrian Veres, M Aurel Nagy, Shervin Tabrizi, Eran Hodis, Douglas A Melton, and Pardis C Sabeti. Identifying Gene Expression Programs of Cell-type Identity and Cellular Activity with Single-Cell RNA-Seq. *bioRxiv*, pages 1–43, November 2018.

[18] Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, Joost van den Oord, Zeynep Kalender Atak, Jasper Wouters, and Stein Aerts. SCENIC: single-cell regulatory network inference and clustering. *Nature methods*, 14(11):1083–1086, October 2017.

[19] Hyun Min Kang, Meena Subramaniam, Sasha Targ, Michelle Nguyen, Lenka Maliskova, Elizabeth McCarthy, Eunice Wan, Simon Wong, Lauren Byrnes, Cristina M Lanata, Rachel E Gate, Sara Mostafavi, Alexander Marson, Noah Zaitlen, Lindsey A Criswell, and Chun Jimmie Ye. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nature Biotechnology*, 36(1):89–94, December 2017.

[20] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008.

[21] Michael Tschannen, Olivier Bachem, and Mario Lucic. Recent advances in autoencoder-based representation learning. *CoRR*, abs/1812.05069, 2018.

[22] Hyunjik Kim and Andriy Mnih. Disentangling by factorising. *arXiv*, 1802.05983, 2018.

[23] Ricky T. Q. Chen, Xuechen Li, Roger Grosse, and David Duvenaud. Isolating sources of disentanglement in variational autoencoders. *arXiv*, 1802.04942, 2018.