Continual single-cell architecture surgery for reference mapping

Soroor Hediyeh-zadeh [1] **Mohammad Lotfollahi** [1 *] **Fabian J. Theis** [1 2 *]

Abstract

The recent emergence of large-scale integrated single cell atlases allow to reformulate many analysis steps in novel single cell transcriptome data as a reference mapping problem. Current deep learning mapping approaches result in a fixed, non-linear function of input gene expression learned from the reference, which is then used to project new query datasets. These methods, therefore, assume that major axes of biological variations are shared between query and the reference. This does not hold when applying such methods to study and compare single cells from perturbation experiments, disease traits or organoids in the context of control cells catalogued in the reference atlas. In this work, we aim to explore continual learning as a means to adapt more flexibly to domain shifts. In particular we introduce a Continual Learner Conditional Variational Autoencoder (CLCVAE), as a architecture surgery optimization strategy for continuously learning new variations in the query to address the challenge of single-cell reference atlas mapping for case-control and perturbation studies, and report improvements over the standard architecture surgery in identification of cell types in the query that are not present in the reference.

1. Motivation: Mapping disease samples to healthy single-cell references is challenging

The existing deep learning frameworks for reference atlas construction and mapping use Conditional Variational Autoencoder (CVAE) (Sohn et al., 2015) models to integrate the data by depleting the lower-dimensional representation, that is the latent representation, of the data learned by the autoencoder from batch effects and dataset-specific variations. During reference construction, the CVAE model learns a non-linear function, that is non-linear weighted combinations, of input gene expression, as well as conditional weights that capture the batches in the reference. During reference mapping, specifically in architecture surgery (aka Transfer Learning) framework proposed by Lotfollahi et al. (2022), the non-linear *projection function* and conditional weights learned on the reference are fixed, and new conditional weights are added to the model to capture new batches added to the data. The new conditional weights are learned such that overall the model results in decent reconstruction of input data. Therefore, learning these new conditional weights serves the purpose of data alignment only, and in practice the new data is projected using a fixed, non-linear function learned from the reference.

The recent Human Lung Cell Atlas (HLCA) study (Sikkema et al., 2022) integrated a vast number of normal lung single-cell datasets and extended this reference atlas with cells from lung disease and cancer cells using the Transfer Learning framework of Lotfollahi et al. (2022) to study disease-associated cell types and states, and compare biomarkers in normal and malignant cells. Projection of cancer and disease cells worked reasonably well in this study and the annotations transferred from the reference to the projected data were found to be mostly correct (e.g. cancer cells were correctly annotated as *unknown* since they were absent in the reference), despite the biological variations not being present in the reference, perhaps owed to the highly non-linear nature of the model. These findings though were based on a, rather modest, uncertainty score cut-off (0.3), which was chosen from ROC curves. One has to note that should they have gone with a higher cut-off, say 0.5, these cancer cells would have been miss-annotated as normal lung or immune cells. Indeed, the study flagged that incorrect annotation transfers (annotations transferred with high certainty not matching the original label) were frequently observed for cell types that are part of a continuum, and those that were not present in the reference, but had high transcriptional similarity to cell types present in the reference. Overall the HLCA effort highlighted challenges in single-cell reference atlas mapping that become very relevant as interest in mapping disease samples to large healthy

[1]Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany [2]Department of Mathematics, Technical University of Munich, Munich, Germany. [*]Correspondence to: Fabian J. Theis, Mohammad Lotfollahi < fabian.theis@helmholtz-muenchen.de, mohammad.lotfollahi@helmholtz-muenchen.de>.

The 2022 ICML Workshop on Computational Biology. Baltimore, Maryland, USA, 2022.

references, identification of disease-specific cell states not present in the reference and data granularity increases.

2. Continual Learner Conditional Variational Autoencoder (CLCVAE) to learn new variations in the query

One approach to allow a machine learning model to learn new variations emerging from the query data is to fine-tune the reference-initialized model on the query dataset. The fine-tuned model, however, suffers from forgetting the variations learned in the reference, a phenomena known as catastrophic forgetting (Kirkpatrick et al., 2017; Nguyen et al., 2017). Benchmarks on single cell data have therefore shown that fixing the projection function, that is freezing the expression weights in architecture surgery, prevents catastrophic forgetting, which also prohibits adaptation to new variations (Lotfollahi et al., 2022).

The goal of continual learning (CL) is to learn new tasks sequentially, without forgetting the knowledge of previously learned tasks (Kirkpatrick et al., 2017). Inspired by this fundamental concept in CL, an alternative less stringent approach to learn new variations, therefore, is to penalize changes in weight estimates of model parameters such that the new estimate for a parameter does not deviate from the old estimate, if the parameter is important for optimal reconstruction of reference, rather than freezing the weights. This is achieved by a Elastic Weight Consolidation regularizer (Kirkpatrick et al., 2017) and presenting the query model with examples from the reference during training.

2.1. Elastic Weight Consolidation

Let θ^* denote the parameters of the reference model. Let θ denote the current state of the parameters for the query model. Let F denote the diagonal Fisher information matrix of θ^*. The Fisher information matrix is the second derivative of the likelihood near an optimum. Therefore, it contains information about which parameters were important in the variational model trained on the reference. Recall that in the standard VAE approach the likelihood is approximated by the variational lower bound. Therefore, the Fisher information matrix here is computed as the gradient with respect to the variational lower bound.

In this work, we add a Elastic Weight (EW) penalty to CVAE models, say, scVI (Lopez et al., 2018) or scANVI (Xu et al., 2021), to minimize deviation of θ from θ^* according to the importance of parameter to the reference model stored in F. The CVAE models are foundational for single-cell reference mapping, where batch effects are widespread. We then consolidate this loss with the loss for the query model at each iteration over the training batch. The loss that is minimized, therefore, is:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{CVAE}}(\theta) \ + \frac{\lambda}{2} \sum_i F_i \ (\theta_i - \theta_i^*)^2,$$

where $\mathcal{L}_{\text{CVAE}}(\theta)$ is the loss from the variational model (i.e. base model), i labels each parameter and λ sets the contribution of EW penalty to the total loss. For example, in the case of the scVI model, $\mathcal{L}_{\text{CVAE}}(\theta)$ is the sum of reconstruction loss and KL-divergence loss.

Our approach, hence, can be described as follows: We initialize the query model with weights learned in the reference model. At each iteration over a training batch, we compute the loss of the model and the EW loss computed for the current state of the model, using the reference model and cells from the reference. We refer to this model as Continual Learner Conditional Variational Autoencoder, CLCVAE, herein.

3. Results

3.1. CLCVAE is as good as architecture surgery in batch effect removal

An important characteristic of a reliable model is that the latent representations should be free from unwanted variations and batch effects. We examined if batch effects are efficiently removed from the latent space of CLCVAE.

We used a standard integration benchmark Pancreas dataset described in Luecken et al. (2022) containing cells sequenced by nine sequencing technologies to assess data integration and batch removal (Fig1. a-c). Cells from two of the sequencing platforms, `celseq2` and `smartseq2` were left out as query datasets, and the reminder were used to construct a reference using scVI model from scvi-tools (Gayoso et al., 2021) using 2000 highly variable genes (HVGs). We trained the query CLCVAE model as described above. We compared the latent space of a surgery model (scVI + scarches) (Fig1.a), with a model fine-tuned on the query (Fig1.b) and the CLCVAE model (Fig1.c).

While the cells from different technologies mix well in the latent space of both surgery and CLCVAE models (Fig1.a,c), some technology-specific patterns were evident in the latent space of the fine-tuned model (Fig1.b). This re-enforces that fine-tuning the reference model on the query leads to forgetting (in this case, the variations due to batch in the reference), and that the EWC regularization in the CLCVAE model is essential to prevent forgetting. We further assessed batch integration and biological conservation by CLCVAE, surgery and fine-tuned models using scIB (Luecken et al., 2022) metrics (Fig1.d). The CLCVAE was marginally better in batch silhouette (sil_batch), cell type silhouette (sil_labels), cell type F1 scores and in preserving graph connectivity.

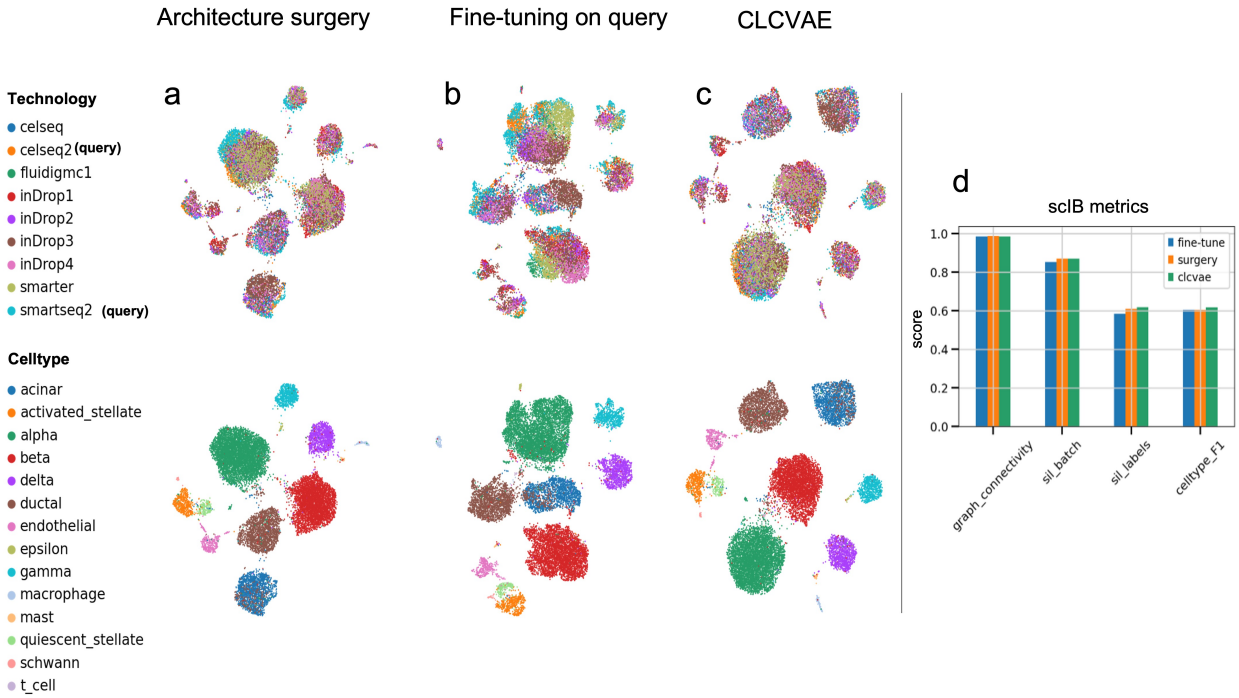

Figure 1. The UMAP representation of latent spaces of pancreas integration benchmark dataset inferred by **a** architecture surgery **b** fine-tuning the reference model on the query **c** CLCVAE model colored by cell type (lower) and sequencing technology (upper). **d** quantitative evaluation of batch correction and biological preservation by scIB integration metrics for each model.

3.2. Mapping cancer samples to a reference of immune cells

Next, we mapped a single-cell RNA-seq dataset of 4645 single cells isolated from 19 melanoma patients, profiling malignant, immune, stromal, and endothelial cells in metastatic melanoma tumors (Tirosh et al., 2016) to a reference of immune cells described in (Lotfollahi et al., 2022) using 4000 HVGs in the reference (Fig2). The query contains cell types that are present in the reference, namely T cells, NK cells, B cells; cell types that are not present in the reference, namely Cancer-associated Fibroblasts (CAFs), endothelial cells (Endo.), malignant cells, as well as Macrophages (Macro.) which are not present in the reference but have high transcriptional similarity to cell types present in the reference. We transferred annotations from reference to query cells using weighted predictions (the highest weight label category, where weight is determined by distance from 50 nearest neighbour cells in the reference). As in the HLCA study, we assigned cells with an uncertainty score greater than 0.3 the *unknown* label.

We observed that the uncertainty scores for CAFs, endothelial and malignant cells not present in the reference were high for a larger proportion of cells in CLCVAE compared to architecture surgery (Fig 2.a-b), indicating that the model has learned query-specific variations. Fig2.c-d is the heatmap of the proportion of cells in the query assigned to each cell type from the reference (rows of the heatmap) for every cell type in the query (columns of the heatmap). We observed less incorrect annotation transfers in the latent space of CLCVAE (Fig2.d) compared to surgery (Fig2.c). In particular, more CAF, endothelial and malignant cells, which are not present in the reference, were labeled as *unknown* by CLCVAE representation. Similar to surgery results, the T cells in Tirosh et al. (2016) were found to be a composite of CD4+ T cells and NKT cells. Although CLCVAE performed well in labeling a larger number of cells in the query not present in the reference as *unknown*, the model could not distinguish Macrophages (Macro.) in the query from CD14+ Monocytes in the reference, most likely due to high transcriptional similarity of these two populations.

3.3. Mapping healthy lung samples to Human Lung Cell Atlas with non-overlapping cell types

We further mapped a dataset of healthy lung cells by Madissoon et al. (2021) to the HCLA using the pre-trained reference scANVI model provided by the authors, performed cell type annotation transfer via weighted predictions as for the melanoma dataset, and compared the distribution of the uncertainty scores for representations learned by architec-

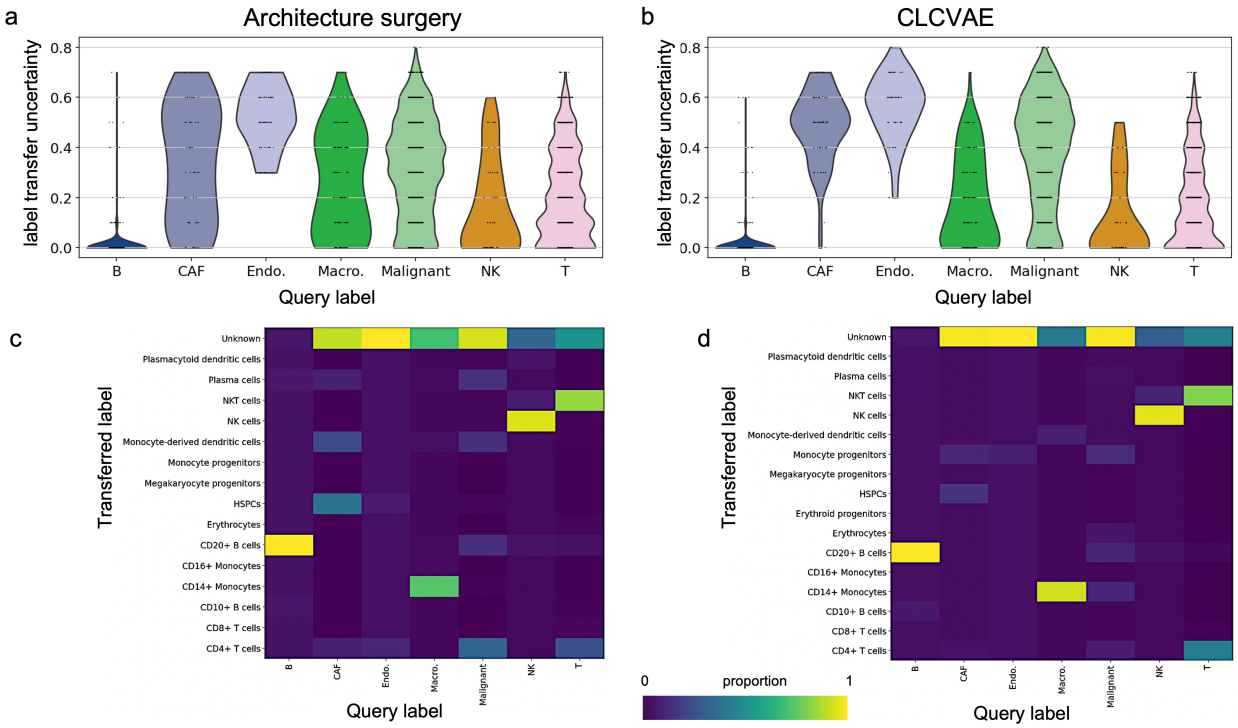

Figure 2. Annotation transfer uncertainty score for single cells from metastatic melanoma tumors mapped to a reference of immune cells for latent representations learned by **a** surgery and **b** CLCVAE. Heatmap of proportion of cells from each cell type in the query assigned to cell type categories in the reference or the "unknown" label by **c** surgery and **d** CLCVAE.

ture surgery and CLCVAE (Fig3). The dataset contains cell types that *are not* present in the reference. For the surgery results, we used the embedding of Madissoon et al. (2021) (Meyer_2021) published by HLCA authors. We transferred level-4 cell type annotations for this analysis. The findings are based on the 2000 HVGs in the reference.

The HLCA study found that the 11 cell types highlighted in green box were annotated incorrectly with high confidence in the query, either due to high transcriptional similarity with cell types in the reference or cells being part of a continuum. For 5/11 of these cell types, the tail distribution of uncertainty scores in Fig3 were fatter in the representation learned by CLCVAE compared to architecture surgery, suggesting that their biological variation is better learned by the proposed approach. These cell types are marked with orange stars. For the rest of the cell type populations shown in Fig3, which are cell types in the query that *are* present in the reference, the score distribution was mostly comparable to standard surgery, although we did observe a fatter tail distribution for DC1 and SMG-DUCT cell types, which could be indicative of query-specific cell type alterations.

4. Discussion and future directions

Single-cell reference atlases are diversifying. The Human Lung Cell Atlas (Sikkema et al., 2022), Human lung Cancer Atlas (Salcher et al., 2022) and cross-tissue disease atlas (Korsunsky et al., 2021) are just examples of reference atlases of growing complexity that enable biomarker discovery, study of biological aberrations in malignancies and drug-resistance mechanisms through mapping of newly acquired samples onto these references. Accurate and robust reference mapping algorithms would, hence, play crucial roles for reliable inference.

In this work we presented a preliminary idea to improve upon mapping disease, cancer and samples from perturbation experiments in general to single-cell reference atlases. We demonstrated that by allowing the model to remember variations in the reference while training on the query data, we could learn models that better distinguish cell types in the query not present in the reference. A few limitations remain to be addressed: 1- resolving cell types and states that are distinct, but have high transcriptional similarity to those in the reference. 2- Currently, raw samples from the reference are required for the training of the query model. Consequently, the computational train time is equivalent to de-novo integration. A potential solution is to present the

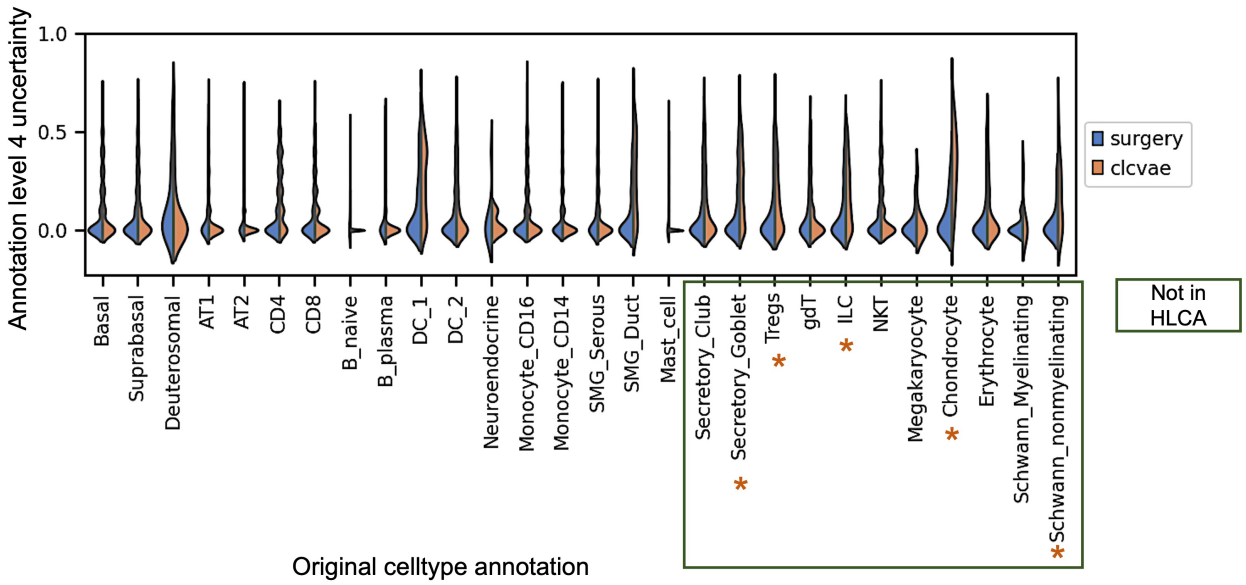

Figure 3. Distribution of annotation transfer uncertainty scores for healthy lung cells mapped to HLCA for overlapping and non-overlapping cell types.

model with latent representations of raw samples from the reference (encoded data) rather than raw samples (Borde, 2021). Evaluations on more datasets should be considered. 3- In case of extending a reference by mapping multiple query datasets sequentially, investigations are required to assess if the reference representation remains unchanged to avoid re-evaluation of integration upon the addition of each query dataset.

Software and Data

All datasets used in this work are public and have been referenced throughout the text.

The jupyter notebooks containing the code for training CLCVAE using modified scvi-tools training plans and results for the analysis presented in this work will be available on github https://github.com/theislab/icml_cbw_2022_clcvae

References

Borde, H. S. d. O. Latent space based memory replay for continual learning in artificial neural networks. *arXiv preprint arXiv:2111.13297*, 2021.

Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Wu, K., Jayasuriya, M., Melhman, E., Langevin, M., Liu, Y., Samaran, J., et al. Scvi-tools: A library for deep probabilistic analysis of single-cell omics data. *bioRxiv*, 2021.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Korsunsky, I., Wei, K., Pohin, M., Kim, E. Y., Barone, F., Kang, J. B., Friedrich, M., Turner, J., Nayar, S., Fisher, B. A., Raza, K., Marshall, J. L., Croft, A. P., Sholl, L. M., Vivero, M., Rosas, I. O., Bowman, S. J., Coles, M., Frei, A. P., Lassen, K., Filer, A., Powrie, F., Buckley, C. D., Brenner, M. B., and Raychaudhuri, S. Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. *bioRxiv*, 2021. doi: 10.1101/2021.01.11.426253. URL https://www.biorxiv.org/content/early/2021/02/18/2021.01.11.426253.

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., et al. Mapping single-cell data to reference atlases by transfer learning. *Nature Biotechnology*, 40(1):121–130, 2022.

Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Müller, M. F., Strobl, D. C., Zappia,

L., Dugas, M., Colomé-Tatché, M., et al. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods*, 19(1):41–50, 2022.

Madissoon, E., Oliver, A. J., Kleshchevnikov, V., Wilbrey-Clark, A., Polanski, K., Orsi, A. R., Mamanova, L., Bolt, L., Richoz, N., Elmentaite, R., et al. A spatial multiomics atlas of the human lung reveals a novel immune cell survival niche. *bioRxiv*, 2021.

Nguyen, C. V., Li, Y., Bui, T. D., and Turner, R. E. Variational continual learning. *arXiv preprint arXiv:1710.10628*, 2017.

Salcher, S., Sturm, G., Horvath, L., Untergasser, G., Fotakis, G., Panizzolo, E., Martowicz, A., Pall, G., Gamerith, G., Sykora, M., Augustin, F., Schmitz, K., Finotello, F., Rieder, D., Sopper, S., Wolf, D., Pircher, A., and Trajanoski, Z. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *bioRxiv*, 2022. doi: 10.1101/2022.05.09.491204. URL https://www.biorxiv.org/content/early/2022/05/10/2022.05.09.491204.

Sikkema, L., Strobl, D. C., Zappia, L., Madissoon, E., Markov, N. S., Zaragosi, L.-E., Ansari, M., Arguel, M.-J., Apperloo, L., Becavin, C., et al. An integrated cell atlas of the human lung in health and disease. *bioRxiv*, 2022.

Sohn, K., Lee, H., and Yan, X. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28, 2015.

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science*, 352(6282):189–196, 2016.

Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M. I., and Yosef, N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular systems biology*, 17(1):e9620, 2021.