Jointly learning T-cell receptor and transcriptomic information to decipher the immune response

Yang An [1*] **Felix Drost** [1 2*] **Fabian Theis** [1 3] **Benjamin Schubert** [1 3*] **Mohammad Lotfollahi** [1 2*]

Abstract

T cells play a pivotal role in the adaptive immune system recognizing foreign antigens through their T-cell receptor (TCR). Although the specificity and affinity of the TCR to its cognate antigen determines the functionality, the phenotypic differentiation and thereby also the fate of the T cell remain poorly understood. Therefore, studying the transcriptional changes of T cells in the context of their TCRs is key to deeper insights into T-cell biology. To this end, we developed a multi-view Variational Autoencoder (mvTCR) to jointly embed transcriptomic and TCR sequence information at a single-cell level to better capture the phenotypic behavior of T cells. We evaluated mvTCR on two datasets showing a clear separation of the cell state and their functionality, thus, providing a more biologically informative representation than models using each modality individually.

1. Introduction

T cells are an integral part of the adaptive immune system. They recognize antigen-derived epitopes that are bound to the Major Histocompatibility Complex of other cells and initiate an immune response to free the body from harmful microorganisms and cancer. This recognition is mainly driven by the T-cell receptor (TCR), a surface protein consisting of an α- and β-chain, which is expressed with a diversity of up to 10^{20} different sequences (Zarnitsyna et al., 2013). The complexity of cell state and functionality makes the study of T cells a challenging endeavor. Recent advances in single-cell technologies allows to jointly determine TCR sequences and transcriptome information (10x Genomics, 2019; Yost et al., 2019; Fischer et al., 2020), which pave the way to a deeper understanding of the phenotypic interplay between gene expression and TCR sequence.

Previous approaches for embedding T cells used TCR sequences for clustering or similarity assessment of groups with common epitope specificity (Glanville et al., 2017) (Dash et al., 2017). Due to greater datasets, new methods were proposed utilizing deep learning models such as DeepTCR, which embeds T cells based on their TCR sequence and VDJ gene usage with an Autoencoder (Sidhom et al., 2021). Recently, two novel methods were proposed that additionally include transcriptomic information by Bayesian Clustering (Zhang et al., 2021) or neighborgraph analysis (Schattgen et al., 2020). However, none of the existing methods have addressed learning a joint representation guided by both functional information from TCR sequences and transcriptional profiling of cells to analyze such paired data.

To address this, we propose mvTCR (https://github.com/SchubertLab/mvTCR) to learn a joint embedding of T cells by using both TCR sequence and the scRNA-seq data. Our model leverages advances is sequence learning using Transformers (Vaswani et al., 2017) combined with generative multi-view learning (Kingma & Welling, 2013; Lee & van der Schaar, 2021) providing a holistic view to analyze T cells. The model is evaluated on two datasets demonstrating the clustering of cells with homogeneous epitope specificity and transcriptional similarity.

2. Methods

The VAE model embeds the transcriptomic data X_{RNA} and the amino acid sequences of the TCR X_{TCR} by its encoding networks E_{RNA} and E_{TCR} to the features H_{RNA} and H_{TCR} of size d_{feat}, respectively. The extracted feature embeddings are subsequently fused by the mixture model M to obtain a joint latent distribution $q(Z_{joint}|X_{RNA}, X_{TCR})$. For downstream tasks, the Gaussian mean of this distri-

[*]Equal contribution [1]Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany [2]School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany [3]Department of Mathematics, Technical University of Munich, Munich, Germany. Correspondence to: Benjamin Schubert, Fabian Theis, Mohammad Lotfollahi <benjamin.schubert@helmholtz-muenchen.de, fabian.theis@helmholtz-muenchen.de, mohammad.lotfollahi@helmholtz-muenchen.de>.

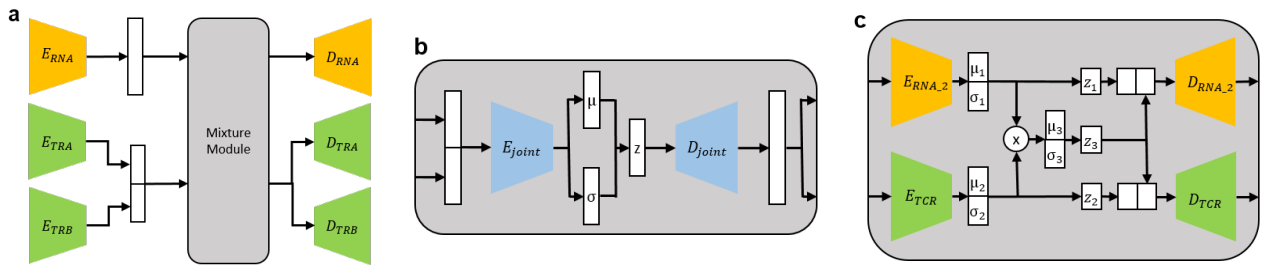


Figure 1. Overview of the mvTCR architecture. a) Composition of the modality-specific encoder and decoder networks. b) Concatenation version of the mixture module. c) PoE version of the mixture module.

bution is used to obtain the T cell embedding. We tested two mixture model schemes: concatenation and Product-of-Expert (PoE) (Lee & van der Schaar, 2021). M's output feeds into the modality-specific decoders D_{RNA} and D_{TCR} reconstructing the corresponding input data (Figure 1).

2.1. Encoder and Decoder Networks

TCR: The sequences of α- and β-chains in X_{TCR} are one-hot-encoded, end-padded to the maximum sequence length, and processed by individual encoder- and decoder-architectures. Since the two chains follow a similar structure but contribute differently, both networks share the same architecture but not their weights. Due to their great ability to encode sequences, the Transformer model is used for the encoders E_{TRA} and E_{TRB} (Vaswani et al., 2017), followed by a linear layer to reduce the output size to $d_{feat}/2$. The features from both chains are concatenated to form H_{TCR}. The decoders D_{TRA} and D_{TRB} adhere to a reverse structure. First, H_{joint} is upsampled by a linear layer, which is followed by the decoding Transformer. Finally, a softmax layer is used to predict the residue at each sequence position.

RNA: Following (Lotfollahi et al., 2019; 2020b), X_{RNA} is passed through the encoder E_{RNA} built from multiple blocks each consisting of a fully-connected layer, batch normalization, Leaky ReLU activation, and a dropout layer. The last linear layer transforms the output to the same dimensionality d_{feat} as the concatenated TCR features to avoid domination of one modality over the other. D_{RNA} follows the reverse architecture of E_{RNA}.

2.2. Mixture Module

The **Concatenation Module** concatenates H_{TCR} and H_{RNA} and passes the joint feature vector through a multi-layer perceptron E_{joint} to estimate the joint posterior $q(Z_{joint}|H_{RNA}, H_{TCR})$, from which the joint embedding Z_{joint} is sampled via the reparameterization trick (Kingma & Welling, 2013)(Figure 1b). Finally, Z_{joint} is passed into the shared decoder D_{joint} and its output is fed into the decoders of both modalities D_{RNA} and D_{TCR}.

The **PoE Module** uses separate encoders E_{RNA_2} and E_{TCR} for each modality to estimate individual marginal latent distribution $q(Z_{RNA}|H_{RNA})$ and $q(Z_{TCR}|H_{TCR})$ (Figure 1c). Here, Z_{joint} is sampled via the reparameterization trick from the closed-form solution of the product of both distributions

$$q(Z_{joint}|Z_{RNA}, Z_{TCR}) = p(Z) \prod_{i}^{N} q(Z_i|X_i), \quad (1)$$

where N represents the number of modalities and $p(Z)$ a zero-mean, univariant Gaussian prior. After sampling, Z_{joint} is passed to both encoders E_{TCR} and E_{RNA_2} in addition to the corresponding marginal distributions Z_{RNA} and Z_{TCR}. Thereby, the model is trained to preserve information specific to each modality individually, while simultaneously capturing features shared across them. All encoders and decoders of the mixture module were built from fully-connected blocks similar to the RNA module.

2.3. Training Procedure

The loss function consists of three weighted parts. The model is optimized towards reconstructing its input data by using the Mean Squared Error loss for X_{RNA} and the Cross-Entropy loss for X_{TCR}. Additionally, the latent distributions $q(Z|X_{RNA}, X_{TCR})$ or $q(Z_1|X_{RNA})$, $q(Z_2|X_{TCR})$, and $q(Z_3|X_{RNA}, X_{TCR})$ are regularized by the Kullback-Leibler divergence to resemble a zero-mean, univariant Normal distribution.

$$\mathcal{L}_{total} = \mathcal{L}_{scRNA} + \lambda_1 \mathcal{L}_{TCR} + \lambda_2 \mathcal{L}_{KDL} \quad (2)$$

The hyperparameters of the network were selected with Optuna (Akiba et al., 2019) to either optimize the reconstruction loss or a dataset-specific downstream task. For task-specific training, the loss weights were included in this search, while for reconstruction they were fixed such that the losses from both modalities have the same magnitude. To avoid overfitting, cells from rare clonotypes (i.e., cells with the same TCR sequence) were oversampled.

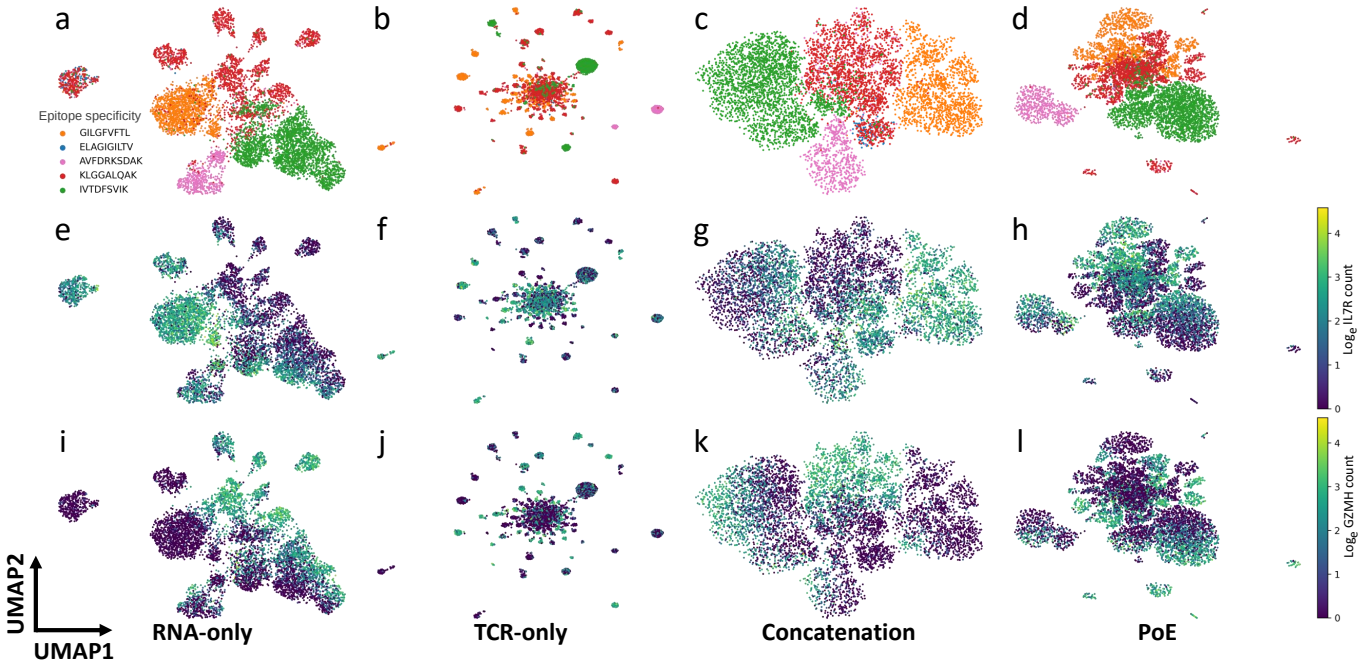


Figure 2. Uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) visualization of the T cell embedding from donor 1 of the 10x Genomics dataset. The columns represent the latent space of the different uni- and dual-modal models. (a-d) are colored by epitope specificity. (e-l) are colored by the normalized and log-transformed gene expression counts of IL7R and GZMH, respectively.

3. Results

We compared our models in various downstream tasks against unimodal baseline models trained on either RNA or TCR information. These baseline models followed the same overall architecture but were trained using only a single modality.

First, we analyzed the dataset provided by 10x Genomics consisting of paired single-cell TCR, transcriptome, and epitope specificity information of over 150,000 CD8$^+$ T cells from four donors (10x Genomics, 2019). To avoid batch effects, the models were trained for donor 1 and 2 separately. Due to limited diversity in epitope specificity, we excluded donor 3 and 4 in our analysis. For model training and downstream tasks, cells with unknown or rare epitope specificity were removed. Second, our models were evaluated on the dataset described in (Fischer et al., 2020), which shows T cell reactivity towards a mix of SARS-CoV-2 epitopes (Gene Expression Omnibus, GSE171037). The cells were split into training, validation, and test set of approximately 70%, 15%, and 15% for the 10x Genomics dataset and 80%, 10%, 10% for the smaller SARS-CoV-2 dataset, respectively. To avoid data leakage, cells from the same clonotype were assigned to the same subset.

3.1. Multimodal learning improved T cell clustering

To investigate whether cell states and functionality are better captured in the joint than in unimodal embeddings, we investigated the latent spaces of donor 1 of the 10X Genomics dataset (Figure 2, donor 2: Suppl. Figure 1, SARS-CoV-2: Suppl. Figure 2). The embedding of the model trained on TCR data was dominated by multiple smaller clusters representing similar or identical clonotypes. However, it failed to accumulate T cells with the same epitope specificity but expressing dissimilar sequences. The RNA-only model formed larger clusters of cells with the same specificity. However, clear separation of subclusters could be observed in T cells specific to the Cytomegalovirus epitope *KLGGALQAK* (Figure 2, red). The joint models embedded T cells with identical cognate epitopes into less fragmented and clearer separated clusters. These findings suggest that the joint models learned to combine complementary information from both modalities, which represent cellular functionality better than separate embedding.

This was also reflected in quantitative evaluations measured by the Average Silhouette Width (ASW) and the Normalized Mutual Information (NMI). ASW measures intra-cluster cohesion compared to inter-cluster separation, where higher values indicate clearer clustering. NMI is an external metric, which evaluates how much information is shared between the assigned cluster and ground-truth labels. In particular,

NMI is the harmonic mean between homogeneity and completeness, thereby measuring the pureness in clusters as well as the fragmentation of labels between clusters. During analysis, Leiden clustering (Traag et al., 2019) was used on the embeddings reporting the maximum metric resulting from three different clustering resolutions $(0.01, 0.1, 1.0$; Table 1). For the 10x Genomics dataset, epitope specificity was used as an external label to calculate the NMI, while for the SARS-CoV-2 dataset, scores were reported for cell type ($CD4^+$ or $CD8^+$ T cells) and reactivity toward a SARS-CoV-2 Spike epitope mix ($CD4^+$ reactive, $CD8^+$ reactive and unreactive T cells). For each dataset, at least one joint model outperformed the single modality models for each metric. This supports the qualitative findings that joint modeling improves clustering of T cells with higher intra-class coherence and inter-class separation, while representing specificity as well as functionality better.

Additionally, T cell embeddings should preserve biological signals such as cell state beyond epitope specificity and cell type. To test this, we performed differential expression analysis (excluding TRA/B genes) to determine the most characteristic marker gene of each Leiden cluster (donor 1: Figure 2 and Suppl. Figure 3, donor 2: Suppl. Figure 1, SARS-CoV-2: Suppl. Figure 2). Here, we showcase how two selected marker genes (IL7R, GZMH) are distributed in the embedding spaces (Figure 2). Note, that these observations also hold true for the top characteristic genes of the remaining clusters as well. The selected genes are expressed in different stages of T cell development. The IL7R (Figure 2 e-f) encodes the IL7 receptor, which interacts with IL17 during T cell development, in particular during V(D)J recombination to form the final TCR-sequence (Schlissel et al., 2000). High expression of this gene, therefore, indicates a T cell during proliferation. GZMH (Figure 2 i-l) gives rise to a cytotoxic protease that induces cell death to infected host cells (Fellows et al., 2007) and is therefore expressed in the effector development stages. Gene expression patterns were clearly preserved in both joint and the RNA-only models. While in the TCR-only embedding, marker genes were arbitrarily distributed in multiple clusters. The joint model seemed also to cluster T cells hierarchically first by epitope specificity and secondly on gene-level creating subpopulations of cells with similar cell states, thus providing a more informed representation when studying human T cell repertoires.

3.2. Joint latent space learning improved specificity prediction

Generally, specificity is mainly attributed to the TCR. However, transcriptome information contains TRA and TRB genes, which determine the set of possible sequences during TCR development. Additionally, previous works (Schattgen et al., 2020) revealed that T cells with a common ancestor express a similar transcriptional profile. Therefore, the antigen-specificity can be inferred from the gene expression data as well. We used this fact to evaluate whether a joint embedding of both modalities can improve TCR-specificity imputation over unimodal models, thereby substantiating the improved expression of the joint latent space. The ability to perform imputation of missing specificity in the embedding space was evaluated by using the cells from the training data as an atlas set and predicting the binding epitopes with a k-Nearest Neighbor (kNN) classifier for the test set. This is especially challenging since the data split prevents common clonotypes between atlas and query set. The joint models perform on-par with the RNA model on donor 1 of the 10x Genomics dataset, while they outperform both unimodal models for donor 2 (Table 2). Interestingly, TCR-specificity could be better inferred by gene expression than by TCR sequence in donor 1 and vice versa in donor 2. The joint model however performed on-par or better than the unimodal models in both donors, indicating that our model was able to represent relevant aspects of T cell functionality by capturing and combining complementary information from both modalities in an intelligent manner.

Table 1. Comparison of single- and joint-modality embedding by cluster evaluation based on ASW and NMI on different datasets.

		RNA	TCR	Concat	PoE
10x Donor 1	ASW	0.14	-0.18	0.30	**0.36**
Specificity	NMI	0.62	0.49	**0.73**	0.49
10x Donor 2	ASW	0.00	-0.10	0.27	**0.38**
Specificity	NMI	0.34	0.28	**0.39**	0.37
SARS-CoV-2	ASW	0.00	0.09	0.13	**0.19**
Cell type	NMI	0.12	0.30	**0.38**	0.37
Reactivity	NMI	0.02	0.18	0.19	**0.24**

Table 2. Comparison of single- and dual-modal embeddings of the 10x Genomics dataset by kNN specificity-prediction between an atlas and a query dataset. All values represent the class-weighted F1-score.

	RNA	TCR	Concat	PoE
10x Donor 1	**0.73**	0.67	**0.73**	0.69
10x Donor 2	0.75	0.86	0.87	**0.90**

4. Conclusion

In this work, we proposed mvTCR, a VAE to embed T cells based on their transcriptomics and TCR-sequence data jointly. Extensive experiments show mvTCR can synergistically combine this complementary information to enrich the latent representation of single-cell data and captures patterns

that would have been otherwise missed when interpreting each level independently. Our model achieves better T cell clustering based on functionality and cell state and reveals subpopulations of different cell states within the epitope specific clustering. We further showed that a joint embedding improves the imputation of unknown epitope specificity. We envision future works will make such representations transferable to new studies (Lotfollahi et al., 2020a) and also improve the interpretability of the latent space (Rybakov et al., 2020). Overall, mvTCR offers a more holistic view of T cell repertoires at single-cell resolution and therefore has the potential of leading to profound new insights into T-cell biology.

Code and data availability

The 10x Genomics datasets can be downloaded from https://support.10xgenomics.com/single-cell-vdj/datasets under the section *Application Note - A New Way of Exploring Immunity*. The SARS-CoV-2 dataset is available at GEO GSE171037.

The source code to recreate the results including training, evaluation, and tutorial scripts is accessible at https://github.com/SchubertLab/mvTCR. The reported models can be downloaded from 10.5281/zenodo.5006839.

References

10x Genomics. A new way of exploring immunity–linking highly multiplexed antigen recognition to immune repertoire and phenotype. *Tech. rep*, 2019. URL https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_AN047_IP_A_New_Way_of_Exploring_Immunity_Digital.pdf.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 2623–2631. Association for Computing Machinery, 2019. ISBN 9781450362016.

Dash, P., Fiore-Gartland, A., Hertz, T., Wang, G., Sharma, S., Souquette, A., Crawford, J. C., Clemens (Day), E., Nguyen, T., Kedzierska, K., La Gruta, N., Bradley, P., and Thomas, P. Quantifiable predictive features define epitope-specific t cell receptor repertoires. *Nature*, 547 (7661):89–93, 06 2017. ISSN 1476-4687.

Fellows, E., Gil-Parrado, S., Jenne, D. E., and Kurschus, F. C. Natural killer cell–derived human granzyme h induces an alternative, caspase-independent cell-death program. *Blood*, 110(2):544–552, 07 2007. ISSN 0006-4971.

Fischer, D. S., Ansari, M., Wagner, K. I., Jarosch, S., Huang, Y., Mayr, C. H., Strunz, M., Lang, N. J., D'Ippolito, E., Hammel, M., et al. Single-cell rna sequencing reveals in vivo signatures of sars-cov-2-reactive t cells through 'reverse phenotyping'. *medRxiv*, 2020.

Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L., Rubelt, F., Ji, X., Han, A., Krams, S., Pettus, C., Haas, N., Lindestam Arlehamn, C., Sette, A., Boyd, S., Scriba, T., Martinez, O., and Davis, M. Identifying specificity groups in the t cell receptor repertoire. *Nature*, 547(7661): 94–98, 06 2017. ISSN 1476-4687.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Lee, C. and van der Schaar, M. A variational information bottleneck approach to multi-omics data integration. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 1513–1521. PMLR, 13–15 Apr 2021.

Lotfollahi, M., Wolf, F. A., and Theis, F. J. scgen predicts single-cell perturbation responses. *Nature methods*, 16 (8):715—721, August 2019. ISSN 1548-7091. doi: 10.1038/s41592-019-0494-8.

Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Buttner, M., Avsec, Z., Misharin, A. V., and Theis, F. J. Query to reference single-cell integration with transfer learning. *bioRxiv*, 2020a.

Lotfollahi, M., Naghipourfar, M., Theis, F. J., and Wolf, F. A. Conditional out-of-distribution generation for unpaired data using transfer vae. *Bioinformatics*, 36 (Supplement_2):i610–i617, 2020b.

McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Rybakov, S., Lotfollahi, M., Theis, F. J., and Wolf, F. A. Learning interpretable latent autoencoder representations with annotations of feature sets. *bioRxiv*, 2020.

Schattgen, S., Guion, K., Crawford, J. C., Souquette, A., Barrio, A. M., Stubbington, M. J., Thomas, P. G., and Bradley, P. Linking t cell receptor sequence to transcriptional profiles with clonotype neighbor graph analysis (conga). *BioRxiv*, 2020.

Schlissel, M. S., Durum, S. D., and Muegge, K. The interleukin 7 receptor is required for t cell receptor γ locus accessibility to the v(d)j recombinase. *The Journal of experimental medicine*, 191(6):1045–1050, 03 2000.

Sidhom, J.-W., Larman, H., Pardoll, D., and Baras, A. Deeptcr is a deep learning framework for revealing sequence concepts within t-cell repertoires. *Nature Communications*, 12(1):1605, 03 2021. ISSN 2041-1723.

Traag, V. A., Waltman, L., and Van Eck, N. J. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, Mar 2019. ISSN 2045-2322.

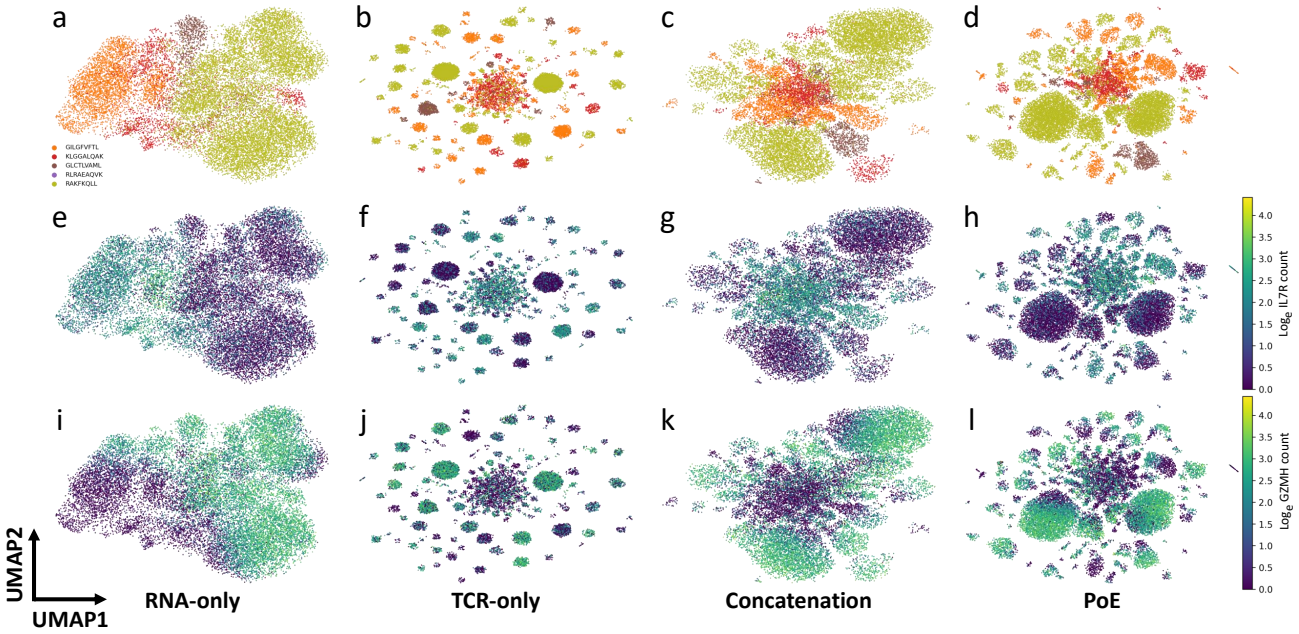
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Yost, K. E., Satpathy, A. T., Wells, D. K., Qi, Y., Wang, C., Kageyama, R., McNamara, K. L., Granja, J. M., Sarin, K. Y., Brown, R. A., et al. Clonal replacement of tumor-specific t cells following pd-1 blockade. *Nature medicine*, 25(8):1251–1259, Aug 2019. ISSN 1546-170X.

Zarnitsyna, V., Evavold, B., Schoettle, L., Blattman, J., and Antia, R. Estimating the diversity, completeness, and cross-reactivity of the t cell repertoire. *Frontiers in Immunology*, 4:485, 2013. ISSN 1664-3224.

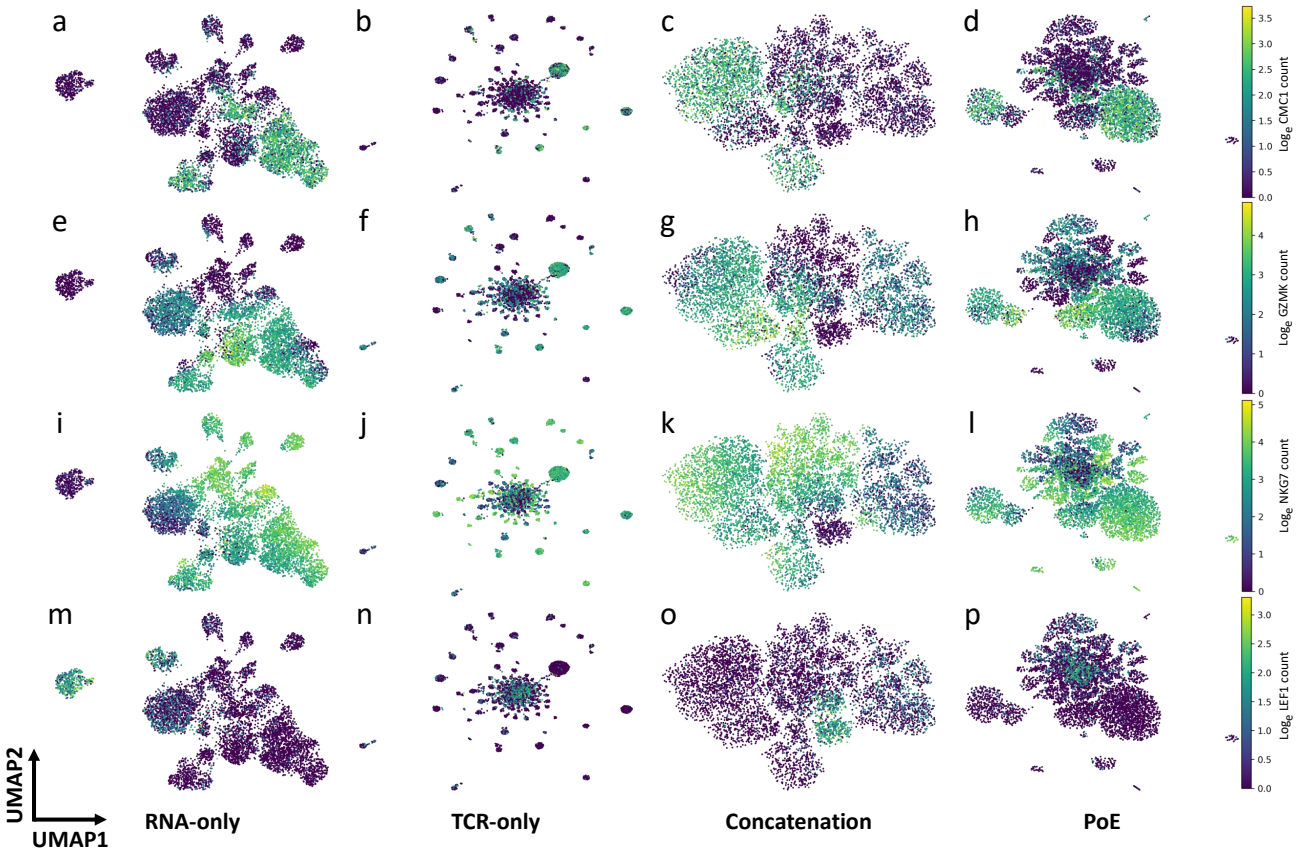
Zhang, Z., Xiong, D., Wang, X., Liu, H., and Wang, T. Mapping the functional landscape of t cell receptor repertoires by single-t cell transcriptomics. *Nature Methods*, 18(1): 92–99, 01 2021. ISSN 1548-7105.

Supplementary Material



Supplementary Figure 1. UMAP visualization of the T cell embedding from donor 2 of the 10x Genomics dataset. The columns represent the latent space of the different uni- and dual-modal models. (a-d) are colored by epitope specificity. (e-l) are colored by the normalized and log-transformed gene expression counts of IL7R and GZMH, respectively.

Supplementary Figure 2. UMAP visualization of the T cell embedding of the SARS-CoV-2 dataset. The columns represent the latent space of the different uni- and dual-modal models. The first two rows are colored by T cell type (a-d) and reactivity (e-h) to the SARS-CoV-2 antigen mix, respectively. The remaining visualizations are colored by the normalized and log-transformed gene expression counts of PABPC1 (i-l), CCL5 (m-p), and CCL4 (q-t).



Supplementary Figure 3. Visualization of additional marker genes from donor 1 of the 10x Genomics dataset. The columns represent the latent space of the different uni- and dual-modal models. The visualizations are colored by the normalized and log-transformed gene expression counts of CMC1 (a-d), GZMK (e-h), NKG7 (i-l), and LEF1 (m-p).