Modeling CAR Response at the Single-Cell Level Using Conditional Optimal Transport

Alice Driessen
IBM Research Zürich
Rüschlikon, Switzerland
D-BSSE, ETH Zürich
Basel, Switzerland
adr@zurich.ibm.com

Jannis Born
IBM Research Zürich
Rüschlikon, Switzerland

Rocio Castellanos Rueda
D-BSSE, ETH Zürich
Basel, Switzerland

Sai T. Reddy
D-BSSE, ETH Zürich
Basel, Switzerland

Marianna Rapsomaniki
Biomedical Data Science Center
Lausanne University Hospital
University of Lausanne
Lausanne, Switzerland
marianna.rapsomaniki@unil.ch

Abstract

Chimeric Antigen Receptor (CAR) T cell therapy is a promising area of cancer immunotherapy. However, many challenges such as loss of persistence, T cell exhaustion, and therapy associated toxicities hamper further advancement of CAR T cell therapy. Therefore, recent efforts have focused on designing improved CARs that show better therapeutic characteristics. However, it is unfeasible to test all CAR variants in lab-based assays as CARs consist of multiple intracellular signalling domains. This results in over 100'000 possible variants. We leverage computational modeling to navigate this vast combinatorial space by learning the relationship between CAR design and T cell functionality, thereby proposing promising CAR T cell designs. CAR T cells expressing different variants can be viewed as cells that underwent different perturbations. Neural Optimal Transport is an upcoming field that can model single cell perturbations and predict unseen cells and conditions. In this work we leverage the conditional Monge Gap to model the response to CAR expression at a single-cell level and generate gene expression of cells that express an unseen CAR design. We show that CAR OT (CAROT) significantly outperforms the baseline for gene expression prediction for in-distribution CAR variants, with distinct gene expression patterns per CAR that capture biological characteristics. When predicting unseen CAR variants, we demonstrate promising results in terms of gene expression prediction and show the model learns gene expression patterns linked to domains in the training set. This work demonstrates that optimal transport may support discovery and development of new CAR T cell designs.

1 Introduction

Chimeric Antigen Receptor (CAR) T cell therapy is a promising area of cancer immunotherapy, with currently six FDA-approved therapies and over a thousand ongoing clinical trials [1]. CARs are synthetic protein receptors consisting of an extracelluar domain linked to an intracellular signalling domains, which carry the necessary T cell activation signals. *Ex vivo* engineered CAR T cells that

AI for New Drug Modalities at NeurIPS 2024.

recognise tumor cells are introduced into the patient, resulting in a living drug capable of a sensitive, target-specific, self-replicating, and long-term response [2, 3]. Despite success, further advancement of CAR T cell therapy is limited by many challenges including loss of persistence, T cell exhaustion and associated toxicities [4]. Additionally, patient responses vary in remission rates and associated toxicities such as cytokine release syndrome [5, 6]. To overcome these challenges, recent efforts have focused on designing novel CARs that show promising therapeutic characteristics, such as memory and cytotoxicity features and limited exhaustion [7]. For example, varying the signalling domains in the CAR was shown to lead to differences in T cell survival and cytotoxicity both *in vitro* and in patients [8, 9]. Still, the understanding of the exact properties of co-stimulatory domains that determine therapeutic efficacy is limited [10]. Despite efforts to establish high-throughput screening, current *in vitro* methods can profile 100 to 200 variants in primary (patient-derived) human T cells [11–14]. Although larger screens have been established, these rely on cell lines that do not completely recapitulate T cell biology [15]. However, currently up to three co-stimulatory domains can be incorporated in a CAR design, given the number of possible signaling domains involved in immunological function, this results in a CAR design space of over 100'000 possible combinations.

Computational modeling can help navigate this vast combinatorial space, mapping the relationship between CAR design and functionality and thus aiding lab-based screening by suggesting promising CAR T cell designs. For example, Daniels *et al.* recently trained a neural network model on a library of around 250 CAR variants to predict stemness and cytotoxicity from the CAR design for over 2000 unmeasured variants [13]. Although the authors were able to point to CAR design rules suggesting efficacy, they relied on summary statistics of the outcome that potentially mask heterogeneity of patient response. Indeed, during a CAR T cell therapy patients are infused with millions of cells, which can be best described by distributions rather than summary statistics [16, 17]. For example, one patient showed at the peak of the response 94% CAR T cells from a single clone, highlighting the importance of taking cellular heterogeneity into account [18]. Additionally, the authors used a one-hot encoding strategy to represent each CAR design as a combination of the thirteen signalling motifs assessed in their study, an approach that cannot generalise to other CAR T cell libraries.

Recently, the growing availability of single-cell RNA sequencing (scRNA-seq) data before and after perturbations has led to the emergence of machine learning models that predict gene expression distribution of single cells in response to drug or genetic perturbations, with early models based on different flavors of autoencoders [19–21]. As scRNA-seq is a cell destructive assay, an inherent difficulty in learning perturbation responses from the resulting data is that the same cell cannot be measured before and after a perturbation. As a result, control and perturbed cell populations are unpaired and direct cell-to-cell comparisons are impossible. Optimal transport (OT) is a field of mathematics concerned with moving mass between probability distributions in a cost-minimizing manner. It poses a natural framework to match the distributions before and after perturbations while accounting for the heterogeneity of the responses [22]. With neural OT, parametric models learn a global OT map through amortized optimization of before/after distributions which allows one to make predictions for unseen initial (cell) distributions. Neural OT has already been successfully employed in matching single-cell distributions over time [23–25], in space [26, 25], or across modalities [27–29]. OT-based methods optimize the transportation map from a distribution of control cells to a distribution of perturbed cells. The resulting map can be used to infer a perturbation response, that is, to predict the change in cell state of the control cells if they had been exposed to this perturbation. The transportation can be learned for the (reduced) gene expression space or other modalities, such as cell surface markers [30]. Beyond the predictions of unseen cells [22], recent frameworks such as CondOT [31] and the conditional Monge Gap [32] allow to make predictions for unseen cells from unseen perturbations, thus extrapolating to new therapeutic options.

Inspired by the above, in this work we treat modeling of CAR T cell response as a perturbation prediction problem, and explicitly model response to a CAR T cell treatment at a single-cell level. We use OT as a generative model, which can generate the gene expression of a cell that expresses a CAR design that is not yet experimentally tested. We introduce a conditional CAR OT (CAROT) model, predicting the single cell distribution of thirty CAR variants. By leveraging recent advances in Protein Language Models (PLMs) that have successfully been used as embeddings for various downstream tasks [33], conditional CAROT (conCAROT) allows us to generalise to unseen CAR designs. We used the model to predict different CAR variants, obtaining distinct and accurate gene expression profiles that capture biological characteristics for in distribution CARs. Additionally, we made predictions for gene expression of unseen CAR T cells, showing that the conCAROT learned

gene expression patterns linked to specific domains. Overall, we show that OT may aid discovery and development of new CAR T cell designs.

2 Methods

2.1 OT fundamentals and Conditional Monge Gap

OT is a mathematical framework that finds the optimal way to transport a source distribution to a target distribution while minimizing the cost of displacement. For a comprehensive overview of OT formalisms and biological applications, we refer to Bunne *et al.* [34]. The Monge formulation of OT [35] finds a push-forward map T that maps the source distribution to the target distribution while allowing splitting of the mass, and is formulated as

$$T \text{ s.t.} T\sharp\mu = \nu, \qquad C(T) = \int_{\mathbf{R}^d} c(x, T(x))\mu(dx) \tag{1}$$

where μ and ν are measures on the source and target distribution, respectively, x is a point in the source distribution and $T(x)$ the transportation of x. $T\sharp\mu$ describes the result of transporting every point in μ and $C(T)$ the cost associated with the transport. In single-cell applications, the Monge formulation translates to finding a map that transforms the gene expression of the control cells (source) to the gene expression of the perturbed cells (target). The mass splitting allows gene expression from a single source cell to be transported to multiple target cells, and a single target cell to receive gene expression from multiple source cells. In the case of neural OT, the learned transport map can then be applied to unseen cells to predict the target distribution at inference time. Such approaches often rely on Brenier's theorem and use input convex neural networks (ICNNs) to parameterize and learn the OT maps [22, 36–39]. In practice, training ICNNs is challenging and can be circumvented e.g., with the Monge Gap [40], a regularizer that can be employed as a loss function in any neural network and estimates the deviation from a proposed transport map to an optimal map. The Monge Gap using n samples of a distribution is:

$$\mathcal{M}_\epsilon(T_\theta) = \frac{1}{n}\sum_{i=1}^{n} c(x_i, T_\theta(x_i)) - W_\epsilon(\mu, T_\theta\sharp\mu), \qquad W_\epsilon(\mu,\nu) := \min_{P\in U_n} \langle P, C\rangle + \epsilon H(P) \tag{2}$$

where c is again a cost function, T_θ a transport map, and μ and ν measures on the distribution (as in Equation 1), W_ϵ is the entropically regularised Kantorovich relaxation, with $H(P)$ as entropy, ϵ the strength of regularisation, $\langle P, C\rangle$ the transportation cost using map P and cost matrix C, and U_n the set off all possible transport maps between the measures μ and ν. Here, we leverage the *conditional* Monge Gap (CMonge, [32]), a recently proposed extension of the Monge Gap that globally parametrizes Equation 2 by optimizing over K conditions simultaneously [32]:

$$\min_\theta \sum_{i=1}^{K} \Delta_\epsilon(T_\theta(p_i)\sharp\mu, \nu_i) + \lambda\mathcal{M}_\epsilon(T_\theta(p_i)). \tag{3}$$

where $\mathcal{M}_\epsilon$ is the Monge Gap defined in Equation 2, Δ_ϵ is the Sinkhorn divergence between source and target samples [41], p_i describes the perturbation or condition, and λ is the strength of the Monge Gap regularization. This setting allows us to learn transport maps *conditioned* on a covariate and extend to unseen perturbations.

2.2 Conditional perturbation modeling of CAR T cell therapy

Chimeric Antigen Receptor library We used a Chimeric Antigen Receptor (CAR) T cell library of 31 constructs, built from six intracellular signalling domains (Figure 1A) and cells of two donors from GEO database number GSE262686 [42]. The library consists of 30 CAR variant designs that contain either one (second generation) or two intracellular co-signalling domains (third generation) in addition to a CD3ζ domain. The control cells express a non-signalling CAR (NS-CAR) that only contains the antigen recognising extracellular domain specific for a cancer antigen, a trans-membrane domain, and no intracellular signalling domains domains. The cells underwent a repeated antigen stimulation study (RAS) to mimic *in vivo* conditions by continued T cell and tumor cell engagement and were subsequently sampled for scRNA-seq at day 0, 6, and 12 (Figure 1B). During the RAS, CAR T cells were co-cultured with a cancer cell line expressing an antigen that is recognised by

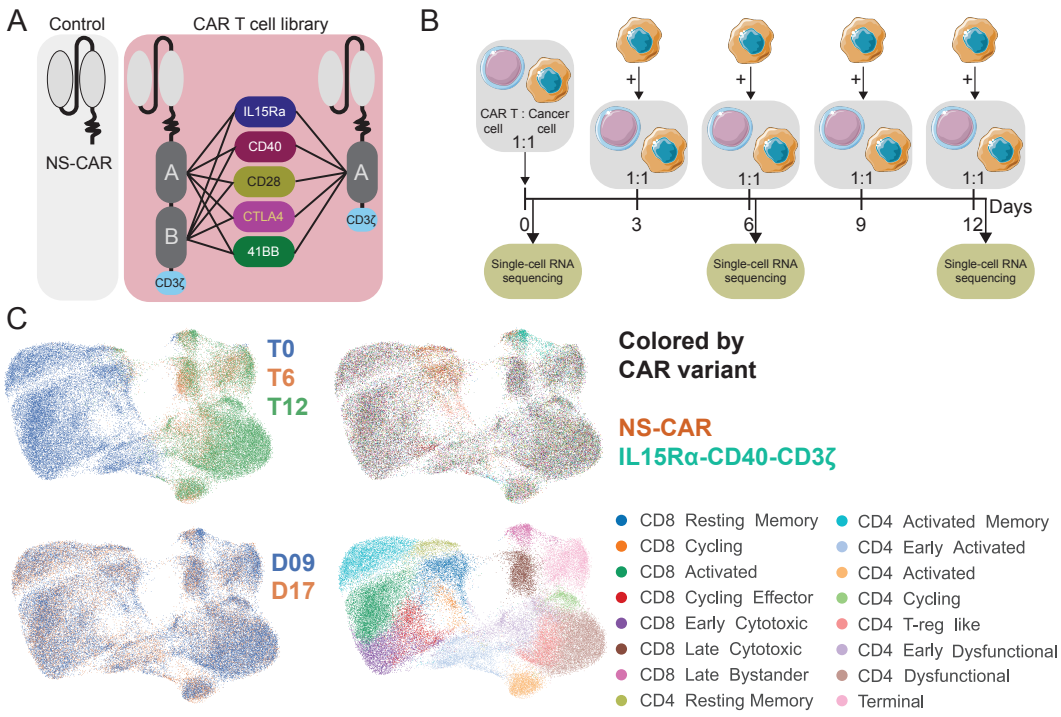

Figure 1: Dataset overview. A) CAR T cell library NS-CAR: Non-signalling CAR. B) Repeated antigen stimulation experiment setup. Cells for sequencing are sampled 6h after co-culture with the (newly added) cancer cells. C) UMAPs showing different covariates in the data based on the HARMONY representation of all genes.

all CAR variants for multiple days, to simulate exposure to cancer cells. The data were corrected for batch effects and annotated for CD4/CD8 cell states, cell cycle phase and functional scores. Functional scores were computed for cytotoxicity, memory, proinflammatory signature, T-helper 1 and T-helper 2 signatures using established marker genes [42] (subsection A.1, Table A1). This dataset is characterized by different sources of variation, including time, donor, cell type, and effect of CAR variant (Figure 1C). As expected, cells harvested at day 0 and days 6 and 12 formed distinct clusters, indicating large shifts in gene expression between short- and long-term activation. Another prominent source of variation is the cell state (phenotype): we observed seven CD8 cell states and eight CD4 cell states, including memory, activated, cytotoxic, bystander and dysfunctional states. The batch effect correction successfully eliminated confounding from donor variation. CAR variants were mixed in the UMAP embedding and only two CAR variants formed distinct clusters. The NS-CAR (brown) mostly occupied the resting memory compartments, showing a clear difference from cells that express CAR variants with signalling domains. The IL15Rα-CD40-CD3ζ variant (turquoise) is prominent in the CD8 late bystander cluster.

CAR T cell OT We build on the CMonge framework of (Harsanyi et al. [32], MIT license) and extend it to model response to CAR T cell therapy. We defined the NS-CAR cells as our source distribution and mapped it to target distributions of CAR-expressing cells (Figure 2A), transporting the non-batch corrected log-counts of 82 genes (see subsection A.2). We trained and applied CAROT on the CD4 or CD8 subsets separately, since transition from CD4 to CD8 states is rare and not the focus of this study. In contrast to previous works that used autoencoders to compress the scRNA-seq data [31, 22, 32, 20, 43], we used a curated functional geneset that contains 82 genes which capture T cell characteristics (e.g., memory, cytotoxicity, exhaustion and proliferation – Table A2), allowing for more interpretability (subsection A.2). We also investigated two baseline settings (Figure 2B): (i) an *identity* or *do-nothing* setting as the lower bound, as it takes cells from the control population without transportation as prediction for the CAR-expressing cells, and (ii) a *within-condition* setting that serves as an upper bound, as it takes CAR-expressing cells from the target CAR variant as prediction. Thus, ideally CAROT performs better than the identity model and close to the within condition model.

We built on the CMonge implementation [32], which has condition-specific dataloaders for efficient GPU-use, meaning that each batch comes from a single condition/CAR variant. We mixed different conditions by using layer normalisation after concatenation of the embedding and gene-expression and take a gradient step after accumulation over four batches. For in distribution settings a train/test split of 0.8/0.2 was used. For out-of-distribution (OOD) experiments we used the same split for training, but, when evaluating the OOD variants, all cells were in the test set (0/1 split). We used the CMonge hyperparameters unless stated differently. We performed single-GPU training/evaluation with a runtime limit of 1h for the unconditional CAROT and 12h for the conCAROT models.

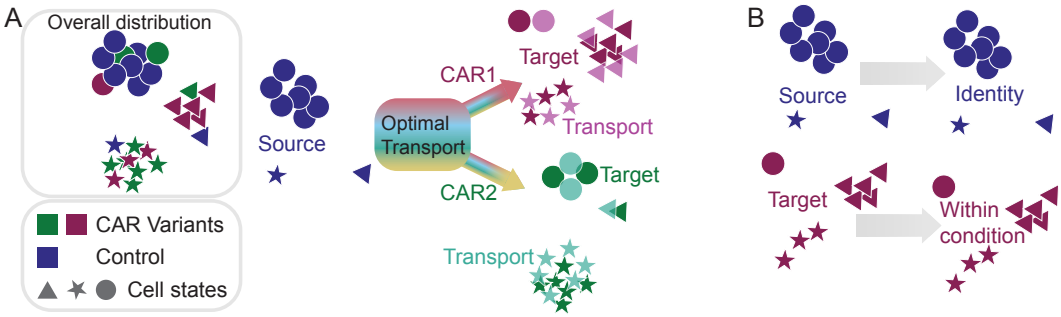

Figure 2: Optimal transport settings. A) The overall distribution shows control and CAR-expressing cells, clustered by cell state. Different variants have a different number of cells in the cell states. Optimal transport model CAROT maps gene expression of control cells (source) to gene expession of CAR predicting cells (target), trying to minimise the transportation cost and the error between the transport and the target expression. B) Two baseline settings. Identiy or do-nothing model uses gene expression of control cells and within condition takes gene expression of car expressing cells as prediction/transport.

CAR embedding To allow conCAROT to condition on a CAR variant and make CAR-specific predictions, we need to construct a generalizable embedding of the different CAR variants. Protein Language Models (PLMs) such as Evolutionary Scale Modeling (ESM), map the amino acid sequence to a numerical embedding and have been pretrained on different proteins and tasks, which allows them to capture biologically relevant information [44]. We embedded the intracellular amino acid sequence of the variants using ESM2 t48_15B_UR50D, referred to as ESM XL [44]. These embeddings were averaged over the sequence, resulting in the length of the embedding dimension of 5120 for ESM XL (subsection A.3).

3 Results

3.1 CAROT projects CAR effect on single cell distribution

We started with the simplest, non-conditional setting of training one OT model per CAR variant, referred to as CAROT (Figure 3A). Models were evaluated using the average R^2 of the mean gene expression per gene over all cells as in [32, 22] (subsection A.5, Figure 3B) and the Maximum Mean Discrepancy (MMD), a distance metric between distributions that compares the gene expression on a finer scale than the mean expression (as in [22], subsection A.5, Figure 3C). For both metrics, OT significantly outperforms the identity model. In terms of R^2 CAROT performs similar to the within condition setting. To evaluate the quality of the prediction, we visualized the source, target and transport cells onto a UMAP projection generated from the functional geneset's non-batch corrected log-counts of all CD4 or CD8 cells (Figure A3). We show as an example two CAR variants, where we observe good mixing between target and transport cells, whereas the source distribution is distinct for the CAR-expressing cells and prediction (Figure 3D, see Figure A4&Figure A5 for all CARs). We also observed distinct transport patterns for different CAR variants, indicating that different transport plans are learned per CAR variant. CD40-CD40-CD3ζ has most cells in the cycling effector and early cytotoxic cell states, which is captured by CAROT. IL15Rα-CD40-CD3ζ has most cells in the late bystander and late cytotoxic cell states, where also most of the transport is located. Taken together, these results establish that CAROT can be used to model CAR-variant effects.

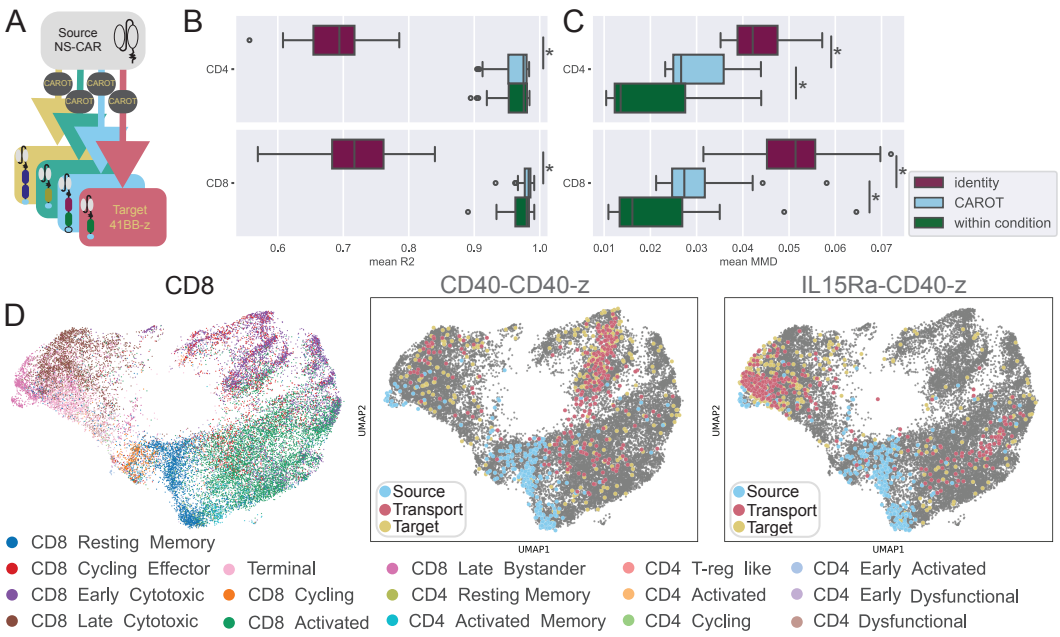

Figure 3: Unconditional optimal transport results. A) CAROT model setting. B) Mean R^2 over 9 samples calculated using all 82 genes in the functional geneset for the CAROT model and the two baselines. Asterisks indicate statistically significant differences (subsection A.6). C) Same as B but showing mean Maximum Mean Discrepancy (MMD). D) UMAP constructed using the non-batch corrected logcounts of the 82 genes in the functional geneset for CD8 cells, colored by cell state (left), and with control cells (source) , CAR-expressing cells (target) and prediction (transport) highlighted for two CARs (middle and right).

3.2 Conditional CAROT maps multiple CAR variants with one model

With the objective of predicting unseen CAR variants in mind, we trained a single OT model conditioned on a CAR variant embedding so that it can predict the effect of all CAR variants, we refer to this model as conditional CAROT (conCAROT) (Figure 4A). We observed that, for many CAR variants, the single cell distribution is quite similar, with few distinctions between CAR variants (Figure A4&Figure A5), which makes it hard for the model to learn to condition on the CAR variant. Also, for many CAR variants, there are only a small number of cells, complicating estimation of the single cell distribution (Figure A1B&C). Therefore, to train conCAROT, we selected only CAR variants for which we have more than 750 cells for CD4 and CD8 separately (Figure A1B&C).

We compared again the MMD and R^2 with the identity and within condition models for both CD4 and CD8 (Figure 4B&C), achieving significant improvement over the baseline in both subsets and metrics. R^2 values were greater than 0.95 for each CAR variant (Figure 4D) and comparable to the condition-specific CAROT models (Figure A10F), indicating no loss in performance compared to CAROT. As before, we see distinct transport patterns for different CAR variants, especially for the CD8 subset, where CD40-CD40-CD3ζ cells are found in the early cytotoxic and cycling effector cell states, and the IL15Rα-CD40-CD3ζ cells found in the late bystander and late cytotoxic cluster (Figure 4E). The CD4 subset CAR-expressing cells differ less, although the predictions look well mixed with the target (Figure A6, CD8 Figure A7).

We further evaluated whether our predictions capture biological characteristics of CAR T cell therapy. We compared the distribution of functional scores, namely cytotoxicity, memory and proinflammatory scores, between the source, target and transport (Figure 5A). We observed similar medians and spread in the score distributions of cytotoxicity and memory of target and transport, unlike the ones between target and source. Furthermore, we compared the percentage of positive cells for each score and each CAR variant and marked that, for most variants, this showed good concordance between target and transport (Figure 5B), even for CAR variants that were not present in the training data (Figure 5B purple dots). However, the percentage of memory signature cells is often underestimated in the target for both in-distribution (ID) and OOD CAR variants, although the ranking of CARs based on the memory signature is maintained. This could indicate a poorer transport of memory genes or

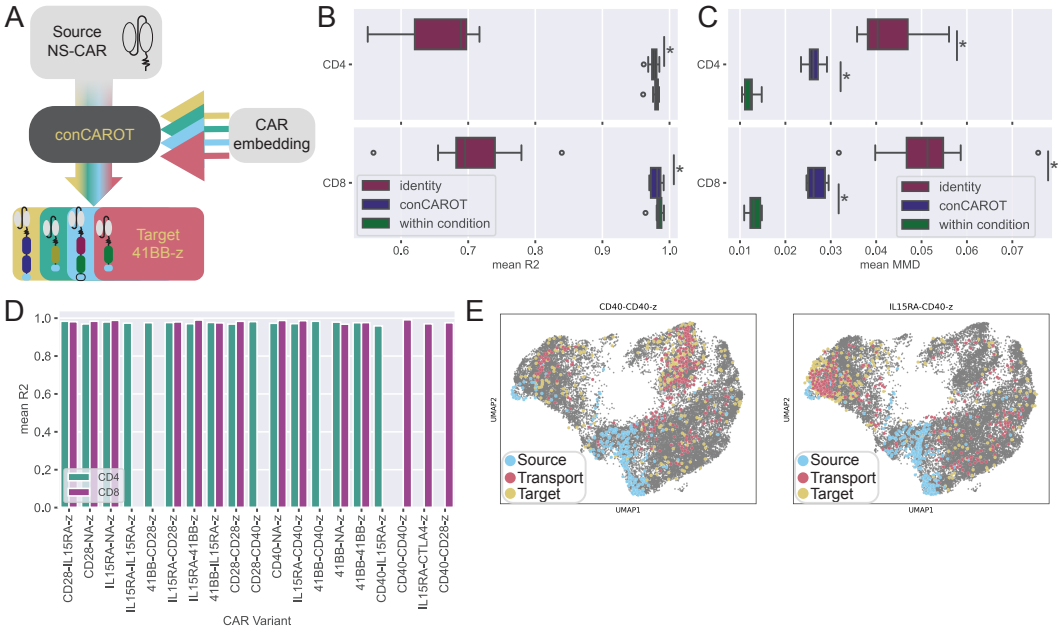

Figure 4: Conditional optimal transport results. A) Conditional CAROT model setting. B) Mean R^2 over 9 samples calculated for the conCAROT model using the ESM XL embedding and the two baselines, and using only CARs with more than 750 cells. Asterisks indicate statistically significant differences (subsection A.6) C) Same as in B, but showing mean MMD. D) Mean R^2 over 9 samples for CARs expressed in over 750 cells, split by CD4/CD8 subtype. Overall distribution is shown in panel B. E) Same UMAP as in Figure 3D (left), highlighted are the control cells (source), CAR -expressing cells (target) and prediction (transport) for two CARs.

more difficult relation between CAR variant and memory signature, as CAR effect on the memory phenotype is less pronounced especially in the CD8 subset [42]. Moreover, we investigated the cell state distribution for different CAR variants and observe an overall similar pattern (Figure 5C). conCAROT accurately captures the CD8 Early Cytotoxic cell state in the CD40-CD40-CD3ζ variant, the CD8 Late Bystander in the IL15Rα-CD40-CD3ζ and the CD8 Resting Memory cell state in the IL15Rα-CD40-CD3ζ variant. However, certain discrepancies between target and transport were also observed. For example, conCAROT tends to predict more cells in the CD4 Early Dysfunctional cell states than present in the target. Also, the CD4 Activated Memory and CD4 Early Activated populations in CTLA4 combined with IL15Rα and CTLA4, respectively, are not well predicted by the model. This is possibly because we had to exclude most CAR variants with CTLA4 in position A due to low cell numbers. Overall, we conclude that conCAROT performs on par with CAROT and it manages to capture biological patterns, with the important benefit of a single model that makes predictions for multiple CAR variants.

3.3 Out-of-distribution prediction of CAR effects

Our goal is to have model that can make accurate predictions of unseen CAR variants, such that it enables *in silico* screening to select promising CAR variants for experimental testing (Figure 6A). Therefore, we further investigated conCAROT in the OOD setting (condCAROT-OOD). When we trained only on CARs with >750 cells, we automatically left out other variants with <750 cells. We observed that the OOD performance on the left out CARs still outperforms the identity setting significantly for the R^2 but less explicit for the MMD (Figure 6B&C). Interestingly, the within condition and identity baselines also achieve a worse MMD on these held out CAR variants, suggesting that those variants are more challenging to predict, which is corroborated by further analysis (Figure A10A&B). Additionally, we observed that conCAROT-OOD transports many cells to the late cytotoxic or late bystander cell state when CD40 is present and many cells to the cycling or resting memory cell state for CTLA4 (Figure 6D). In the training data, the variants IL15Rα-CD40-

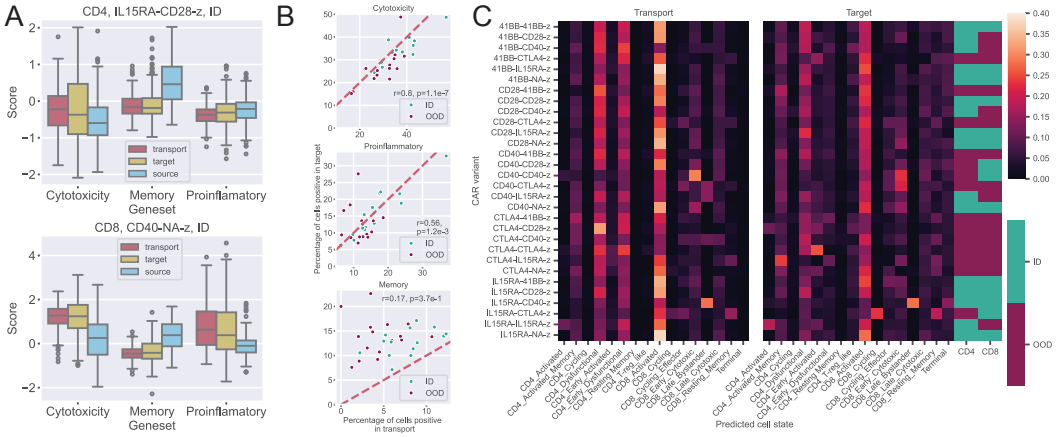

Figure 5: Biological features of transported gene expression. A) Geneset scores for two in-distribution (ID) CAR variants, colored by source, target and transport. B) Percentage of positive CD8 cells for each geneset in the target and transport, colored by ID and OOD variants. C) Heatmaps of predicted cell type fractions for transport and target, annotated with a label that indicates whether the CAR was ID or OOD for each subset.

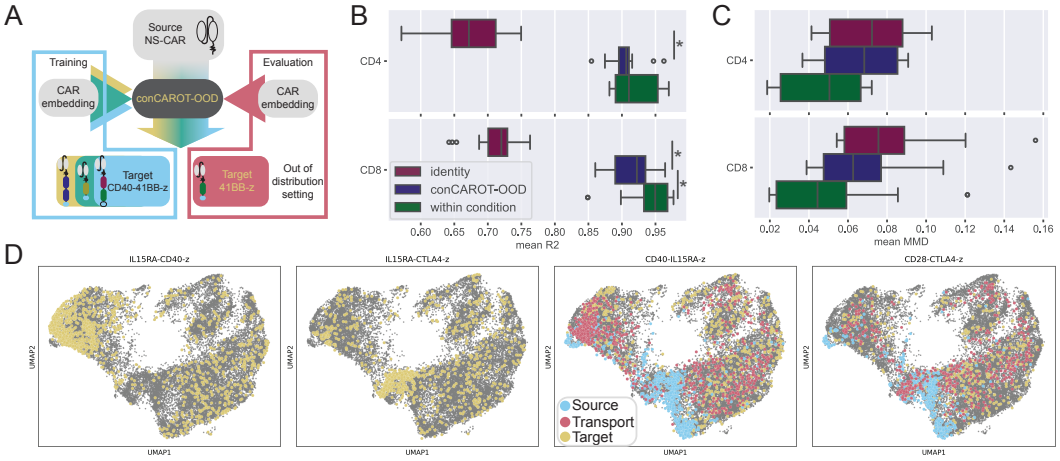

Figure 6: Out of distribution conditional optimal transport results. A) conCAROT OOD model setting. B) Mean R^2 over 9 samples calculated for conCAROT and the two baselines using the ESM XL embedding for CARs not seen during training (CARs with <750 cells). Asterisks indicate statistically significant differences (subsection A.6). C) Same as in B, but showing the mean MMD. D) Same UMAPs as in Figure 3D, with all CD8 target cells for IL15Rα-CD40-CD3ζ and IL15Rα-CTLA4-CD3ζ shown in yellow in the first two subplots. These CAR variants were in the training data and show the CD40 and CTLA4 patterns that were learned. The third and fourth UMAPs highlight control cells (source), CAR-expressing cells (target) and prediction (transport) for two CARs, showing the OOD predictions for variants with CD40 or CTLA4.

CD3ζ and IL15Rα-CTLA4-CD3ζ clearly show clustering of cells in these regions as well, showing that conCAROT learns to condition on specific domains (All OOD CARs in Figure A8&Figure A9).

To determine if conCAROT benefited from training on more CAR variants in the OOD setting, we trained 30 conCAROT models by leaving one CAR variant out of the training data and evaluating on the left out CAR variant. Again, conCAROT outperforms the identity model for both CD4 and CD8 in terms of the R^2 and MMD, even on the OOD setting (Figure A10C&D). The UMAP projections for source, target and transport show good mixing of target and transport for CD4 CAR variants (Figure A11). For the CD8 subsets, we observed that it is difficult to capture the distinct patterns for different CAR variants, as variants with clear expression pattern show similar transport maps (Figure A12). This could again be due to many CAR variants having similar gene expression, hindering the model to condition on the CAR variant. However, the OOD setting is also harder and more CAR variants are likely needed to improve generalising to unseen CAR variants.

4 Limitations

A current limitation of this work is a lack of benchmarking to other perturbation prediction methods, which we are currently working on and will include in the future [21, 20]. Additionally, the results shown here are limited to one dataset based on healthy donors. Generalisation to other datasets or to CAR T cells used in cancer patients needs to be verified in future work. We would also like to point out that we did not specifically evaluate unseen domains. For the CD8 conCAROT-OOD models, all domains were always in the training data. However, for the CD4 conCAROT-OOD model trained only on CARs with >750 cells the CTLA4 domain was not observed at all during training. Nevertheless, we did not do an extensive analysis of this model's performance on CTLA4 CARs specifically nor ran an experiment with deliberately held out domains. This setting is also planned in future work.

5 Discussion

We presented a novel and promising approach for modeling the effect of CAR design on the CAR-expressing T cell population that can exend to unseen CAR variants. Using optimal transport, a method which naturally handles unpaired measurements under multiple conditions, we map control cells with a non-signalling CAR to cells expressing different CAR variants. We showed that we can accurately infer perturbations of multiple CAR variants when conCAROT is conditioned on ESM embeddings of the CAR design. These predictions capture biological characteristics of CAR variants, making *in silico* CAR screening feasible. Additionally, we showed promising results of using the conCAROT to predict gene expressing for cells expressing unseen CAR variants. This model used domain-specific patterns learned from CARs in the training data to prediction for unseen variants.

This approach is adaptable to future indicators of clinical relevance and future CAR variants. Currently it is unclear which characteristics to use for *in vitro* CAR variant selection for clinical outcome [45]; many possible genesets and downstream analyses are possible. A future study could point to new clinically relevant indicators that can be derived from gene expression, not just the functionalities or cell states investigated here. Different downstream analyses on the predicted gene expression could facilitate easy adaption to novel CAR T cell characteristics. Additionally, we could also retrain the model using novel genes if the novel indicators are not well captured in the current geneset. As we use a protein language model for embedding the CAR variants, all future CAR variants can be encoded as long as they can be represented with an amino acid sequence. Previous CAR T cell models focused only on stemness and cytotoxicity and encoded variants with a binary encoding [13]. This prevents the previous models from extending to CARs with a different library design and adaptation to other readouts without retraining.

Increasing CAR library size could improve OOD generalisation, especially when it includes more CARs with distinct patterns. We observed that many CAR variants responded similarly in the repeated antigen stimulation setting. This hinders the model's ability to condition on the CAR embedding, since the variable that it is conditioned on can have little effect. Together, this hampers the accurate prediction of CAR variants with distinct effect. A larger library could overcome some of these hurdles, as it gives the model more examples from which to learn more general patterns. Also, a larger library would most likely have more variants with distinct effects, improving the model's prediction of distinct CAR variants.

A CAR-specific model could help screen patients for approved CAR T cell therapies. Patients have different responses and success rates for given (approved) CAR T cell therapies [2]. A model specialized on making predictions for one CAR variant from different control populations could make valuable predictions given a simple blood sample from patients. These predictions could give indication whether or not the patients might have a promising T cell phenotype distribution given the CAR variant, for example, indications of memory response or cytotoxicity of the resulting CAR T cells. In this study, we did not have enough data to investigate this approach, as splitting by donor leads to too few cells to effectively train the optimal transport model.

6 Data and code availability

Code is available under https://github.com/AI4SCR/CAR-conditional-monge/ and data will be accessible upon publication of the data paper under in the GEO database under accession number GSE262686 [42].

7 Acknowledgements

This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 955321, IBM Research and NCCR Molecular Systems Engineering.

8 Competing interests

S.T.R. holds shares of Alloy Therapeutics and Engimmune Therapeutics. S.T.R. is on the scientific advisory board of Alloy Therapeutics and Engimmune Therapeutics.

References

[1] Valentine Wang, Mélanie Gauthier, Véronique Decot, Loïc Reppel, and Danièle Bensoussan. Systematic Review on CAR-T Cell Clinical Trials Up to 2022: Academic Center Input. *Cancers*, 15(4):1003, January 2023. ISSN 2072-6694. doi: 10.3390/cancers15041003. URL https://www.mdpi.com/2072-6694/15/4/1003. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[2] Kathryn M. Cappell and James N. Kochenderfer. Long-term outcomes following CAR T cell therapy: what we know so far. *Nature Reviews Clinical Oncology*, 20(6):359–371, June 2023. ISSN 1759-4782. doi: 10.1038/s41571-023-00754-1. URL https://www.nature.com/articles/s41571-023-00754-1. Publisher: Nature Publishing Group.

[3] J. Joseph Melenhorst, Gregory M. Chen, Meng Wang, David L. Porter, Changya Chen, McKensie A. Collins, Peng Gao, Shovik Bandyopadhyay, Hongxing Sun, Ziran Zhao, Stefan Lundh, Iulian Pruteanu-Malinici, Christopher L. Nobles, Sayantan Maji, Noelle V. Frey, Saar I. Gill, Alison W. Loren, Lifeng Tian, Irina Kulikovskaya, Minnal Gupta, David E. Ambrose, Megan M. Davis, Joseph A. Fraietta, Jennifer L. Brogdon, Regina M. Young, Anne Chew, Bruce L. Levine, Donald L. Siegel, Cécile Alanio, E. John Wherry, Frederic D. Bushman, Simon F. Lacey, Kai Tan, and Carl H. June. Decade-long leukaemia remissions with persistence of CD4+ CAR T cells. *Nature*, 602(7897):503–509, February 2022. ISSN 1476-4687. doi: 10.1038/s41586-021-04390-6. URL https://www.nature.com/articles/s41586-021-04390-6. Publisher: Nature Publishing Group.

[4] Robert C. Sterner and Rosalie M. Sterner. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer Journal*, 11(4):1–11, April 2021. ISSN 2044-5385. doi: 10.1038/s41408-021-00459-7. URL https://www.nature.com/articles/s41408-021-00459-7. Publisher: Nature Publishing Group.

[5] Nikhil C. Munshi, Larry D. Anderson, Nina Shah, Deepu Madduri, Jesús Berdeja, Sagar Lonial, Noopur Raje, Yi Lin, David Siegel, Albert Oriol, Philippe Moreau, Ibrahim Yakoub-Agha, Michel Delforge, Michele Cavo, Hermann Einsele, Hartmut Goldschmidt, Katja Weisel, Alessandro Rambaldi, Donna Reece, Fabio Petrocca, Monica Massaro, Jamie N. Connarn, Shari Kaiser, Payal Patel, Liping Huang, Timothy B. Campbell, Kristen Hege, and Jesús San-Miguel. Idecabtagene Vicleucel in Relapsed and Refractory Multiple Myeloma. *New England Journal of Medicine*, 384(8):705–716, February 2021. ISSN 0028-4793. doi: 10.1056/NEJMoa2024850. URL https://doi.org/10.1056/NEJMoa2024850. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa2024850.

[6] Daniel W. Lee, James N. Kochenderfer, Maryalice Stetler-Stevenson, Yongzhi K. Cui, Cindy Delbrook, Steven A. Feldman, Terry J. Fry, Rimas Orentas, Marianna Sabatino, Nirali N. Shah, Seth M. Steinberg, Dave Stroncek, Nick Tschernia, Constance Yuan, Hua Zhang, Ling Zhang,

Steven A. Rosenberg, Alan S. Wayne, and Crystal L. Mackall. T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. *Lancet (London, England)*, 385(9967):517–528, February 2015. ISSN 1474-547X. doi: 10.1016/S0140-6736(14)61403-3.

[7] Rebecca C. Larson and Marcela V. Maus. Recent advances and discoveries in the mechanisms and functions of CAR T cells. *Nature Reviews Cancer*, 21(3):145–161, March 2021. ISSN 1474-1768. doi: 10.1038/s41568-020-00323-z. URL https://www.nature.com/articles/s41568-020-00323-z. Publisher: Nature Publishing Group.

[8] Benjamin I. Philipson, Roddy S. O'Connor, Michael J. May, Carl H. June, Steven M. Albelda, and Michael C. Milone. 4-1BB costimulation promotes CAR T cell survival through noncanonical NF-B signaling. *Science Signaling*, 13(625), March 2020. ISSN 1945-0877, 1937-9145. doi: 10.1126/scisignal.aay8248. URL https://stke.sciencemag.org/content/13/625/eaay8248. Publisher: American Association for the Advancement of Science Section: Research Article.

[9] Xiangyu Zhao, Junfang Yang, Xian Zhang, Xin-An Lu, Min Xiong, Jianping Zhang, Xiaosu Zhou, Feifei Qi, Ting He, Yanping Ding, Xuelian Hu, Floris De Smet, Peihua Lu, and Xiaojun Huang. Efficacy and Safety of CD28- or 4-1BB-Based CD19 CAR-T Cells in B Cell Acute Lymphoblastic Leukemia. *Molecular Therapy Oncolytics*, 18:272–281, June 2020. ISSN 2372-7705. doi: 10.1016/j.omto.2020.06.016. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7378699/.

[10] Omkar U. Kawalekar, Roddy S. O'Connor, Joseph A. Fraietta, Lili Guo, Shannon E. McGettigan, Avery D. Posey, Prachi R. Patel, Sonia Guedan, John Scholler, Brian Keith, Nathaniel W. Snyder, Ian A. Blair, Michael C. Milone, and Carl H. June. Distinct Signaling of Coreceptors Regulates Specific Metabolism Pathways and Impacts Memory Development in CAR T Cells. *Immunity*, 44(2):380–390, February 2016. ISSN 1074-7613. doi: 10.1016/j.immuni.2016.01.021. URL https://www.cell.com/immunity/abstract/S1074-7613(16)30008-5. Publisher: Elsevier.

[11] Rocío Castellanos-Rueda, Raphaël B. Di Roberto, Florian Bieberich, Fabrice S. Schlatter, Darya Palianina, Oanh T. P. Nguyen, Edo Kapetanovic, Heinz Läubli, Andreas Hierlemann, Nina Khanna, and Sai T. Reddy. speedingCARs: accelerating the engineering of CAR T cells by signaling domain shuffling and single-cell sequencing. *Nature Communications*, 13(1):6555, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34141-8. URL https://www.nature.com/articles/s41467-022-34141-8. Number: 1 Publisher: Nature Publishing Group.

[12] Daniel B. Goodman, Camillia S. Azimi, Kendall Kearns, Alexis Talbot, Kiavash Garakani, Julie Garcia, Nisarg Patel, Byungjin Hwang, David Lee, Emily Park, Vivasvan S. Vykunta, Brian R. Shy, Chun Jimmie Ye, Justin Eyquem, Alexander Marson, Jeffrey A. Bluestone, and Kole T. Roybal. Pooled screening of CAR T cells identifies diverse immune signaling domains for next-generation immunotherapies. *Science Translational Medicine*, 14(670):eabm1463, November 2022. ISSN 1946-6242. doi: 10.1126/scitranslmed.abm1463.

[13] Kyle G. Daniels, Shangying Wang, Milos S. Simic, Hersh K. Bhargava, Sara Capponi, Yurie Tonai, Wei Yu, Simone Bianco, and Wendell A. Lim. Decoding CAR T cell phenotype using combinatorial signaling motif libraries and machine learning. *Science*, 378(6625):1194–1200, December 2022. doi: 10.1126/science.abq0225. URL https://www.science.org/doi/10.1126/science.abq0225. Publisher: American Association for the Advancement of Science.

[14] Jong Keon Jang, Junhee Pyo, Chong Hyun Suh, Hye Sun Park, Young Kwang Chae, and Kyung Won Kim. Safety and Efficacy of Chimeric Antigen Receptor T-Cell Therapy for Glioblastoma: A Systemic Review and Meta-Analysis. *Frontiers in Oncology*, 12:851877, 2022. ISSN 2234-943X. doi: 10.3389/fonc.2022.851877.

[15] Khloe S. Gordon, Taeyoon Kyung, Caleb R. Perez, Patrick V. Holec, Azucena Ramos, Angela Q. Zhang, Yash Agarwal, Yunpeng Liu, Catherine E. Koch, Alina Starchenko, Brian A. Joughin, Douglas A. Lauffenburger, Darrell J. Irvine, Michael T. Hemann, and Michael E.

Birnbaum. Screening for CD19-specific chimaeric antigen receptors with enhanced signalling via a barcoded library of intracellular domains. *Nature Biomedical Engineering*, 6(7):855–866, July 2022. ISSN 2157-846X. doi: 10.1038/s41551-022-00896-0. URL https://www.nature.com/articles/s41551-022-00896-0. Number: 7 Publisher: Nature Publishing Group.

[16] Luise Fischer, Nora Grieb, Patrick Born, Ronald Weiss, Sabine Seiffert, Andreas Boldt, Stephan Fricke, Paul Franz, Simone Heyn, Anne Sophie Kubasch, Ronny Baber, Heike Weidner, Song Yau Wang, Enrica Bach, Sandra Hoffmann, Jule Ussmann, Janine Kirchberg, Saskia Hell, Sebastian Schwind, Klaus H. Metzeler, Marco Herling, Madlen Jentzsch, Georg-Nikolaus Franke, Ulrich Sack, Kristin Reiche, Ulrike Köhl, Uwe Platzbecker, Vladan Vucinic, and Maximilian Merz. Cellular dynamics following CAR T cell therapy are associated with response and toxicity in relapsed/refractory myeloma. *Leukemia*, 38(2): 372–382, February 2024. ISSN 1476-5551. doi: 10.1038/s41375-023-02129-y. URL https://www.nature.com/articles/s41375-023-02129-y. Number: 2 Publisher: Nature Publishing Group.

[17] Tina Sarén, Mohanraj Ramachandran, Gustav Gammelgård, Tanja Lövgren, Claudio Mirabello, Åsa K. Björklund, Kristina Wikström, Jamileh Hashemi, Eva Freyhult, Håkan Ahlström, Rose-Marie Amini, Hans Hagberg, Angelica Loskog, Gunilla Enblad, and Magnus Essand. Single-Cell RNA Analysis Reveals Cell-Intrinsic Functions of CAR T Cells Correlating with Response in a Phase II Study of Lymphoma Patients. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 29(20):4139–4152, October 2023. ISSN 1557-3265. doi: 10.1158/1078-0432.CCR-23-0178.

[18] Joseph A. Fraietta, Christopher L. Nobles, Morgan A. Sammons, Stefan Lundh, Shannon A. Carty, Tyler J. Reich, Alexandria P. Cogdill, Jennifer J. D. Morrissette, Jamie E. DeNizio, Shantan Reddy, Young Hwang, Mercy Gohil, Irina Kulikovskaya, Farzana Nazimuddin, Minnal Gupta, Fang Chen, John K. Everett, Katherine A. Alexander, Enrique Lin-Shiao, Marvin H. Gee, Xiaojun Liu, Regina M. Young, David Ambrose, Yan Wang, Jun Xu, Martha S. Jordan, Katherine T. Marcucci, Bruce L. Levine, K. Christopher Garcia, Yangbing Zhao, Michael Kalos, David L. Porter, Rahul M. Kohli, Simon F. Lacey, Shelley L. Berger, Frederic D. Bushman, Carl H. June, and J. Joseph Melenhorst. Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. *Nature*, 558(7709):307–312, June 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0178-z. URL https://www.nature.com/articles/s41586-018-0178-z. Publisher: Nature Publishing Group.

[19] Yuge Ji, Mohammad Lotfollahi, F. Alexander Wolf, and Fabian J. Theis. Machine learning for perturbational single-cell omics. *Cell Systems*, 12(6):522–537, June 2021. ISSN 2405-4712, 2405-4720. doi: 10.1016/j.cels.2021.05.016. URL https://www.cell.com/cell-systems/abstract/S2405-4712(21)00202-7. Publisher: Elsevier.

[20] Mohammad Lotfollahi, F. Alexander Wolf, and Fabian J. Theis. scGen predicts single-cell perturbation responses. *Nature Methods*, 16(8):715–721, August 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0494-8. URL https://www.nature.com/articles/s41592-019-0494-8. Number: 8 Publisher: Nature Publishing Group.

[21] Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, Yuge Ji, Ignacio L. Ibarra, F. Alexander Wolf, Nafissa Yakubova, Fabian J. Theis, and David Lopez-Paz. Compositional perturbation autoencoder for single-cell response modeling. April 2021. doi: 10.1101/2021.04.14.439903. URL http://biorxiv.org/lookup/doi/10.1101/2021.04.14.439903.

[22] Charlotte Bunne, Stefan G. Stark, Gabriele Gut, Jacobo Sarabia del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, pages 1–10, September 2023. ISSN 1548-7105. doi: 10.1038/s41592-023-01969-x. URL https://www.nature.com/articles/s41592-023-01969-x. Publisher: Nature Publishing Group.

[23] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S.

Lander. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell*, 176(4):928–943.e22, February 2019. ISSN 1097-4172. doi: 10.1016/j.cell.2019.01.006.

[24] Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics. *Proceedings of Machine Learning Research*, 119:9526–9536, July 2020. ISSN 2640-3498.

[25] Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, Laetitia Meng-Papaxanthos, Michael Sterr, Aimée Bastidas-Ponce, Marta Tarquis-Medina, Heiko Lickert, Mostafa Bakhti, Mor Nitzan, Marco Cuturi, and Fabian J. Theis. Mapping cells through time and space with moscot, May 2023. URL https://www.biorxiv.org/content/10.1101/2023.05.11.540374v2. Pages: 2023.05.11.540374 Section: New Results.

[26] Noa Moriel, Enes Senel, Nir Friedman, Nikolaus Rajewsky, Nikos Karaiskos, and Mor Nitzan. NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nature Protocols*, 16(9):4177–4200, September 2021. ISSN 1750-2799. doi: 10.1038/s41596-021-00573-7.

[27] Federico Gossi, Pushpak Pati, Panagiotis Chouvardas, Adriano Luca Martinelli, Marianna Kruithof-de Julio, and Maria Anna Rapsomaniki. Matching single cells across modalities with contrastive learning and optimal transport. *Briefings in Bioinformatics*, 24(3):bbad130, April 2023. ISSN 1467-5463. doi: 10.1093/bib/bbad130. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10199774/.

[28] Pinar Demetci, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. SCOT: Single-Cell Multi-Omics Alignment with Optimal Transport. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 29(1):3–18, January 2022. ISSN 1557-8666. doi: 10.1089/cmb.2021.0446.

[29] Geert-Jan Huizing, Ina Maria Deutschmann, Gabriel Peyré, and Laura Cantini. Paired single-cell multi-omics data integration with Mowgli. *Nature Communications*, 14(1):7711, November 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-43019-2. URL https://www.nature.com/articles/s41467-023-43019-2. Publisher: Nature Publishing Group.

[30] Eustasio del Barrio, Hristo Inouzhe, Jean-Michel Loubes, Carlos Matrán, and Agustín Mayo-Íscar. optimalFlow: optimal transport approach to flow cytometry gating and population matching. *BMC Bioinformatics*, 21(1):479, October 2020. ISSN 1471-2105. doi: 10.1186/s12859-020-03795-w. URL https://doi.org/10.1186/s12859-020-03795-w.

[31] Charlotte Bunne, Andreas Krause, and Marco Cuturi. Supervised Training of Conditional Monge Maps. 2022.

[32] Benedek Harsanyi, Marianna Rapsomaniki, and Jannis Born. Learning Drug Perturbations via Conditional Map Estimators. In *ICLR 2024 Workshop on Machine Learning for Genomics Explorations*, 2024. URL https://openreview.net/forum?id=FE7lRuwmfI.

[33] Jiajia Liu, Mengyuan Yang, Yankai Yu, Haixia Xu, Kang Li, and Xiaobo Zhou. Large language models in bioinformatics: applications and perspectives. *ArXiv*, page arXiv:2401.04155v1, January 2024. ISSN 2331-8422. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10802675/.

[34] Charlotte Bunne, Geoffrey Schiebinger, Andreas Krause, Aviv Regev, and Marco Cuturi. Optimal transport for single-cell and spatial omics. *Nature Reviews Methods Primers*, 4(1):1–21, August 2024. ISSN 2662-8449. doi: 10.1038/s43586-024-00334-2. URL https://www.nature.com/articles/s43586-024-00334-2. Publisher: Nature Publishing Group.

[35] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. Imprimerie royale, 1781. URL https://books.google.ch/books?id=IG7CGwAACAAJ.

[36] Charlotte Bunne, Laetitia Papaxanthos, Andreas Krause, and Marco Cuturi. Proximal Optimal Transport Modeling of Population Dynamics. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pages 6511–6528. PMLR, May 2022. URL https://proceedings.mlr.press/v151/bunne22a.html. ISSN: 2640-3498.

[37] Ashok Vardhan Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason D. Lee. Optimal transport mapping via input convex neural networks, June 2020. URL http://arxiv.org/abs/1908.10962. arXiv:1908.10962 [cs, stat].

[38] David Alvarez-Melis, Yair Schiff, and Youssef Mroueh. Optimizing Functionals on the Space of Probabilities with Input Convex Neural Networks, November 2021. URL http://arxiv.org/abs/2106.00774. arXiv:2106.00774 [cs, math, stat].

[39] Petr Mokrov, Alexander Korotin, Lingxiao Li, Aude Genevay, Justin M Solomon, and Evgeny Burnaev. Large-Scale Wasserstein Gradient Flows. In *Advances in Neural Information Processing Systems*, volume 34, pages 15243–15256. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/810dfbbebb17302018ae903e9cb7a483-Abstract.html.

[40] Théo Uscidda and Marco Cuturi. The Monge Gap: A Regularizer to Learn All Transport Maps. In *Proceedings of the 40th International Conference on Machine Learning*, pages 34709–34733. PMLR, July 2023. doi: 10.48550/arXiv.2302.04953. URL https://proceedings.mlr.press/v202/uscidda23a.html. ISSN: 2640-3498.

[41] Aude Genevay, Gabriel Peyre, and Marco Cuturi. Learning Generative Models with Sinkhorn Divergences. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR, March 2018. URL https://proceedings.mlr.press/v84/genevay18a.html. ISSN: 2640-3498.

[42] Rocío Castellanos-Rueda, Kai-Ling K. Wang, Juliette L. Forster, Alice Driessen, Jessica A. Frank, María Rodríguez Martínez, and Sai T. Reddy. Dissecting the role of CAR signaling architectures on T cell activation and persistence using pooled screening and single-cell sequencing, January 2024. URL http://biorxiv.org/content/early/2024/02/26/2024.02.26.582129.abstract.

[43] Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, December 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0229-2. URL https://www.nature.com/articles/s41592-018-0229-2. Number: 12 Publisher: Nature Publishing Group.

[44] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, and others. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021. doi: 10.1073/pnas.2016239118. URL https://www.pnas.org/doi/full/10.1073/pnas.2016239118. Publisher: National Acad Sciences.

[45] Nicholas J. Haradhvala and Marcela V. Maus. Understanding Mechanisms of Response to CAR T-cell Therapy through Single-Cell Sequencing: Insights and Challenges. *Blood Cancer Discovery*, pages OF1–OF4, February 2024. ISSN 2643-3230, 2643-3249. doi: 10.1158/2643-3230.BCD-23-0212. URL https://aacrjournals.org/bloodcancerdiscov/article/doi/10.1158/2643-3230.BCD-23-0212/734109/Understanding-Mechanisms-of-Response-to-CAR-T-cell.

[46] Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, 16(12):1289–1296, December 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0619-0. URL https://www.nature.com/articles/s41592-019-0619-0. Publisher: Nature Publishing Group.

[47] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. Van Der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa,

Paul Van Mulbregt, SciPy 1.0 Contributors, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius De Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3): 261–272, March 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0686-2. URL https://www.nature.com/articles/s41592-019-0686-2.

[48] Moshe Sade-Feldman, Keren Yizhak, Stacey L. Bjorgaard, John P. Ray, Carl G de Boer, Russell W. Jenkins, David J. Lieb, Jonathan H. Chen, Dennie T. Frederick, Michal Barzily-Rokni, Samuel S. Freeman, Alexandre Reuben, Paul J. Hoover, Alexandra-Chloé Villani, Elena Ivanova, Andrew Portell, Patrick H. Lizotte, Amir R. Aref, Jean-Pierre Eliane, Marc R. Hammond, Hans Vitzthum, Shauna M. Blackmon, Bo Li, Vancheswaran Gopalakrishnan, Sangeetha M. Reddy, Zachary A Cooper, Cloud P. Paweletz, David A. Barbie, Anat Stemmer-Rachamimov, Keith T. Flaherty, Jennifer A. Wargo, Genevieve M. Boland, Ryan J. Sullivan, Gad Getz, and Nir Hacohen. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell*, 175(4):998–1013.e20, November 2018. ISSN 0092-8674. doi: 10.1016/j.cell.2018.10.038. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6641984/.

[49] Xiaonan Wang, Carlotta Peticone, Ekaterini Kotsopoulou, Berthold Göttgens, and Fernando J. Calero-Nieto. Single-cell transcriptome analysis of CAR T-cell products reveals subpopulations, stimulation, and exhaustion signatures. *OncoImmunology*, 10(1): 1866287, January 2021. ISSN null. doi: 10.1080/2162402X.2020.1866287. URL https://doi.org/10.1080/2162402X.2020.1866287. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/2162402X.2020.1866287.

[50] Qing Deng, Guangchun Han, Nahum Puebla-Osorio, Man Chun John Ma, Paolo Strati, Beth Chasen, Enyu Dai, Minghao Dang, Neeraj Jain, Haopeng Yang, Yuanxin Wang, Shaojun Zhang, Ruiping Wang, Runzhe Chen, Jordan Showell, Sreejoyee Ghosh, Sridevi Patchva, Qi Zhang, Ryan Sun, Frederick Hagemeister, Luis Fayad, Felipe Samaniego, Hans C. Lee, Loretta J. Nastoupil, Nathan Fowler, R. Eric Davis, Jason Westin, Sattva S. Neelapu, Linghua Wang, and Michael R. Green. Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. *Nature Medicine*, 26(12):1878–1887, December 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-1061-7. URL https://www.nature.com/articles/s41591-020-1061-7. Number: 12 Publisher: Nature Publishing Group.

[51] Rachel C. Lynn, Evan W. Weber, Elena Sotillo, David Gennert, Peng Xu, Zinaida Good, Hima Anbunathan, John Lattin, Robert Jones, Victor Tieu, Surya Nagaraja, Jeffrey Granja, Charles F. A. de Bourcy, Robbie Majzner, Ansuman T. Satpathy, Stephen R. Quake, Michelle Monje, Howard Y. Chang, and Crystal L. Mackall. c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature*, 576(7786):293–300, December 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1805-z. URL https://www.nature.com/articles/s41586-019-1805-z. Number: 7786 Publisher: Nature Publishing Group.

[52] Alyssa Sheih, Valentin Voillet, Laïla-Aïcha Hanafi, Hannah A. DeBerg, Masanao Yajima, Reed Hawkins, Vivian Gersuk, Stanley R. Riddell, David G. Maloney, Martin E. Wohlfahrt, Dnyanada Pande, Mark R. Enstrom, Hans-Peter Kiem, Jennifer E. Adair, Raphaël Gottardo, Peter S. Linsley, and Cameron J. Turtle. Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy. *Nature Communications*, 11(1):

219, January 2020. ISSN 2041-1723. doi: 10.1038/s41467-019-13880-1. URL https://www.nature.com/articles/s41467-019-13880-1. Number: 1 Publisher: Nature Publishing Group.

[53] Alexander I. Salter, Richard G. Ivey, Jacob J. Kennedy, Valentin Voillet, Anusha Rajan, Eva J. Alderman, Uliana J. Voytovich, Chenwei Lin, Daniel Sommermeyer, Lingfeng Liu, Jeffrey R. Whiteaker, Raphael Gottardo, Amanda G. Paulovich, and Stanley R. Riddell. Phosphoproteomic analysis of chimeric antigen receptor signaling reveals kinetic and quantitative differences that affect cell function. *Science Signaling*, 11(544), August 2018. ISSN 1945-0877, 1937-9145. doi: 10.1126/scisignal.aat6753. URL https://stke.sciencemag.org/content/11/544/eaat6753. Publisher: American Association for the Advancement of Science Section: Research Article.

[54] Gregory M. Chen, Changya Chen, Rajat K. Das, Peng Gao, Chia-Hui Chen, Shovik Bandyopadhyay, Yang-Yang Ding, Yasin Uzun, Wenbao Yu, Qin Zhu, Regina M. Myers, Stephan A. Grupp, David M. Barrett, and Kai Tan. Integrative bulk and single-cell profiling of premanufacture T-cell populations reveals factors mediating long-term persistence of CAR T-cell therapy. *Cancer Discovery*, January 2021. ISSN 2159-8274, 2159-8290. doi: 10.1158/2159-8290.CD-20-1677. URL https://cancerdiscovery.aacrjournals.org/content/early/2021/04/05/2159-8290.CD-20-1677. Publisher: American Association for Cancer Research Section: Research Article.

[55] Peter A. Szabo, Hanna Mendes Levitin, Michelle Miron, Mark E. Snyder, Takashi Senda, Jinzhou Yuan, Yim Ling Cheng, Erin C. Bush, Pranay Dogra, Puspa Thapa, Donna L. Farber, and Peter A. Sims. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, 10(1):4706, October 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-12464-3. URL https://www.nature.com/articles/s41467-019-12464-3. Number: 1 Publisher: Nature Publishing Group.

[56] Emma C. Morris, Sattva S. Neelapu, Theodoros Giavridis, and Michel Sadelain. Cytokine release syndrome and associated neurotoxicity in cancer immunotherapy. *Nature Reviews Immunology*, pages 1–12, May 2021. ISSN 1474-1741. doi: 10.1038/s41577-021-00547-6. URL https://www.nature.com/articles/s41577-021-00547-6. Publisher: Nature Publishing Group.

[57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[58] Hengshi Yu and Joshua D. Welch. PerturbNet predicts single-cell responses to unseen chemical and genetic perturbations, July 2022. URL https://www.biorxiv.org/content/10.1101/2022.07.20.500854v2. Pages: 2022.07.20.500854 Section: New Results.

A Appendix

A.1 scRNA-seq preprocessing

The scRNA-seq data was quality controlled with nFeature_RNA > 300, nCount_RNA < 50000, percent.mt < 20 and nCount_ADT < 30000. Because of large differences in cell numbers per CAR variant, we applied a subsampling strategy. If a CAR variant has more than 6.25% of the cells in a sample, then these cells downsampled to maximum 6.25% of all cells. (6.25% is deduced from twice the average number of cells per variant 100/32 =3.123%). Cells were scaled and regression was done for cell cycle phase (S and G2M) and percentage of mitochondrial genes. Batch correction for donor and timepoint was done with Harmony, using lambda=c(1,200) for UMAPs in Figure 1 only [46]. This preprocessing was done before downloading the data and cell numbers per variant after quality control and sampling are shown in Figure A1. For detailed experimental set up and scRNA-seq preprocessing see [42].

A.2 Genesets and geneset scoring

The functional geneset was obtained by selecting 77 functional genes from various published research with scRNA-seq of CAR T cells and listed in Table A2. We included five additional genes needed to compute the functional scores listed in Table A1, resulting in the functional geneset of 82 genes. 82 random genes were sampled from genes that are expressed by at least one cell in each CAR variant. The highly variable genes (HVG) were determined by first taking the top 100 HVG of each CAR variant with the control, yielding 30 HVG genesets of 100 genes. Then we counted how often each gene occurred over all the 30 genesets and took the genes that occurred at least 22 times, giving us 81 unique genes. The number 22 was chosen to get a geneset size in the same range as the functional and score geneset size. We decided the final geneset based on the performance of unconditional CAROT and used this geneset for the conCAROT experiments.

To evaluate if the geneset captured differences between control and CAR-expressing cells, we calculated, for each gene, its information content in terms of entropy between CAR-expressing cells and control cells (Figure A2) and compared it with the one obtained from 82 randomly sampled genes and 81 HVG. We observed that the functional and HVG genesets scored similarly, indicating that our functional geneset can differentiate between control and CAR populations equally well as the HVG. To compute the information content of each gene across control and CAR treatment, we computed the relative entropy of the distribution of the sum of library-size normalised counts as follows

$$D = \sum_{k=0}^{G} (p_k * log(p_k/q_k))$$

where p_k and q_k are the counts of a geneset (of length G) for the control and the CAR-expressing cells respectively, normalized such that $\sum p_k = \sum q_k = 1$ [47].

Table A1: Genes used for geneset scoring. Taken from [42]

Cytotoxicity	Proinflamatory	Memory	CD4_Th1	CD4_Th2
GZMB	IFNG	TCF7	IL2	IL5
PRF1	TNF	SELL	IFNG	IL13
FASLG	CRTAM	CCR7	TNF	IL4
	CSF2	LEF1		
	XCL1	IL7R		
	XCL2			
	CCL1			
	CCL4			

A.3 CAR embeddings

To construct the CAR embeddings, we compared different approaches to encode the CAR variant: (i) two binary embeddings, (ii) two ESM-based embeddings, and (iii) a metadata embedding. The two

Table A2: Prior knowledge functional genes from literature. CRS: Cytokine Release Syndrome

Function	Gene	Refs	Function	Gene	Refs
Memory	CD3E	[48]	Exhaustion	ADORA2A	[48]
Memory	CCR7	[49, 48]	Exhaustion	BATF	[50, 48, 51, 49]
Memory	CD28	[48]	Exhaustion	BATF3	[51]
Memory	CD27	[48]	Exhaustion	BTLA	[48, 51, 52, 49]
Memory	SELL	[49, 48]	Exhaustion	CCL1	[49]
Memory	IL7R	[53]	Exhaustion	CCL3	[49, 52]
Memory	TCF7	[54]	Exhaustion	CCL4	[49]
Memory	LEF1	[54]	Exhaustion	CCL5	[49]
Memory	KLF2	[53]	Exhaustion	CD160	[55]
Cytotoxicity	GNLY	[56, 50]	Exhaustion	CD2	[48]
Cytotoxicity	GZMK	[56, 52]	Exhaustion	CD244	[52, 49, 55]
Cytotoxicity	GZMA	[50, 52]	Exhaustion	CD3E	[48]
Cytotoxicity	GZMB	[50, 52]	Exhaustion	CTLA4	[55]
Cytotoxicity	PRF1	[52]	Exhaustion	ENTPD1	[49]
Cytotoxicity	LAG3	[52]	Exhaustion	GZMB	[51]
Cytotoxicity	NKG7	[52]	Exhaustion	HAVCR1	[48]
Cytotoxicity	ZEB2	[56, 54]	Exhaustion	HAVCR2	[48, 52]
Cytotoxicity	EOMES	[56]	Exhaustion	ID2	[50, 49]
Cytotoxicity	ZNF683	[56]	Exhaustion	IFNG	[51]
Cytotoxicity	TBX21	[54]	Exhaustion	IL13	[51]
Cytotoxicity	PRDM1	[54]	Exhaustion	IL17RA	[51]
Proliferation	IL2	[56, 48, 53]	Exhaustion	IL2RA	[51]
Proliferation	LIF	[56]	Exhaustion	IRF4	[51]
Proliferation	CENPV	[56]	Exhaustion	KIR3DL1	[48]
Proliferation	G0S2	[56]	Exhaustion	KLF2	[51]
Proliferation	ORC6	[56]	Exhaustion	KLRG1	[52]
Proliferation	CD3E	[48]	Exhaustion	LAG3	[55]
Proliferation	CD2	[48]	Exhaustion	LAYN	[55]
Proliferation	CD28	[48]	Exhaustion	LEF1	[51, 49]
Proliferation	IL2RA	[48, 51]	Exhaustion	NCAM1	[48]
Proliferation	CD69	[48]	Exhaustion	NCR1	[48]
Proliferation	ICOS	[48]	Exhaustion	PDCD1	[50, 55]
Proliferation	TNFRSF4	[48]	Exhaustion	TCF7	[51]
Proliferation	TNFRSF9	[48]	Exhaustion	TIGIT	[52]
Proliferation	CD27	[48]	Exhaustion	TNFRSF18	[51]
Proliferation	TNF	[48, 53]	CRS	IL1B	[50]
Proliferation	IFNG	[48, 51, 53]	CRS	CXCL8	[50]
Proliferation	GZMB	[51, 53]	CRS	CCL3	[50]
Proliferation	MKI67	[52]	CRS	CCL4	[50]
Proliferation	CDK1	[52]	CRS	IL13	[50]
Proliferation	CCNA2	[52]	CRS	CD69	[50]
Proliferation	CDCA2	[52]	CRS	LEF1	[50]
Proliferation	FOS	[49]	CRS	IL7R	[50]
Proliferation	CCL3	[53]	CRS	STAT1	[50]
Proliferation	CCL4	[53]	CRS	FOXP1	[50]
Proliferation	NCR1	[48]	CRS	CD27	[50]
Proliferation	NCAM1	[48]	CRS	IL16	[50]
Cytokines	CCL3	[55]	CRS	GZMB	[50]
Cytokines	CCL4	[55]	CRS	GZMA	[50]
Cytokines	CCL20	[55]	CRS	BATF	[50]
Cytokines	IFNG	[55]	CRS	GZMH	[50]
Cytokines	IL10	[55]	CRS	IL13	[51]
Cytokines	TNF	[55]	CRS	IL1A	[51]
Cytokines	LAG3	[55]	CRS	CSF2	[51]
Cytokines	CD226	[55]			
Cytokines	HAVCR2	[55]			
Cytokines	HOPX	[55]			

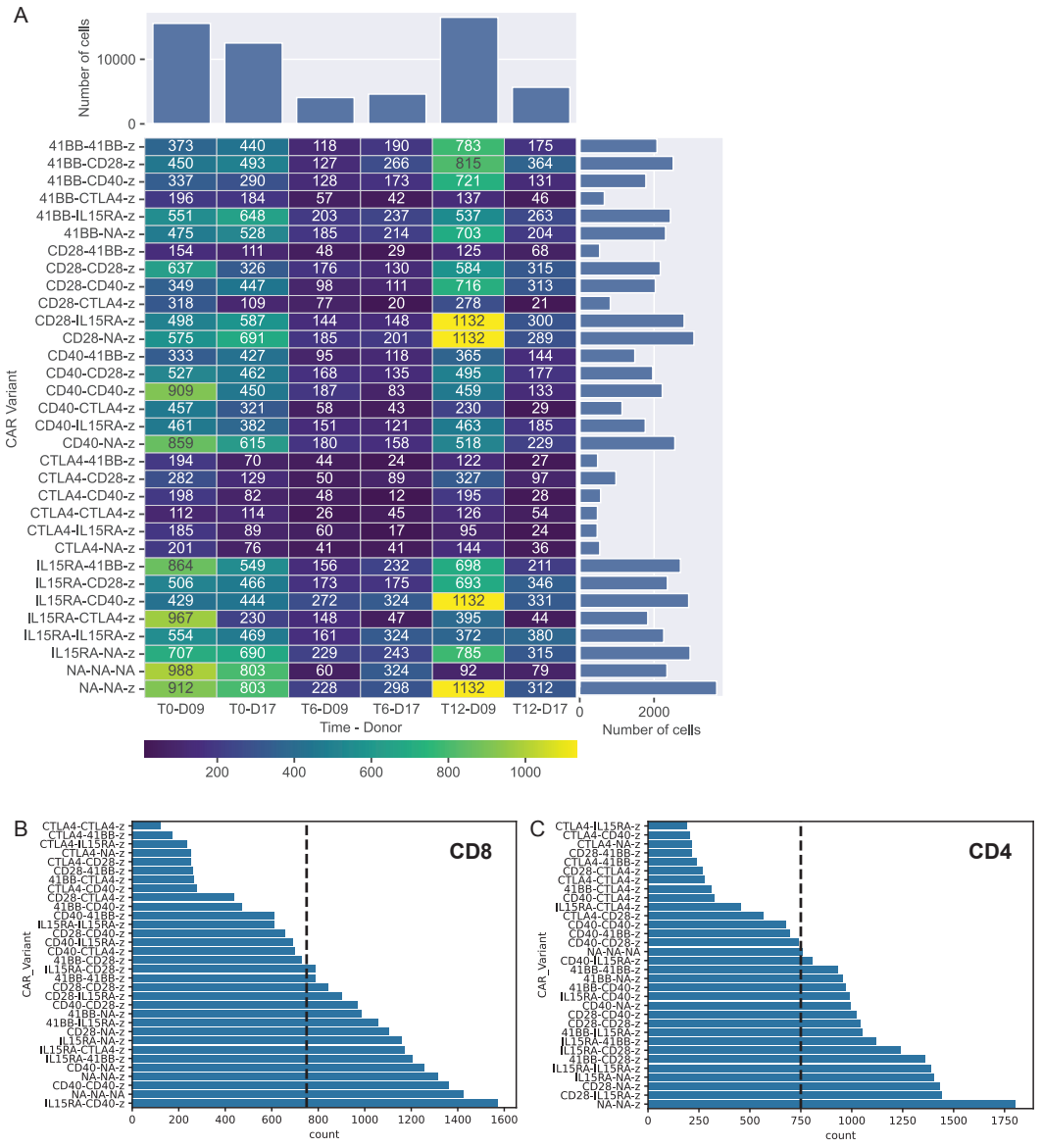

Figure A1: Dataset overview and CAR selection thresholds. A) Overview over cell counts per CAR variant, donor and timpoint. Cells are colored and annotated with cell counts, barplots on top and right show the marginal counts. B) Cell counts per CAR variant in the CD8 subset. Dashed line indicates threshold for selecting the variants for training data. C) Cell counts per variant in the CD4 subset. Dashed line indicates threshold for selecting the variants for training data.

binary CAR embeddings are based on the absence or presence of the five signalling domains and the CD3ζ domain. For the 11-dimensional embedding, the first five bits describe which domain is in position A. Each position corresponds to one of the five possible domains, and of these five bits only the bit corresponding to the domain will have a 1. Then there are five bits that in a similar manner describe which domain is in position B. The last and 11th bit describes the presence (1) or absence (0) of the CD3ζ domain. Similarly, the last bit of the 16-dimensional binary embedding also indicates CD3ζ presence. The 16d embedding has three bits for each of the five signalling domains. Of these three bits, the first position is 1 if the domain is at all present in the CAR variant, irrespective of position. The second bit describes whether the domain is in position A and the third bit position B.

We used Facebook's ESM2 models, either t6_8M_UR50D or t48_15B_UR50D, referred to as ESM small or ESM XL respectively to compare large and small ESM embeddings [44]. We embedded only the intracellular signalling tail, since this is where the CAR variants differ. Then we averaged these embeddings over the sequence, resulting in the length of the embedding dimension of 5120 for ESM XL and 256 for ESM small.

Additionally we used a metadata embedding, which includes information cells that express the CAR variant. For each variant the embedding consists of the means and standard deviation of the scores "Cytotoxicity_1", "Proinflamatory_2", "Memory_3", "CD4_Th1_4", "CD4_Th_5", "S.Score", "G2M.Score". Also, the fraction of cells from each of the donors, timepoints, cell cycle phases and cell states were included, resulting in a 40 dimensional embedding.

The binary embeddings are based on the CAR library used here and cannot embed novel CAR variants, whereas the ESM embeddings are based on the amino acid sequence of the variants and can therefore be extended to novel CAR variants. The metadata embedding captures considerable information that is derived from the scRNA-seq and should therefore greatly facilitate the conditional OT problem. We observed that a larger ESM embedding (ESM XL) performs better than a smaller ESM model (Figure A2). Interestingly, the binary embeddings don't show significantly worse performance than the metadata or ESM embeddings, despite having a lower dimension and only information about the presence/absence of domains. We continued with the ESM XL embedding, since it can be used in an out-of-distribution setting.

A.4 OT UMAPs and biological scores

For comparing biological scores between source, target and transport we sampled the same number of cells from source and target. The number of cells is the minimum of number of cells in the target or source validation set, as the target validation set is CAR variant dependent. Then we transported the target cells to obtain the same number of transport cells. All three datatypes (source, target and transport) were then mapped onto a UMAP based on all CD4 or CD8 cells.

Additionally, we calculated geneset scores for the cytotoxicity, memory and proinflammatory genesets. The thresholding for cells positive for a certain scores are taken from [42] and are 0 for the memory geneset and 1 otherwise. We trained a cell-typing model on all CD4 and CD8 cells, since this seemed to work better dan CD4 and CD8 separately. This Support Vector Machine model was then used to predict the celltypes for source, target and transport. This model was implemented with scikit-learn [57].

A.5 Metrics and evaluation

We leveraged the coefficient of determination (R^2) and the Maximum Mean Discrapancy (MMD) for model evaluation, as these metrics are often used in perturbation single cell models and optimal transport models [22, 31, 32, 20, 58].

For the R^2 the average expression per gene over all cells is computed, for the prediction and target separately. Then the R2 is determined on the average expression of all genes.

$$R2 = 1 - \frac{\sum_{i=1}^{g}(x_i - y_i)^2}{\sum_{i=1}^{g}(x_i - \bar{x})^2}$$

Where g indicated all genes, and x_i is the average target gene expression for gene i and y_i the average predicted gene expression for gene i. $\bar{x}$ is the average gene target gene expression over all genes.

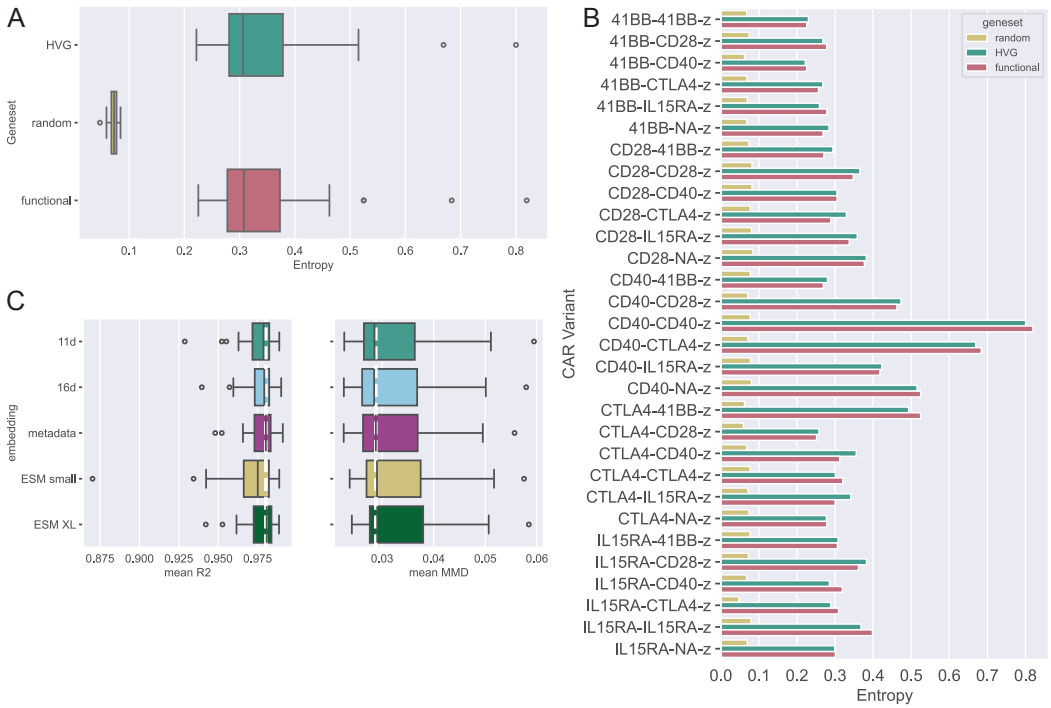

Figure A2: Genesets and CAR embedding evaluation. A) Distribution of entropy between the control gene expression counts and CAR-expressing cells gene expression counts, for genes in the indicated genesets. B) Entropy between control and CAR-expressing cells as in A, only now shown per CAR. C) Comparing different CAR embeddings using the conditional CAROT model, trained on all CAR variants and makes predictions for all CAR variants using a test split. The white line shows the median of the metadata embedding. Left shows the R2 and right the MMD.

We used the MMD to find the maximum distance between these distributions in the kernel space.

$$MMD(T, P) = \|\mu_T - \mu_P\|_K$$

where T is the target distribution over all cells and genes and P the predicted distribution. We take an average over the MMD with a RBF-kernel with $\gamma = [2, 1, 0.5, 0.1, 0.01, 0.005]$.

A.6 Statistical analyses

We tested (con)CAROT(-OOD) versus identity and (con)CAROT(-OOD) versus the within condition using a two-sided Mann-Whitney U test using [47]. This non-parametric test assumes independence between the two groups and ordinal observations. The independence assumption might not hold for our data, since the target difficulty and subsequent score might influence the performance of (con)CAROT(-OOD). We did multiple hypothesis correction using the Bonferroni method, by lowering the significance threshold $\alpha = 0.05$ to $\alpha = 0.05/n$ with $n = 8$ since we tested for each experiment two subsets, two scores and two comparisons ($2^3 = 8$). Significant differences are indicated with an '*' in the manuscript.

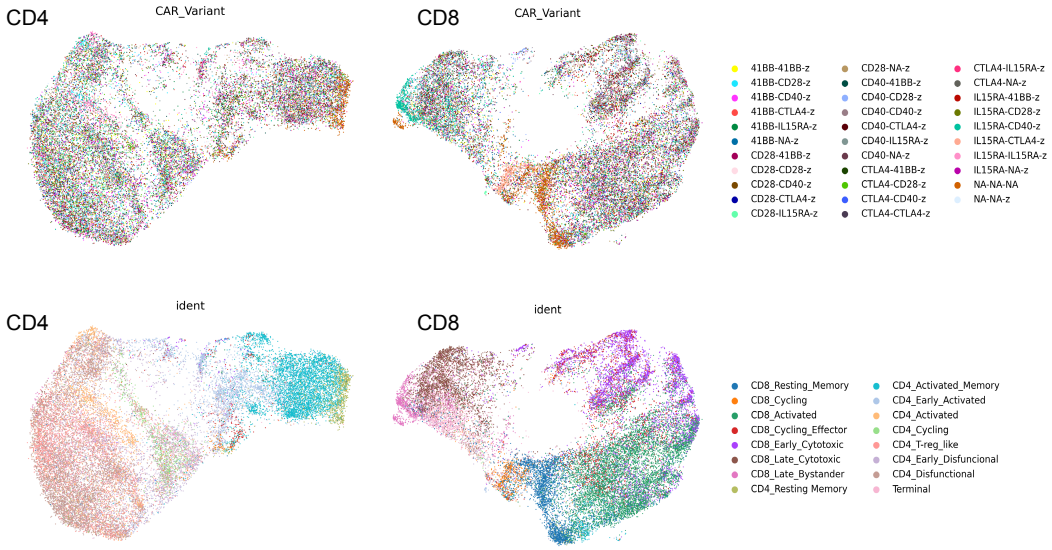

Figure A3: UMAPs of CD8 and CD4 subsets based on the logcounts of the 82 genes in the functional geneset. Colored by CAR variant and cell state.

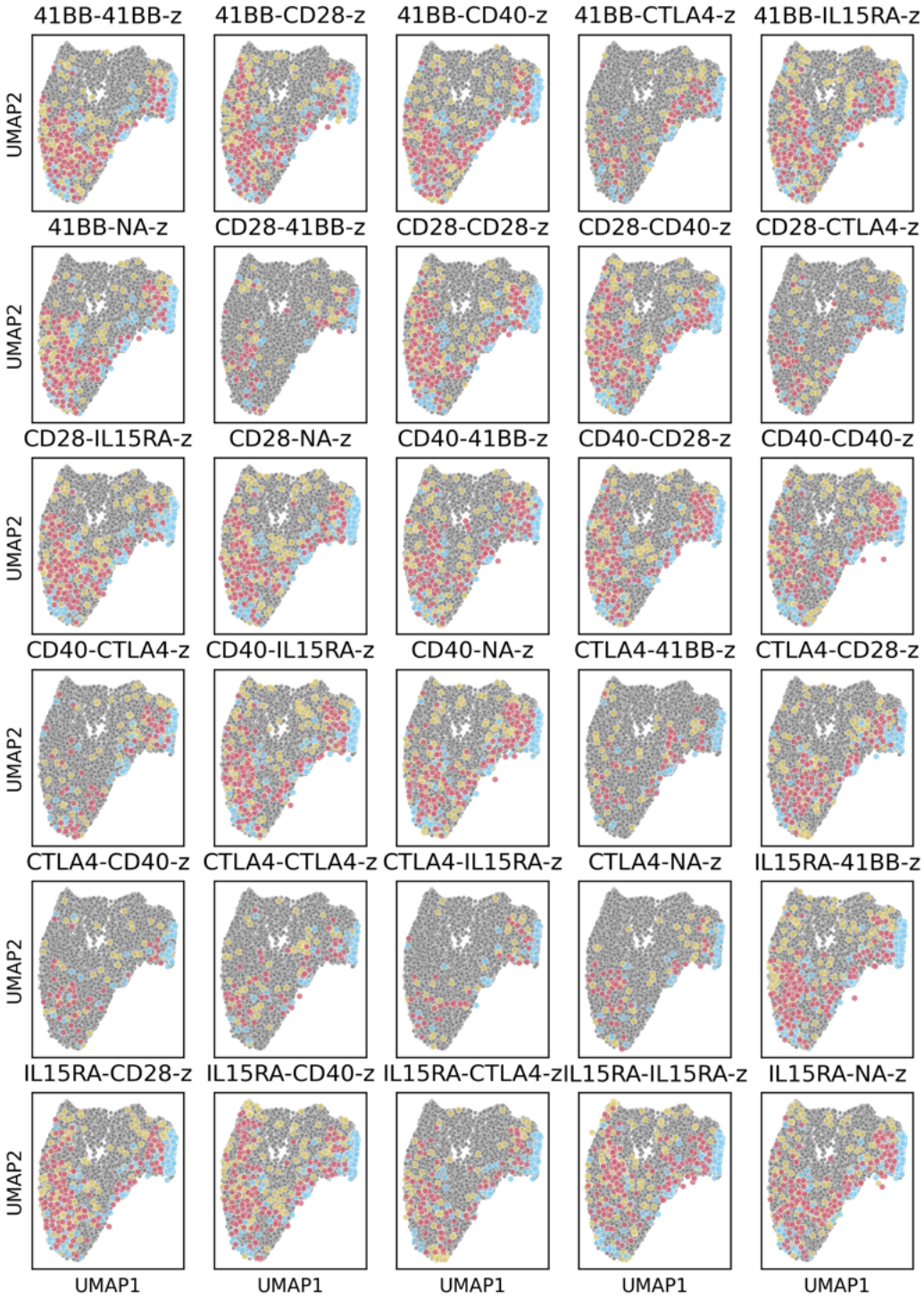

Figure A4: UMAPs of CD4 subset cells for the unconditional OT model (one model per CAR variant and subset).

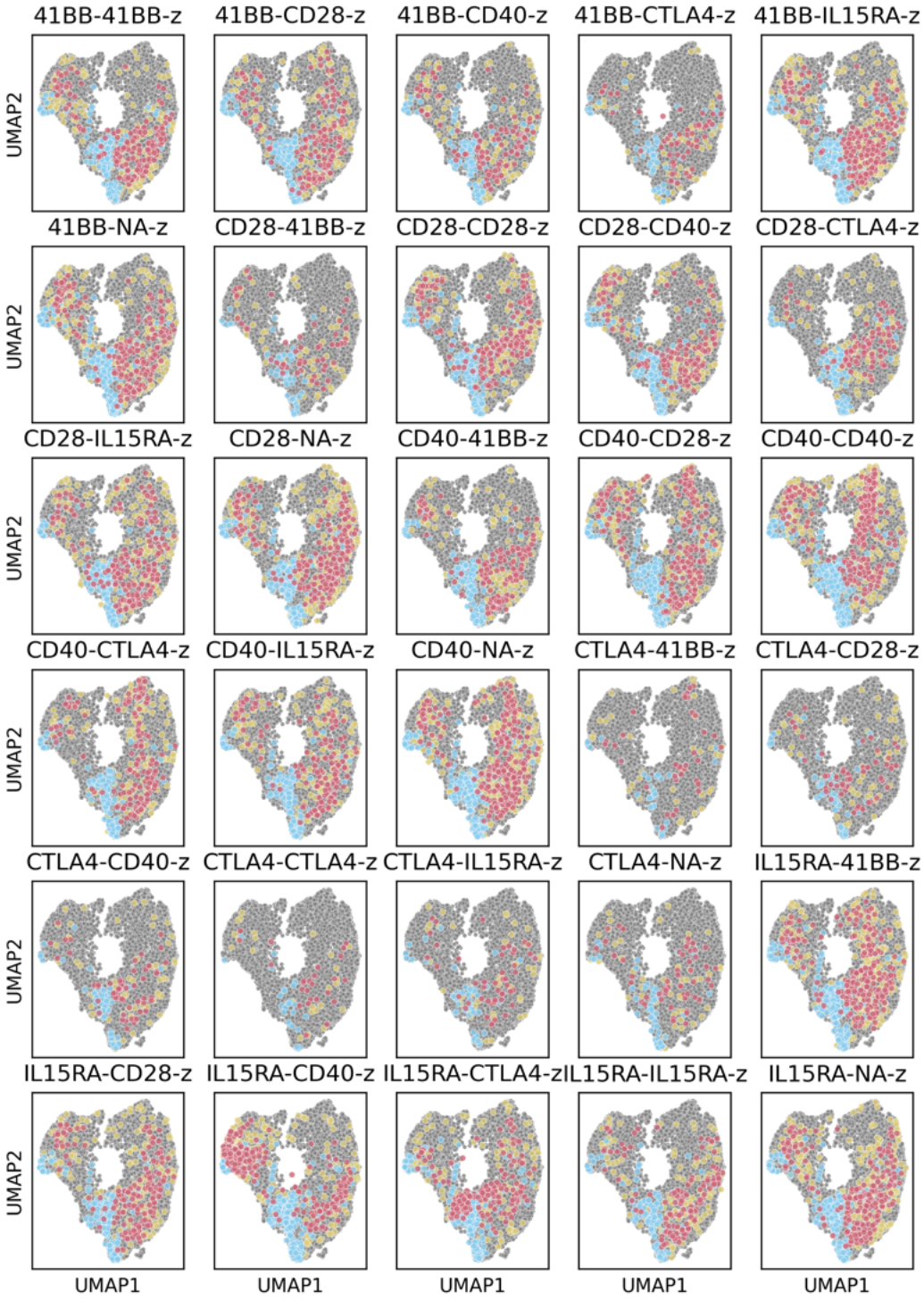

Figure A5: UMAPs of CD8 subset cells for the unconditional OT model (one model per CAR variant and subset).

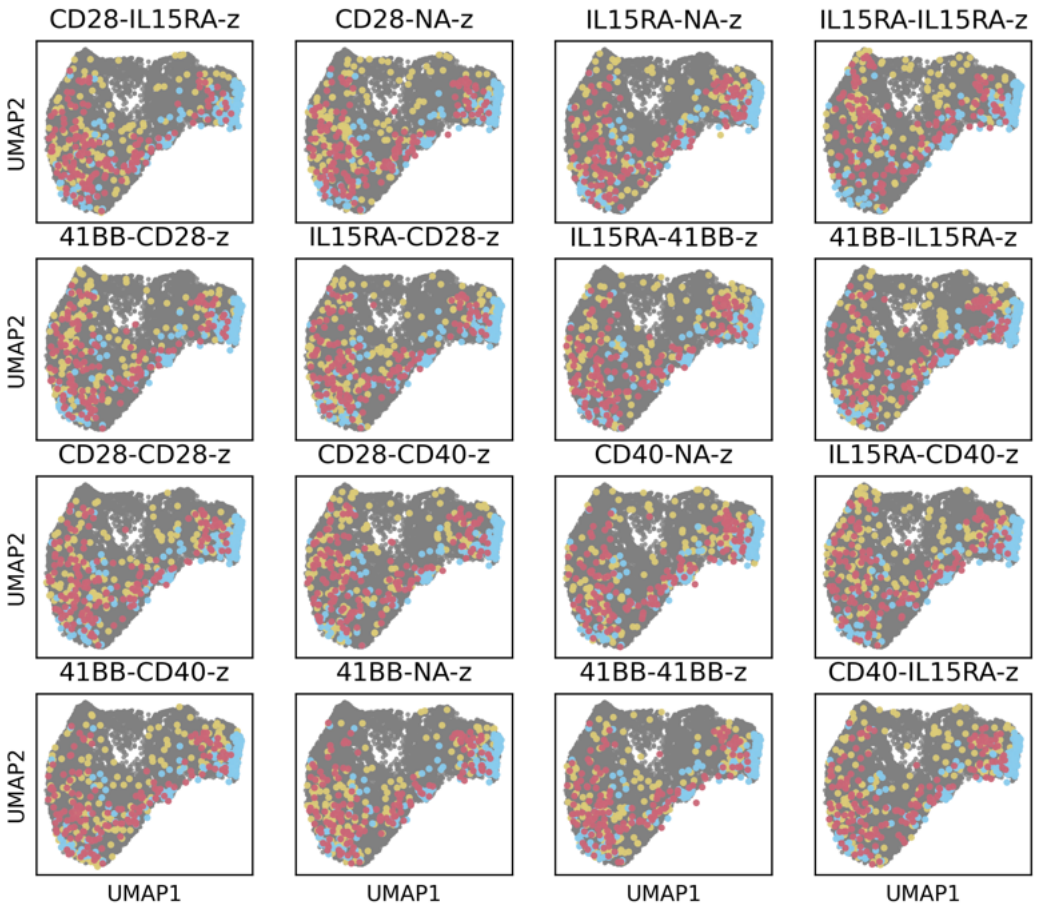

Figure A6: UMAPs of CD4 subset cells for the conditional OT model for CAR variants with >750 cells in the subset, all variants shown here were present on in the training set. Evaluation on a held-out test set.

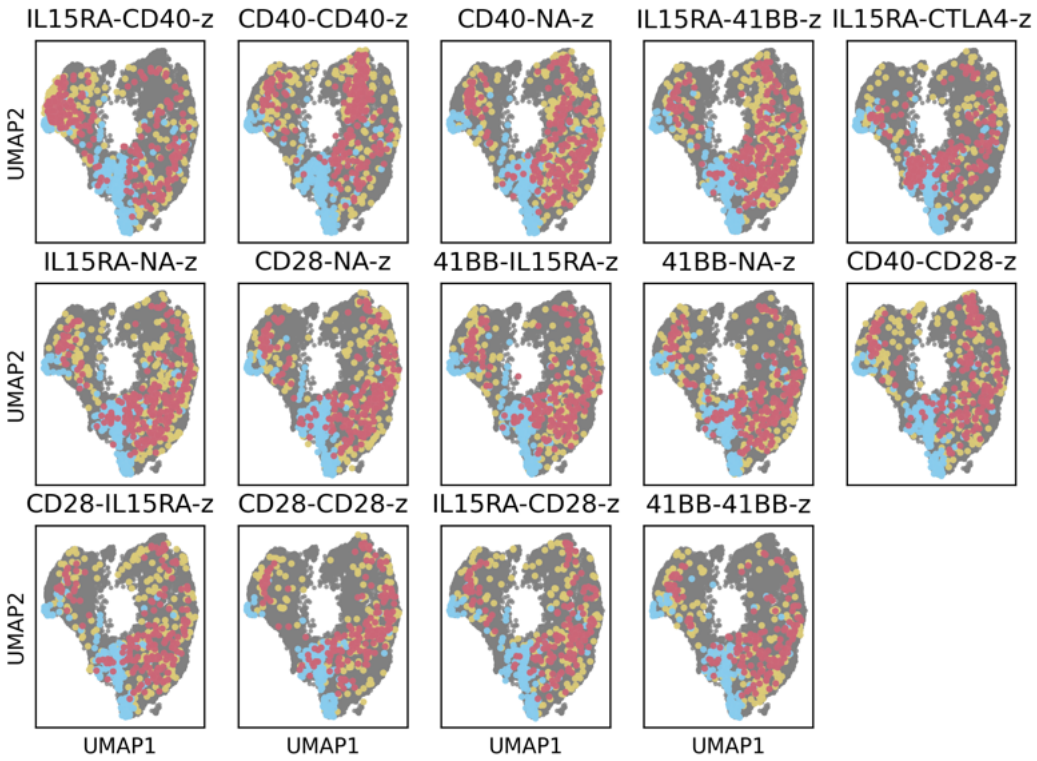

Figure A7: UMAPs of CD8 subset cells for the conditional OT model for CAR variants with >750 cells in the subset, all variants shown here were present on in the training set. Evaluation on a held-out test set.

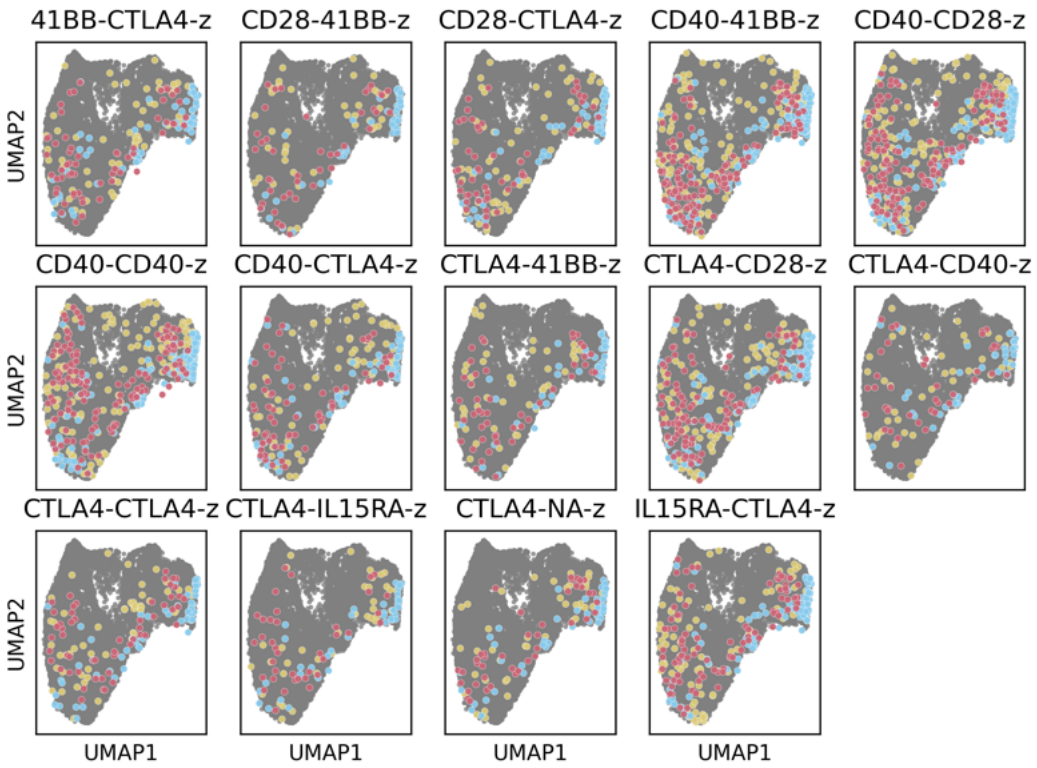

Figure A8: UMAPs of CD4 subset cells for the conditional OT model trained on CAR variants with >750 cells in the subset. All variants shown here were not present on in the training set, shown are all available cells for the variant

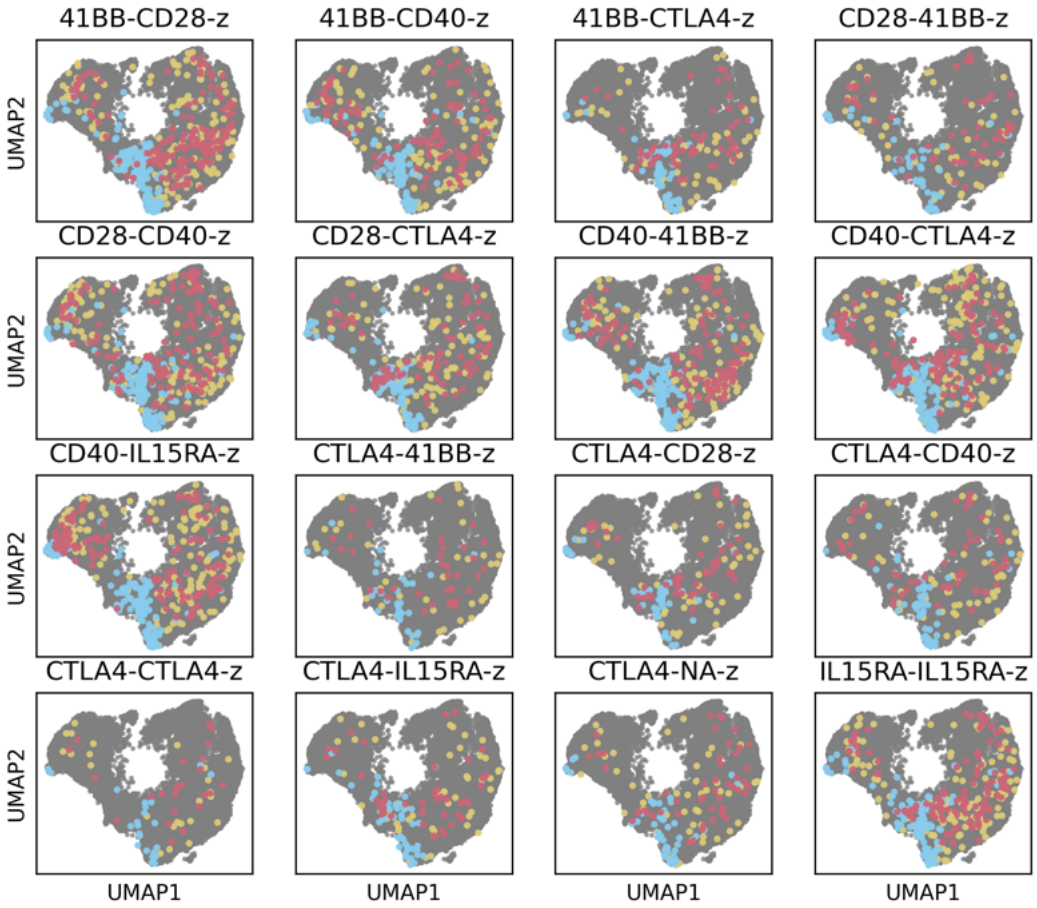

Figure A9: UMAPs of CD8 subset cells for the conditional OT model trained on CAR variants with >750 cells in the subset. All variants shown here were not present on in the training set, shown are all available cells for the variant

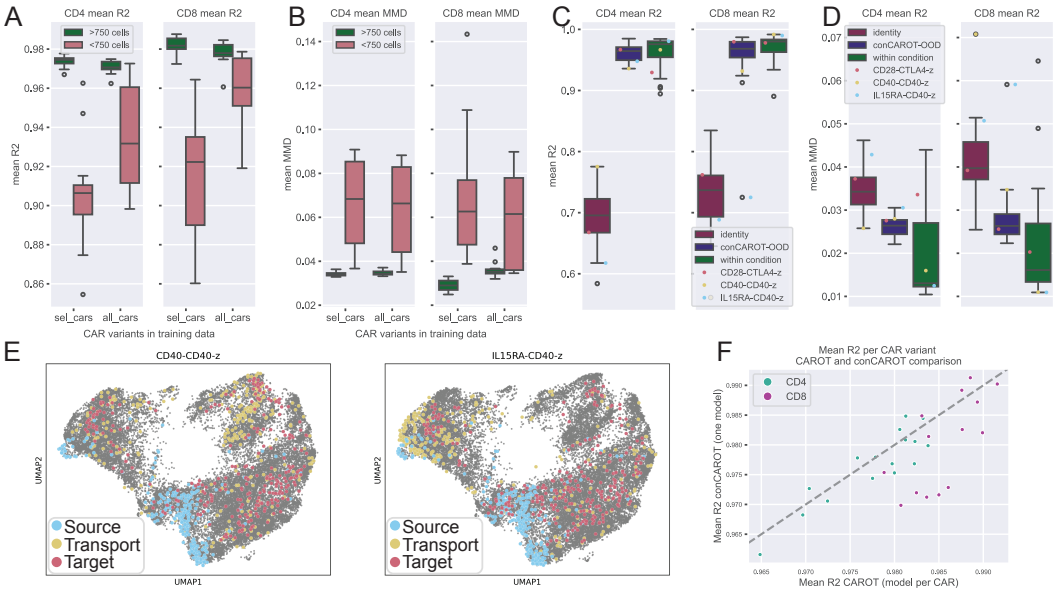

Figure A10: Additional OOD experiment results. A-B) Comparing performance of the model trained with all CAR variants (all_cars) and the model trained with only variants with >750 cells (sel_cars) (x-axis). The evaluation is split by CAR variants with >750 cells and variants with <750 cells. The variants with >750 cells are in distribution for both models, whereas the variants with <750 cells are OOD for the sel_cars model. Plots show the average score over nine samples per variant with the R^2 in A and the MMD in B. C-D) OOD performance of a model trained on all in-distribution CARs, leaving out one CAR at a time. Highlighted are CARs also shown in the UMAPs in the main text and in panel E that show a distinct response. Performance is again averaged over nine samples, with R^2 in C and MMD in D. E) UMAPs of CAR variants also shown in the main text for the OOD model evaluated in C&D. F) Comparison of CAROT (one model per CAR variant) and conditional CAROT (one model trained on all >750 cells CAR Variants), for all in-distribution CARs used in the conditional CAROT training. Variants for which conditional CAROT outperforms CAROT: CD4 - CD28-IL15RA-z, 41BB-CD28-z, IL15RA-41BB-z, CD28-CD40-z, and 41BB-CD40-z. CD8 - 41BB-41BB-z, IL15RA-41BB-z, and IL15RA-NA-z, CD28-NA-z

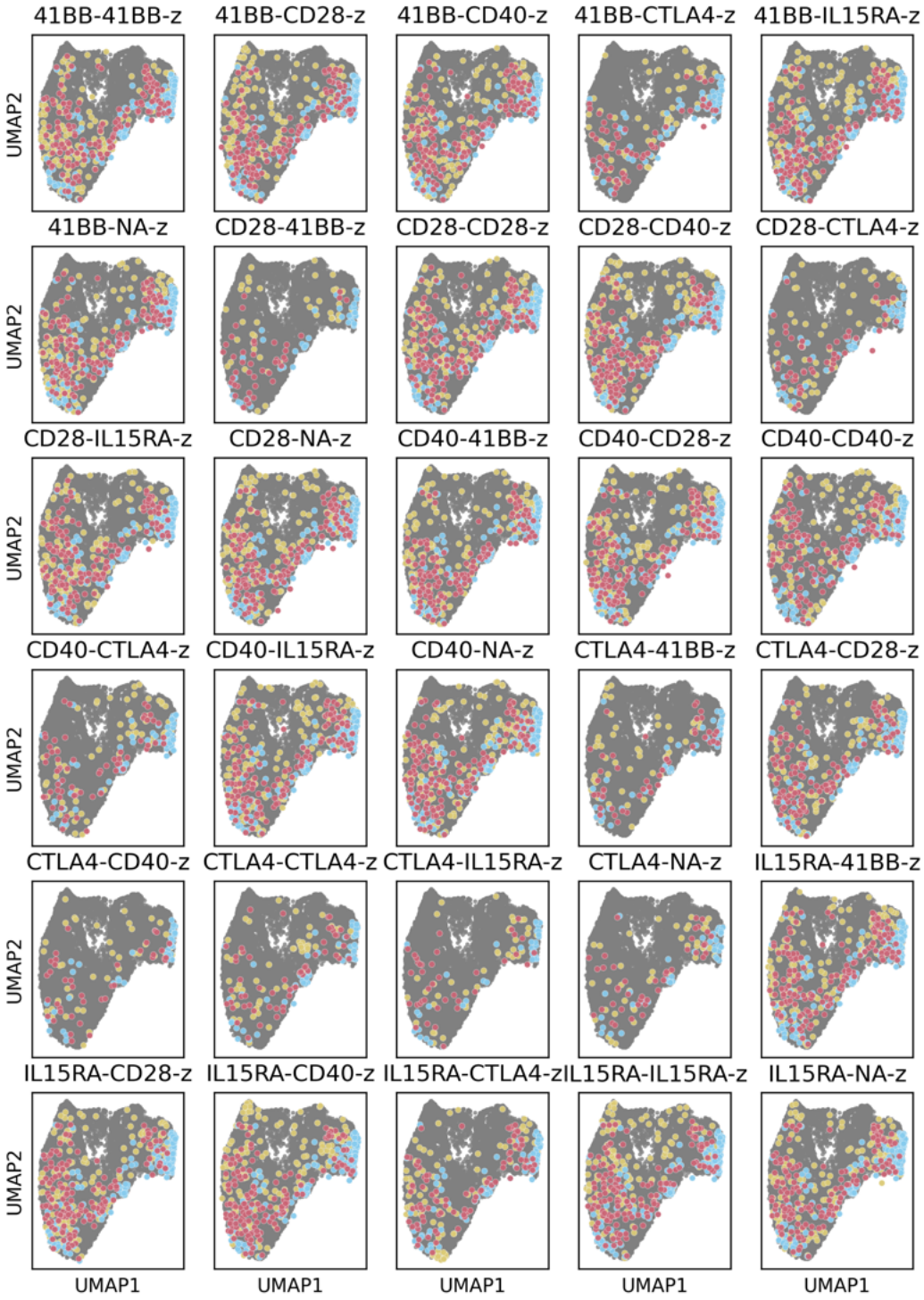

Figure A11: UMAPs of CD4 subset cells for the conditional OT model trained all CAR variants in the OOD setting, leaving the indicated CAR variant out during training.

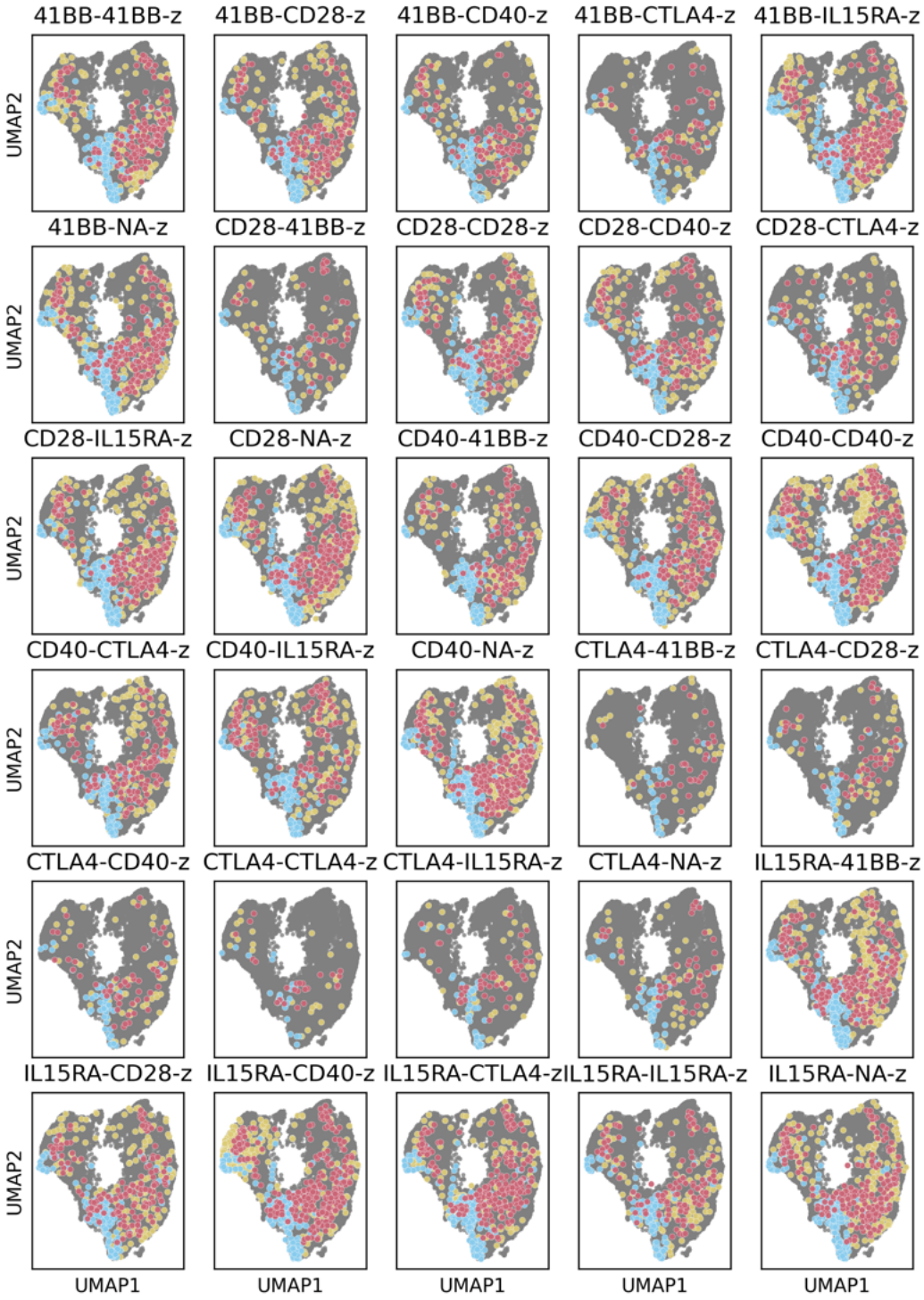

Figure A12: UMAPs of CD8 subset cells for the conditional OT model trained all CAR variants in the OOD setting, leaving the indicated CAR variant out during training.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The results are shown in the Results section.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are mentioned in the section 4. Limitations to the statistical test are mentioned in subsection A.6.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   [NA] .

Justification: The paper does not include theoretical results. Methods have been described in cited works.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All details can be found in the methods section, appendix or the cited works. Code is publicly available on GitHub and we can include the link after review (not before due to anonymity).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Specific instruction are given in subsection 2.2. GitHub link with all code to reproduce the experiments and figures will be made available after review, due to anonymity not during review. Data will be made available after acceptance of the data-related paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: All settings are mentioned in the methods/appendix or otherwise in cited works. Config files will be included in available code after review.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Most quantative figures are boxplots with whiskers. Significance is indicated with an asterisk, methods for calculating significance are reported in the appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Resources are described in subsection 2.2, we submitted jobs on a HPC so spefics of GPUs and CPUs varied per job. We don't mention overall compute time or compute time including experimental runs.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Data and code are both public work and used with consent.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: The paper only discusses potential positive outcomes on healthcare.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA] .

    Justification: This paper doesn't release a model or data. Both model and data are already public.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Both data and code reported with correct citations. Data accession number is given such that data can be accessed after data-paper publication. No license was available for the data, but GEO data is free to use unless otherwise stated (which it is not) and written consent was obtained from the original authors. License of the code is reported, no version was provided.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA] .

    Justification: No new assets are introduced in the paper

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA] .

    Justification: No crowdsourcing experiments or research with human subject.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA] .

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.