Conditional Normalizing Flows for the Design of T Cell Therapies

Samarth E. Kadaba [*1 2 3] **Alexander K. Eapen** [*1 2 4] **Kristin C.Y. Tsui** [1 2] **Kuan Pang** [5] **Theodore L. Roth** [4 6] **Zinaida Good** [1 2 6]

Abstract

Designing effective cell therapies requires understanding how transcriptional regulation within infusion products influence patient outcomes. Here, we propose a generative model that leverages single-cell RNA sequencing (scRNA-seq) data paired with clinical outcomes to learn the gene regulatory networks within engineered Chimeric Antigen Receptor (CAR) T cells. Using conditional normalizing flows, our model captures the high-dimensional distribution of gene activity while conditioning on patient and response-specific features. This approach enables patient response prediction with 73% accuracy and accurate simulations of gene knockdown, knockouts, and over-expression experiments. Our model identified function-recovering genetic modifications for CAR T infusion products, which were validated experimentally in the context of a genetic screen.

1. Introduction

1.1. Cell Therapy is a Complex Drug Design Problem

Chimeric Antigen Receptor (CAR) T cell therapies are engineered T cells used in cancer treatment. Their efficacy depends on complex factors, including infusion product CAR T cell features (cell state), tumor features (*e.g.* target expression), and clinical features (*e.g.* tumor burden). Understanding the precise transcriptional profile of an effective CAR T cell is difficult because of complex T cell differentiation states, engineering trade-offs between cytotoxicity and long-term persistence, and tumor antigen heterogeneity (Srivastava & Riddell, 2018). Recent work aims to enhance CAR T cell persistence by targeting relevant genes (Labib et al., 2022; Institute, 2024; Zhou et al., 2025), such as metabolic regulators (McPhedren et al., 2024), requiring an understanding of gene interactions and perturbation effects. However, identifying the determinants of a durable clinical response based on the state of CAR T cells and a rational design of CAR T cell infusion products based on primary clinical data are emerging areas of research, with no end-to-end methods available at the time of writing. Here we seek to address this gap by introducing computational methods for generalizable design of CAR T cell therapies.

1.2. Data

We use single-cell RNA sequencing data (scRNA-seq data), represented as sparse matrices of normalized read counts, to learn gene regulatory networks (GRNs) underlying CAR T efficacy. Using patient outcome data collected from a commercial therapy (Deng et al., 2020; Li et al., 2023), we train generative models to impute masked expression and estimate perturbation effects by conditioning these models on whether a patient achieves a durable response (assessed at 3 months) following CAR T cell therapy or not (i.e. no treatment effect).

1.3. Contributions

We introduce ***tcellNF***, a normalizing flow model framework trained on gene expression of CAR T cells and treatment outcomes. Our model combines established ML methods to address the challenges arising with scRNA-seq data (from CAR T cell infusion products). It (i) mitigates confounding variables by orthogonalizing them in an embedding space, (ii) maps binary outcomes to a learnable continuum of efficacy, and (iii) imputes gene expression to predict perturbation impacts. Our model demonstrates that posing therapeutic efficacy and drug design as a joint optimization problem in cell embedding spaces of CAR T transcriptome profiles can help rationally design therapies with desirable infusion product phenotypes.

[*]Equal contribution [1]Division of Immunology and Rheumatology, Department of Medicine, Stanford University, Stanford, CA [2]Center for Biomedical Informatics Research, Department of Medicine, Stanford University, Stanford, CA [3]Department of Electrical Engineering, Stanford University, Stanford, CA [4]Department of Pathology, Stanford University, Stanford, CA [5]Department of Computer Science, Stanford University, Stanford, CA [6]Parker Institute for Cancer Immunotherapy, San Francisco, CA. Correspondence to: Zinaida Good <zinaida@stanford.edu>, Theodore L. Roth <troth@stanford.edu>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

2. Methods

Because of the high-dimensional and complex probabilistic landscape of gene expression, we implement a modified RealNVP (Dinh et al., 2017) architecture to disentangle patient-specific and biological variation from scRNA-seq data. We first provide a primer on normalizing flows and then discuss adaptations of RealNVP to our specific task.

2.1. Normalizing Flow Framework

Normalizing flows are a class of generative models that learn invertible mappings between a simple base distribution (typically a normal distribution) and a complex target distribution. Let $\mathbf{x} \in \mathbb{R}^D$ represent a gene expression vector with 1152 genes that were selected based on a differential expression analysis between the two clinical response groups, and $\mathbf{z} \in \mathbb{R}^D$ represent a latent variable following a standard normal distribution $p_{\mathbf{z}}(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. The normalizing flow defines an invertible transformation $f_\theta : \mathbb{R}^D \to \mathbb{R}^D$ parameterized by θ, such that $\mathbf{z} = f_\theta(\mathbf{x})$. The log probability density of the observed data can be computed using the change of variables formula (1). where $\left|\det \frac{\partial f_\theta(\mathbf{x})}{\partial \mathbf{x}}\right|$ is the absolute value of the determinant of the Jacobian of f_θ.

$$\log p_{\mathbf{x}}(\mathbf{x}) = \log p_{\mathbf{z}}(f_\theta(\mathbf{x})) + \log \left|\det \frac{\partial f_\theta(\mathbf{x})}{\partial \mathbf{x}}\right| \quad (1)$$

2.2. Patient-Invariant RealNVP Architecture

RealNVP consists of a sequence of coupling layers where each layer transforms only a subset of the input dimensions, keeping the others fixed. Each coupling layer follows:

$$\mathbf{y}_{1:d} = \mathbf{x}_{1:d}$$
$$\mathbf{y}_{d+1:D} = \mathbf{x}_{d+1:D} \odot \exp(s(\mathbf{x}_{1:d}, \mathbf{c})) + t(\mathbf{x}_{1:d}, \mathbf{c})$$

where $\mathbf{x}$ is the input, $\mathbf{y}$ is the output, s and t are scale and translation functions implemented as neural networks which map the subsetted input, $x_{1:d}$ into an internal representation $\mathbf{h}$, $\odot$ denotes element-wise multiplication, and $\mathbf{c}$ is a condition vector. The Jacobian determinant of this transformation is

$$\det \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \prod_{i=d+1}^{D} \exp(s(\mathbf{x}_{1:d}, \mathbf{c})_i)$$

The log-determinant is simply:

$$\log \left|\det \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right| = \sum_{i=d+1}^{D} s(\mathbf{x}_{1:d}, \mathbf{c})_i$$

Successive coupling layers use alternating partitioning patterns to ensure that all dimensions can be transformed.

2.3. Disentanglement of Biological and Patient-Specific Variation

To disentangle patient-specific from biological variation, we split the latent space $\mathbf{z}$ into two components:

$$\mathbf{z} = [\mathbf{z}_{\text{bio}}; \mathbf{z}_{\text{patient}}]$$

where $\mathbf{z}_{\text{bio}} \in \mathbb{R}^{d_{\text{bio}}}$ represents biological variation and $\mathbf{z}_{\text{patient}} \in \mathbb{R}^{d_{\text{patient}}}$ represents patient-specific variation, with $d_{\text{bio}} + d_{\text{patient}} = D$. This separation is achieved through a latent splitting network:

$$[\mathbf{z}_{\text{bio}}; \mathbf{z}_{\text{patient}}] = g_\phi(\mathbf{z})$$

where g_ϕ is a neural network parameterized by ϕ.

2.4. Class Conditioning with Feature-wise Linear Modulation (FiLM)

To condition the model on patient outcome we assign each cell belonging to a given patient a binary outcome, "overall response" (OR) or "no response" (NR) . Specifically, patients achieving complete or partial response at 3 months were categorized as OR while those with stable or progressive disease were labeled NR. We use FiLM (Perez et al., 2018) conditioning in the coupling layers. For a class label $y \in \{0, 1\}$ (where 0 represents OR and 1 represents NR), the condition embedding $\mathbf{c} = h_\psi(y)$ is obtained through an embedding layer parameterized by ψ.

The FiLM conditioning modifies the internal features $\mathbf{h}$ of the scale and translation networks:

$$\mathbf{h}' = \gamma(\mathbf{c}) \odot \mathbf{h} + \beta(\mathbf{c})$$

where γ and β are learnable functions that generate scale and bias terms from the condition embedding.

2.5. Pretraining Objective

To ensure $\mathbf{z}_{\text{bio}}$ is invariant to patient-specific information while retaining class information, we employed an adversarial training approach. Two discriminator networks were used:

1. A patient discriminator $D_{\text{patient}}(\mathbf{z}_{\text{bio}})$ that predicts patient identity 2. A class discriminator $D_{\text{class}}(\mathbf{z}_{\text{bio}})$ that predicts cell state (OR or NR). The total loss function for training is given by (2).

$$\mathcal{L} = \mathcal{L}_{\text{NLL}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{class}} \mathcal{L}_{\text{class}} \quad (2)$$

where $\mathcal{L}_{\text{NLL}} = -\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}}[\log p_{\mathbf{x}}(\mathbf{x})]$ is the negative log-likelihood, $\mathcal{L}_{\text{adv}} = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}}[D_{\text{KL}}(D_{\text{patient}}(\mathbf{z}_{\text{bio}}) || \mathbf{u})]$ penal-

izes deviation from a uniform patient distribution, and $\mathcal{L}_{\text{class}} = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[-\log D_{\text{class}}(\mathbf{z}_{\text{bio}})_y]$ is the class prediction loss.

The patient discriminator minimizes the classification loss of the patient-specific embeddings with patient id. This adversarial setup encourages $\mathbf{z}_{\text{bio}}$ to contain biological information relevant to the cell state while being uninformative about patient identity.This enforces that patient-specific information is contained in an isolated portion of the embedding.

The class discriminator enables conditional guidance on $\mathbf{z}_{\text{bio}}$ by minimizing the classification loss of response outcome with a transform of the biological embedding. To motivate the various loss terms including both FiLM-based class conditioning and patient-adversarial training we conduct ablation studies on model performance with and without these components on several downstream tasks as detailed below.

2.6. Conditional Generation and Inpainting with Langevin Dynamics

We used Langevin dynamics, a gradient-based MCMC method, to inpaint missing gene expression values by iteratively refining samples using the score function (the gradient of the log-likelihood).

We chose Langevin dynamics over deterministic optimization for gene expression imputation because it samples from the posterior distribution $P(\mathbf{x}_m|\mathbf{x}_o, \mathbf{c})$, where $\mathbf{x}_m = \mathbf{x} \odot \mathbf{m}$, $\mathbf{x}_o = \mathbf{x} \odot (\mathbf{1} - \mathbf{m})$ with $\mathbf{m}$ a binary mask and $\mathbf{c}$ a class conditioning embedding. This provides uncertainty quantification which is essential for biological interpretation. Additionally, the stochastic exploration naturally handles multimodal posteriors that arise from multiple stable gene regulatory states, while respecting the complex correlation structures learned by our Real-NVP model.The update rule is:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \frac{\epsilon_t}{2}\nabla_{\mathbf{x}} \log p(\mathbf{x}^{(t)}|y) + \sqrt{\epsilon_t}\boldsymbol{\omega}_t$$

where $\mathbf{x}^{(t)}$ is the gene expression vector at step t, ϵ_t is the step size, $\boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and y is the class label.

To inpaint only missing entries defined by a binary mask $\mathbf{m}$ ($m_i = 1$ if observed, 0 if missing), we use:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + (1 - \mathbf{m}) \odot \Big(\frac{\epsilon_t}{2}\nabla_{\mathbf{x}} \log p(\mathbf{x}^{(t)}|y) + \sqrt{\epsilon_t}\boldsymbol{\omega}_t\Big) + \mathbf{m} \odot \mathbf{x}_{\text{obs}}$$

where $\mathbf{x}_{\text{obs}}$ contains the observed gene values.

To improve sample quality, we use annealed Langevin dynamics with a temperature schedule that gradually reduces noise:

$$\sqrt{\epsilon_t} = \sigma_{\text{max}} \left(1 - \frac{t}{T}\right)^2$$

where σ_{max} is the initial noise scale, t is the current step, and T is the total number of steps.

When inpainting with only the biological component, the gradient is computed as:

$$\nabla_{\mathbf{x}} \log p_{\text{bio}}(\mathbf{x}|y) = \nabla_{\mathbf{z}_{\text{bio}}} \log p(\mathbf{z}_{\text{bio}}) \frac{\partial \mathbf{z}_{\text{bio}}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

where $p(\mathbf{z}_{\text{bio}})$ is the prior over the biological latent space, and the Jacobians account for transformations from input space to latent space.

2.7. Gene Perturbation Analysis

For gene perturbation analysis, we examined how altering the expression level of a specific gene affects the likelihood of a sample under the different class conditions. Hence, $P(c|\mathbf{x})$, where c is the class-conditioning (either "OR" or "NR"), gives the posterior probability of the data $\mathbf{x}$ being observed in the conditioned flow. For a gene expression vector $\mathbf{x}$ and perturbation factor α, the perturbed expression for gene i is $\mathbf{x}'_i = \alpha \cdot \mathbf{x}_i$. The effect is quantified by the change in log-likelihood: $\Delta \log p(\text{OR}|\mathbf{x}) = \log p(\mathbf{x}'|y = 0) - \log p(\mathbf{x}|y = 0)$ and $\Delta \log p(\text{NR}|\mathbf{x}) = \log p(\mathbf{x}'|y = 1) - \log p(\mathbf{x}|y = 1)$. The overall effect of the perturbation is given by (3).

$$\text{Effect} = \mathbb{E}_{\mathbf{x}}[\Delta \log p(NR|\mathbf{x}) - \Delta \log p(OR|\mathbf{x})] \quad (3)$$

A positive Effect indicates the perturbation shifts gene expression patterns toward the NR state, while a negative Effect indicates a shift toward the OR state. The magnitude of this effect is arbitrarily classified as:

- **Strong:** $|\text{Effect}| \geq 0.5$
- **Moderate:** $0.1 \leq |\text{Effect}| < 0.5$
- **Weak:** $|\text{Effect}| < 0.1$

In the experiments discussed below, we negate this value since we care about *recovering* cell function, and hence maximizing $\log p(OR|\mathbf{x})$.

3. Experiments

In the subsequent experiments, we benchmark *tcellNF* on both masked gene expression prediction, cell-level label classification, and patient-level outcome classification. Importantly, we note that for generalization to new patients,

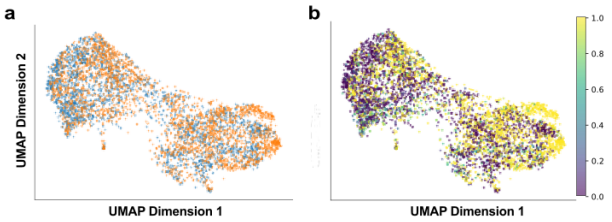

Figure 1. (a) UMAP of cells colored by original binary labels: orange for $P(\text{OR}) = 1$, blue for $P(\text{OR}) = 0$. (b) Same UMAP colored by *tcellNF* soft-labels under the OR condition using Eq. (2). Log-likelihoods from class-conditioned flows serve as proxies for patient response by averaging cell-level scores and applying a threshold. Data includes 266,152 cells from 36 patients covering 1,152 differentially expressed genes.

we use only the learned biologically-relevant embeddings ($\mathbf{z}_{\text{bio}}$) for the normalizing flow generative process as well as cell-level classification.

3.1. Soft-labeling of Clinical Outcomes

We evaluate the predictive power of *tcellNF* on resolving binary labels ("OR" or "NR") at the three-month time point. *tcellNF* successfully represents infusion-product cells on a continuum of "goodness" (Fig. 1b), and can reconstruct the original binary labels based on taking the dominant log-likelihoods assigned to each cell under each condition. (Table. 1).

Training	Class	Precision	Recall	F1	Acc.	Support
w/o Conditioning.	OR	0.000	0.000	0.000	0.457	19338
	NR	0.457	1.000	0.627	0.457	16253
w/o Adv.	OR	0.75	0.75	0.75	0.73	19338
	NR	0.70	0.69	0.70	0.73	16253
tcellNF	OR	0.83	0.85	0.84	0.82	19338
	NR	0.81	0.79	0.80	0.82	16253

Table 1. Comparison of classification metrics for distinguishing OR (0) and NR (1) on cell-level classification. Models trained with and without adversarial training are evaluated on 35,591 cells from a withheld test set of 11 patients. Without FiLM-conditioning signals, OR and NR samples are equally likely under the normalizing flow yielding 50% classification accuracy.

3.2. Masked Gene Expression Prediction

We next evaluate the regression outcomes of predicting masked gene expression. We evaluate the reverse process, via Langevin dynamics, on a set of unseen cells to impute randomly masked genes (30%). We use 10-fold cross validation to fit predicted scores to the actual scores using a linear model and report the average R^2 for all folds (Fig. 2). *tcellNF* shows the ability to generalize to unseen cells (from unseen patients) in the masked gene expression task.

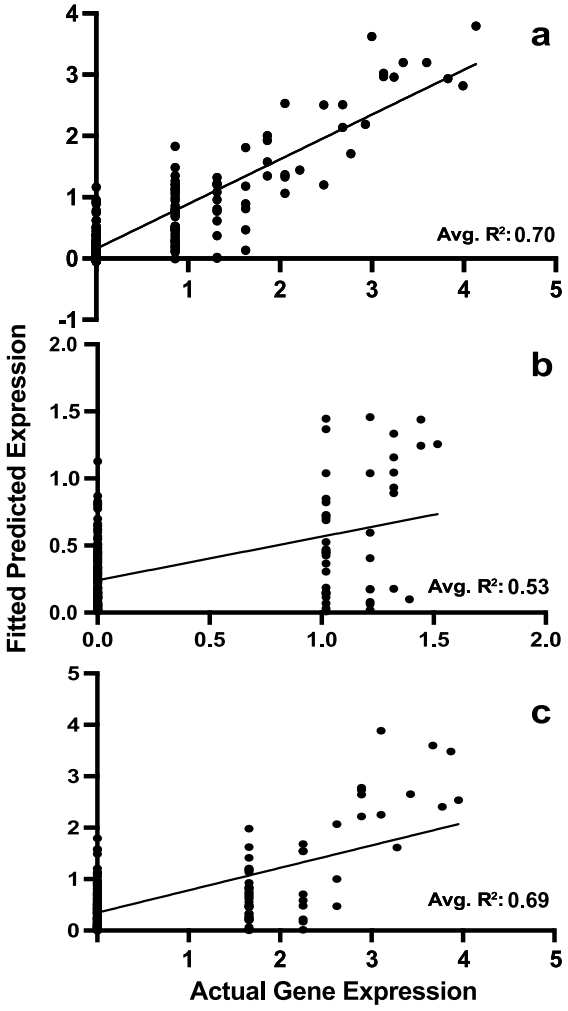

Figure 2. We plot a subset of a representative fold for clarity. a) ***tcellNF*** yielded an Average R^2 value ≈ 0.7 and average Pearson correlation $r \approx 0.84$. All gene expression values are normalized, $\log_2$ transformed. (c) Removing conditioning signals generally does not seem to decrease performance on masked gene expression prediction. By comparison, our ablations with removing adversarial loss terms (b) significantly decrease regression accuracy suggesting patient-specific features, if not correct for, are strongly prominent in learned cell embeddings. Hence, generalization to out-of-distribution (OOD) data remains poor without adjusting for these effects.

3.3. Patient Outcome Classification

We trained *tcellNF* with conditional information to capture gene expression variation linked to clinical outcomes. To evaluate patient-level performance, we averaged log-likelihoods across all cells for each patient under both OR and NR conditions, using this to compute an aggregate outcome score. To maintain a balanced outcome distribution, we fixed the test patient set. tcellNF outperformed all baselines (Table 3), showing that outcome variation can be effectively captured at the cell level with appropriate soft-labeling.

Classifier	Accuracy	F1	AUROC	AUPRC
tcellNF	**0.73**	**0.80**	**0.73**	**0.72**
LogisticRegression	0.60	0.65	0.62	0.63
KNeighborsClassifier	0.57	0.63	0.59	0.60
SVC	0.57	0.62	0.58	0.62
DecisionTreeClassifier	0.52	0.55	0.52	0.55
RandomForestClassifier	0.55	0.59	0.56	0.60
GaussianNB	0.57	0.66	0.60	0.64

Table 2. Validation metrics (Accuracy, F1, ROC AUC, AUPRC) for various classifiers on held-out test set. Bold indicates best-performing model. We train on cells from 26 patients and test on 11 unseen patients. During training we subsample the number of cells from each patient so as to avoid patient-specific batch effects. Dataset specification and cell counts for training and testing are listed in Appendix B.

3.4. Perturbation Analysis and Validation

We simulate *in silico* perturbations by multiplying a gene's expression: $0.5\times$ for knockdowns, $0.0\times$ for knockouts, and $100\times$ for overexpression. To account for GRN effects, we mask 30% of other genes and run reverse inference. Perturbation impact is measured by the shift in log-likelihoods under OR vs. NR; a shift favoring OR is considered function-recovering (Fig. 3). We validate predictions against 25 experimentally perturbed genes (Table 3). *tcellNF* predicts directional effects and nominates candidates with known CAR T relevance (Strijker et al., 2025; Wang et al., 2024; Wheeler et al., 2023; Li et al., 2013; Sun et al., 2023). Though limited in predicting effect size (Spearman $\rho = 0.24$, $p = 0.06$, Fig. 3b), predictions track experimental trends. See Appendix A for methods.

Limitations. This study has several limitations. First, we note that our validation protocol was based on a fixed subset of held-out patients, which may overestimate generalizability. Future work requires more robust evaluation strategies, such as repeated random subsampling of both patients and cells. Second, our gene set was selected based on differential expression analysis, which may have inadvertently introduced data leakage and biased classification performance.

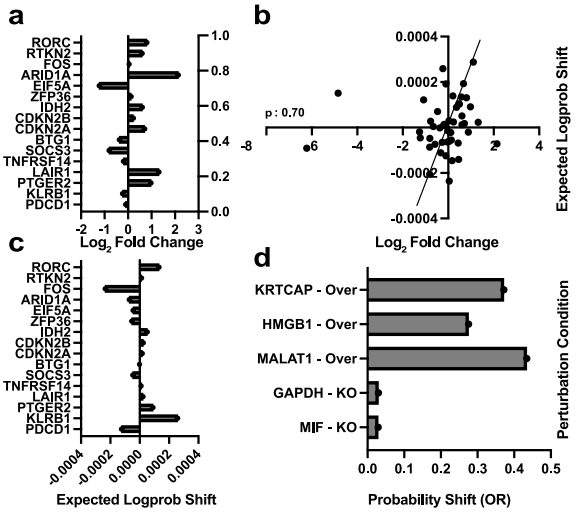

Figure 3. (a) Experimentally observed $\log_2$ fold changes from various perturbations. Here we only show knockdowns. (b) Comparison of observed $\log_2$ fold changes and computationally predicted shifts in log-likelihood. (c) Computationally predicted directional outcomes for knockdown perturbations. (d) 5 of the top nominations for perturbations to recover cell function (i.e. increase the likelihood of OR).

Perturbation Type	tcellNF	w/o Adv. (%)	w/o Conditioning (%)	Support
Knockdown	70.6	35.3	58.8	17
Knockout	47.1	52.9	52.9	17
Overexpression	66.7	41.7	58.3	12
Overall	60.9	43.5	56.5	46

Table 3. Directional accuracy (%) and support (number of perturbations) by perturbation type, comparing models trained with adversarial training, without adversarial training, and a degenerate model that always predicts OR. While the degenerate model achieves moderately high accuracy by exploiting class imbalance, it fails to capture true directionality.

Adapting the model to include more genes and particularly those relevant to T cell metabolism or function as opposed to arbitrary ribosomal, mitochondrial, and mitotic genes is likely to nominate more interesting and candidates with regard to the study of CAR T cell behavior. Finally, we also note that the dataset analyzed was relatively limited in scope and does not fully represent of the broader CAR T therapy patients. Validation across larger and more demographically diverse cohorts will be critical for assessing the robustness and translational utility of our approach.

Acknowledgements

We thank the reviewers for their feedback, the Stanford Research Computing Center for access to the Marlowe High Performance Compute resources, and Dr. Kameron Rodrigues for contributing to the ideas. This work was supported by the National Institutes of Health Office of the Director (1OT2OD038101 to C.L.M., O.G., Z.G.), the Parker Institute for Cancer Immunotherapy (C-02895 Parker Bridge Fellow to Z.G., C-04134 Kona Innovation Challenge to Z.G.). Z.G. and T.L.R. are members of the Parker Institute for Cancer Immunotherapy, which supports the Stanford University Cancer Immunotherapy Program.

Impact Statement

Interpreting clinical correlative data. The work presented here details methods for inferring CAR T design principles from low-sample clinical correlative datasets. Without requiring explicit training on masked gene expression, *tcellNF* can effectively impute gene expression values and shows strong predictive ability on characterizing directional outcomes of *in vitro* perturbations. By disentangling confounding patient-specific variables from biologically-relevant features in the learned embedding space, our framework enables soft-labeling of binary clinical outcomes and the recovery of binary label classification at the cell-level and patient-level. This work highlights the underutilized power of clinical correlative data in guiding therapeutic and experimental design. More broadly, we demonstrate that efficient, high dimensional models can be built from limited patient samples to nominate actionable perturbations for improving next-generation cell and gene therapies.

References

Deng, Q., Han, G., Puebla-Osorio, N., Ma, M. C. J., Jiang, N., Yang, H., Sun, J., Gu, S., Zhang, Q., Pan, S., Chen, G., Chen, M., Ma, J., Chen, B., Song, P., Song, R., Davis, E., Diehl, M., Wang, Y., Zhang, S., Winer, R. E., Champlin, R., Davis, R. E., Neelapu, S. S., Wang, L., and Green, M. R. Characteristics of anti-cd19 car t cell infusion products associated with efficacy and toxicity in patients with large b cell lymphomas. *Nature Medicine*, 26:1878–1887, 2020. doi: 10.1038/s41591-020-1061-7. URL https://www.nature.com/articles/s41591-020-1061-7.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. Density estimation using real nvp. In *5th International Conference on Learning Representations (ICLR)*, 2017. URL https://arxiv.org/abs/1605.08803.

Institute, N. C. Gene fusion from cancerous t cells may boost t-cell therapy. *Cancer Currents Blog*, 2024. URL https://www.cancer.gov/news-events/cancer-currents-blog/2024/gene-alteration-boosts-t-cell-therapy. Published March 20, 2024.

Labib, I., El Hajj, H., El Hajj, R., Khalife, J., El Sabban, M., Bazarbachi, A., and El Hajj, H. Advances in car-t cell genetic engineering strategies to overcome hurdles in solid tumors treatment. *Frontiers in Immunology*, 13: 830292, 2022. doi: 10.3389/fimmu.2022.830292. URL https://www.frontiersin.org/articles/10.3389/fimmu.2022.830292/full.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Li, G., Liang, X., and Lotze, M. T. Hmgb1: The central cytokine for all lymphoid cells. *Frontiers in Immunology*, 4:68, 2013. doi: 10.3389/fimmu.2013.00068.

Li, X., Henderson, J., Gordon, M. J., Sheikh, I., Nastoupil, L. J., Westin, J., Flowers, C., Ahmed, S., Wang, L., Neelapu, S. S., Strati, P., Deng, Q., and Green, M. R. A single-cell atlas of cd19 chimeric antigen receptor t cells. *Cancer Cell*, 41(11):1835–1837, 2023. doi: 10.1016/j.ccell.2023.08.015. URL https://www.sciencedirect.com/science/article/pii/S1535610823003161?via%3Dihub.

McPhedren, T. et al. Metabolic engineering for optimized car-t cell therapy. *Nature Reviews Immunology*, 2024. URL https://pubmed.ncbi.nlm.nih.gov/38388705/.

Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. URL https://arxiv.org/abs/1709.07871.

Roth, T. L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P. J., Hiatt, J. B., Saco, J., Krystofinski, P., Li, H., Tobin, V., Nguyen, D. N., Gate, D., and Marson, A. Reprogramming human t cell function and specificity with non-viral genome targeting. *Nature*, 559(7714):405–409, 2018. doi: 10.1038/s41586-018-0326-5. URL https://www.nature.com/articles/s41586-018-0326-5.

Srivastava, S. and Riddell, S. R. Car t cell therapy: Challenges to bench-to-bedside efficacy. *The Journal of Immunology*, 200(2):459–468, 2018. doi: 10.4049/jimmunol.1701155.

Strijker, J. G., Pascual-Pasto, G., Grothusen, G. P., Kalmeijer, Y. J., Kalaitsidou, E., Zhao, C., McIntyre, B., Matlaga, S., Visser, L. L., Barisa, M., Himsworth, C., Shah, R., Muller, H., Schild, L. G., Hains, P. G., Zhong, Q., Reddel, R. R., Robinson, P. J., Catena, X., Soengas, M. S., Margaritis, T., Dekker, F. J., Anderson, J., Molenaar, J. J., Bosse, K. R., Wu, W., and Wienke, J. Blocking mif secretion enhances car t-cell efficacy against neuroblastoma. *European Journal of Cancer*, 218:115263, 2025. doi: 10.1016/j.ejca.2025.115263.

Sun, P., Zhang, H., Shi, J., Xu, M., Cheng, T., Lu, B., Yang, L., Zhang, X., and Huang, J. Krtcap2 as an immunological and prognostic biomarker of hepatocellular carcinoma. *Colloids and Surfaces B: Biointerfaces*, 222: 113124, 2023. doi: 10.1016/j.colsurfb.2023.113124.

Takacsi-Nagy, O. and Satpathy, A. T. Non-viral intron knockins enable simplified and flexible targeting of endogenous genes in primary human t cells. *bioRxiv*, 2024. doi: 10.1101/2024.03.05.582227. URL https://www.biorxiv.org/content/10.1101/2024.03.05.582227v2.

Wang, X., Fu, S.-Q., Yuan, X., Yu, F., Ji, Q., Tang, H.-W., Li, R.-K., Huang, S., Huang, P.-Q., Qin, W.-T., Zuo, H., Du, C., Yao, L.-L., Li, H., Li, J., Li, D.-X., Yang, Y., Xiao, S.-Y., Tulamaiti, A., Wang, X.-F., Dai, C.-H., Zhang, X., Jiang, S.-H., Hu, L.-P., Zhang, X.-L., and Zhang, Z.-G. A gapdh serotonylation system couples cd8+ t cell glycolytic metabolism to antitumor immunity. *Molecular Cell*, 84(4):760–775.e7, 2024.

Wheeler, B. D., Gagnon, J. D., Zhu, W. S., Muñoz-Sandoval, P., Wong, S. K., Simeonov, D. R., Li, Z., Debarge, R., Spitzer, M. H., Marson, A., and Ansel, K. M. The lncrna malat1 inhibits mir-15/16 to enhance cytotoxic t cell activation and memory cell formation. *eLife*, 12:RP87900, 2023. doi: 10.7554/eLife.87900.

Zhou, Y., Wang, S., Wang, Y., Zhang, Y., and Zhang, J. Car-t cell therapy: developments, challenges and expanded applications from cancer to autoimmunity. *Frontiers in Immunology*, 15:1519671, 2025. doi: 10.3389/fimmu.2024.1519671. URL https://www.frontiersin.org/articles/10.3389/fimmu.2024.1519671/full.

A. Experimental Methods

Primary Human T cell Culture and Editing

Following an approved Institution Blood Center IRB protocol, human T cells were isolated from peripheral blood mononuclear cells (PBMCs) collected from healthy human donors. Density centrifugation and CD3-based selection were applied in the isolation procedure, using Ficoll (Lymphoprep, StemCell) within SepMate tubes (StemCell) and the Human CD3 T Cell Enrichment kit (StemCell), respectively, according to manufacturer instructions.

The isolated T cells were counted (Countess, Thermo) and activated with anti-human CD3/CD28 Dynabeads (Cell Therapy Systems, Thermo) at a 1:1 ratio, in XVivo 15 media (Lonza) supplemented with 5% fetal bovine serum (FBS; MilliporeSigma) and 50 U mL^{-1} of human IL-2 (Peprotech). Forty-eight hours after activation, T cells were de-beaded (EasySep Magnet, StemCell), pelleted, and resuspended in Lonza P3 Buffer at a density of 2×10^6 cells / 20 μL buffer.

A total of 50×10^6 bulk T cells from two donors were each electroporated (Gen2 Lonza 4D electroporation system, Lonza) on day 0 with 0.5 μg of 1 μg μL^{-1} Midiprep-generated (Zymo) HDRT plasmid DNA library and Cas12a-encoding mRNA. Edited cells were recovered in XVivo 15 media and incubated for 15 minutes at 37 °C before being transferred to tissue culture flasks.

In Vitro Toledo DLBCL Challenge Assay

CD19-expressing Toledo Diffuse Large B-cell Lymphoma (DLBCL) cells, generously donated by the Crystal Mackall group, were expanded and cultured in RPMI supplemented with 10% FBS, HEPES, and 100× Pen/Strep, and passaged every 2–3 days to maintain a density of approximately 1.5×10^6 cells per mL.

An in vitro assay, as described in (Takacsi-Nagy & Satpathy, 2024), was performed in which edited T cells were chronically stimulated over the course of 14 days via repetitive co-culture with target cells, simulating chronic antigen challenge within tumors. The chronic stimulation assay was conducted between days 6 and 20, with re-seeding at an effector-to-target ratio of 1:8 for each donor, based on pilot screen results showing efficient clearing of $CD19^+$ Toledo cells at lower ratios.

Pooled Screening

Pooled knockin screens were conducted via a non-viral electroporation editing system first described in (Roth et al., 2018). Primary human T cells were transduced with 652 unique CAR constructs from the CRISPR-All Cell Therapy Universal Screening (CACTUS) library. CACTUS comprises a pooled library of approximately 130 overexpression

targets, 230 knockdown targets, and 230 knockout targets, identified from literature on effector function-enhancing genetic modifications in both native and engineered human T cells.

Approximately 37 additional constructs were added to the original 652-member library based on clinical profiling of responders and non-responders in CD19-directed CAR clinical trials. Gene sequences were sourced from the Ensembl genome database and codon-optimized. These were synthesized into a high-copy, kanamycin-resistant cloning vector of pMB1 origin (Twist Bioscience).

Genomic DNA was isolated from the expanded T cell populations at both the start (D6) and end (D21) of the assay. Next-generation sequencing (NGS) was used to compare barcode readouts before and after the chronic stimulation challenge.

B. Data

B.1. Splits and Cell Counts

Below we list the train:test split used to evaluate all models. We randomly crafted a split that would yield an approximately equal balance of "OR" cells and "NR" cells. Because we downsampled patients to avoid overfitting to those patients with significantly more cells, our final training data ended up using much fewer cells that was actually available in the dataset. It is interesting to note that even in this low-sample paradigm, *tcellNF* performs well on classification and regression tasks.

Patient ID	Split	Cell Count	3-Month Response
ac05	Test	2090	OR
ac09	Test	2090	OR
ac10	Test	2090	OR
ac11	Test	2090	NR
ac17	Test	2090	NR
ac18	Test	2090	NR
ac22	Test	2090	NR
ac24	Test	2090	NR
ac28	Test	2090	OR
ac37	Test	2090	OR
ac39	Test	2090	OR
ac01	Train	3574	OR
ac02	Train	3574	NR
ac03	Train	3574	NR
ac04	Train	3574	NR
ac07	Train	3574	OR
ac08	Train	3574	OR
ac12	Train	3574	OR
ac13	Train	3574	NR
ac14	Train	3574	OR
ac15	Train	3574	NR
ac16	Train	3574	OR
ac19	Train	3574	NR
ac20	Train	3574	OR
ac21	Train	3574	NR
ac23	Train	3574	NR
ac25	Train	3574	NR
ac26	Train	3574	NR
ac27	Train	3574	OR
ac29	Train	3574	OR
ac30	Train	3574	OR
ac32	Train	3574	OR
ac33	Train	3574	OR
ac34	Train	3574	OR
ac38	Train	3574	NR

Table 4. 3-Month patient responses across training and test sets. Train set contains 39,314 NR cells and 46,462 OR cells. Test set contains 10,450 NR cells and 12,540 OR cells.