AI-driven single-cell data analysis identifies a cell signature predictive of neurotoxicity and clinical response in CAR-T cell therapy of DLBCL

Adoptive cell therapies represent a groundbreaking approach in cancer treatment, leveraging endogenous tumor-infiltrating lymphocytes (TIL) or genetically modified T cells to target specific antigens expressed by cancer cells. Among these therapies, chimeric antigen receptor T cell (CAR-T cell) therapy has emerged as a prominent strategy for combating hematologic malignancies.[1]

CAR-T cells are engineered to express chimeric antigen receptors (CARs) that enable them to recognize and eliminate cancer cells expressing specific antigens, such as CD19 and BCMA. This approach has shown remarkable success in inducing prolonged remissions in about half of patients with B cell malignancies, albeit with associated adverse events such as cytokine release syndrome (CRS) and immune effector cell-associated neurotoxicity syndrome (ICANS). While CRS is often low grade in adult lymphoma patients, high-grade ICANS (grades 3-4) occurs in 30-45% of the patients.[2,3]

Understanding the mechanisms underlying treatment response and toxicities is crucial for optimizing the therapeutic potential of CAR-T cell therapy. Single-cell genomics, a rapidly advancing technology, offers a comprehensive view of cellular diversity and heterogeneity, thereby facilitating the elucidation of disease mechanisms, identification of therapeutic targets, and prediction of therapy outcomes and toxicities.

In a foundational study by Deng et al. (2020)[4], single-cell transcriptomic analysis of CAR-T cell infusion products in patients with diffuse large B cell lymphoma (DLBCL) revealed distinct cellular signatures associated with treatment response and the development of ICANS. Memory T cell enrichment was observed in responders, while a novel ICANS-associated cell population (IACs) was identified in patients with high-grade neurotoxicities.

Building upon these findings, our study harnesses the power of single-cell RNA sequencing data from Deng et al. to train our proprietary AI platform, ScaiVision. By integrating scRNA-seq data with clinically relevant endpoints, our platform aims to find cellular and molecular features that predict treatment response and neurotoxicity development in CAR-T cell therapy recipients with high accuracy. Such biosignatures could serve the basis of predictive clinical biomarker assays and enable precision medicine for CAR-T-treated patients.

Data

Publicly available scRNA-seq data was sourced from Deng et al. The dataset comprised of 137,326 cells extracted from 32 patients, including 16 cases of diffuse large B cell lymphoma (DLBCL), 6 instances of transformed follicular lymphoma (tFL), and 2 occurrences of primary mediastinal B cell lymphoma (PMBCL). These cells were isolated from standard-of-care axi-cel product bags subsequent to the infusion of CAR T cells. At the 3-month follow-up, disease progression (PD) was observed in 13 patients (50%), with one patient showing partial response (PR) (4%) and nine patients achieving complete response (CR) (38%). Additionally, one patient's response was deemed not evaluable (NE). Notably, no patient experienced severe symptoms of cytokine release syndrome (CRS). However, 12 patients developed low-grade immune effector cell-associated neurotoxicity syndrome (ICANS) categorized as grade 1-2, while another 12 patients experienced high-grade ICANS categorized as grade 3-4. (Figure 1)

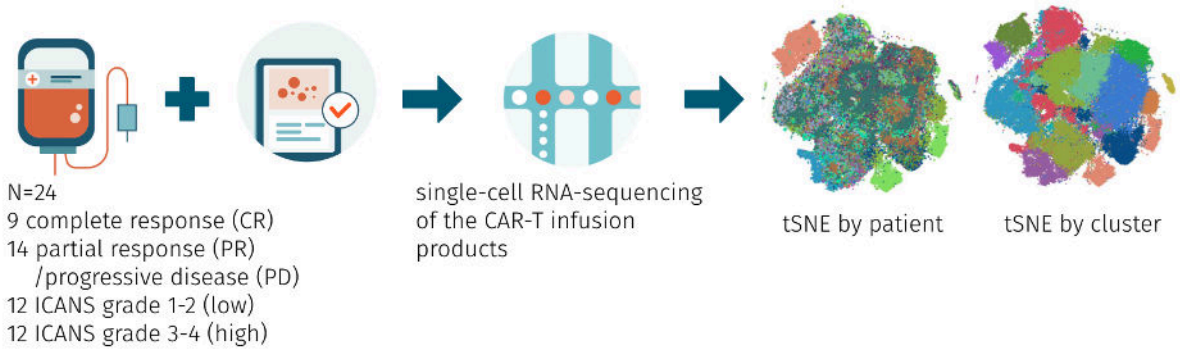

Figure 1. A schematic overview of the data by Deng et al.

Results: ScaiVision uncovers predictors of clinical response

Due to the considerable variability observed in the treatment response efficacy of CAR-T therapy among patients, we utilized our proprietary AI platform, ScaiVision, to identify defining features of treatment endpoint. By employing this innovative approach, we aimed to unravel the key biological factors influencing treatment efficacy, ultimately enhancing our understanding and predictive capabilities of CAR-T therapies.

In the first instance, we split the data into 4 cross-validation splits of 75% for training and 25% for testing of the model performance in several iterations to find the best and most robust model predicting therapy response (Figure 2A). We subsequently applied ScaiVision to each training cohort and extracted trained models from each splt. We selected the best performing models for downstream analysis, characterized by their superior predictive capabilities above a threshold of AUC of 0.8, in predicting CAR-T cell therapy response in the validation cohorts. These models are informed by different filters highlighting different cell populations and will need further analysis to consolidate the cell signatures of the best performing models and validate them on an independent dataset.

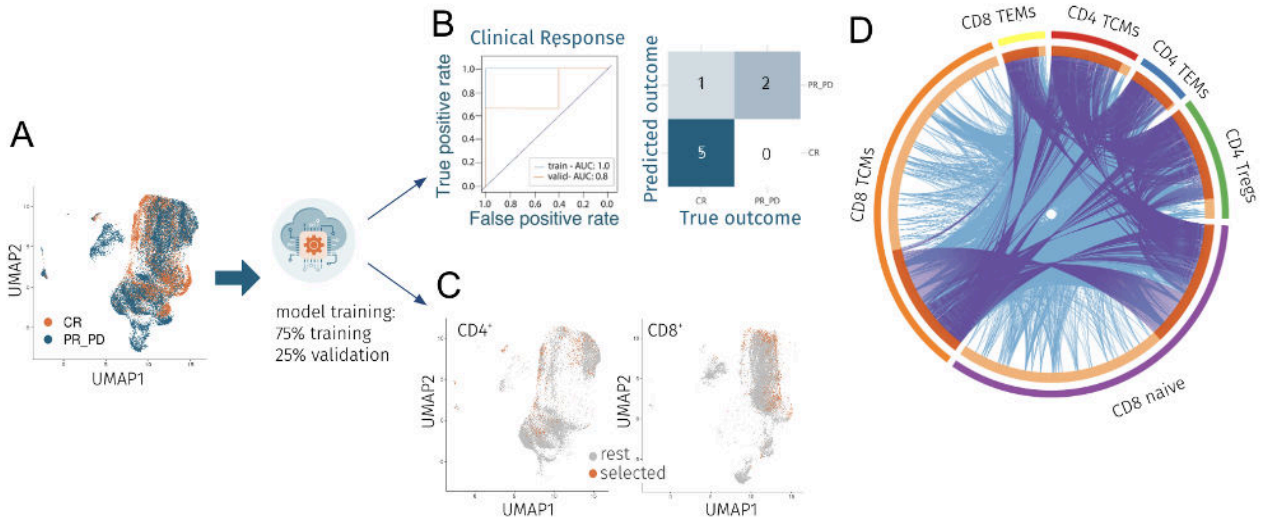

Figure 2. A) UMAP Projection of scRNA-seq labeled for treatment endpoint B) Representative model showcasing predictive capability of train AUC: 1.0 and Validation AUC: 0.8, and prediction results on validation samples C) UMAP projection showcasing CD4+ and CD8+ cells identified by ScaiVison D) Differentially-expressed genes in predictive cells: purple lines indicate common genes between cell types, blue lines indicate common pathway annotations.

In this report, for subsequent analysis we focused on a top performing model (Train AUC=1 Valid=0.8) (Figure 2B). Due to the shallowness of the neural network, we were able to extract a signature defying this model which finds an enrichment of memory markers such as CCR7 and SELL in CD8+ T cells and DDIT4, TPI1, MKI67 TK1, TYMS and STMN1 in selected CD4 T cells. The ScaiVision model also finds a depletion of inflammatory markers such as CCL3, CCL4 and CCL5, effector molecules such as GNLY and GZMH, IFNG, and MHC II genes in predictive CD8 and CD4 T cells (Figure 2C and D). These signatures are consistent with the published findings of an enrichment of memory T cell clusters and a depletion of exhausted CD8+ and CD4+ T cell clusters in CR vs PD/PR patients, respectively.

Results: ScaiVision reveals a highly accurate predictive signature of high-grade ICANS

To address the considerable variability in treatment-associated toxicity observed among patients receiving CAR-T therapy, we once again employed our proprietary AI platform, ScaiVision, to identify defining features of toxicity endpoints. As previously described, we split the same data into several CV splits of 18 samples to train ScaiVision on cell signatures that correspond to patients developing high-grade (n=9) vs low-grade (n=9) ICANS and evaluated the model performance on the leftover 6 samples to find the best and most robust models predicting high-grade toxicities (Figure 3A). ScaiVision found multiple models that were able to predict the development of high-grade ICANS across different CV splits with AUC = 1 both in the training as well as the validation dataset. (Figure 3B)

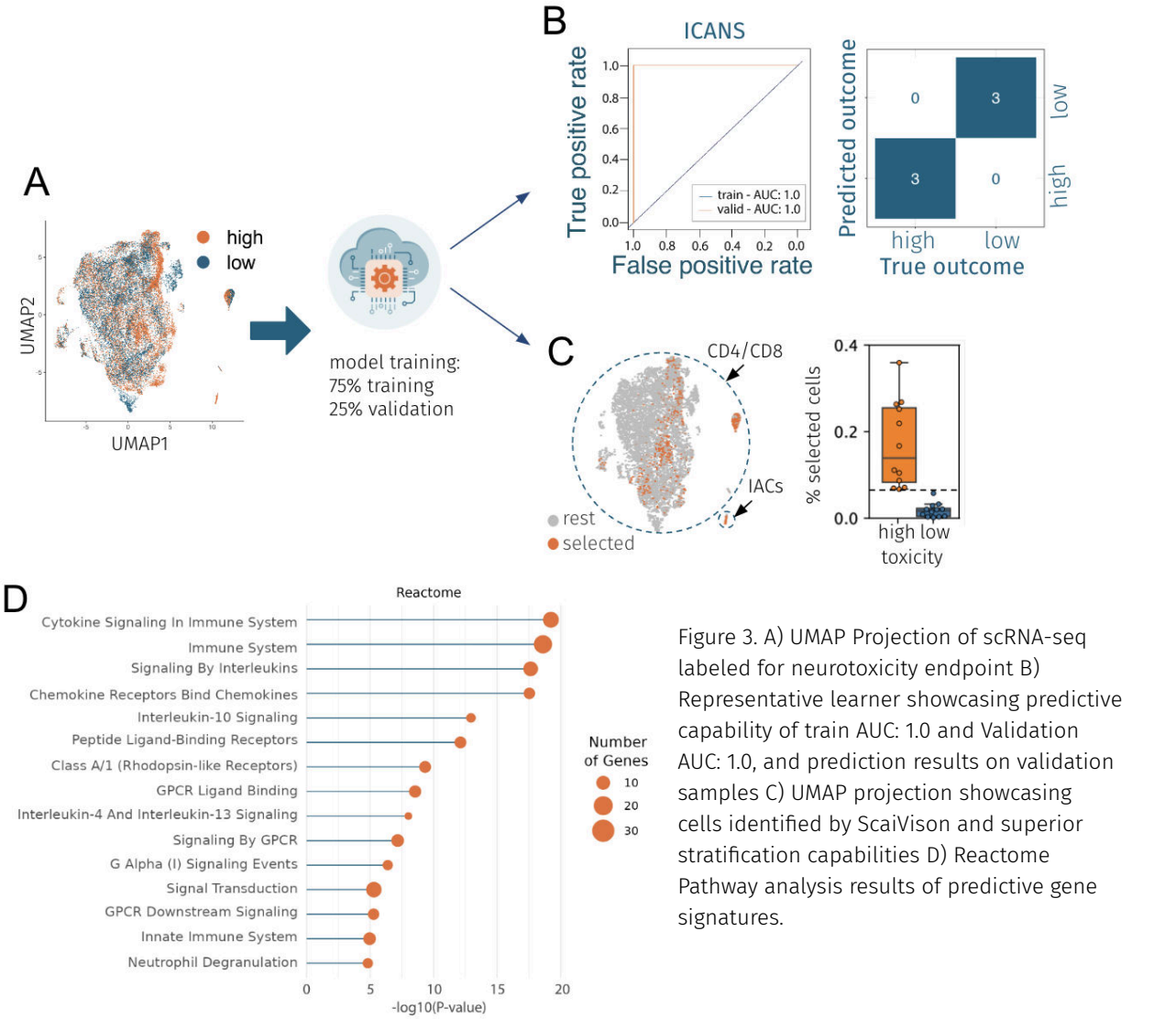

Figure 3. A) UMAP Projection of scRNA-seq labeled for neurotoxicity endpoint B) Representative learner showcasing predictive capability of train AUC: 1.0 and Validation AUC: 1.0, and prediction results on validation samples C) UMAP projection showcasing cells identified by ScaiVison and superior stratification capabilities D) Reactome Pathway analysis results of predictive gene signatures.

These models highlighted a distinct cell population of ICANS-associated cells (IACs) which was described in the original publication characterized both with T and myeloid cell lineage markers. However, while Deng et al could not find any significant DEG within the CD8+ and CD4+ T cells, the predictive models from ScaiVision included additional CD8+ and CD4+ cells which together with the IACs could predict the development of high-grade ICANS with 100% accuracy. (Figure 3C)

We next conducted pathway analysis to gain an insight into the underlying biological processes associated with our predictive signature. Among the genes predictive of CAR-T therapy-induced neurotoxicity, we identified enriched pathways through Reactome analysis. Notably, cytokine signaling in the immune system, interleukin signaling, and GPCR signaling pathways were significantly associated with our predictive signature. This discovery suggests the potential involvement of these pathways in the molecular mechanisms underlying CAR-T therapy-induced neurotoxicity and further study is required to gain further insights into neurotoxicity mechanisms and to aid in identifying potential therapeutic targets. (Figure 3D)

Conclusions

Our analysis confirms that the diversity in cellular and molecular characteristics of infused CAR-T cell products significantly impacts treatment effectiveness and side effects in lymphoma patients undergoing CD19-targeting CAR-T therapy. Looking ahead, single-cell genomic technologies are poised to become indispensable in both CAR-T treatment research and clinical decision-making. Integrating these technologies into clinical processes will revolutionize drug development, improve trial outcomes, and enhance patient care through routine CAR-T cell monitoring. As cell engineering progresses, single-cell methods will offer deeper insights into CAR-T dynamics, cellular interactions, and molecular pathways. However, overcoming limitations is crucial to integrating single-cell genomics into standard treatment protocols, necessitating adaptation for clinical settings and improving assay stability, sensitivity, robustness, and affordability. ScaiVision provides a unique AI-driven approach to single-cell data, enabling the training of explainable AI models that extract maximum information and lead to the discovery of clinical biomarker assays based on cellular and molecular features.

References

1. Barboy, O., et al. In Synergy: Optimizing CAR T Development and Personalizing Patient Care Using Single-Cell Technologies. Cancer Discov. (2023)
2. Lee, D. W. et al. ASTCT consensus grading for cytokine release syndrome and neurologic toxicity associated with immune effector cells. Biol. Blood Marrow Transplant. 25, 625–638 (2019).
3. Neelapu, S. S. et al. Chimeric antigen receptor T-cell therapy - assessment and management of toxicities. Nat. Rev. Clin. Oncol. 15, 47–62 (2018).
4. Deng, Q. et al. Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. Nat Med. 1878-1887 (2020)

Method

The standard approach to the analysis of scRNA-seq data would rely on standard preprocessing steps followed by the definition of clusters of different cell populations that share common characteristics before starting to compare the differentially expressed genes (DEG) between the corresponding clusters of complete responders vs non-responders. This approach is inspired by the analysis of bulk RNA-seq data, requires the definition of a biased number of clusters and does not make use of the richness of the single-cell resolution. Additionally, by comparing selected clusters, we might miss the discovery of complex signatures that are composed of multiple cell clusters.

The development of new technologies and the generation of new data types requires the development of new analytical tools. Our in-house developed software platform, ScaiVision, includes a rather shallow convolutional neural network which uses supervised representation learning to identify prediction-relevant features from single-cell data. The platform can be trained on different single-cell data types and on a variety of endpoint types (numeric vs text, or quality metrics vs outcomes). Through the training process, ScaiVision identifies signals that are common across individuals but distinct to a certain disease state or condition. Thanks to its shallow architecture, the results of a ScaiVision analysis consist of both a trained model capable of predicting the endpoint for new, unknown samples, as well as the molecular characterization of the most relevant cells associated with that endpoint. ScaiVision approaches the data in a cluster-free manner keeping the single-cell resolution throughout the analysis. This allows it to extract the maximum information from very complex single-cell data and is very sensitive to rare cell populations or complex signatures composed of different cell populations.