Analysis of patient responses to CAR T therapy using single-cell multiomics profiling

Je-Won Im, Dr. Zhiliang Bai, Dr. Rong Fan

Department of Biomedical Engineering, Yale School of Engineering & Applied Science, New Haven, CT

Introduction

Adoptive T cell transfer (ACT) is the use of lymphocytes to combat diseases. One emerging type of ACT developed for cancer therapy is chimeric antigen receptor (CAR) T cells, which uses the substitution of a single-chain variable fragment (scFv) with a TCRς chain from the regular double-chain T cell receptors (TCRs). While TCRs have an antigen-specific response based on their variable antigen-binding site, the scFv of all CAR T cells target the same set of tumor-specific proteins, and the single chain facilitates genetically engineering and modifications. CAR T therapy is effective because it does not target tumors through the major histocompatibility complex (MHC) of cells, unlike other ACT treatments, reducing off-tumor targeting [1]. Once activated, CAR T cells proliferate and activate non-CAR T cells for tumor responses, which overcomes the autoimmunity challenges of cancer treatment. Targeting the CD19 surface molecule on tumor cells has been effective in treating B cell malignancies, especially B-cell acute lymphoblastic leukemia (B-ALL).

While the recent FDA commercial approval of CAR T cell therapy shows successes and efficacy in treating cancer, several limitations still pose problems in treating a greater number of patients successfully. There are four main types of issues: first, CAR T cells may not proliferate properly once transferred back into the patient's body, resulting in short-term or no response. Second, tumors show adaptive resistance such as CD19 loss, which is present in up to 28% of acute leukemia cases [1]. Third, toxicities arise from immunotherapy because of the number of immune cells activated, which can result in severe cytokine release syndrome (CRS) or neurotoxicity [1]. Last, CAR T therapy is currently limited to targeting cancers of the B cell lineage and cannot target solid tumors, even though new cellular engineering with adding co-stimulatory molecules and secondary target molecules has widened the possibilities [2]. These along with a number of different factors result in various patient responses, ranging from no response, relapse, and complete remission. CAR T cell therapy, once fully developed, potentially can provide a cheaper alternative to high-cost therapies and even introduce an off-the-shelf treatment for cancer, as scFv engineering stays constant between different T cells [2]. In this research, we analyzed differentially expressed genes (DEGs) between various patient CAR T cell samples and T cell subtypes within these samples in order to find markers of complete remission and no response patients.

Methods

Patient samples

Samples of pediatric patients with resistant or refractory B-ALL were collected from two separate trials at the University of Pennsylvania and the Children's Hospital of Pennsylvania. One set of samples were collected from trials aimed to determine the safety and feasibility of CART-19 cell therapy and the proliferation and duration in vivo (ClinicalTrials.gov, NCT01626495). Another set of samples were obtained from a pilot trial on the effect of tocilizumab on the risk of CART-19 associated cytokine release syndrome (ClinicalTrials.gov, NCT02906371). All laboratory operations involving the collection and analysis of samples followed the International Conference on Harmonization Guidelines for Good Clinical Practice, and ethical standards were ensured.

Single-cell multiomics profiling

Protocols follow those described in Bai *et al.* 2022 [3]. Patient T cell samples were transduced using a lentiviral vector containing a CD19-specific CAR with 4-1BB/CD3 transgene. An *in vitro* coculture assay was created through the NIH3T3 mouse fibroblast line. The CD19 environment was NIH3T3 cells transduced with human CD19 (CD19-3T3), and the negative control expressed mesothelin (MSLN-3T3). CTL019 cells were cocultured in the CD19-3T3 and MSLN-3T3 cells to produce stimulated and control samples. CAR+ cells were stained for CITE-seq, measuring 17 types of cell surface protein levels through antibody-derived tags (ADTs), then prepared for single-cell RNA sequencing (scRNA-seq). A unique molecular identifier (UMI) count matrix filtered mouse cells and low-quality mRNAs from the sequenced mRNA data.

Data analysis

The Seurat v4.0 software was used to perform integration, quality control (QC), and differential expression analysis of the multimodal data. The integration of different samples followed the standard Seurat workflow using anchor points and canonical correlation analysis (CCA) [4]. To reduce the memory usage, reverse principal component analysis (rPCA) was done beforehand, and the integration using the SCTransform function involved the first 30 rPCA dimensions. RNA expression was log normalized, and the ADT assay was normalized through centered log transformation. Clustered data was visualized in two dimensions using UMAP. Top markers for each cluster were found using the FindMarkers function. The ShinyGO 0.76 gene enrichment analysis tool was used in determining gene ontology (GO) terms and enriched pathways.

High Performance Computing Clusters

The Yale University high performance computing (HPC) clusters were used to accommodate for the high memory usage of the integration and analysis. The general partition from the Ruddle cluster of the Yale Center for Genome Analysis was used to store patient data and run analysis through interactive RStudio Server. With memory reduction algorithms, the environment required approximately 100 to 115 gigabytes along with a single node and core.

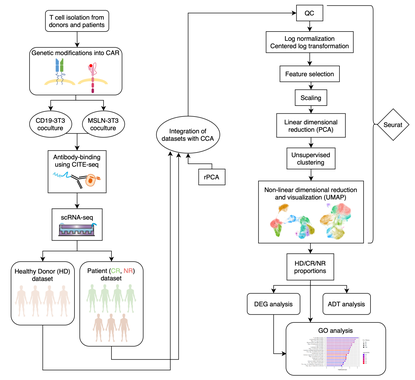

Figure 1: Schematic workflow of data collection and analysis.
Left shows the collection of T cells, CAR engineering and co-culture with 3T3 cells to simulate an *in vitro* tumor environment, and the CITE-seq and scRNA-seq producing multimodal data. HD, CR, and NR samples were integrated through CCA, with reverse PCA (rPCA) reducing memory usage.
Right shows the standard workflow of Seurat, including QC, normalization, feature selection, scaling, PCA, clustering, and visualization. Clustered data was then analyzed with its HD/CR/NR proportions through differentially expressed gene (DEG) and cell surface protein expression analysis. DEGs were then used for GO analysis through ShinyGO.

Results

The data analysis was performed in three phases: an analysis of the 4 healthy donor samples, the 7 patient samples, and then one of a fully integrated dataset of both healthy donors and patients. Healthy donor (HD) data showed that stimulated cells had higher cytotoxic activity, such as the marker GZMB, and higher helper activities, especially with T cell type 1 helper cell markers IFNG and IL13. Basal state cells had higher regulatory activities, especially with the feature TGFB1. Chemokine activity varied between stimulated and basal states, with CCL3, XCL1, and XCL2 levels higher on stimulated cells and CCL5 higher on basal state cells.

To analyze the markers distinguishing complete remission (CR) and no response (NR) patients, HD data was integrated into the patient dataset. It was assumed that HD features showed similar expression to CR patients, thus making the sample size larger. The UMAP projection showed less divergent clustering than the HD UMAP, with 13 clusters total (Figure 2a). Analysis of the proportions of CR, HD, and NR samples in each clusters showed that cluster 3, with the lowest NR percentage, and 12, with the highest, corresponded to the most divergent clusters in the visualization (Figure 2b).

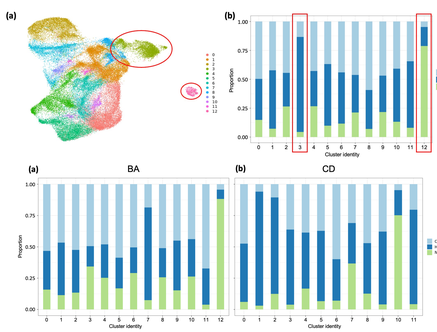

Figure 2: UMAP of integrated data and and proportions of CR/HD/NR cells of clusters.
(a) 13 (0-12) clusters of the fully integrated dataset (rPCA dim=50, PCA dim=30). Divergent clusters 3 and 12 are circled in red.
(b) Proportions of the CR (light blue), HD (navy), and NR (green) cells within the unsupervised clusters. Clusters 3 and 12 from (a) are boxed in red.

Figure 3: Proportion of patient/donor type in clustering of BA and CD subsets (res = 0.4).
(a) Percentages for basal state cells (BA), 13 clusters total (0-12).
(b) Percentages for stimulated cells (CD), 12 clusters total (0-11).

Cell surface protein analysis was performed on the overall profiles of CR, HD, and NR samples, as protein expression level differences between UMAP clusters were insignificant. The Fas apoptotic protein (CD95) and the apoptotic regulator CD28 were expressed higher in NR cells (Figure 4). The CD62L protein, found to be a regulation marker in the HD dataset, was expressed higher on NR cells, while two activation markers, HLA-DR and CD69, were higher in CR and HD cells.

One key regulatory gene, TGFB1, which produces the protein transforming growth factor beta-1, recurringly appeared in markers of basal state as well as NR cells. Overall comparison of the patient/donor type showed that expression was markedly higher in NR cells (Figure 5a). The expression levels in the unsupervised clusters of the overall integrated dataset showed little expression in cluster 3, which was also the cluster with the highest CR/HD proportion (Figure 5b,2b).

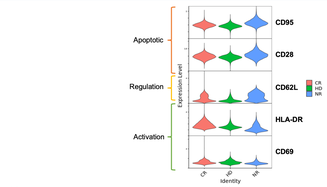

Figure 4: Violin plot of key cell surface proteins differentially expressed between total CR/HD and NR data.
CR samples (red) and HD samples (green) are compared against NR samples (NR). Top two: apoptotic protein CD95 (Fas), CD28. Third: regulator protein CD62L. Last two: activation proteins HLA-DR and CD69.

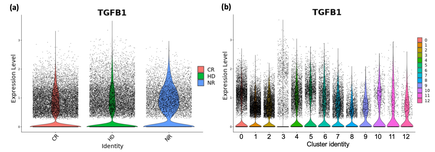

Figure 5: Violin plot of TGFB1 expression.
(a) TGBF1 expression in total CR, HD, NR data.
(b) TGBF1 expression in unsupervised 13 clustering from complete integrated data. Cluster 3 showed a higher proportion of non-NR samples in previous analysis.

Because HD analysis showed that T cells diverged greatly depending on whether they were stimulated, we further separated and clustered the two stimulation states. Basal state cells clustered into 13 clusters and stimulated cells into 12 at the same UMAP resolution (Figure 3). To find CR and NR markers, clusters from these subsets with a high percentage of CR/HD or NR cells were run through GO analysis (Table 1).

Table 1: Key GO pathways in notable basal state (BA) and stimulated (CD) clusters.
Rows highlighted in green show clusters with high CR/HD proportions, and orange with high NR proportions. 2 clusters from BA and 7 clusters from CD showed the highest CR/HD percentages, while 1 cluster from each stimulation type had a high NR percentage.

BA cluster ID#	Key Pathways
7	T cell/leukocyte/lymphocyte activation, cell cycle activities
11	Cell migration, locomotion, differentiation and development
12	Protein targeting, localization, ER-related activities
CD cluster ID#	**Key Pathways**
0	Lymphocyte activation, cytokine regulation, neg. regulation of apoptosis
1	Mitosis, cell division, cell cycle
3	Lymphocyte/leukocyte activation, cell-cell adhesion
5	Cytokine production, proliferation
6	CD4+ T cell activation, cell-cell adhesion
9	Cell-cell adhesion, response to cytokines
10	Protein targeting, localization, ER-related activities, mRNA catabolic production
11	Neg. regulation of transcription, cell cycle

Discussion

We show through DEG and cell protein analysis that complete remission patient CAR T cells performed a greater efficacy of CAR T cell function such as immune cell activation and cytokine production. CR cells also showed greater cell migration, proliferation, and a negative regulation of apoptosis, which would allow a larger response against B-ALL. On the other hand, no response patient cells showed less activation under CD19 stimulation, expressing less genes involved in immune cell activation, being more susceptible to apoptosis limiting proliferation, and greater focus on biological activities of somatic cells rather than CAR T-specific functions. A set of NR cells did not change their pathways much at all between basal and stimulated cells, meaning that overall, NR cells are unresponsive, underactivated, and understimulated. Future work will involve performing similar analysis on relapse patients and find differences to no response patients and testing the direct effect of specific genes enhancing or deprecating CAR T function through gene insertion or knockouts. This project will contribute to guiding the further genetic engineering improvements to CAR T cells and therapeutic response.

Acknowledgements

I would like to thank Dr. Bai and Dr. Fan for the incredible opportunity of working at the lab and conducting research over the summer, as well as other members of the Fan Lab. I would also like to thank Dr. Gell, Lena, the SRP cohort, and family and friends for support me throughout this journey.

References

[1] June CH, O'Connor RS, Kawalekar OU, Ghassemi S, Milone MC. 2018. CAR T cell immunotherapy for human cancer. Science. 359(6382):1361–1365. doi:10.1126/science.aar6711.
[2] June CH, Sadelain M. 2018. Chimeric Antigen Receptor Therapy. N Engl J Med. 379(1):64–73. doi:10.1056/NEJMra1706169.
[3] Bai Z, Woodhouse S, Zhao Z, Arya R, Govek K, Kim D, Lundh S, Baysoy A, Sun H, Deng Y, et al. 2022. Single-cell antigen-specific landscape of CAR T infusion product identifies determinants of CD19-positive relapse in patients with ALL. Sci Adv. 8(23):eabj2820. doi:10.1126/sciadv.abj2820.
[4] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. Cell. 177(7):1888-1902.e21. doi:10.1016/j.cell.2019.05.031.