The tumor microenvironment (TME) is comprised of the cells and extracellular matrix (ECM) that surround the tumor to promote growth and metastasis1, including immune cells such as CD8+ T cells. T cells induce adaptive immune responses in tissue through cytokines and effector molecules. CD8+ T cells acquire cytotoxic functions that attack tumor and infected cells. In tumors, the state and subtype of CD8+ T cells has been shown to have an effect on immunotherapy response, such as immune checkpoint blockade (ICB)2,3, chimeric antigen receptor(CAR) T cell therapy4. These therapies have become increasingly popular targets since they do not target the tumor cells directly and have been shown to lead to long term clinical responses in a subset of patients. One main area of interest is determining what makes these patients good candidates to try and increase patient response as well as develop better prediction for response type. Specifically, large populations of exhausted, also known as dysfunctional, T cells no longer have anti-tumor properties and are thought to be linked to poor immunotherapy response5–10. The progression from naive to exhausted T cells is highly regulated by transcription factors and epigenetic regulators, suggesting that scRNA seq could reveal underlying phenotype composition and progression of T cells along the pathway to exhaustion. Exhausted T cells are categorized by their distinct genetic signatures that result in overexpression of inhibitory receptors and dysregulated signaling pathways11–15. The composition and state of T cells in the TME could influence the effect that immunotherapy treatments have on the tumor cells.

Previous studies on T cell state and TME composition try to subtype the T cells present in one type of tumor or tissue6,9, reestablishing T cell subtypes and states specific to their tumor type of interest. Recent studies have aimed to create a pan-cancer map of T cell heterogeneity and dynamics in the TME in order to establish a generalized baseline for T cell studies across cancer types9,16; however, there is currently no easy way to compare the map they created to other datasets and studies. The creation of a new method aimed at easily compare and query the state of T cells in a new sample could better help researchers understand the overall state of T cell populations and their relationship to immunotherapy treatment response. I hypothesize that a distance measurement that defines the “closeness” of a T cell population of interest to established T cell states could help inform tumor response in the TME to immunotherapy treatments. I will address this hypothesis using the following specific aims:

Specific Aim 1: To give meaningful information about T cell subtype state by creating a spherical mapping system and distance metric based on established CD8+ T cell subtypes

Specific Aim 2: To confirm that T cell subtypes correlate to immunotherapy response by performing differential gene analysis on an unrelated immunotherapy dataset

Significance:

The presence and state of T cells in tumors has long been correlated with patient prognosis and is more recently a common target for therapies. However, little is known about the tumor immune environment and the corresponding mechanisms that affect prognosis and reaction to therapies. Recent studies further looking into these underlying mechanisms often focus on a specific therapy and tumor type. This limited focus can only reveal part of the greater biological picture. It is largely unknown how these results tie into other tumor types and T cell responses. Without being able to connect T cell subtypes across multiple cancers and tissues much can be left out. In addition, this could lead to the replication of previous work and reveal similar conclusions without tying together the bigger picture of T cell regulatory influence to tumor response.

Since recent studies have established important genes in the exhaustion phenotype, it becomes critical to connect T cell state across tissue and cancer type. It is important for researchers to be able to easily determine how their T cell populations of interest relate to previous work. This understanding of where T cell state lies in relation to defined phenotypes could lay the groundwork for more interesting and informed biological questions and results.

One such area implicated in differential T cell states is response to therapies, specifically ones that are immune targeted. It is well known that patients can have differing responses to immunological treatments; however, the mechanisms for these differences are not well categorized. Local T cell response to these treatments is one area lacking in mechanistic understanding. Studies of the TME are still developing and provide a possible explanation for the differing response to these therapeutic agents despite similarity in tumor classification.

Many recent studies have focused on the genetic differences of T cell populations in cancer patients related to their treatment response. These studies often focus on the main differences in their cohort and define new T cell states and subtypes. This new definition leaves much to be desired in the way of comparisons to existing defined states. One such paper has aimed to define the T cell states in hopes of connecting tumor response across tissue type. The newly defined markers

Being able to relate a patients T cell population status to others could lead the way to informed therapeutic choices in the clinic. If the underlying T cell populations could be revealed before treatment decision, researchers and clinicians could make informed decisions based on previous research and patient response. It is well known that tumors vary from patient to patient, as such cancer treatment should be as individual as tumor makeup. This level of therapeutic designation could lead the way to even more personalized decisions that take immune and tumor cell composition into account.

My project seeks to more effectively connect independent studies by allowing biologists to relate the states of their T cells to those defined by X in the paper.

Innovation:

Currently, T cell state has been heavily implicated in tumor response to immune and other therapies. Recent studies have aimed to characterize the differences in T cell populations that determine prognosis and response. They have revealed that there are many different pathways that lead to the exhaustion phenotype, as well as different subtypes of the the exhaustion phenotype. These differences are determined using cancer from a specific tissue type, usually in relation to a specific therapeutic as well. One recent study has aimed to tie together the defining genetic markers from all cancer types that distinguish the different T cell states. They were able to classify many different subtypes of T cells, both naïve and exhausted in phenotype by different genetic markers. However, they did so by showing relative effect size of each gene. While highly informative, this data that distinguishes the states is not easily accessible for most researchers.

This project aims to create an easy-to-use tool for biologists to compare their respective studies to previous studies and their defined states. Currently, no such tool exists making it difficult for researchers to utilize the mapping of these T cell states ~~as recently defined~~. This tool shall easily compare T cell states regardless of tissue and cancer origin. The focus of this tool will be on creating a multi-dimensional space to allow information from all implicated genes to influence the distance metric. Since there are multiple known existing pathways from naive to exhausted T cells, it is imperative that the analysis allows for this variability. In addition, this tool will define a “closeness” metric to inform users of similar their population is to a defined state. The use of this “closeness” metric instead of a more stringent clustering metric will allow researchers to make more informed decisions about the composition of their sample. The “closeness” metric will act as a confidence measure to prevent stringent definitions of the current state of the population. The purpose of avoiding classical clustering allows for the discovery of intermediate phenotypes. Investigators can interrogate how a sample relates to a specific state instead of trying to give distinct labels.

Most existing tools aim to place cells along a trajectory of their differentiation pathways. Since not all of these pathways are well defined, the focus of this new tool will be on key genes to allow a comprehensive cross tissue study that other, currently existing tools do not. In addition, if the pathway is not well defined, this tool still allows the researcher to query the relatedness of a cell to a defined state using only gene expression. In this way, this project focuses on how a sample relates to others and defined states instead of the pathway a cell may take to get to the defined state.

The key use of visualization and an easily interpretable distance metric will make this easy to use and interpret. Researchers will confidently be able to compare their results to those of others and make more informed biological questions and conclusions.

Specific Aim 1: To give meaningful information about T cell subtype state by creating a spherical mapping system and distance metric based on established CD8+ T cell subtypes

T cells are known to play an important role in the TME and response of tumor cells to immunotherapy treatments. Populations comprised of mostly exhausted T cells often have a poor response. However, most studies reestablish what exhausted, naïve T cells look like in regards to gene expression, without comparing to other tumor and tissue types. Most studies define the most differentially expressed genes in their cohort that elicited the different in response. This redefinition makes it difficult to compare immunotherapy responses due to T cell composition across studies. To overcome this limitation there has been a recent push to define T cell sub states across cancer types. Despite this, there still does not exist an easy way to compare the important genes in these states. Here we define the idea of “closeness” between two states of T cells which will be determined by the gene expression of a subset of genes that… have designated as important in the exhaustion pathways.

Approach:

Our approach shall first normalize the new input data in order to compare it to the data utilized in… In order to make comparisons between the defend T cell states and new unrelated data, it must be processed following the same pipeline as defined in…. The pipeline will take scRNA seq data from CD8+ T cells as input and partition the cells into smaller groups called miniclusters. Harmony will then be used for batch effect correction. Seurat will be further used to identify clusters and estimate moderated effect size. Once the data has been processed endpoints of exhausted and naïve T cells will be defined either according to user or program specification. In…. they identified four subtypes of T cells that correspond to four naïve states in …. These four states will be used to define the naive endpoint. The four endpoints of exhausted T cells defined shall be used as the others. Using UMAP’s spherical embeddings these endpoints will be located at opposite ends of the sphere.

Specific Aim 2: To confirm that T cell subtypes correlate to immunotherapy response by doing differential gene analysis on an unrelated immunotherapy dataset

Although these subtypes have been defined, there is no existing study to directly tie these states to phenotypic outcomes of therapeutics. Therefore, it is important to investigate the correlation of therapeutic outcomes to these defined states.

Since this project largely relies upon the assumption that these chosen genes have the potential to predict therapy response it is necessary to test on independent datasets2,4,17,18. These datasets are from unrelated, independent studies on varying tumor and tissue types. All four can be found on TISCH, which means they have been normalized to allow cross study comparisons and have CD8+T cell scRNA seq. They contain information about patient response, tumor, and treatment type. Since they are from the same data resource and have been normalized similarly, this should remove bias. In addition, the initial studies done on these datasets were aimed at adjacent biological questions but were not directed at CD8+ Tcell states.

In an attempt to compare T cell state effect on patient response, the same pipeline will be conducted on all datasets. A series of clustering and feature analysis shall be conducted on the studies to look at differential gene expression between the different groups of responders in each individual study. Potential clustering approaches include UMAP, PCA, hierarchical clustering, and logistic regression. The top genes from each study will be pulled out and compared to the different states as defined by… this is to ensure that there is a distinguishable difference between patients with different responses that can be quantified by T cell population composition.

* Hypothesis:
  + CD8+ T cells can be subtypes according to a subset of genes
  + Unique and non-overlapping
  + If youre null hypothesis instead of true, then T cells are more of a continuous gradient
  + Can be given a distance measurement to “closeness” to subtype definition
  + Distance can help inform how T cell will react in the TME
  + CD8+ T cells can be subtypes according to a subset of selected genes
  + CD8+ T cells can be mapped
  + Meaningful distance metric can reveal meaningful information about CD8 +T cell subtypes
* Propose:
  + Using reference map as defined by…
  + Create a mapping and distance metric in order to provide researchers with an easy way of comparing their data and samples to defined CD8+ T cell subtypes from different tissue and cancer types
* IF Incorrect:
  + Means may be way more subtypes possibly infinite space of continuous gene expression that would be difficult to represent or cluster in a meaningful way
  + In this instance having a distance metric could be more important clinically
  + Could be a case-by-case basis where you have to stratify by tumor/cancer type as well
  + Would have to evaluate each new population and corresponding therapy response independently
  + Markers may not be as important as initially thought, could be more correlatory than causatory
* Immunotherapies have been shown to have promising results in various cancers. However, response is varied among patients and cancer types
* Suggesting there is an underlying mechanism that contributes to the success of the treatment
* Its possible this variation in response is linked to the T cell state
* Progression to exhaustion state seems to be highly regulated by transcription factors and epigenetic regulators suggesting that scRNA seq could reveal the underlying phenotype and progression of the T cell along the pathway to exhaustion
* Recent studies have shown that T cell population makeup can have a large effect on immunotherapy success and tumor progression
* Tumors with populations largely comprised of exhausted T cells can have a worse prognosis
* Recent studies have aimed to further subtype CD8+ T cells in tumors and further elucidate exhausted T cell subtypes and lineage
* Descriptive subtyping of T cell populations and composition in the TME could improve future immunotherapy aims and decision making
* Unfortunately, although subtypes have been defined there is not an easy and standardized way to represent the composition and subtype of T cell populations in a sample
* Currently most studies illustrate the differences in their cohort instead of comparing to CD8+ T cell subtypes as a whole
* Difference in population has been shown to have impact on outcome
* *Define naive vs exhausted T cell*
* Exhausted:
  + Dysfunctional
  + No longer have anti-tumor effector potential
* Understanding the relationship of T cell population to the definition of naive and exhausted t cells can help us better understand the mechanisms of immune response in tumor tissue
* *Figure UMAP actually does with multi-dimensional data and how that transforms onto spherical space*
* Significance:

* + Immunotherapy is becoming a big target for cancer treatments
  + T cells and other immune type cells are known to influence how effective the treatment is
  + Different treatments could be more/less effective due to composition of T cell population
  + Being able to easily compare studies and T cells could help make initial hypotheses and drive biological questions
  + Easy way to stratify the patients who have a response vs the ones who do not
  + Easily accessible tool for biologists
  + Not a lot of computation required
  + Can lead to more interesting analyses focused on widely defined T cell subtypes
  + T cell states have been defined
  + Papers are defining the same subtypes
  + Would remove the issue of constantly redefining subtypes
  + Could expand to other disease/subtype problems
  + Easy way for biologists to initially examine/query their data
  + Could allow for more interesting biological questions
    - Immunotherapy response
    - Easier to compare T cell populations across samples
  + Easy way to map new data to previously established definitions instead of creating a new subtype for each project
  + Provides genes of interest to interrogate
  + Builds upon previous work
  + Personalize immunotherapies
* Specific Aim 1: To give meaningful information about T cell subtype we will create a spherical mapping and distance metric based on CD8 + T cell subtype
  + Create mapping and distance metric
  + Partition cells into small groups called miniclusters
  + Harmony for batch effect correction
  + Seurat to identify clusters
  + Estimate moderated effect size
  + Define endpoints of exhausted and naieve T cells
  + Map onto spherical coords using UMAP
  + Map new data onto it
  + Measure distance from endpoints
  + Do for different T cell endpoints
  + do for indiv T cells and for T cells as a population
  + See if patients who responded better can be determined using the distance metric
  + Visualization to show where T cells lie between endpoints
  + Can map multiple endpoints
  + Multiple T cell populations

Specific Aim 2: To confirm that the defined T cell subtypes correlate to immunotherapy response

* Use unrelated immunotherapy dataset with response
* Pick out differential genes in responsive and non-responsive patients
* See if corresponds to subtypes defined in map
* Compare genes to genes picked out in paper
* See if correspond
* See if distance works for unrelated dataset

Specific Aim 2: To confirm that the defined T cell subtypes correlate to immunotherapy response

* Want to make sure that the genes are truly distinguishable and able to be pulled out
* Basically going to cluster the two sets based on response
* See if gene sets correspond to the ones defined in the paper
* If aim 1 worked use that created tool to get distance metric for each patient and see how it corresponds to phenotype of treatment response
* Most of these studies look at differentiating patients based on various metrics but don’t focus specifically on the differences in CD8+ T cell populations
* Great because they are unrelated and independent which will allow for a largely unbiased study of the relationship between T cell state and treatment outcome.
* Want to ensure that there is a distinguishable difference in at least some patients between those who respond and those that don’t that can be quantified by T cell population composition.

TODO:

1. Finish SA2
2. Finish SA1
3. Finish intro
4. Incorporate figures
5. Add references
6. Go through make sure not missing shit

A picture containing calendar

Description automatically generatedDiagram

Description automatically generated

1. Arneth, B. Tumor microenvironment. *Medicina (Lithuania)* vol. 56 (2020).

2. Yuen, K. C. *et al.* High systemic and tumor-associated IL-8 correlates with reduced clinical benefit of PD-L1 blockade. *Nature Medicine* **26**, 693–698 (2020).

3. Sharma, P. & Allison, J. P. *The future of immune checkpoint therapy*. https://www.science.org.

4. Sheih, A. *et al.* Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy. *Nature Communications* **11**, (2020).

5. Golubovskaya, V. & Wu, L. Different subsets of T cells, memory, effector functions, and CAR-T immunotherapy. *Cancers* vol. 8 (2016).

6. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).

7. Wu, T. D. *et al.* Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* **579**, 274–278 (2020).

8. Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications* **10**, (2019).

9. Lu, Y. C. *et al.* Single-cell transcriptome analysis reveals gene signatures associated with T-cell persistence following adoptive cell therapy. *Cancer Immunology Research* **7**, 1824–1836 (2019).

10. Pauken, K. E. & Wherry, E. J. Overcoming T cell exhaustion in infection and cancer. *Trends in Immunology* vol. 36 265–276 (2015).

11. Sen, D. R. *et al.* The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).

12. Khan, O. *et al.* TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* **571**, 211–218 (2019).

13. Wherry, E. J. *et al.* Molecular Signature of CD8+ T Cell Exhaustion during Chronic Viral Infection. *Immunity* **27**, 670–684 (2007).

14. Wherry, E. J. & Ahmed, R. Memory CD8 T-Cell Differentiation during Viral Infection. *Journal of Virology* **78**, 5535–5545 (2004).

15. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nature Reviews Immunology* vol. 15 486–499 (2015).

16. Zheng, L. *et al.* Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, (2021).

17. Rendeiro, A. F. *et al.* Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nature Communications* **11**, (2020).

18. Ledergor, G. *et al.* Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nature Medicine* **24**, 1867–1876 (2018).