TCR Dist

TCRDist was used to calculate the distance metric between alpha and beta chains and alpha+Beta chain combined.

In this methodology, T-cell receptor (TCR) repertoire analysis conducted using TCRdist, a computational tool designed to quantify the similarity between TCR sequences. This analysis is focused on the alpha and beta chains of TCRs from mouse and human samples, which are part of the immune system's mechanism for recognizing and responding to pathogens.

In the context of immunology, TCRs are highly variable receptors that enable T cells to recognize a vast array of antigens presented by the major histocompatibility complex (MHC) on the surface of other cells. The unique variability in the TCR regions, especially the Complementarity Determining Regions (CDRs), is crucial for the immune system's ability to respond to diverse pathogens.  
  
Regarding the images which represent the results from the UMAP dimensionality reduction applied to the TCRdist matrices:

The mouse data shows clusters of TCR sequences with similar characteristics, which likely correspond to TCRs that recognize similar antigens. The clustering pattern is relatively spread out, indicating diversity in the TCR repertoire.

The human data also shows clustering, which indicates that certain TCR sequences are more similar to each other than others, likely due to common antigen recognition.

Weights (weights\_a and weights\_b) are assigned to different regions, such as the CDR3, to emphasize their importance in antigen recognition. In the provided code, the cdr3 regions are given a weight of 3, reflecting their critical role in specificity, as the CDR3 region is the most variable and directly interacts with the peptide presented by MHC molecules. The weights of the hypterparameter was chosen after tedious process of predictions from the KNN algorithm.

Hamming distance is used for its simplicity and effectiveness in capturing sequence similarity.

Other distance metrics were tried out like the default and Euclidean distance , hamming distance had given the better accuracy of them all.

The ability to discern subtle differences between TCR sequences is critical because small changes can significantly alter antigen specificity. UMAP's preservation of both local and global structures helps in identifying clusters that might represent TCRs targeting the same antigen or similar antigens. This is crucial in understanding the immune response, which is often the goal of such analyses. This clearly shown in the clusters formed and shown in the UMAP dimension reduction algorithm.

The choice between UMAP and other dimensionality reduction methods should be based on the specific requirements of the data and the goals of the analysis. For the visualization of complex immunological datasets where both local similarity (fine-grained clustering of very similar TCRs) and broader trends (such as the overall diversity of the TCR repertoire) are important, UMAP is often the preferred choice.

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9719034/

2. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2021.640725/full

3. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2020.01803/full

4.https://www.annualreviews.org/content/journals/10.1146/annurev-immunol-082119-124838

5.https://elifesciences.org/reviewed-preprints/88837v1#abstract

6. Vujovic M, Degn KF, Marin FI, Schaap-Johansen AL, Chain B, Andresen TL, Kaplinsky J, Marcatili P. T cell receptor sequence clustering and antigen specificity. Comput Struct Biotechnol J (2020) 18:2166–2173. doi:10.1016/j.csbj.2020.06.041

A map of different colored dots

Description automatically generated

A diagram of a number of dots

Description automatically generated with medium confidence