Interpretation and analysis of dimensionality reduction results

In the dimensionality reduction results considering both Alpha and Beta chains, the human TCRs are widely distributed across the graphs, showing a more diverse clustering with some tightly clustered groups and some scattered points. There are some overlapping point, but overall, the data distribution is more dispersed. The mouse TCR data shows a relatively compact cluster structure, with obvious boundaries and distances between TCR clusters. For example, the TCR clusters corresponding to antigens such as "SSYRRVPGI" have distinct boundaries and distances between them and other data points. Therefore, when considering both Alpha and Beta chains in combination, mouse TCRs show more concentrated specificity compared to human.

Compared to the dimensionality reduction results of combining Alpha and Beta chains, the single Alpha chain TCR clustering results for both humans and mouses were more dispersed, but the number of points appeared to become fewer as a large number of data points overlapped. Overall, the data points form a large number of small clusters that are more dispersed from each other and more tightly packed internally. The data points in the small clusters have many highly correlated features.

The dimensionality reduction of the Beta chains resulted in tighter clusters with more distinct boundaries than the previous two. The human TCR data formed several larger clusters, and these clusters did not show domination by specific epitopes within the clusters, but rather a uniform distribution of data points for various epitopes. The mouse single Beta chain data showed one large cluster and some discrete points, and unlike the human data, the clusters showed relatively obvious clustering of data points of different epitopes.

Overall, the human TCR is more dispersed and more difficult to form a single cluster compared to the mouse. It is possible that the high genetic diversity of the human population, the more complex immune system, and the exposure to a wide variety of pathogens in different environments have caused the human TCR to exhibit a more diverse clustering behaviour. Single Alpha chain data exhibit a large amount of overlap and form small dispersed clusters; single Beta chain data are denser with distinct cluster boundaries. This may be due to the fact that during gene rearrangement, the Alpha chains does not have a D region and only undergoes VJ rearrangement **[Rosati, E., Dowds, C.M., Liaskou, E. et al. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol 17, 61 (2017). https://doi.org/10.1186/s12896-017-0379-9]**, which has a lower antigenic specificity compared to Beta. And after combining Alpha and Beta, the data points appeared to be more dispersed overall. We believe the reason is different TCRs have different combinations of features on the Alpha and Beta chains, resulting in a more diverse distribution, making the data points more scattered after combination.

