# Literature Review

## Introduction

As part of adaptive immunity, T cells detect and eliminate infected or cancerous cells in the body. T cell receptors (TCR) express on the membrane of T cells recognize the target. As a result, a wide range of TCRs are produced. Since TCR repertoires respond very strongly to health status, TCR specificity can be very useful for therapeutic approaches. Empirical methods do not offer a reliable way of predicting specificity for TCRs.

In this project, machine learning will be used to predict TCR based on the VDJdb TCR database.

## Theoretical Framework

As part of the study (Dash et al., 2017), quantifiable predictive features were defined for certain epitopes within the TCR library. TCR epitope specificity was predicted based on sequence features alone in a novel approach to sequencing and analysing TCRs. The analysis they produce is useful for grouping receptors that are related and selecting representatives to study specificity in further experiments. Using parameterised antigen-specific immune libraries based on different epitopes, they propose developing generalised TCR-pMHC recognition models, which can be applied to a variety of research areas, including cancer immunotherapy and infectious diseases diagnosis and treatment.

The study (Vujovic et al., 2020) explored TCR sequence clustering based on antigen specificity, an important step towards understanding how TCRs detect different antigens and the potential for targeted therapies based on TCRs. Based on TCR sequences, this paper proposes a clustering algorithm for classifying antigen specificity. In addition to enhancing prediction specificity and accuracy, deep learning techniques are also highly effective. There is no single tool available to unambiguously classify TCR receptor specificity because of the TCR's cross-reactivity, its ability to bind multiple antigens with different affinities, and its inability to induce T-cell activation merely by binding to the TCR.

The DeepTCR framework was developed by (Sidhom et al., 2021). Deep learning algorithms are used for analysing TCR sequencing data. A structure based on TCR sequences is proposed for revealing the complex pattern of antigenic specificity found in TCR sequences. Analyzing TCR sequences from human and mouse datasets with DeepTCR uses both supervised and unsupervised deep learning methods. In order to efficiently learn these complex patterns, the framework extracts features from sequencing data and employs a variety of neural network architectures. Using deep learning from conventional T cells culture combined with TCR sequencing, they identify antigen-specific responses that identify both an immune response and its TCR sequence diversity. Their results demonstrate the power of detecting an immune response as well as its TCR sequence diversity. In their work, they demonstrate the value of this approach in generating previously unrecognized hypotheses.

The book Multidimensional Scaling (Kruskal & Wish, 1993) describes the statistical technique of multidimensional scaling (MDS), which transforms complex, multidimensional data into a more understandable two- or three-dimensional space. The MDS method can be applied to immunology, including the modeling of TCR specificity. The objective is to reduce dimensionality while maintaining similarity or dissimilarity in the distance between data points. The use of this technique is essential when analyzing TCRs. This is because receptors and their target antigens are both represented in multidimensional space, and their proximity indicates a likelihood of interaction or similarity in function. Using MDS, one can visualise and explore the high-dimensional space of TCR interactions for TCR specificity prediction. TCR sequences can be visually grouped into functional or specific groups by MDS's effective reduction of dimensionality.

**Reference：**

1. Dash, P. et al. (2017) ‘Quantifiable predictive features define epitope-specific T cell receptor repertoires’, Nature, 547(7661), pp. 89–93. doi:10.1038/nature22383.
2. Sidhom, J.-W. et al. (2021) ‘DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires’, Nature Communications, 12(1). doi:10.1038/s41467-021-21879-w.
3. Vujovic, M. et al. (2020) ‘T cell receptor sequence clustering and antigen specificity’, Computational and Structural Biotechnology Journal, 18, pp. 2166–2173. doi:10.1016/j.csbj.2020.06.041.
4. Kruskal, J.B. and Wish, M. (1993) Multidimensional scaling. Newbury Park, CA: SAGE publ.