The framework of our model is depicted in Figure 1. The model includes two key components: (1) TCR classifier which is used to calculate the edit distance matrix of TRBs between the target TCR and other TCRs in the training database and select the antigen of the closest training data as the target antigen. (2) Specificity Classifier Based on Random Forest. First, the selected target antigen and CDR3 are encoded by N-Gram respectively. Then the encoded CDR3 and antigen sequences are used as feature input, and the labels are used as a model output to build a random forest model (RF) which finally produces prediction results. In the following section, we will provide detailed explanations of the model.

图1先空着，等周四我确定之后再补上来