Repertoire-based classifiers and disease-specific diagnostics

Abbie Olson[1], Travers Ching[2], Bryan Howie[2], Cara Forsberg[2]
Adaptive Biotechnologies, University of Oregon

Abstract

The ability of T cells to produce αβ-heterodimeric antigen-specific T cell receptors (TCRs) through V(D)J recombination is a paramount component of the adaptive immune system's ability to identify and clear pathogenic infections. An individual's T cell repertoire encodes their dynamic immune history. Profiling T cell repertoires for enrichment of various pathogens via immunosequencing could be generalized to a wide variety of diseases and could ultimately be utilized in a universal diagnostic[1]. A combination of lab work and machine learning methods may be the key to this diagnostic of the future. The following review covers virus-specific classifiers, specifically for cytomegalovirus (Adaptive Biotechnologies) and SARS-CoV-2 (Adaptive Biotechnologies and Microsoft)[2].

Figure 1: ImmuneCODE™ partnerships.

Introduction

In this review, we will cover two independent studies, both with similar goals of identifying specific morbidities based on dynamic T cell responses. TCR specificity is mediated by primary sequence diversity. When a TCR-beta (TCRβ) gene matures, it is randomly rearranged at its complementary determining region 3 (CDR3) by combining noncontiguous variable (V), diversity (D), and joining (J) region gene segments of the germline locus[1]. Indels at the V-D and D-J junctions add to further diversification of the receptors, which results in incredibly diverse TCRβ CDR3 regions. When a T cell recognizes an antigen, it proliferates via clonal expansion and some of those T cells become part of the memory compartment (memory T cells), in which they may reside for years. Healthy adults express around 10^7 unique TCRβ chains on around 10^{12} circulating T cells[3].

A phenomenon can occur in public T cells in which a particular antigen is targeted by the same TCR sequence in multiple individuals, thus potentially eliminating the need for sequencing in the case of some diseases. However, additional examples of public T cell responses to infectious diseases also include Epstein-Barr virus (EBV), *Clostridium tetani*, parvovirus, herpes

simplex virus (HSV), HIV, influenza, and more recently, SARS-CoV-2. Additionally, public T cell responses have also been recorded in certain malignancies and autoimmune disorders[1].

Cytomegalovirus

The first example that we will explore are public T cell responses in patients who had been infected with cytomegalovirus (CMV) at some point. Generally, these public T cell responses are studied in the context of specific antigens in a human leukocyte antigen (HLA) context by isolation of antigen-bound T cells followed by low-throughput sequencing of the variable regions of TCRβ chains[1]. The method to actually finding these enriched TCRβ chains requires finding the statistical significance between sets of TCRβ sequences and phenotypes of interest. The idea is that the following approach may have the potential to be applied as a diagnostic strategy for a range of immune-related phenotypes[1].

After immunosequencing, TCRβ sequences were analyzed *in silico*. The presence or absence of individual TCRβ sequences were tabulated for each sample in a large investigational cohort (Fig. 2) of healthy bone marrow donors and their concordance with phenotypes of interest was assessed. Approximately 190,000 unique TCRβ sequences were found in each cohort subject. In this study, unique TCRβ sequences were defined as a unique combination of V-genes, J-genes, and CDR3 amino acid sequences, also called a BioIDs (Fig. 3).

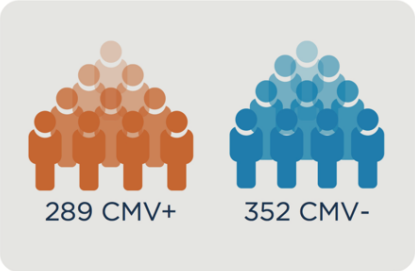

Figure 2: Serostatus distribution of testing cohort.

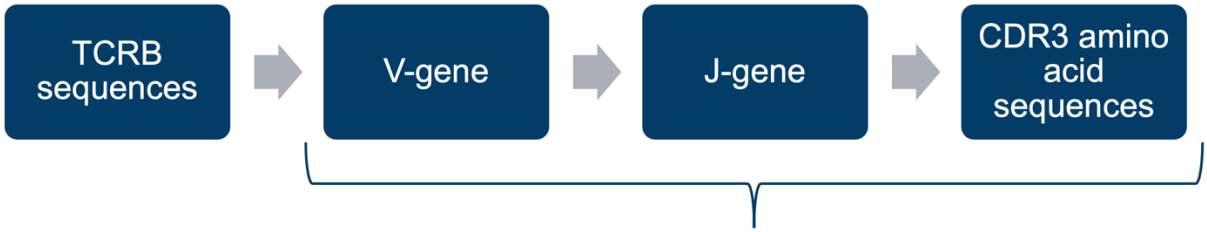

Figure 3: Schematic of BioID components.

CMV was used in the proof-of-principal study, as it's a chronic virus that is considered a model organism for public T cell responses. It infects anywhere from 30 to 90% of adults. The cohort was serotyped for CMV status (CMV+ or CMV-), and a smaller, independent cohort was also sequenced for validation purposes. Each sample in the testing cohort was screened for

unique TCRβ chains and CMV+ subjects were compared to CMV- subjects using a significance threshold of $P < 1 \times 10^{-4}$, which was established in a cross-validation procedure[1].

164 TCRβ chains were determined via these thresholds to be CMV-associated, most of which were hypothesized to be from CMV+ patients. A generative binary classifier was then constructed to test the serostatus hypothesis, functioning to infer CMV serostatus. The classifier was trained on the large cohort and tested on the smaller cohort and yielded impressive accuracy: a AUROC value of 0.94, a sensitivity of 0.90, a specificity of 0.88, and a diagnostic odds ratio of 0.70 (Fig. 4). These results demonstrate an excellent diagnostic approach and could be utilized in the case of other morbidities, in fact a similar method is being utilized by Adaptive to identify SARS-CoV-2 infection.

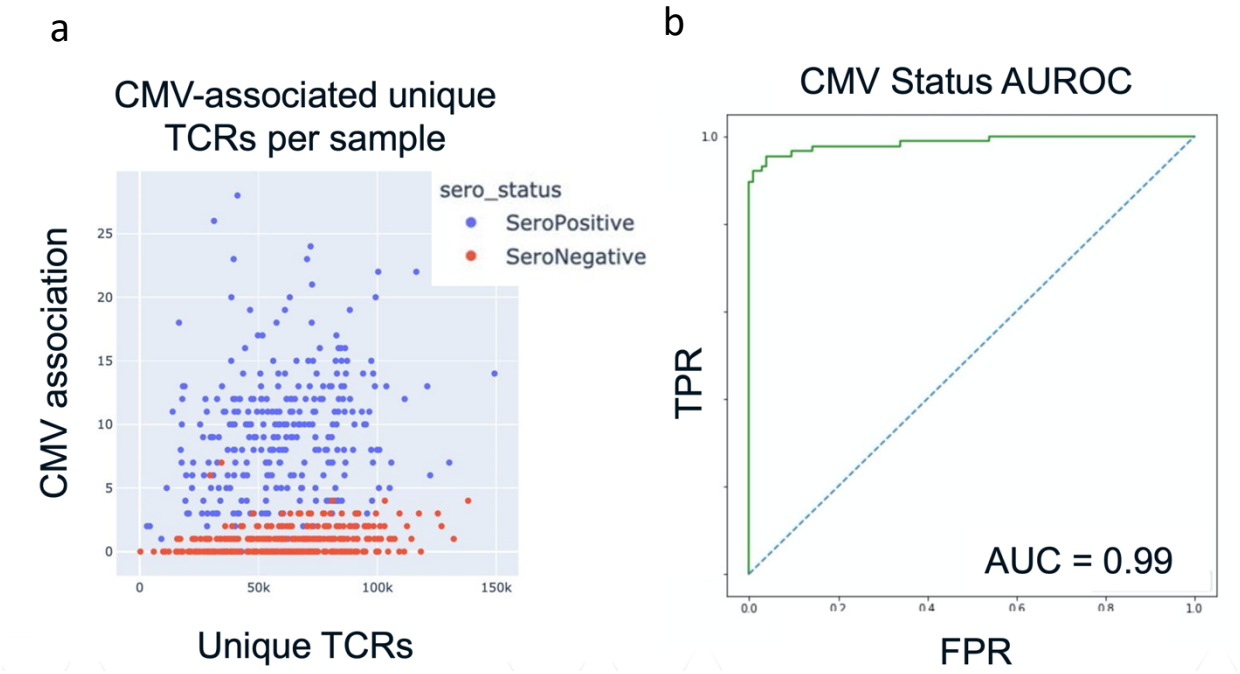

Figure 4: (a) Scatterplot of the distinct separation between CMV+ and CMV- samples. The x-axis displays counts of unique TCRs per sample, while the y-axis displays the number of CMV-associated TCRβs compared to the total number of unique TCRβs sampled for each subject by serostatus. **(b)** ROC curve showing the performance of the classifier trained and tested on the large cohort.

ImmuneCODE™

With a widespread interest in understanding and ultimately contributing to the treatment of SARS-CoV-2, Adaptive has built a TCR platform to track disease-associated TCRs during the course of infection. The central role of T cells in the early detection of viral infections makes them a great target for assessing immune response to SARS-CoV-2. 61 total COVID-19 subjects were selected for the study: 58 recovered and 3 acutely infected individuals. Each subject was characterized via CD8 T cell response mapping to antigen stimulation. 545 HLA class I presented viral peptides (class II data is forthcoming) were introduced to the CD8 T cells[2].

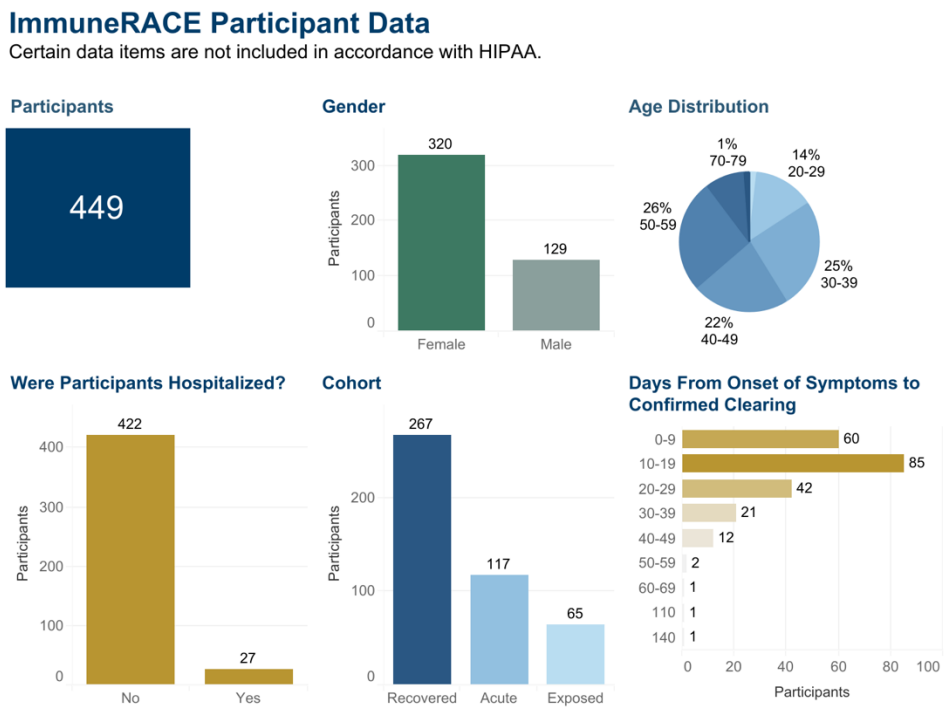

Figure 4: ImmuneCODE/ImmuneRACE participant data.

T cell repertoires were then sequenced on 1,815 samples from 1,521 COVID-19 subjects. 3,500 controls were also employed to identify public TCRs associated with SARS-CoV-2 infection from CD8 (killer) and CD4 (helper) T cells. The data reveals that T cell responses to SARS-CoV-2 infection peaks one to two weeks after infection (an expected response) and can be detected for several months after convalescence3. A classifier similar to that of the CMV subjects was trained to diagnose SARS-CoV-2 infection via TCR sequencing from blood samples. Like the CMV status classifier, this one had a whopping 99.8% specificity, in addition to high early sensitivity after infection and lasting sensitivity after convalescence. This classifier is remarkably accurate, and much of that can be attributed to scientists at both Adaptive and Microsoft, who originally partnered on a Lyme disease diagnostic. The goal of this initial partnership (which also focused on mapping an atlas of immune responses for multiple diseases) was to create a universal blood diagnostic to read a subject's immune responses and diagnose or treat diseases. However, the partnership has almost exclusively shifted its current focus (as of November 2020) to COVID-19 subjects.

Conventional immune monitoring assays like ELISpot and ICS require live T cells, which significantly limits throughput. Adaptive's scalable molecular assay utilizes a multiplexed experimental platform to interrogate T cell repertoires with query antigens, thus leading to the identification of SARS-CoV-2-specific TCRs in an HLA context[2]. This approach allows for the characterization of numerous antigens involved in T cell immune responses, effectively becoming a map between TCR sequences and SARS-CoV-2-specific antigens. Clonal depth and breadth were also captured via this method, in addition to immune response dynamics to SARS-CoV-2 over time. T cell responses were reported to be durable for at least 3 months after infection, though it could be longer.

CD8 T cell responses were directly characterized with Adaptive's MIRA (Multiplex Identification of T cell Receptor Antigen Specificity) assay, which maps TCRs to antigens at a high specificity. The 545 chosen query peptides were then selected from HLA-I-NetMHCpan predictions across multiple representative HLA types from the literature[3]. The peptides were synthesized and assigned either individually or as groups of related peptides to one of 269 unique MIRA pools. MIRA was then performed on T cells derived from peripheral blood mononuclear cells (PBMCs) and public T cells that were shared between many individuals. *In silico* processes were quite similar to those of the CMV work. Public COVID-19-associated TCRs were then identified via a Fisher's Exact test.

51 samples were paired with immunosequencing and COVID-19 MIRA data, and concordance between breadth and depth was estimated by MIRA with remarkable accuracy (Fig. 5). Clonal depth and breadth in TCRs were measured at an order of magnitude higher than public clones, demonstrating MIRA's strength in identifying disease associated TCRs.

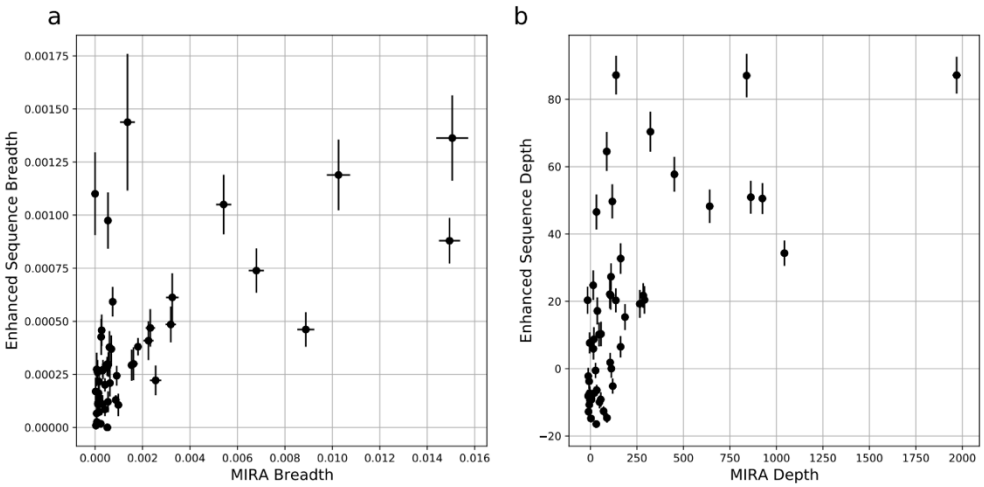

Figure 5: MIRA-estimated clonal depth and breadth distinguished disease-associated TCRs.

Conclusion

While functional T cell assays are difficult to perform, they will likely continue to be critical in understanding cellular immune responses to SARS-CoV-2, with the benefits of mapping specific TCRs diseases outweighing the challenges. Analyses on CMV and SARS-CoV-2 both suggest that specific T cell responses to infection are clinically insightful. Some might consider an endeavor like this to emulate the failed "Theranos" diagnostic, though this evidence is undeniable. T cells hold a key to identifying morbidities quickly and much less expensively than current methods require. Though this research is ongoing, the importance of characterizing immune responses will only expand opportunities for vaccine and therapeutic developments for a myriad of diseases. Nevertheless, a reproducible, high-throughput molecular approach to assess T cell responses will serve a currently unmet need to characterize the adaptive immune system's response to SARS-CoV-2 antigens.

References

1. Emerson, Ryan O., et al. "Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire." *Nature genetics* 49.5 (2017): 659-665.
2. Sidhom, John-William, and Alexander S. Baras. "Analysis of SARS-CoV-2 specific T-cell receptors in ImmuneCode reveals cross-reactivity to immunodominant Influenza M1 epitope." *bioRxiv* (2020).
3. Snyder, Thomas M., et al. "Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels." *medRxiv* (2020).