**Overcoming the Limitations Posed by TCR-beta Repertoire Modeling through In-Silico DNA Recombination Algorithm**

The adaptive immune system protects vertebrates by detecting and neutralizing foreign invaders (antigens) using T-cell receptors (TCRs), which are placed on the surface of a T-cell.

A TCR recognizes an antigen by detecting the small protein fragments that are on the surface of that antigen, and then sends a message to the nucleus of its T-cell. This successful recognition induces a response to eliminate the antigens. The diversity in the TCR pool increases the chance of detecting a variety of antigens for the adaptive immune system, which is the first step of a successful recovery from diseases. Analysis of TCR pool (repertoire) is crucial for understanding the functionality of a healthy immune system, determining the nature of successful and unsuccessful immune responses, and understanding the immune mechanism in the presence of different diseases such as type 1 diabetes, various cancers (blood, breast, colorectal, etc.), rheumatoid arthritis (an autoimmune disease), and multiple sclerosis. The response of the immune system to a specific antigen often leaves evidence in the form of repertoire sequence patterns (signatures) that are common across individuals. Ability to detect such patterns is critical for understanding the correlation between the immune receptors and different diseases, and identifying immune receptor clones that can be converted into precision vaccines.

A diverse set of TCRs is required for the adaptive immune system to detect a wide variety of antigens successfully. The immune systems of the vertebrates achieve this diversity through the DNA recombination process, known as the V (D)J recombination. This process involves a rearrangement of the variable (V ), diversity (D), and joining (J) gene segments in a combinatorial way chosen from members of each gene family. The V(D)J recombination is the primary mechanism for generating a diverse repertoire of T-cell receptors (TCRs) essential to the adaptive immune system for recognizing a wide variety of diseases. However, modeling TCR repertoire is computationally challenging as the total number of TCRs to be generated and processed can exceed 1018 sequences.

The earliest clinical contributions in the field have already saved countless lives while garnering a Nobel Prize, yet the field is in its infancy and progress is hampered by a lack of understanding of the nature of which immune receptors are 1) created by the immune system, 2) bind to which antigenic targets, 3) become recruited into an immune response, and 4) retained in a memory anamnestic response. The main hindrance to systematically developing new immunotherapies, new immunodiagnostics, and novel immunobiomarkers is the enormous magnitude (>1018 unique species) of the repertoire of TCR species that can be made by the immune system. Replicating the recombination process in a simulation environment allows immunologists to test different hypothesis on immune system response analysis. Computational tools with modeling and predictive power which can handle this massive scale are sorely needed. In this project our goal is to leverage the GPU-based synthesis-on-the-fly approach developed by Akoglu’s lab and conduct a three phase investigation that covers:

1. Modeling all the potential TCR repertoire of the mouse using GPU-based TCR synthesis implementations on the UA HPC system
2. Investigating performance analysis of GPU-based TCR synthesis implementations in terms of execution time and scalability
3. Collecting number of unique recombination pathways for each in-vivo sequence based on mouse data set and applying statistical analysis on correlating to V, D, J and n-nucleotide sequences
4. Developing visualization tools for studying the immune receptors

Simulations will be carried out on the UA HPC system, scalability analysis will be conducted by varying the number of GPUs, and execution time analysis will be based on two versions of TCR synthesis in reference to serially executing code.

Student will automate the process of launching experiments for various simulation configurations and collecting simulation outputs, and then apply statistical methods and visualization approaches for understanding the significance of the increased number of unique recombination pathways. .