CALIFORNIA SCIENCE & ENGINEERING FAIR
2019 PROJECT SUMMARY

Name(s)

Cynthia Chen

Project Number

S0807

Project Title

Decoding Neural Networks: Novel Computational Methods to Discover Anti-Tumor B Cell Receptor Binding Motifs

Abstract

Objectives

Cancer accounts for over 8 million deaths worldwide each year. Studying the binding interaction between the B cell receptor (BCR) and the tumor antigen has become a promising field for building a better understanding of how the human adaptive immune system attacks cancer cells. Currently, the recurring protein sequence patterns (termed motifs) for the BCR binding region remain largely undifferentiated between cancer types. Determining these cancer-specific BCR binding motifs is crucial, as they can inform more targeted immunotherapies for cancer patients.

Methods

I developed novel computational methods to uncover the BCR binding motifs encoded in a deep neural network. In previous research, the deep learning model was trained on 3 million BCR sequences from the TCGA database and achieved an 0.8 average AUC in predicting cancer-specific BCR binding affinities for 13 cancer types. To decode the key motif information that allowed the model to accurately predict cancer types, I implemented a computational pipeline, developed with 3500+ lines of Python and R code. My pipeline consists of several algorithms: generating random input sequences, running the model to rank sequences, visualizing top sequences to distinguish binding patterns, and clustering to identify motifs.

Results

Using my pipeline, I discovered 65 BCR binding motifs among all 13 cancer types and identified the 12 most significant motifs overall. The robustness of the motifs was validated through a synthetic data simulation and extensive correlation analyses. Last, I demonstrated the versatility of my computational pipeline by applying it to antigen-specific sequences and full-length CDR3 sequences.

Conclusions

My research is the first to reveal and validate anti-tumor B cell receptor binding motifs for specific cancer types. This discovery is a key step towards future synthesis of new motif-based antibody drugs and more powerful and precise cancer treatments. The approaches and methods that I developed are versatile and applicable to other types of cancer and disease. Furthermore, the novel computational pipeline that I propose in my research can be reused and employed to decode a wide range of deep learning models and ultimately lead to more transparent and understandable AI.

Summary Statement

I developed a novel computational pipeline to decode deep neural networks, and I employed my pipeline to discover anti-tumor B cell receptor binding motifs.

Help Received

This research was conducted at Harvard Medical School under the guidance of Dr. Sherlock Hu. My science teacher, Mr. Christopher Spenner, advised me on the presentation aspect of my project.