Project Proposal

Project Title: Analyzing HLA Sequences to Predict Organ Rejection and Find Optimal Targets for Precise Immunosuppression

Author: Samhitha Bodangi

Date: 13 December 2023

Project Description:

The overall aim of this project is to decrease organ rejection using a machine-learning model that analyzes donor and recipient HLA sequences to find personalized targets for immunosuppression.

Current treatments to prevent organ rejection include life-long immunosuppressors that can severely weaken the immune system. Even with a perfect match, the polymorphism of HLA molecules can initiate an organ rejection response, forcing doctors to increase the dosage of immunosuppressors. The engineering goal of this project is to create a machine-learning model that can analyze donor and recipient HLA sequences and find specific allele differences to provide precise targets to prevent rejection while maintaining the strength of the rest of the immune system.

For doctors to classify an organ match, they check for similarities in the Human Leukocyte Antigen HLA sequence (*Matching and Compatibility | Transplant Center | UC Davis Health*, n.d.). Understanding the exact differences in HLAs between the donor and recipient can result in a better treatment method that is personalized and accurate for the recipient. Therefore, analyzing donor and recipient HLAs can predict the most prominent signal or pathway can provide evidence of the most effective antibody that can be used to block them. This way, the only T-cells that are suppressed are the ones that participate in organ rejection and will allow the rest of the immune system to be competent against other viruses and diseases.

In a clinical setting, the model could be used after analyzing donor and recipient samples and selecting immunosuppressors that inhibit the predicted pathways or proteins that would be present in the rejection response. Finding personalized targets can decrease the side effects of these medications and provide new insight into newer targets for more precise medications.

Background:

Problem Statement:
Chronic organ rejection affects about 50% of kidney transplants five years post-transplant. Due to chronic rejection occurring over a long period of time, there are limited methods to diagnose and treat chronic rejection. Even though Human Leukocyte Antigen (HLA) mismatches are the primary cause of rejection, HLA genes are very polymorphic, and current HLA typing methods do not account for the diverse amino acid variations within each allele that can initiate rejection.

Engineering Objective:
The objective is to make a machine learning model that can predict rejection, given donor and recipient HLA alleles. The model will work by identifying solvent-accessible amino acid mismatches. Then, the model will use these mismatches to predict donor-derived peptides that would bind to recipient MHC class II molecules. Using public databases and open-source servers, this model will predict the risk of rejection and provide information on targets for personalized immunosuppression.

Supply vs. Demand

Organ transplants are among the greatest advances in modern medicine, saving tens of thousands of lives every year. By increasing life expectancies and improving the quality of life, they remain the best therapy for terminal and irreversible organ failure (Grinyó, 2013). However, there is currently a major problem in the organ transplant industry: the demand is vastly greater than the supply. Over 100,000 people in the United States are on an organ transplant list, with another person being added to the national transplant waiting list every nine minutes. Unfortunately, many patients are unable to receive an organ transplant due to a lack of organ donations. About seventeen people die each day while waiting for an organ transplant (*Organ, Eye and Tissue Donation Statistics*, n.d.). The immense demand emphasizes that every donated organ has the potential to change lives, and it is crucial to maintain the long-term health of each organ, for the sake of the patient and the organ as well.

Overview of Organ Rejection:

Even if a patient is successful in receiving an organ transplant, many medical complications may occur after the transplant, the most common being organ rejection. The immune system is a body system that destroys foreign cells to protect the body from harm. In the case of organ rejection, the immune system recognizes the transplanted organ as foreign and attempts to attack it by producing cells or antibodies that invade the organ (*Understanding Transplant Rejection | Stony Brook Medicine*, n.d.). Currently, all transplant patients are prescribed immunosuppressors to decrease the risk of organ rejection. These immunosuppressors have revolutionized the field of organ transplantation and remain the standard of care to avoid rejection (Azzi et al., 2013). However, recipients must take immunosuppressive drugs for their entire lives for their bodies to accept a donated organ. While these medications prevent organ rejection to an extent, about 10-20% of patients will still experience at least one episode of rejection within the first 3 months to 1 year after a transplant (*Organ Rejection after Renal*

Transplant | Columbia Surgery, n.d.). Additionally, they can also severely weaken the immune system, increasing the risk of cancer, infections, and other diseases (Kelly, 2022). New treatments are necessary to prevent organ rejection without using broad immunosuppressors that weaken the entire immune system.

Chronic Rejection

Depending on the mechanisms and timeframe of the rejection episode, rejection can be categorized into many different types. Acute and chronic rejection are categorized based on the time rejection occurred after the transplant. Acute rejection occurs within the first three months to a year after the transplant, while chronic rejection can occur after the first year of the transplant. Chronic rejection is often irreversible and can lead or graft failure or death (Hunt & Saab, 2012). Immunosuppressors are effective in decreasing the risk of acute rejection, but not against chronic rejection. By 5 years posttransplant, chronic rejection affects up to 50% of kidney transplants (Gautreaux, 2017). Due to chronic rejection being often asymptomatic and occurring over an extended period, there is currently no medicine to date that can treat chronic rejection symptoms (*Understanding Transplant Rejection | Stony Brook Medicine*, n.d.). The common treatment method is to increase the dosage of immunosuppressive drugs, which can exacerbate the dangerous side effects. Therefore, it is imperative to understand and target the mechanisms that are involved in chronic rejection to maintain long-term allograft health.

MHC-Peptide Presentation and T-Cell Activation

Early chronic organ rejection is primarily caused by T-cell-mediated rejection (Chong, 2020). T-cells are a type of immune cell that plays a crucial role in identifying and eliminating foreign cells. When T-cells misinterpret donated organ cells as foreign, it can lead to T-cell activation and an attack on the transplanted organ. MHC peptide presentation plays a vital role in T-cell activation and can lead to developing strategies to prevent transplant rejection. The major histocompatibility complex (MHC) is a group of genes that code for MHC molecules found on the surface of cells. These proteins play a vital role in the immune system's ability to distinguish between "self" and "non-self" (King, 2007). There are two main types of MHC molecules: MHC class I and MHC class II molecules. As MHC class I molecules are present on all nucleated body cells and directly interact with T-cells, this project will focus on MHC class I peptide presentation (Hewitt, 2003). In the case of organ transplantation, intracellular proteins are broken down into smaller peptides inside the organ cells. These peptides are transported into the endoplasmic reticulum and bind to a groove in the MHC class I molecule, forming a peptide-MHC complex. This complex then travels to the cell surface and is displayed for T-cells to recognize. If

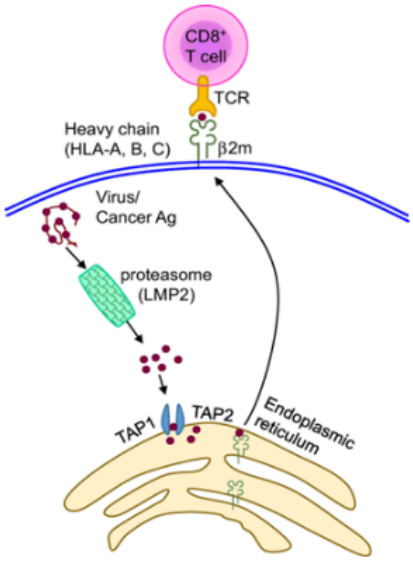

Figure 1: Intracellular antigenic peptides are presented to CD8+ T cells by MHC class I. Even though this figure depicts antigens of virus or cancer origins, the process is synonymous with organ cells. The antigens are processed by the immunoproteasome into peptides that are transported into the endoplasmic reticulum through TAP1 and TAP2 transporter proteins. These peptides are then loaded onto the MHC class I molecule (β2m). The MHC-peptide complex is presented on the cell surface to CD8+ T-cells (Vijayan et al., 2019).

T-cell receptors (TCRs) recognize a peptide from the transplanted organ on an MHC molecule, it activates, starting the immune response against the transplanted organ. Once T-cells recognize foreign antigens displayed on antigen-presenting cells (APC), proteins called cytokines synthesize and allow for the proliferation and differentiation of T-cells that attack the organ (Ingulli, 2010). Multiple cytokines or pathways may be active during rejection and may be different from person to person (Chen et al., 2019).

Tissue Typing and Immune Profiling

This document proposes selective T-cell inhibition with a personalized medicine approach to induce donor-specific tolerance. When looking for organ matches, doctors perform Human Leukocyte Antigen (HLA) typing to understand the similarity in antigens between the donor and the recipient. The HLA is a group of genes that provide instructions to make antigens that are present on the surface of cells (Manski et al., 2019). Six specific HLAs are looked for, and a higher similarity results in a likely chance of an organ match (Matching and Compatibility | Transplant Center | UC Davis Health, n.d.). Therefore, understanding the exact differences in HLAs between the donor and recipient can result in a better treatment method that is personalized and accurate for the recipient. With successful results, the supply and demand issue can be reduced, allowing more patients to receive a transplant.

Benefits of Machine Learning

Machine learning is a subset of artificial intelligence that uses statistical techniques that allow computer systems to automatically learn and develop from experience without being explicitly programmed (Costa, 2019). Previous studies have employed machine learning techniques to sift through massive datasets of gene expression data. Machine learning algorithms can analyze data to identify patterns and establish relationships from complex datasets. For this project, machine learning would allow HLA sequence data to be used to make a prediction model. By training the model on datasets of HLA sequences and peptide binding affinities, the algorithm can predict these complexes with high accuracy, paving the way for personalized and targeted immunosuppression. There have been many studies that employ machine learning to predict organ rejection. However, those models focus on “whole” HLA mismatches, which do not account for HLA polymorphism or the peptide sequences. Therefore, by focusing on HLA sequences and peptides, a more accurate and robust model can be created to prevent organ rejection. This way, we can protect the patient and the organ from harm.

Research Plan:

IDV: The machine-learning model, HLA sequence testing, datasets
DV: Donor and recipient HLA combinations, the type of model constructed

Materials:

- Public datasets
    - Gene Expression Omnibus (GEO)
    - United Network for Organ Sharing (U.N.O.S.) STAR Files
    - IPD-IMGT/HLA from the European Bioinformatics Institute (EBI)
- Computer and software
    - Google Colaboratory (Python programming language)
    - Rstudio (R programming language)
    - GitHub
    - Microsoft Excel
    - Statistical Analysis System (SAS)
    - Visual Studio Code
- Servers
    - NetMHCpan 4.0
    - Immune Epitope Database (IEDB)
    - NetSurfP

Procedure:

1. Collect data from publicly available databases, preferably in CSV or table-like format that can be imported as a data frame
    a. HLA allele typing data for donor and recipient and the status of rejection from U.N.O.S. to be used for testing and model validation
    b. HLA amino acid sequence data from IPD-IMGT/HLA database
2. Align amino acid sequences for each HLA loci and find amino acid mismatched
    a. Any position where the amino acids differ between donor and recipient are identified as mismatches
3. Find solvent-accessible amino acids using the NetSurfP server
4. Generate donor-derived peptides using solvent-accessible amino acid mismatches with NetMHCIIpan and find the most significant peptides
    a. Peptide length is 15-20 amino acids (for MHC class II)
5. Immunosuppressive targets are the most significant peptides that might cause rejection
6. Construct multiple models, such as SVM, KNN, RF
7. Test model using U.N.O.S. donor and recipient HLA typing samples to validate accuracy
8. Create user user-friendly, open-source web application that holds the model
    a. Using Visual Studio Code and HTML to handle UI
9. Compare U.N.O.S. samples with competitor models (HLA-EMMA, PIRCHE-II, and HLAMatchmaker) to assess increased or decreased accuracy

Risk/Safety Concerns:

As the model will focus on using public data sets, there are not many safety concerns with this project. However, if medical records are used, the data will only be used for the model and will not include any revealing details about the patient.

Data Analysis:

The data will primarily be analyzed with a decision matrix that contains multiple criteria, along with the weightage and scores for each respective model. Below is the list of criteria that will be used to evaluate the outcomes of each model type.

ROC Curve and AUC will provide information on how efficient the model is. The ROC curve is the graph and the area under the curve will range from 0 to 1, where 0.5 represents random guessing and 1 indicates perfect performance.

Confusion Matrix will be used to analyze the false positives and false negatives of the model.

Accuracy, Precision, Recall, and F1 Score will be used to evaluate the accuracy and precision of the model.

External/Cross-validation will be used to evaluate the accuracy of the model outside of the training dataset. It will give more information about the universality of the model which will be helpful when understanding the clinical applications of the model

Simulations will be used to see how the model could potentially be used clinically and the accuracy of the model in predicting rejection.

Timeline:

October: (reach at least 10 journal articles by the middle of this month)

- Continue researching
- Reach out to professors and labs

November:

- Finalize Lab
- Work on Project Documents
- Get datasets
- Start preliminary project testing
- Download software and research model types

December: (reach 20 journal articles by the middle of this month)

- Continue initial preliminary project testing (finish by the middle of this month)
- Begin basic model construction
- Analyze cytokine profile and choose monoclonal antibodies
- Create a poster for preliminary data and practice presenting for December Fair (12/15)
- Finish Z fair forms

January:

- Continue collecting data (finish all testing by the end of the month)
- Adjust model and train based on evaluation
- Cross-validate models with alternative datasets
- Try different algorithms and models and determine which model has the best outcomes
- If feasible, conduct lab testing, if not, use SimUNet or other research simulations

February:

- Analyzing data and creating graphs
- Create a design matrix of results across all models
- Create final poster and practice presenting for February Fair (2/15)

References:

Azzi, J. R., Sayegh, M. H., & Mallat, S. G. (2013). Calcineurin Inhibitors: 40 Years Later, Can't Live Without …. *The Journal of Immunology*, *191*(12), 5785–5791. https://doi.org/10.4049/jimmunol.1390055

Chen, H., Yang, J., Zhang, S., Qin, X., Jin, W., Sun, L., Li, F., & Cheng, Y. (2019). Serological cytokine profiles of cardiac rejection and lung infection after heart transplantation in rats. *Journal of Cardiothoracic Surgery*, *14*(1), 26. https://doi.org/10.1186/s13019-019-0839-5

Chong, A. S. (2020). B cells as antigen-presenting cells in transplantation rejection and tolerance. *Cellular Immunology*, *349*, 104061. https://doi.org/10.1016/j.cellimm.2020.104061

Costa, C. D. (2019, August 26). What Is Machine Learning & Deep Learning? *Medium*. https://medium.com/@clairedigitalogy/what-is-machine-learning-deep-learning-7788604004da

Gautreaux, M. D. (2017). Chapter 17 - Histocompatibility Testing in the Transplant Setting. In G. Orlando, G. Remuzzi, & D. F. Williams (Eds.), *Kidney Transplantation, Bioengineering and*

Regeneration (pp. 223–234). Academic Press. https://doi.org/10.1016/B978-0-12-801734-0.00017-5

Grinyo, J. M. (2013). Why Is Organ Transplantation Clinically Important? *Cold Spring Harbor Perspectives in Medicine*, *13*(11). https://doi.org/10.1101/cshperspect.a014985

Hewitt, E. W. (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology*, *110*(2), 163–169. https://doi.org/10.1046/j.1365-2567.2003.01738.x

Hunt, D., & Saab, S. (2012). Post–Liver Transplantation Management - ScienceDirect. In *Zakim and Boyer's Hepatology* (Sixth, pp. 869–882). https://www.sciencedirect.com/science/article/abs/pii/B9781437708813000498

Ingulli, E. (2010). Mechanism of cellular rejection in transplantation | Pediatric Nephrology. *Pediatric Nephrology*, *25*. https://doi.org/10.1007/s00467-008-1020-x

Kelly, J. (2022, April 27). *End of anti-rejection transplant drugs? A clinical trial at Hume-Lee hopes so.* VCU Health. https://www.vcuhealth.org/news/end-of-anti-rejection-transplant-drugs-a-clinical-trial-at-hume-lee-hopes-so

King, T. C. (2007). 2 - Inflammation, Inflammatory Mediators, and Immune-Mediated Disease. In T. C. King (Ed.), *Elsevier's Integrated Pathology* (pp. 21–57). Mosby. https://doi.org/10.1016/B978-0-323-04328-1.50008-5

Manski, C. F., Tambur, A. R., & Gmeiner, M. (2019). Predicting kidney transplant outcomes with partial knowledge of HLA mismatch. *Proceedings of the National Academy of Sciences*, *116*(41), 20339–20345. https://doi.org/10.1073/pnas.1911281116

Matching and Compatibility. (n.d.). UC Davis Health. Retrieved November 8, 2023, from https://health.ucdavis.edu/transplant/livingkidneydonation/matching-and-compatibility.html

Organ, Eye and Tissue Donation Statistics. (n.d.). Donate Life America. Retrieved November 8, 2023, from https://donatelife.net/donation/statistics/

Organ Rejection after Renal Transplant. (n.d.). Columbia Surgery. Retrieved November 8, 2023, from https://columbiasurgery.org/kidney-transplant/organ-rejection-after-renal-transplant

Understanding Transplant Rejection. (n.d.). Stony Brook Medicine. Retrieved November 8, 2023, from https://www.stonybrookmedicine.edu/patientcare/transplant/rejection

Vijayan, S., Sidiq, T., Yousuf, S., van den Elsen, P. J., & Kobayashi, K. S. (2019). Class I transactivator, NLRC5: a central player in the MHC class I pathway and cancer immune surveillance. *Immunogenetics*, *71*(3), 273–282. https://doi.org/10.1007/s00251-019-01106-z