

# Analyzing HLA Sequences to Predict Organ Rejection and Find Targets for Precise Immunosuppression

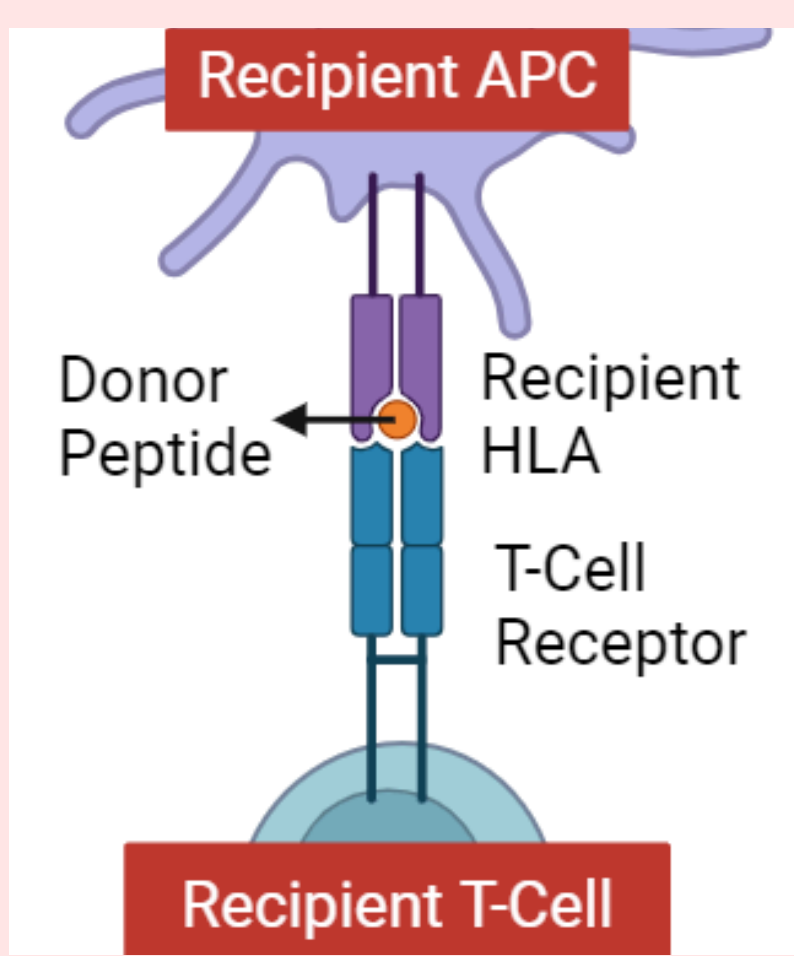
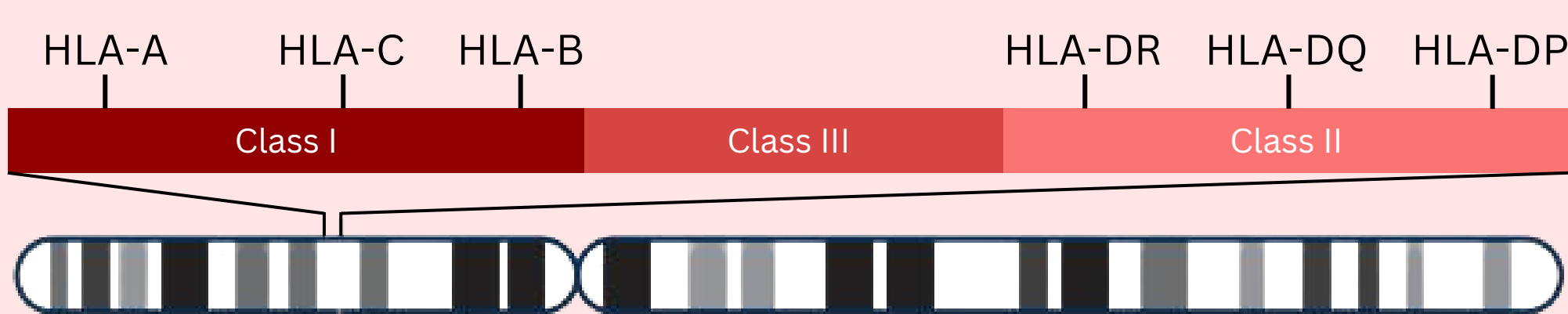
Samhitha Bodangi  
Advisor: Dr. Kevin Crowthers

## Background



- Organ Rejection is caused by the DNA or gene differences in the donated organ and the recipient's body
- Clinicians prescribe Immunosuppressors to prevent rejection, which are life-long drugs that weaken the immune system
- Side effect is a severely weakened Immune System, unable to protect body against other diseases or viruses
- Medications not as effective at preventing chronic rejection

## Indirect Allorecognition



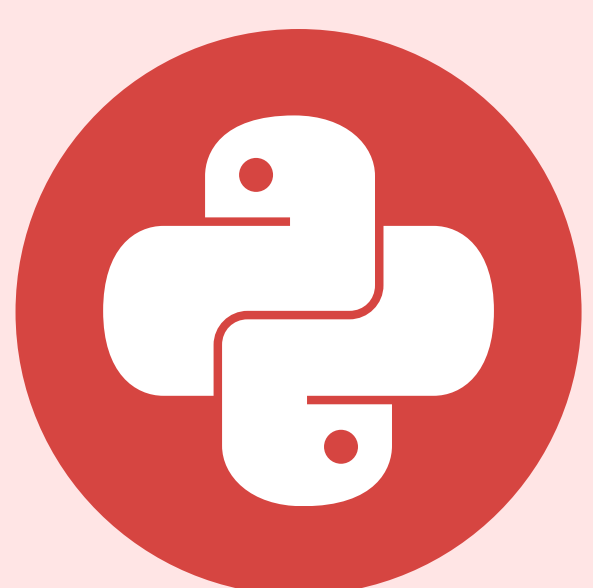
- Human Leukocyte Antigen (HLA) genes located on chromosome 6
- Genes code for the HLA surface molecules on immune cells (APC)
- Multiple allele variations in loci
- HLA peptide presentation:
- Recipient APC processes and presents donor-derived peptide
- T-cells recognize foreign peptide
- Initiates immune response

## Methodology



### 1. Data Collection and Analysis

- Collect HLA typing data from the past donor and the recipients, with the rejection outcome and match scores
- Obtain amino acid sequences for commonly typed HLA alleles



### 2. Machine Learning Model

- Develop models using sequences
- Use NetSurfP server to find solvent accessible amino acids
- NetMHCIIpan to generate donor-derived peptides and find strong binding, significant peptides



### 3. Model Testing and Revision

- Test algorithm iterations with HLA typing data and compatibility scores
- Analyze model's performance using accuracy metrics and decision matrix
- Compare model with current research models and findings

## Engineering Problem

Chronic organ rejection affects about 50% of all kidney transplants 5 years post-transplant.

Broad immunosuppressors can severely weaken the immune system, making the body unable to defend itself against other diseases.

## Engineering Goal

Make a model that identifies solvent-accessible amino acid mismatches and predict donor peptides that would bind to recipient HLA class II molecules to predict rejection and find targets for precise immunosuppression.

## Main Takeaways

- Peptides with solvent-accessible mismatches were predicted to be strong binding to recipient HLA class II molecules
- Greater number of solvent-accessible mismatches correlated with a higher number of strong-binding peptide targets
- Focusing on indirect MHC-peptide presentation can be beneficial in predicting precise immunosuppressive targets
- Certain HLA loci may have a greater influence on the rejection outcome and the compatibility score than others

## Rejection and Output Scores

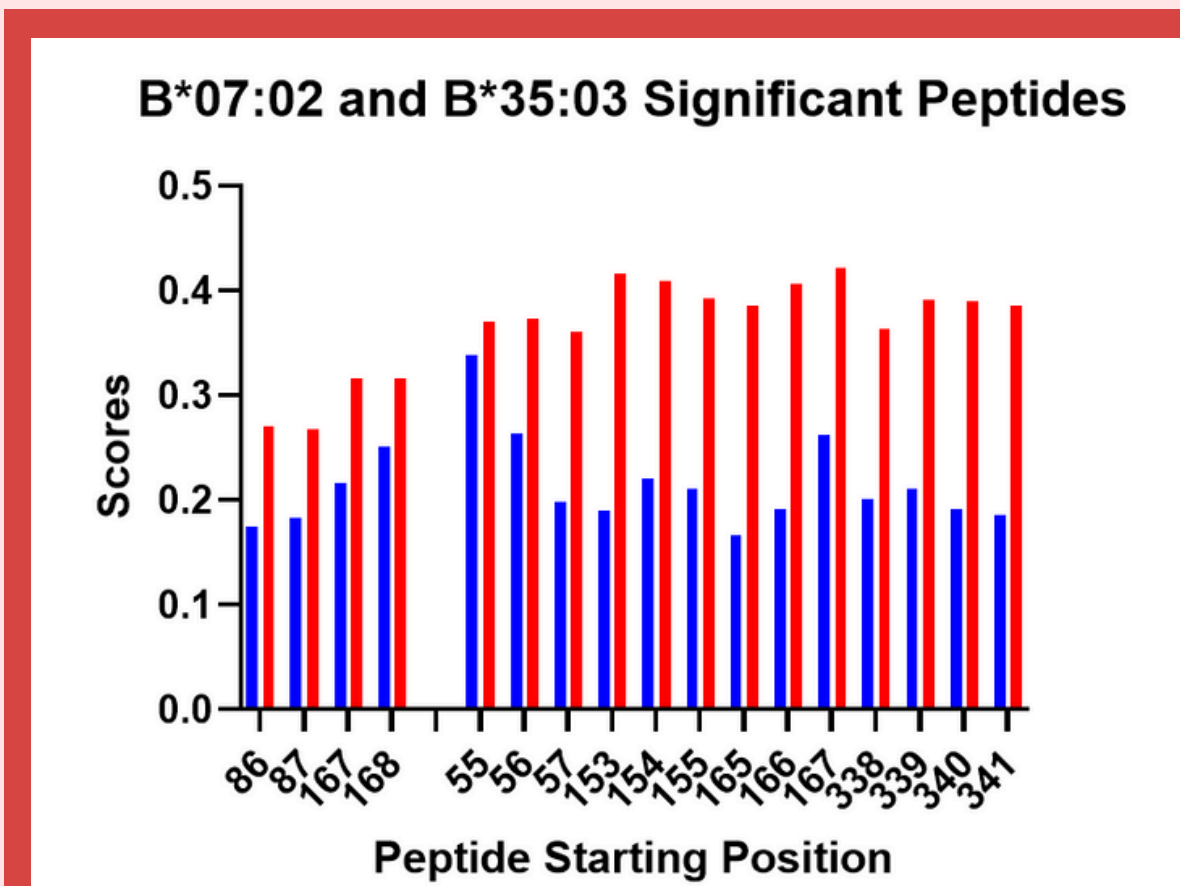


Figure 1: Sample output of significant peptides for donor alleles B\*07:02 and B\*35:03. Blue is binding affinity score and red is eluted ligand scores.

Info	Allele	33	91	93	94	95	138	176	180	202
Recipient	B*08:01	D	F	T	N	T	N	V	D	T
Recipient	B*40:02	H	S	T	N	T	N	V	L	T
Donor	B*07:02	Y	Y	A	Q	A	D	E	R	K
Total MM		9								
SA MM		3								

Info	Allele	33	48	69	118	119	121	127	138	140	155	187	218	306	329	349
Recipient	B*08:01	D	S	E	T	L	S	V	N	Y	R	T	I	V	A	C
Recipient	B*40:02	H	T	K	T	L	S	V	N	Y	R	E	I	V	A	C
Donor	B*35:03	Y	A	T	I	I	R	L	D	F	S	L	V	I	T	S
Total MM		14														
SA MM		7														

Figure 2: Sample output of amino acid mismatches and solvent-accessible mismatches with the amino acid position for donor alleles B\*07:02 and B\*35:03 compared to recipient alleles B\*08:01 and B\*40:02

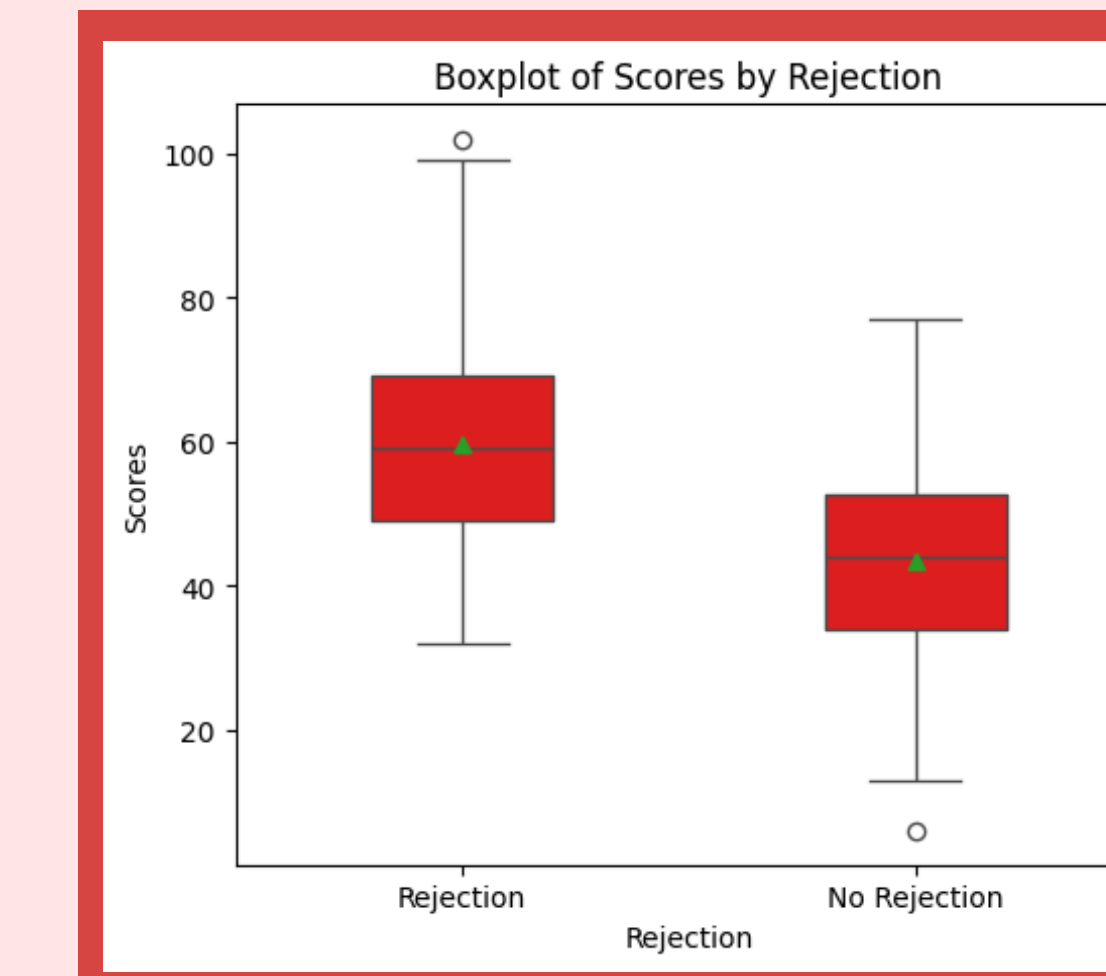


Figure 3: Box and Whisker Plot for Output Scores and Rejection

## Compatibility Score Regression Models

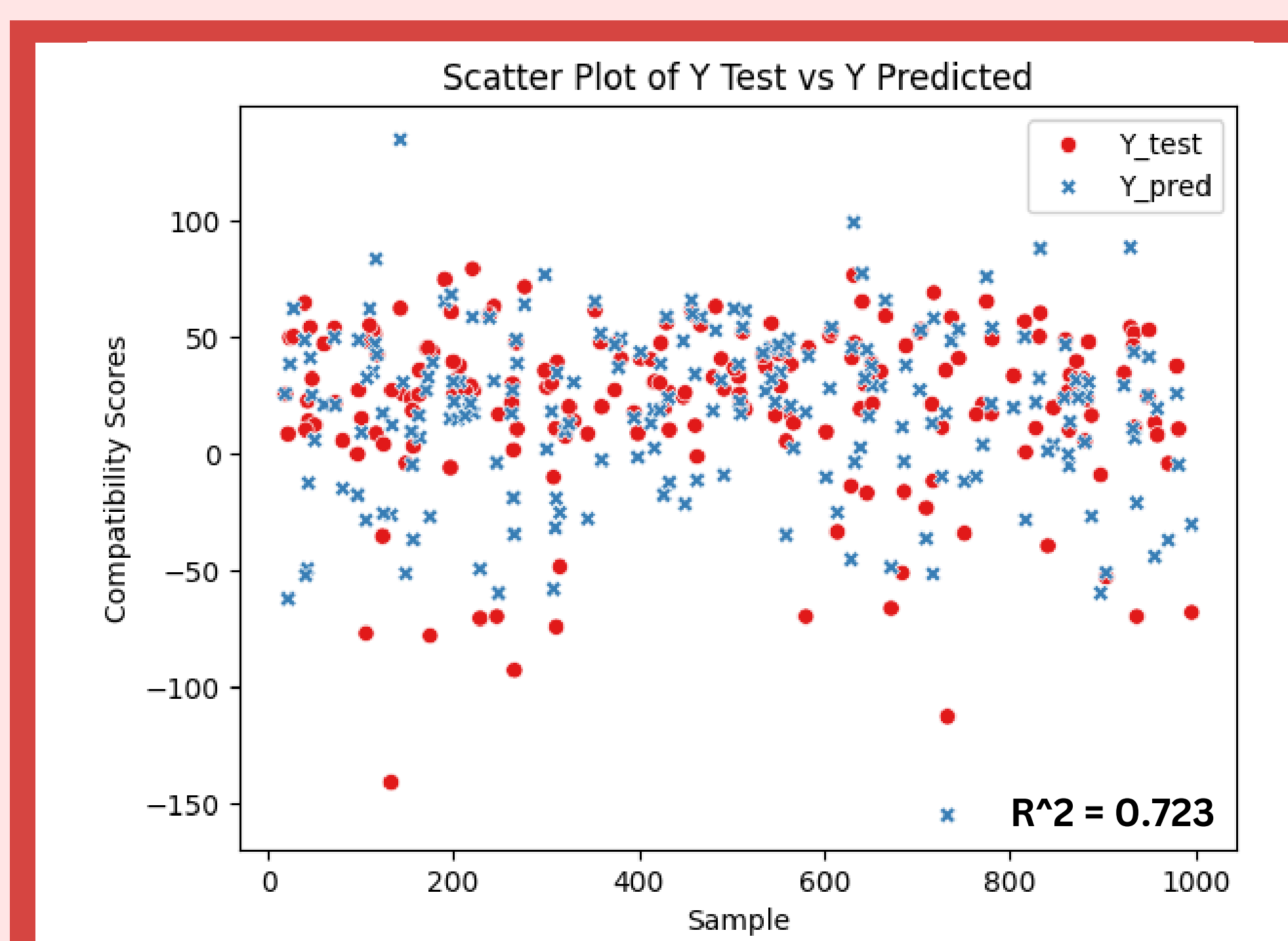
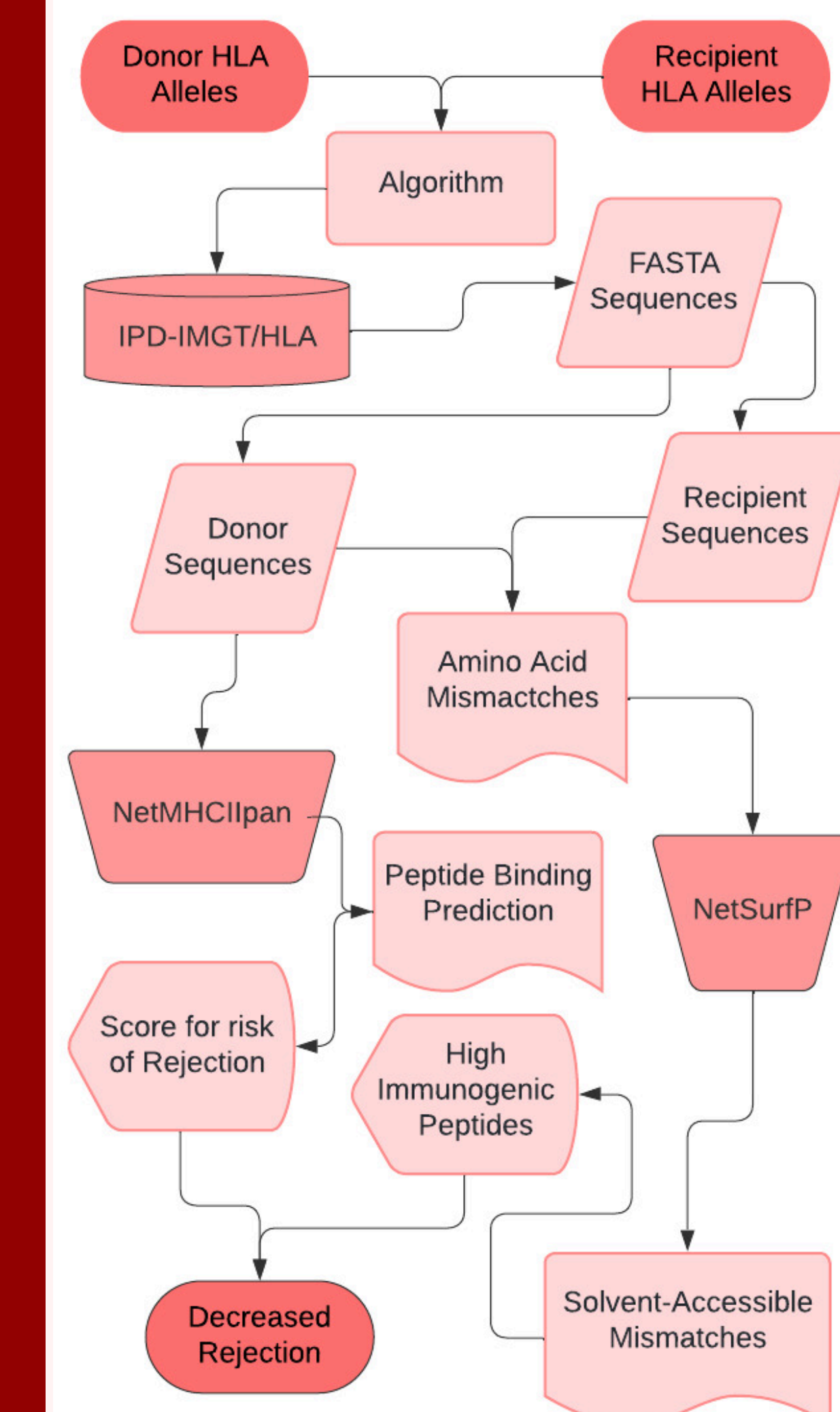


Figure 4: Scatter Plot for Ridge Regression Model after Feature Selection

Criteria	Linear Regression	Ridge Regression	Random Forest	Lasso Regression	Polynomial Regression
Mean Squared Error	487.682	485.305	531.052	488.244	485.564
Mean Absolute Error	17.028	16.782	15.068	16.967	16.886
Root Mean Squared Error	22.084	22.030	23.045	22.096	22.036
R Squared Value	0.624	0.626	0.590	0.623	0.625

Table 1: Decision Matrix for Compatibility Score Regression Models

## Flow Chart



## Materials



## Analysis

- Higher overall scores for rejection samples show a correlation between peptide targets and rejection
- Samples with more solvent-accessible mismatches had more strong-binding peptide targets
- Rejection groups were statistically significant, with a t-statistic of **6.269** and p-value of **1.646e-9**
- There was no statistically significant difference between the predicted scores and actual scores, with a t-statistic of **1.918** and a p-value of **0.057**
- Ridge regression performed the best, but outliers show potential for some HLA loci having more weightage

## Future Studies

- Create a web application that holds the model, giving clinicians more information on significant peptide targets and the risk of rejection
- Improve model by investigating which loci are the greatest indicators of rejection and assigning weights
- Modify the model to analyze rejection risk and targets for other organs, such as heart or lung transplants
- Experiment with other prediction software for most efficient model predictions