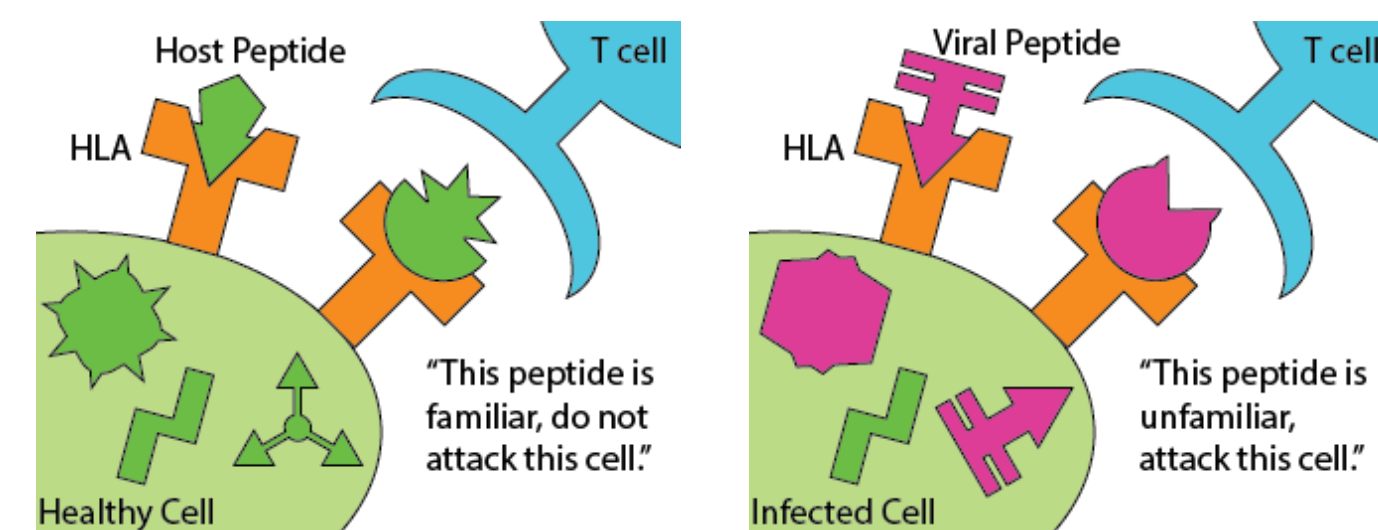


Modeling HLA ligands for binding prediction of new peptides

Alex Han

Introduction

Human leukocyte antigens (HLA) are a family of proteins that are an integral part of the human immune system. They serve as “windows” for cells in the human body. HLA bind short pieces of proteins, or peptides, from the cell interior and present them at the cell exterior. Host human peptides presented at the HLA show that a cell is healthy; bacterial or viral peptides presented by an infected cell show that a cell is unhealthy.



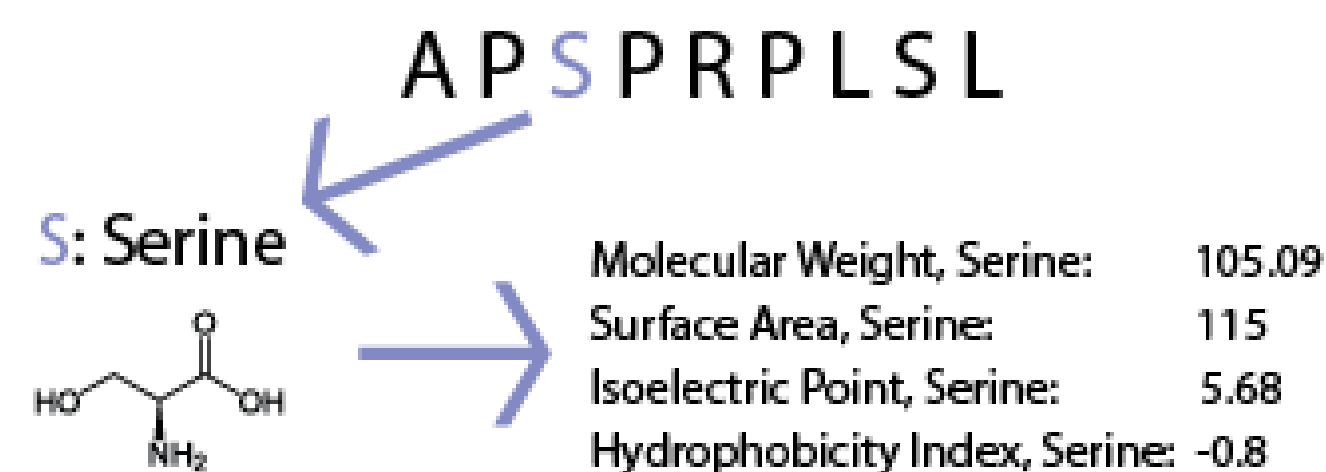
My goal in this project was to

- Model how different HLA have different binding affinities for different types of peptides.
- Use these models to predict whether new pathogenic peptides may bind to an HLA.

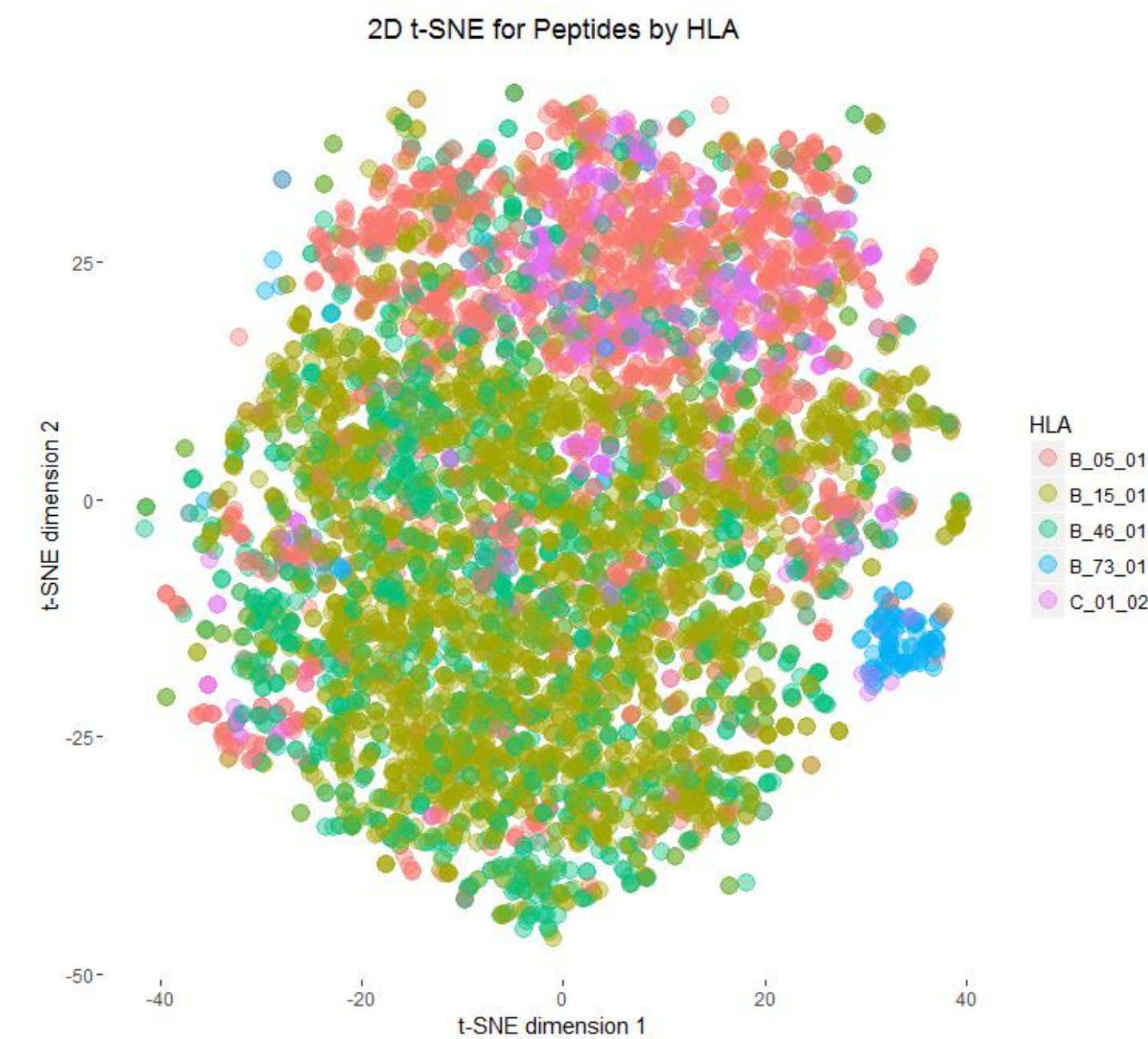
Data

- 9 amino acid length peptides, or nonamers, which were shown to bind to 5 different HLA (*positive class*)
- Randomly generated nonamers (*negative class*)
- Proteome of the H1N1 Influenza A virus, cut into all possible nonamers (*for prediction*)

All nonamer sequence strings were transformed into 4 physiochemical properties at each amino acid to make 36 features per peptide.



Visualization

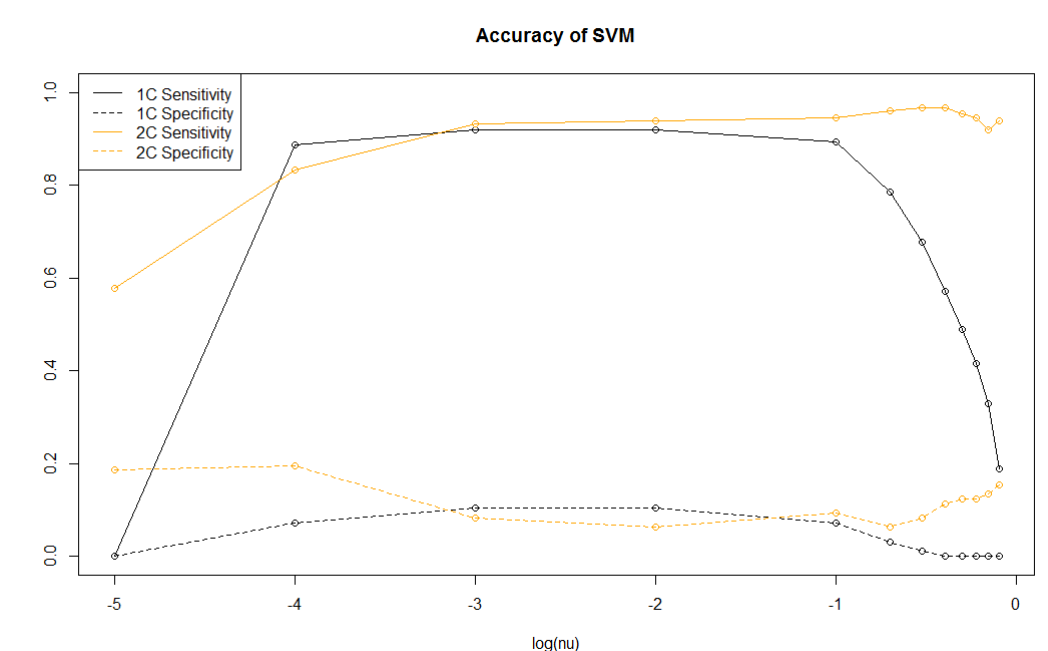
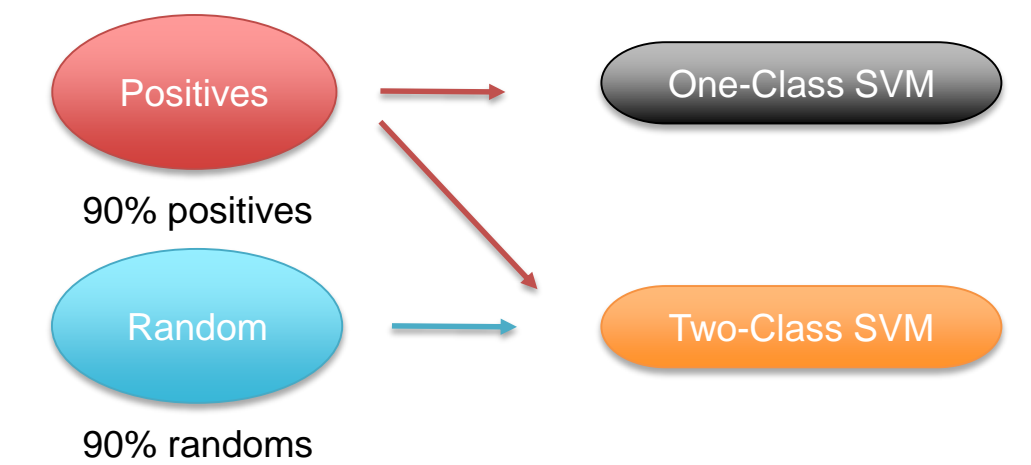


KL distances

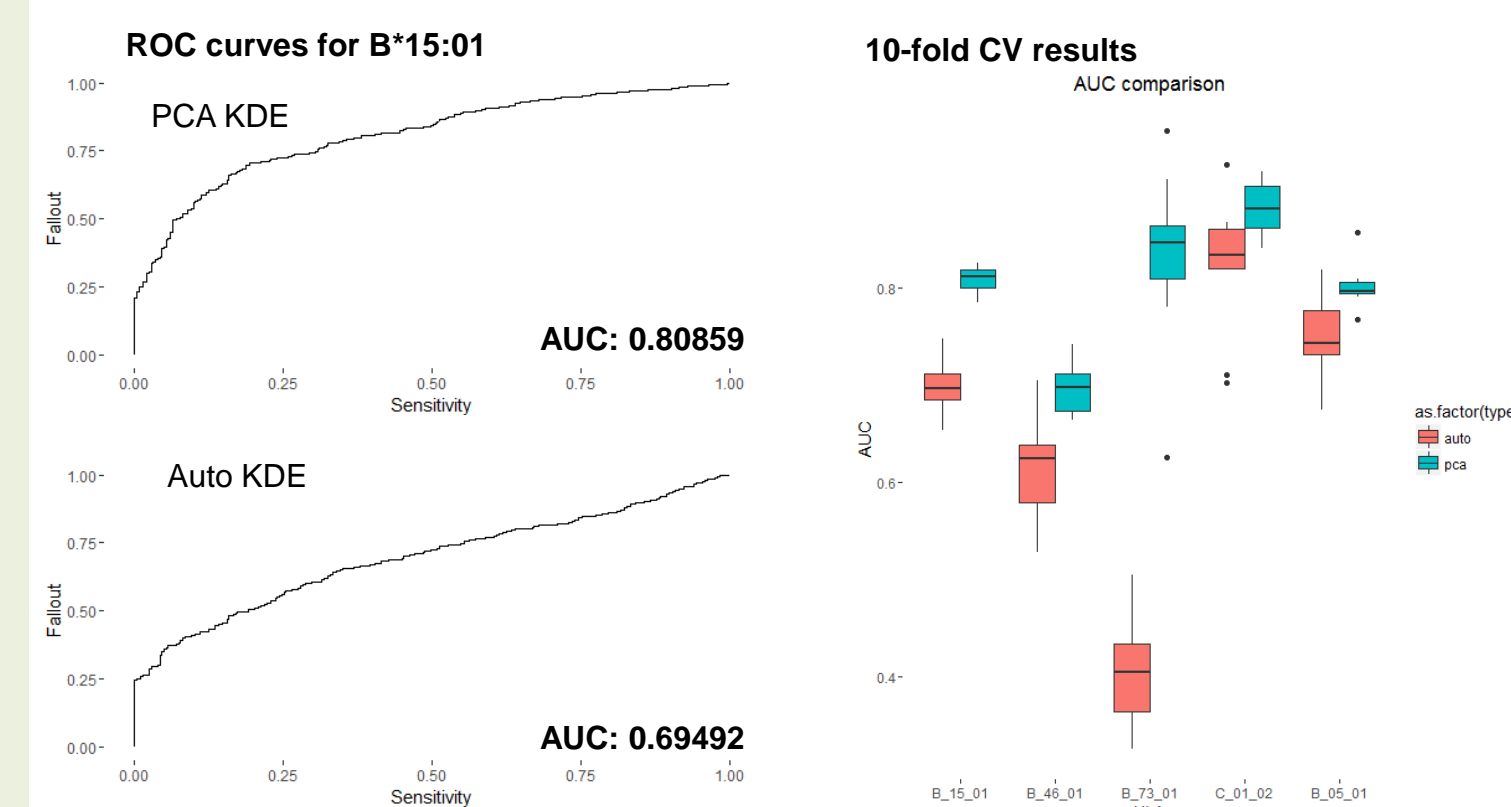
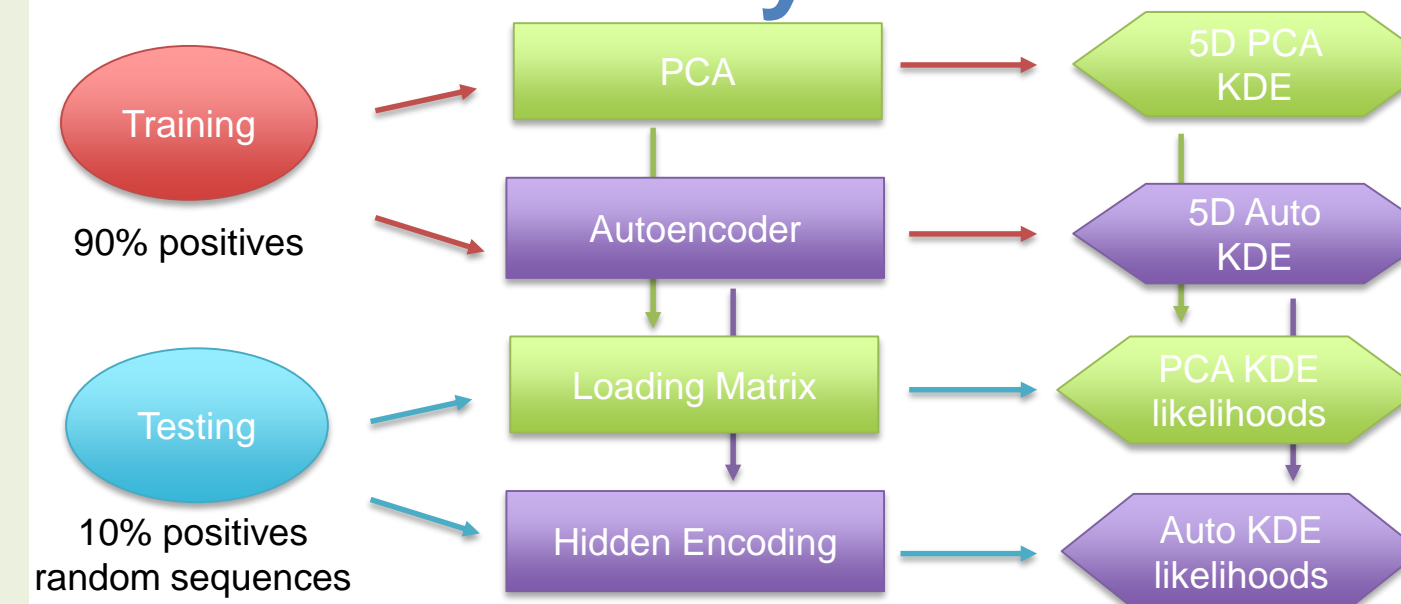
PCA	B*05	B*15	B*46	B*73	C*01
B*05	0	0.51	0.24	0.40	0.83
B*15	0.37	0	0.05	0.24	0.53
B*46	0.47	0.18	0	0.29	0.58
B*73	0.19	0.18	0.09	0	0.30
C*01	0.25	0.15	0.07	0.17	0

auto	B*05	B*15	B*46	B*73	C*01
B*05	0	13.8	12.3	10.8	12.2
B*15	9.69	0	2.62	4.68	1.80
B*46	29.7	14.2	0	13.9	10.3
B*73	540	401	402	0	335
C*01	80.9	39.9	31.5	22.9	0

One- & Two-Class v-SVM



Kernel Density Estimation



Prediction with H1N1 (swine flu)

A small fraction of viral sequences were obtained from an assay involving a cell with HLA B*05:01 infected by a strain of Influenza Virus A subtype H1N1. All four models were recreated using all of the data and optimal parameters for total accuracy before using them to predict whether the sequences would be obtained or not.

Sensitivity: # binders assigned as binders / # binders
Fallout: # nonbinders assigned as nonbinders / # nonbinders

	5D PCA KDE	5D Auto KDE	One-Class SVM	Two-Class SVM
Sensitivity			1	1
Specficity			0.1002	0.1536

Acknowledgments

I'd like to thank Curtis McMurtrey and William Hildebrand for the data, and the Parham Lab and the Bustamante Lab for advice, in particular, Hugo Hilton and Elena Sorokin.