Predicting pMHC-I Binding from LC-MS/MS Data
Using Hidden Markov Models

Jordan Force

August 23, 2017
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Background</td>
<td>Predicting pMHC-I Binding from LC-MS/MS Data</td>
</tr>
<tr>
<td><strong>2</strong> Data</td>
<td></td>
</tr>
<tr>
<td><strong>3</strong> Prediction Methods</td>
<td></td>
</tr>
<tr>
<td><strong>4</strong> Future Work</td>
<td></td>
</tr>
</tbody>
</table>
# Table of Contents

1. **Background**

2. **Data**

3. **Prediction Methods**

4. **Future Work**
MHC-I Function

Figure: The function of the MHC-I
A More Memorable Diagram

Figure: The function of the MHC-I, with chicken
Measuring Affinity

1. Produce MHC-I Protein
2. Fold it \textit{in-vitro} with peptide
3. Measure stability of peptide-MHC complex
4. Or competition with a standard known binder
Disadvantages to this Approach

1 Binding not done in normal environment
Note: This is a work in progress!
# Table of Contents

1. Background
2. Data
3. Prediction Methods
4. Future Work
MS/MS Data Source

1. Data came from Abelin et al.
2. Used cell lines expressing a single HLA allele
3. Sequenced presented peptides (from endogenous proteins) using LC-MS/MS
Improvements

Figure: Improvement
Predicting pMHC-I Binding from LC-MS/MS Data

Background

Figure: How Mass Spec Works
# Table of Contents

1. Background
2. Data
3. Prediction Methods
4. Future Work
PWM

1. Construct a matrix with frequency of amino acid at each position
2. Compute likelihood of a given peptide under this matrix
Left-to-right HMM, with $n$ rows, and $w$ columns. $w$ is length of peptides.

Figure: HMM with $w = 4$ and $n = 2$
HMM Training

1. Train HMM *only* on the binder dataset
2. Used EM algorithm included in *hmmlearn* package
3. Ran EM algorithm 10 times, picked parameters that gave highest likelihood.
HMM Scoring

1. Compute probability $\log_2(P(x_1x_2\ldots x_w))$ of sequence under HMM model
2. Adjust classification threshold and generate ROC curve
(some) Results

1. Generated random peptides, uniform distribution of amino acids
Predicting pMHC-I Binding from LC-MS/MS Data

Jordan Force

Result for peptides of length 9, HLA-A0101

HLA: HLA-A0101, Num positives: 447
length: 9

Fold 0, AUC: 1.00
Fold 1, AUC: 0.99
Fold 2, AUC: 1.00
Fold 3, AUC: 1.00
Fold 4, AUC: 1.00
1. **Selected random peptides from human proteome**
Result for peptides of length 9, HLA-A0101
(some) Results

1. Used peptides that were presented by HLA-A0201, HLA-A0203, HLA-A0204 and HLA-A0207, but not HLA-A0101, as negative

2. Still performed very well
Result for peptides of length 9, HLA-A0101
UConn recently bought an MS/MS system appropriate for protein work; will generate data for mouse MHC alleles
Flanking Residues

Predicting pMHC-I Binding from LC-MS/MS Data

Jordan Force

Background
Data
Prediction Methods
Future Work
### Flanking Residues (continued)

<table>
<thead>
<tr>
<th>Flanking Residues</th>
<th>Downstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNTSRKHPAGEDCMLVYFW</td>
<td>D1 D2 D3 D4 D5 D6 D7 D8 D9 D10</td>
</tr>
<tr>
<td></td>
<td>+11 +2 0 +1 -2 0 -1 +3 +0</td>
</tr>
<tr>
<td></td>
<td>+17 +16 +13 +9 +9 -4 +7 +5 +4</td>
</tr>
<tr>
<td></td>
<td>-5 -8 +1 0 -4 +1 -6 -8 -4</td>
</tr>
<tr>
<td></td>
<td>+18 -6 -4 -10 -7 -10 -10 -7 -10</td>
</tr>
<tr>
<td></td>
<td>+27 +13 +0 +1 +1 -2 -5 -2 -4</td>
</tr>
<tr>
<td></td>
<td>+51 +29 +13 +11 +8 +10 +10 +9 +8</td>
</tr>
<tr>
<td></td>
<td>+21 -17 +4 +5 -1 -2 +0 -12 -3</td>
</tr>
<tr>
<td></td>
<td>-41 -24 -18 +1 -3 -6 -7 -6 -3</td>
</tr>
<tr>
<td></td>
<td>+42 +2 +8 +6 -2 -1 -1 -2 0</td>
</tr>
<tr>
<td></td>
<td>+6 +17 +8 +2 +4 +7 +5 -2 0</td>
</tr>
<tr>
<td></td>
<td>-20 -10 +9 -1 +9 +7 +4 +7</td>
</tr>
<tr>
<td></td>
<td>-18 +7 +13 +6 +8 +15 +12 +9 +1 +12</td>
</tr>
<tr>
<td></td>
<td>+1 -12 -13 +15 -17 +24 -20 +17</td>
</tr>
<tr>
<td></td>
<td>+17 +16 -3 +1 -5 +3 +0 +5 +10 +2</td>
</tr>
<tr>
<td></td>
<td>-23 +3 +14 -12 -10 -5 -6 +0 +2</td>
</tr>
<tr>
<td></td>
<td>-10 -8 -8 +7 0 -1 +4 +6 +7</td>
</tr>
<tr>
<td></td>
<td>-41 -2 -2 -1 +7 +2 +10 +5 +2</td>
</tr>
<tr>
<td></td>
<td>-19 -20 +6 +9 +3 -3 +3 +4 +2 +2</td>
</tr>
<tr>
<td></td>
<td>-27 -23 +3 -13 -6 -8 +3 +4</td>
</tr>
<tr>
<td></td>
<td>-55 -24 +3 -22 -19 -21 -8</td>
</tr>
<tr>
<td></td>
<td>+271 +114 +79 +61 +47 +39 +30 +26 +21 +20</td>
</tr>
</tbody>
</table>
Higher Order Models

Allow for relationships between residues that are non-adjacent
Mass spectrometry profiling of hla-associated peptidomes in mono-allelic cells enables more accurate epitope prediction.