Immunoinformatics resources for the understanding of immunological information

A case study in personalized cancer immunotherapy

Bjoern Peters
La Jolla Institute for Allergy and Immunology
La Jolla Institute for Allergy and Immunology (LIAI)
Overview

• Part I - Lecture: Biological Background
  – T cell immune responses target non-self entities
  – Cancer cells bear somatic mutations
  – Cancer immunotherapy aims to target immune responses to cancer cells

• Part II – Lecture: Bioinformatic guided approaches
  – Sequencing approaches identify tumor specific somatic mutations
  – HLA binding predictions can identify which of these will be immunogenic

• Part III – Hands on session: Design a personalized cancer vaccine
HLA molecules as sensors of non-self

HLA = Human MHC molecules
CD8⁺ T cell epitopes in viral infection

Mouse

Virus

MHC-I

APC

Hexagon shapes represent viral antigens.
CD8⁺ T cell epitopes in viral infection

- How do peptides get loaded on MHC molecules?
- How do T cells distinguish self- from non-self peptides?
MHC I - Antigen processing and presentation pathway

MHC:peptide binding mode

• Each human has 6 types of MHC molecules (alleles)
• >3000 alleles are known
• Distinct binding specificities → individual epitope repertoire

X-Ray Structure: Madden, Cell 1993.
Viewer: Beaver and Ponomarenko, Immunome Research, 2007
Self–reactive T cells are deleted during maturation

The repertoire of T cells is shaped by both positive and negative selection

Expert Reviews in Molecular Medicine © 1999 Cambridge University Press
Background: Cancer
What is cancer?

- All cancers derive from single cells that have acquired the characteristics of continually dividing in an unrestrained manner and invading surrounding tissues.
- Cancer cells behave in this abnormal manner because of changes in the DNA sequence of key genes, which are known as cancer genes. Therefore all cancers are genetic diseases.
What is a mutation?

• **Germline mutation**
  – A change in the DNA sequence that can be inherited from either parent

• **Somatic mutation**
  – A change in the DNA sequence in cells other than sperm or egg
  – The mutation is present in the cancer cell and its offspring, but not in the patient’s healthy cells
Mutations & cancer genes

- Cancer genes are causally implicated in oncogenesis.
- Mutations in cancer genes can occur somatically or can be inherited.
- Mutations in some cancer genes can be inherited from parents, in which case they are present in every cell of the body. Such people are at a higher risk of developing cancer.
- Somatic mutations can occur in any of the cells of the body except the germ cells (sperm and egg) and therefore are not passed on to children.
Importance of somatic DNA changes in human cancer

Only 5–10% of cancer cases have a clear hereditary component, e.g. BRCA1 and BRCA2 in breast cancer.

Even in those cases where susceptibility is clearly inherited, somatic changes are required for cancer to develop.
Examples of mutations

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTCGTTAGGCA</td>
<td>ACTCCTTCTAGGCA</td>
<td>Substitution</td>
</tr>
<tr>
<td>ACTCGTTAGGCA</td>
<td>ACTCGGCA</td>
<td>Deletion</td>
</tr>
<tr>
<td>ACTCGTTAGGCA</td>
<td>ACTCGTATCAGGCA</td>
<td>Insertion</td>
</tr>
<tr>
<td>ACTCGTTAGGCA</td>
<td>ACTCTTGCAGGCA</td>
<td>Inversion</td>
</tr>
<tr>
<td>ACTCGTTAGGCA</td>
<td>ACTCGTTAGGCA</td>
<td>Duplication</td>
</tr>
</tbody>
</table>
Cancer progression

Mutations in multiple cancer genes are required for the development and progression of a single cancer.

- Benign Tumour
- \textit{In situ} cancer
- Invasive cancer
- Metastatic cancer
Neoepitopes (Neoantigens)

- Cancers genomes accumulate mutations
- Mutations in coding regions are translated in mutated protein sequences
- Mutated peptides can be presented as epitopes on MHC to T cells

**Neoepitopes** are presumably recognized by tumor-infiltrating lymphocytes (TILs)

**Neoepitopes** are highly tumor-specific!

Schumacher & Schreiber, Science. 2015 Apr 3;348(6230):69-74
Cancer Immunotherapy

• **Vaccination**: Introduce or boost an immune response against a specific target (antigen)

• Cancer cells contain non-self antigens that *could* be recognized by T cells, but presence of cancer means this mechanism has failed, typically by the tumor suppressing immune responses

• **Checkpoint blockade treatments**: Block immune suppressive mechanisms to boost T cell immune responses against cancer cells.

• Problem: Checkpoint blockade is unspecific, and will also boost unwanted autoimmune responses

• **Personalized Cancer Immunotherapy**: Boost anti-tumor response with vaccine containing peptides corresponding to cancer mutations that can be recognized by T cells.

→ How can such a vaccine be designed?
Overview

• Part I - Lecture: Biological Background
  – T cell immune responses target non-self entities
  – Cancer cells bear somatic mutations
  – Cancer immunotherapy aims to target immune responses to cancer cells

• Part II – Lecture: Bioinformatic guided approaches
  – Sequencing approaches identify tumor specific somatic mutations
  – HLA binding predictions can identify which of these will be immunogenic

• Part III – Hands on session: Design a personalized cancer vaccine
DNA and RNA sequencing identifies tumor specific somatic mutations

Which mutations can be recognized by the patient’s T cells?
→ Resulting peptides have to bind HLA molecules of the patient
HLA Typing: Targeted sequencing of HLA locus

DNA Isolation

PCR Primary Amplification (exons 1-5)

PCR Primary Amplification
Product Purification

Sequencing Reactions (forward & reverse orientations)

Sequencing Reaction
Precipitation

Utilization of 96 sample sequencing instrument

Sequencing Analysis

*http://www.ashi-hla.org/publicationfiles/ASHI_Quarterly/25_2_2001/highthrusbt3.htm*
Measuring and predicting MHC:peptide binding

**Experimental Basis: MHC Binding Assay**

- List of peptides with allele specific binding affinity

<table>
<thead>
<tr>
<th>Sequence</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIVTMFEAL</td>
<td>3.6</td>
</tr>
<tr>
<td>LKGPDYIKG</td>
<td>308</td>
</tr>
<tr>
<td>NFCNLTSAF</td>
<td>50,000</td>
</tr>
<tr>
<td>AQSQCRFTR</td>
<td>38,000</td>
</tr>
<tr>
<td>CTAGPFGM</td>
<td>143</td>
</tr>
<tr>
<td>CFGNTAVAK</td>
<td>50,000</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

log(IC<sub>50</sub>) ~ Binding free Energy

low IC<sub>50</sub> → high affinity

**Impossible to measure all peptides**

→ Predict binding peptides using machine learning

Find function F<sub>i</sub> in F<sub>i</sub> (Sequence) ≈ Affinity

Many different approaches (ANN, SVM, HMM, LP, ...)

**T cell epitope mapping**

<table>
<thead>
<tr>
<th>ORF</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MGQIVTMFEALPHIDEVINIVILIVITGIKAVYN...</td>
</tr>
<tr>
<td>2</td>
<td>MLKGPDYIKGVYQFKSVEDMSHNLNTMPNACSAN...</td>
</tr>
<tr>
<td>3</td>
<td>MHNFCNLTSAFNKKTFDHTLMSIVSSSLHLSIDGN...</td>
</tr>
<tr>
<td>4</td>
<td>MSAQSCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSD...</td>
</tr>
<tr>
<td>5</td>
<td>MHCTYAGPFGMSRILLSQETKTFTRRLAGTFWTLSS...</td>
</tr>
<tr>
<td>6</td>
<td>MKCFGTAVAKCNYHDAEFCDMMLRLLIDYNAALSKF...</td>
</tr>
<tr>
<td>7</td>
<td>MLMRNHLDDLMGVYPYCNYSKFWYLEHAKTGETSPKC...</td>
</tr>
</tbody>
</table>
Machine learning PSSM = Minimize the difference between predicted and measured binding affinities by varying the matrix values

N peptides with measured binding affinities

<table>
<thead>
<tr>
<th>log (IC50)</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>FQPQNGSFI</td>
</tr>
<tr>
<td>0.72</td>
<td>ISVANKIYM</td>
</tr>
<tr>
<td>2.37</td>
<td>RVYEALYYV</td>
</tr>
<tr>
<td>3.42</td>
<td>FQPQSQGQFI</td>
</tr>
<tr>
<td>3.46</td>
<td>LYEKVSQL</td>
</tr>
<tr>
<td>4.07</td>
<td>FKSVEFDMS</td>
</tr>
<tr>
<td>4.18</td>
<td>FQPQNGQFH</td>
</tr>
<tr>
<td>4.24</td>
<td>VLMLPVWFL</td>
</tr>
<tr>
<td>4.39</td>
<td>YMTLGQVVF</td>
</tr>
<tr>
<td>4.40</td>
<td>EDVKNAVGV</td>
</tr>
<tr>
<td>4.90</td>
<td>VFYEQMKRF</td>
</tr>
</tbody>
</table>

Offset: 4.3
Predictions available as webserver

- Immune Epitope Database (IEDB) Analysis resource
- http://tools.iedb.org/mhci/
### MHC-I Binding Predictions

#### Prediction Method Version
- **2013-02-22** [Older versions]

#### Specify Sequence(s)
- Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. *(Browse for sequences in NCBI)*

- Or select file containing sequence(s): **Choose File** - No file chosen

- Choose sequence format: **auto detect format**

#### Choose a Prediction Method
- **IEDB recommended** - [Help on prediction method selections](#)

#### Specify what to make binding predictions for
- **MHC source species**: **human**
- **Show only frequently occurring alleles**: ✔️

#### Specify Output
- **Sort peptides by**: **Percentile Rank**
- **Show**: **All predictions**
- **Output format**: **XHTML table**
- **Email address (optional)**

---

*Note: The image contains a web interface for the MHC-I Binding Predictions tool, allowing users to input sequences and select prediction methods.*
<table>
<thead>
<tr>
<th>Specify Sequence(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. <em>(Browse for sequences in NCBI)</em></td>
</tr>
<tr>
<td>Region 1: SPLPSQAMLDLMLSPDD</td>
</tr>
<tr>
<td>Region 2: DPGPDEAPWPEAAPPV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Or select file containing sequence(s)</th>
<th>Choose File</th>
<th>No file chosen</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Choose sequence format</th>
<th>auto detect format</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Choose a Prediction Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Method: IEDB recommended</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specify what to make binding predictions for</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC source species: human</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Show only frequently occuring alleles:</th>
<th>Allele</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select MHC allele(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Select HLA allele reference set:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specify Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sort peptides by: Percentile Rank</td>
</tr>
<tr>
<td>Show: All predictions</td>
</tr>
<tr>
<td>Output format: XHTML table</td>
</tr>
<tr>
<td>Email address (optional):</td>
</tr>
</tbody>
</table>

© 2005-2017 | IEDB Home
Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)

OR select file containing sequence(s) Choose File No file chosen

Choose sequence format auto detect format

Choose a Prediction Method

Prediction Method IEDB recommended Help on prediction method selections

Specify what to make binding predictions for

MHC source species human

Show only frequently occurring alleles: ✔️
Select MHC allele(s)
Select HLA allele reference set: ❌

Sort peptides by Percentile Rank 10 11 12 13 14

Show All predictions

Output format XHTML Table All lengths

Email address (optional)
### Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. *(Browse for sequences in NCBI)*

```
>Region 1
SPLPSQAMLDMLSPDD
>Region 2
DPGPDEAPWMPFEEAPPV
```

Or select file containing sequence(s)  
Choose File  No file chosen

Choose sequence format  
auto detect format

### Choose a Prediction Method

Prediction Method  IEDB recommended  Help on prediction method selections

### Specify what to make binding predictions for

MHC source species  
human

Show only frequently occurring alleles: 
Select MHC allele(s)

Select HLA allele reference set:  

<table>
<thead>
<tr>
<th>Allele</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A*02:01</td>
<td>9</td>
</tr>
</tbody>
</table>

### Specify Output

Sort peptides by  
Percentile Rank

Show  
All predictions

Output format  
XHTML table

Email address (optional)  

[Submit]  [Reset]
MHC-I Binding Prediction Results

Input Sequences

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reg 1</td>
<td>SPLPSQAMLDDLMLSPDD</td>
</tr>
<tr>
<td>2</td>
<td>Reg 2</td>
<td>DPGPDEAPWMPEAAPPV</td>
</tr>
</tbody>
</table>

Prediction method: IEDB recommended | Low percentile_rank = good binders

Download result

Citations
Check to expand the result:

<table>
<thead>
<tr>
<th>Allele</th>
<th>#</th>
<th>Start</th>
<th>End</th>
<th>Length</th>
<th>Peptide</th>
<th>Method used</th>
<th>Percentile_rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A*02:01</td>
<td>2</td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>WMPEAAPPV</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>0.4</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>8</td>
<td>16</td>
<td>9</td>
<td>MLDMLMLSPD</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>2.9</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>7</td>
<td>15</td>
<td>9</td>
<td>AMLDMLMLSP</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>4.0</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>5</td>
<td>13</td>
<td>9</td>
<td>SQAMLDMLML</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>7.7</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>6</td>
<td>14</td>
<td>9</td>
<td>QAMLDMLMS</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>26.0</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>2</td>
<td>5</td>
<td>13</td>
<td>9</td>
<td>DEAPWMPEA</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>32.0</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>SPLPSQAML</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>33.0</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>9</td>
<td>LPSQAMLSDL</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>39.0</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>PSQAMLDLM</td>
<td>Consensus (ann/comlib_sidney2008/smm)</td>
<td>43.0</td>
</tr>
</tbody>
</table>
Evaluating binding predictions

• Percentile rank < 0.5% = high affinity binder
• Percentile rank 0.5%-1% = intermediate binder
• Percentile rank 1% - 2% = low affinity binder
• Percentile rank 2% - 5% = borderline
• Percentile rank >5% is a non-binder
Overview

• Part I - Lecture: Biological Background
  – T cell immune responses target non-self entities
  – Cancer cells bear somatic mutations
  – Cancer immunotherapy aims to target immune responses to cancer cells

• Part II – Lecture: Bioinformatic guided approaches
  – Sequencing approaches identify tumor specific somatic mutations
  – HLA binding predictions can identify which of these will be immunogenic

• Part III – Hands on session: Design a personalized cancer vaccine
Input data from actual patient

>P53_HUMAN Cellular tumor antigen p53 - Healthy Tissue
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAAPAAPTAPAAPAPAPSWPLSSSVPSQKTYQGSYGFRILGFLHSGTAK
SVTCTYSPALNMFCQLAKTCPVQLWVDSTPPPGTRVRAIAMYKQSQHMTEVVRRCPHHE
RCSDSDGAPPQHLIRVEGNLRVEYLDHDRNTFRHSVVVYPYEPEVGSCTTIIHYNYMCNS
SCMGGMNRRPILTIIITLESSGNNLGRIDNSFEVRCAPGDRDRRTEEENLRKKGEHPHELP
PGSTKRALPNNNTSSSPQPKKPPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKQSTSRRHKKLPMKTEGPDS

>P53_HUMAN Cellular tumor antigen p53 - Tumor Tissue
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPWMPEAAPPVAPAAPAAPTAPAAPAPAPSWPLSSSVPSQKTYQGSYGFRILGFLHSGTAK
SVTCTYSPALNMFCQLAKTCPVQLWVDSTPPPGTRVRAIAMYKQSQHMTEVVRRCPHHE
RCSDSDGAPPQHLIRVEGNLRVEYLDHDRNTFVHSVVVYPYEPEVGSCTTIIHYNYMCNS
SCMGGMNRRPILTIIITLEV

HLA typing results:
HLA-A*02:01, HLA-A*68:01
HLA-B*07:02, HLA-B*35:01
Steps

• Step 1: Identify sequence regions that contain all 9-mer peptides that are only found in the tumor
• Step 2: Run HLA binding prediction to identify 9-mer peptides in the sequence regions unique to the tumor that can be presented to T cells
• Step 3: Select the top peptide for each HLA allele
• Step 4: What is the un-mutated form of the chosen peptides in the patient? What is their MHC binding affinity?
• Step 5: Are the peptides really specific for the tumor? Examine this using NCBI BLAST
• Step 6: Decide: Which peptide would you choose?
backup
Cancer genes

- There are two types of cancer genes:
  - Tumour suppressor genes
  - Oncogenes

- To date, we know of approximately 400 somatic “cancer genes” * but there are almost certainly more to be found

- COSMIC is a catalogue of somatic mutations found in cancer genes in human tumours and is available at: http://www.sanger.ac.uk/genetics/CGP/cosmic/

*(COSMIC v47release. July 2010)
Tumour suppressor gene

These genes normally function to PREVENT cell growth/division
Oncogene

Genes which normally function to PROMOTE cell growth/division in a controlled manner

Ras