Prediction of B-Cell Epitopes

14 February 2019 / Maximilian Collatz
The problem

• Unknown vaccination status of patients
• Vaccination very expensive
  ⇒ Testing vaccination status is less expensive but still far from cheap
  ⇒ Time consuming
The vision

- One chip for all
- All genes of the STIKO list
- Less expensive and much faster

STIKO – Ständige Impfkommission
(Standing Committee on Vaccination)
### STIKO list December 2018

- Influenza
- Tetanus
- Measles
- Mumps
- HPV
- ...

#### Table 1: Immunisation schedule (vaccinations) for infants, children, adolescents and adults

<table>
<thead>
<tr>
<th>Vaccination against</th>
<th>Age in weeks</th>
<th>Age in months</th>
<th>Age in years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Tetanus (T)</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Diphtheria (D/d)</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Pertussis (aP/aP)</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td><em>H. influenzae b</em> (Hib)</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Poliomyelitis (IPV)</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Hepatitis B (HB)</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Pneumococcus A</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Rotavirus (RV)</td>
<td>P1</td>
<td>P2</td>
<td></td>
</tr>
</tbody>
</table>

- Meningococcal C   | P1 (from 12 months) |       | C |     |      |      |
- Measles            | P1 | P2  |       | C |      |      |
- Mumps, Rubella     | P1 | P2  |       | C |      |      |
- Varicella          | P1 | P2  |       | C |      |      |
- Influenza          |       |     | C   |     |      |      |
- HPV                | P1  |     | P2  | C   |      |      |

*Notes:
- e: dose, s: second dose, f: fifth dose
- B: booster dose
- C: conventional dose
- (yearly): dose given yearly

Maximilian Collatz – Prediction of B-Cell Epitopes
Background

Immunodominance
- Proteins that evoke the immune response
- Often exposed surface proteins

Epitope
- Region of the immunodominant protein that is recognized by the immune system
- Continuous or discontinuous
- Length: 8-21 amino acids
Properties of antigens

- Protein disorder (IUPRED)
- Surface accessibility (RaptorX)
- Protein secondary structure (RaptorX)
- Localization (DeepLoc)
- Polymorphic regions (own pipeline)
Epitope prediction by protein structure

IUPRED – disorder tendency
- Intrinsically Unstructured proteins without secondary structure
- Epitope regions are IU-regions

RaptorX – protein structure
- β-sheets
- α-helixes
- Random coil
- Surface accessibility

DeepLoc
- Cellular localization prediction
Detection of polymorphism scores

Protein sequence ➔ BLAST (NCBI NR database) ➔ BLAST result table ➔ Filter hits (80% id/len) ➔ Homologue protein sequences ➔ Remove duplicates ➔ Unique homologs ➔ FAMSA ➔ MSA ➔ Remove gaps in query ➔ Gapless MSA ➔ Detect polymorph regions ➔ Polymorphism scores
Visualization requirements

- Large proteins
- Protein disorder
- Accessibility
  - Buried
  - Medium
  - Exposed
- Structure
  - β-sheets
  - α-helixes
  - Coil
- Polymorphism score

→ Interactive HTML plot
Maximilian Collatz – Prediction of B-Cell Epitopes
Training the Deep Neural Network

- **Input**
  - Epitope sequence
  - Predicted structure information

- **Output**
  - Epitope, or non-epitope
The dataset

- 630 confirmed epitopes
- 771 confirmed non-epitopes and random sequences
Accuracy

[Graph showing an ROC curve with an area of 0.83]
Outlook

• Fine-tune input and DNN parameters
• Evaluate prediction by spotting of predicted epitope sequences and verification with antibody sera
• Evaluation against other published epitopes
Thanks
The Deep Neural Network

• Deep Neural Network
  • Protein sequence
    • Two layers of LSTM with 50 nodes
    • One layer of 50 Dense nodes
  • Other parameters
    • One layer of 50 Dense nodes
  • Combined
    • One layer of 50 Dense nodes
    • One layer of 2 Dense nodes (epitope or non-epitope)