Protein Prediction I Exercise
Summer Term 2020
## (Preliminary) Project Outline

<table>
<thead>
<tr>
<th>Date</th>
<th>Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.04.20</td>
<td>Introduction</td>
</tr>
<tr>
<td>07.05.20</td>
<td>General information on SARS-CoV-2 and COVID-19</td>
</tr>
<tr>
<td>14.05.20</td>
<td>First results on literature research</td>
</tr>
<tr>
<td>21.05.20</td>
<td>No exercise (Christi Himmelfahrt)</td>
</tr>
<tr>
<td>28.05.20</td>
<td>Assessment of current literature</td>
</tr>
<tr>
<td>04.06.20</td>
<td>Distribution of protein features</td>
</tr>
<tr>
<td>11.06.20</td>
<td>No exercise (Fronleichnam)</td>
</tr>
<tr>
<td>18.06.20</td>
<td>Prediction of features for SARS-CoV-2 proteins</td>
</tr>
<tr>
<td>25.06.20</td>
<td>Refinement of results</td>
</tr>
<tr>
<td>02.07.20</td>
<td>Refinement of results</td>
</tr>
<tr>
<td>09.07.20</td>
<td>Final presentations</td>
</tr>
<tr>
<td>16.07.20</td>
<td>Paper Review</td>
</tr>
</tbody>
</table>
Some general remarks

- Finish the tasks/analyses you still wanted to perform
- Combine your predictions + annotations to one overall annotation for the protein
  - How would you do that?
- Interpret your results
  - Are annotations missing for your feature?
  - Do the predictions match the annotations?
  - Are your results as expected?
- Again, present your results in 15 minutes next week
Until next exercise - Group 1

● Secondary structure/Disorder:
  ○ For ratio of helix-sheet-coil, make a stacked barplot
  ○ Combine analysis of secondary structure and disorder

● Effect of point mutations:
  ○ Make sure that your pictures are big enough
  ○ Check the output scores for SIFT
  ○ Compare your results to another predictor (use SNAP2 if you can’t find anything else)
Until next exercise - Group 2

● Conservation:
  ○ Can you think of any explanation why Nsp12 is so highly conserved to the rest? (I would have expected the structural proteins to be more conserved)

● Binding sites:
  ○ Instead of Protein Length vs Pocket Size and color by protein, just show relativ pocket size per protein
  ○ Think of a way to summarize the heatmaps in one number (or so)
Until next exercise - Group 3

- Transmembrane protein:
  - Can you support your predictions for Nsp3 with other tools?

- Metal binding site:
  - By cutoff-plots show some colours that are not present in the legend
  - How do the results look if you combine them?
    - Is there any strong indication (supported by multiple tools) for certain binding sites?
    - Also take the reliability into account
Until next exercise - Group 4

- **DNA/RNA/Protein binding:**
  - Summarize results for all proteins as a small table (e.g. number of residues predicted as DNA/RNA/Protein bindings)
  - Maybe adapt a visualization similar to posttranslational modification to allow easy comparison between annotations and predictions

- **Posttranslational modification:**
  - Add color legend to your visualizations
  - Check for e.g. an enrichment of posttranslational modification
    - What are the overall statistics for all proteins?
    - Are there certain outliers (protein with a lot of acetylations,...)
For the final presentation

- 15-20 minutes per group + 10 minutes discussion
- You can only earn the bonus for the exam if you present something
- Only present your sequence analysis results
  - What was your feature (short introduction)
  - What predictors did you use (short introduction + estimated performance for proteins where we know the annotations)
  - For (some of) the proteins present the combined “annotations” (combination of annotations and predictions)
    - Think about a meaningful way to pick proteins
  - Interpretation of your results
    - Are certain annotations missing/should be further investigated?
    - Where do predictions (dis)agree with the annotations?
    - Are predictions as expected?
- Think about a meaningful way to visualize and interpret your results!