AliNA – a deep learning-based program for prediction of secondary structure without specification of thermodynamic parameters

Shamsudin S. Nasaev, Artem R. Mukanov, Ivan I. Kuznetsov, Alexander V. Veselovsky

nasaev.shs@gmail.com
Why is it important to predict RNA structure?

RNA participate in numerous functions in cells, such as:

- Translation of hereditary information
- Participation in synthesis of proteins and regulation of genes activity

The spatial structure necessary for RNA to fulfill their function is determined by the secondary structure.

tRNA secondary structure
tRNA tertiary structure
### Secondary structure prediction methods

<table>
<thead>
<tr>
<th>Homology-based</th>
<th>Dynamic programming-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>CentroidAlifold</td>
<td>RNAFold</td>
</tr>
<tr>
<td>MXSCARNA</td>
<td>UNAFold</td>
</tr>
<tr>
<td>TurboFold II</td>
<td>DotKnot</td>
</tr>
</tbody>
</table>

**Diagram:**
- Kissing Hairpin pseudoknot
- Hairpin loop
- Internal loop
- Bulged loop
- H-type pseudoknot
AliNA is a convolutional segmentation neural network based on the UNet architecture. It operates on a two-dimensional matrix. Dinucleotide is an elementary unit in analysis. The method predicts the probability of complementary bonds in each cell. There is no necessity for any additional parameters. Sequence length is limited to 256 nucleotides.
Output processing

Resetting to zero of the column Y and row Y as well as column X and row X

Choice of the maximal number in column Y (row X)
Comparison of prediction software

- SPOT-RNA is based on deep learning but it uses single nucleotides interaction.
- MxFold is also based on deep learning but it utilizes thermodynamic parameters.
- Dotknot is an addon for ViennaRNA program. It is based on dynamic programming and makes it possible to predict pseudoknots.
Similarity dependence test
Length dependence test
Prediction example

https://github.com/Arty40m/AliNA
Thank you for your attention!