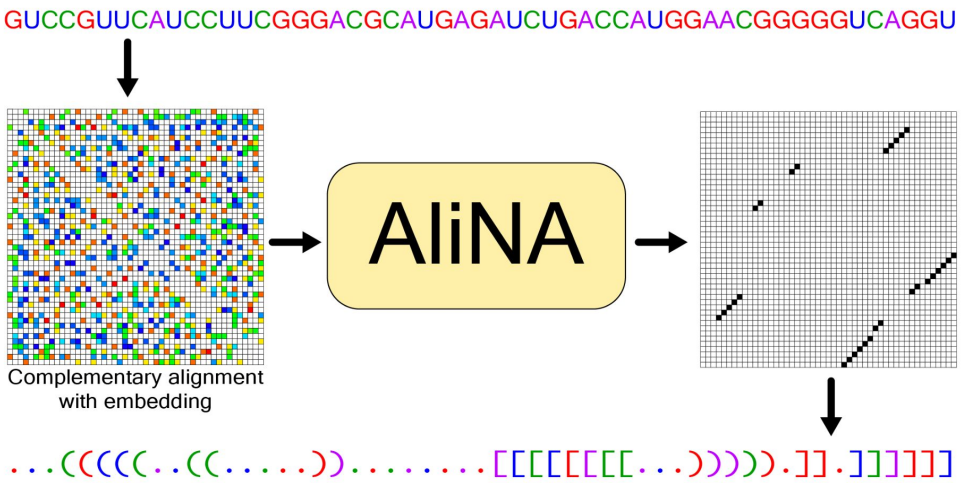
AliNA – a deep learning-based program for prediction of secondary structure without specification of thermodynamic parameters

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



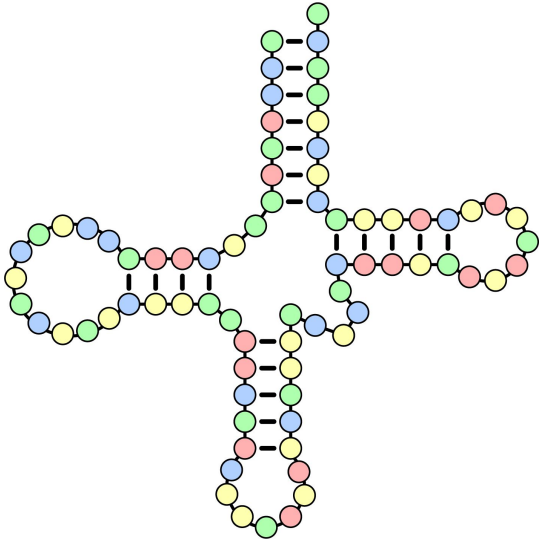
nasaev.shs@gmail.com

Why is it important to predict RNA structure?

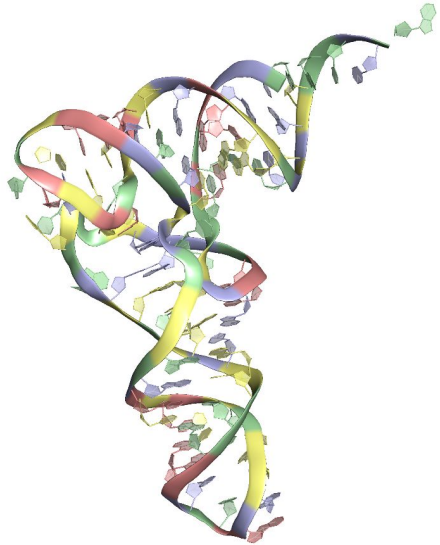
RNA participate in numerous functions in cells, such as:

- 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

Homology-based

CentroidAlifold

MXSCARNA

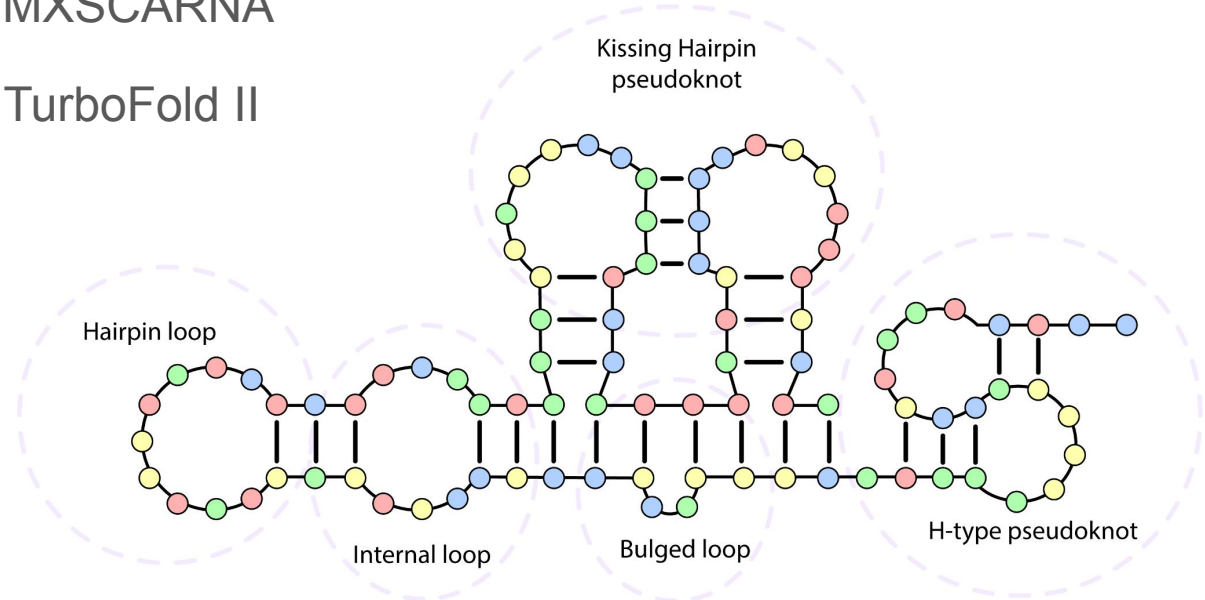
TurboFold II

Dynamic programming-based

RNAFold

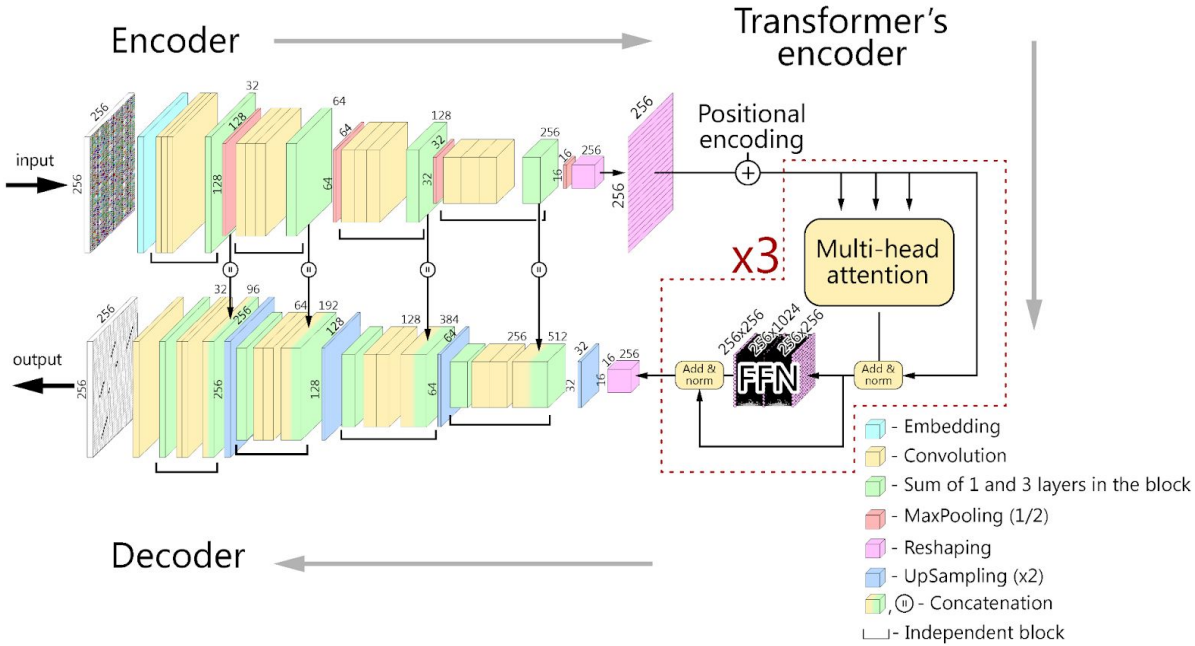
UNAFold

DotKnot

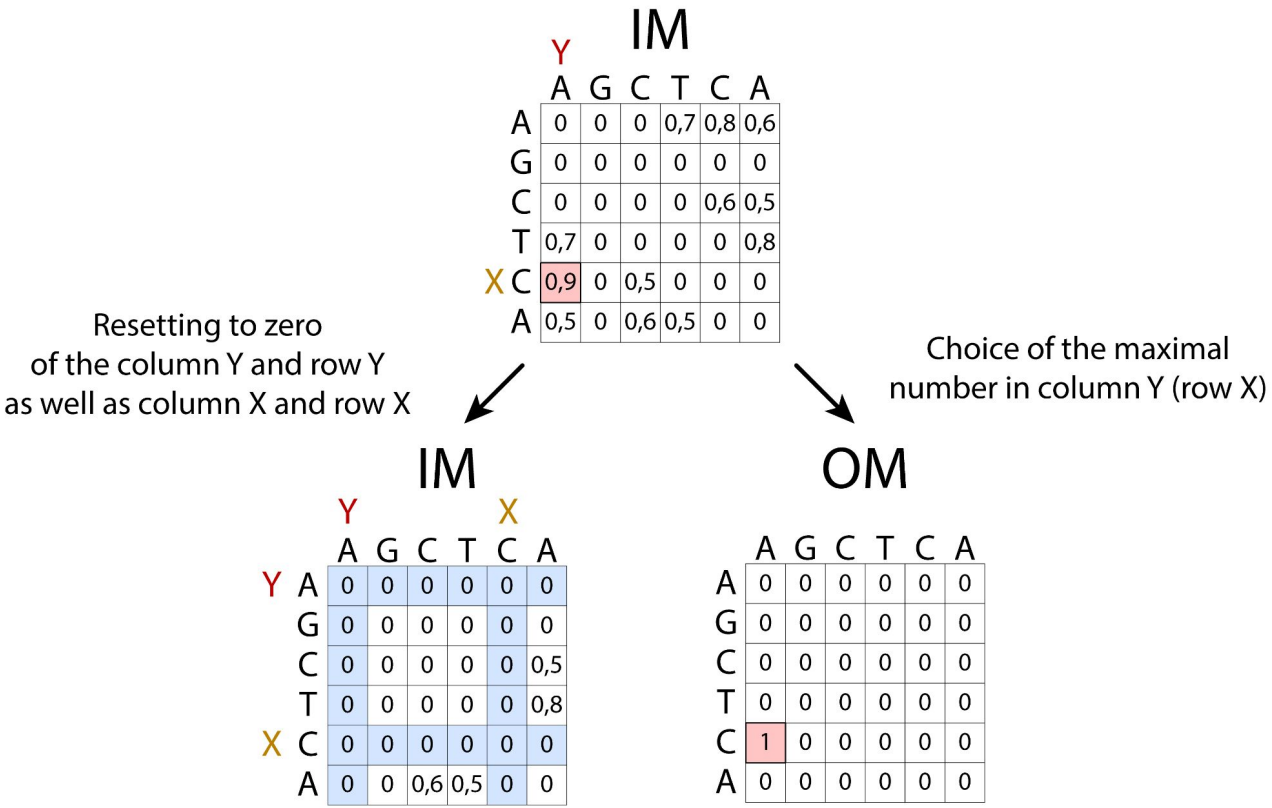


Neural network architecture

- 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.

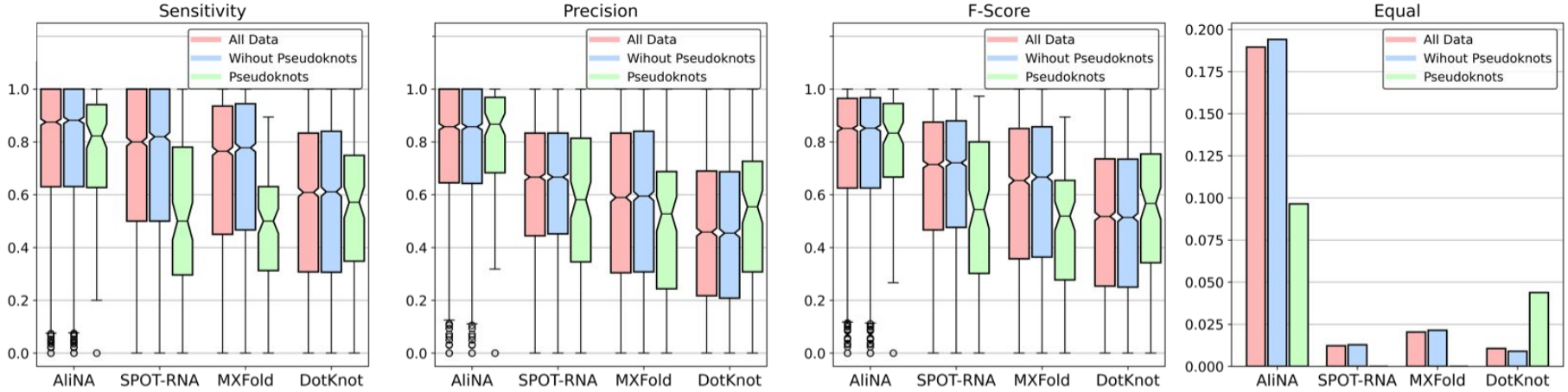


Output processing

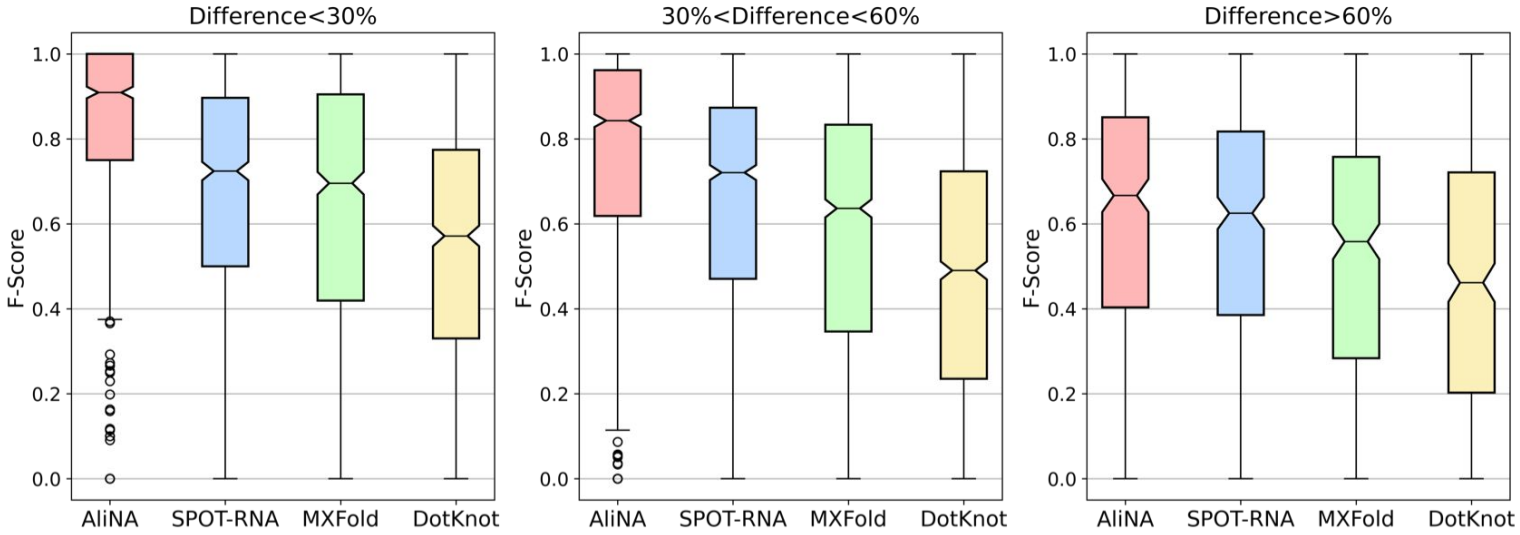


Comparison of prediction software

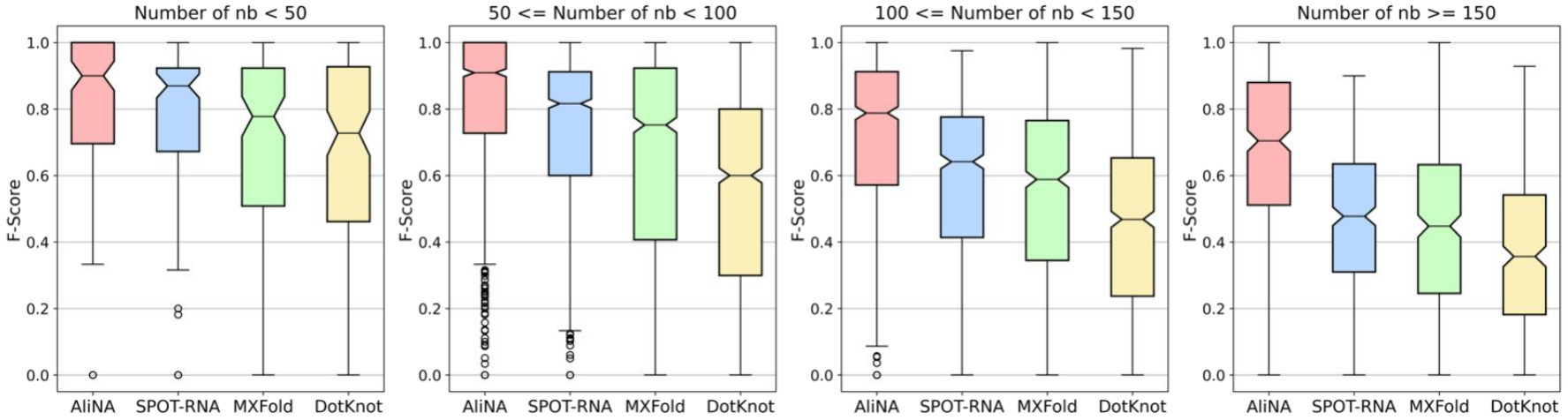
- SPOT-RNA is based on deep learning but it uses single nucleotides interaction.
- MxFold is also based on deep learning but it utilizes thermodynamic parameters.
- Dotknot is an addon for ViennaRNA program. It is based on dynamic programming and makes it possible to predict pseudoknots.



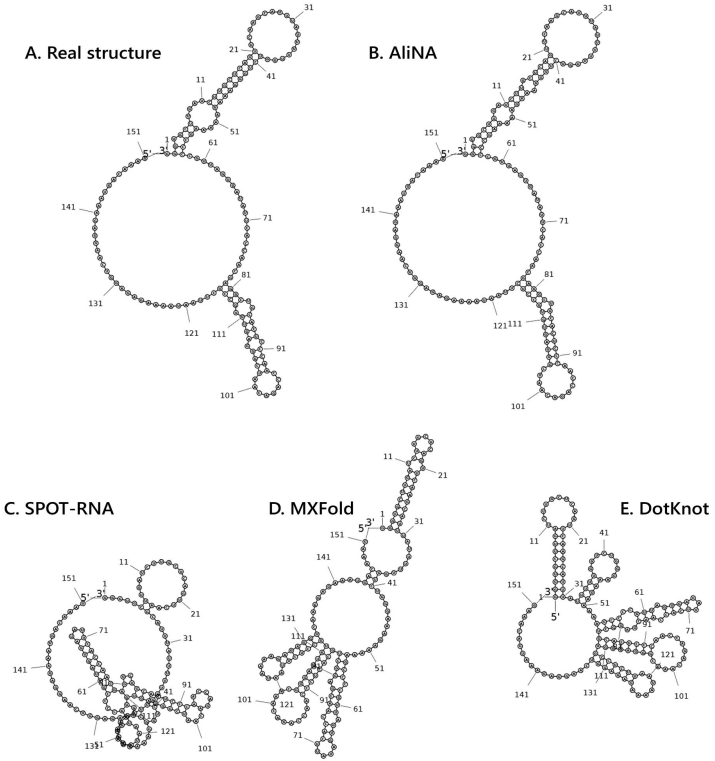
Similarity dependence test



Length dependence test



Prediction example



https://github.com/Arty40m/AliNA

Thank you for your attention!