# Deep Learning the binding patterns of RNA-binding proteins using ENNGene

Eliška Chalupová1,2, Ondřej Vaculík1,2, Jakub Poláček3, Filip Jozefov3, Tomáš Majtner,2 and Panagiotis Alexiou2

1 Faculty of Science, National Centre for Biomolecular Research, Masaryk University, Brno, Czechia

2 Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czechia

3 Faculty of Informatics, Masaryk University, Brno, Czechia

Human genome encodes more than 1500 known RNA-binding proteins (RBPs). Research in the last decades clearly showed that the protein-RNA binding is driven by a combination of multiple factors. Although the RNA sequence is often one of the major binding determinants, other aspects such as site availability, binding competitors, or multimerization of the protein are also crucial. Whether the binding event will occur or not thus can not be easily predicted by classical motif search.

Deep Learning (DL) models now commonly outperform the older Machine Learning (ML) and other methods in RBP target site classification and present a prospective approach able to learn complex patterns given enough data. We show that we can easily reach or even outperform the state-of-the-art results using the ENNGene tool by quickly searching for optimal network architecture and including evolutionary conservation as an additional input feature.

ENNGene is a tool developed in our lab that allows training of custom Convolutional or hybrid Convolutional-Recurrent Neural Networks on any Genomic data through a Graphical User Interface. The tool allows multiple streams of input information, including sequence, evolutionary conservation, and predicted RNA secondary structure. ENNGene deals with all steps of data preprocessing, model training, and evaluation, exporting useful metrics and graphs. To facilitate interpretation of the predicted results, Integrated Gradients provide the user with a graphical representation of an attribution level of each nucleotide of a predicted sequence.