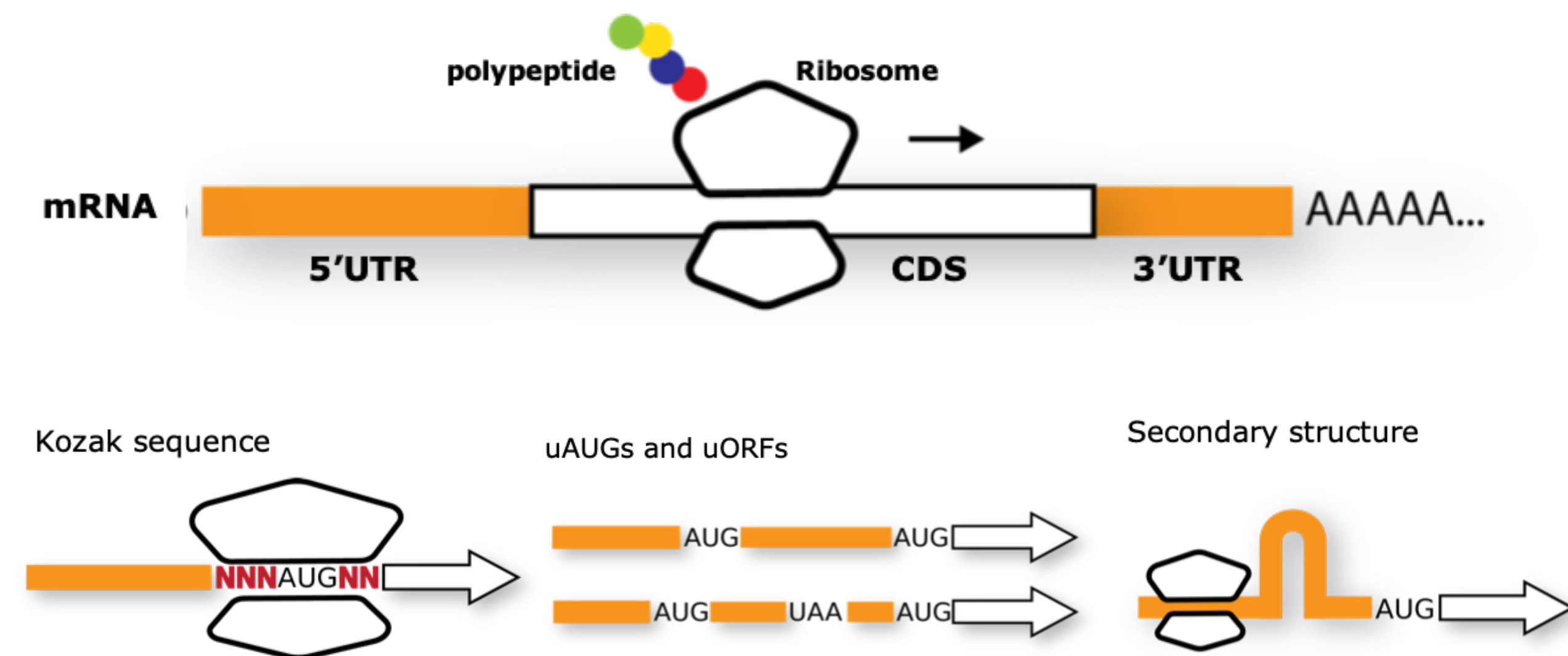


STUDENTS: BAN WANG

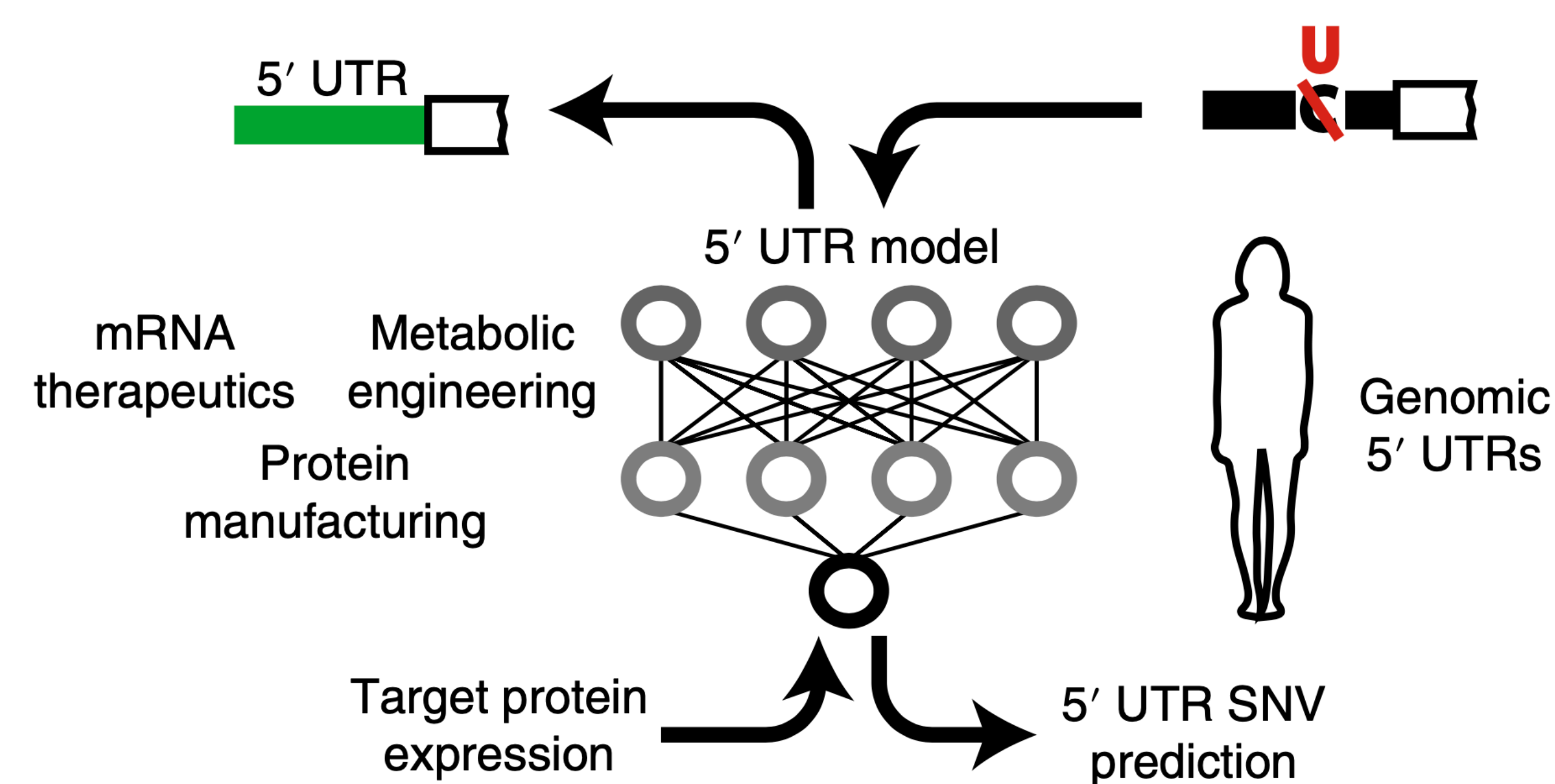
Abstract

The ability to predict the impact of *cis*-regulatory sequences on gene expression would facilitate discovery in fundamental and applied biology. Here we combine polysome profiling of a library of 280,000 randomized 5' untranslated regions (UTRs) with deep learning to build a predictive model that relates human 5' UTR sequence to translation. Our data clearly showed the expected decrease in ribosome loading for sequences with upstream start codons (uAUGs) and each translation initiation site (TIS) or the Kozak sequence can uniquely tune translation initiation to a fine degree. The predictive model we built performed extremely well on our held-out dataset and sequence motifs are calculated. We also perform in silico saturation mutagenesis for human genes and provide evidence of single-nucleotide variants (SNVs) associated with human diseases that substantially change ribosome loading and thus may represent a molecular basis for disease.

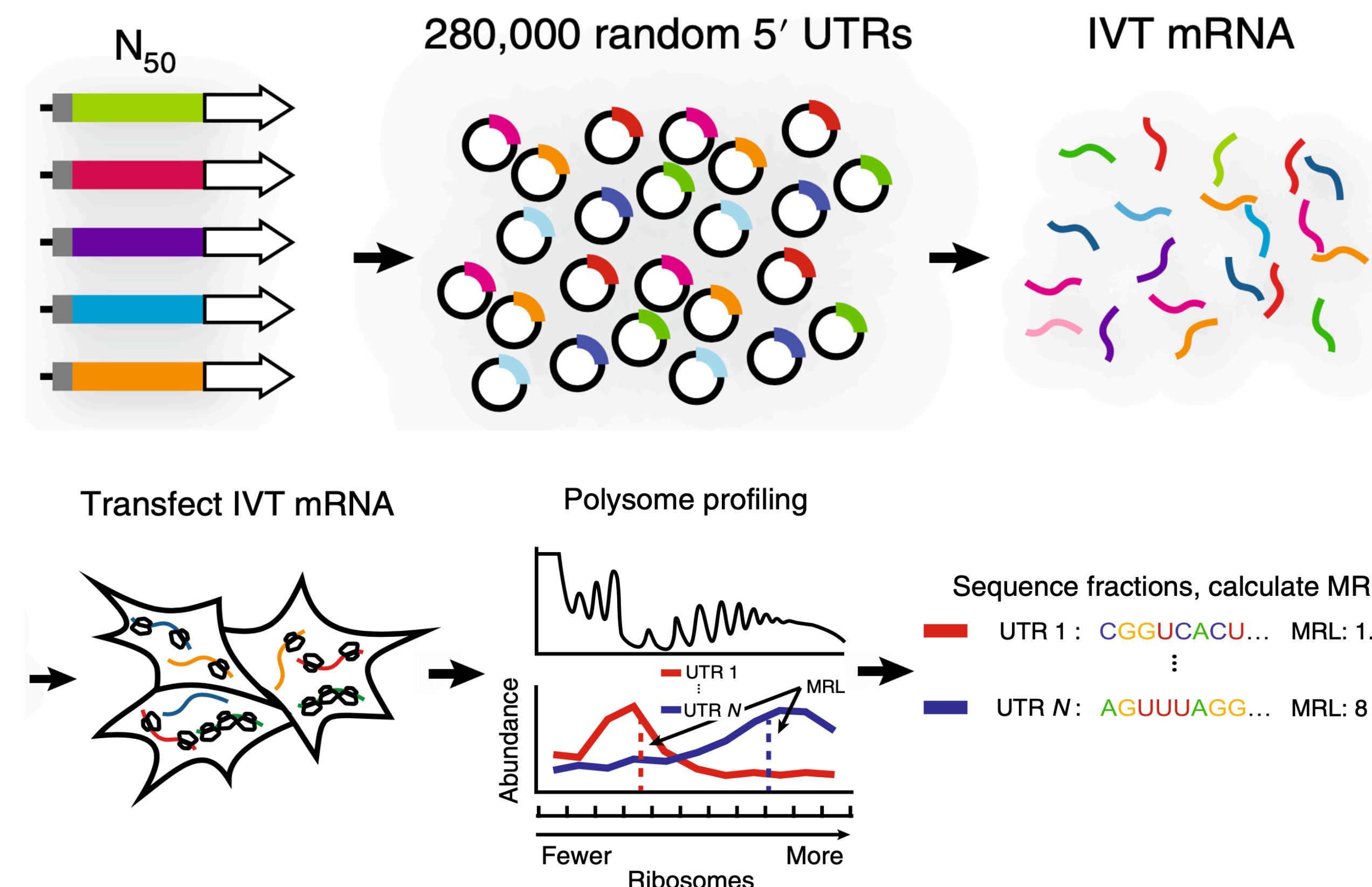
5' UTR (Untranslated Region)



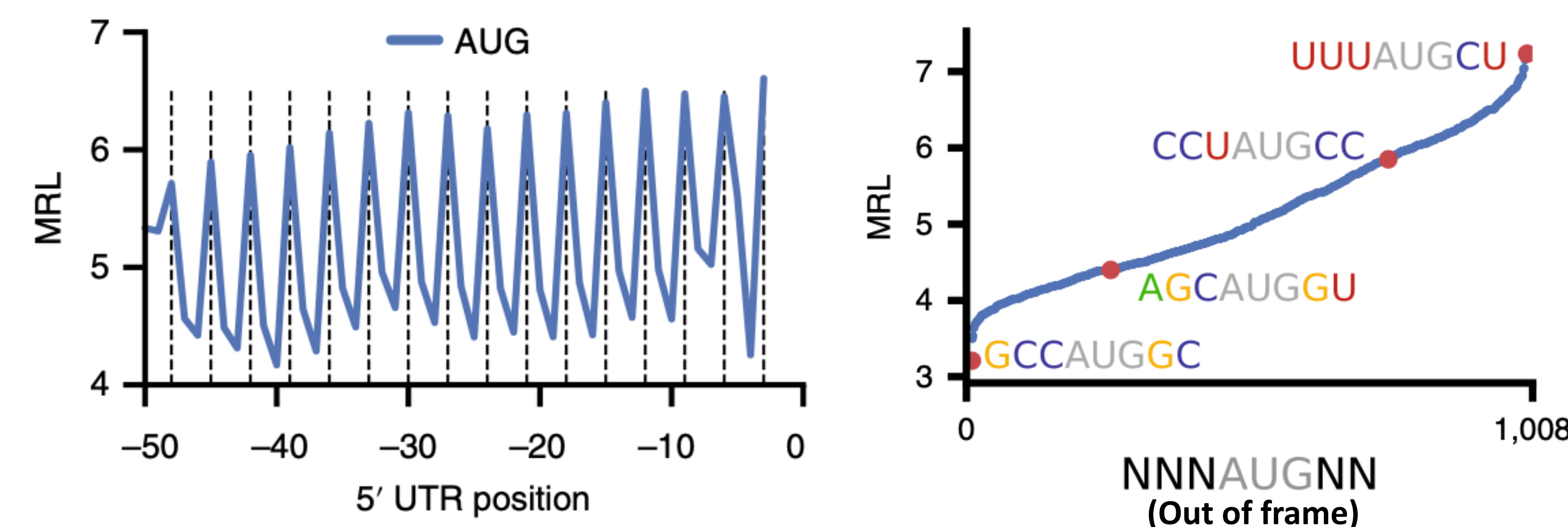
Motivation



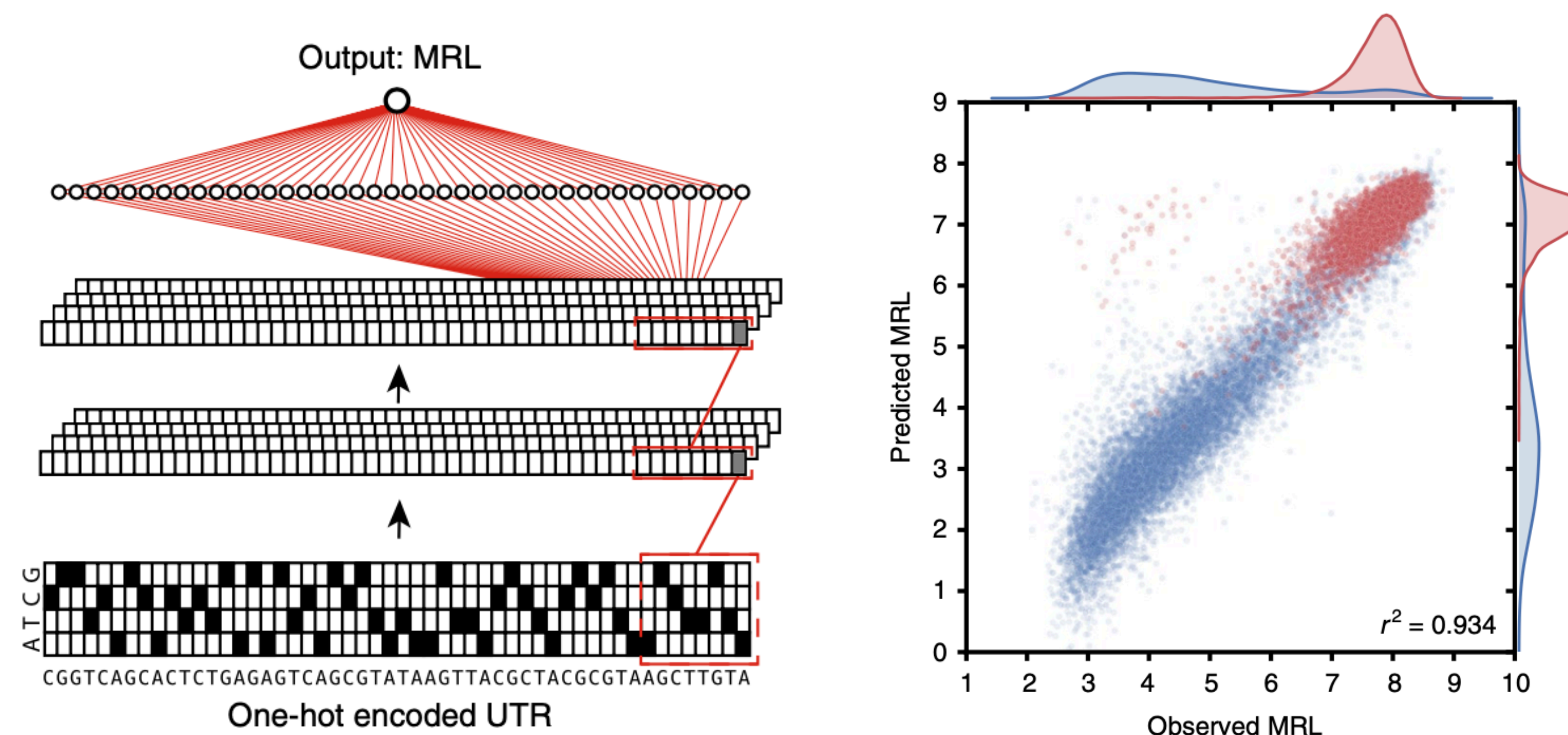
Massively Parallel Reporter Assay (MPRA)



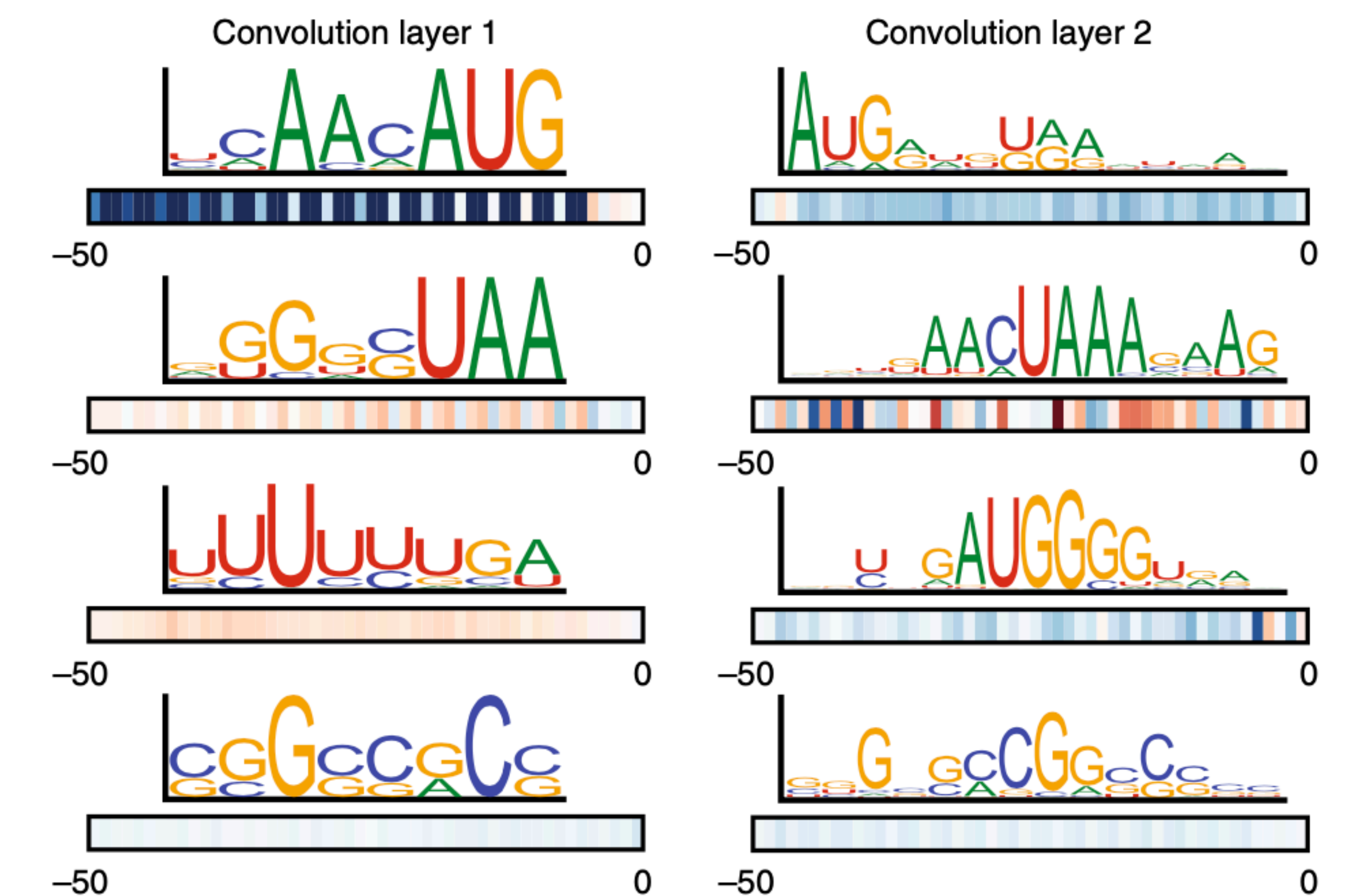
Upstream AUGs (uAUGs)



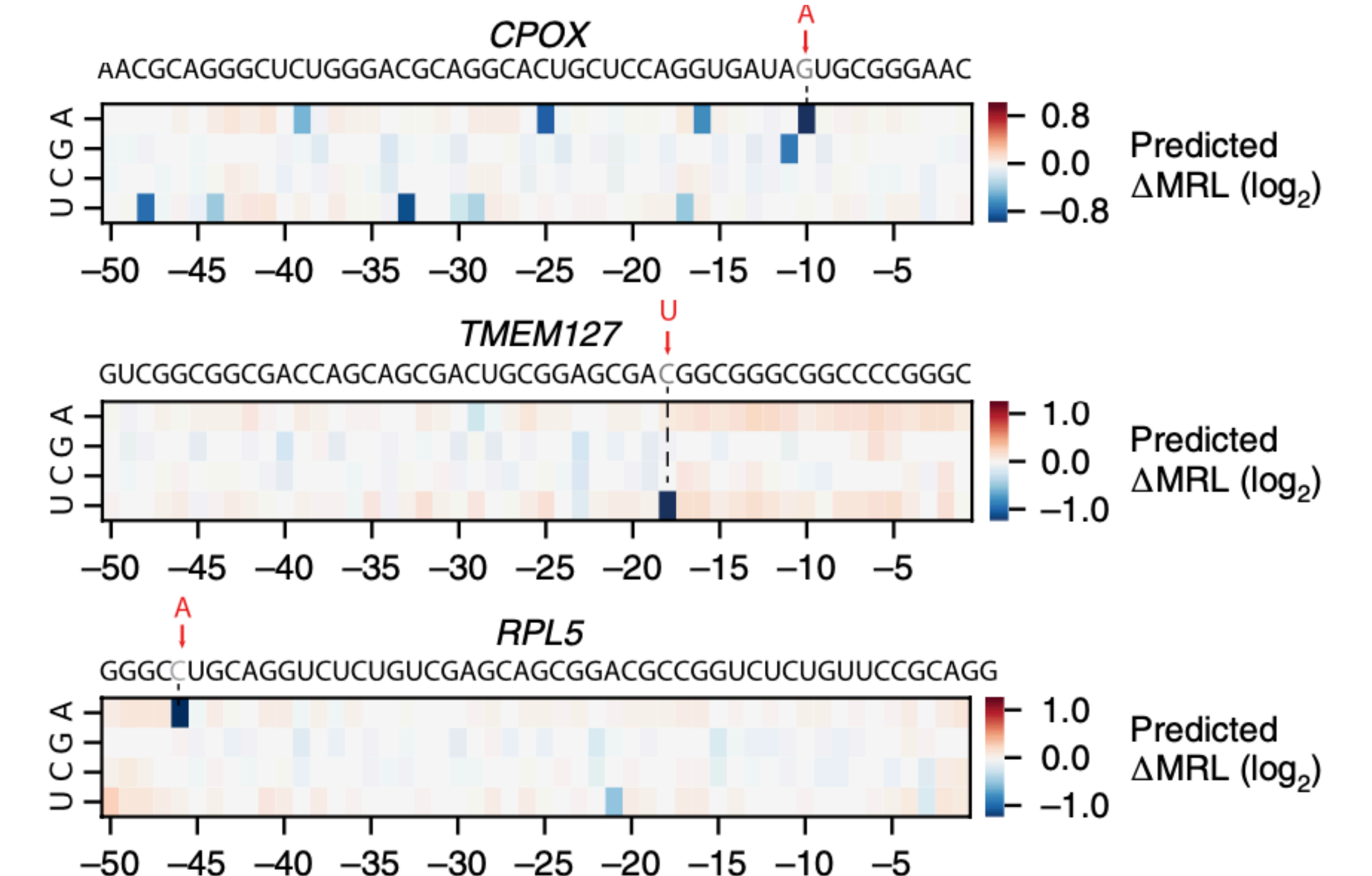
Convolutional Neural Network (CNN) Modeling



CNN Filters / Motifs



In Silico Saturation Mutagenesis for Human Genes



Acknowledgments

Collaborators: Paul J. Sample (*equal contribution), David W. Reid, Vlad Presnyak, Iain J. Mcfadyen, David R. Morris, Georg Seelig
 Acknowledgements: This work was supported by a sponsored research agreement by Moderna and National Institutes of Health grant R01HG009892 to G.S.