

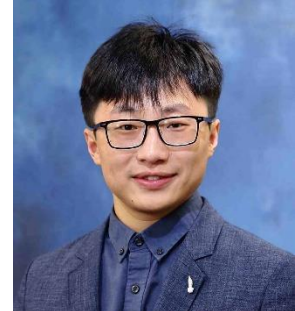
RNA structure prediction and design with large language model-based deep learning methods

Prof. Yu Li

Assistant Professor

Department of Computer Science and Engineering

The Chinese University of Hong Kong



Abstract

RNAs represent a class of programmable biomolecules capable of performing diverse biological functions. However, the structural flexibility of RNA molecules presents critical challenges that lead to a scarcity of experimentally determined RNA structures, in contrast to protein structures. Here, we develop a language model-based deep learning method, RhoFold+, to accurately predict the 3D structures of single-chain RNAs from sequences. We integrate a language model, pre-trained on about 23.7 million RNA sequences without structural information leakage, and multiple techniques for overcoming data scarcity into a fully automated end-to-end pipeline, enabling RhoFold+ to model RNA folding. Retrospective evaluations on RNA-Puzzles and natural RNA targets in CASP15 indicate RhoFold+'s ability to outperform other methods, including human expert groups and computational approaches. RhoFold+'s efficacy and generalizability are further supported by cross-family and cross-type validations, as well as by benchmarks using a time-censored dataset. The above progress may enable new RNAs to be designed in a structure-guided manner with the help of RNA language models. Consequently, we develop a structure-to-sequence platform for the de novo generative design of RNA aptamers. We show that our approach can design RNA aptamers that are structurally similar, yet sequence dissimilar, to light-up aptamers that fluoresce in the presence of small molecules. We experimentally validate several generated RNA aptamers to have fluorescent activity, show that these aptamers can be optimized for activity in silico, and find that they exhibit a mechanism of fluorescence similar to that of known light-up aptamers. Our results demonstrate how language model-based deep learning methods can potentially promote the RNA field. The works have been published or in press at *Nature Methods* and *Nature Computational Science*.

Biography

Yu Li is an Assistant Professor in the Department of Computer Science and Engineering at CUHK, leading the Artificial Intelligence in Healthcare (AIH) group. He is also the Visiting Assistant Professor at MIT/Harvard, working with Prof. James Collins. He works at the intersection between machine learning, healthcare and bioinformatics, developing new machine learning methods to resolve the computational problems in biology and healthcare, which leads to

works published in top venues, such as *Nature Biotechnology*, *Nature Methods*, *Nature Computational Science*, and *Nature Communications*. In 2022, he was selected to the Forbes 30 Under 30 Asia list, Healthcare & Science. His team won the Championship in the CASP15-RNA 3D structure prediction competition in 2022. He obtained his Ph.D. in computer science from KAUST in 2020, after which he was nominated KAUST Alumni Change Makers Awards in 2022. Before that, he got the Bachelor degree in Biosciences with the First-class Honor from University of Science and Technology of China (USTC). He received Department Exemplary Teaching Award in 2024.