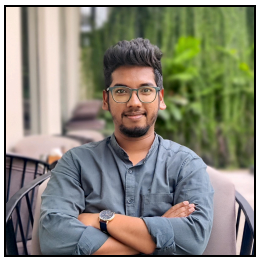
Predicting Gene Expression from DNA Sequences Using Neural Networks

Dr. Divya Tej Sowpati*

Centre for Cellular & Molecular Biology, Habsiguda, Uppal Road, Hyderabad

Email: tej@ccmb.res.in



The genome contains a complex regulatory code written in the language of DNA, which dictates the precise expression of thousands of genes. Deciphering this code is fundamental to understanding biology. However, this remains a key challenge due to the underlying complexity. Artificial intelligence offers a powerful new toolkit for identifying and learning subtle patterns within sequences that are linked to specific biological functions.

This talk and the subsequent session will explore the core concept of treating gene regulation as a predictive problem: can we learn a function that maps a DNA sequence to its functional output, such as gene expression? We will discuss how computational models, specifically neural networks, act as powerful pattern recognizers. By analyzing thousands of examples, these systems can learn to associate specific features in a promoter sequence with high or low gene expression.

The goal is to provide an intuitive understanding of this new data-driven paradigm. While we will not be developing a “state-of-the-art” predictive model, we will delve into the foundational concepts required to understand and eventually build more complex and powerful models.