

Your Paper

You

November 29, 2023

Abstract

Your abstract.

1 Introduction

TODO

2 DNABERT

For short DNA sequences

3 Enformers

The Enformer model [1] utilizes transformer modules, known for their effectiveness in natural language processing, to analyze DNA sequences. Its key features include:

- Transformer Layers: These enable the model to consider each part of the DNA sequence in relation to the entire sequence, crucial for integrating distant genomic elements.
- Extended Receptive Field: Enformer can analyze elements up to 100 kb from the transcription start site, much further than previous models, allowing it to capture a broader range of regulatory elements like distant enhancers.
- Attention Mechanism: This allows the model to weigh different parts of the sequence differently, depending on their relevance to gene expression.

By utilizing existing gene expression data, Enformer can be adapted to Dicty's genome, enabling predictions of gene expression levels based on genomic sequences. This method allows for a deeper understanding of gene regulation in Dicty, identifying key regulatory elements and exploring the impact of genetic variations on gene expression.

References

- [1] Žiga Avsec et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods*, 18(10):1196–1203, 2021.