

AI Is Creating Lifesaving Medical Insights From DNA Secrets

Posted by [Okachinepa](#) 11/20/2024



Courtesy of SynEvol
Credit: Los Alamos Research Lab

Researchers at Los Alamos National Laboratory have created the ground-breaking multimodal deep learning model EPBDxDNABERT-2 to better comprehend the function of DNA in disease. This model is intended to accurately detect connections between DNA and transcription factors, which are proteins that control gene activity. In order to capture these fine movements, EPBDxDNABERT-2 uses a mechanism called "DNA breathing," in which the DNA double-helix opens and shuts on its own. This ability could improve medication development for illnesses caused by gene activity.

The study's lead author, Anowarul Kabir, a researcher at Los Alamos, noted, "The human genome is incomprehensibly large, and there are many types of transcription factors." Therefore, it is essential to determine which transcription factor attaches to which part of the extraordinarily lengthy DNA helix. We attempted to use artificial intelligence, namely deep learning methods, to address that issue.

Each human cell has three billion English letters, or DNA, which serves as a blueprint for development and operation. By binding to certain DNA sequences, transcription factors control gene expression, which is how genes direct the growth and operation of cells. Drug development may be greatly impacted by precisely anticipating transcription factor binding sites because this control is involved in disorders like cancer.

DNA sequences were used to train the study team's basic model. EPBDxDNABERT-2, which can process genome sequences across chromosomes and incorporate related DNA dynamics as input, is the outcome of the team's development of a DNA simulation program that captures a variety of DNA dynamics and integrating it with the genomic foundation model. DNA breathing, or the local, spontaneous opening and shutting of the DNA double-helix structure, is one such input that is correlated with transcriptional activity, including the binding of transcription factors.

"The transcription factor-binding predictions were significantly improved by integrating the DNA breathing features with the DNABERT-2 foundational model," stated Manish Bhattarai, a researcher at Los Alamos. We provide the model with segments of DNA code as input and ask it to determine whether or not it binds to a transcription factor in a variety of cell types. The outcomes increased the likelihood that numerous transcription factors would bind to particular gene sites.

Venado, the Laboratory's newest supercomputer, which combines a central processing unit and a graphics processing unit to power artificial intelligence capabilities, was used by the team to run their deep learning model. A deep-learning model uses text and images to find intricate patterns that produce predictions and insights, much like the neural networks in the brain.

The team used gene sequencing data from 690 experimental outcomes, including 91 human cell types and 161 different transcription factors, to train the model. They discovered that EPBDxDNABERT-2 greatly enhances the prediction of the binding of more than 660 transcription factors, improving it by 9.6% in one important parameter. The in-nature datasets, or the data directly derived from study with living species, like mice, were supplemented by additional experiments on in vitro datasets, which were obtained from tests conducted in a controlled setting.

The study discovered that although DNA breathing alone can nearly precisely predict transcriptional activity, the multimodal model can also extract binding motifs, which are unique DNA sequences that transcription factors bind to. These motifs are an essential component in the explanation of transcription processes.

"Our multimodal foundational model demonstrates versatility, robustness, and efficacy as evidenced by its performance across multiple, diverse datasets," Bhattarai stated. "This model represents a significant breakthrough in computational genomics, offering an advanced instrument for deciphering intricate biological processes."

SynEVOL® | S & S FIRM® ∅ ∅ ∅ - Research & Development

https://synevol.rf.gd/index.php?file=News&op=index_comment&news_id=837.