[BIOINFO 2024]

High-Fidelity Synthetic DNA Generation via GANs
: Enhancing Genetic Regulation and Therapeutic Potential

Myeongryun Lee[1]

[1]*Department of Psychology B.S., University of Utah*

Generative Adversarial Networks (GANs) have emerged as a groundbreaking tool in synthetic biology, capable of generating synthetic DNA sequences with specific genetic motifs. This study investigates the potential of GANs in creating DNA sequences containing splice site motifs, crucial for gene regulation. We trained a GAN model on an extensive and diverse dataset of over 10,000 genomic sequences, followed by rigorous evaluation using advanced bioinformatics tools such as SpliceRover and WebLogo. Th e GAN model underwent approximately 4000 epochs of training, achieving a mean ac curacy of 95% in replicating splice site motifs.

Our evaluation revealed that the synthetic sequences generated by the GAN closely mimic the patterns of real splice sites, particularly in the presence of AG dinucleotides at specific positions, demonstrating high fidelity with a recall of 93% and precision o f 92%. These findings underscore the GAN’s ability to capture and reproduce intricate genetic motifs, marking a significant advancement in synthetic biology. However, challenges such as mode collapse, where the diversity of generated sequences is limited, and the need for improved sequence diversity were identified.

Addressing these challenges is crucial for further progress in the field. Future research will focus on refining the GAN architecture and training processes to enhance the diversity and biological relevance of the generated sequences. The implications of this re search are profound. By demonstrating the ability of GANs to generate biologically re levant synthetic DNA sequences, we open new avenues for genetic research, biotechnology applications, and the development of novel therapeutic strategies. For instance, synthetic DNA sequences with accurately replicated splice sites could be utilized in gene therapy to correct splicing defects or in the design of synthetic genes with desired regulatory features.

This study provides compelling evidence of the capabilities of GANs in synthetic biology. The ability to generate high-fidelity synthetic DNA sequences with specific motif s not only advances our understanding of genetic regulation but also paves the way for innovative applications in biotechnology and medicine. The insights gained from this research will guide future efforts to improve GAN models and their applications, ensuring that they meet the demands of cutting-edge genetic research and therapeutic development.