

# *Advancing Genetic Engineering Through AI: Sequencing and Editing Innovations*

Shili Sun<sup>1,a,\*</sup>

<sup>1</sup>*Department of Economics, University of Washington, Seattle, United States*

*a. ssl0226@asu.edu.pl*

*\*corresponding author*

**Abstract:** As an important part of modern biotechnology, genetic engineering is widely used in fields such as disease treatment, agricultural improvement, and environmental protection. Gene sequencing technology, especially next-generation sequencing technology, provides a powerful tool for studying biological genetic information. However, with the rapid growth of genomic data, how to efficiently and accurately analyze and apply this huge data has become a major challenge facing genetic engineering. In recent years, artificial intelligence (AI) technology, especially deep learning, has been widely used in the automated processing of large-scale data and has shown great potential in genetic engineering. AI technology not only shows advantages in gene editing optimization, genetic variation detection and genome association analysis, but also significantly improves the efficiency and accuracy of genetic data analysis. Although AI has brought many conveniences in genetic engineering, challenges such as technology transparency, data quality issues, and ethics and privacy protection still need to be solved. This article explores the application of artificial intelligence in genetic engineering sequencing and data analysis, analyzes how AI can improve the efficiency and accuracy of genetic data analysis, and discusses the potential contribution of AI in gene editing and precision medicine. As AI continues to develop, it is expected to play an increasingly important role in fields such as genomics, gene editing, and precision medicine, and provide more effective strategies for future disease treatment.

**Keywords:** AI, genetic engineering sequencing, gene editing engineering.

## 1. Introduction

As the core of modern biotechnology, genetic engineering has been widely used in many fields such as disease treatment, agricultural improvement and environmental protection. As the basic technology of genetic engineering, gene sequencing technology provides a deep understanding of the genetic information of organisms, which is crucial to the field of biological genetic engineering. With fast and accurate characteristics, Next-generation sequencing technology (NGS) technology has become a basic tool in genomics research. However, with the explosive growth of genomic data, how to efficiently analyze, interpret and apply these complex data has become an urgent problem in the field of genetic engineering. Traditional data analysis methods often face problems such as slow calculation speed and poor data accuracy when facing such huge and diverse data.

In recent years, the rapid development of artificial intelligence (AI) technology has made it possible to process large-scale data automatically through deep learning and has already brought

revolutionary changes in some fields. For example, the Nobel Prize in Chemistry in 2024 was won for AI-assisted research ---- "Demis Hassabis and John Jumper have developed an AI model to solve a 50-year-old problem: predicting proteins' complex structures. " [1]. In the field of genetic engineering, AI also has great potential, especially in improving the efficiency, accuracy and automation of data analysis, predicting gene function and so on. For example, AI has been used to optimize the guide RNA design of CRISPR-Cas9 systems to improve the precision and efficiency of gene editing [2]. In addition, AI has also been applied to variation detection of genetic data and genome association analysis to help identify genetic variants associated with specific diseases [3].

While AI technology brings many conveniences and advances, it also faces a series of complex challenges. On the technical side, AI models are often limited by poor data quality, and lack of transparency. In terms of practicality, AI has not yet developed to handle some complex problems and can only do some simple analysis based on data, and cannot replace human brain thinking. In addition, public trust in AI is a problem, if all genetic data are provided to AI learning, AI improvement will be rapid, but at the same time, the direction of AI development will be unknown. These issues require a combination of technological innovation, policy adjustments, and ethical guidance to promote the healthy development of AI and maximize its social benefits.

Overall, even if AI technology has so many problems, the advantages of introducing AI technology into the field of genetic engineering must outweigh the disadvantages. AI can not only improve the efficiency of research in the field of genetic engineering, but also accelerate the application of gene editing technology and precision medicine. As AI technology continues to advance, it is expected that AI will play an increasingly important role in many fields, such as genomics, gene editing, and precision medicine. This paper will discuss the combination of AI technology and genetic engineering through the following aspects: the application of artificial intelligence in genetic engineering sequencing and data analysis, how AI can improve the efficiency and accuracy of genetic data analysis, and its practical application in gene editing.

## **2. The Application of AI in Genetic Engineering Sequencing**

### **2.1. Overview of Gene Sequencing Technology**

Gene sequencing technology is the core tool of modern biology and genomics, which is widely used in genomics research, clinical diagnosis and precision medicine. Sequencing DNA means determining the order of the four chemical building blocks - called "bases" - that make up the DNA molecule [4]. Since the completion of the Human Genome Project in the 1990s, gene sequencing technology has made remarkable progress, from the early Sanger sequencing method to the current NGS, which has promoted the rapid development of genomics research.

The traditional Sanger sequencing method is known for its high accuracy, but due to its high cost and low sequencing speed, it is too limited to handle large-scale genomic data [5]. In contrast, NGS is able to sequence millions to billions of DNA fragments simultaneously by parallelizing sequencing methods, greatly increasing the yield and accuracy of data, which makes large-scale genome analysis possible [6]. In addition, NGS technology has significantly reduced the cost of sequencing and greatly improved the speed of data processing, providing a new perspective and method for genomics research.

### **2.2. The Role of AI in Gene Sequencing**

However, as gene sequencing technology continues to evolve, the amount of data generated has grown exponentially. In particular, the genetic data generated by NGS is not only a huge amount of data, but also a complex data structure, including gene sequence, gene expression, epigenetic modification and other multi-dimensional information. How to extract valuable information from

massive genetic data for effective analysis and interpretation has become a major challenge in the field of genomics. In the face of such complex and huge data, traditional statistical methods often face the problem of slow calculation speed and limited data processing ability. In addition, genetic data often contains noise, missing values, and potential error information, which makes the processing and analysis of genetic data more difficult. AI technology, especially machine learning and deep learning technology, can efficiently process large amounts of data in an automated way, overcoming the limitations of traditional methods in dealing with complex data. Even, by training on massive amounts of genetic data, AI can detect variations. With the help of AI, scientists can analyze genetic data more accurately, identify disease-related genes, design personalized treatment programs, and accelerate drug research and development [3].

First, AI can improve the efficiency of data preprocessing in an automated way, removing low-quality reads, fixing errors, and standardizing data. For example, AI can use deep learning models to de-noise gene sequence data and automatically identify and remove possible technical errors and low-quality sequencing data [7]. Gene sequencing data usually need to go through complex pre-processing and quality control steps to ensure the accuracy and validity of sequencing data. With AI technology, researchers can clean data more quickly and accurately, reduce human intervention, and improve the reliability of subsequent analyses. In addition, AI can also help design personalized genome sequencing schemes, adjust sequencing strategies according to individual genome characteristics, improve the depth and coverage of sequencing, and further improve sequencing accuracy.

Furthermore, the AI can be trained to spot low-frequency or covert variants that are difficult to identify with traditional methods. For example, deep learning models can automatically identify mutation sites with pathological significance by learning the features of gene sequences, thus improving the accuracy of variation detection [8]. Genome sequencing usually produces a large number of single nucleotide variation (SNPs), insertion deletion (Indels) and other variation data, AI models can automatically extract valuable information from these data to identify disease-related variants, which will greatly improve the efficiency of genomics research.

An important application of AI in genetic data analysis is drug discovery. Genomic information provides valuable clues for the discovery of drug targets. With AI technology, scientists can mine genomic data for potential drug targets and predict how those targets will interact with drugs. AI can also predict an individual's response to a drug by analyzing a patient's genetic information, supporting the development of precision drugs. AI can identify potential drug targets by modeling the relationship between drugs and genetic data. For example, AI has already played an important role in drug development for multiple cancer types, driving the development of targeted therapeutic drugs by analyzing genetic data and drug responses [7]. With the continuous development of AI technology, future genomic drug research and development will be more accurate and efficient, and the research and development cycle and cost of drugs are also expected to be greatly shortened.

### **3. The Application of AI in Gene Editing Engineering**

#### **3.1. Overview of CRISPR-Cas9 in Gene Editing Engineering**

CRISPR-Cas9 is a groundbreaking gene-editing technology that has revolutionized the ability to alter DNA with unprecedented precision. CRISPR-Cas9 is a unique technology that enables geneticists and medical researchers to edit parts of the genome by removing, adding or altering sections of the DNA sequence. Originating from a natural defense mechanism found in bacteria, the system utilizes the Cas9 protein, which acts like molecular scissors to make cuts in DNA, and a guide RNA (gRNA) that navigates Cas9 to a specific genetic sequence. This guide RNA is designed to match the DNA sequence at the desired editing site, ensuring Cas9 cuts at the exact location where modifications are intended [9]. The technology facilitates fundamental research in genetics, allowing scientists to

easily disable or modify genes to study their function. As research progresses, CRISPR-Cas9 continues to be a pivotal tool in biotechnology, promising to drive significant advancements in gene therapy and genetic engineering.

### **3.2. The Role of AI in Gene Editing Engineering**

In the field of modern gene editing, the use of Artificial Intelligence (AI) has significantly enhanced the efficiency of designing and applying CRISPR-Cas systems, especially in the design of guide RNAs (gRNAs).

Now, AI models have been successfully applied to design gRNAs for the CRISPR-Cas systems. For instance, Sanjana's lab collaborated with the lab of machine learning expert David Knowles to develop a deep learning model named TIGER (Targeted Inhibition of Gene Expression via guide RNA design). This model was trained on data from CRISPR screens and is capable of predicting both on-target and off-target activities. By comparing the predictions generated by the deep learning model with laboratory tests conducted in human cells, TIGER was found to accurately predict both the desired on-target activity and potential off-target effects, which is crucial for optimizing gRNA design [10].

Moreover, other tools like DeepCRISPR, CRISTA, and DeepHF also showcase the breadth of AI applications in this field. These tools can predict optimal guide RNAs (gRNAs) for a specified target sequence, taking into account multiple factors. These include genomic context, the type of Cas protein, the desired type of mutation, on-target/off-target scores, potential off-target sites, and the potential impacts of genome editing on gene function and cell phenotype [7].

These advancements highlight how AI and machine learning are not just supporting but significantly propelling forward the capabilities and precision in genome editing. As AI continues to evolve, it promises to further refine the specificity and efficiency of CRISPR applications, making gene editing more reliable and accessible for research and therapeutic purposes.

### **3.3. The Role of AI in Gene Expression Regulation Analysis**

Gene expression regulation is one of the core issues in genomics research, involving many complex factors such as transcription factors, epigenetic modification, non-coding RNA, etc. Although traditional gene expression analysis methods can reveal the expression patterns of certain genes, they often have limitations when dealing with complex gene regulatory networks. AI, especially deep learning technology, can automatically identify key factors in gene regulatory networks by learning large amounts of gene expression data and reveal the complex mechanisms behind gene expression. By using AI, researchers were able to discover new transcription factors, regulatory elements, and the effects of epigenetic modifications on gene expression. AI models can analyze the expression patterns of genes in different cell types and environmental conditions, further helping scientists understand how genes function in different biological processes. For example, AI technology has been successfully applied to analyze the abnormal expression of genes in cancer cells, reveal the changes in gene regulation during the occurrence of cancer, and provide new ideas for the early diagnosis and treatment of cancer [11].

## **4. AI-Driven Genomics Tools and Platforms**

### **4.1. Application of Open Source AI Platform**

Open source AI platforms have become indispensable tools in genomics research, providing powerful resources to process and analyze vast genetic data sets. These platforms enable researchers to develop and apply advanced machine learning models to explore complex patterns in genetic sequences.

Google's DeepVariant, for example, uses deep learning to identify genetic variants more accurately than traditional methods. In addition, MIT's open source AI system has demonstrated its potential to generate complex biological models that can help scientists better understand disease mechanisms and biological processes. [7,12].

## 4.2. The Role of AI Platforms in the Automation of Genetic Engineering Laboratories

Laboratory automation is being greatly enhanced by AI technology. AI systems can optimize experimental processes, automate sample processing, data analysis, and interpretation of results. This automation technology not only improves the efficiency of the experiment, but also ensures the consistency and reliability of the experimental results. The applications of AI in automation range from basic data entry to complex experimental design and implementation, greatly reducing the burden on experimenters and improving the speed and quality of scientific research [7].

## 5. Challenges and future prospects

### 5.1. Technical Challenges: Data Heterogeneity and Privacy Issues

The heterogeneity of genomics data and the sheer volume of data present significant technical challenges for AI applications. AI models need to be able to process and integrate data from different technologies, platforms, and experimental conditions, which requires algorithms to be powerful and flexible. At the same time, genetic data involves personal privacy, and finding a balance between promoting scientific research and protecting privacy is an ongoing challenge. AI models need to be developed and deployed while ensuring data security and user privacy [10].

### 5.2. Ethical and Legal Challenges

With the application of AI technology in sensitive areas such as gene editing and gene therapy, ethical and legal issues become particularly important. How to ensure that the use of technology does not cause ethical and legal disputes in society, especially when it involves research that changes the human genome, requires the international community to jointly develop strict guidelines and legal frameworks [3].

## 6. Conclusion

The application of AI in the field of genomics has proven its significant value in improving research efficiency, facilitating interdisciplinary collaboration, and advancing precision medicine. Through the sharing of open source platforms, the implementation of laboratory automation, and the optimization of data analysis, AI is changing the research and application landscape of genomics. Future research should aim to address the challenges of AI in data processing and privacy protection, ensuring that the development of the technology is consistent with ethical regulations. In addition, strengthening the applied research of AI in gene editing and therapy can provide more effective strategies and methods for treating major diseases such as genetic diseases and cancer. Through international cooperation and policy development, the healthy development of AI technology can be ensured to contribute to the well-being of all mankind.

## References

- [1] Nobel Prize. (2024). *The Nobel Prize in Chemistry 2024*. The Nobel Prize. Retrieved from <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>
- [2] Mishra, R., & Li, B. (2020). *The application of artificial intelligence in the genetic study of Alzheimer's disease*. *Aging and Disease*, 11(6), 1567–1584.

- [3] Vilhekar, R. S., & Rawekar, A. (2024). *Artificial intelligence in genetics*. *Cureus*, 16(1).
- [4] National Human Genome Research Institute. (n.d.). *DNA sequencing fact sheet*. Retrieved from <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Fact-Sheet>
- [5] Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). *Sequencing technologies and genome sequencing*. *Journal of Applied Genetics*, 52, 413-435.
- [6] Dias, R., & Torkamani, A. (2019). *Artificial intelligence in clinical and genomic diagnostics*. *Genome Medicine*, 11(1), 70.
- [7] Dixit, S., Kumar, A., Srinivasan, K., et al. (2024). *Advancing genome editing with artificial intelligence: Opportunities, challenges, and future directions*. *Frontiers in Bioengineering and Biotechnology*, 11, 1335901.
- [8] Redman, M., King, A., Watson, C., & King, D. (2016). *What is CRISPR/Cas9?* *Archives of Disease in Childhood-Education and Practice*, 101(4), 213-215.
- [9] New York University. (2023). *AI and CRISPR used to tame gene expression*. Retrieved from <https://www.nyu.edu/about/news-publications/news/2023/july/ai-crispr-gene-expression.html>
- [10] Heather, J. M., & Chain, B. (2016). *The sequence of sequencers: The history of sequencing DNA*. *Genomics*, 107(1), 1-8.
- [11] Ouyang, A., & Jameel, A. L. (2023). *MIT scientists build a system that can generate AI models for biology research*. *MIT News ON CAMPUS AND AROUND THE WORLD*. Retrieved from <https://news.mit.edu/2023/bioautomated-open-source-machine-learning-platform-for-research-labs-0706>
- [12] Xu, J., Yang, P., Xue, S., et al. (2019). *Translating cancer genomics into precision medicine with artificial intelligence: Applications, challenges and future perspectives*. *Human Genetics*, 138(2), 109-124.