Computational Molecular Biology 2024, Vol.14, No.3, 106-114
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

# AI in Biology: Transforming Genomic Research with Machine Learning

Qiang Zhang, Yu Wang
Biotechnology Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China
Corresponding author: yu.wang@cuixi.org

Computational Molecular Biology, 2024, Vol.14, No.3 doi: 10.5376/cmb.2024.14.0013
Received: 08 Apr., 2024
Accepted: 23 May, 2024
Published: 10 Jun., 2024



**Preferred citation for this article**:
Zhang Q., and Wang Y., 2024, AI in biology: transforming genomic research with machine learning, Computational Molecular Biology, 14(3): 106-114 (doi: 10.5376/cmb.2024.14.0013)

**Abstract** With the rapid development of artificial intelligence (AI) and machine learning (ML) technologies, the field of biology, particularly genomic research, is undergoing profound transformations. This study explores how AI and ML are redefining genomic data analysis and functional genomics research, while emphasizing the critical role these technologies play in enhancing research efficiency, improving accuracy, and advancing personalized medicine. The application of AI in biology has expanded from basic data processing to complex tasks such as gene function prediction, identification of regulatory elements, and understanding epigenetic modifications. Through an in-depth analysis of key machine learning techniques, including supervised learning, unsupervised learning, and deep learning, this study demonstrates how these methods are revolutionizing traditional genomic data analysis workflows, significantly improving the efficiency of sequence alignment, variant calling, and gene expression profiling. Additionally, it discusses the future prospects of AI-driven genomic tools, cloud computing, big data integration, and open-source platform collaboration, aiming to provide valuable insights for future research and technological development.

**Keywords** Artificial intelligence (AI); Machine learning (ML); Genomic research; Functional genomics; Personalized medicine

## 1 Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized numerous fields, and biology is no exception. The advent of high-throughput technologies has led to an explosion of biological data, necessitating advanced computational methods to analyze and interpret these vast datasets. Machine learning, which involves developing algorithms that improve through experience, has shown immense potential in handling complex biological data. Techniques such as supervised, semi-supervised, and unsupervised learning, as well as deep learning, are being increasingly applied to genomic data to uncover hidden patterns and make accurate predictions (Angermueller et al., 2016). These methods have been particularly effective in tasks such as annotating sequence elements, predicting gene expression levels, and identifying genomic elements like promoters and enhancers (Wu and Zhao, 2019; Liu et al., 2020).

Genomic research is pivotal in understanding the fundamental mechanisms of life and disease. By studying the genome, researchers can identify genetic variations that contribute to diseases, understand gene function, and develop targeted therapies. The ability to analyze large-scale genomic data has opened new avenues in precision medicine, where treatments can be tailored to an individual's genetic makeup (Koumakis et al., 2020). The integration of machine learning in genomic research has further accelerated discoveries, enabling the modeling of complex biological networks and the prediction of disease risks based on genetic information (Leung et al., 2016; Camacho et al., 2018).

This study will provide a comprehensive overview of the current applications of artificial intelligence and machine learning in genomics research, explore various machine learning techniques and their practical applications in genomics, discuss the challenges and limitations of these methods, and emphasize the future development directions in this field. I hope to clarify the transformative impact of machine learning on genomic research and its potential to further advance biology and medicine.

## 2 Overview of Machine Learning Techniques in Genomics

### 2.1 Supervised learning

Supervised learning involves training a model on a labeled dataset, where the input data is paired with the correct

output. This approach is widely used in genomics for tasks such as variant calling, gene expression prediction, and classification of genomic sequences. For instance, supervised learning techniques have been applied to annotate sequence elements and analyze epigenetic, proteomic, and metabolomic data (Libbrecht and Noble, 2015). These methods are particularly effective in scenarios where a large amount of labeled data is available, allowing the model to learn the mapping from inputs to outputs accurately.

### 2.2 Unsupervised learning

Unsupervised learning techniques are used to identify patterns and structures in data without the need for labeled outputs. In genomics, these methods are often employed for clustering and dimensionality reduction tasks. Clustering approaches, such as hierarchical, centroid-based, and density-based methods, help in understanding the natural structure inherent in genomic data, such as gene expression profiles and cellular processes (Figure 1) (Karim et al., 2020). Unsupervised learning is crucial for exploratory data analysis, where the goal is to uncover hidden patterns and relationships within the data.

This image illustrates the use of a convolutional autoencoder for unsupervised learning to perform clustering analysis on microscope images. Clustering analysis is conducted after image processing, utilizing clustering algorithms such as K-means to group the feature space. This approach helps uncover hidden patterns and relationships within the data, such as distinct clusters of gene expression or differences between cell types. To enhance clustering performance, the network jointly optimizes both the reconstruction loss and the Cluster Assignment Hardening (CAH) loss, refining the clustering results by continuously adjusting the network parameters. This application of unsupervised learning in genomics is particularly suited for exploratory data analysis. Through clustering methods, it can help us understand the intrinsic natural structure of genomic data, thereby revealing hidden patterns in gene function and cellular processes.

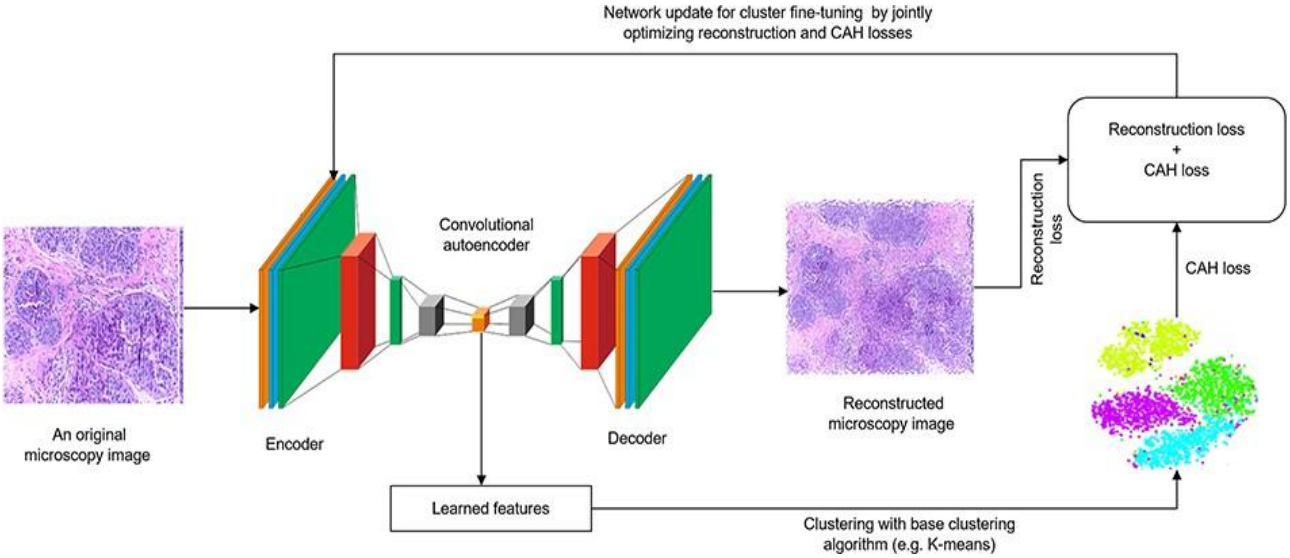


Figure 1 Schematic representation of a VAE used for clustering GE data (Adopted from Karim et al., 2020)

### 2.3 Deep learning and neural networks

Deep learning, a subset of machine learning, has revolutionized the analysis of genomic data by leveraging multilayered artificial neural networks (ANNs) to model complex patterns. Deep learning techniques, including convolutional neural networks (CNNs) and deep neural networks (DNNs), have shown remarkable success in various genomic applications. These methods are particularly adept at handling high-dimensional data and have been used to predict the structure and function of genomic elements, such as promoters and enhancers (Li, 2018; Liu et al., 2020; Schmidt and Hildebrandt, 2020). Deep learning models have also been applied to next-generation sequencing (NGS) data for tasks such as variant calling, metagenomic classification, and genomic feature detection. Despite their success, one of the challenges with deep learning models is their interpretability. Efforts are being made to develop methods for interpreting the predictions of DNNs to better understand the underlying molecular and cellular mechanisms (Talukder et al., 2020).

# 3 Applications of Machine Learning in Genomic Data Analysis

## 3.1 Sequence alignment and assembly

### 3.1.1 Traditional methods vs. AI-enhanced techniques

Traditional methods for sequence alignment and assembly, such as Burrows-Wheeler Alignment (BWA) and Genome Analysis ToolKit (GATK), have been foundational in genomic research. However, these methods often struggle with the massive volume and complexity of next-generation sequencing data. AI-enhanced techniques, such as those implemented in Findmap and Findvar, offer significant improvements. Findmap, for instance, integrates known variant locations during alignment, which enhances both speed and accuracy compared to traditional methods like BWA and SNAP. Similarly, GotCloud employs machine learning for efficient variant calling and genotyping, automating several steps and reducing computational resource requirements (Jun et al., 2015).

### 3.1.2 Accuracy and efficiency improvements

AI-enhanced techniques have demonstrated substantial improvements in both accuracy and efficiency. For example, Findmap correctly mapped 92.9% of reads, outperforming traditional methods like BWA (90.5%) and SNAP (92.6%). Additionally, Findvar showed high accuracy in calling single nucleotide variants (99.8%), insertions (79%), and deletions (67%), surpassing traditional tools like SAMtools in certain aspects. The graph genome reference implementation also enhances read mapping sensitivity and variant calling accuracy, achieving a 0.5% increase in recall without compromising specificity (Rakocevic et al., 2019).

### 3.1.3 Case studies in genomic assembly

Several case studies highlight the practical applications of AI in genomic assembly. For instance, the 1000 Bull Genomes Project utilized Findmap and Findvar to process large datasets efficiently, significantly reducing clock times compared to traditional methods (VanRaden et al., 2019). Another example is the use of GotCloud in the 1000 Genomes Project and the NHLBI Exome Sequencing Project, where it effectively filtered false positives and detected true variants with high power (Jun et al., 2015). These case studies underscore the potential of AI-enhanced techniques to handle large-scale genomic data more effectively.

## 3.2 Variant calling and mutation analysis

Machine learning has revolutionized variant calling and mutation analysis by improving the accuracy and efficiency of detecting genetic variants. Tools like GotCloud and Findvar leverage machine learning to automate variant calling, filter artifacts, and refine genotypes, thereby enhancing the reliability of genomic data analysis (Zou et al., 2018). Deep learning methods have also been applied to predict the effects of genetic variants on gene expression, further advancing our understanding of genomic variations.

## 3.3 Gene expression profiling

Gene expression profiling has benefited significantly from machine learning, particularly deep learning techniques. The Enformer model, for example, integrates long-range interactions in the genome to predict gene expression with high accuracy. This model has outperformed traditional methods in predicting the effects of noncoding variants on gene expression, demonstrating the potential of deep learning to enhance our understanding of gene regulation (Avsec et al., 2021). Deep learning frameworks have been applied to various aspects of gene expression analysis, including the identification of sequence motifs and promoter-enhancer interactions (Talukder et al., 2020; Routhier and Mozziconacci, 2022).

# 4 Machine Learning in Functional Genomics

## 4.1 Predicting gene function

Machine learning has become an indispensable tool in predicting gene function, leveraging vast amounts of genomic data to uncover insights that traditional methods might miss. For instance, machine learning algorithms have been employed to integrate heterogeneous data and detect patterns that are not easily discernible through rule-based approaches. This has been particularly useful in plant genomics, where predicting gene function and organismal phenotypes remains a significant challenge (Leung et al., 2016). Deep learning models have shown promise in predicting the structure and function of genomic elements, such as promoters and enhancers, which are crucial for understanding gene expression levels (Liu et al., 2020).

### 4.2 Identifying regulatory elements

Identifying regulatory elements within the genome is another area where machine learning has made substantial contributions. Convolutional neural networks (CNNs) have been developed to predict cell type-specific epigenetic and transcriptional profiles from DNA sequences alone. These models can identify promoters and distal regulatory elements, synthesizing their content to make effective gene expression predictions (Kelley et al., 2017). Machine learning algorithms trained on multiple genomes have improved the accuracy of gene expression predictions and facilitated the analysis of human genetic variants associated with molecular phenotypes and diseases (Kelley, 2019). These advancements highlight the potential of machine learning to enhance our understanding of gene regulation and its implications for human health.

### 4.3 Understanding epigenetic modifications

Epigenetic modifications play a crucial role in gene expression and are linked to various cellular processes, including differentiation, development, and tumorigenesis. Machine learning methods have been widely applied to study these modifications, providing insights into the regulatory mechanisms that rely on epigenetic changes. For example, a unified deep learning model called ZayyuNet has been proposed for identifying various epigenetic modifications, such as DNA N6-Methyladenine (6mA) and RNA N6-Methyladenosine (m6A). This model has demonstrated superior performance compared to current state-of-the-art models (Abbas et al., 2021). Deep neural networks have been utilized to interpret genomic and epigenomic data, focusing on tasks such as sequence motif identification and gene expression prediction (Talukder et al., 2020). These approaches have significantly advanced our understanding of how epigenetic modifications influence gene regulation and cellular function.

## 5 Integrating Multi-Omics Data with AI

### 5.1 Challenges in multi-omics integration

Integrating multi-omics data presents several challenges due to the inherent complexity and heterogeneity of the data. High-dimensionality, data heterogeneity, and noise are significant obstacles that need to be addressed to effectively combine data from different omics layers such as genomics, proteomics, and metabolomics (Mirza et al., 2019). The curse of dimensionality, where the number of features far exceeds the number of samples, complicates the analysis and integration process. Missing data and class imbalance further exacerbate these challenges, necessitating specialized computational approaches to manage these issues effectively. The lack of universal analysis protocols and the need for interpretability and explainability in models also pose significant hurdles (Wörheide et al., 2021).

### 5.2 AI models for multi-omics analysis

Machine learning and deep learning models have shown great promise in addressing the challenges of multi-omics data integration. Various models, including Bayesian models, tree-based methods, kernel methods, network-based fusion methods, and matrix factorization models, have been employed to integrate and analyze multi-omics data (Li et al., 2016). Deep learning, in particular, has gained prominence due to its ability to capture complex, non-linear relationships in large-scale datasets (Kang et al., 2021; Saha et al., 2023). Autoencoders and other deep neural networks have been used to learn cross-modality interactions and provide interpretable results in a multi-source setting (Benkirane et al., 2023). These models have been applied to tasks such as disease subtype classification, biomarker discovery, and drug response prediction, demonstrating their potential in advancing precision medicine.

### 5.3 Applications in personalized medicine

The integration of multi-omics data using AI models has significant implications for personalized medicine. By combining data from various omics sources, researchers can gain a comprehensive understanding of the molecular mechanisms underlying diseases, leading to more accurate disease prediction, patient stratification, and the development of personalized treatment plans (Figure 2) (Kang et al., 2021; Reel et al., 2021). For instance, deep learning models have been used to classify tumor types and breast cancer subtypes, as well as predict survival outcomes in cancer patients. The ability to integrate and analyze multi-omics data also facilitates the discovery of new biomarkers, which can be used to monitor disease progression and response to treatment, ultimately

improving patient outcomes (Nicora et al., 2020; Reel et al., 2021). As the field continues to evolve, the integration of multi-omics data with AI will play a crucial role in the development of precision medicine strategies, paving the way for more targeted and effective therapies.

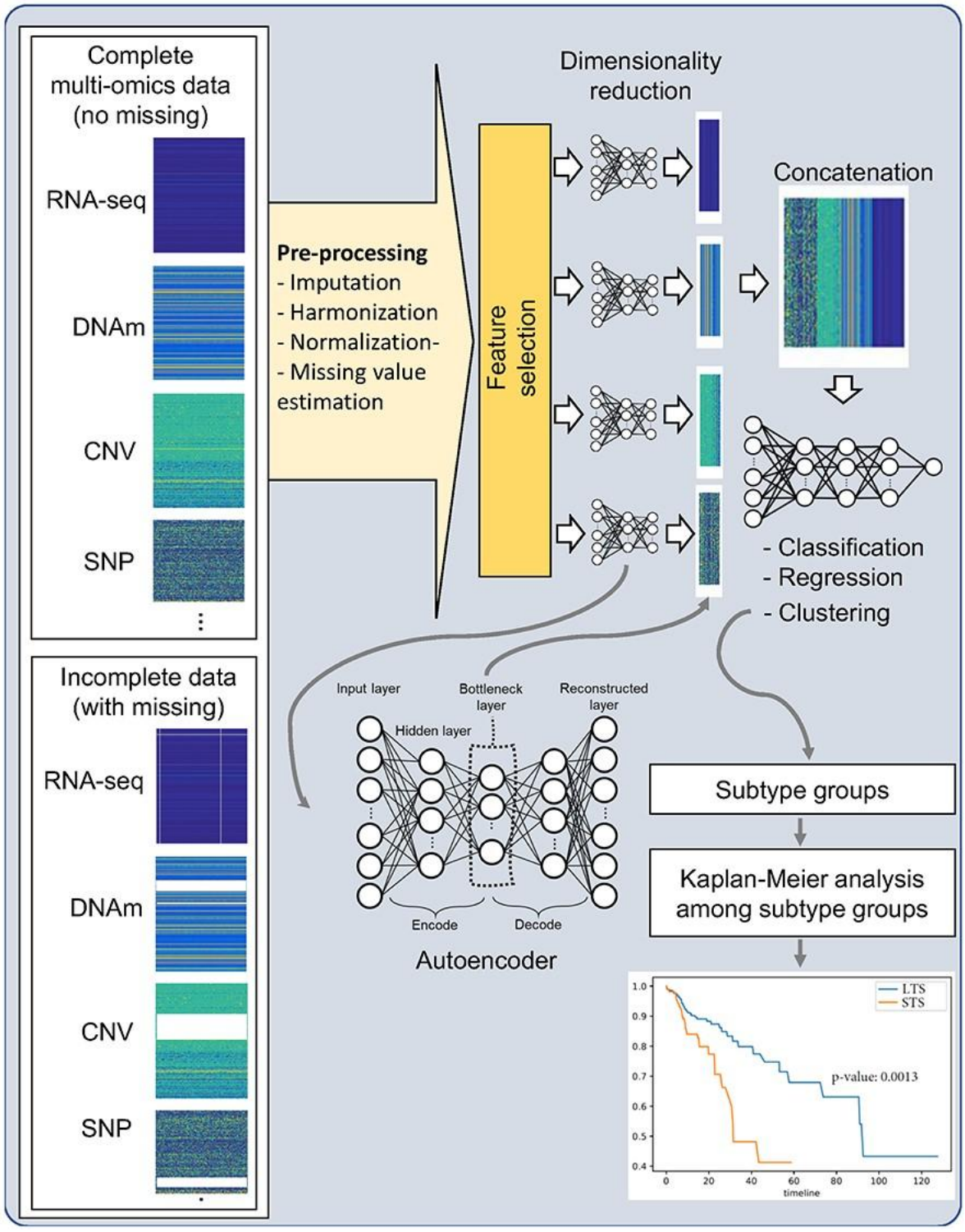


Figure 2 Pipeline of multi-omics data integration analyses (Adopted from Kang et al., 2021)

## 6 Ethical Considerations and Challenges

### 6.1 Data privacy and security

The integration of AI in genomic research necessitates the collection and sharing of vast amounts of sensitive genomic data, raising significant concerns about data privacy and security. The potential for breaches in patient

privacy is a major issue, as unauthorized access to genomic data can lead to misuse and exploitation (Azencott et al., 2018). Techniques such as fully homomorphic encryption (FHE) have been developed to enable the secure analysis of encrypted data, allowing for privacy-preserving applications in genomics without compromising the confidentiality of the data (Wood et al., 2020). Privacy-preserving AI techniques, including federated learning, have been proposed to protect individual privacy while enabling collaborative research (Torkzadehmahani et al., 2020). These methods aim to balance the need for data sharing in scientific research with the imperative to protect participant privacy.

### 6.2 Bias in machine learning models

Bias in AI models is a critical ethical concern, particularly in the context of genomic research where biased data can lead to prejudiced outcomes. AI systems trained on biased datasets may produce results that disproportionately affect certain demographic groups, leading to issues of fairness and discrimination (Ntoutsi et al., 2020). For instance, AI tools developed using data from a homogenous population may not perform well when applied to diverse populations, exacerbating health disparities (Char, 2022). Addressing bias requires embedding ethical and legal principles in the design, training, and deployment of AI systems to ensure equitable outcomes. This includes deliberate efforts to minimize bias through diverse and representative data collection and rigorous validation processes.

### 6.3 Regulatory and legal issues

The combination of AI and genomics introduces complex regulatory and legal challenges. The rapid advancement of these technologies often outpaces the development of regulatory frameworks, leading to uncertainties in their governance (Botes, 2023). The precautionary principle, which aims to prevent irreversible harm, has been suggested as a regulatory approach to manage the uncertainties associated with AI and genomics. There is a need for clear regulations to address data privacy, security, and ethical concerns in clinical AI systems (Gedefaw et al., 2023). The development of comprehensive and adaptive regulatory frameworks is essential to ensure the safe and ethical use of AI in genomic research, balancing innovation with the protection of individual rights and societal values.

## 7 Advances in AI Tools and Platforms for Genomic Research

### 7.1 Development of AI-driven genomic tools

The integration of artificial intelligence (AI) into genomic research has led to the development of sophisticated tools that enhance the analysis and interpretation of complex genomic data. Machine learning (ML) applications have been particularly transformative, enabling the annotation of sequence elements and the analysis of epigenetic, proteomic, and metabolomic data. Platforms like PrismML exemplify this advancement by allowing users to perform multivariate machine learning on large genomic datasets, facilitating the identification of genotype-phenotype patterns and the prediction of clinical outcomes (Reddy et al., 2020). These tools leverage the computational power of cloud computing to handle the intensive processing requirements, making analyses faster and more scalable.

### 7.2 Cloud computing and big data integration

The advent of cloud computing has revolutionized the way genomic data is processed and analyzed. Platforms such as the Genomics Virtual Laboratory (GVL) and Sherlock provide scalable, flexible, and accessible computational resources that are essential for handling the vast amounts of data generated by next-generation sequencing technologies (Afgan et al., 2015). These platforms offer a range of analysis and visualization tools, workflow management systems, and the ability to add or remove compute nodes as needed, thereby meeting the diverse demands of genomic researchers. The elasticity, reproducibility, and privacy features of cloud computing make it ideally suited for large-scale reanalysis of publicly available archived data, including privacy-protected datasets (Langmead and Nellore, 2018)。

### 7.3 Open-source platforms and collaboration

Open-source platforms play a crucial role in fostering collaboration and innovation in genomic research. Tools like Sherlock and Machado provide comprehensive frameworks for storing, querying, and analyzing large

genomic datasets, facilitating data sharing and collaborative efforts among researchers (Yukselen et al., 2020). These platforms are designed to be user-friendly and accessible, with features such as graphical user interfaces and modular process designs that simplify the creation and execution of complex data processing pipelines. By leveraging modern big data technologies, these open-source platforms empower researchers to manage, analyze, and integrate diverse genomic data, driving advancements in precision medicine and personalized healthcare (Vadapalli et al., 2022).

## 8 Concluding Remarks

The integration of artificial intelligence (AI) and machine learning (ML) into genomic research has significantly transformed the field, enabling the analysis of large, complex datasets with unprecedented accuracy and efficiency. ML methods, including supervised, semi-supervised, and unsupervised learning, have been effectively applied to genome sequencing data, aiding in the annotation of sequence elements and the analysis of epigenetic, proteomic, and metabolomic data. Deep learning models have shown superior performance in specific genomic tasks, such as predicting gene expression levels and identifying genomic elements like promoters and enhancers. These models are particularly effective in handling high-dimensional data, which is common in genomics. AI and ML approaches are crucial in precision medicine, where they help integrate genetic, environmental, and lifestyle factors to diagnose and treat diseases more accurately. These methods facilitate the analysis of whole genome and exome sequencing data, contributing to personalized treatment plans. Bibliometric analyses reveal that AI applications in biotechnology and applied microbiology are rapidly evolving, with significant contributions from global institutions. Key research areas include deep learning, prediction models, and systems biology.

The future of AI in genomic research is promising, with several potential advancements on the horizon. The integration of deep learning with multi-scale and multimodal data analysis is expected to drive significant advancements in precision medicine, enabling more comprehensive and accurate models of disease progression and treatment. Future research will likely focus on integrating AI-generated predictive knowledge with traditional causal concepts in molecular genetics. This integration is essential for developing robust scientific understanding and effective policies in genomic medicine. As AI applications in biology continue to grow, there will be a need for improved standards and practices in publishing and experimental design. This will ensure the reliability and reproducibility of AI-driven research findings.

To fully harness the potential of AI in genomic research, the following recommendations are proposed. Encourage collaboration between AI experts, biologists, and medical researchers to develop more sophisticated models and algorithms that can address complex biological questions. Emphasize the importance of high-quality data and robust pre-processing techniques to avoid the pitfalls associated with poor data quality, which can lead to inaccurate models and predictions. Address ethical concerns related to the use of AI in genomics, particularly regarding data privacy, consent, and the potential for bias in AI models. Developing ethical guidelines and frameworks will be crucial for the responsible use of AI in this field. Continued investment in AI research and development is essential to drive innovation and maintain the momentum in genomic research. This includes funding for both basic and applied research, as well as support for training and education in AI and genomics. By following these recommendations, the field of genomic research can continue to benefit from the transformative potential of AI, leading to more accurate, efficient, and personalized approaches to understanding and treating genetic diseases.

### Acknowledgments

Cuixi Academy of Biotechnology provided critical resources that facilitated this research, and we express our gratitude.We also would like to thank two anonymous peer reviewers for their careful review and valuable comments.

### Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

Abbas Z., Tayara H., and Chong K., 2021, ZayyuNet–a unified deep learning model for the identification of epigenetic modifications using raw genomic sequences, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19: 2533-2544.
https://doi.org/10.1109/TCBB.2021.3083789

Afgan E., Sloggett C., Goonasekera N., Makunin I., Benson D., Crowe M., Gladman S., Kowsar Y., Pheasant M., Horst R., and Lonie A., 2015, Genomics virtual laboratory: a practical bioinformatics workbench for the cloud, PLoS ONE, 10(10): e0140829.
https://doi.org/10.1371/journal.pone.0140829

Angermueller C., Pärnamaa T., Parts L., and Stegle O., 2016, Deep learning for computational biology, Molecular Systems Biology, 12(7): 878.
https://doi.org/10.15252/msb.20156651

Avsec Ž., Agarwal V., Visentin D., Ledsam J., Grabska-Barwinska A., Taylor K., Assael Y., Jumper J., Kohli P., and Kelley D., 2021, Effective gene expression prediction from sequence by integrating long-range interactions, Nature Methods, 18: 1196-1203.
https://doi.org/10.1038/s41592-021-01252-x

Azencott C.A., 2018, Machine learning and genomics: precision medicine versus patient privacy, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128): 20170350.
https://doi.org/10.1098/rsta.2017.0350

Benkirane H., Pradat Y., Michiels S., and Cournède P.H., 2023, CustOmics: a versatile deep-learning based strategy for multi-omics integration, PLOS Computational Biology, 19(3): e1010921.
https://doi.org/10.1371/journal.pcbi.1010921

Botes M., 2023, Regulating scientific and technological uncertainty: the precautionary principle in the context of human genomics and AI, South African Journal of Science, 119(5-6): 1-6.
https://doi.org/10.17159/sajs.2023/15037

Camacho D., Collins K., Powers R., Costello J., and Collins J., 2018, Next-generation machine learning for biological networks, Cell, 173: 1581-1592.
https://doi.org/10.1016/j.cell.2018.05.015

Char D., 2022, Challenges of local ethics review in a global healthcare AI market, The American Journal of Bioethics, 22: 39-41.
https://doi.org/10.1080/15265161.2022.2055214

Gedefaw L.F., Liu C., Ip R.K.L., Tse H.F., Yeung M.H.Y., Yip S.L., and Huang C., 2023, Artificial intelligence-assisted diagnostic cytology and genomic testing for hematologic disorders, Cells, 12(13): 1755.
https://doi.org/10.3390/cells12131755

Jun G., Wing M.K., Abecasis G.R., and Kang H.M., 2015, An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data, Genome Research, 25(6): 918-925.
https://doi.org/10.1101/gr.176552.114

Kang M., Ko E., and Mersha T.B., 2021, A roadmap for multi-omics data integration using deep learning, Briefings in Bioinformatics, 23(1): bbab454.
https://doi.org/10.1093/bib/bbab454.

Karim M.R., Beyan O., Zappa A., Costa I.G., Rebholz-Schuhmann D., Cochez M., and Decker S., 2020, Deep learning-based clustering approaches for bioinformatics, Briefings in Bioinformatics, 22(1): 393-415.
https://doi.org/10.1093/bib/bbz170

Kelley D.R., 2019, Cross-species regulatory sequence activity prediction, PLoS Computational Biology, 16(7): e1008050.
https://doi.org/10.1371/journal.pcbi.1008050

Kelley D., Reshef Y., Bileschi M., Belanger D., McLean C., and Snoek J., 2017, Sequential regulatory activity prediction across chromosomes with convolutional neural networks, Genome Research, 28: 739-750.
https://doi.org/10.1101/161851

Koumakis L., 2020, Deep learning models in genomics; are we there yet, Computational and Structural Biotechnology Journal, 18: 1466-1473.
https://doi.org/10.1016/j.csbj.2020.06.017

Langmead B., and Nellore A., 2018, Cloud computing for genomic data analysis and collaboration, Nature Reviews Genetics, 19: 208-219.
https://doi.org/10.1038/nrg.2017.113.

Leung M., Delong A., Alipanahi B., and Frey B., 2016, Machine learning in genomic medicine: a review of computational problems and data sets, Proceedings of the IEEE, 104: 176-197.
https://doi.org/10.1109/JPROC.2015.2494198

Li Y., Wu F., and Ngom A., 2016, A review on machine learning principles for multi-view biological data integration, Briefings in Bioinformatics, 19: 325-340.
https://doi.org/10.1093/bib/bbw113

Libbrecht M., and Noble W., 2015, Machine learning applications in genetics and genomics, Nature Reviews Genetics, 16: 321-332.
https://doi.org/10.1038/nrg3920

Liu J., Li J., Wang H., and Yan J., 2020, Application of deep learning in genomics, Science China Life Sciences, 63: 1860-1878.
https://doi.org/10.1007/s11427-020-1804-5

Mirza B., Wang W., Wang J., Choi H., Chung N.C., and Ping P., 2019, Machine learning and integrative analysis of biomedical big data, Genes, 10(2): 87.
https://doi.org/10.3390/genes10020087

Nicora G., Vitali F., Dagliati A., Geifman N., and Bellazzi R., 2020, Integrated multi-omics analyses in oncology: a review of machine learning methods and tools, Frontiers in Oncology, 10: 1030.
https://doi.org/10.3389/fonc.2020.01030

Ntoutsi E., Fafalios P., Gadiraju U., Iosifidis V., Nejdl W., Vidal M., Ruggieri S., Turini F., Papadopoulos S., Krasanakis E., Kompatsiaris I., Kinder-Kurlanda K., Wagner C., Karimi F., Fernández M., Alani H., Berendt, B., Kruegel T., Heinze C., Broelemann K., Kasneci G., Tiropanis T., and Staab S., 2020, Bias in data-driven artificial intelligence systems-an introductory survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3): e1356.
https://doi.org/10.1002/widm.1356

Rakocevic G., Semenyuk V., Lee W., Spencer J., Browning J., Johnson I., Arsenijević V., Nadj J., Ghose K., Suciu M., Ji S., Demir G., Li L., Toptas B., Dolgoborodov A., Pollex B., Spulber I., Glotova I., Kómár P., Stachyra A., Li Y., Popovic M., Källberg M., Jain A., and Kural D., 2019, Fast and accurate genomic analyses using genome graphs, Nature Genetics, 51: 354-362.
https://doi.org/10.1038/s41588-018-0316-4

Reddy A., Flemming D., Selitsky S., Pavel A., Alexe G., and Bhanot G., 2020, Abstract 858: PrismML: a machine learning platform to query genotype-phenotype patterns in large genomics studies, Cancer Research, 80: 858-858.
https://doi.org/10.1158/1538-7445.am2020-858

Reel P.S., Reel S., Pearson E., Trucco E., and Jefferson E., 2021, Using machine learning approaches for multi-omics data analysis: a review, Biotechnology Advances, 49: 107739.
https://doi.org/10.1016/j.biotechadv.2021.107739

Routhier E., and Mozziconacci J., 2022, Genomics enters the deep learning era, PeerJ, 10: e13613.
https://doi.org/10.7717/peerj.13613

Saha, G., Babur M., Khan M., Saha H., Kumar D., 2023, Integrative analysis of multi-omics data with deep learning: challenges and opportunities in bioinformatics, Journal of Propulsion Technology, 23(1): bbab454.
https://doi.org/10.52783/tjjpt.v44.i3.488

Schmidt B., and Hildebrandt A., 2020, Deep learning in next-generation sequencing, Drug Discovery Today, 26: 173-180.
https://doi.org/10.1016/j.drudis.2020.10.002

Talukder A., Barham C., Li X., and Hu H., 2020, Interpretation of deep learning in genomics and epigenomics, Briefings in Bioinformatics, 22(3): bbaa177.
https://doi.org/10.1093/bib/bbaa177

Torkzadehmahani R., Nasirigerdeh R., Blumenthal D., Kacprowski T., List M., Matschinske J., Späth J., Wenke N., Bihari B., Frisch T., Hartebrodt A., Hauschild A., Heider D., Holzinger A., Hötzendorfer W., Kastelitz M., Mayer R., Nogales C., Pustozerova A., Röttger R., Schmidt H., Schwalber A., Tschohl C., Wohner A., and Baumbach J., 2020, Privacy-preserving artificial intelligence techniques in biomedicine, Methods of Information in Medicine, 61: e12-e27.
https://doi.org/10.1055/s-0041-1740630

Vadapalli S., Abdelhalim H., Zeeshan S., and Ahmed Z., 2022, Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine, Briefings in Bioinformatics, 23(5): bbac191.
https://doi.org/10.1093/bib/bbac191

VanRaden P., Bickhart D., and O'Connell J., 2019, Calling known variants and identifying new variants while rapidly aligning sequence data, Journal of Dairy Science, 102(4): 3216-3229.
https://doi.org/10.3168/jds.2018-15172

Wood A., Najarian K., and Kahrobaei D., 2020, Homomorphic encryption for machine learning in medicine and bioinformatics, ACM Computing Surveys (CSUR), 53: 1-35.
https://doi.org/10.1145/3394658

Wörheide M., Krumsiek J., Kastenmüller G., and Arnold M., 2021, Multi-omics integration in biomedical research-a metabolomics-centric review, Analytica Chimica Acta, 1141: 144-162.
https://doi.org/10.1016/j.aca.2020.10.038

Wu J., and Zhao, Y., 2019, Machine learning technology in the application of genome analysis: a systematic review, Gene, 705: 149-156.
https://doi.org/10.1016/j.gene.2019.04.062

Yukselen O., Turkyilmaz O., Ozturk A., Garber M., and Kucukural A., 2020, Dolphin next: a distributed data processing platform for high throughput genomics, BMC Genomics, 21: 1-16.
https://doi.org/10.1186/s12864-020-6714-x

Zou J., Huss M., Abid A., Mohammadi P., Torkamani A., and Telenti A., 2018, A primer on deep learning in genomics, Nature Genetics, 51(1): 12-18.
https://doi.org/10.1038/s41588-018-0295-5