

## Bioinformatics in the Age of Big Data: Leveraging Computational Tools for Biological Discoveries

Xiaoming Liu, Wei Zhang ✉

WuXi AppTec Co., Ltd, Wuxi, 518083, Jiangsu, China

✉ Corresponding author: [zhangwei0901@126.com](mailto:zhangwei0901@126.com)

Computational Molecular Biology, 2024, Vol.14, No.4 doi: [10.5376/cmb.2024.14.0020](https://doi.org/10.5376/cmb.2024.14.0020)

Received: 20 Jun., 2024

Accepted: 05 Aug., 2024

Published: 25 Aug., 2024

**Copyright** © 2024 Liu and Zhang. This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Preferred citation for this article:

Liu X.M., and Zhang W., 2024, Bioinformatics in the age of big data: leveraging computational tools for biological discoveries, Computational Molecular Biology, 14(4): 173-181 (doi: [10.5376/cmb.2024.14.0020](https://doi.org/10.5376/cmb.2024.14.0020))

**Abstract** The rise of big data has changed the landscape of bioinformatics, providing new opportunities for biological discoveries, but also bringing significant computational challenges. This study provides an in-depth analysis of bioinformatics in the era of big data, focusing on the evolution of computing tools and their role in modern biology. It reviews the usage process from early bioinformatics tools to current high-throughput data analysis, as well as the expansion of public biological databases. In the context of genomics, proteomics, and multi omics integration, key computing methods, including machine learning algorithms, data mining, and high-performance computing, are discussed. Explore future development directions such as artificial intelligence, cloud computing, and open source collaboration platforms, in order to provide new perspectives for researchers and promote further innovation and development in bioinformatics.

**Keywords** Bioinformatics; Big data; Machine learning; Genomics; High-performance computing

## 1 Introduction

Bioinformatics, an interdisciplinary field that merges biology, computer science, and information technology, has become indispensable in modern biological research. The advent of high-throughput technologies has led to an unprecedented accumulation of biological data, often referred to as "Big Data". This data encompasses various types, including genomic sequences, protein structures, and complex biological networks, which require sophisticated computational tools for effective analysis and interpretation (Gauthier et al., 2018; Shoaib et al., 2021). The exponential growth of biological data presents both opportunities and challenges, necessitating the development of new methodologies and tools to manage, analyze, and derive meaningful insights from these vast datasets (Khan et al., 2022).

Computational tools have revolutionized the field of bioinformatics by enabling the systematic organization, analysis, and understanding of complex biological data. These tools range from traditional statistical methods to advanced machine learning and deep learning algorithms, which are particularly adept at handling large-scale data (Gupta et al., 2021; Raina, 2023). For instance, graph neural networks (GNNs) have been employed to analyze biological networks, aiding in protein function prediction and drug discovery (Muzio et al., 2020). Similarly, deep learning techniques have shown promise in various bioinformatics applications, including genomic data analysis and disease diagnosis. The integration of computational tools in bioinformatics not only accelerates data processing but also enhances the predictive power and reproducibility of biological research.

This study provides a comprehensive overview of the current state of bioinformatics in the context of big data. It explores various computational tools and methods developed to address the challenges posed by large-scale biological data, analyzing the applications of deep learning and other advanced computational technologies in bioinformatics. Special emphasis is placed on their impact in fields such as genomics, proteomics, and systems biology. We address both theoretical and practical issues associated with these tools and propose future research directions to further advance the field.

## 2 Evolution of Bioinformatics in the Big Data Era

### 2.1 Historical background and early bioinformatics tools

The origins of bioinformatics can be traced back over 50 years, long before the advent of next-generation sequencing technologies. The field began in the early 1960s with the application of computational methods to protein sequence analysis, including *de novo* sequence assembly and the creation of biological sequence databases. Early bioinformatics tools were developed to handle the increasing amount of biological data generated by molecular biology methods and the miniaturization of computers. These foundational tools laid the groundwork for the integration of computational techniques into biological research, enabling the systematic organization, analysis, and understanding of biological data (Gauthier et al., 2018; Shoaib et al., 2021).

### 2.2 Shift towards high-throughput data analysis

The rapid development of high-throughput sequencing (HTS) techniques has significantly transformed bioinformatics, ushering in the era of big data in biology. High-throughput technologies have expanded the availability and quantity of molecular data, necessitating the development of new computational tools for data analysis. The emergence of next-generation sequencing programs has led to unparalleled growth in whole-genome sequencing projects, such as the 100 000 human genomes and 1 000 plant species initiatives. This shift towards high-throughput data analysis has also seen the rise of deep learning and machine learning methodologies, which are now commonly used to identify patterns, make predictions, and model biological processes (Koumakis, 2020). Tools like TBtools have been developed to provide user-friendly interfaces for wet-lab biologists, facilitating the analysis of large-scale datasets (Figure 1) (Chen et al., 2020).

The graphical features of TBtools, combined with large-scale data generated by HTS (High-Throughput Sequencing) technologies, have greatly enhanced the efficiency of biological research, enabling biologists to better understand complex genomic structures and functional patterns. This also reflects the growing importance of methods such as machine learning and deep learning in recognizing patterns, making predictions, and simulating biological processes as biological data continues to increase.

### 2.3 Growth of public biological databases

The exponential growth of biological data has necessitated the creation and expansion of public biological databases. Institutions like the European Bioinformatics Institute (EMBL-EBI) have played a crucial role in maintaining comprehensive data resources, which stored over 390 petabytes of raw data by the end of 2020 (Khan et al., 2022). Databases such as KEGG have become essential for the biological interpretation of genome sequences and other high-throughput data, providing practical value for researchers. The integration of digital information, biological data, electronic medical records, and clinical information has created a tsunami of opportunities for knowledge discovery, emphasizing the need for open data sources, open access to software, and the implementation of machine learning and artificial intelligence. The continuous development and maintenance of these databases are critical for supporting the ever-growing demands of bioinformatics research and applications (Solanki et al., 2020).

## 3 Key Computational Tools for Big Data in Biology

### 3.1 Machine learning algorithms in bioinformatics

#### 3.1.1 Application in gene expression and regulation studies

Machine learning algorithms have become indispensable in the analysis of gene expression and regulation. These algorithms facilitate the automatic extraction and selection of features from large datasets, enabling the generation of predictive models that can efficiently study complex biological systems. For instance, machine learning techniques are integrated with bioinformatics methods to enhance training and validation processes, identify interpretable features, and investigate models (Auslander et al., 2021). Probabilistic graphical models have been employed to reconstruct gene regulatory networks from transcriptomics and genomics data, providing a concise representation of complex gene regulatory relationships (Cheng, 2020).

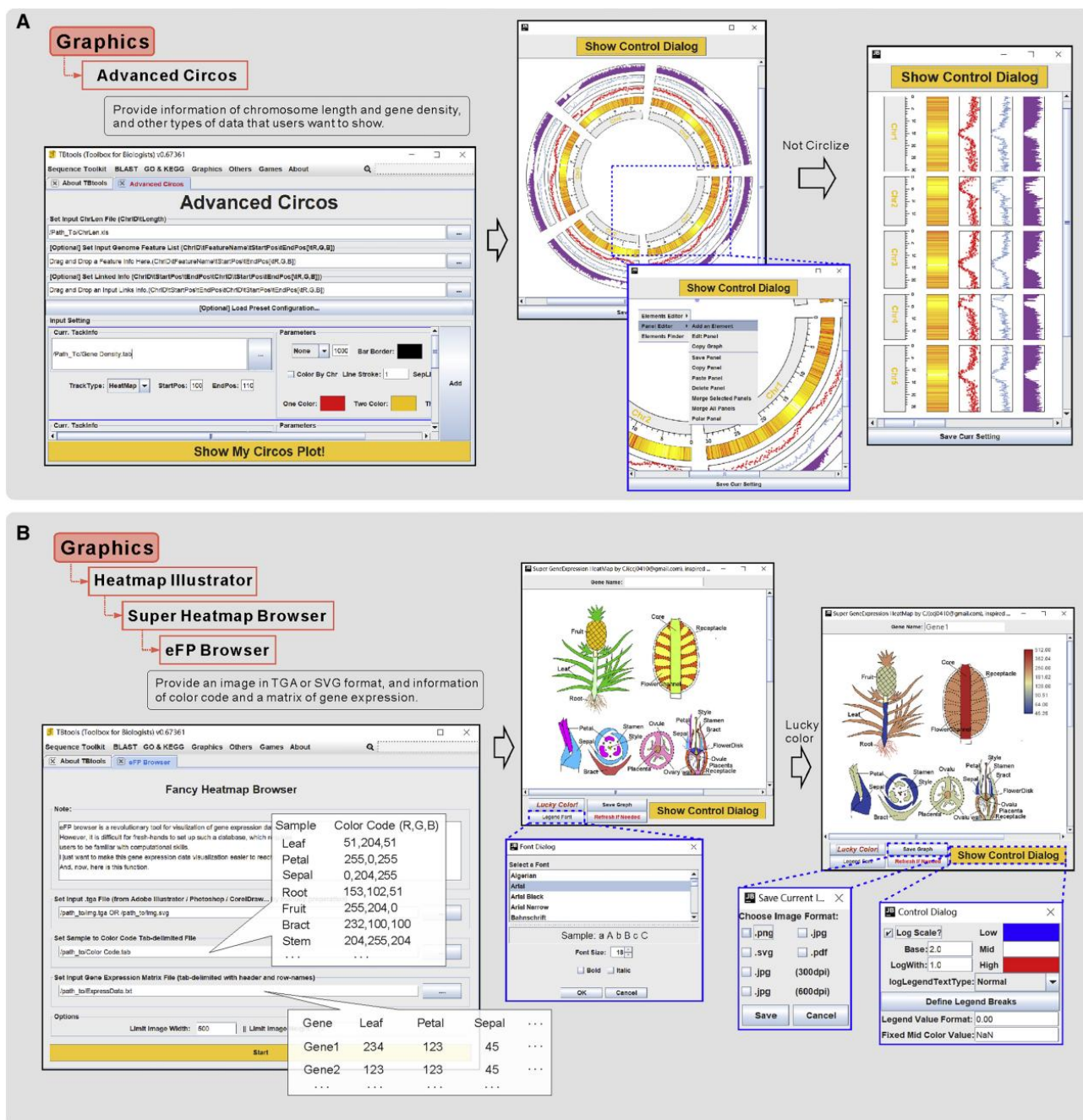


Figure 1 Examples Demonstrating the “Advanced Circos” and “eFP Browser” Functions in TBtools (Adopted from Chen et al., 2020)

### 3.1.2 Predictive models for protein structure and function

Predicting protein structure and function is a major challenge in bioinformatics, which has seen significant advancements through the application of machine learning. Deep learning methods, such as convolutional neural networks, have been used to predict residue-residue contacts and reconstruct protein tertiary structures from sequence data, achieving top rankings in critical assessments of protein structure prediction. Bioinformatics tools continue to evolve, improving the accuracy of predictions regarding protein functionality, homology, mutations, and evolutionary processes (Hernández-Domínguez et al., 2019).

### 3.1.3 Integration of multi-omics data

The integration of multi-omics data is crucial for a comprehensive understanding of biological systems. High-performance computing (HPC) infrastructure has empowered machine learning and optimization algorithms to analyze and integrate large-scale omics data. For example, large-scale data-driven optimization algorithms have been developed to reconstruct high-resolution 3D genome structures from Hi-C data, which can be used to study

gene function, gene expression, and genome methylation (Kanehisa, 2019). Deep learning architectures have been applied across various bioinformatics domains, including omics, biomedical imaging, and signal processing, to transform biomedical big data into valuable knowledge (Min et al, 2016).

### 3.2 Data mining and pattern recognition

Data mining and pattern recognition are essential for extracting meaningful insights from large biological datasets. Formal concept analysis (FCA) is one such method that allows the examination of structural properties of data, facilitating applications such as gene data analysis, biomarker discovery, and protein-protein interaction analysis (Roscoe et al., 2022). Graph neural networks (GNNs) have been employed to analyze biological networks, predicting protein functions, protein-protein interactions, and aiding in drug discovery and development.

### 3.3 High-performance computing (HPC) for genomic data

High-performance computing (HPC) plays a pivotal role in managing and analyzing the vast amounts of genomic data generated by next-generation sequencing technologies. HPC infrastructure, such as GPUs and HPC clusters, supports the execution of large-scale machine learning and optimization algorithms, enabling the fast analysis of massive DNA, RNA, and protein sequence data (Kashyap et al., 2016; Cheng, 2020). These computational resources are critical for addressing bioinformatics problems, such as the construction of co-expression and regulatory networks, detection of protein complexes, and querying heterogeneous disease networks.

## 4 Challenges in Bioinformatics and Big Data Management

### 4.1 Data storage and accessibility issues

The exponential growth of biological data, driven by advancements in high-throughput sequencing and other technologies, has created significant challenges in data storage and accessibility. For instance, the European Bioinformatics Institute (EMBL-EBI) stored over 390 petabytes of raw data by the end of 2020, and this volume is expected to reach the exascale within the next few years (Shahid, 2023). Platforms like Sherlock have been developed to address these challenges by providing cloud-based solutions for storing, converting, querying, and sharing large datasets, thereby streamlining bioinformatics data management. However, the sheer volume and complexity of the data necessitate continuous improvements in storage technologies and data management practices to ensure that researchers can efficiently access and utilize these vast resources (Gauthier et al., 2018).

### 4.2 Managing data complexity and integration

The complexity of biological data, which often includes diverse data types such as genomic sequences, protein structures, and interaction networks, poses significant challenges for integration and analysis. Tools like TBtools have been developed to facilitate the handling of such complex datasets by providing a user-friendly interface and a wide range of functions for data processing and visualization (Chen et al., 2020). The integration of deep learning techniques has shown promise in transforming biomedical big data into valuable knowledge, although it also introduces new challenges related to data heterogeneity and the need for specialized computational resources. Platforms like Sherlock further aid in managing data complexity by converting various structured data into optimized formats, enabling efficient distributed analytical queries (Bohár et al., 2022).

### 4.3 Ethical concerns and data privacy

The management of large-scale biological data also raises significant ethical concerns and data privacy issues. The sensitive nature of personal health and genomic data necessitates robust privacy protections to prevent unauthorized access and data breaches. Advances in cryptography, such as homomorphic encryption, offer potential solutions by allowing data to be stored and computed on in encrypted form, without the need for decryption keys (Dowlin et al., 2017). This approach enables researchers to outsource data storage to untrusted clouds while maintaining data privacy. Additionally, the integration of ethical guidelines and best practices is crucial to ensure the responsible use of bioinformatics tools and data, particularly as the field continues to evolve with new technologies and methodologies (Shahid, 2023).

## 5 Applications of Bioinformatics in Biological Discoveries

### 5.1 Drug discovery and development

Bioinformatics has revolutionized the field of drug discovery and development by providing computational tools and techniques that accelerate the identification of drug targets and the screening of drug candidates. High-throughput data, such as genomic, epigenetic, transcriptomic, and proteomic data, have significantly contributed to mechanism-based drug discovery and drug repurposing (Xia, 2017). The integration of bioinformatics in drug discovery allows for more realistic protein-ligand docking experiments and more informative virtual screening, which are essential for identifying nontoxic and efficient drugs (Ramharack and Soliman, 2018; Chen, 2024). Bioinformatics tools facilitate the characterization of side effects and the prediction of drug resistance, making the drug development process more efficient and targeted.

### 5.2 Functional genomics and systems biology

#### 5.2.1 Identifying gene networks and pathways

Bioinformatics plays a crucial role in identifying gene networks and pathways by analyzing high-throughput sequencing data. Tools and techniques developed in bioinformatics help in the systematic organization and analysis of biological data, which is essential for understanding complex biological pathways and mechanisms involved in systems biology. The use of formal concept analysis (FCA) and graph neural networks (GNN) has shown promise in identifying influential nodes in gene regulatory networks and predicting gene interactions. These computational approaches enable researchers to map out gene networks and understand their regulatory mechanisms, which is vital for functional genomics studies (Figure 2) (Muzio et al., 2020; Roscoe et al., 2022).

Graph convolutional networks (GCN) is a subset of graph neural network (GNN) that adopts the highly successful architecture of convolutional neural networks (CNN) to handle graph-structured data. In biological network analysis, this hierarchical integration of information through GNN allows the precise identification of key nodes and interactions between genes that influence gene expression within complex gene regulatory networks. By utilizing this approach, researchers can gain a better understanding of gene regulatory mechanisms and predict potential regulatory relationships or the impact of gene mutations, which is of great significance for disease diagnosis and drug development.

#### 5.2.2 Analyzing transcriptomic and proteomic data

The advent of next-generation sequencing (NGS) and mass spectrometry has generated vast amounts of transcriptomic and proteomic data, necessitating the development of bioinformatics tools for data analysis. These tools are essential for detecting sequence variation, gene expression, and protein interactions (Mohanasundaram et al., 2023). Bioinformatics pipelines for NGS data analysis include sequence generation, alignment to a reference genome, and interpretation of results, which are crucial for understanding gene-to-gene interactions and identifying phenotype-differentiating pathways (Koumakis et al., 2017). The integration of omic approaches in bioinformatics allows for a comprehensive analysis of transcriptomic and proteomic data, providing insights into the functional roles of genes and proteins in various biological processes.

#### 5.2.3 Systems-level understanding of cellular functions

Bioinformatics enables a systems-level understanding of cellular functions by integrating data from various omic technologies. The analysis of high-throughput data, such as genomic, transcriptomic, and proteomic data, helps in constructing detailed models of cellular functions and interactions (Cheba, 2019; Shoaib et al., 2021). Systems biology approaches in bioinformatics involve the use of computational tools to analyze gene regulatory networks, protein-protein interactions, and metabolic pathways, providing a holistic view of cellular processes. This systems-level understanding is crucial for identifying potential therapeutic targets and understanding the molecular basis of diseases.

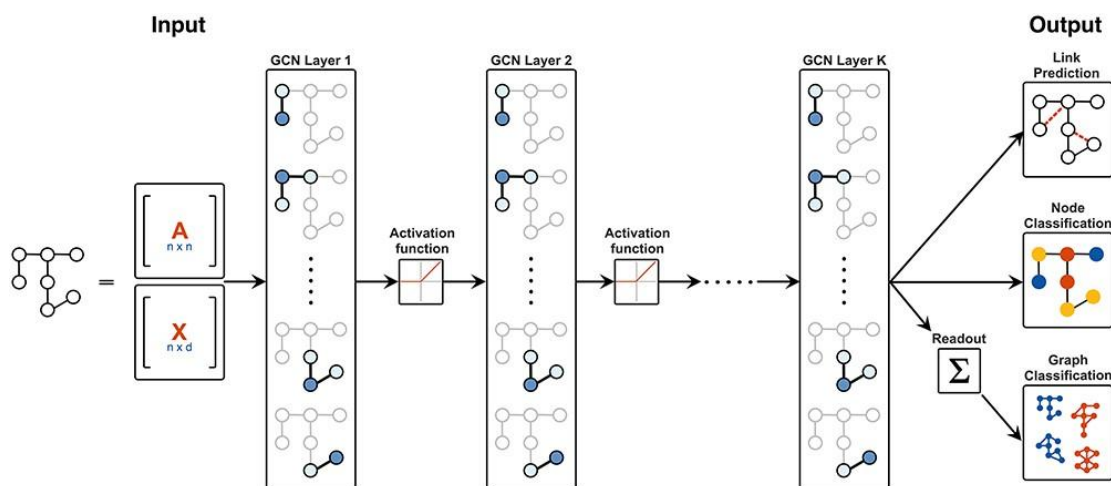


Figure 2 A visual depiction of a k-layer GCN (Adopted from Muzio et al., 2020)

Image caption: Each layer of the GCN aggregates over the neighborhood of each node, using the node representations from the previous layer in the network. The aggregations in each layer then pass through an activation function before going to the next layer. This network can be used to produce various different outputs: for predicting new edges in the input network (link prediction), classifying individual nodes in the input graph (node classification), or classifying the entire input graph (graph classification) (Adopted from Muzio et al., 2020)

### 5.3 Evolutionary studies and phylogenetics

Bioinformatics has significantly advanced the field of evolutionary studies and phylogenetics by providing tools for the analysis of genetic sequences and the construction of phylogenetic trees. The availability of high-throughput sequencing data has enabled researchers to study the evolutionary relationships between species and understand the genetic basis of adaptation and speciation. Bioinformatics tools facilitate the comparison of genomes, identification of conserved sequences, and reconstruction of evolutionary histories, which are essential for phylogenetic studies (Roscoe et al., 2022). The integration of bioinformatics in evolutionary studies allows for the analysis of large-scale genomic data, providing insights into the evolutionary dynamics of populations and species.

## 6 Future Directions and Innovations in Bioinformatics

### 6.1 Artificial intelligence and deep learning in bioinformatics

Artificial intelligence (AI) and deep learning (DL) have revolutionized bioinformatics by providing powerful tools to analyze and interpret vast amounts of biological data. Deep learning, a subset of machine learning, has shown remarkable success in various bioinformatics applications, including omics data analysis, biomedical imaging, and signal processing (Tang et al., 2019). The flexibility and adaptability of deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have enabled researchers to uncover complex relationships within large-scale biological datasets (Li et al., 2019).

Recent advancements in deep learning have led to the development of ensemble deep learning methods, which combine multiple models to improve accuracy, stability, and reproducibility in bioinformatics research. These methods have been applied to a wide range of bioinformatics tasks, from basic sequence analysis to systems biology, demonstrating their potential to address diverse challenges in the field (Cao et al., 2020). The integration of AI with molecular databases has paved the way for novel applications and improved user-friendly interfaces. By incorporating deep learning and deep reasoning, molecular databases can partially self-maintain and perform comparative analyses of newly submitted data against existing datasets, thereby enhancing the efficiency and accuracy of bioinformatics analyses.

### 6.2 Cloud computing and distributed databases

The rapid growth of big data in bioinformatics has necessitated the adoption of cloud computing and distributed databases to manage and analyze large datasets efficiently. Cloud computing offers scalable and flexible resources that can handle the computational demands of bioinformatics applications, enabling researchers to perform

complex analyses without the need for extensive local infrastructure (Li et al., 2020). Edge computing, which involves processing data closer to its source, has also gained traction in bioinformatics. The collaboration between cloud and edge computing, known as edge-cloud polarization, allows for efficient data processing and analysis in resource-constrained environments, such as Internet of Things (IoT) scenarios. This collaborative approach leverages the strengths of both computing paradigms, providing a robust framework for bioinformatics research.

### 6.3 Collaborative platforms for open-source bioinformatics

The open-source movement has significantly impacted bioinformatics by fostering collaboration and innovation among researchers worldwide. Collaborative platforms enable scientists to share data, tools, and methodologies, accelerating the pace of discovery and reducing redundancy in research efforts. These platforms also promote transparency and reproducibility, which are essential for the validation and verification of bioinformatics findings. Artificial intelligence-enhanced molecular databases exemplify the potential of collaborative platforms in bioinformatics. By integrating AI and deep learning, these databases can offer more intuitive and user-friendly interfaces, making bioinformatics tools accessible to researchers with varying levels of expertise (Kashangura, 2021). This democratization of bioinformatics resources can drive innovation and facilitate the development of new applications in the field.

## 7 Concluding Remarks

The field of bioinformatics has undergone significant transformations over the past few decades, driven by advancements in computational methods and the exponential growth of biological data. Initially, bioinformatics focused on protein sequence analysis and the development of biological sequence databases in the 1960s. The advent of next-generation sequencing (NGS) technologies in the 1990s and 2000s further accelerated the field, enabling the generation of vast amounts of genomic data at reduced costs. This era of 'Big Data' has necessitated the development of sophisticated bioinformatics tools and software for data analysis, including sequence submission, retrieval, and structure prediction tools. The integration of artificial intelligence and machine learning has further enhanced the ability to analyze and interpret complex biological datasets, leading to novel insights in biomedicine. Visualization techniques have also evolved, providing essential tools for understanding the intricate relationships within large-scale biological data.

The future of bioinformatics holds immense potential for groundbreaking biological discoveries. The continuous improvement in sequencing technologies and computational tools will enable more comprehensive and detailed analyses of genomes, transcriptomes, and proteomes. High-throughput data from large-scale projects, such as the 100 000 Genomes Project, will facilitate the identification of new drug targets and the development of personalized medicine. The integration of diverse data types, including omics data and electronic medical records, will enhance our understanding of complex biological systems and disease mechanisms. Moreover, the application of machine learning and artificial intelligence will drive the discovery of new biomarkers and therapeutic strategies, ultimately advancing precision medicine.

Effective data management and integration are crucial for maximizing the potential of bioinformatics in the age of big data. Standardization and interoperability of data formats are essential to ensure seamless data sharing and collaboration across different research groups and institutions. The development of user-friendly bioinformatics tools and platforms, such as TBtools, can facilitate data analysis for researchers with varying levels of computational expertise. Additionally, the adoption of open data sources and open-source software will promote transparency and reproducibility in bioinformatics research. To address the challenges posed by the sheer volume and complexity of biological data, it is recommended to leverage distributed and parallel computing technologies, as well as graph-based architectures, to optimize data processing and analysis. Finally, integrating data visualization best practices will enhance the interpretability and communication of bioinformatics findings, aiding in the translation of data-driven discoveries into practical applications.

### Acknowledgments

Thanks to the two anonymous peer reviewers for their reading and valuable feedback, and also thank our colleague Yu from the research team for providing the information.

## 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

- Abed R., and Al-Najjar Y., 2021, Bioinformatics storing databases, *Technium BioChemMed*, 2(4): 96-105.  
<https://doi.org/10.47577/biochemmed.v2i4.5335>
- Auslander N., Gussow A.B., and Koonin E.V., 2021, Incorporating machine learning into established bioinformatics frameworks, *International Journal of Molecular Sciences*, 22(6): 2903.  
<https://doi.org/10.3390/ijms22062903>
- Bohár B., Fazekas D., Madgwick M., Csabai L., Olbei M., Korcsmáros T., and Szalay-Beko M., 2022, Sherlock: an open-source data platform to store analyze and integrate big data for computational biologists, *F1000Research*, 10: 409.  
<https://doi.org/10.12688/f1000research.52791.2>
- Cao Y., Geddes T.A., Yang J.Y.H., and Yang P., 2020, Ensemble deep learning in bioinformatics, *Nature Machine Intelligence*, 2(9): 500-508.  
<https://doi.org/10.1038/s42256-020-0217-y>
- Cheba B., 2019, Biotechnological applications of bioinformatics in the post genomic ERA, 2019 International Conference on Computer and Information Sciences (ICIS), 2019: 1-6.  
<https://doi.org/10.1109/ICCISCI.2019.8716439>
- Chen S.Y., 2024 Crossing disease boundaries: how AI drives rare disease drug discovery, *Bioscience Evidence*, 14(1): 21-28.  
<https://doi.org/10.5376/be.2024.14.0003>
- Chen C.J., Chen H., Zhang Y., Thomas H., Frank M.H., He Y., and Xia R., 2020, TBtools - an integrative toolkit developed for interactive analyses of big biological data, *Molecular Plant*, 13(8): 1194-1202.  
<https://doi.org/10.1016/j.molp.2020.06.009>
- Cheng J., 2020, Large-scale machine learning and optimization for bioinformatics data analysis, *Proceedings of the 11th ACM International Conference on Bioinformatics Computational Biology and Health Informatics*, 2020: 1-2.  
<https://doi.org/10.1145/3388440.3415587>
- Dowlin N., Gilad-Bachrach R., Laine K., Lauter K., Naehrig M., and Wernsing J., 2017, Manual for using homomorphic encryption for bioinformatics, *Proceedings of the IEEE*, 105: 552-567.  
<https://doi.org/10.1109/JPROC.2016.2622218>
- Gauthier J., Vincent A.T., Charette S.J., and Derôme N., 2018, A brief history of bioinformatics, *Briefings in Bioinformatics*, 20(6): 1981-1996.  
<https://doi.org/10.1093/bib/bby063>
- Gupta A., Gangotia D., and Mani I., 2021, Bioinformatics tools and software, *Advances in Bioinformatics*, 7(3): 115-122.  
[https://doi.org/10.1007/978-981-33-6191-1\\_2](https://doi.org/10.1007/978-981-33-6191-1_2)
- Hernández-Domínguez E.M., Castillo-Ortega L., García-Esquivel Y., Mandujano-González V., Díaz-Godínez G., and Álvarez-Cervantes J., 2019, Bioinformatics as a tool for the structural and evolutionary analysis of proteins, *Computational Biology and Chemistry*, 2019: 37.  
<https://doi.org/10.5772/intechopen.89594>
- Kanehisa M., 2019, Toward understanding the origin and evolution of cellular organisms, *Protein Science*, 28(11): 1947-1951.  
<https://doi.org/10.1002/pro.3715>
- Kashangura C., 2021, Artificial intelligence enhanced molecular databases can enable improved user-friendly bioinformatics and pave the way for novel applications, *South African Journal of Science*, 117(1-2): 1-2.  
<https://doi.org/10.17159/SAJS.2021/8151>
- Kashyap H., Ahmed H.A., Hoque N., Roy S., and Bhattacharyya D.K., 2016, Big data analytics in bioinformatics: architectures techniques tools and issues, *Network Modeling Analysis in Health Informatics and Bioinformatics*, 5: 1-28.  
<https://doi.org/10.1007/s13721-016-0135-4>
- Khan A.M., Ranganathan S., and Suravajhala P., 2022, Editorial: bioinformatics and the translation of data-driven discoveries, *Frontiers in Genetics*, 13: 902940.  
<https://doi.org/10.3389/fgene.2022.902940>
- 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>
- Koumakis L., Mizzi C., and Potamias G., 2017, Bioinformatics tools for data analysis, *Business Media*, 339-351.  
<https://doi.org/10.1016/B978-0-12-802971-8.00019-5>
- Li H.Y., Tian S.Y., Li Y., Fang Q.M., Tan R.B., Pan Y., Huang C., Xu Y.J., and Gao X., 2020, Modern deep learning in bioinformatics, *Journal of Molecular Cell Biology*, 12(11): 823-827.  
<https://doi.org/10.1093/jmcb/mjaa030>
- Li Y., Huang C., Ding L.Z., Li Z.X., Pan Y.J., and Gao X., 2019, Deep learning in bioinformatics: introduction application and perspective in the big data era, *Methods*, 166: 4-21.  
<https://doi.org/10.1016/J.YMETH.2019.04.008>



- Min S., Lee B., and Yoon S., 2016, Deep learning in bioinformatics, Briefings in Bioinformatics, 18: 851-869.  
<https://doi.org/10.1093/bib/bbw068>
- Mohanasundaram S., Dhatwalia D., Vijayaraghavan P., Alzubaidi L.H., and Makhzuna K., 2023, Bioinformatics: computational approaches for genomics and proteomics, E3S Web of Conferences, 399: 04042.  
<https://doi.org/10.1051/e3sconf/202339904042>
- Muzio G., O'Bray L., and Borgwardt K., 2020, Biological network analysis with deep learning, Briefings in Bioinformatics, 22: 1515-1530.  
<https://doi.org/10.1093/bib/bbaa257>
- Raina K.S., 2023, Role of bioinformatics in analysing big data using statistical computing and computer science, Interantional Journal of Scientific Research in Engineering and Management, 2023.  
<https://doi.org/10.55041/ijserem18829>
- Ramharack P., and Soliman M., 2018, Bioinformatics-based tools in drug discovery: the cartography from single gene to integrative biological networks, Drug Discovery Today, 23(9): 1658-1665.  
<https://doi.org/10.1016/j.drudis.2018.05.041>
- Roscoe S., Khatri M., Voshall A., Batra S., Kaur S., and Deogun J., 2022, Formal concept analysis applications in bioinformatics, ACM Computing Surveys, 55(8): 1-40.  
<https://doi.org/10.1145/3554728>
- Shahid U., 2023, Leveraging fine-tuned large language models in bioinformatics: a research perspective, Qeios, 2023.  
<https://doi.org/10.32388/we7umn.2>
- Shoaib M., Singh A., Gulati S., and Kukreti S., 2021, Mapping genomes by using bioinformatics data and tools, Academic Press, 2021: 245-278.  
<https://doi.org/10.1016/B978-0-12-821748-1.00002-6>
- Solanki V., Hashmi N., and Raghuvanshi D., 2020, Computational of bioinformatics, International Journal of Trend in Scientific Research and Development, 36(5): 601-615.
- Tang B.H., Pan Z.X., Yin K., and Khateeb A., 2019, Recent advances of deep learning in bioinformatics and computational biology, Frontiers in Genetics, 10: 214.  
<https://doi.org/10.3389/fgene.2019.00214>
- Xia X.H., 2017, Bioinformatics and drug discovery, Current Topics in Medicinal Chemistry, 17(15): 1709-1726.  
<https://doi.org/10.2174/1568026617666161116143440>

---

#### Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---