

Applications of Bioinformatics in Biology and Medicine

Yuxing Zhang

¹School of Network Security and Information Technology, Yili Normal University, Yining 835000, Xinjiang, China.

²Key Laboratory of Intelligent Computing Research and Application, Yili Normal University, Yining 835000, Xinjiang, China.

How to cite this paper: Yuxing Zhang. (2024). Applications of Bioinformatics in Biology and Medicine. *Advance in Biological Research*, 5(2), 42-48.

DOI: 10.26855/abr.2024.12.002

Received: December 12, 2024

Accepted: January 10, 2025

Published: February 7, 2025

Corresponding author: Yuxing Zhang, School of Network Security and Information Technology, Yili Normal University, Yining 835000, Xinjiang, China; Key Laboratory of Intelligent Computing Research and Application, Yili Normal University, Yining 835000, Xinjiang, China.

Abstract

With the advent of the post-genome era and the emergence of various high-throughput research methods, the acquisition and analysis of high-throughput data in the field of life science and medical health are increasing, and bioinformatics, which takes biological big data as the research object, has also received increasing attention. In this paper, the concept, development, and application of bioinformatics in the field of biology and medicine, including genomics, proteomics, transcriptomics, and other aspects of research, as well as the important contributions to disease diagnosis, drug development, and personalized medicine are described in detail. At the same time, the challenges faced by bioinformatics are analyzed, and the future development trend of bioinformatics is forecasted. The aim is to provide a valuable reference for researchers and students to understand the relevant knowledge in the field of bioinformatics, and for researchers, policy makers, and biomedical industry personnel to quickly grasp the general situation and development trend of the field of bioinformatics.

Keywords

Bioinformatics; Biology; Medicine; Applied research

1. Introduction

With the rapid development of life science and information technology, bioinformatics comes into being. Bioinformatics integrates the knowledge and technology of biology, computer science, mathematics, and statistics, aiming to collect, store, analyze, and interpret biological data to reveal the nature and law of life phenomena. In the field of biology and medicine, the application of bioinformatics provides researchers with powerful tools and methods to promote the progress of life science and the development of medicine [1].

2. Concept and development of bioinformatics

2.1 Concept

Bioinformatics is a subject formed by the integration of life science and computer science, information science, mathematics, physics, chemistry, systems science, and other disciplines. Bioinformatics uses computer science, information science, statistics, and physics as research methods. Its core system is the integration and mining of multi-modality data of biological systems. Its research contents include designing new methods and new algorithms to reveal the connections between biological big data, developing databases and software tools for storing and analyzing biological big data, as well as analyzing and interpreting biological big data. The ultimate goal of bioinformatics is to discover new biological knowledge, and to decipher the regulatory mechanisms and laws of important life processes and life homeostasis maintenance as well as disease development [2].

2.2 Development

In the 1950s, with the discovery of the double helix structure of DNA, life science entered the era of molecular biology. In the 1970s, computer technology began to be applied to the processing and analysis of biological data. In the 1980s, the concept of "bioinformatics" was first proposed. In the 1990s, bioinformatics developed rapidly with the implementation of the Human Genome Project. Since the 21st century, with the development and widespread application of genome sequencing and various omics research methods, bioinformatics has developed into a very important interdisciplinary subject in the field of modern life science research and has become the driving force for many new research directions and scientific discoveries. Nowadays, bioinformatics has become an indispensable and important discipline in the field of life science and medicine [3].

3. Application of bioinformatics in biology

3.1 Genomics research

(1) Genome sequence analysis

Bioinformatics plays a key role in genome sequence analysis. By comparing, splicing, and annotating genome sequences, the location, structure, and function of genes can be determined. For example, sequence alignment algorithms can be used to identify homologous genes between different species, thereby revealing the relationship between species evolution. At the same time, bioinformatics can also predict gene coding regions, promoters, enhancers, and other regulatory elements, providing important clues for further study of gene expression regulation mechanisms [4].

(2) Genome evolution research

Bioinformatics can study the evolution of the genome by comparing the genome sequences of different species. For example, by analyzing repeated sequences in the genome, the expansion and contraction of gene families, and other phenomena, it is possible to understand the adaptive changes in species during evolution. In addition, bioinformatics can also use phylogenetic analysis methods to construct species evolutionary trees and reveal the relationships between species [5].

(3) Functional genomics research

Functional genomics aims to study the function of genes and their role in life. Bioinformatics can predict the function of genes by analyzing information such as gene expression data and protein interaction networks. For example, gene expression profiling can be used to find genes associated with specific biological processes or diseases, providing clues for further study of the function of genes. At the same time, bioinformatics can also use methods such as protein structure prediction and molecular docking to study the function of proteins and their interactions with other molecules [6].

3.2 Proteomic research

(1) Protein structure prediction

The structure of a protein determines its function, so the prediction of protein structure is one of the important contents of proteomics. Bioinformatics can use homology modeling, folding recognition, and prediction to predict the three-dimensional structure of proteins. Based on known protein structure information, these methods predict the structure of unknown proteins by comparison and simulation. Protein structure prediction is of great significance for understanding protein function, designing new drugs, and developing biotechnology [7].

(2) Protein function prediction

Bioinformatics can predict the function of a protein by analyzing information such as its sequence, structure, and expression pattern. For example, sequence similarity search can be used to find unknown proteins that are similar to known functional proteins and thus to infer their function. At the same time, bioinformatics can also use information such as protein interaction networks and gene expression profiles to further verify and improve protein function prediction [8].

(3) Proteomic data analysis

Proteomics research has produced a large number of experimental data, such as mass spectrometry data, two-dimensional gel electrophoresis data, etc. Bioinformatics can use data analysis software and algorithms to process and analyze these data. For example, mass spectrometry data can be used to identify the type and quantity of proteins, and to analyze the expression level and modification state of proteins. At the same time, bioinformatics can also use cluster analysis, principal

component analysis, and other methods to reduce and visualize proteomic data in order to better understand the characteristics and changes of the proteome [9].

3.3 Transcriptomic research

(1) Gene expression analysis

Transcriptomics research the expression levels and changes of genes under different conditions. Bioinformatics can use the expression data generated by gene chips, transcriptome sequencing, and other technologies to analyze gene expression. For example, differentially expressed genes associated with specific biological processes or diseases can be found through differential expression analysis. At the same time, bioinformatics can also use gene expression network analysis to study the regulatory relationships and synergies between genes [10].

(2) Transcriptional regulation

Transcriptional regulation is an important part of gene expression regulation. Bioinformatics can study the transcriptional regulation mechanism by analyzing the information on transcription factor binding sites and promoter regions. For example, transcription factor binding site prediction algorithms can be used to predict the binding sites of transcription factors to DNA, thereby revealing transcriptional regulatory networks. At the same time, bioinformatics can also use the data generated by chromatin immunoprecipitation sequencing and other technologies to further verify the binding sites and regulatory roles of transcription factors [11].

(3) Non-coding RNA research

Non-coding RNA plays an important role in gene expression regulation. Bioinformatics can use non-coding RNA data generated by transcriptome sequencing and other technologies to identify and predict the function of non-coding RNA. For example, by analyzing the sequence characteristics and expression patterns of non-coding RNA, it is possible to predict their role in gene expression regulation, cell differentiation, and disease occurrence. At the same time, bioinformatics can also use bioinformatics tools to study the interaction of non-coding RNA with proteins, DNA, and other molecules [12].

4. Application of bioinformatics in medicine

4.1 Disease diagnosis

(1) Biomarker discovery

Bioinformatics can discover biomarkers associated with disease by analyzing data from biological samples of disease patients and healthy controls. For example, gene expression profiling can be used to identify differentially expressed genes associated with specific diseases, which can be used as biomarkers for disease diagnosis. At the same time, bioinformatics can also use proteomics, metabolomics, and other technologies to find proteins, metabolites, and other biomarkers related to diseases [13].

(2) Disease risk assessment

Bioinformatics can use data generated by genomics, transcriptomics, and other technologies to assess disease risk. For example, by analyzing an individual's genome sequence, it is possible to predict their risk of developing certain genetic diseases. At the same time, bioinformatics can also use gene expression profiles, proteomics, and other data to assess an individual's risk of developing certain complex diseases [14].

(3) Disease classification and prognosis

Bioinformatics can make disease classification and prognosis judgments by analyzing the biological sample data of disease patients. For example, gene expression profiling can be used to classify disease patients into different subtypes, which may have different clinical characteristics and treatment responses. At the same time, bioinformatics can also use biomarkers and other information to predict the prognosis of disease patients and provide references for clinical treatment [15].

4.2 Drug research

(1) Drug target discovery

Bioinformatics can identify potential drug targets by analyzing disease-related biological data. For example, genomics, proteomics, and other technologies can be used to discover genes and proteins associated with diseases, which can be used as targets for drug development. At the same time, bioinformatics can also use biological network analysis methods to study disease-related signaling pathways and networks and find key node proteins as drug targets [16].

(2) Drug design

Bioinformatics can make use of protein structure prediction, molecular docking, and other methods for drug design. For example, by predicting the binding patterns of drug molecules to target proteins, drug molecules with high affinity and specificity can be designed. At the same time, bioinformatics can also use information such as pharmacokinetics and pharmacodynamics to optimize drug design and development [17].

(3) Drug screening

Bioinformatics can use the data generated by high-throughput screening techniques for drug screening. For example, drug activity data generated by cell experiments and animal experiments can be used to screen out drug molecules with potential therapeutic effects. At the same time, bioinformatics can also use virtual screening and other methods to screen drugs on the computer to improve the efficiency and success rate of drug research and development [18].

4.3 Personalized medicine

(1) Genomic medicine

Genomic medicine is an important part of personalized medicine. Bioinformatics can provide clinicians with personalized medical recommendations by analyzing an individual's genome sequence. For example, by detecting an individual's drug metabolism genes, disease susceptibility genes, and other information, it can be used to develop personalized medication protocols and disease prevention strategies for doctors. At the same time, bioinformatics can also use genomic data for disease risk assessment and early diagnosis to support personalized medicine [19].

(2) Transcriptome medicine

Transcriptome medicine can understand an individual's physiological state and disease risk by analyzing their gene expression profile. Bioinformatics can use transcriptome data generated by technologies such as transcriptome sequencing for transcriptome medical research. For example, by analyzing transcriptome data from patients with a disease, differentially expressed genes associated with the disease can be identified, providing targets for personalized treatment. At the same time, bioinformatics can also use transcriptome data to monitor the therapeutic effect and prognosis of diseases, and provide references for clinical treatment [20].

(3) Proteome medicine

Proteomic medicine can understand an individual's physiological state and disease risk by analyzing his or her proteome. Bioinformatics can use proteomic data generated by proteomic technology to conduct proteomic medical research. For example, by analyzing proteomic data from disease patients, protein markers associated with the disease can be found to provide a basis for personalized diagnosis and treatment. At the same time, bioinformatics can also use proteomic data to study the mechanism of action and adverse reactions of drugs and provide references for personalized medicine [21].

5. Challenges in bioinformatics

5.1 Data quality and standardization issues

Bioinformatics research relies on large amounts of biological data, which are of varying quality and standardization. Data quality issues include accuracy, completeness, and reliability of data, while data standardization issues involve data format, naming conventions, and annotations. These problems bring great challenges to the analysis and application of bioinformatics [22].

5.2 Data analysis and algorithm problems

Bioinformatics data analysis requires the use of complex algorithms and software tools, however, the performance and accuracy of these algorithms and tools need to be improved. For example, in genome sequence analysis, the accuracy and efficiency of sequence alignment algorithms remain a challenge; In protein structure prediction, the accuracy and reliability of the prediction algorithm also need to be further improved. In addition, bioinformatics data analysis is also faced with high dimensions, small samples, and other problems, need to develop more effective data analysis methods [23].

5.3 Data security and privacy issues

Bioinformatics research involves a large amount of personal biological data, and the security and privacy of these data are increasingly concerned. How to protect the security and privacy of personal biodata and prevent data leakage and abuse is an important challenge facing bioinformatics [24].

5.4 Issues of interdisciplinary cooperation

Bioinformatics is an interdisciplinary discipline that requires the cooperation of multiple disciplines such as biology, computer science, mathematics, and statistics. However, there are still some obstacles in the communication and cooperation between different disciplines, and it is necessary to strengthen the mechanism and platform construction of interdisciplinary cooperation to promote the exchange and cooperation between different disciplines [25].

6. Future trends in bioinformatics

6.1 Application of big data and artificial intelligence

With the continuous accumulation of biological data, big data, and artificial intelligence technology will be more and more widely used in bioinformatics. Big data technology can help bioinformaticians process and analyze large-scale biological data and improve the efficiency and accuracy of data analysis. Artificial intelligence technology can use methods such as machine learning and deep learning to automatically analyze and predict biological data, providing new ideas and methods for biological and medical research [26].

6.2 Multi-omics data integration analysis

A biological system is a complex network, a single omics data often cannot fully reflect the nature and law of life phenomena. Therefore, multi-omics data integration analysis will become an important development trend in bioinformatics. By integrating multiple omics data such as genomics, proteomics, transcriptomics, metabolomics, etc., we can gain a more comprehensive understanding of the structure and function of biological systems, and provide more powerful support for disease diagnosis, drug development, and personalized medicine [27].

6.3 Development of precision medicine

Precision medicine is a personalized medical model based on individual genomes, transcriptomes, proteomes, and other omics information. Bioinformatics will play a key role in the development of precision medicine by analyzing multi-omics data from individuals to provide clinicians with personalized medical recommendations and treatment options. In the future, precision medicine will become an important direction of medical development, and bioinformatics will also usher in a broader space for development [28].

6.4 Strengthening of interdisciplinary cooperation

The development of bioinformatics requires interdisciplinary cooperation, and in the future, cooperation between multiple disciplines such as biology, computer science, mathematics, and statistics will be strengthened. At the same time, it will also strengthen cooperation with clinical medicine, pharmacy, and other fields to promote the application of bioinformatics technology in clinical practice. In addition, international cooperation will be strengthened to jointly promote the development of bioinformatics [29].

7. Conclusion

As a new interdisciplinary subject, bioinformatics has been applied more and more widely in the fields of biology and medicine. It provides powerful tools and methods for the research of life science and promotes the progress of life science and the development of medicine. However, bioinformatics also faces challenges in data quality, data analysis, data security, and interdisciplinary collaboration. In the future, with the development of big data, artificial intelligence, and other technologies, bioinformatics will usher in a broader space for development. With the promotion of bioinformatics, life science and medicine will achieve more brilliant achievements [30].

Funding

This research is supported by the Scientific Research Project of Yili Normal University (NO. 2021YSYB091); Science and Technology Plan Project of Ili Kazakh Autonomous Prefecture (NO. YZ2022Y014); Open Project of Yili Normal University Belt and Road Development Research Institute (NO. YDYL2022YB028); Open Project of Yili Normal University Frontier Chinese Literature and History Research Center (NO. BJWSY202312).

References

- [1] Akalin PK. Introduction to bioinformatics. *Molecular Nutrition & Food Research*. 2006;50(7):610-619.
- [2] Azad RK, Shulaev V. Metabolomics technology and bioinformatics for precision medicine. *Briefings in Bioinformatics*. 2019;20(6):1957-1971.
- [3] Gauthier J, Vincent AT, Charette SJ, et al. A brief history of bioinformatics. *Briefings in Bioinformatics*. 2019;20(6):1981-1996.
- [4] Simpson RC, Shanahan ER, Scolyer RA, et al. Towards modulating the gut microbiota to enhance the efficacy of immune-checkpoint inhibitors. *Nature Reviews Clinical Oncology*. 2023;20:697-715.
- [5] Zhao Y, Cheng MY, Zou L, et al. Hidden link in gut-joint axis: gut microbes promote rheumatoid arthritis at early stage by enhancing ascorbate degradation. *Gut*. 2022;71(5):1041-1043.
- [6] Deng L, Gao BB, Zhao L, et al. Diurnal RNAPII-tethered chromatin interactions are associated with rhythmic gene expression in rice. *Genome Biology*. 2022;23:7.
- [7] Almeida A, Mitchell AL, Boland M, et al. A new genomic blueprint of the human gut microbiota. *Nature*. 2019;568:499-504.
- [8] Chen XW, Hao YJ, Cui Y, et al. LncVar: deciphering genetic variations associated with long noncoding genes. *Methods in Molecular Biology*. 2019;1870:189-198.
- [9] Jiang Y, Sun AH, Zhao Y, et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*. 2019;567:257-261.
- [10] Zhong SJ, Zhang S, Fan XY, et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*. 2018;555:524-528.
- [11] Zhu GT, Wang SC, Huang ZJ, et al. Rewiring of the fruit metabolome in tomato breeding. *Cell*. 2018;172:249-261.
- [12] Zhang ZY, Wang M, Xie DF, et al. METTL3-mediated N6-methyladenosine mRNA modification enhances long-term memory consolidation. *Cell Research*. 2018;28:1050-1061.
- [13] Gates AJ, Gysi DM, Kellis M, et al. A wealth of discovery built on the Human Genome Project-by the numbers. *Nature*. 2021;590:212-215.
- [14] Martinez-Jimenez F, Muinos F, Sentis I, et al. A compendium of mutational cancer driver Genes. *Nature Reviews Cancer*. 2020;20(10):555-572.
- [15] Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nature Medicine*. 2019;25:24-29.
- [16] Zhou Y, Zhang YT, Lian XC, et al. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Research*. 2022;50:1398-1407.
- [17] Su X, Wang HX, Zhao N, et al. Trends in innovative drug development in China. *Nature Reviews Drug Discovery*. 2022;21(10):709-710.
- [18] Zeng XX, Zhu SY, Lu WQ, et al. Target identification among known drugs by deep learning from heterogeneous networks. *Chemical Science*. 2020;11:1775-1797.
- [19] Shang J, Wan YS, Luo CM, et al. Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*. 2020;117(21):11727-11734.
- [20] Zhou C, Chen ZK, Zhang L, et al. SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Research*. 2019;47:388-394.
- [21] Qiu TY, Yang YY, Qiu JX, et al. CE-BLAST makes it possible to compute antigenic similarity for newly emerging pathogens. *Nature Communications*. 2018;9:1772.
- [22] CNCB-NGDC Members & Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformatics in 2022. *Nucleic Acids Research*. 2022;50:27-38.
- [23] Koppad S, Basava A, Gkoutos GV, et al. Cloud computing enabled big multi-omics data analytics. *Bioinformatics Biology Insights*. 2021;15:1-16.
- [24] Chen F, Zhang ZQ, Ding CJ, et al. Analysis of global biosafety strategy and recommendations to China. *Bulletin of Chinese Academy of Sciences*. 2020;35(2):204-211.
- [25] Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*. 2022;40:1458-1466.
- [26] Chowdhury R, Bouatta N, Biswas S, et al. Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotechnology*. 2022;40:1617-1623.
- [27] Vamathevan J, Apweiler R, Birney E. Biomolecular data resources: bioinformatics infrastructure for biomedical data science. *Annual*

Review of Biomedical Data Science. 2019;2:199-222.

- [28] Kundu S, Ashinsky BG, Bouhrara M, et al. Enabling early detection of osteoarthritis from presymptomatic cartilage texture maps via transport-based learning. *Proceedings of the National Academy of Sciences of the National Academy of Sciences*. 2020;117:24709-24719.
- [29] Liu XF, Lu Z. Progress of bioinformatics studies for multi-omics and multi-modal data in complex diseases. *Chinese Science Bulletin*. 2024;69(30):4432-4446.
- [30] Sayers EW, Bolton EE, Brister JR, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. 2022;50:20-26.