Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v31i6.6723

INFORMATION TECHNOLOGIES DRIVING INNOVATION IN BIOLOGICAL SCIENCE: A FOCUS ON BIOINFORMATICS APPLICATION DEVELOPMENT

Zeshan Anwar[1], Abdur Rahman[2], Bonface Obare[3], Mohanapriya N[4], Tariq Rafique[5*], Hrishitva Patel[6], Rabia Tehseen[7]

[1]M. Phil Biotechnology, University of Agriculture Faisalabad, Pakistan
[2]Scholar Accounting and Technology, School of Business, Emporia State University, USA
[3]Research Associate, Department of Biotechnology, University of Pretoria, South Africa
[4]Lecturer, Department of Computer Science and ICT, Cavendish University Lusaka, Zambia
[5*]Assistant Professor, Dadabhoy Institute of Higher Education, Karachi, Pakistan
[6]PhD (Pursuing), Information Systems, University of Texas at San Antonio, United States of America
[7]Assistant professor, Department of Computer Science, University of Central Punjab, Lahore, Pakistan

***Corresponding Author:** Tariq Rafique
*Email: dr.tariq1106@gmail.com

ABSTRACT:
Background: DNA sequences contain vast amounts of information, necessitating advanced computing methods and data modeling techniques for analysis. The University of Technology's Systems Engineering and Computer Science program's (Iraqi research group/university) research group aims to leverage information technologies to drive significant progress in biological science.
Objective: This study explores computational methods conducive to developing bioinformatics applications to expedite computation, enhance inference times, and bolster the reliability of analyses derived from DNA sequence data.
Methods: The research scrutinizes various computational methods suitable for bioinformatics applications within the framework of the University of Technology's Systems Engineering and Computer Science program's (IRAQI RESEARCH GROUP/UNIVERSITY) research group.
Results: Identified computational methods exhibit promise in accelerating analysis processes and improving the reliability of results derived from DNA sequence data. These methods serve as foundational tools for advancing scientific inquiry in biological science.
Conclusion: By employing sophisticated computational methods, such as those investigated within the University of Technology's Systems Engineering and Computer Science program's IRAQI RESEARCH GROUP/UNIVERSITY research group, bioinformatics applications can achieve significant advancements, facilitating progress in biological science.

Keywords: sequences, proteins, DNA, RNA, bioinformatics, biology, biological data, and technologies.

INTRODUCTION:

The scientific community conducting biological research is confronted with ever-increasing challenges daily. These challenges include managing massive volumes of data that are expanding exponentially in size and complexity due to technological advancements that enable more accurate calculations. The community seeks answers to studies on DNA sequences and molecular structure. Fortunately, significant advancements in intelligent data processing and analysis techniques have been made possible by technological growth in electronics, software development, and telecommunications. These developments have benefited scientific investigations that aim to fully comprehend the data structures of live beings (Maljkovic Berry et al., 2020).

Large-scale data management is difficult and thus calls for computational processes that are highly performant in terms of both response times and space. To create useful tools for future research, this article examines some of the most popular computational methods and approaches for analyzing DNA sequences. Section 2 of this page discusses bioinformatics's concepts, applications, and scope. The alignment of DNA sequences is discussed in Section 3. The computational tools for creating and implementing bioinformatics solutions are covered in Section 4 and include Matlab usage, biological databases, data warehouses, data mining, and learning machines. Section 5 outlines the conclusions and next steps (Baxevanis et al., 2020).

BIOINFORMATICS

Understanding connections, structures, and patterns in biological data is one of bioinformatics' most crucial responsibilities. In recent years, several fields, including computer science, mathematics, statistics, chemistry, and non-traditional biological sciences, have been drawn to bioinformatics. This is a result of the abundance of both public and private biological data available and the urgent need to convert the data into valuable biological knowledge and information. Using sophisticated computational approaches in conjunction with these fields, initiatives including biological control, genomic analysis, and drug discovery and development are created. This entails managing and analyzing massive amounts of biological data on DNA, RNA, protein patterns, protein structures, genetic expression profiles, and protein interactions using computer technology and statistical techniques (Tyagi et al., 2022).

Bioinformatics's Range.

The subfields of bioinformatics consist of two complementary areas: The creation of databases and I.T. tools and the use of these to produce biological knowledge to comprehend living systems more fully. Software for recording sequences, structural and functional analysis of the sequences, and the creation and preservation of biological databases are all examples of the tools being developed. Examining biological data frequently raises fresh issues and difficulties, propelling the advancement of more advanced computational instruments (Wood et al., 2020).

In what ways is bioinformatics applicable?

In addition to being a crucial science for fundamental studies in molecular biology and genomics, bioinformatics is also significantly influencing many fields in biotechnology and the biomedical sciences. Its applications necessitate expertise in agricultural biotechnology, forensic DNA analysis, and drug formulation. Compared to traditional trial and error, a bioinformatics-based method greatly decreases the time and expense required to produce medications with greater potency, fewer side effects, and reduced toxicity. Molecular phylogenetic analysis results have been admitted as evidence in criminal courts in forensic medicine. Forensic identity analysis has used certain complex Bayesian statistics and plausibility-based DNA analysis techniques (Liu et al., 2020).

Genomics and bioinformatics are poised to transform healthcare systems by advancing personalized therapy. The combination of high-speed genomic sequencing and advanced information technology enables a doctor at a clinic to quickly sequence a patient's DNA. This allows them to discover potential detrimental mutations in the genome, enabling them to play a crucial role in early and successful diagnosis (Chen et al., 2023).

Medical intervention for illnesses.

Agriculture also utilizes bioinformatics techniques. Using plant genome databases and gene expression analysis has significantly contributed to advancing novel crop varieties characterized by enhanced yield and increased disease resistance. Bioinformatics encompasses the creation of databases or knowledge bases to store and retrieve biological data, algorithms for analyzing and determining these data's linkages, and statistical tools for identifying and interpreting data sets (Vaisvila et al., 2021).

Analysis of DNA Sequences

DNA sequence analysis involves identifying and examining functional and structural disparities among biological sequences. This can be achieved by comparing novel (unidentified) sequences with extensively researched and annotated (established) sequences. This approach encompasses aligning sequences, examining sequence databases, identifying patterns, recreating evolutionary relationships, and comparing genome development. Scientists have determined that two analogous sequences possess identical functional roles. The comparison can be conducted by considering either the biochemical behaviour or the protein structure (Jovic et al., 2022).

Homologous sequences are referred to as such when two sequences originating from distinct organisms exhibit similarity. The process of aligning DNA sequences. The comparison of DNA sequences serves as the fundamental basis for bioinformatic analysis. This is a crucial initial stage in analyzing the newly identified sequences regarding their structure and function. Sequence alignment is the essential technique involved in this type of comparison. The approach described involves the comparison of sequences through the identification of shared character patterns and the establishment of corresponding residues across related sequences. The process of aligning sequence pairs is crucial to identify commonalities within a database and perform alignment on numerous sequences (van der Loos & Nijland, 2021).

Sequence homology is a crucial notion in sequence analysis. Homologous relationship or homology refers to the connection between two sequences that share a shared evolutionary origin. Sequence similarity, although frequently used interchangeably and inaccurately, refers to the proportion of aligned residues that exhibit similar physicochemical features, including size, cost, and hydrophobicity (Miller et al., 2020).

Application of Computational Technologies in the Field of Bioinformatics.

Like other scientific fields, biology yields substantial amounts of data that necessitate sophisticated computing methods for real-time processing, depending on the study goals. Numerous approaches are encompassed within I.T. research and development, specifically in data storage and processing. These techniques include relational and semantic databases (D.B.), data warehouses, data mining, and various artificial intelligence methodologies (Urban et al., 2021).

Biological databases.

In contemporary biological databases, three predominant database architectures are commonly employed: flat files, relational databases, and object-oriented databases, despite the evident drawbacks associated with their utilization. The failure to effectively scale real models with the necessary data volume is the underlying cause. Biological databases can be classified into three distinct types based on their content. Primary databases are repositories that house authentic biological data. The entities above refer to raw sequence files or structural data, such as GenBank and Protein Data Bank. Secondary databases are repositories that store information that has been computationally processed or manually curated, derived from primary data sources. Protein sequence databases that have been translated encompass functional annotations that fall within this particular category, such as Swiss-Prot and PIR (Beck et al., 2022).

Specialized databases, like Flybase, cater to a specific area of research. Examples of databases that focus on a certain organism or kind of data are the Ribosomal Database Project and the HIV Sequence Database. The necessity of integrating secondary and specialized databases with main databases is the root cause of many issues that arise in scientific research. Cross-referencing and linking, or being "linked," to relevant entries in other databases with more information is advised for entries in one database. The primary obstacle to establishing connections across distinct biological databases is the lack of compatibility between the three types of database formats outlined above, which restricts their ability to communicate with one another. Using the Common Object Request Broker Architecture (CORBA) specification language, which enables database programmes in various locations to communicate over a network through a "broker interface" without needing to understand each structure independently, is one way to standardize communication between databases in distributed systems (Sigsgaard et al., 2020).

Databases are also linked using a related protocol known as eXtensible Markup Language (XML). Each biological record in this format is broken down into smaller, fundamental parts labelled using hierarchical clustering labels. Sequences from cloning vectors and primary data redundancy brought on by repeatedly acquiring identical or matched sequences can contaminate gene sequences. The National Centre for Biotechnology Information (NCBI) developed the nonredundant database RefSeq, which merges related sequence fragments and identical sequences from the same organism into a single record to lessen this redundancy. Developing sequence cluster databases, such as UniGene, that connect expressed tag sequences (ESTs) derived from the same gene is an additional strategy to tackle redundancy. Due to many entries, the genetic sequence is frequently discovered under different names. Reannotation of genes and proteins using a common vocabulary to describe them is required to mitigate this issue. All genes and proteins have a uniform and clear naming scheme thanks to Gene Ontology (Qian et al., 2020).

The Data Warehouse.

A collection of integrated, subject-oriented, time-varying, non-transient data that aids in managerial decision-making is called a data warehouse (D.W.). Examining bioinformatics projects reveals that the field's needs include storing vast amounts of data with various dimensions for long periods of time, in various formats, and from various sources. Based on a data warehouse known as BioStar schema, the author discusses their suggested multidimensional modelling for biomedical data that can capture the complex semantics of biomedical data and offer higher extensibility and flexibility for fast-expanding biological research approaches (Novák et al., 2020).

To maintain many-to-many links between the central entity and the dimensions, it is necessary to store the various measures in distinct n-tables, which can be made to support particular attributes of a measure. Ligand Depot is a comprehensive knowledge resource on nucleic acids, proteins, and tiny compounds. Its main goal is to give tiny molecules chemical and structural details. In addition, it supports keyword-based searches, offers a graphical search interface for chemical substructures, and grants access to many online resources. In subsequent work, we suggest including a more advanced graphical user interface and enhancing search capabilities (Chen et al., 2023; Thareja & Chhillar, 2022).

***Bioinformatics Data Mining*.**

The focus of data mining is the study of methods for obtaining useful information from vast amounts of biological data. Effective software tools are required for data retrieval, biological sequence comparison, pattern recognition, and knowledge discovery visualization to accomplish this. We can highlight a few of the most popular data mining methods in bioinformatics:

KDD, which is the entire process of extracting non-trivial, previously undiscovered, and possibly helpful knowledge from a dataset (Shen et al., 2022);

Textual mining, also known as KDT, is a process that focuses on extracting knowledge from unstructured data contained in textual databases. It involves the identification and exploration of knowledge within texts. Furthermore

Statistics in data mining can be categorized into two groups: Supervised learning, which involves making inferences about future observations based on existing knowledge of sequence groups, and Unsupervised learning, which involves identifying previously unknown groups of similar cases in the data (Phan et al., 2006). Bioinformatics research software tools can be categorized into four distinct classes:

Tools for recovering data.

An instance of an integrated data retrieval system is Entrez, developed by the National Centre for Biotechnology Information (NCBI). This system offers comprehensive access to several data domains, encompassing literature sequences, nucleotides and proteins, whole genomes, 3D architectures, and other relevant information. The present study aims to compare sequencing and alignment methods. One example is BLAST, known for its high speed and ability to search a complete nonredundant database quickly. GenBank and EMBL are prominent tools for managing biological databases and aligning local sequences in pairs. FASTA is a useful tool for efficiently comparing proteins or nucleotides. Employing a substitution matrix effectively enhances the sensitivity of similarity search by executing optimized searches for local alignments (Ge et al., 2018).

ClustalW is a tool that may be utilized for multiple sequence alignment. It can align DNA or protein sequences to reveal their relationships and evolutionary origin. Pattern detection techniques identify and analyze patterns or characteristics within a given dataset. Cluster analysis is a statistical technique employed to identify distinct groups within a provided dataset, wherein objects within the same group exhibit similarities to one another while displaying dissimilarities from objects in other groups. GeneQuiz is a comprehensive, integrated tool that proves to be valuable in the exploration of expression patterns. It operates on a wide scale and employs diverse search and analysis techniques to analyze DNA and protein sequences (Ge et al., 2018).

Tools for seeing.

These tools facilitate interactive and graphical visualization of genetic data. Expression Profiler and GeneQuiz, more extensive analytic programs, provide an integrated visualization tool. In addition, many visualization software packages can be accessed at no cost through online platforms. Examples of software tools that can be utilized include Protein Explorer, which offers a three-dimensional visualization of protein structure through an interactive system, and TreeView, which presents a graphical depiction of clustering outcomes and image navigation through hierarchical trees (Jung et al., 2011).

Machine learning.

A learning machine is a dynamic mechanism that enables computers to acquire knowledge through experience, exemplification, and analogy. The neural network is a machine learning technique effectively used to address various bioinformatics challenges. An instance of applying a neural network system, which relies on brain knowledge, can be observed in the analysis of DNA sequences. The emergence of artificial neural networks can be attributed to the prevalence of two prominent genetic search engines. GRAIL is a pioneering gene search programme that utilizes a neural network integration of predictive coding algorithms to detect genes, exons, and other properties in DNA sequences. It recognizes coding potential inside fixed-length windows without extra features (Shanbehzadeh et al., 2021).

Gene Parser is an additional gene search method specifically developed to ascertain and analyze the intricate composition of protein genes within genomic DNA sequences. GenCANS, an artificial neural system, was created to analyze and handle substantial molecular sequencing data obtained from the Human Genome Project. The genetic algorithm has demonstrated significant efficacy in

addressing several practical challenges across multiple fields, with a special emphasis on bioinformatics. Notably, it has been effectively employed in resolving multiple sequence alignment difficulties. SAGA, a widely recognized method, involves the random production of an initial population of forms, which then undergoes evolutionary changes over several generations, resulting in a progressive enhancement of the population's fitness (Qi et al., 2019).

Soft computing within the field of bioinformatics.

Expert systems, a prevalent soft computing technology, are constructed by aggregating knowledge from specialized human experts. Experts frequently encounter challenges in determining the specific guidelines they employ. The resolution of problems involves the extraction of a comprehensive depiction of the concealed circumstances, encompassing the various aspects and rules that align with the conduct of the human expert. The rapid progress in biotechnology has led to the generation of vast quantities of biological data. Moreover, the data may contain significant linkages and correlations that are not readily apparent. Certain soft computing techniques are specifically intended to process extensive data sets effectively and can also be employed to extract such correlations. Fuzzy systems are crucial in bioinformatics for constructing knowledge-based systems (Ge et al., 2018).

Section	**Summary**
Introduction	The scientific community faces mounting challenges in biological research due to the exponential growth in data complexity. Technological advancements enable more accurate calculations, benefiting investigations into DNA sequences and molecular structures [MaljkovicBerry2020].
Bioinformatics	Bioinformatics integrates various fields to understand biological data, driving initiatives like genomic analysis and drug development. It encompasses database creation, sequence analysis, and knowledge extraction from vast biological datasets [Tyagi2022].
Analysis of DNA Sequences	DNA sequence analysis involves identifying functional and structural differences between biological sequences. It includes aligning sequences, identifying patterns, and reconstructing evolutionary relationships, crucial for understanding their structure and function [Jovic2022].
Application of Computational Technologies	Biology generates vast amounts of data requiring sophisticated computing methods. IT approaches include relational databases, data mining, and artificial intelligence methodologies, facilitating data storage, retrieval, and analysis in bioinformatics [Urban2021].
Biological Databases	Various database architectures are used in biology, storing primary, secondary, and specialized biological data. Challenges include scalability and integration across databases, addressed through standardization protocols like CORBA and XML [Beck2022].
Data Warehouse	Data warehouses store and manage diverse biological data types, offering multidimensional modeling to support complex biomedical research needs. Examples include Ligand Depot for chemical data and BioStar schema for biomedical data integration [Novák2020].
Bioinformatics Data Mining	Data mining extracts valuable insights from biological datasets, employing techniques like KDD and textual mining. Supervised and unsupervised learning methods aid in pattern recognition and knowledge discovery, crucial for advancing bioinformatics research [Shen2022].
Tools for Biological Data Recovery	Software tools like BLAST and ClustalW facilitate data retrieval and sequence alignment, essential for understanding genetic relationships. Visualization tools like Protein Explorer offer interactive genetic data representations [Ge2018].
Machine Learning	Machine learning techniques, including neural networks and genetic algorithms, are applied in bioinformatics for gene prediction and sequence alignment. These methods offer efficient solutions to various bioinformatics challenges [Shanbehzadeh2021].
Soft Computing	Soft computing methods, such as fuzzy logic and expert systems, extract knowledge from complex biological data. Fuzzy logic enhances protein motif flexibility, gene expression

Section	Summary
	analysis, and genetic network deciphering, advancing bioinformatics research [Yue2019].

Fuzzy logic approaches can be effectively employed in various biomedical science and bioinformatics domains. Several significant applications of fuzzy logic include enhancing the flexibility of protein motifs, investigating variations among polynucleotides, analyzing experimental expression data through the application of the fuzzy theory of adaptive resonance, aligning sequences using a fuzzy recast of a dynamic programming algorithm, employing fuzzy systems for genetic sequencing of DNA, analyzing gene expression data, examining gene relationships, deciphering genetic networks, and classifying amino acid sequences into distinct superfamilies (Yue et al., 2019).

MATLAB is utilized in the field of Bioinformatics.

MATLAB is an abbreviated form of the acronym "MATrix Laboratory". The software is a comprehensive application processing and development environment designed to facilitate the execution of projects that require complex mathematical computations and their corresponding graphical representation. Additionally, the organization presently offers a diverse array of specialized support programmes called Toolbox. Bioinformatics offers a flexible and adaptable platform for molecular biologists and other scientific researchers to investigate concepts, develop prototype algorithms, and establish applications in medicinal research, genetic engineering, and genomic and proteomic endeavours. The Toolbox allows users to access various genomic and proteomic data formats, analysis methodologies, specialized visualizations specifically designed for genomic and proteomic sequences, and microarray research (Li et al., 2021; Maljkovic Berry et al., 2020).

CONCLUSION:

Data storage plays a crucial role in bioinformatics by enabling the exploration of biological knowledge and facilitating information interchange for research purposes.
The advent of collaborative biological web apps is widely regarded as a transformative force in biological research. The assistance of an I.T. professional is crucial for scientists to effectively utilize and fully leverage the technological features of the B.D. The current body of research predominantly focuses on the life and health sciences, thereby necessitating the adoption of a research methodology rooted in computer science within this domain. It is important to establish clear definitions for the technology platforms and implementation procedures linked to the suggested solution to advance the field. During this phase, researchers see that their interests and needs are being acknowledged, which is a crucial prerequisite for the effectiveness of the proposed system.

REFERENCES:

1. Baxevanis, A. D., Bader, G. D., & Wishart, D. S. (2020). *Bioinformatics*. John Wiley & Sons.
2. Beck, D., Ben Maamar, M., & Skinner, M. K. (2022). Genome-wide CpG density and DNA methylation analysis method (MeDIP, RRBS, and WGBS) comparisons. *Epigenetics, 17*(5), 518-530.
3. Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., Liu, Y., Feng, J., Chen, H., & He, Y. (2023). TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Molecular Plant*, *16*(11), 1733-1742.
4. Ge, H., Yan, Y., Wu, D., Huang, Y., & Tian, F. (2018). Potential role of LINC00996 in colorectal cancer: a study based on data mining and bioinformatics. *OncoTargets and therapy*, 4845-4855.

5. Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., & Luo, Y. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine*, *12*(3), e694.
6. Jung, Y., Lee, S., Choi, H.-S., Kim, S.-N., Lee, E., Shin, Y., Seo, J., Kim, B., Jung, Y., & Kim, W. K. (2011). Clinical validation of colorectal cancer biomarkers identified from bioinformatics analysis of public expression data. *Clinical Cancer Research*, *17*(4), 700-709.
7. Li, Y., Ma, L., Wu, D., & Chen, G. (2021). Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. *Briefings in Bioinformatics*, *22*(5), bbab024.
8. Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological entomology*, *45*(3), 373-385.
9. Maljkovic Berry, I., Melendrez, M. C., Bishop-Lilly, K. A., Rutvisuttinunt, W., Pollett, S., Talundzic, E., Morton, L., & Jarman, R. G. (2020). Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *The Journal of infectious diseases*, *221*(Supplement_3), S292-S307.
10. Miller, S., Chiu, C., Rodino, K. G., & Miller, M. B. (2020). Point-counterpoint: should we be performing metagenomic next-generation sequencing for infectious disease diagnosis in the clinical laboratory? *Journal of clinical microbiology*, *58*(3), 10.1128/jcm. 01739-01719.
11. Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, *15*(11), 3745-3776.
12. Phan, J. H., Quo, C.-F., & Wang, M. D. (2006). Functional genomics and proteomics in the clinical neurosciences: data mining and bioinformatics. *Progress in Brain Research*, *158*, 83-108.
13. Qi, Y., Qi, H., Liu, Z., He, P., & Li, B. (2019). Bioinformatics analysis of key genes and pathways in colorectal cancer. *Journal of Computational Biology*, *26*(4), 364-375.
14. Qian, X.-B., Chen, T., Xu, Y.-P., Chen, L., Sun, F.-X., Lu, M.-P., & Liu, Y.-X. (2020). A guide to human microbiome research: study design, sample collection, and bioinformatics analysis. *Chinese Medical Journal*, *133*(15), 1844-1855.
15. Shanbehzadeh, M., Nopour, R., & Kazemi-Arpanahi, H. (2021). Comparison of four data mining algorithms for predicting colorectal cancer risk. *Journal of Advances in Medical and Biomedical Research*, *29*(133), 100-108.
16. Shen, W., Song, Z., Zhong, X., Huang, M., Shen, D., Gao, P., Qian, X., Wang, M., He, X., & Wang, T. (2022). Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. *Imeta*, *1*(3), e36.
17. Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Møller, P. R., Hansen, M. M., & Thomsen, P. F. (2020). Population-level inferences from environmental DNA—Current status and future perspectives. *Evolutionary Applications*, *13*(2), 245-262.
18. Thareja, P., & Chhillar, R. S. (2022). A detailed survey on data mining based optimization schemes for bioinformatics applications. *ECS Transactions*, *107*(1), 4689.
19. Tyagi, A., Sharma, A., & Bhardwaj, M. (2022). Future of bioinformatics in India: A survey. *International Journal of Health Sciences*(II), 431187.
20. Urban, L., Holzer, A., Baronas, J. J., Hall, M. B., Braeuninger-Weimer, P., Scherm, M. J., Kunz, D. J., Perera, S. N., Martin-Herranz, D. E., & Tipper, E. T. (2021). Freshwater monitoring by nanopore sequencing. *Elife*, *10*, e61504.
21. Vaisvila, R., Ponnaluri, V. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., & Marks, K. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome research*, *31*(7), 1280-1289.
22. van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*, *30*(13), 3270-3288.

23. Wood, A., Najarian, K., & Kahrobaei, D. (2020). Homomorphic encryption for machine learning in medicine and bioinformatics. *ACM Computing Surveys (CSUR)*, *53*(4), 1-35.
24. Yue, C., Liang, C., Li, P., Yan, L., Zhang, D., Xu, Y., Wei, Z., & Wu, J. (2019). DUXAP8 a pan-cancer prognostic marker involved in the molecular regulatory mechanism in hepatocellular carcinoma: a comprehensive study based on data mining, bioinformatics, and in vitro validation. *OncoTargets and therapy*, 11637-11650.