# The Role of Bioinformatics in Drug Discovery: Accelerating Target Identification and Therapeutic Development

Aneeqa Munawar[1,*], Maryam Zain[2], Saif ur Rahman[3], Muhammad Farooq[4], Willan Antonio Muñoz-Chamba[5] and Sehrish Sarwar[1]

[1]Department of Biochemistry, Faculty of Sciences, University of Agriculture, Faisalabad, Pakistan
[2]Department of Biochemistry and Biotechnology, Faculty of Life Sciences, The Women University, Multan, Pakistan
[3]Department of Environmental Science and Engineering, School of Energy and Environmental Engineering, University of Science and Technology, Beijing, China
[4]Department of Biochemistry, Faculty of Life Sciences, Government College and University, Faisalabad, Pakistan
[5]Department of Biochemistry, School of Medicine, Faculty of Medical Sciences, University of Guayaquil, Guayas, Ecuador
*Corresponding author: aneeqamunawar152@gmail.com

## Abstract

Bioinformatics has revolutionized drug discovery by accelerating the identification of drug targets, enhancing the screening and optimization of drug candidates, and enabling the prediction of side effects and drug resistance. Traditional drug discovery methods, though successful, often require over 12 years and an average investment of USD 1.8 billion to progress from lead identification to clinical trials. In contrast, in silico approaches have emerged as promising alternatives, significantly reducing the time, cost, and labor associated with drug development. More realistic protein-ligand docking experiments, along with instructive virtual screening, were made possible by the accumulation of protein as well as RNA structures, the creation of homology modeling along with protein structure simulation, and the large structure databases regarding small molecules and metabolites. Hence, computational methods have already facilitated the successful design of novel drug compounds, offering precise and reliable results for informed decision-making. Integrating bioinformatics into research has also advanced our understanding of the human genome, paving the way for discovering innovative drugs and targeted therapies. Furthermore, bioinformatics continues to impact biological research by identifying biomarkers, driver mutations, oncogenic pathways, and therapeutic targets. Future advancements, including the incorporation of multi-omics data, network-based analyses, and machine learning, will be crucial for biomarker discovery and improving patient prognosis. Here, the limitations of data mining software, popular soft wares, and databases for drug discovery, and the conceptual framework driving data collection, highlighting its potential and usefulness, is discussed.

**Keywords:** Bioinformatics, Genes, Alternatives, Drug discovery, Targeted identification, Treatment

**Cite this Article as:** Munawar A, Zain M, Rahman SU, Farooq M, Muñoz-Chamba WA and Sarwar S, 2025. The role of bioinformatics in drug discovery: accelerating target identification and therapeutic development. In: Farooqi SH, Kholik K and Zaman MA (eds), One Health Horizons: Integrating Biodiversity, Biosecurity, and Sustainable Practices. Unique Scientific Publishers, Faisalabad, Pakistan, pp: 271-279. https://doi.org/10.47278/book.HH/2025.315

A Publication of Unique Scientific Publishers

**Chapter No:** 25-037

**Received:** 04-Feb-2025
**Revised:** 09-Apr-2025
**Accepted:** 16-May-2025

## Introduction

Bioinformatics began over fifty years ago with computational techniques for protein sequence analysis in the early 1960s, setting the groundwork for DNA sequencing and desktop computers. DNA analysis emerged due to molecular biology techniques and computer science advancements, simplifying DNA manipulation and sequencing, and developing powerful computers and software for bioinformatics tasks (Gauthier et al., 2019). By facilitating more accurate target identification as well as computational drug design, bioinformatics' development from its early roots of protein sequence analysis towards sophisticated DNA analysis has greatly sped drug discovery, lowering costs and increasing success rates (Clark & Lillard Jr, 2024).

Traditional drug discovery and development are costly and time-consuming, with an average cost of USD 1.8 billion and a high failure rate of 96% (Shaker et al., 2021). Traditional methods like whole plant extracts and single-compound-based medicines face limitations like inefficiency, high costs, and lengthy processes, making discovery challenging despite industrial R&D efforts (Berdigaliyev & Aljofan, 2020). Drug candidates often fail due to issues like insufficient efficacy (40-50%) and suboptimal absorption, distribution, metabolism, excretion, and toxicity profiles (Sun et al., 2022). Advancements in drug discovery have led to the development of therapeutic agents targeting critical proteins, overcoming time-consuming, laborious, high-priced, and chemically disparate in vitro and high-throughput screening methods (Bossé, 2018; Shaker et al., 2021). Molecular docking and bioinformatics have introduced a new era into pharmaceutical experimentation, helping in drug development, and providing essential tools for drug discovery (Okpo et al., 2024). Bioinformatics serves as a fundamental pillar in the process of discovering new drugs, separating the innovative ones, pinpointing omics-based biomarkers, and finding new applications for already existing drugs, as they are considered to be the drugs of the future (Singh et al., 2022). Molecular docking is a very powerful computer technique

that allows researchers to predict, study, and speed up the identification of drug candidates as well as to choose promising lead compounds for further research (Muhammed & Aki-Yalcin, 2024). By lowering the need for costly and time-consuming experimental procedures, molecular docking accelerates the procedures of drug development. This chapter highlights the importance of bioinformatics tools for effective drug delivery at the targeted site for the effective treatment of various ailments.

**Bioinformatics Tools and Core Areas**

Bioinformatics is a multidisciplinary field that enhances research efficiency by developing tools for managing growing biological data, offering potential in genomic analysis and protein studies through software, algorithms, and databases (Ali et al., 2021). It is essential for drug discovery, understanding how diseases work, and studying molecular interactions. The field is separated into various domains. These domains have specialized tools for studying different biological information. These tools are discussed in Table 1 and mentioned in Figure 1.

**Table 1:** Different domains, tools, and their applications in drug discovery

| Domains | Tools | Applications | References |
|---|---|---|---|
| Sequence Analysis | BLAST | Finds homologous sequences in many organisms | (Hung & Weng, 2016) |
| | MEGA | Creates phylogenetic trees to analyze evolutionary connections | (Kumar et al., 2018) |
| | Clustal Omega | For multiple sequence alignments | (Sievers & Higgins, 2020) |
| Genomics and Genome Annotation | Ensembl | Offers genomic information on a range of animals | (Harrison et al., 2024) |
| | UCSC Genome Browser | Provides functional insights by visualizing genomic data | (Lee et al., 2020) |
| | NCBI GenBank | holds DNA sequences | (Sayers et al., 2019) |
| Proteomics and Protein Structure Analysis | UniProt | database of protein sequences and functions | (Zaru et al., 2023) |
| | Protein Data Bank | Organizes 3D protein structures | (Burley et al., 2022) |
| | I-TASSER | Predicts 3D protein models | (Muthiah et al., 2021) |
| | SwissDock | Perform molecular docking to investigate interactions between proteins and ligands | (Patil & Rohane, 2021) |
| Structural Bioinformatics and Molecular Docking | Autodock | Runs simulations of automatic docking. | (Sriramulu & Lee, 2021) |
| | Patchdock | For geometric shape complementarity molecular docking | (Ravi & Paramasivam) |
| | PyMol | 3D molecular structure viewer | (Yuan et al., 2017) |
| Transcriptomics and Gene Expression Analysis | Gene Expression Omnibus | Gene expression database | (Alameer & Chicco, 2022) |
| | STRING | Protein-protein interaction analyzer | (Szklarczyk et al., 2021) |
| | ArrayExpress | makes transcriptome datasets accessible. | (Athar et al., 2019) |
| | DAVID | carries out enrichment analysis and functional annotation. | (Sherman et al., 2022) |
| Metagenomics | QIIME | Examines the makeup of the microbiome. | (Estaki et al., 2020) |
| | MG-RAST | Analyze data from metagenomics sequences. | (Keegan et al., 2016) |
| | Kraken2 | Effectively categorizes microbial sequences. | (Lu et al., 2022) |
| Pharmacogenomics and Drug Discovery | DrugBank | Offer comprehensive information on the drug and its target. | (Knox et al., 2024) |
| | SwissADME | Rorecasts pharmacokinetics and drug likeness | (Bakchi et al., 2022) |
| | ADMETlab | Evaluates a drug candidate's toxicity, distribution, metabolism, excretion, as well as absorption. | (Xiong et al., 2021) |
| | PubChem | Includes information on chemical structures and bioactivity. | (Kim et al., 2023) |
| Computational Systems Biology | KEGG | Maps the signaling and metabolic pathways. | (Kanehisa et al., 2023) |
| | Rectome | Offers databases of pathways for functional genomics. | (Southern et al., 2023) |
| | Cytoscape | Shows networks of molecular interactions. | (Piñero et al., 2021) |
| Evolutionary Bioinformatics and Phylogenetics | MEGA | Keeps phylogenetic trees and related information | (Kumar et al., 2018) |
| | PhyML | Maximizes the likelihood of phylogenetic analysis | (Lefort et al., 2017) |
| | TreeBase | keeps phylogenetic trees as well as related information in storage | (Carbone et al., 2019) |
| Biomedical Informatics and Clinical Data Analysis | ClinVar | Explains genetic variations and their implications for clinical practice | (Landrum et al., 2020) |
| | Online Mendelian Inheritance in Man (OMIM) | Includes details on human genes along with genetic illnesses | (Hamosh et al., 2021) |
| | GNOMAD | Provides genetic data at the population level for variation analysis | (Gudmundsson et al., 2022) |

**Integrating Bioinformatics in Every Phase of Drug Discovery**

**Importance of Target Identification in Drug Discovery**

Target identification is very important in the discovery of novel drugs, which enables researchers to understand complicated drug mechanisms. Over the past two centuries, advances in the identification of specific targets show that the technologies have improved drug selectivity as well as minimized the side effects associated with it (Tabana et al., 2023). A wide range of biomolecules can serve as therapeutic targets, including enzymes, cellular receptors, ion channels, DNA, along various transcription factors (Cui et al., 2022; Moumné et al., 2022; Picci et al., 2022). Bioinformatics, a multidisciplinary field involving molecular phylogenetics, population genetics, proteomics, transcriptomics, and genomics, helps identify effective, affordable, and safe drug targets, prevent drug resistance, and evaluate environmental health impacts (Xia, 2017).


**Fig. 1:** Different domains and software of bioinformatics along with their functions

**Bioinformatics Tools for Target Identification**

Bioinformatics is crucial in drug discovery, aiding in the systematic selection of therapeutic targets. Tools like BLAST identify genetic mutations linked to diseases, predicting drug targets based on genetic evidence and variant effects. The Open Targets platform blends the data from genome-wide association studies and functional genomic studies (McDonagh et al., 2024). In proteomics, the databases like UniProt provides extensive information on protein sequences along with their functions that are essential for understanding protein-target interactions and identifying biomarkers for drug efficacy (Zhang et al., 2022). In transcriptomics, RNA sequencing technologies enable gene expression analysis, helping identify dysregulated genes in disease conditions, such as the tools like STRING build protein-protein interaction networks that reveal the biological pathways involved in diseases, aiding in target prioritization for drug development (Clark & Lillard Jr, 2024). Bioinformatics resources help identify and validate potential drug targets, accelerating drug discovery. Public databases like OMIM, Genetic Association Database (GAD), genome-wide association study (GWAS) Catalog, and DisGeNET provide direct information on oncogenes and tumor suppressor gene sets for various cancer types. The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) contain raw high-throughput data for bioinformatics analysis. Drug-responsive genes are crucial for analysis, and databases like DrugBank, Therapeutic Target Database, Pharmacogenomics Knowledge Base (PharmGKB), and Naturally Occurring Plant-based Anti-cancer Compound-Activity-Target (NPACT) provide direct drug-target gene interactions as given in Table 2. Researchers can use gene set enrichment analysis to identify cancer-drug relationships, using advanced techniques like machine learning or network analysis. Integrating multi-omics data aids in drug repositioning, personalized treatments, and combinations. A meta-analysis of Non-small-cell lung cancer (NSCLC) mRNA expression profiles identified biomarkers and prognostic mechanisms.

**Role of Bioinformatics in Lead Discovery and Optimization**

Bioinformatics is rapidly expanding in genomics, proteomics, metabolomics, transcriptomics, and molecular phylogenomics, aiding drug design and identifying new substances like lead compounds through screening techniques (Arya & Coumar, 2021; Branco & Choupina, 2021). Along with the adaptation techniques, bioinformatics also builds virtual chemical libraries for the identification of new compounds (Raslan et al., 2023). The quantitative structure-activity relationship (QSAR) method offers an opportunity to know specifically the chemical configurations that are pharmacologically active (Gini, 2018). They are the epitope candidates that induce the immune response, and this is the main reason why cancer immunoinformatics has provided pointers to the new way of vaccine development (Muhammed & Aki-Yalcin, 2024). In the context of drug discovery, virtual screening, and repurposing, cost-effective and target genes' potential evaluation options, such as three-dimensional structure searches (Zitnik et al., 2018), molecular docking, and bioinformatics methods, are mostly used (Muhammed & Aki-Yalcin, 2024). The epitope candidates' predictions, which then induce the immune response, are the main reason cancer immunoinformatics has pointed to a new direction of vaccine design.

The druggable targets are evaluated by using molecular dynamics simulations, predictive tools such as fPocket (Kochnev & Durrant, 2022), etc. to the effect of the technology on the project failure risks and the accuracy of the drug target predictions that are necessary for

the research and development of vaccines (Li et al., 2020). These computer simulations of molecular dynamics allow studying the behavior of the various properties of the molecular systems in the knowledge of their structure, dynamics, and thermodynamics (Salo-Ahen et al., 2020). In the same way, the molecular docking computational method, such as the case computationally produced 3D structures of small ligands, is used to position them to interact with receptors (Jakhar et al., 2020). Repositioning drugs in the biopharma industry allows for a reconsideration of therapeutic purposes, leading to groundbreaking cancer and Parkinson's therapies, cost savings, and reduced preclinical safety testing (Ballard et al., 2020). Bioinformatics, which is the invention of technology, is proving to be the game changer, as it facilitates finding new drug targets and the optimization of drug candidates (Agamah et al., 2020). The main way to go about it is to utilize synthetic methods through which natural not found in nature are the leading structured changes absorption, distribution, metabolism, excretion, and toxicity (ADMET) performance, potency, and selectivity enhancements apart from the initial condition, the patentability of the new structure is not influenced (Ashenden, 2021).

**Table 2:** Online databases for oncogenes and drug-responsive genes

| Database | URL | References |
|---|---|---|
| The Cancer Genome Atlas (TCGA) | http://cancergenome.nih.gov | (Ganini et al., 2021) |
| Online Mendelian Inheritance in Man (OMIM) | https://www.omim.org/ | (Hamosh et al., 2021) |
| GWAS Catalog | https://www.ebi.ac.uk/gwas/ | (Sollis et al., 2023) |
| Gene Expression Omnibus (GEO) | http://www.ncbi.nlm.nih.gov/geo | (Wang et al., 2019) |
| DisGeNet | https://disgenet.com/ | (Piñero et al., 2021) |
| The Comparative Toxicogenomics Database (CTD) | https://ctdbase.org/ | (Davis et al., 2025) |
| Genetic Association Database (GAD) | https://maayanlab.cloud/Harmonizome/resource/Genetic+Association+Database | (Canela-Xandri et al., 2018) |
| ClinVar | https://www.ncbi.nlm.nih.gov/clinvar/ | (Landrum et al., 2020) |
| Catalogue of Somatic Mutations in Cancer (COSMIC) | https://cancer.sanger.ac.uk/cosmic | (Tate et al., 2019) |
| PharmGKB | https://www.pharmgkb.org/ | (Gong et al., 2021) |
| DrugBank | https://go.drugbank.com/ | (Knox et al., 2024) |
| Clinicaltrials.gov | https://clinicaltrials.gov/ | (Gresham et al., 2022) |
| The Connectivity Map (CMap) | https://www.broadinstitute.org/connectivity-map-cmap | (Zhao et al., 2023) |
| Cancer Cell Line Encyclopedia (CCLE) | https://sites.broadinstitute.org/ccle | (Nusinow & Gygi, 2020) |
| Genomics of Drug Sensitivity in Cancer (GDSC) | https://www.cancerrxgene.org/ | (Pant et al., 2024) |
| NCI60 | https://discover.nci.nih.gov/cellminer/ | (Hernández-Hernández & Ballester, 2023) |

**Drug Repositioning and Repurposing through Bioinformatics**

The biopharma industry's venture into the repurposing of drugs is identifying possible alternative medicines for existing drugs and has resulted in breakthroughs in the field of cancer and Parkinson's disease (Ballard et al., 2020). Bioinformatics advances might assist in drug repurposing by pointing out targets of choice and optimizing candidates of the drug (Patel et al., 2024). This drug, as the one used in thalidomide, has been identified by means of the bioinformatics instruments for its anti-angiogenic, anti-inflammatory, and cytokine modulation mechanisms, which show its capacity to treat cancer, erythema, autoimmune diseases, as well as bowel diseases (Amare et al., 2021). The connectivity map, a valuable resource with 18,000 user submissions, aids in researching diseases like osteoclast differentiation, diabetes, and inflammatory bowel disease by analyzing primary cell types and mRNA expression profiling (Subramanian et al., 2017). A connectivity map identified 16 compounds similar to thalidomide, providing insight into its mechanism of action and suggesting potential combinations for reprogramming in inflammatory bowel disease treatment (Pugnetti et al., 2023).

Riluzole, a glutamate antagonist that has received FDA approval, is found to have dramatically impacted the survival rate of those suffering from amyotrophic lateral sclerosis, which is a devastating and untreatable neurodegenerative disorder (Andrews et al., 2020; Hollingworth et al., 2024). Glutamate neurotransmission inhibition; which Riluzole is known for, is being considered now as a supportive therapy in Alzheimer's disease treatment (Hollingworth et al., 2024). The research indicates that riluzole treatment might also have applicability in other diseases through the same pathway (Ihara & Saito, 2020).

Bioinformatics methods have been the tools to investigate the targeted therapies for example the cancer management programs and the routes connected to specific diseases (Mahfauz et al., 2024), such as acute myeloid leukemia (AML) for which protein-protein (Gini) interaction pathways, gene ontology and Kyoto Encyclopedia of Genes and Genomes were utilized (Liu et al., 2024) Some of the diseases like cancer, Alzheimer's and metabolic dysfunction that are significantly driving the new developments of steatosis-associated liver disease are addressed on the cellular scale by means of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, MD simulation, and a compound-target-pathway association (Shi et al., 2023). The primary objective of the research was to assess the positive influence of Metformin on health span which the authors associated with the high possibility of drug-likeness, low toxicity, KEGG network correlation scores, and also the predictions made through the computational modeling (Petrie, 2024; Salekeen et al., 2024).

Dexamethasone (DEX) is a widely used medication for treating various conditions such as autoimmune diseases, allergies, ocular disorders, cancer, and COVID-19 (Madamsetty et al., 2022). Bioinformatics analyses using databases such as PubMed, GEO, and GREIN have indicated the potential repurposing of DEX for treating COVID-19. Visualization platforms such as Heatmapper, Enrichr, MSigDB, and Chip-Atlas assist with data visualization, disease pathway analysis, and transcription factor binding site localization (Sharma, 2021). DEX is likely to reduce inflammation and immune responses in critical (Ahmed & Hassan, 2020; Noreen et al., 2021).

Modern sequencing technologies and bioinformatics analyses have significantly improved our understanding of normal tissues and tumors, particularly in the study of the effects of aspirin on platelets and its potential cancer treatment (Menter & Bresalier, 2023). The study used GO and KEGG enrichment analyses to analyze metastasis-related genes expressed differentially in cancerous and normal tissues, identifying genes that could act effectively as receptors for aspirin treatment (Li et al., 2022).

**Application of Machine Learning in Drug Discovery**

Advances in computational science have significantly accelerated drug discovery and development. Artificial Intelligence (AI), especially Machine learning algorithms, significantly contribute to drug discovery, generating data, predicting properties, and optimizing bioactivity. High-throughput screening, sequencing, and online databases have created a data-driven environment, enabling drug development. Software and web tools are important enablers for the innovation of data analytics (Mateo et al., 2015).

The well-known algorithms Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest are also on the basis of ML strategies (Trigueros et al., 2019). Random Forest utilizes the uncorrelated decision trees strategy, that is, it does not consider the outliers, it classifies the data and it decreases the error rate. It also is better than SVM and Naive Bayes in collaborative predictions, which means that it can reduce the number of false positives. Naive Bayes is the Drug Information Resource best for handling biomedical data and for predicting ligand-target interactions. For example, the tools extended-connectivity fingerprint-6 and SVM make the prediction of molecules that act against the estrogen receptor and HIV much more accurate (Patel et al., 2020).

Introducing machine learning in health care has transformed the aspect of disease identification from a reactive approach to a proactive approach. ML focuses on early detection of diseases, predicting the course of disease and revolutionizing treatment planning (Rasool et al., 2023). ML methods such as SVM, logistic regression (LR), and clustering are widely used in medicine to classify and diagnose chronic diseases and will have even more importance in medical practice in the future (Battineni et al., 2020). Some of the widely used applications are mentioned in Figure 2.

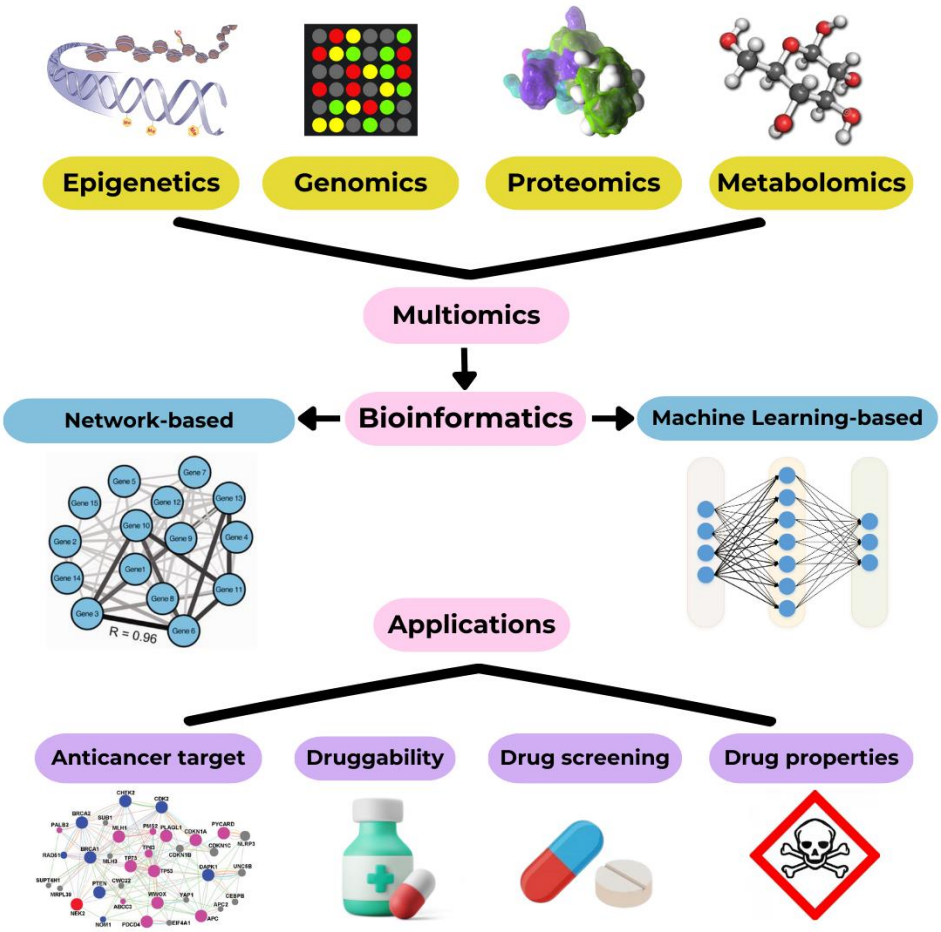


**Fig. 2:** Bioinformatics linking multi-omics to computational methods providing diverse applications in drug discovery

**Challenges and Limitations**

Bioinformatics offers great potential for drug discovery, but it will only be effective when properly directed, with some potential drawbacks. Some of the challenges lie in large data volumes (Branco & Choupina, 2021), managing biological data, General and specific metabolic changes, and drug-induced pharmacological activity on the target. With increased usage of genomic data analysis, concerns of data privacy and ethics have risen, calling for strong ethical concerns (Xia, 2017; Kullappan Malathi & Sudha Ramaiah, 2018). Collaborative efforts among bioinformaticians, healthcare professionals, ethicists, and policymakers are essential to establish responsible guidelines and legislation.

**Future Directions and Innovations**

In order to make effective drug designs, bioinformatics plays a vital role in modeling biological systems and docking proteins, mainly for antibiotic resistance, cancer treatments, as well as antiviral drugs for diseases like COVID-19. Additionally, to predict the immune system responses and omics data, bioinformatics implements mathematical models such as Swiss-Model, Autodock Vina, Avogadro, and Chimera. High-throughput technologies are also being used in bioinformatics to assist and prevent antibiotic resistance. By facilitating the rapid detection of pathogens, elucidation of transmission pathways, virulence, and antimicrobial resistance, as well as guiding marker and low-cost molecular assay discovery.

Advanced genomics analysis tools such as PlasmidFinder, ResFinder, plasmidSPAdes, and BLAST will be critical for tracking resistance traits and helping explain resistance mechanisms. These will help understand the things leading to resistance to allow accurate mapping of resistance profiles that can be used for drug development, disease monitoring, and environmental health. Engaging with these challenges is crucial for boosting the contribution of bioinformatics in the development of safe and effective therapeutics.

**Conclusion**

Bioinformatics plays a pivotal role in drug discovery, gene sequencing, gene alignment, and proteomics studies. Genomic data analysis is modernized by technological innovations in biology and medicine and offers a vision of hope for future advancements. The proposed new instruments to interpret clinical data, to dock molecules, to analyze sequences, have improved the efficiency and accuracy of these tools in drug development. Metagenomics and systems biology expand bioinformatics by considering diverse microbial populations and complex biological pathways. Collectively, the future of drug development appears bright as bioinformatics tools will be using AI and machine learning increasingly. There is a hope that treatment will become vastly more effective, much more affordable, and too personalized. Interdisciplinary collaborations, the use of new technologies, and the management of ethical questions are the main aspects through which personalized genomics and precision medicine will develop. Through advanced computing algorithms, drug targets can now be shuffled, and new drug therapies that work on the same molecular interactions are proposed by research studies. Consequently, the application insights from bioinformatics enable researchers to design novel medicines, which in the long run will ameliorate global healthcare.

## References


Agamah, F. E., Mazandu, G. K., Hassan, R., Bope, C. D., Thomford, N. E., Ghansah, A., & Chimusa, E. R. (2020). Computational/in silico methods in drug target and lead prediction. *Briefings in bioinformatics*, *21*(5), 1663-1675.

Ahmed, M. H., & Hassan, A. (2020). Dexamethasone for the treatment of coronavirus disease (COVID-19): a review. *SN comprehensive clinical Medicine*, *2*(12), 2637-2646.

Alameer, A., & Chicco, D. (2022). geoCancerPrognosticDatasetsRetriever: a bioinformatics tool to easily identify cancer prognostic datasets on Gene Expression Omnibus (GEO). *Bioinformatics*, *38*(6), 1761-1763.

Ali, M. M., Hamid, M., Saleem, M., Malik, S., Mian, N. A., Ihsan, M. A., & Awan, F. (2021). [Retracted] Status of Bioinformatics Education in South Asia: Past and Present. *BioMed Research International*, *2021*(1), 5568262.

Amare, G. G., Meharie, B. G., & Belayneh, Y. M. (2021). A drug repositioning success: The repositioned therapeutic applications and mechanisms of action of thalidomide. *Journal of Oncology Pharmacy Practice*, *27*(3), 673-678.

Andrews, J. A., Jackson, C. E., Heiman-Patterson, T. D., Bettica, P., Brooks, B. R., & Pioro, E. P. (2020). Real-world evidence of riluzole effectiveness in treating amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *21*(7-8), 509-518.

Arya, H., & Coumar, M. S. (2021). Lead identification and optimization. In *The design & development of novel drugs and vaccines* (pp. 31-63). Elsevier.

Ashenden, S. K. (2021). Lead optimization. In *The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry* (pp. 103-117). Elsevier.

Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., & Papatheodorou, I. (2019). ArrayExpress update–from bulk to single-cell expression data. *Nucleic Acids Research*, *47*(D1), D711-D715.

Bakchi, B., Krishna, A. D., Sreecharan, E., Ganesh, V. B. J., Niharika, M., Maharshi, S.,...Shaik, A. B. (2022). An overview on applications of SwissADME web tool in the design and development of anticancer, antitubercular and antimicrobial agents: a medicinal chemist's perspective. *Journal of Molecular Structure*, *1259*, 132712.

Ballard, C., Aarsland, D., Cummings, J., O'Brien, J., Mills, R., Molinuevo, J. L.,...Corbett, A. (2020). Drug repositioning and repurposing for Alzheimer disease. *Nature Reviews Neurology*, *16*(12), 661-673.

Ballard, C. G., Kreitzman, D. L., Isaacson, S., Liu, I.-Y., Norton, J. C., Demos, G., & Ferreira, J. J. (2020). Long-term evaluation of open-label pimavanserin safety and tolerability in Parkinson's disease psychosis. *Parkinsonism & Related Disorders*, *77*, 100-106.

Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of Personalized Medicine*, *10*(2), 21.

Berdigaliyev, N., & Aljofan, M. (2020). An overview of drug discovery and development. *Future medicinal chemistry*, *12*(10), 939-947.

Bossé, R. (2018). The future of drug discovery. *Drug Discovery*, 9.

Branco, I., & Choupina, A. (2021). Bioinformatics: new tools and applications in life science and personalized medicine. *Applied Microbiology and Biotechnology*, *105*, 937-951.

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., & Feng, Z. (2022). RCSB Protein Data Bank: Celebrating 50 years of the PDB with new tools for understanding and visualizing biological macromolecules in 3D. *Protein Science*, *31*(1), 187-208.

Canela-Xandri, O., Rawlik, K., & Tenesa, A. (2018). An atlas of genetic associations in UK Biobank. *Nature Genetics*, *50*(11), 1593-1599.

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Magain, N., & Lutzoni, F. (2019). T-BAS Version 2.1: Tree-Based Alignment

Selector toolkit for evolutionary placement of DNA sequences and viewing alignments and specimen metadata on curated and custom trees. *Microbiol Resour Announc*, 8: e00328-19. In.

Clark, A. J., & Lillard Jr, J. W. (2024). A comprehensive review of bioinformatics tools for genomic biomarker discovery driving precision oncology. *Genes*, *15*(8), 1036.

Cui, Z., Li, C., Chen, P., & Yang, H. (2022). An update of label-free protein target identification methods for natural active products. *Theranostics*, *12*(4), 1829.

Davis, A. P., Wiegers, T. C., Sciaky, D., Barkalow, F., Strong, M., Wyatt, B., & Mattingly, C. J. (2025). Comparative toxicogenomics database's 20th anniversary: update 2025. *Nucleic Acids Research*, *53*(D1), D1328-D1334.

Estaki, M., Jiang, L., Bokulich, N. A., McDonald, D., González, A., Kosciolek, T., & Vázquez-Baeza, Y. (2020). QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Current protocols in bioinformatics*, *70*(1), e100.

Ganini, C., Amelio, I., Bertolo, R., Bove, P., Buonomo, O. C., Candi, E., & Mauriello, A. (2021). Global mapping of cancers: The Cancer Genome Atlas and beyond. *Molecular Oncology*, *15*(11), 2823-2840.

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in bioinformatics*, *20*(6), 1981-1996.

Gini, G. (2018). QSAR: what else? *Computational Toxicology: Methods and Protocols*, 79-105.

Gong, L., Whirl-Carrillo, M., & Klein, T. E. (2021). PharmGKB, an integrated resource of pharmacogenomic knowledge. *Current Protocols*, *1*(8), e226.

Gresham, G., Meinert, J. L., Gresham, A. G., Piantadosi, S., & Meinert, C. L. (2022). Update on the clinical trial landscape: analysis of ClinicalTrials. gov registration data, 2000–2020. *Trials*, *23*(1), 858.

Gudmundsson, S., Singer-Berk, M., Watts, N. A., Phu, W., Goodrich, J. K., Solomonson, M., & O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*, *43*(8), 1012-1030.

Hamosh, A., Amberger, J. S., Bocchini, C., Scott, A. F., & Rasmussen, S. A. (2021). Online mendelian inheritance in man (OMIM®): Victor MCKUSICK's magnum opus. *American Journal of Medical Genetics Part A*, *185*(11), 3259-3265.

Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., & Bhai, J. (2024). Ensembl 2024. *Nucleic acids research*, *52*(D1), D891-D899.

Hernández-Hernández, S., & Ballester, P. J. (2023). On the best way to cluster NCI-60 molecules. *Biomolecules*, *13*(3), 498.

Hollingworth, D., Thomas, F., Page, D. A., Fouda, M. A., De Castro, R. L.-R., Sula, A., & Ruben, P. C. (2024). Structural basis for the rescue of hyperexcitable cells by the amyotrophic lateral sclerosis drug Riluzole. *Nature Communications*, *15*(1), 8426.

Hung, J.-H., & Weng, Z. (2016). Sequence alignment and homology search with BLAST and ClustalW. *Cold Spring Harbor Protocols*, *2016*(11), pdb-prot093088.

Ihara, M., & Saito, S. (2020). Drug repositioning for Alzheimer's disease: finding hidden clues in old drugs. *Journal of Alzheimer's Disease*, *74*(4), 1013-1028.

Jakhar, R., Dangi, M., Khichi, A., & Chhillar, A. K. (2020). Relevance of molecular docking studies in drug designing. *Current Bioinformatics*, *15*(4), 270-278.

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, *51*(D1), D587-D592.

Keegan, K. P., Glass, E. M., & Meyer, F. (2016). MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Microbial environmental genomics (MEG)*, 207-233.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., & Yu, B. (2023). PubChem 2023 update. *Nucleic acids research*, *51*(D1), D1373-D1380.

Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., & Strawbridge, S. A. (2024). DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic acids research*, *52*(D1), D1265-D1275.

Kochnev, Y., & Durrant, J. D. (2022). FPocketWeb: protein pocket hunting in a web browser. *Journal of Cheminformatics*, *14*(1), 58.

Kullappan Malathi, K. M., & Sudha Ramaiah, S. R. (2018). Bioinformatics approaches for new drug discovery: a review.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), 1547-1549.

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., & Liu, C. (2020). ClinVar: improvements to accessing data. *Nucleic acids research*, *48*(D1), D835-D844.

Lee, C. M., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Gonzalez, J. N., & Powell, C. C. (2020). UCSC Genome Browser enters 20th year. *Nucleic acids research*, *48*(D1), D756-D761.

Lefort, V., Longueville, J.-E., & Gascuel, O. (2017). SMS: smart model selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422-2424.

Li, J., Shen, C., Jiang, X., Cui, X., Wu, Z., Liu, Y.,...Shen, A. (2022). Establishment of metastasis-related prognostic models for colon cancer and bioinformatics analysis of the benefits of aspirin chemoprophylaxis.

Li, K., Du, Y., Li, L., & Wei, D.-Q. (2020). Bioinformatics approaches for anti-cancer drug discovery. *Current Drug Targets*, *21*(1), 3-17.

Liu, S., Xu, M., Yang, Z., Li, Y., Wu, D., & Tang, X. (2024). Network pharmacology-based investigation and experimental validation of the mechanism of metformin in the treatment of acute myeloid leukemia. *European Journal of Medical Research*, *29*(1), 475.

Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., & Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature protocols*, *17*(12), 2815-2839.

Madamsetty, V. S., Mohammadinejad, R., Uzieliene, I., Nabavi, N., Dehshahri, A., Garcia-Couce, J., & Makvandi, P. (2022). Dexamethasone: insights into pharmacological aspects, therapeutic mechanisms, and delivery systems. *ACS Biomaterials Science & Engineering*, *8*(5), 1763-1790.

Mahfauz, M., Yuruker, O., & Kalkan, R. (2024). Repurposing metformin as a potential anticancer agent using in silico technique. *DARU Journal of Pharmaceutical Sciences*, 1-7.

Mateo, J., Carreira, S., Sandhu, S., Miranda, S., Mossop, H., Perez-Lopez, R., & Tunariu, N. (2015). DNA-repair defects and olaparib in metastatic prostate cancer. *New England Journal of Medicine*, *373*(18), 1697-1708.

McDonagh, E. M., Trynka, G., McCarthy, M., Holzinger, E. R., Khader, S., Nakic, N., & Hulcoop, D. (2024). Human Genetics and Genomics for Drug Target Identification and Prioritization: Open Targets' Perspective. *Annual Review of Biomedical Data Science*, *7*.

Menter, D. G., & Bresalier, R. S. (2023). An aspirin a day: new pharmacological developments and cancer chemoprevention. *Annual Review of Pharmacology and Toxicology*, *63*(1), 165-186.

Moumné, L., Marie, A.-C., & Crouvezier, N. (2022). Oligonucleotide therapeutics: from discovery and development to patentability. *Pharmaceutics*, *14*(2), 260.

Muhammed, M. T., & Aki-Yalcin, E. (2024). Molecular docking: Principles, advances, and its applications in drug discovery. *Letters in Drug Design & Discovery*, *21*(3), 480-495.

Muthiah, I., Rajendran, K., & Dhanaraj, P. (2021). In silico molecular docking and physicochemical property studies on effective phytochemicals targeting GPR116 for breast cancer treatment. *Molecular and Cellular Biochemistry*, *476*, 883-896.

Noreen, S., Maqbool, I., & Madni, A. (2021). Dexamethasone: Therapeutic potential, risks, and future projection during COVID-19 pandemic. *European journal of pharmacology*, *894*, 173854.

Nusinow, D. P., & Gygi, S. P. (2020). A guide to the quantitative proteomic profiles of the cancer cell line encyclopedia. *bioRxiv*, 2020-2002.

Okpo, E. A., Agboke, A. A., Udobi, C. E., John, G. E., & Andy, I. E. (2024). The synergy of molecular docking and bioinformatics: an in depth review in drug discovery. *Biotechnology Journal International*, *28*(4), 119-136.

Pant, L., Al Mukaddim, A., Rahman, M. K., Sayeed, A. A., Hossain, M. S., Khan, M. T., & Ahmed, A. (2024). Genomic predictors of drug sensitivity in cancer: Integrating genomic data for personalized medicine in the USA. *Computer Science & IT Research Journal*, *5*(12), 2682-2702.

Patel, B., Gelat, B., Soni, M., & Rathaur, P. (2024). Bioinformatics Perspective of Drug Repurposing. *Current Bioinformatics*, *19*(4), 295-315.

Patel, L., Shukla, T., Huang, X., Ussery, D. W., & Wang, S. (2020). Machine learning methods in drug discovery. *Molecules*, *25*(22), 5277.

Patil, N. S., & Rohane, S. H. (2021). Organization of Swiss Dock: in study of computational and molecular docking study.

Petrie, J. R. (2024). Metformin beyond type 2 diabetes: Emerging and potential new indications. *Diabetes, Obesity and Metabolism*, *26*, 31-41.

Picci, G., Marchesan, S., & Caltagirone, C. (2022). Ion channels and transporters as therapeutic agents: from biomolecules to supramolecular medicinal chemistry. *Biomedicines*, *10*(4), 885.

Piñero, J., Saüch, J., Sanz, F., & Furlong, L. I. (2021). The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Computational and Structural Biotechnology Journal*, *19*, 2960-2967.

Pugnetti, L., Curci, D., Bidoli, C., Gerdol, M., Celsi, F., Renzo, S., & Maestro, A. (2023). Gene expression profiling in white blood cells reveals new insights into the molecular mechanisms of thalidomide in children with inflammatory bowel disease. *Biomedicine & Pharmacotherapy*, *164*, 114927.

Raslan, M. A., Raslan, S. A., Shehata, E. M., Mahmoud, A. S., & Sabri, N. A. (2023). Advances in the Applications of Bioinformatics and Chemoinformatics. *Pharmaceuticals*, *16*(7), 1050.

Rasool, S., Husnain, A., Saeed, A., Gill, A. Y., & Hussain, H. K. (2023). Harnessing predictive power: exploring the crucial role of machine learning in early disease detection. *JURIHUM: Jurnal Inovasi dan Humaniora*, *1*(2), 302-315.

Ravi, K., & Paramasivam, G. (2023). In silico screening of inhibitors for programmed death ligand 1 (PDL-1) protein from anti-metabolite and plant alkaloid sources using patchdock software for the therapy of colorectal adenocarcinoma.

Salekeen, R., Lustgarten, M. S., Khan, U., & Islam, K. M. D. (2024). Model organism life extending therapeutics modulate diverse nodes in the drug-gene-microbe tripartite human longevity interactome. *Journal of Biomolecular Structure and Dynamics*, *42*(1), 393-411.

Salo-Ahen, O. M. H., Alanko, I., Bhadane, R., Bonvin, A. M. J. J., Honorato, R. V., Hossain, S., & Larsen, A. S. (2020). Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, *9*(1), 71.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic acids research*, *47*(D1), D94-D99.

Shaker, B., Ahmad, S., Lee, J., Jung, C., & Na, D. (2021). In silico methods and tools for drug discovery. *Computers in Biology and Medicine*, *137*, 104851.

Sharma, A. (2021). Inferring molecular mechanisms of dexamethasone therapy in severe COVID-19 from existing transcriptomic data. *Gene*, *788*, 145665.

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, *50*(W1), W216-W221.

Shi, X., Li, L., Liu, Z., Wang, F., & Huang, H. (2023). Exploring the mechanism of metformin action in Alzheimer's disease and type 2 diabetes based on network pharmacology, molecular docking, and molecular dynamic simulation. *Therapeutic Advances in Endocrinology and Metabolism*, *14*, 20420188231187493.

Sievers, F., & Higgins, D. G. (2020). The clustal omega multiple alignment package. In *Multiple sequence alignment: Methods and protocols* (pp. 3-16). Springer.

Singh, S. K., Kumar, A., Singh, R. B., Ghosh, P., & Bajad, N. G. (2022). Recent Applications of Bioinformatics in Target Identification and Drug Discovery for Alzheimer's Disease. *Current Topics in Medicinal Chemistry*, *22*(26), 2153-2175. https://doi.org/https://doi.org/10.2174/1568026623666221026091010

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., & Hayhurst, J. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, *51*(D1), D977-D985.

Southern, J., Gonzalez, G., Borgas, P., Poynter, L., Laponogov, I., Zhong, Y., & Veselkov, K. (2023). Genomic-driven nutritional interventions

for radiotherapy-resistant rectal cancer patient. *Scientific Reports*, *13*(1), 14862.
Sriramulu, D. K., & Lee, S.-G. (2021). Effect of molecular properties of the protein-ligand complex on the prediction accuracy of AutoDock. *Journal of Molecular Graphics and Modelling*, *106*, 107921.
Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., & Golub, T. R. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, *171*(6), 1437-1452.e1417. https://doi.org/10.1016/j.cell.2017.10.049
Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, *12*(7), 3049-3062.
Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., & Bork, P. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, *49*(D1), D605-D612.
Tabana, Y., Babu, D., Fahlman, R., Siraki, A. G., & Barakat, K. (2023). Target identification of small molecules: an overview of the current applications in drug discovery. *BMC biotechnology*, *23*(1), 44.
Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., & Dawson, E. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Research*, *47*(D1), D941-D947.
Trigueros, S., B. Domènech, E., Toulis, V., & Marfany, G. (2019). In vitro gene delivery in retinal pigment epithelium cells by plasmid DNA-wrapped gold nanoparticles. *Genes*, *10*(4), 289.
Wang, Z., Lachmann, A., & Ma'ayan, A. (2019). Mining data and metadata from the gene expression omnibus. *Biophysical reviews*, *11*, 103-110.
Xia, X. (2017). Bioinformatics and drug discovery. *Current Topics in Medicinal Chemistry*, *17*(15), 1709-1726.
Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., & Lu, A. (2021). ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research*, *49*(W1), W5-W14.
Yuan, S., Chan, H. C. S., & Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *7*(2), e1298.
Zaru, R., Orchard, S., & UniProt, C. (2023). UniProt tools: BLAST, align, peptide search, and ID mapping. *Current Protocols*, *3*(3), e697.
Zhang, X., Wu, F., Yang, N., Zhan, X., Liao, J., Mai, S., & Huang, Z. (2022). In silico methods for identification of potential therapeutic targets. *Interdisciplinary Sciences: Computational Life Sciences*, 1-26.
Zhao, Y., Chen, X., Chen, J., & Qi, X. (2023). Decoding Connectivity Map-based drug repurposing for oncotherapy. *Briefings in Bioinformatics*, *24*(3), bbad142.
Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, *34*(13), i457-i466