Proceedings of Scientific Research Universidad Anáhuac, 2022, 2(3): 27-35.
ISSN: in process.

Review Article

Bioinformatics approaches for Biomedical Research

Kasandra Aguilar-Cázarez[a1], Ernesto Andrade-Collantes[b2], Marisol Verdugo-Meza[b3], Claudia María de-la-Rocha-Morales[b4], Cruz Fernando López-Carrera[c5], Paúl Alexis López-Durán[d6]*

[a]Centro de Investigación de Estudios Avanzados, Ciudad de México, México.
[b]Universidad Autónoma de Sinaloa, Facultad de Ciencias Químico Biológicas, Sinaloa, México.
[c]Instituto Politécnico Nacional, Escuela Nacional de Ciencias Biológicas, Ciudad de México, México.
[d]Universidad Anáhuac México, Campus Norte, Facultad de Ciencias de la Salud, Estado de México, México.

ID ORCID:
[1]https://orcid.org/0000-0002-5740-7296, [2]https://orcid.org/0000-0002-7924-0132, [3]https://orcid.org/0000-0002-8300-049X,
[4]https://orcid.org/0000-0002-4505-5393, [5]https://orcid.org/0000-0003-2643-7713, [6]https://orcid.org/0000-0002-2577-6599.

https://doi.org/10.36105/psrua.2022v2n3.04

ABSTRACT

An enormous amount of data is generated and compiled in several databases every year. Along with this, comes a demand for the analysis and interpretation of the entirety of this biological information. Taking care of this task, bioinformatics promises breakthroughs in research and development in complex biomedical areas. In just a few years since its beginning, bioinformatics has led to great progress and demonstrated its potential. It has created an opportunity to solve arising medical and molecular issues faster and more efficiently, as compared to the traditional approach. The present review aims to present some of the main applications of bioinformatics in the field of biomedicine, such as comparative genomics, biomarker identification, computer-aided drug design, vaccine design, and personalized medicine. In addition, we also cover some of its steadily reduced limitations.

Key words: bioinformatics; biomedicine; comparative genomics; biomarkers; computer-aided drug design; vaccine design; personalized medicine.

RESUMEN

Cada año se genera una enorme cantidad de datos recopilados en bases de datos. Junto con esto surge una demanda de análisis e interpretación de la totalidad de la información biológica; esta importante tarea es atendida por la bioinformática, la cual promete grandes avances en la investigación y desarrollo de áreas biomédicas complejas. En tan sólo unos años desde su inicio ha llevado a un gran progreso y ha demostrado su potencial, creando una oportunidad para resolver los problemas médicos y moleculares de manera más rápida y eficiente en comparación con el enfoque tradicional. El presente artículo pretende discutir algunas de las principales aplicaciones de la bioinformática en el campo de la biomedicina, tales como, genómica comparativa, identificación de biomarcadores, diseño de fármacos, diseño de vacunas y medicina personalizada. De manera adicional, se enlistan algunas de sus principales limitaciones, las cuales están en constante disminución.

Palabras clave: bioinformática; biomedicina; genómica comparativa; biomarcadores; diseño computarizado de fármacos; diseño de vacunas; medicina personalizada.

* *Corresponding author:* Paúl Alexis López-Durán. Universidad Anáhuac México, Campus Norte, Facultad de Ciencias de la Salud. Address: Av. Universidad Anáhuac núm. 46, Lomas Anáhuac, C.P. 52786, Huixquilucan, Estado de México, México. Tel.: +52 55 5627 0210. Email: paul.lopezdu@anahuac.mx

Received: 19 November 2021.
Accepted: 14 February 2022.

INTRODUCTION

With the increasing development of next-generation sequencing (NGS), large-scale high-throughput molecular profiling studies and omics sciences, we now have a better understanding of the complex nature of numerous illnesses.[1] Each year, large amounts of data are generated and compiled in freely accessible databases[2], bringing along a high demand for their analysis and interpretation. This is managed by evolving bioinformatics.

Bioinformatics is defined as an interdisciplinary science that applies computational tools and analysis to the capture and interpretation of biological data, which brings together computer science, mathematics, physics, and biology.[3] It requires research, development, and application of informatics tools and approaches to acquire, store, visualize, and interpret biological and medical data.[4] It allows to test hypotheses through *in silico* methods to have a better knowledge before proceeding with expensive studies. In addition, bioinformatics provides results that are more accurate when assembling reliable interpretations. Therefore, it promises breakthroughs in research and development in complex biomedical systems as well as public health, drug design, comparative genomics, and personalized medicine, among others.[5]

From the first use of molecular sequences for evolutionary studies by Zuckerkandl and Pauling (1965)[4] to the massive up-to-date sequencing programs, bioinformatics has evolved just as much as its matter of study.[6] It has branched out into bioinformatics tools applicable in genomics, proteomics, drug design, and simulations of molecular dynamics. Combined, they provide a more complete understanding of how diseases work.[7] As an emerging field, it is essential for managing data in modern biology and medicine; for instance, it is a major protagonist in the development of all vaccines against COVID-19.[8] (Fig. 1).

In the present review, the main points on topics of novel interest in the area of applied bioinformatics in biomedicine are discussed.

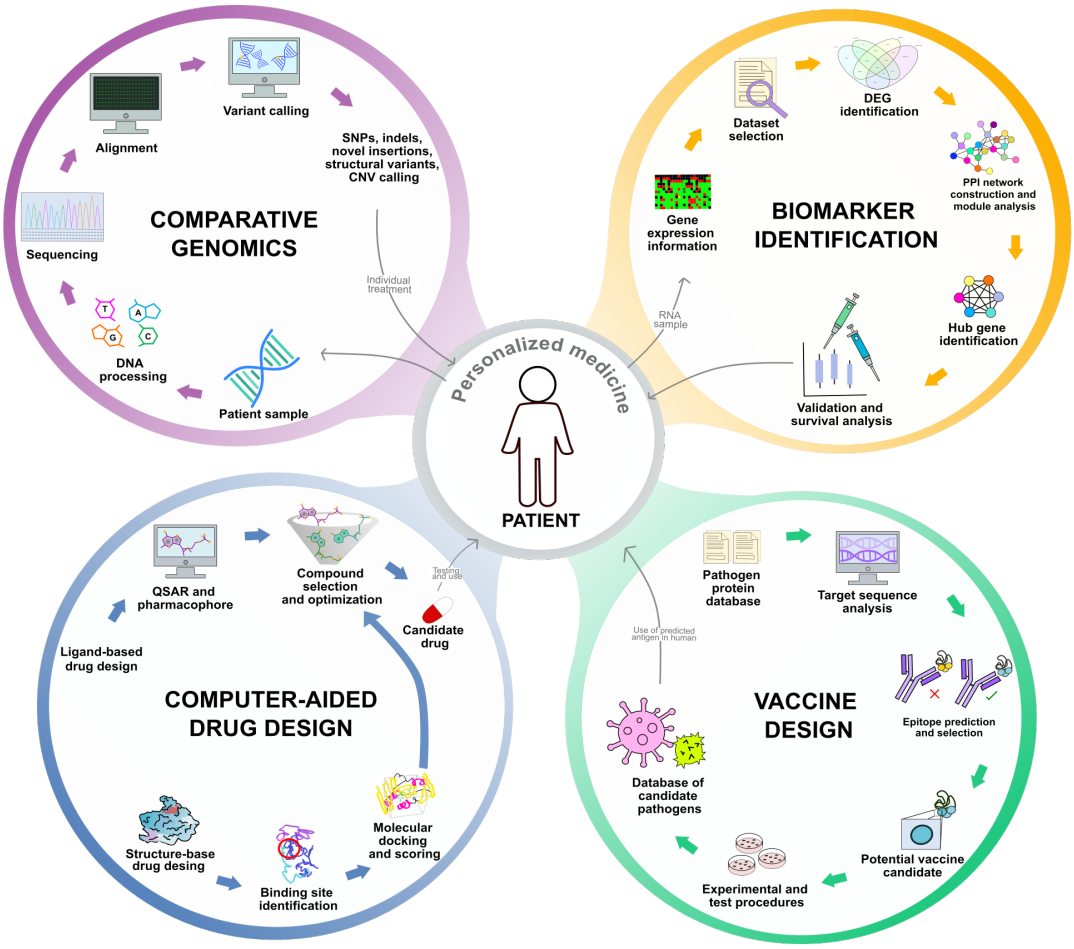

FIGURE 1. Applications of bioinformatics in biomedicine. The general workflow starting from the patient is summarized for each bioinformatics application: Comparative genomics, biomarker identification, Computer-aided drug design, and vaccine design. SNPs: single nucleotide polymorphisms; CNV: copy number variation; DEGs: differentially expressed genes; PPI: protein-protein interaction; QSAR: quantitative structure–activity relationship.

https://doi.org/10.36105/psrua.2022v2n3.04

Comparative genomics

Since the sequencing of the human genome, scientists have developed a special interest in the field of biological research called comparative genomics (CG).[9] CG can be defined as the comparison of biological information derived from whole genome sequences.[10] It aims to determine similarities and differences between genomes to provide information about the biology of the respective organisms, describing the structure and identifying coding and non-coding regions of the genome.[11] The comparison of DNA sequences can be applied in several areas. This information reveals the molecular basis of individuality, uncovers novel regulatory mechanisms, predicts the metabolic capabilities of an organism, and enables the prediction of gene function, among others.[10] Assuming that you have a complete genome, the steps in CG start with genome annotation; that is, gene finding and function assignment.[12] The process of gene search predicts the section of the genome that contains genes while the function assignment seeks to predict the function of the coded proteins. This process is performed by automated software algorithms, called pipelines, using resources from databases and biological data. Annotation allows for clustering genes in homolog or ortholog families.[12] Diverse applications can be inferred with comparative genomics; in definition, unique signatures can be detected by forensic microbiology.[13] For instance, virulence and antibiotic resistance genes can be identified as targets for genetic manipulation in public health. Sequence comparison can be used to identify specific types of parasites for treatment or diagnosis.[14] Nowadays, the comparison of SARS-COV-2 distribution, spread, and evolution uses CG for multiple purposes.[15] It can also be used in the advanced molecular characterization of infectious agents in clinical environments. This could improve both the identification and genetic characterization including resistance profiles, promoting outbreak investigations and molecular surveillance.[16] For instance, the comparison of metagenomes could help to understand the dynamics of a given microbiota, disease versus healthy states (e.g., cancer tumor versus normal tissue)[17], different diets[18], and different geographical locations.[19]

Biomarker identification

Biomarkers can be defined as "an indicator of normal biological processes, pathogenic processes or pharmacological response to a therapeutic intervention"[20], which may have a monitoring, diagnostic, or predictive value[21]. They can be single nucleotide polymorphisms (SNPs), structural variants, circulating DNAs, methylated DNAs, mRNAs, microRNAs (miRNAs), proteins, and metabolites.[22]

In genomics, there are two approaches, the search for unique genetic variants that can act as predictors of disease and the composite predictor in the form of a genetic risk score.[23] On their behalf, the RNA and protein profiling strategies have an important biological impact. Despite current practical limitations, transcriptomic assays together with proteomics can provide essential biological information.[24] This kind of analysis is performed thanks to high-throughput technologies, including RNA sequencing and microarrays.[24] Bioinformatics analysis is a valuable tool with great potential, especially when there is a discrepancy between genomic alterations and gene expression. Statistical tools are used to differentiate signal and noise in the output of gene expression.[25] Gene expression meta-analysis is increasingly used in many fields to improve the reproducibility of a study and discover new robust biomarkers.[26] The necessary data are frequently obtained from the gene expression omnibus (GEO) database. Additionally, metabolomics, the profiling of metabolites in biofluids, cells and tissues, is routinely applied as a tool for biomarker discovery.[27] A metabolome-based strategy for identifying candidates with biological activity would consist of statistically analyze a list of metabolites generated from databases. Metabolic pathways and bioinformatics analysis of interaction networks can be used to reduce data complexity.[28] The most important step in determining the activity is to use an appropriate screening strategy. This includes gene expression, protein expression, and protein activity, the modulation of desired cellular phenotypes.[28]

The protein/gene-protein/gene interaction networking, hub gene identification, gene enrichment, and functional gene annotation analyses are powerful tools for the identify potential diagnosis and treatment biomarkers in diseases such as cancer[1,2,29-31], bipolar disorder[2], depression[32], diabetes[33], arteriosclerosis[34], and others. For instance, Wan et al. in 2020 applied bioinformatics analysis to identify eight candidate genes that could be a potential prognostic marker of thyroid carcinoma based on expression analysis profiles from the GEO database.[31] Liu et al. in 2019 identified several hub genes and key pathways associated with bipolar disorder (BD) based on a gene co-expression network analysis, which might have important clinical implications for BD treatment and diagnosis.[35]

Computer-aided drug design

The traditional methodology for novel drug discovery is a costly, extremely risky, and time-consuming process. Indeed, 75% of the total cost of drug development corresponds to lead molecules in clinical trials that never enter the consumer clinical market.[36] The steps of traditional drug discovery can be summarized as follows: target identification, target validation, lead identification, lead optimization, preclinical stage, and clinical trial stage.[37]

To address the challenges of traditional drug discovery, new strategies such as computer-aided drug design (CADD) have been developed. CADD comprises approaches such as *ab*

initio design, toxicity profile, quantitative structure–activity relationship (QSAR), docking, molecular dynamics (MD), quantitative free-energy calculations, and homology modeling.[38]

These approaches can be classified in two categories: structure-based drug design (SBDD) and ligand-based drug design (LBDD). The first depends on the availability of the 3D target structure to screen promising ligand molecules by calculating the interaction energies between the target and compound. The knowledge of the binding site in the target protein structure is used to identify and evaluate ligands based on their interactions with the residues present in the active site.[39] LBDD is used when the drug target structure is unavailable. The knowledge of the molecules interacting with the drug target is applied to develop pharmacophore and 3D-QSAR models to understand the characteristics a molecule must possess to bind to the target.[40]

CADD has emerged as a crucial ally in the design of therapeutic agents against COVID-19 since the beginning of the outbreak. For instance, Elfiky (2020) used homology modeling, MD, and molecular docking to report the relevance of sofosbuvir, ribavirin, galidesivir, remdesivir, and others as candidate drugs for clinical trials.[41]

Vaccine design

Vaccine design is supported by bioinformatics, which identifies and predicts the essential components of a vaccine (target antigen, B and T epitopes, linker, and adjuvant). These could be coupled with molecular modeling, MD simulations, and protein–protein docking, among others, to help scientists predict the adequate properties of vaccine construction in a shorter time by making conscientious decisions. Vaccine design allows the analysis of pathogenic organisms that cannot be cultivated *in vitro* and improves the allergen filter process for a better selection.[3,42]

Reverse vaccinology (RV) and Structural vaccinology (SV), two bioinformatics procedures, result in fundamental steps to achieve an optimum vaccine. RV defines the process of antigen discovery for further vaccine development, starting from genome information.[43] It involves the identification of novel antigens and the design of B- and T-cell epitopes through the study of the genome of an organism.[44] It is also useful to find genes encoding proteins that could reveal adequate epitopes.[45-47] RV employs software to determine antigenic and physicochemical characteristics associated with antigenic epitopes.[48] On the other hand, SV uses 3D structural information to design novel and/or improved peptide-based vaccine antigens.[49] It assesses the 3D molecular structure of the epitope in the antigen–antibody complex, creates MD simulations to predict and model the epitope and its interactions, tests reengineered antigen into immunoinformatics platforms (i.e., epitome), and finally assesses vaccine candidates (VC) for efficacy and safety *in vivo*.[50]

Diverse platforms exist with the purpose of exploring epitopes and simulating antigen-protein docking, present antibody structures, and antigen-antibody reactions, among others. Many of these pipelines have been created or refocused before the advent of the pandemic.[44,51]

Recently, the procedures mentioned above have been used extensively in the face of the emergence of the COVID-19 pandemic. Due to the current context, whole genome sequences have been available since December 2019 and, therefore, were perfect candidates for vaccine design. In this sense, Grifoni et al. (2020) made predictions for effective epitopes that can facilitate a vaccine design against the SARS-COV-2 virus, especially supported by the platform IEDB and its incorporated tools as ViPR[52] designed a multi-epitope peptide vaccine against COVID-19 using immunoinformatics. They used IEDB for T-cell epitope prediction and RaptorX[5] for 3D structural modeling.[53]

Personalized medicine

Personalized precision medicine is a new strategy where different methods, diagnoses, and therapies are applied to the peculiarities each patient presents. It also serves as prevention of emerging diseases, considering the patient's genome, lifestyle, and environment. This may improve side effects related to medications and generate successful therapies. Thanks to the advances in bioinformatics in recent years, changes have been generated in medicine and treatments for patients. An earlier detection of diseases has been achieved, along with a better targeted therapy. In addition, genome sequencing and data analysis have played a key role in the development of efficient personalized medicine.[22]

The main tools of precision medicine are multi-omics techniques and the data obtained from these, such as DNA sequences, transcriptomes, proteomes, metabolomes, epigenomes, and microbiomes. There are different bioinformatic tools for the data processing and analysis of genomic sequences, while the software to be used depends on the origin of the genetic information.[54]

The new sequencing technologies and their rapid development gave rise to the so-called new P4 medicine (predictive, preventive, personalized, and participatory), with novel approaches, opportunities, and challenges.[55] Firstly, because of prediction and a deeper knowledge of biology and diseases using genomic technologies, individuals prone to certain diseases can be identified even before they show symptoms. Secondly, prevention depends on early detection. People can be

https://doi.org/10.36105/psrua.2022v2n3.04

guided to improve their lifestyle, due to the risk of contracting certain diseases, avoid health problems, and choose the best treatment for their disease. Thirdly, personalization is related to the knowledge of the patient's genome and lifestyle. By understanding the biology of the disease in a better way, an adequate therapy can be prescribed, resulting in fewer unwanted side effects.

Finally, the participatory approach uses the patient's information. Platforms and tools are created for the discussion of diseases and allow for general knowledge.[55,56]

Advances in bioinformatics in precision medicine improve efficiency in the way people are characterized and the way personalized medicines are made and tested to demonstrate their usefulness. With the big data obtained, there is a better understanding of the diseases of certain groups of patients, a better diagnosis, as well as a better development of specific drugs. It is possible to reduce the adverse effects of aggressive treatments and generate more information on rare and complex diseases. The holistic approach of personalized precision medicine will help to gain more information to improve people's quality of life.[56]

Limitations of bioinformatics

Bioinformatics has proven to be a useful tool in scientific research fields. It provides an exceptional and outstanding platform and opportunity for scientists to integrate omics science[57], bioinformatics tools and data, clinical research, disease-specific biomarkers[58], dynamic networks, and precision medicine.[59] It creates an opportunity to solve arising issues faster and more efficiently.

Nevertheless, technological biases and limitations may not have been sufficiently considered in the development of bioinformatics tools. Herein, we mention and further explain what we consider to be the main limitations bioinformatics face and how technology must move forward in order to bridge these gaps.

1. Amounts of data

Contemporary sequencing technologies are capable of generating as much as 600–1000 GB per run. Further analysis of raw sequences will increase the amount of data by a factor of ten to twenty in each analysis. This general development of sequencing techniques has led to an increase in the depth of sequencing, generating a critical problem of data storage.[60] The rapid development of sequencing techniques has brought sequencing-based research to bear in all areas of life sciences. However, as the technology grew faster than its computational counterpart did, unforeseen amounts of sequencing data were generated, posing a real challenge for data analysis and bioinformatics in general.

2. Dependence on reference databases

Most bioinformatics analysis pipelines depend on sequence comparison against reference databases.[61,62] This can be problematic considering the potential incompleteness of databases. The increased error rate of the current emerging long-read technologies can have a negative impact on biological interpretations. For example, errors in protein-coding regions can affect the accuracy of protein predictions.[63] Recent studies apply correction methods (hybrid error correction) to show that the original error rate of 19% is the limit for perfect correction; beyond that, long reads are too error-prone to be corrected by these methods. Therefore, updated databases with better quality data and innovative bioinformatics tools are necessary to support a stable and effective analysis.[64]

3. Genetic information privacy

Genetic testing companies that offer direct-to-consumer (DTC) services are in constant growth, raising new privacy and ethical concerns. These companies collect data from increasing numbers of people, some of whom may have shared their data without knowing the consequences for themselves and their blood relatives.Privacy breaches can have serious social implications and adversely affect genomic-driven research by decreasing data collection and data sharing.[65,66] Then, it is essential to ensure privacy as both a fundamental right for individuals and a strategy to support responsible data sharing.

4. Experimental proof and lab work are still required

Bioinformatics findings and research may provide evidence for the progress of potential biomarkers and help understand the role of certain genes and/or proteins in any given ailment. Still, these theoretical breakthroughs are still to be proven experimentally if they are used as therapeutic treatments.[67-69] So, public and private investment for the generation and transfer of knowledge could support the development of high-quality translational and collaborative research. This, in turn, will prevent and help to face menacing situations, as the current COVID-19 pandemic, or older issues such as obesity and cancer.[70]

5. Bioinformatics: not for everyone

Every novel or inexperienced user of bioinformatics, from ecologists to population geneticists or cell biologists, deal with bioinformatics matters. As stated above, bioinformatics plays an increasingly central role in biological research, medicine, and other areas of human life. Then, we need sophisticated and user-friendly bioinformatics resources to accompany it. The

ability to access and study molecular sequence data should not be reserved for those with exceptional computer skills but be made available to all scientists, medical professionals, and the general population.[71]

CONCLUSION

Bioinformatics is a relatively young field compared to other areas. It is rapidly growing and has a great impact in biomedicine. Despite its limitations, it has led to great progress over the years and has demonstrated its potential use for a better understanding of molecular mechanisms of diseases. This allows us to identify biomarkers, create 3D molecular models, screen drugs, predict vaccines, and achieve personalized medicine. In brief, the rapid development of bioinformatics tools, software, genomics, transcriptomics, proteomics, and metabolomics data has greatly facilitated biomedical research. It has also enhanced the understanding of the biological meaning of DNA/RNA/protein modifications, interactions within complex organism networks, and the discovery of new ways to accurately apply this knowledge, enabling new therapeutic measures. This helps us to better face diseases from an increasingly broader perspective, safer, more effective, and more specific.

REFERENCES

1. Hao S, Lv J, Yang Q, Wang A, Li Z, Guo Y, et al. Identification of Key Genes and Circular RNAs in Human Gastric Cancer. Med Sci Monit. 2019 Apr 5;25:2488–504.
   https://doi.org/10.12659/MSM.915382
2. Liu W, Ouyang S, Zhou Z, Wang M, Wang T, Qi Y, et al. Identification of genes associated with cancer progression and prognosis in lung adenocarcinoma: Analyses based on microarray from Oncomine and The Cancer Genome Atlas databases. Mol Genet Genomic Med [Internet]. 2019 Feb [cited 2021 Oct 9];7(2).
   https://onlinelibrary.wiley.com/doi/10.1002/mgg3.528
3. Bayat A. Science, medicine, and the future: Bioinformatics. BMJ. 2002 Apr 27;324(7344):1018–22.
   https://doi.org/10.1136/bmj.324.7344.1018
4. Akalın PK. Introduction to bioinformatics. Mol Nutr Food Res. 2006 Jul;50(7):610–9.
   https://doi.org/10.1002/mnfr.200500273
5. Xu J, McPartlon M, Li J. Improved protein structure prediction by deep learning irrespective of co-evolution information. Nat Mach Intell. 2021 Jul;3(7):601–9.
   https://doi.org/10.1038/s42256-021-00348-5
6. Padilla-Rojas C, Jimenez-Vasquez V, Hurtado V, Mestanza O, Molina IS, Barcena L, et al. Genomic analysis reveals a rapid spread and predominance of lambda (C.37) SARS-COV-2 lineage in Peru despite circulation of variants of concern. J Med Virol. 2021 Dec;93(12):6845–9.
   https://doi.org/10.1002/jmv.27261
7. Branco I, Choupina A. Bioinformatics: new tools and applications in life science and personalized medicine. Appl Microbiol Biotechnol. 2021 Feb;105(3):937–51.
   https://doi.org/10.1007/s00253-020-11056-2
8. Li Y, Tenchov R, Smoot J, Liu C, Watkins S, Zhou Q. A Comprehensive Review of the Global Efforts on COVID-19 Vaccine Development. ACS Cent Sci. 2021 Apr 28;7(4):512–33.
   https://doi.org/10.1021/acscentsci.1c00120
9. NIH N. Comparative Genomics [Internet]. Genome.gov. 2007.
   https://www.genome.gov/11006946/comparative-genomics
10. de Crécy-Lagard V, Hanson AD. Comparative Genomics. In: Reference Module in Biomedical Sciences [Internet]. Elsevier; 2018 [cited 2021 Oct 8]. p. B9780128012383660000.
    https://linkinghub.elsevier.com/retrieve/pii/B9780128012383660956
11. Wei L, Liu Y, Dubchak I, Shon J, Park J. Comparative genomics approaches to study organism similarities and differences. J Biomed Inform. 2002 Apr;35(2):142–50.
    https://doi.org/10.1016/S1532-0464(02)00506-3
12. Setubal JC, Almeida NF, Wattam AR. Comparative Genomics for Prokaryotes. In: Setubal JC, Stoye J, Stadler PF, editors. Comparative Genomics [Internet]. New York, NY: Springer New York; 2018 [cited 2021 Oct 9]. p. 55–78. (Methods in Molecular Biology; vol. 1704).
    http://link.springer.com/10.1007/978-1-4939-7463-4_3
13. Hunt SY, Barnaby NG, Budowle B, Morse S. Forensic Microbiology. In: Encyclopedia of Microbiology [Internet]. Elsevier; 2009 [cited 2021 Oct 13]. p. 22–34.
    https://linkinghub.elsevier.com/retrieve/pii/B9780123739445002911
14. Díaz-Camacho SP, Parra-Unda JR, Ríos-Sicairos J, Delgado-Vargas F. Molecular Identification of the Etiological Agent of Human Gnathostomiasis in an Endemic Area of Mexico. Jpn J Infect Dis. 2020;73(1):44–50.
    https://doi.org/10.7883/yoken.JJID.2019.180
15. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol. 2020 Nov;5(11):1408–17.
    https://doi.org/10.1038/s41564-020-0771-4
16. López-Durán PA, Fonseca-Coronado S, Lozano-Trenado LM, Araujo-Betanzos S, Rugerio-Trujillo DA, Vaughan G, et al. Nosocomial transmission of extensively drug resistant

https://doi.org/10.36105/psrua.2022v2n3.04

Acinetobacter baumannii strains in a tertiary level hospital. Ozer EA, editor. PLOS ONE. 2020 Apr 17;15(4):e0231829. https://doi.org/10.1371/journal.pone.0231829

17. Mizutani S, Yamada T, Yachida S. Significance of the gut microbiome in multistep colorectal carcinogenesis. Cancer Sci. 2020 Mar;111(3):766–73. https://doi.org/10.1111/cas.14298
18. De Angelis M, Ferrocino I, Calabrese FM, De Filippis F, Cavallo N, Siragusa S, et al. Diet influences the functions of the human intestinal microbiome. Sci Rep. 2020 Dec;10(1):4247. https://doi.org/10.1038/s41598-020-61192-y
19. Cao S, Zhang W, Ding W, Wang M, Fan S, Yang B, et al. Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. Microbiome. 2020 Dec;8(1):47. https://doi.org/10.1186/s40168-020-00826-9
20. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001 Mar;69(3):89–95. https://doi.org/10.1067/mcp.2001.113989
21. Capecchi R, Puxeddu I, Pratesi F, Migliorini P. New biomarkers in SLE: from bench to bedside. Rheumatology. 2020 Dec 5;59(Supplement_5):v12–8. https://doi.org/10.1093/rheumatology/keaa484
22. Azad RK, Shulaev V. Metabolomics technology and bioinformatics for precision medicine. Brief Bioinform. 2019 Nov 27;20(6):1957–71. https://doi.org/10.1093/bib/bbx170
23. Telenti A. Integrating metabolomics with genomics. Pharmacogenomics. 2018 Dec;19(18):1377–81. https://doi.org/10.2217/pgs-2018-0155
24. Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R. Review of precision cancer medicine: Evolution of the treatment paradigm. Cancer Treat Rev. 2020 Jun;86:102019. https://doi.org/10.1016/j.ctrv.2020.102019
25. Grau J, Ben-Gal I, Posch S, Grosse I. VOMBAT: prediction of transcription factor binding sites using variable order Bayesian trees. Nucleic Acids Res. 2006 Jul 1;34(Web Server):W529–33. https://doi.org/10.1093/nar/gkl212
26. Toro-Domínguez D, Martorell-Marugán J, López-Domínguez R, García-Moreno A, González-Rumayor V, Alarcón-Riquelme ME, et al. ImaGEO: integrative gene expression meta-analysis from GEO database. Kelso J, editor. Bioinformatics. 2019 Mar 1;35(5):880–2. https://doi.org/10.1093/bioinformatics/bty721
27. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016 Jul;17(7):451–9. https://doi.org/10.1038/nrm.2016.25
28. Rinschen MM, Ivanisevic J, Giera M, Siuzdak G. Identification of bioactive metabolites using activity metabolomics. Nat Rev Mol Cell Biol. 2019 Jun;20(6):353–67. https://doi.org/10.1038/s41580-019-0108-4
29. Maharjan M, Tanvir RB, Chowdhury K, Duan W, Mondal AM. Computational identification of biomarker genes for lung cancer considering treatment and non-treatment studies. BMC Bioinformatics. 2020 Dec;21(S9):218. https://doi.org/10.1186/s12859-020-3524-8
30. Song X, Du R, Gui H, Zhou M, Zhong W, Mao C, et al. Identification of potential hub genes related to the progression and prognosis of hepatocellular carcinoma through integrated bioinformatics analysis. Oncol Rep [Internet]. 2019 Nov 6 [cited 2021 Oct 13]. http://www.spandidos-publications.com/10.3892/or.2019.7400
31. Wan Y, Zhang X, Leng H, Yin W, Zeng W, Zhang C. Identifying hub genes of papillary thyroid carcinoma in the TCGA and GEO database using bioinformatics analysis. PeerJ. 2020 Jul 9;8:e9120. https://doi.org/10.7717/peerj.9120
32. Gururajan A, Clarke G, Dinan TG, Cryan JF. Molecular biomarkers of depression. Neurosci Biobehav Rev. 2016 May;64:101–33. https://doi.org/10.1016/j.neubiorev.2016.02.011
33. Li W, Luo C, Xie X, Xiao Y, Zhao F, Cai J, et al. Identification of key genes and pathways in syphilis combined with diabetes: a bioinformatics study. AMB Express. 2020 Dec;10(1):83. https://doi.org/10.1186/s13568-020-01009-3
34. Chen P, Chen Y, Wu W, Chen L, Yang X, Zhang S. Identification and validation of four hub genes involved in the plaque deterioration of atherosclerosis. Aging. 2019 Aug 26;11(16):6469–89. https://doi.org/10.18632/aging.102200
35. Liu Y, Gu H-Y, Zhu J, Niu Y-M, Zhang C, Guo G-L. Identification of Hub Genes and Key Pathways Associated With Bipolar Disorder Based on Weighted Gene Co-expression Network Analysis. Front Physiol. 2019 Aug 20;10:1081. https://doi.org/10.3389/fphys.2019.01081
36. Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 2016 Dec 12;12:2694–718. https://doi.org/10.3762/bjoc.12.267
37. Silverman RB, Holladay MW. The organic chemistry of drug design and drug action. Third edition. Amsterdam ; Boston: Elsevier/AP, Academic Press, is an imprint of Elsevier; 2014. 517 p. https://doi.org/10.1016/C2009-0-64537-2

38. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chem Biol Interact. 2008 Jan;171(2):165–76.
https://doi.org/10.1016/j.cbi.2006.12.006

39. Grinter S, Zou X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. Molecules. 2014 Jul 11;19(7):10150–76.
https://doi.org/10.3390/molecules190710150

40. Acharya C, Coop A, E. Polli J, D. MacKerell A. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. Curr Comput Aided-Drug Des. 2011 Mar 1;7(1):10–22.
https://doi.org/10.2174/157340911793743547

41. Elfiky AA. SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: an *in silico* perspective. J Biomol Struct Dyn. 2020 May 6;1–9.
https://doi.org/10.1080/07391102.2020.1761882

42. Arya H, Bhatt TK. Role of Bioinformatics in Subunit Vaccine Design. In: Molecular Docking for Computer-Aided Drug Design [Internet]. Elsevier; 2021 [cited 2021 Oct 9]. p. 425–39.
https://doi.org/10.1016/B978-0-12-822312-3.00013-8

43. Del Tordello E, Rappuoli R, Delany I. Reverse Vaccinology. In: Human Vaccines [Internet]. Elsevier; 2017 [cited 2021 Oct 13]. p. 65–86.
https://doi.org/10.1016/B978-0-12-802302-0.00002-9

44. Chukwudozie OS, Duru VC, Ndiribe CC, Aborode AT, Oyebanji VO, Emikpe BO. The Relevance of Bioinformatics Applications in the Discovery of Vaccine Candidates and Potential Drugs for COVID-19 Treatment. Bioinforma Biol Insights. 2021 Jan;15.
https://doi.org/10.1177/11779322211002168

45. Adu-Bobie J. Two years into reverse vaccinology. Vaccine. 2003 Jan 30;21(7–8):605–10.
https://doi.org/10.1016/S0264-410X(02)00566-2

46. Rappuoli R. Reverse vaccinology. Curr Opin Microbiol. 2000 Oct;3(5):445–50.
https://doi.org/10.1016/s1369-5274(00)00119-3

47. Seib KL, Zhao X, Rappuoli R. Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clin Microbiol Infect. 2012 Oct;18:109–16.
https://doi.org/10.1111/j.1469-0691.2012.03939.x

48. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Dec;8(1):4.
https://doi.org/10.1186/1471-2105-8-4

49. Dormitzer PR, Ulmer JB, Rappuoli R. Structure-based antigen design: a strategy for next generation vaccines. Trends Biotechnol. 2008 Dec;26(12):659–67.
https://doi.org/10.1016/j.tibtech.2008.08.002

50. María RR, Arturo CJ, Alicia JA, Paulina MG, Gerardo AO. The Impact of Bioinformatics on Vaccine Design and Development. In: Afrin F, Hemeg H, Ozbak H, editors. Vaccines [Internet]. InTech; 2017 [cited 2021 Oct 13].
https://doi.org/10.5772/INTECHOPEN.69273

51. Ishack S, Lipner SR. Bioinformatics and immunoinformatics to support COVID-19 vaccine development. J Med Virol. 2021 Sep;93(9):5209–11.
https://doi.org/10.1002/jmv.27017

52. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. Cell Host Microbe. 2020 Apr;27(4):671-680.e2.
https://doi.org/10.1016/j.chom.2020.03.002

53. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, et al. Design of multi epitope-based peptide vaccine against E protein of human COVID-19: An immunoinformatics approach [Internet]. Bioinformatics; 2020 Feb [cited 2021 Oct 13].
https://doi.org/10.1155/2020/2683286

54. Naithani N, Sinha S, Misra P, Vasudevan B, Sahu R. Precision medicine: Concept and tools. Med J Armed Forces India. 2021 Jul;77(3):249–57.
https://doi.org/10.1016/j.mjafi.2021.06.021

55. Carrasco-Ramiro F, Peiró-Pastor R, Aguado B. Human genomics projects and precision medicine. Gene Ther. 2017 Sep;24(9):551–61.
https://doi.org/10.1038/gt.2017.77

56. Sunil Krishnan G, Joshi A, Kaushik V. Bioinformatics in Personalized Medicine. In: Singh V, Kumar A, editors. Advances in Bioinformatics [Internet]. Singapore: Springer Singapore; 2021 [cited 2021 Oct 13]. p. 303–15.
https://doi.org/10.1007/978-981-33-6191-1_15

57. Liu L, Song B, Ma J, Song Y, Zhang S-Y, Tang Y, et al. Bioinformatics approaches for deciphering the epitranscriptome: Recent progress and emerging topics. Comput Struct Biotechnol J. 2020;18:1587–604.
https://doi.org/10.1016/j.csbj.2020.06.010

58. Chen L, Lu D, Sun K, Xu Y, Hu P, Li X, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. Gene. 2019 Apr;692:119–25.
https://doi.org/10.1016/j.gene.2019.01.001

59. Bhuvaneshwar K, Belouali A, Singh V, Johnson RM, Song L, Alaoui A, et al. G-DOC Plus – an integrative bioinformatics platform for precision medicine. BMC Bioinformatics. 2016 Dec;17(1):193
https://doi.org/10.1186/s12859-016-1010-0

60. Jünemann S, Kleinbölting N, Jaenicke S, Henke C, Hassa J, Nelkner J, et al. Bioinformatics for NGS-based metagenomics and the application to biogas research. J Biotechnol. 2017 Nov;261:10–23.

https://doi.org/10.1016/j.jbiotec.2017.08.012

61. Muthiah I, Rajendran K, Dhanaraj P, Vallinayagam S. In silico structure prediction, molecular docking and dynamic simulation studies on G Protein-Coupled Receptor 116: a novel insight into breast cancer therapy. J Biomol Struct Dyn. 2021 Sep 2;39(13):4807–15.

https://doi.org/10.1080/07391102.2020.1783365

62. Purkayastha A, Ditty SE, Su J, McGraw J, Hadfield TL, Tibbetts C, et al. Genomic and Bioinformatics Analysis of HAdV-4, a Human Adenovirus Causing Acute Respiratory Disease: Implications for Gene Therapy and Vaccine Vector Development. J Virol. 2005 Feb 15;79(4):2559–72.

https://doi.org/10.1128/JVI.79.4.2559-2572.2005

63. Watson M, Warr A. Errors in long-read assemblies can critically affect protein prediction. Nat Biotechnol. 2019 Feb;37(2):124–6.

https://doi.org/10.1038/s41587-018-0004-z

64. Wang A, Au KF. Performance difference of graph-based and alignment-based hybrid error correction methods for error-prone long reads. Genome Biol. 2020 Dec;21(1):14.

https://doi.org/10.1186/s13059-019-1885-y

65. McGuire AL, Oliver JM, Slashinski MJ, Graves JL, Wang T, Kelly PA, et al. To share or not to share: A randomized trial of consent for data sharing in genome research: Genet Med. 2011 Nov;13(11):948–55.

https://doi.org/10.1097/gim.0b013e3182227589

66. Oliver JM, Slashinski MJ, Wang T, Kelly PA, Hilsenbeck SG, McGuire AL. Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives. Public Health Genomics. 2012;15(2):106–14.

https://doi.org/10.1159/000334718

67. Bejleri J, Jirström E, Donovan P, Williams DJ, Pfeiffer S. Diagnostic and Prognostic Circulating MicroRNA in Acute Stroke: A Systematic and Bioinformatic Analysis of Current Evidence. J Stroke. 2021 May 31;23(2):162–82.

https://doi.org/10.5853/jos.2020.05085

68. Ghadamyari F, Heidari MM, Zeinali S, Khatami M, Merat S, Bagherian H, et al. Mutational screening through comprehensive bioinformatics analysis to detect novel germline mutations in the APC gene in patients with familial adenomatous polyposis (FAP). J Clin Lab Anal [Internet]. 2021 May [cited 2021 Nov 17];35(5).

https://doi.org/10.1002/jcla.23768

69. Sufyan M, Ali Ashfaq U, Ahmad S, Noor F, Hamzah Saleem M, Farhan Aslam M, et al. Identifying key genes and screening therapeutic agents associated with diabetes mellitus and HCV-related hepatocellular carcinoma by bioinformatics analysis. Saudi J Biol Sci. 2021 Oct;28(10):5518–25.

https://doi.org/10.1016/j.sjbs.2021.07.068

70. Dopazo J, Maya-Miles D, García F, Lorusso N, Calleja MÁ, Pareja MJ, et al. Implementing Personalized Medicine in COVID-19 in Andalusia: An Opportunity to Transform the Healthcare System. J Pers Med. 2021 May 26;11(6):475

https://doi.org/10.3390/jpm11060475

71. Smith DR. The battle for user-friendly bioinformatics. Front Genet [Internet]. 2013 [cited 2021 Nov 17];4.

https://doi.org/10.3389/fgene.2013.00187