INTERNATIONAL JOURNAL OF SCIENTIFIC INFORMATION

www.jsiinternational.com

Review Article

A review of bioinformatics applications in cancer research

Samos*
Department of Zoology
Rajasthan University, Rajasthan
*Corresponding Author
Email: pkaswan007@gmail.com

Article Received on: 25/03/23 Revised on: 03/04/23 Approved for publication: 6/04/23

ABSTRACT

The study of epigenetic changes, particularly the acetylation, and deacetylation of histones, is becoming a more crucial topic in the fight against cancer. The ultimate objective of cancer research is the organizing and analysis of the ever-growing data as well as the development of novel treatment or diagnostic approaches. The different effective bioinformatics approaches for cancer research are reviewed in this article. The fundamental ideas and tenets of the significant and widely employed bioinformatics methods are introduced. A list of software and databases that are accessible is also given. Finally, obstacles and future directions for the creation and use of diverse methodologies are highlighted. We propose that assessing such technologies could be one of the crucial phases in the future creation of effective cancer therapy options.

Keywords: Bioinformatics, Cancer Research, Cheminformatics, Epigenetics, Interactome, and SNP

INTRODUCTION

One group of disorders called cancer is mostly brought on by epigenetic changes. Gene alterations that impact gene activity and expression are known as epigenetic modifications. It has taken a lot of work to identify the epigenetic mechanism. Among these, histone acetylation and deacetylation are crucial in the emergence of cancer. Histone deacetylases (HATs) and histone deacetylases (HDACs), which are the crucial enzymes and regulators in this process, have been the subject of extensive

research. A significant amount of data is being produced by an expanding variety of biological tests and laboratory procedures since epigenetics is a characteristic of cancer. Therefore, big data analysis techniques are applied to research and development in cancer research to provide more clarity and improved decision-making. One such an approach that has been investigated for a while and has success stories is bioinformatics. This is a multidisciplinary field that combines information technology (IT) and all parts of the life sciences to address issues in the biological sciences. The development of bioinformatics as a platform for cancer research has made significant progress to date. Omics is the most significant of the several bioinformatics approaches, including genomes, proteomics, transcriptomics, metabolomics, etc. In this way, we aimed to clarify the function of bioinformatics and some of its possible uses in cancer research.

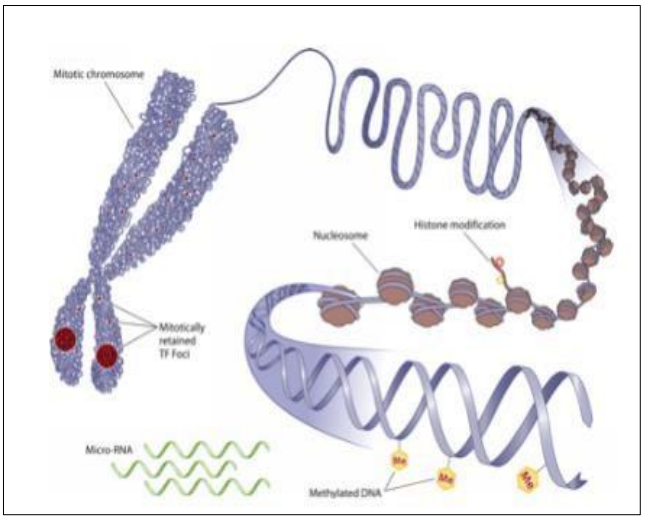

Figure 1: Mechanisms for epigenetic modification. Gene expression is regulated by DNA methylation, which is carried out by DNA methyltransferase enzymes, histone modification and acetylation in lysine residues, and inactivation [1]

Methodology

Starting with simple search terms like "bioinformatics," "cancer research," and "cancer bioinformatics," we used PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and other literature searches, such as google scholar, to find publications on our search criteria. Each page's documents were examined and arranged according to relevance, year, content, and importance. The duplicates were identified manually, and those were taken out of the final list. The structure of our paper is as follows: We start by giving an overview of some key bioinformatics techniques, then we look at their significance in cancer research and discuss how to use them. Thereafter, a conclusion follows.

Genome-wide association studies (GWAS)

Single nucleotide polymorphisms, or SNPs, are tiny variations that are more common in cancer, and this field helps to investigate the genetic architecture of disease [2]. SNP studies have been reported to provide information on mutations, which helps researchers to identify deletions, insertions, and copy number variations [3, 4], in addition to the inherent genetic aberrations. GWAS eventually sheds insight into the therapeutic significance of these processes in the development of cancer.

Phylogenetics analysis

Phylogenetic studies serve as the foundation for comparative genomics, which compares genes from different animal species to understand the mechanism and evolution of disease genes [5]. To understand the function of genes and their origin, phylogenetic analysis is useful. Recent years have seen significant progress in the creation of evolutionary studies for cancer research [6, 7].

Interactome and pathway studies

The network of genes and proteins in a cell is called the interactome [8]. Studies of protein-protein interactions (PPIs) elucidate the disease's underlying molecular mechanisms and pinpoint its true origin. This enables the creation of hub genes for the gene involved in the development of cancer. Because of all these factors, it is necessary to interpret networks clearly to determine their clinical value and networking. Researchers can find multiple genes that were differently expressed in clinical samples and assist in determining matching biomarkers by performing Interactome and pathway analysis [9].

Structural Bioinformatics

The prediction and analysis of the structures of biological molecules like DNA, RNA, and proteins are known as structural bioinformatics. By comparing database structures and validating them, this method allows the function of molecules to be deduced from their sequence or structural details [10]. As a result, homology modeling has gained widespread acceptance as a

technique for clarifying the theoretical model of molecules, particularly proteins. The use of this for receptor-drug interaction studies would enable further validation.

Cheminformatics and drug discovery

It can be difficult to design and manufacture effective therapeutic components for cancer. Cheminformatics is a contemporary field that analyses chemical compounds' complex structures to determine which ones have the potential to be therapeutic molecules. These methods for drug discovery support the development of possible cancer treatments by pharmaceutical corporations and medical researchers [11]. However, compounds discovered through docking, dynamics, and quantitative structure-activity relationship (QSAR) studies need additional validation through adsorption, distribution, metabolism, and excretion (ADME), Lipinski's rule analysis, and additional wet-lab biological testing, where only actual trials are conducted. According to QSAR, drug development is a key area [12]. For clinical testing and drug validation, ADME testing is essential [13], and Lipinski's rule determines if a molecule is orally active or not using a set of rules [14].

Table 1: Overview of the important bioinformatics methods and list of available software/databases

S.No	Method	Softwares /databases
1	GWAS	PLINK/GPLINK[15], METAL[16], GWAMA[17], MANTRA[18].
2	Phylogenetic analysis	Clustalw/X[19], Phylip[20], MEGA[21], BEAST[22], PAUP[23].
3	Interactome and pathway studies	IntAct[24], PANTHER[25], KEGG[26], STRING[27], BioGrid[28].
4	Structural Bioinformatics	SWISS-MODEL[29], Phyre2[30], PDB[31], Modeller[32],
5	Cheminformatics and drug discovery	Schrodinger[33], BioVia DS visualize (Biovia, 2016), Patchdock[34],

Discussion

The findings of this study describe the function of bioinformatics in the study of cancer. In addition to what we have described, studies reveal that new fields like systems biology and precision medicine are developing domains that have a significant impact on the advancement of cancer research. A big dataset and gene expression data might be produced using new sequencing methods, which could save time and money while improving the outcomes of future studies.

Conclusion

It has been established that bioinformatics is a crucial area of study for cancer research. A huge number of chemical compounds can be screened using developing bioinformatics approaches like machine learning, which allows for the discovery of innovative and targeted therapeutic components for the treatment of cancer targets. The advancement of these tools for cheminformatics-based drug discovery is a significant step toward the quicker development of future therapeutic candidates. Our findings in this review may help people better comprehend the various bioinformatics specialties and suggest the most effective approaches for cancer research. However, there is still a need to look into further applications.

References

1. Zaidi SK, Young DW, Montecito M, Lain JB, Stein JL, van Wijnen AJ, et al. Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. Mol Cell Biol. 2010; 30:4758–66.

2. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005; 6:95–108.

3. Easton DF, Eeles RA. Genome-wide association studies in cancer. Hum Mol Genet. 2008; 17:R109–15.

4. He Q, He Q, Liu X, Wei Y, Shen S, Hu X, et al. Genome-wide prediction of cancer driver genes based on SNP and cancer SNV data. Am J Cancer Res. 2014; 4:394–410.

5. Lake JA, Moore JE. Phylogenetic analysis and comparative genomics. Trends Biotechnol. 1998; 16:22–3.

6. Brown D, Smeets D, Szekely B, Larsimont D, Szasz AM, Adnet P-Y, et al. Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. Nat Commun. England; 2017; 8:14944.

7. Somarelli JA, Ware KE, Kostadinov R, Robinson JM, Amri H, Abu-Asab M, et al. PhyloOncology: Understanding cancer through phylogenetic analysis. Biochim Biophys Acta. 2017; 1867:101–8.

8. Vidal M, Cusick ME, Barabási A-L. Interactome Networks and Human Disease. Cell. 2011; 144:986–98.

9. Coulombe B. Mapping the Disease Protein Interactome: Toward a Molecular Medicine GPS to Accelerate Drug and Biomarker Discovery. J Proteome Res. 2011; 10:120–5.

10. Chandra N, An, and P, Yeturu K. Structural bioinformatics: deriving biological insights from protein structures. Interdiscip Sci. Germany; 2010;2:347–66.

11. Begam BF, Kumar JS. A Study on Cheminformatics and its Applications on Modern Drug Discovery. Procedia Eng. 2012;38:1264–75.

12. Gini G. QSAR Methods. Methods Mol. Biol. 2016. p. 1–20.

13. Vugmeyster Y, Harrold J, Xu X. Absorption, Distribution, Metabolism, and Excretion (ADME) Studies of Biotherapeutics for Autoimmune and Inflammatory Conditions. AAPS J. Boston; 2012; 14:714–27.

14. Pollastri MP. Overview of the Rule of Five. CurrProtocPharmacol. 2010; Chapter 9: Unit 9.12.

15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and PopulationBased Linkage Analyses. Am J Hum Genet. 2007; 81:559–75.

16. Willer CJ, Li Y, Abecasis GR. METAL: a fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1.

17. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics. 2010; 11:288.

18. Morris AP. Transethnic meta-analysis of genome wide association studies. Genet Epidemiol. 2011; 35:809–22.

19. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–80.

20. Retief JD. Phylogenetic analysis using PHYLIP. Methods Mol Biol. 2000; 132:243–58.

21. Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bio informs. 2008; 9:299–306.

22. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214.

23. Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP. CurrProtocBioinforma. 2003; Chapter 6: Unit 6.4.

24. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, et al. IntAct—open source resource for molecular interaction data. Nucleic Acids Res. 2007; 35:D561–5.

25. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013; 8:1551.

26. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000; 28:27–30.

27. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 2017; 45:D362–8.

28. Chatr-aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, et al. The BioGRID interaction database: 2017 update. Nucleic Acids Res. 2017; 45:D369–79.

29. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014; 42:W252-8.

30. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modelling, prediction, and analysis. Nat Protoc. 2015; 10:845–58.

31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–42.

32. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using Modeller. CurrProtocBioinforma. 2006; Unit-5.6.

33. Bhachoo J, Beuming T. Investigating Protein-Peptide Interactions Using the Schrodinger Computational Suite. Methods Mol Biol. 2017; 1561:235–54.

34. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 2005; 33:W363-7.

Cite this article as: *Samos. A review of bioinformatics applications in cancer research. Int. J. Sci. Info. 2023; 1(1):1-7*

Source of support: Nil, Conflict of interest: None Declared