

Applications of Bioinformatics in Cancer Research- A review

Sabeena M^{1 2}, Kaiser Jamil^{1*}, A.V.N Swamy³

¹ Center for Biotechnology and Bioinformatics, School of Life Sciences, Jawaharlal Nehru Institute of Advanced Studies (JNIAS), Secunderabad-500003, Telangana, India

^{1*}Head of Genetics Department, Bhagwan Mahavir medical Research Centre, 10-1-1, Mahavir Marg, Hyderabad-500004, Telangana, India.

² Department of Biotechnology, Jawaharlal Nehru Technological University Anantapur (JNTUA), Ananthapuramu-515002, Andhra Pradesh, India

³Department of Chemical Engineering, Jawaharlal Nehru Technological University Anantapur (JNTUA), Ananthapuramu-515002, Andhra Pradesh, India

***Corresponding author**

Received: May 13, 2018

Accepted: June 20, 2018

ABSTRACT

Epigenetic modifications, and in particular the acetylation and deacetylation of histones are major contributors to carcinogenesis and form an increasingly important area in cancer research. The ultimate goal of cancer research lies in the analysis and organization of the increasing amount of data and also the discovery of novel therapeutic or diagnostic options. This article reviews the various efficient bioinformatics methods for cancer research. The main concepts and principles for the important and commonly used bioinformatics methods are introduced. In addition, a list of available software and databases are provided. Finally, future directions and challenges for the development and applications of various techniques are discussed. We suggest that, the evaluation of such technologies could form the important steps in the future development of successful cancer therapeutic options.

Keywords: Cancer research; epigenetics; bioinformatics; interactome; SNP; cheminformatics.

Introduction

Cancer is a class of diseases that occurs mainly by epigenetic modifications. Epigenetics modifications are the changes which occurs in the genes and affects the activity and expression. Much effort has been made to determine the epigenetic mechanism. Among this, histone acetylation and deacetylation play a major role in cancer development. Extensive research has been conducted in histone deacetylases (HATs) and histone deacetylases (HDACs) which are the essential enzymes and regulators in this process. Since, epigenetics is a hallmark of cancer, large amount of data is being generated by an increasing number of biological tests and laboratory procedures. Thus, big data analysis strategies are used to provide more clarity and a better decision for research and development in cancer research. Bioinformatics is one such method which has been studied for several years and reports success stories. This is a multidisciplinary subject which encompasses all the aspects of life sciences and information technology (IT) to solve problems in biological sciences. To date, much progress is there in the development of bioinformatics as a platform for cancer research. Across several methodologies designed and developed for bioinformatics, omics is the most important such as genomics, proteomics, transcriptomics, metabolomics etc. In this review, we sought to elucidate the role of bioinformatics and its potential applications in cancer research.

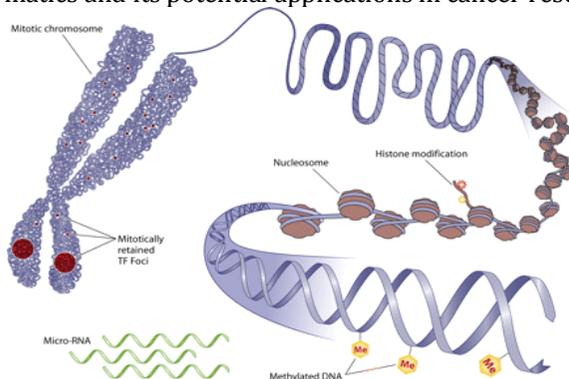


Figure 1: Epigenetic modification mechanisms. DNA methylation mediated by DNA methyltransferase enzymes, Histone modification and acetylation in lysine residues of histone and miRNA begins activation to regulate gene expression [1].

Methodology

We used PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>) and other literature searches such as google scholar to identify publications on our search criteria, starting with basic search queries such as 'bioinformatics', 'cancer research', 'cancer bioinformatics' etc.to generate adequate number of research papers. We analyzed all the obtained papers on each page and sorted by relevance, year, content and importance. A manual comparison was done to find the duplication and these were excluded from the final list.Our paper is organized as follows: first we provide an overview of some of the important bioinformatics methods which is followed by examining its importance in cancer research and discussed the application of these methods. This is followed by a conclusion.

Genome-wide association studies (GWAS)

This field helps to investigate the genetic architecture of disease and identifies single and minute variations called single nucleotide polymorphisms or SNPs that occur more frequently in cancer [2]. In addition to the inherent genetic aberrations, SNP studies have been reported to provide information of mutations which helps the researchers to identify deletions, insertions and copy number variations [3,4]. GWAS eventually highlights the mechanisms in the cancer progression and its clinical implications.

Phylogenetics analysis

Phylogenetic studies are basis for comparative genomics which involves the comparative studies of genes across the animal kingdom to underpin the mechanism, and evolution of disease genes [5]. Thus Phylogenetic analysis are helpful in determining the role of genes and their origin. Significant advances have been made in last few years towards the development of evolutionary studies for cancer research [6,7].

Interactome and pathway studies

Interactome is the network of genes/proteins in the cell [8]. Protein-protein interaction (PPI) studies reveal the molecular pathways associated with the disease and identify its real cause. This allows the generation of hub genes for the gene taking part in cancer progression. All these factors mean that, a clear interpretation of network is required to uncover their clinical value and their networking. By executing Interactome and pathway analysis, researchers can identify numerous genes that were differentially expressed in clinical samples and help to determine corresponding biomarkers [9].

Structural bioinformatics

Structural bioinformatics is the prediction and analysis of the structure of biological molecules such as DNA, RNA and proteins. In this method, the function of molecules can be inferred from their sequence or structural details by comparing database structures and validates them [10]. Therefore, homology modeling has become a widely accepted method for elucidation of theoretical model of molecules especially proteins. This would allow further validation and use for receptor-drug interaction studies.

Cheminformatics and drug discovery

Design and development of novel and efficient drug components for cancer is a challenging process. Cheminformatics is a modern field which uses the complex structure of chemical compounds and identify its potential as drug compounds. These drug discovery techniques which helps pharma companies and medical researchers to make potential medicines for cancer research [11]. However, compounds identified through quantitative structure activity relationship (QSAR), docking and dynamics studies require further validation such as adsorption, distribution, metabolization and excretion (ADME), lipinski's rule analysis, including further wetlab and biological testing, where only real trials are made. QSAR appeared to be as important area in drug development [12]. ADME testing is crucial for clinical testing and validation of drugs [13] and lipinski's rule validates whether a molecule is orally active or not based on certain rules [14]

Table 1: Overview of the important bioinformatics methods and list of available softwares/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 (<i>Biovia</i> , 2016), Patchdock [34].

Discussion

The results obtained in this study describe the role of bioinformatics in cancer research. In addition to the description we have provided, studies show that new areas such as systems biology, precision medicine that serves as emerging fields which contribute effectively in speeding up the cancer research. Emerging sequencing technologies offer the possibility of producing large datasets and gene expression data which can help save time and revenue by producing better results for future researches.

Conclusion

Bioinformatics has been shown as an important field of study for cancer research. In addition, emerging bioinformatics methodologies such as machine learning offers the possibility of screening large number of chemical compounds in which novel and specific drug components could be identified for cancer targets. The evolution of such technologies for cheminformatics-based drug discovery is an important step in the future development of potential drug compounds at a faster pace. Our findings in this review may enhance the understanding of the different branches of bioinformatics and suggest the best methodology according to the need of research in cancer. But still there is a need to explore for more applications.

References

- Zaidi SK, Young DW, Montecino M, Lian JB, Stein JL, van Wijnen AJ, et al. Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. *Mol Cell Biol.* 2010;30:4758–66.
- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6:95–108.
- Easton DF, Eeles RA. Genome-wide association studies in cancer. *Hum Mol Genet.* 2008;17:R109–15.
- 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.
- Lake JA, Moore JE. Phylogenetic analysis and comparative genomics. *Trends Biotechnol.* 1998;16:22–3.
- Brown D, Smeets D, Szekeley 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.
- 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.
- Vidal M, Cusick ME, Barabási A-L. Interactome Networks and Human Disease. *Cell.* 2011;144:986–98.
- 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.
- Chandra N, Anand P, Yeturu K. Structural bioinformatics: deriving biological insights from protein structures. *Interdiscip Sci. Germany;* 2010;2:347–66.
- Begam BF, Kumar JS. A Study on Cheminformatics and its Applications on Modern Drug Discovery. *Procedia Eng.* 2012;38:1264–75.
- Gini G. QSAR Methods. *Methods Mol. Biol.* 2016. p. 1–20.
- 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.
- Pollastri MP. Overview on the Rule of Five. *Curr Protoc Pharmacol.* 2010;Chapter 9:Unit 9.12.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* 2007;81:559–75.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26:2190–1.
- Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010;11:288.
- Morris AP. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol.* 2011;35:809–22.
- 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.
- Retief JD. Phylogenetic analysis using PHYLIP. *Methods Mol Biol.* 2000;132:243–58.
- Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* 2008;9:299–306.
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:214.
- Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP. *Curr Protoc Bioinforma.* 2003;Chapter 6:Unit 6.4.
- 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.
- 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. *Curr Protoc Bioinforma.* 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.