International Journal of Theoretical & Applied Sciences, 14(2): 52-56(2022)

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

An Overview of Cancer Research using Bioinformatics

Piyusha Sharma*
Assistant Professor, Desh Bhagat University, Mandi Gobindgarh (Punjab), India.

(Corresponding author: Piyusha Sharma)*

(Received 11 August 2022; Accepted 26 October 2022)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: A significant amount of biological data and information have been made available to the scientific community by the Human Genome Project. This scenario shows the myriad arenas in which bioinformatics have found utility. The use of bioinformatics in cancer study and treatment is now widespread, and it is evident that scientists, academics and researchers have rapidly and thoroughly explored the bioinformatics tools that are regarded to be essential for the therapies. The advancement of scientific research has been aided by publicly available resources. In order to exploit this new knowledge effectively, the era of big data and genomics has brought up a necessity for collaboration. Here, we provide a description of the web resources for cancer genomics research and assign ratings based on the variety of cancer kinds, sample size, comprehensiveness of the omics data, and user experience. We anticipate that this introduction will raise knowledge of these resources and make it easier for the cancer research community to use them. The resources we assessed include data repositories and analysis tools.

Keywords: Cancer, Bioinformatics, databases, web tools, genomics, research.

I. INTRODUCTION

Cancer is categorised as a genetic condition in which the cells are unable to follow the cell cycle's successive phases and divide normally in a way. That is, cells lose control over the cell cycle and begin to divide uncontrollably. As a result, the chromosomes of cancer cells will be misarranged or may be missing major portions. Numerous efforts are made to find out how to detect, diagnose, and treat such dangerous diseases as a result of big and quick advancements in medical sector study. Additionally, increased pressure to use bioinformatics in cancer therapy came from the 2003 finding of the Human Genome Project. Bioinformatics has emerged in recent years as a result of the need to comprehend the world around us better. It is a multidisciplinary field that has resulted by the combination of numerous disciplines, including biology, computing, statistics, chemistry, mathematics, and more. For the purposes of gathering, compiling, organising, analysing, and digitising biological data, each of those disciplines is playing a significant role. By using highly developed algorithms, computational and statistical techniques, the design and construction of software tools, and theoretical frameworks, it has become achievable to manage, analyse, organise and classify the vast amount of biological data. In order to find biological data and use bioinformatics in cancer research and treatment, it has been observed that experts and physicians attempt to use the numerous databases that are available and the various search engines like Google. Fully automated workflows and servers have made homology modelling simpler and more effective while also enabling users without specialised computational skills to build precise protein models and have easy access to modelling results, their display, and interpretation [25]. Utilizing the protein's basic sequence, which is readily available in databases, and previous knowledge gleaned through structural similarity with other proteins, homology modelling is used to generate protein 3D structures [10]. Studying the underlying theory and implementation methods of the database's structure, storage, design, and maintenance makes processing and analysis of the data contained within the database simple [28]. High-resolution protein 3D structures can be produced for scientific purposes by employing internet computing resources [9]. The integration of bioinformatics databases, data types, and structures is a key consideration when deciding how bioinformatics will be used in the medical profession, especially in the treatment and therapies for cancer. In a broader sense, using machine learning-driven informatics approaches, that is useful in enhancing statistical analysis of integrated histopathologic datasets has improved predictive models that can be created using cutting-edge bioinformatics methods, data from high throughput small molecule screening, and/or results information. The decision tree classifier is used for discovering important miRNA based biomarkers for breast cancer [20]. Recent study highlights LMTK3 phosphorylation by CDK5 results in breast cancer tumourogenesis [12].

The inexorably developing number of uses of Machine Learning in medicinal services permits us to look at a future where information, investigation, and advancement work connected at the hip to help incalculable patients without them consistently acknowledging it [1].

Collectively, the publications show how machine learning techniques can advance cancer biology significantly. In fact, we can eventually hope to improve research efficiency and make significant improvements to patients' overall health as we gain a better understanding of how various machine learning approaches are best suited to pursue the important questions that are described in the articles of this series [14].

II. BIOINFORMATICS' IMPACT ON CANCER RESEARCH

A potent tool for biological research is microarray technology. They enable the simultaneous study of hundreds to thousands of DNA expression sequences for genomic research and diagnostic applications, which ensures that the way gene expression is studied will be revolutionised. Microarrays are extensively used to comprehend the genetic and epigenetic makeup of cancer cells in addition to monitoring and analysing gene expression patterns [16]. For more than ten years, microarray technology has been extensively employed in cancer research, diagnosis, and tumour categorization. Its widespread adoption is a result of the shortcomings of traditional cancer gene study procedures, which are mostly time and money consuming. Due to their compact size, microarrays are substantially advancing and can be used to quickly survey a large number of genes or when the study sample is tiny. Finding differences in gene expression between healthy and cancerous cells was one of the initial uses of microarray technology. The utilisation of oligonucleotide, cDNA, and genomic chips is involved in DNA microarray analysis. Single Nucleotide Polymorphism (SNP), Mutation, and Genotyping analyses are performed on cDNA microarrays, which are typically used for gene expression analysis. Oligonucleotide microarrays are used to examine gene expression [16]. The scientific understanding of how cancer-relevant transcription factors affect gene networks and eventually cancer formation has advanced thanks to their application in deciphering the signal pathways of these transcription factors. Additionally, microarray technology enables molecular analysis of a cell's condition and may recognise a certain species of cell based on its gene expression profile. Future cancer diagnosis will greatly benefit from this because conventional techniques cannot differentiate between cancers that are morphologically identical but molecularly distinct. The clinical progression of a disease is greatly influenced by the molecular variations [16]. Despite the fact that the microarray technique for tumour categorization and diagnosis shows promise as a future diagnostic tool still, there are a number of drawbacks, one of which is the difficulty to precisely diagnose certain tumours by gene expression profile alone because no particular group of biomarkers has been developed for the diagnosis of particular malignancies. This is also related to the requirement for the development of data analysis tools and techniques. Additionally, microarray tests are expensive, the gene expression data reveal striking differences within a single tumour, and the gene expression profile does not allow for early tumour diagnosis. Despite these drawbacks, DNA microarrays are most effective for molecular classification based on biological and genetic alterations. For all facets of human disease and biomedical research, microarrays are a significant tool for the investigation of global gene expression [15]. Bioinformatics approaches are needed because large amounts of data generated by microarray technology must be analysed using computational methods. In order to comprehend and analyse data on a big scale, bioinformatics is an interdisciplinary technique that combines information technology and biological research. It involves building databases, developing software, and handling data. Aims of bioinformatics include first structuring data so that researchers may quickly access existing data as well as input data into the database. It can also be used to create tools and resources that facilitate data analysis, and then use those tools to do it in a biologically meaningful way [18]. Agriculture, medicine, forensic science, pharmaceuticals, and the biotech industry are just a few of the industries where bioinformatics is used [19]. This field of bioinformatics has evolved quickly, keeping up with the expansion of genome sequences by concentrating on medical applications, especially cancer, which is one of the major diseases that claim lives worldwide. Cancer research and diagnosis are aided by the use of bioinformatics tools and technologies, such as web technology, Cytoscape, Gene Expression Profiling Interactive Analysis (GEPIA), and databases like the Kyoto Encyclopedia of Genes and Genomes (KEGG), National Center for Biotechnology Information (NCBI), gene omnibus databases, Surveillance, Epidemiology, and End Results (SEER) database. It is used in the diagnosis of various different forms of cancer, including cervical cancer, pancreatic cancer, breast cancer, and lung cancer. Faster cancer diagnosis, detection, and prevention have been made possible by advances in bioinformatics technology; as a result, a sustainable solution has been transformed by this technology [14].

III. BIOINFORMATICS TOOLS AND DATABASES CONTAINING CANCER RESEARCH

The vast, enormous, and highly complex amount of biological data required effective and powerful storage, access, and manipulation. Thus, the development of bioinformatics databases—which can be further divided into sequence databases, microarray databases, genome databases, databases of protein structures, and many more was necessary. The most significant databanks are the Gene Bank, the EMBL DNA database, the DNA Data Bank Japan (DDNJ), and the Protein databases at SWISS-PORT (Protein sequence database). Micro array databases contain information on micro array organic phenomena under various biological circumstances. Examples of this type of database include Array Express. Genome databases compile gene (DNA) sequences from many organisms. The databases Xenbase, Corn, SEED, and RGD11 are examples of this category. Another example of a bioinformatics database created by combining cheminformatics and bioinformatics is the Drug Bank database [3].

As a result of big and quick advancements in medical field a substantial efforts are being made to figure out how to detect, recognise, and treat such life - threatening disease. Furthermore, the 2003 finding of the Human Genome Project had increased pressure on bioinformatics to be used in the treatment of cancer. The use of bioinformatics in cancer research and treatment is currently widespread, and it is obvious that professionals and researchers have conducted extensive research on the bioinformatics tools that are thought to be essential for treatment [5].

The Database for Annotation, Visualization and Integrated Discovery (DAVID) The functional bioinformatics tool employs a variety of algorithms to condense a huge number of genes with related biological terms into families that are arranged in a logical and relevant manner. These families are referred to as biological modules [13]. Surveillance, Epidemiology, and End Results Program (SEER) this initiative attempts to collect data on its diagnosis, treatment, and trends [6]. The programme tracks differences in cancer survival rates by age, ethnicity, and stage at diagnosis for each form of cancer. Gene ontology (GO) this comprehensive bioinformatics source offers details on functional genomics as a way to express biological knowledge. The website (http://www.geneontology.org) hosts this community-based initiative. Gene Expression Profiling Iterative Analysis (GEPIA), this online server enables biologists and clinicians to carry out thorough and sophisticated data mining jobs with easy clicking, allowing data mining in research fields, scientific discussions, and the development of cancer treatment options. It is a web server used to profile and examine both normal and cancerous gene expression [23]. University of Alabama Cancer Database (UALCAN) is a comprehensive, approachable, and interactive online tool for studying cancer omics data is the UALCAN database. It is a complete data-mining platform that is integrated to make the analysis of the cancer transcriptome easier. The Gene Omnibus Database (GEO) High-throughput gene expression and other functional genomics data are freely archived and distributed globally via the GEO database, a public resource. Instead of focusing solely on gene expression investigations, the GEO has expanded to cover several other data uses with the rapid advancement of technology, such as looking at chromatin structure and genome-protein interactions [4]. The Cancer Genome Atlas (TCGA) is one of the most comprehensive and effective cancer genomics initiatives is the Cancer Genome Atlas. Over 11,000 people representing over 30 different cancer types have had their genomic sequence, expression, methylation, and copy number variation data gathered, evaluated, and made available via the database programme [24]. In order to improve cancer diagnosis, bioinformatics has so emerged as a very potent and revolutionary tool. By using a variety of tools and databases, as well as the enormous volumes of data produced by microarray technology, it assists in the early cancer diagnosis of the many types of malignancies. The UCSC Cancer Genomics Browser is an online tool for hosting, viewing, and analysing data on clinical research and cancer genetics online [8]. NONCODE is a constantly updated knowledge base of ncRNAs from several species, including human and mouse, with the exception of tRNAs and rRNAs [17]. The Cancer Genome Work Bench (CGWB) focuses on gene expression, mutations, CNV, and methylation and is a useful tool [30]. The Genomics of Drug Sensitivity in Cancer (GDSC) database is the greatest publicly available database of information on medication sensitivity and response in cancer cells [28]. CanSAR is an open, cross-disciplinary knowledgebase that focuses on cancer and seeks to promote translational cancer research [2]. The cBioPortal for Cancer Genomics is a platform available to researchers for exploring, visualising, and analysing multimodal cancer genomics data [7]. SomamiR is a database created to explore into the relationship between somatic and germline mutations and miRNA function in cancer. The paediatric cancer genome project (PCGP) and the gene expression omnibus (GEO) are the sources of the mutation data, while miRBase release 18 is the source of the miRNA data. MethyCancer houses information on DNA methylation, cancer, CNV, CpG Island clones, and the relationships between these databases. The user-friendly MethyView display makes it simple to search the database [11]. The SNP500Cancer database, a part of the Cancer Genome Anatomy Project (CGAP) (http://cgap.nci.nih.gov/Tools) is a database used to verify the sequencing and genotype of single nucleotide polymorphisms (SNPs) in cancer and other complicated disorders [21]. The primary objective of the SNP500Cancer project is to re-sequence reference samples in order to identify known or novel SNPs for molecular epidemiology investigations in cancer. The Catalogue of Somatic Mutations In Cancer (COSMIC) is the largest database of somatic mutations and the impact they have on cancer in humans [19]. The European Genome-phenome Archive (EGA) is a repository for all kinds of genotyping and sequencing research. Nearly 58% of all EGA studies are related to cancer, including research from the International Cancer Genome Consortium (ICGC). The Cancer Genomics Hub (CGHub) is a central repository for the genomic data generated by three distinct projects at the National Cancer Institute (NCI) of the United States, namely the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Cancer Cell Line encyclopedia (CCLE) projects and The Cancer Genome Atlas (TCGA) [26].

CONCLUSIONS

The current review collection provides in-depth analyses of the key bioinformatics resources and databases utilised in the inquiry. Recent advances in genomics and related information technologies have accelerated the connection between scientific study and therapeutic application through the creation of a public data repository and analytical tools [29]. Scientific research has advanced thanks to readily accessible resources. The genomics and big data era has created a demand for collaboration and data sharing in order to successfully utilise this new information. Here, we describe the online resources for cancer genomics research and assess them based comprehensiveness of the omics data, and user-friendliness. We believe that by introducing the resources now, the cancer research community will be better aware of them and be able to use them more readily. The resources reviewed include data repositories and analysis tools. Although the field of bioinformatics has demonstrated to have a big impact on the medical industry, it can also be extremely important in fields like agriculture, raising livestock, and even space exploration. This collection of articles emphasises the variety of methods and data sets that are being developed in bioinformatics techniques to address difficult problems like how to prioritise lead compounds with the potential to disrupt the tumor-immune microenvironment and how to more accurately predict clinical outcomes.

Acknowledgement: I extend my sincere thanks to Prof. (Dr.) Deepinderpal Singh for constant encouragement and support. I acknowledge the research content published by authors.

Conflict of Interest. None.

REFERENCES

[1]. Arjun K. P. and Kumar, K. S. (2020). Machine Learning - A Neoteric Medicine to Healthcare. *International Journal on Emerging Technologies, 11*(3): 195–201.

[2]. Bulusu, K. C., Tym, J. E., Coker, E. A., Schierz, A. C., and Al-Lazikani, B. (2014). CanSAR: updated cancer research and drug discovery knowledgebase. *Nucleic acids research*, *42*(D1): D1040-D1047.

[3]. Chowdhary, M., Rani, A., Parkash, J., Shahnaz, M., and Dev, D. (2016). Bioinformatics: an overview for cancer research. *Journal of Drug Delivery and Therapeutics*, *6*(4): 69-72.

[4]. Clough, E., and Barrett, T. (2016). The gene expression omnibus database. *In Statistical genomics. Humana Press, New York, NY,* 93-110

[5]. Cohen, J. (2005). Computer science and bioinformatics. *Communications of the ACM*, *48*(3): 72-78.

[6]. Duggan, M. A., Anderson, W. F., Altekruse, S., Penberthy, L., and Sherman, M. E. (2016). The surveillance, epidemiology and end results (SEER) program and pathology: towards strengthening the critical relationship. *The American journal of surgical pathology*, *40*(12): e94.

[7]. Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., and Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*, *6*(269), pl1-pl1.

[8]. Goldman, M., Craft, B., Swatloski, T., Ellrott, K., Cline, M., Diekhans, M., and Zhu, J. (2013). The UCSC cancer genomics browser: update 2013. *Nucleic acids research*, *41*(D1): D949-D954.

[9]. Haddad, Y., Adam, V., and Heger, Z. (2020). Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLoS computational biology*, 16(4): e1007449.

[10]. Hameduh, T., Haddad, Y., Adam, V., and Heger, Z. (2020). Homology modeling in the time of collective and artificial intelligence. *Computational and Structural Biotechnology Journal*, *18*: 3494-3506.

[11]. He, X., Chang, S., Zhang, J., Zhao, Q., Xiang, H., Kusonmano, K., and Wang, J. (2007). Methy Cancer: the database of human DNA methylation and cancer. *Nucleic acids research*, *36*(suppl_1): D836-D841.

[12]. Himakshi Sarma and Venkata Satish Kumar Mattaparthi (2017). Effect of Activation Loop Phosphorylation on Lemur Tyrosine Kinase 3 (LMTK3) activity: A Molecular Dynamics Simulation Study. Biological Forum – An International Journal 9(1), 194-206.

[13]. Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., and Lempicki, R. A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, *8*(9): 1-16.

[14]. Kihara, D., Yang, Y. D., and Hawkins, T. (2006). Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. *Cancer Informatics*, *2*: 117693510600200020.

[15]. Kim, H. (2004). Role of microarray in cancer diagnosis. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, *36*(1): 1-3.

[16]. Kim, I. J., Kang, H. C., and Park, J. G. (2004). Microarray applications in cancer research. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, *36*(4): 207-213.

[17]. Liu, C., Bai, B., Skogerbø, G., Cai, L., Deng, W., Zhang, Y., and Chen, R. (2005). NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic acids research*, *33*(suppl_1): D112-D115.

[18]. Luscombe, N. M., Greenbaum, D., and Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*, *40*(04): 346-358.

[19]. Martincorena, I., and Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science*, *349*(6255): 1483-1489.

[20]. Mehta, A. A. and Mazumdar, H. S. (2020). Discovery of Significant miRNA-biomarkers for Breast Cancer using Decision Tree Classifier. International *Journal on Emerging Technologies, 11*(2): 453–460.

[21]. Packer, B. R., Yeager, M., Burdett, L., Welch, R., Beerman, M., Qi, L., and Chanock, S. J. (2006). SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic acids research*, *34*(suppl_1): D617-D621.

[22]. Singh, S. K., Kumar, A., Singh, R. B., Ghosh, P., and 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.

[23]. Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research*, *45*(W1): W98-W102.

[24]. Wang, Z., Jensen, M. A., and Zenklusen, J. C. (2016). A practical guide to the cancer genome atlas (TCGA). In *Statistical Genomics*. Humana Press, New York, NY. 111-141.

[25]. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., and Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, *46*(1): 296-303.

[26]. Wilks, C., Cline, M. S., Weiler, E., Diehkans, M., Craft, B., Martin, C., and Maltbie, D. (2014). The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database*, *2014*.

[27]. Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., and Lyu, J. (2020). Brief introduction of medical database and data mining technology in big data era. *Journal of Evidence-Based Medicine*, *13*(1): 57-69.

[28]. Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., and Garnett, M. J. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, *41*(D1): D955-D961.

[29]. Yang, Y., Dong, X., Xie, B., Ding, N., Chen, J., Li, Y., and Fang, X. (2015). Databases and web tools for cancer genomics study. *Genomics, proteomics & bioinformatics*, *13*(1): 46-50.

27[30]. Zhang, J., Finney, R. P., Rowe, W., Edmonson, M., Yang, S. H., Dracheva, T., and Buetow, K. H. (2007). Systematic analysis of genetic alterations in tumors using Cancer Genome Work Bench (CGWB). *Genome research*, *17*(7): 1111-1117.