International Journal of Biotechnology and Clinical Medicine ISSN 2583-4665
Vol.2, No.1, Jan - Mar 2023, pp. 01~10

ROLE OF BIOINFORMATICS IN CANCER DIAGNOSIS, PROGNOSIS AND THERAPIES

Md. Selim Reza[1]*, Mst. Ayesha Siddika[2], Khalilur Rahman[1], Jagannath Adhikary[1], Md. Nurul Haque Mollah[1*]
[1]Bioinformatics Lab, Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh
[2]Microbiology Lab, Department of Veterinary and Animal Sciences, University of Rajshahi, Rajshahi-6205, Bangladesh

Article Info

Article history:

Received March 2nd, 2023
Revised March 12th, 2023
Accepted March 27th, 2023

Keyword:

Cancer therapy
Clinical case studies
Drug discovery
Computational chemistry
Diagnosis

ABSTRACT

According to the World Health Organization, cancer is projected to be among the main causes of death worldwide in 2020, accounting for roughly 10 million fatalities. Globally, different cancers claim the most lives, including cervical, endometrial, lung, breast, stomach, rectum, and colon cancers. Several factors, ranging from the type of diet to the type of virus infection, might increase the risk of developing cancer. With proper treatment, cancer is highly likely to be cured. Hence, early discovery of cancer can help to minimize the number of lives lost to cancer each year. The advancement of microarray technology has made it possible to collect a lot of information on the genes expressed differently in malignant cells. This sheer volume of data, computational tools, and databases must be used to store, process, and extract useful information from the data gathered, such as new biomarkers for cancer detection. Hence necessitates the use of bioinformatics tools to accomplish this task. Bioinformatics is paramount in enhancing cancer diagnosis by finding biomarkers that may be used to diagnose cancer. Moreover, bioinformatics aids in the discovery of shared biomarkers and differentially expressed genes, as well as candidate drug agents in various cancer types, boosting the cancer diagnosis process.

Corresponding Author:

Dr. Md. Selim Reza,
Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh.
Email: selim.ru4778@gmail.com

How to Cite:

Md. Selim Reza et al., Role of Bioinformatics in Cancer Diagnosis, Prognosis, and Therapies. IJBTCM. 2023; Volume 2 (Issue 1): Page 01-10.

1. INTRODUCTION

The unregulated or uncontrolled spread of abnormal (tumor or malignant) cells originating from cells of a particular organ across the body is known as cancer. Cancer is brought on by substances like chemical carcinogens that cause DNA mutations, periodic injury, ionizing radiation like ultraviolet radiation, hormones that promote uncontrollable cell growth, genetic abnormalities, immunological dysfunction, and viruses like the human papillomavirus, hepatitis B, and hepatitis C [1,2]. Despite considerable advancements in its diagnosis, cancer remains a commonly lethal disease in humans (the second-leading cause of death). It has become a significant threat to humankind due to its rapid growth rate and genetics [3]. Colon, cervical, breast, ovarian, gastric, pancreatic, esophageal, oral, bladder, lung, lymphoma, leukemia, testicular, melanoma, glioma, prostate, and hepatoma cancer are some forms of cancer [4]. One of the most frequent gynecological cancers is cervical cancer, related to intensive sexual activity with numerous partners, infrequent condom usage, and immunosuppression [5]. Pancreatic cancer is a widespread malignant tumor of the digestive tract with a high degree of resistance and a miserable prognosis [6]. Combining computational and non-computational techniques is used in cancer diagnosis [7–14].

Microarray technology has been successfully applied for over a decade in oncology, tumor categorization, and diagnosis. Its widespread adoption results from the shortcomings of conventional cancer biomarker study methods, which are costly and time-consuming. Microarrays are significantly progressing due

to their compact volume and may be used to research a small number of biomarkers or to rapidly scan a huge number of biomarkers.. Moreover, microarray technology makes it possible to examine a cell's condition at the molecular level and recognize a certain cell species based on its gene expression profile. Bioinformatics techniques are required due to the large amount of data produced by microarray techniques that must be analyzed through computational schemes [5–7,14–17]. Computer technology and biomedical study are combined in the multidisciplinary field of bioinformatics [6,15–19]. To interpret and analyze data on a large scale, it entails designing tools, processing data, and creating databases. The mentioned biomarkers are validated using a heat map based on the co-regulation scores utilizing a co-expression network analysis. Protein-Protein Interactions (PPI) are employed to locate the disease-causing hub genes [3,18–20]. Protein three-dimensional (3D) structures are essential in human biology and bioscience fields like protein function prediction and therapeutic development [21].

This area of bioinformatics has rapidly developed while maintaining up with the increase of genome sequences, emphasizing on therapeutic diagnostics, notably cancer, which is one of the main reasons for mortality globally [22]. The cancer study and diagnosis are being aided by the use of bioinformatics tools like web technology, Gene Expression Profiling Interactive Analysis (GEPIA), Cytoscape, and databases like the National Center for Biotechnology Information (NCBI), Kyoto Encyclopedia of Genes and Genomes (KEGG), gene omnibus (GEO) databases, Surveillance, Epidemiology, and End Results (SEER) database. It is employed to identify numerous cancer types, such as cervical cancer (CC), hepatocellular carcinoma (HCC), colorectal cancer (CRC), pancreatic cancer (PC), breast cancer (BC), lung cancer (LC), etc. Because of this, cancer may now be diagnosed, identified, and prevented quickly. As a result, bioinformatics technology has changed the efficient solution [22].

2. BIOINFORMATICS DATABASES AND TOOLS USED IN CANCER DIAGNOSIS AND THERAPIES

Before using bioinformatics techniques, it is crucial to gather pertinent data related to the field of research. The procedure for gaining this data is known as data mining. Massive volumes of data are analyzed using data mining to find patterns, correlations, & unknowns. This is crucial for procedures like gene discovery, protein function domain and motif recognition, disease diagnosis and prognosis, reconstruction of gene and protein interaction networks, prediction of protein subcellular placement, and data cleansing [23] [5]. The platform used in the data mining process is called Oncomine. Oncomine is a database and integrated mining platform for cancer microarrays that systematically curates, examines, and makes all publicly accessible cancer microarray data available [23].

2.1.The Gene Omnibus (GEO) Database

The GEO database is a public asset that stores and makes accessible high-throughput gene expression data as well as other functional genomics data. The GEO expands to include new data applications as a result of the rapid technological progress, including the study of genome-protein interaction and gene regulation, in addition to studies on gene expression [24]. The database provides users with access to information from thousands of research and offers them with many web-based data analysis resources. With the help of GEO (http://www.ncbi.nlm.nih.gov/geo/), individuals may view and examine data relevant to their interests while receiving thorough descriptions [25].

2.2.The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) is considered to be among the most comprehensive and successful genomic data studies. The database initiative has created, evaluated, and publicly released genomic sequence, copy number variation (CNV), expression, and methylation, data on more than 11,000 individuals from more than 30 distinct forms of cancer [26]. TCGA was a joint effort of the National Human Genome Research Institute (NHGRI) and the National Cancer Institute (NCI) [26]. TCGA is a systematic and well-organized initiative to employ genomic analysis tools, namely large-scale genome sequencing, to improve our comprehension of the molecular causes of cancer.

2.3.The Human Protein Atlas (HPA)

The Human Protein Atlas (HPA) is a Swedish-based database that was introduced in 2003 to map all the human proteins in organs, cells, and tissues by integrating different omics tools, including mass spectrometry based transcriptomics, proteomics, systems biology, and antibody based imaging. The HPA is divided into three sub-atlases: the Cell Atlas, Pathology Atlas, and Tissue Atlas [27]. The algorithm uses sensitive and highly specific antibodies to offer a precise assessment of protein expression. The HPA puts all of its antibodies through a stringent validation process that includes immunohistochemical staining, immunofluorescence testing, and western blot analysis on a selection of carefully chosen sample materials.

3. BIOINFORMATICS TOOLS USED IN CANCER DIAGNOSIS AND THERAPIES

3.1.Identification of differentially expressed genes

Gene expression (GE) is the procedure through which information from a gene is used to synthesize a functional gene product, which might be a protein. Suppose the difference is statistically significant in the read counts of two experimental conditions or a change in the gene expression levels. In that case, that gene is said to be a differentially expressed gene (DEG). To detect DEGs between two circumstances, it is essential to identify statistical distributional features of the data to approximate the nature of differential genes (DGs). Recently, many schemes have been developed to identify DEGs from RNA-seq data. Among them, edgeR [28], edgeR (robust) [29], DESeq [30], DESeq2 [31], NBPSeq[32], EBSeq [33], and baySeq [34] have been widely used with their individual R packages.

3.2.Protein-Protein Interaction (PPI) Network Analysis

Several biological functions, including cell-to-cell interactions, as well as the regulation of metabolism and development, are handled by protein-protein interactions (PPIs) [35]. PPIs are becoming one of the crucial steps of system biology, and the PPI network of DEGs is identified through the STRING online database (https://string-db.org/) [36]. The PPI network's quality is improved using the Cytoscape program [37]. To choose the Hub Genes (HubGs) from the PPI network, one must utilize the Cytoscape plugin cytoHubba [37,38]. The most significant modules from the PPIs networks are detected using the Molecular Complex Detection (MCODE) plugin of the Cytoscape program (http://apps.cytoscape.org/apps/mcode). By MCODE clustering, highly interconnected areas are discovered, assisting the study in efficient drug development [39].

3.3.Regulatory Network Analysis (NetworkAnalyst)

A gene regulatory network is an assemblage of regulatory interactions among transcription factors (TFs) and TF binding sites of particular mRNA to direct certain expression levels of mRNA and their resulting proteins [40]. So, the TFs–hub Genes and miRNAs–hub Genes interaction network analysis is important to explore key transcriptional regulatory TFs and miRNAs of potential biomarkers by using the NetworkAnalyst web server [41].

3.4.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 concepts into well-ordered, relevant groups or biological modules [42]. The program is frequently used for intricate biological tasks like connecting gene-disease associations, identifying enriched biological themes, particularly GO terms, discovering functionally related gene groups, clustering redundant annotation terms, listing interacting proteins, and visualizing genes using Bio Carta and KEGG pathway maps. It primarily utilizes four data analysis modules: (i) GO charts, which show how genes are represented in terms of biological processes (BPs) as well as cellular components (CCs) and molecular functions (MFs); (ii) Domain Charts, which show how differentially expressed genes (DEGs) are distributed over gene families members; (iii) The annotation tool, which autonomously adds comments to gene lists; and (iv) KEGG Charts, which show the DEGs among KEGG biological pathways [43].

3.5. Surveillance, Epidemiology, and End Results Program (SEER)

This program, launched in January 1973, proposes assembly data on cancer diagnosis, treatment, and trends for over 30% of the American (U.S.) population [44]. The program tracks the different cancer types and variations in survival by age, ethnicity, and stage at diagnosis. Over a thousand researchers, doctors, and lawmakers have used the program to study and interpret the variations and development of cancer in the U.S., turning cancer data into discoveries (NCI, 2018).The SEER program has shown to be a very useful instrument for observing molecular subtyping data and histopathologic cancer subtypes. The SEER database and tools have been used in bioinformatics studies to analyze and evaluate early deaths, survival rates, and prognostic survival factors, observe cancer patterns, and enhance overall results.

3.6.Gene Ontology (GO)

This extensive bioinformatics source offers details on how functional genomics can be used to describe biological knowledge. This collaborative effort is accessible at (http://www.geneontology.org). Three categories—molecular function, cell component, and biological process are used to define biological knowledge. The functions carried out by gene products at the molecular level, such as transfer and catalysis, are referred to as molecular functions. The GO details the activities of the gene products rather than describing the intricate

structures where the activities occur. Information on the cellular compartment or steady macromolecular complexes, which are the places where the gene products carry out their functions, is provided by the components of the cell. This is the cellular structure, to put it another way. The more extensive biological processes are those that involve numerous molecular actions. Transmembrane transfer of glucose, as an illustration [45]. The GO could also be combined with the KEGG path like in [46], which discovered this conjunction to analyze the cancer-related long non-coding RNAs.

3.7.Gene Expression Profiling Iterative Analysis (GEPIA)

The GEPIA is a popular online database for profiling cancer and normal gene expression [47]. With just a few clicks, biologists and clinicians can complete comprehensive and complicated data mining chores using this web server, facilitating data mining for research projects, academic discussions, and the development of cancer treatments. GEPIA offers a tool for resolving bulk RNA datasets in the TCGA and Genotype-Tissue Expression (GTEx) projects to study expression profiles across cancer and healthy patient groups. This is accomplished using various methods, including examining the cell type and the traits of various cancer cell types [48]. It offers a better comprehension of gene functions and opens up new possibilities for data mining in cancer studies. You can access the website at http://gepia.cancer-pku.cn/.

3.8.The University of Alabama Cancer Database (UALCAN)

A complete, user-friendly, and interactive online tool for studying cancer omics data is the UALCAN database [49]. It is a combined data-mining tool that makes it easier to analyze the entire cancer transcriptome. Using patient clinical data from 33 distinct cancer types and TCGA RNA-sequencing, UALCAN also incorporates many metastatic tumors. The relative expression study of a query gene or genes in tumor and normal samples is made easier by UALCAN. Also, it lists the top genes that are over- and under-expressed in various cancers. Through UALCAN, one can examine or confirm the pan-cancer expression pattern of numerous user-defined genes. As a result, it functions as a one-stop-shop by making easy access to outside resources like Gene Cards, the Human Protein Reference Database, PubMed, Target Scan, and Human Protein Atlas that are used to research protein expression in different cancers. UALCAN makes it simple for users to find publicly accessible cancer OMICS data, spot biomarkers, conduct in silico gene validation on potential genes of interest, and provide graphs and plots showing patient survival data and expression profile information.

3.9.Molecular Docking

Molecular interactions, including enzyme-substrate, drug-nucleic acid, protein-nucleic acid, protein-protein, and drug-protein, play key roles in several crucial biological processes, like cell regulation, transport, antibody-antigen recognition, signal transduction, enzyme inhibition, gene expression control, and even the assembly of multi-domain target proteins [50–52]. When these interactions occur, stable protein-ligand or protein-protein complexes are frequently formed, crucial for the involved proteins' biological functions. The 3D structure of a protein is essential to understand the binding mechanism and affinities among the interacting molecules. However, using experimental techniques like X-ray crystallography or NMR to produce complex structures is frequently challenging and costly [51]. Consequently, molecular docking is crucial for comprehending protein-ligand or protein-protein interactions [53–55]. Molecular docking is a broadly utilized computer simulation process to compute the conformation of a protein-ligand complex.

To evaluate and explain protein-ligand or protein-protein interactions, a wide variety of algorithms are available, and their number is continually improving. In molecular docking procedures, precision and speed are essential for achieving good outcomes. Many methods share common methodologies with new enhancements designed to produce a quick method with the highest level of accuracy. The most common molecular docking platforms are Glide [56], DOCK [57], ICM [58], FlexX [59], DockVision [60], AutoDock-vina [61], and AutoDock [62], etc.

4. Molecular Dynamics Simulations

Molecular dynamics (MD) is an intelligent computing approach that permits us to simulate the interactions of molecules and atoms of a scheme over a certain time by solving classical equations of motion. In biological sciences, MD simulations are often utilized to understand the molecular dynamics of protein and ligand-protein complexes to study the molecular mechanisms of action and their basic processes. The wide range of MD simulation allows us to examine a protein's structural flexibility and stability to study its native conformational space or study perturbations induced by allosteric inducers, changes in potential, native ligands, etc. It can offer a dynamic view of the binding procedure and an understanding of the function of protein regions [63,64]. It also reveals how a protein interacts with its environment, including binding partners, localization, etc. [65]. Some of the more popular MD simulation methods are AMBER [66], GROMACS [67], CHARMM [68],

NAMD [69], and YASARA [70], etc. For visualization aims, some standard software is UCSF Chimera [71], visual molecular dynamics (VMD) [72], PyMOL [73], Rasmol [74], and Discovery Studio Visualizer [75], etc. For analysis of numerical and mathematical data, some common software is SciDAVis (https://scidavis.sourceforge.net/), xmgrace (https://plasma-gate.weizmann.ac.il/Grace/), Gnuplot (http://www.gnuplot.info/), R [76], Origin (https://www.originlab.com/) or even Excel.

5. Case studies on the use of bioinformatics databases and tools for cancer diagnosis

The diagnostic procedure for many cancer kinds has been greatly enhanced due to the presented bioinformatics tools and databases. Three case studies were examined for comprehending how the bioinformatics tools and databases provided have enhanced the cancer detection process. The examined case studies focused on increasing the accuracy of cancer diagnoses for three of the most prevalent cancer types, including pancreatic, breast, and cervical cancer.

5.1.Cervical Cancer

The symptoms of cervical cancer (CC), a form of malignancy that develops from the cervix (lower portion of the uterus), include vaginal discharge and irregular bleeding, pelvic discomfort, and pain during sexual activity [77]. According to reports, human papillomavirus (HPV) infection nearly always results in CC [78]. CC is now ranked as the 2nd most frequent malignancy in women in middle- and low-income countries (MLICs) and the 4th most common kind of cancer in women, with a high death rate globally [79,80]. Around 569,847 new CC cases with 311,365 fatalities are reported each year, according to the 2018 Globocan data [80]. In the US, CC affects 14,065 female patients annually, resulting in 5266 fatalities [81]. Around 84–90% of these fatalities occurred in MLICs, including South Africa [82]. Nevertheless, various novel gene properties and signal pathways that may be utilized to identify cervical cancer have been found by using high-throughput sequencing skills and bioinformatics techniques to evaluate the data obtained. So, by detecting cell diseases at an early stage using gene features and signal pathways, illness diagnosis, prognosis, and recurrence may all be enhanced [83].

Our earlier study used many well-known bioinformatics techniques to identify candidate genes, emphasizing their regulatory elements and the dysregulated molecular activities and pathways that were in charge of the development of CC [3,5]. Using the GPEA database to evaluate their distinct patterns of expression between CC and normal samples, studied find out four candidate genes (CDK1, AURKA, TOP2A, and CHEK1) as the key genes (KGs). Finally identified five FDA-approved candidate medications (Docetaxel, Temsirolimus, vincristine, vinorelbine, and paclitaxel) based on the proposed potential genes using molecular docking analysis.

5.2.Breast Cancer

Being the most common disease in females and one with rising death rates in recent years, breast cancer (BC) is a major topic for biomedical research. The significance of early detection and treatment for breast cancer cannot be overstated. Biopsy, Magnetic resonance imaging (MRI), positron emission tomography (PET), mammography, and ultrasound, are all used to diagnose breast cancer, but they are costly, insensitive, and time-consuming. By identifying BC biomarkers for early detection and, subsequently breast cancer diagnosis, bioinformatics offers a faster and more effective method [9].

By bioinformatics tools, our earlier research revealed 13 DEGs (IRF9, AKR1C1, ANGPT1, OAS1, SLCO2A1, OAS3, NQO1, FN1, TP53INP1, HPGD, BCL11A, and ATF6B) as the potential genes that cause BC [9]. The KEGG pathway enrichment analysis and the GO terms (BPs, MFs, and CCs) study both identified certain key GO functions from each of the BPs, MFs, and CCs that DEGs, including potential genes, considerably enrich. This research also identified seven small molecules as the top-ranked potential medications for treating BC: nilotinib, NVP-BHG712, AP-24534, GSK2126458, TG-02, YM201636, and CX-5461.

5.3.Hepatocellular Carcinoma

Hepatocellular carcinoma (HCC), a kind of primary liver cancer, is the third leading cause of cancer-related death globally, with a substantially greater mortality rate than its incidence[79]. In Southeast Asia and Africa, where the hepatitis B virus is endemic, HCC mortality and incidence rates are high[84,85]. In contrast, except for Thailand, the incidence of HCC is relatively low in Europe, Australia, North America, and most Asian countries[86,87]. Over the last decade, there has been a slight improvement in systemic treatment for HCC. Despite significant advancements in HCC treatment, such as liver transplantation, interventional therapy, and radical surgical resection[88–90], global long-term HCC survival rates remain low.

Our previous study discovered ten potential genes (CDKN3, TK1, NCAPG, CDCA5, RACGAP1, AURKA, PRC1, UBE2T, MELK, and ASPM) for HCC [12]. The DEG-set enrichment analysis with the GO-terms and KEGG pathways revealed HCC-related few crucial molecular functions, biological processes, and signaling pathways. This study also identified three candidate drug agents (Dactinomycin, Vincristine, and Sirolimus) for the treatment of HCC through molecular docking analysis.

5.4.Pancreatic Cancer

Cancer that arises in the pancreas is known as pancreatic cancer (PC). It is one of the crucial obstacles in deaths related to cancer globally[91–93]. In both sexes, PC occurrence and mortality rate increase with age, and it is usually diagnosed in people over 70 [94]. At most, 10% of PC cases live more than 5 years.

Our previous study identified the top-ranked eight key genes/proteins (ADAM10, COL1A1, COL1A2, COL3A1, FBN1, FN1, LAMC1, and P4HB) as genomic biomarkers through PPI network analysis [6]. The top-ranked 5 TFs proteins (FOXC1, FOXL1, YY1, STAT1, STAT3) and 5 miRNAs (hsa-mir-29c-3p, hsa-mir-29b-3p, hsa-mir-6752-5p, hsa-mir-6842-5p, hsa-mir-7110-5p) were identified using the GRN analysis. Furthermore, this study selected the top-rated six candidate drugs (Linsitinib, NVP-BHG712, Timosaponin A-III, Irinotecan, CX5461, and Olaparib) for the treatment against PC based on molecular docking and dynamic simulation analysis.

5.5.Gastric Cancer

One of the most prevalent malignant tumors and the third biggest cause of cancer-related death worldwide is gastric cancer (GC) [95]. Despite having access to cutting-edge therapy, the prognosis for GC remains dismal, and the overall survival rate has not risen over 30% [96]. Clinical diagnosis and therapy development are hampered by the molecular heterogeneity of GC patients [97]. Using bioinformatics analysis, our earlier work [13] found three genes for GC patients: CDH2, COL4A1, and COL5A2. Efficacious supplementary medications (Everolimus, Docetaxel, Lanreotide, Venetoclax, Temsirolimus, and Nilotinib) for treating GC patients were also suggested in this research.

5.6.Colorectal Cancer

The second most lethal tumor globally and the third most prevalent solid malignancy is colorectal cancer (CRC) [79]. By 2030, there will be 2.2 million new instances of CRC and 1.1 million fatalities worldwide, a 60% increase in the prevalence of the disease [98]. Due to a lack of data on diagnostic biomarkers and the molecular basis of CRC, the number of new cases and fatalities are rising [98]. Early CRC diagnosis is linked to reduced morbidity and death rates and a greater survival rate than late detection. For instance, with CRC, early identification boosts the five-year survival rate from 11% (late detection) to 90% [99]. The top 10 DEGs in our study— CDC20, CDKN3, CKS2, PTTG1, CDK1, TOP2A, AURKA, MELK, TPX2, and MAD2L1,—were deemed to be the core genes (CGs), which were highly predictive of prognosis in CRC's early stages [100]. The enrichment analysis also discovered several significant GO keywords and signaling pathways that cause CRC. Lastly, this work used molecular docking analysis to choose seven potential medications (Cardidigin, Manzamine A, Staurosporine, Benzo[a]pyrene, Sitosterol, Riccardin D, Nocardiopsis sp.) for the therapy of CRC.

6. CONCLUSION

In summary, being the top cause of mortality worldwide, cancer research is still extremely challenging. Early and effective diagnosis is crucial, but non-computational procedures can be difficult since they are more costly, less effective, and expose patients to radiation like CT scans. Therefore, bioinformatics has emerged as a highly potent and revolutionary way to improve cancer diagnosis, prognosis, and therapies. It aids in the early cancer detection, prognosis, and treatment of several forms of cancer by utilizing a multitude of techniques and databases in combination with the large amounts of data provided by microarray technology.

FUNDING

Nil

ETHICAL APPROVAL

Nil

COMPETING INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Kumar P, Pawaiya RVS. Advances in cancer diagnostics. Braz J Vet Pathol. 2010; 3(2): 142-153.
2. Wang W, Luo J, Wang S. Recent Progress in Isolation and Detection of Extracellular Vesicles for Cancer Diagnostics. Adv Healthcare Mat. 2018; 7(20): 10800484.
3. Reza MS, Harun-Or-Roshid M, Islam MA, Hossen MA, Hossain MT, Feng S, et al. Bioinformatics Screening of Potential Biomarkers from mRNA Expression Profiles to Discover Drug Targets and Agents for Cervical Cancer. Int J Mol Sci. 2022; 23(7): 3968.
4. Ou HT, Chung WP, Su PF, Lin TH, Lin JY, Wen YC, et al. Health-related quality of life associated with different cancer treatments in Chinese breast cancer survivors in Taiwan. Eur J Cancer Care (Engl). 2019; 28 (4): e13069.
5. Reza MS, Hossen MA, Harun-Or-Roshid M, Siddika MA, Kabir MH, Mollah MNH. Metadata analysis to explore hub of the hub-genes highlighting their functions, pathways and regulators for cervical cancer diagnosis and therapies. Discov Oncol. 2022; 13(1): e13069.
6. Hossen MB, Islam MA, Reza MS, Kibria MK, Horaira MA, Tuly KF, et al. Robust identification of common genomic biomarkers from multiple gene expression profiles for the prognosis, diagnosis, and therapies of pancreatic cancer. Comput Biol Med. 2023; 152: 106411.
7. Mosharaf P, Reza S, Gov E, Mahumud RA. Disclosing Potential Key Genes , Therapeutic Targets and Agents for Non-Small Cell Lung Cancer : Evidence from Integrative Bioinformatics Analysis. Vaccines 2022; 10(5): 771.
8. Mosharaf MP, Reza MS, Kibria MK, Ahmed FF, Kabir MH, Hasan S, Mollah MNH. Computational identification of host genomic biomarkers highlighting their functions, pathways and regulators that influence SARS-CoV-2 infections and drug repurposing. Scientific Reports 2022; 12(1):4279.
9. Alam MS, Sultana A, Reza MS, Amanullah M, Kabir SR, Mollah MNH. Integrated bioinformatics and statistical approaches to explore molecular biomarkers for breast cancer diagnosis, prognosis and therapies. PLoS One. 2022; 17(5): e0268967.
10. Hossain MT, Li S, Reza MS, Feng S, Zhang X, Jin Z, et al. Identification of circRNA Biomarker for Gastric Cancer through Integrated Analysis. Front Mol Biosci. 2022; 9: 175.
11. Rahman MM, Hossain MT, Reza MS, Peng Y, Feng S, Wei Y. Identification of Potential Long Non-Coding RNA Candidates that Contribute to Triple-Negative Breast Cancer in Humans through Computational Approach. Int J Mol Sci. 2021; 22: 12359.
12. Hossen MA, Reza MS, Harun-Or-Roshid M, Islam MA, Siddika MA, Mollah MNH. Identification of Drug Targets and Agents Associated with Hepatocellular Carcinoma through Integrated Bioinformatics Analysis. Curr Cancer Drug Targets. 2023.
13. Hossain MT, Reza MS, Peng Y, Feng S, Wei Y. Identification of Key Genes as Potential Drug Targets for Gastric Cancer. Tsinghua Sci Technol. 2023; 28(4): 649–664.
14. Mosharaf MP, Kibria MK, Hossen MB, Islam MA, Reza MS, Mahumud RA, Alam K, Gow J, Mollah MNH. Meta-Data Analysis to Explore the Hub of the Hub-Genes That Influence SARS-CoV-2 Infections Highlighting Their Pathogenetic Processes and Drugs Repurposing. Vaccines 2022;10(8):1248.
15. Saravanan KM, Zhang H, Hossain MT, Reza MS, Wei Y. Deep Learning-Based Drug Screening for COVID-19 and Case Studies. In: Methods in Pharmacology and Toxicology [Internet]. 2021. p. 631–60. Available from: https://link.springer.com/10.1007/7653_2020_58
16. Hossain T, Zhang J, Reza S, Peng Y, Feng S, Wei Y. Reconstruction of Full-Length circRNA Sequences Using Chimeric Alignment Information. Int J Mol Sci. 2022; 23(12): 6776.
17. Reza MS, Cai YP, Zhang L, Zhang X, Wei Y. Computational Solutions for Microbiome and Metagenomics Sequencing Analyses. Front Mol Biosci. 2021; 8: 698384.
18. Ahmed FF, Reza MS, Sarker MS, Islam MS, Mosharaf MP, Hasan S, et al. Identification of host transcriptome-guided repurposable drugs for SARS-CoV-1 infections and their validation with SARS-CoV-2 infections by using the integrated bioinformatics approaches. PLoS One. 2022;17(4): e0266124.
19. Mosharaf MP, Reza MS, Kibria MK, Ahmed FF, Kabir MH, Hasan S, et al. Computational identification of host genomic biomarkers highlighting their functions, pathways and regulators that influence SARS-CoV-2 infections and drug repurposing. Sci Rep. 2022; 12(1): 4279.
20. Chang JW, DIng Y, Tahir Ul Qamar M, Shen Y, Gao J, Chen LL. A deep learning model based on sparse auto-encoder for prioritizing cancer-related genes and drug target combinations. Carcinogenesis. 2019; 40(5): 624-632.
21. Reza MS, Zhang H, Hossain MT, Jin L, Feng S, Wei Y. Comtop: Protein residue–residue contact prediction through mixed integer linear optimization. Membranes. 2021; 11(7): 503.
22. Chowdhary M, Rani A, Parkash J, Shahnaz M, Dev D. Bioinformatics: an overview for cancer research. J Drug Deliv Ther. 2016; 6(4): 69-72.
23. Ushasri K, Prasad AR, Reddy JKK, Saravana S. Significance of Data Mining in Bioinformatics. Int J Eng Res Technol. 2014; 1: 86–88.
24. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. Neoplasia. 2004; 6(1): 1-6.
25. Clough E, Barrett T. The Gene Expression Omnibus database. In: Methods in Molecular Biology. 2016; 1418: 93-110.
26. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, et al. NCBI GEO: Mining millions of expression profiles - Database and tools. Nucleic Acids Res. 2005; 33: 562-566.
27. Wang Z, Jensen MA, Zenklusen JC. A practical guide to The Cancer Genome Atlas (TCGA). Methods in Molecular Biology. 2016; 1418: 111-141.

28. Lindskog C. The Human Protein Atlas – an important resource for basic and clinical research. Expert Rev Proteomics. 2016; 13(7): 627-629.
29. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1): 139-140.
30. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Res. 2014; 42(11): e91.
31. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11(10): R106.
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12): 550.
33. Di Y, Schafer DW, Cumbie JS, Chang JH. The NBP negative binomial model for assessing differential gene expression from RNA-Seq. Stat Appl Genet Mol Biol. 2011; 10(1).
34. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, et al. EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics. 2013; 29(8): 1035–1043.
35. Hardcastle TJ, Kelly KA. BaySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics. 2010; 11: 1-14.
36. Braun P, Gingras AC. History of protein-protein interactions: From egg-white to complex networks. Proteomics. 2012; 12(10): 1478-1498.
37. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39: D561-568.
38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11): 2498–2504.
39. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014; 8(4): 1-7.
40. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003; 4(1): 1-27.
41. Yachie-Kinoshita A, Kaizu K. Cell modeling and simulation. Encycl Bioinforma Comput Biol ABC Bioinforma. 2018; 1(3): 864–873.
42. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019; 47(W1): W234–W241.
43. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007; 8(9): R183.
44. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003; 4(5): R60.
45. Duggan MA, Anderson WF, Altekruse S, Penberthy L, Sherman ME. The surveillance, epidemiology, and end results (SEER) program and pathology: Toward strengthening the critical relationship. Am J Surg Pathol. 2016; 40(12): e94-e102.
46. Blake JA, Christie KR, Dolan ME, Drabkin HJ, Hill DP, Ni L, et al. Gene ontology consortium: Going forward. Nucleic Acids Res. 2015;43(D1): D1049-1056.
47. Chen L, Zhang YH, Lu G, Huang T, Cai YD. Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways. Artif Intell Med. 2017; 76: 27-36.
48. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017; 45(W1): 98-102.
49. Li C, Tang Z, Zhang W, Ye Z, Liu F. GEPIA2021: Integrating multiple deconvolution-based analysis into GEPIA. Nucleic Acids Res. 2021; 49(W1): 242-246.
50. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017; 19(8): 649-658.
51. Rohs R, Bloch I, Sklenar H, Shakked Z. Molecular flexibility in ab initio drug docking to DNA: Binding-site and binding-mode transitions in all-atom Monte Carlo simulations. Nucleic Acids Res. 2005; 33(22): 7048–7057.
52. Hernndez-Santoyo A, Yair A, Altuzar V, Vivanco-Cid H, Mendoza-Barrer C. Protein-Protein and Protein-Ligand Docking. Protein Eng Technol Appl. 2013.
53. Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. Biophys Rev. 2014; 6(1): 75–87.
54. Huang SY, Zou X. Advances and challenges in Protein-ligand docking. Int J Mol Sci. 2010; 11(8): 3016–3034.
55. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. Proteins Struct Funct Genet. 2006; 65(1): 15–26.
56. Meza Menchaca T, Juárez-Portilla C, C. Zepeda R. Past, Present, and Future of Molecular Docking. Drug Discov Dev - New Adv. 2020.
57. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem. 2004; 47(7): 1739-1749.
58. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol. 1982; 161(2): 269-288.
59. Abagyan R, Totrov M, Kuznetsov D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. J Comput Chem. 1994; 15(5): 488–506.

60. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol. 1996; 261(3): 470–489.
61. Hart TN, Read RJ. A multiple-start Monte Carlo docking method. Proteins Struct Funct Bioinforma. 1992; 13(3): 206–22.
62. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2009; 31(2): 455-461.
63. Morris GM, Huey R Fau - Lindstrom W, Lindstrom W Fau - Sanner MF, Sanner Mf Fau - Belew RK, Belew Rk Fau - Goodsell DS, Goodsell Ds Fau - Olson AJ, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009; 30(16): 2785-2791.
64. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron. 2018; 99(6): 1129–1143.
65. Latorraca NR, Fastman NM, Venkatakrishnan AJ, Frommer WB, Dror RO, Feng L. Mechanism of Substrate Translocation in an Alternating Access Transporter. Cell. 2017; 169(1): 96-107.
66. Latorraca NR, Wang JK, Bauer B, Townshend RJL, Hollingsworth SA, Olivieri JE, et al. Molecular mechanism of GPCR-mediated arrestin activation. Nature. 2018; 557(7705): 452–456.
67. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci. 2013; 3(2): 198–210.
68. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015; 1(2): 19–25.
69. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. J Comput Chem. 2009; 30(10): 1545–1614.
70. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26(16): 1781–1802.
71. Krieger, Elmar GV, Spronk C. YASARA - Yet Another Scientific Artificial Reality Application. YASARA.org. 2013;
72. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem. 2004; 25: 1605-1612.
73. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996; 14(1):33-38.
74. DeLano WL. The PyMOL Molecular Graphics System, Version 2.3. Schrödinger LLC. 2020.
75. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci. 1995; 20(9): 374–376.
76. Studio D. Discovery Studio Visualizer. Discovery. 2014; 3–5.
77. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. J Comput Graph Stat. 1996; 5(3): 299–314.
78. Urasa M, Darj E. Knowledge of cervical cancer and screening practices of nurses at a regional hospital in Tanzania. Afr Health Sci. 2011; 11(1): 48-57.
79. Small W, Bacon MA, Bajaj A, Chuang LT, Fisher BJ, Harkenrider MM, et al. Cervical cancer: A global health crisis. Cancer. 2017; 123(13): 2404-2412.
80. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68(6): 394-424.
81. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Heal. 2020; 8(2): 191-203.
82. Bruni L, Alberto G, Serrano B, Mena M, Gómez D, Muñoz J, et al. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in India. Summ Rep 10 December 2018. 2017;27(July).
83. Allanson ER, Schmeler KM. Cervical Cancer Prevention in Low- And Middle-Income Countries. Clin Obstet Gynecol. 2021; 64(3): 501-518.
84. Martínez-Rodríguez F, Limones-González JE, Mendoza-Almanza B, Esparza-Ibarra EL, Gallegos-Flores PI, Ayala-Luján JL, et al. Understanding cervical cancer through proteomics. Cells. 2021; 10(8): 1854.
85. Popper H, Shafritz DA, Hoofnagle JH. Relation of the hepatitis B virus carrier state to hepatocellular carcinoma. Hepatology. 1987; 7(4): 764–772.
86. Tanaka M, Katayama F, Kato H, Tanaka H, Wang J, Qiao YL, et al. Hepatitis B and C virus infection and hepatocellular carcinoma in China: A review of epidemiology and control measures. J Epidemiol. 2011; 21(6): 401–416.
87. Wu J, Yang S, Xu K, Ding C, Zhou Y, Fu X, et al. Patterns and Trends of Liver Cancer Incidence Rates in Eastern and Southeastern Asian Countries (1983–2007) and Predictions to 2030. Gastroenterology. 2018; 154(6): 1719-1728.
88. Turdean S, Gurzu S, Turcu M, Voidazan S, Sin A. Current data in clinicopathological characteristics of primary hepatic tumors. Rom J Morphol Embryol. 2012; 53(3): 719–724.
89. Reig M, da Fonseca LG, Faivre S. New trials and results in systemic treatment of HCC. J Hepatol. 2018; 69(2): 525–533.
90. Cauchy F, Zalinski S, Dokmak S, Fuks D, Farges O, Castera L, et al. Surgical treatment of hepatocellular carcinoma associated with the metabolic syndrome. Br J Surg. 2013; 100(1): 113–121.
91. Luo JJ, Zhang ZH, Liu QX, Zhang W, Wang JH, Yan ZP. Endovascular brachytherapy combined with stent placement and TACE for treatment of HCC with main portal vein tumor thrombus. Hepatol Int. 2016; 10(1): 185–195.
92. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. The Lancet. 2016; 388: 73–85.
93. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020; 70(1): 7–30.
94. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021; 71(1): 7–33.

95. Ilic M, Ilic I. Epidemiology of pancreatic cancer. World J Gastroenter. 2016; 22: 9694–9705.
96. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J Clin. 2018; 68(6): 394-424.
97. Rugge M, Genta RM, Di Mario F, El-Omar EM, El-Serag HB, Fassan M, Hunt RH, Kuipers EJ, Malfertheiner P, Sugano K, Graham DY. Gastric cancer as preventable disease. Clinical Gastroenter Hepatol. 2017; 15(12): 1833-1843.
98. Shi Y, Qi L, Chen H, Zhang J, Guan Q, He J, et al. Identification of Genes Universally Differentially Expressed in Gastric Cancer. Biomed Res Int. 2021: 7326853.
99. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017; 66(4): 683-691.
100. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics. CA Cancer J Clin. 2020; 70(3): 145-164.
101. Islam MA, Hossen MB, Horaira MA, Hossen MA, Kibria MK, Reza MS, Tuly KF, Faruqe MO, Kabir F, Mahumud RA, Mollah MN. Exploring Core Genes by Comparative Transcriptomics Analysis for Early Diagnosis, Prognosis, and Therapies of Colorectal Cancer. Cancers. 2023; 15(5): 1369.