



Review Article

Bioinformatics in Early detection and Prognosis of Cancer

Md Soleyman Hossain^{1*}, Farha Matin Juliana²

¹Department of Biochemistry & Molecular Biology, Jahangirnagar University, Dhaka-1342, Bangladesh

²Professor, Department of Biochemistry & Molecular Biology, Jahangirnagar University, Dhaka-1342, Bangladesh

ARTICLE INFO

Received: 18 June 2024

Accepted: 22 June 2024

Published: 01 July 2024

Keywords:

Bioinformatics; Proteomics; multi-omics; computational; Prognosis; LUAD, LUSC, PCa, BRCA, TAP1.

DOI:

10.5281/zenodo.12610166

ABSTRACT

A relatively new area of biology called bioinformatics deals with the processing and interpretation of biological data using computational, analytical, and algebraic methods. The broad term "bioinformatics" describes the application of digital technology to the study of biological processes through the collection of high-dimensional data from several sources. The main focus of bioinformatics research, which is mostly done in silico and usually entails synthesizing new knowledge from existing data, is the design and testing of the software tools needed to assess the data. A cancer patient's prognosis is enhanced by early detection; nonetheless, early diagnosis might be challenging to accept. Proteomics research and DNA microarrays have been utilized for extensive gene expression studies. In order to facilitate researchers' rapid acquisition and comprehension of these tools' functions, this review explores the use, features, and web servers, including KM plotter, GEPIA2, TCGA, PrognoScan, UALCAN, UCSC Xena, HPA, GENT2, cBioportal, GDC portal, TIMER, others multi-omics statistical tools.

INTRODUCTION

The majority of this review explains to the researchers what bioinformatics is, how it's used in cancer detection and therapy, and how to use various kinds of bioinformatic tools and databases. One of the leading causes of death worldwide, cancer is extremely challenging to diagnose in its early stages. The utilization of many biomarkers in various cancer types at the genetic level is the primary topic of this review, which gives a clinical

practitioner a clear concept for improved treatment and prognosis. Bioinformatics is the study of biological data acquisition and interpretation through the use of computational and analysis tools [1]. A relatively new area of biology called bioinformatics deals with the processing and interpretation of biological data using computational, analytical, and algebraic methods. During the "genomics" and "omics" eras, the rapid growth of internet communication knowledge

*Corresponding Author: Md Soleyman Hossain

Address: Department of Biochemistry & Molecular Biology, Jahangirnagar University, Dhaka-1342, Bangladesh

Email ✉: shjubio0000@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



made it overwhelming for laboratory researchers to examine data from experiments conducted during the preceding ten years. The increasing demand for in-depth biological research means that traditional gene-by-gene approaches are insufficient to fully satisfy genuine biology. The massive volumes of data produced by emerging sciences like genome sequencing and microarray chips demand data management and integration across multiple platforms. After that, in order to acquire biotic knowledge for therapeutic outcomes, information analysis and reporting are required [2]. A wide term, "bioinformatics" refers to the use of digital technology to study biological processes using high-dimensional data collected from many resources. The design and testing of the software tools required to evaluate the information are the core of bioinformatics research, which is conducted in large portions and typically involves the synthesis of new learning from available data negatively impacts the other's safety, therapeutic effectiveness, appearance, or elegance [1].

Background:

Due to the disease's poor diagnosis, cancer has been characterized as a complex sickness that affects a large number of people and is thought to be a common prominent cause of individual mortality. Prognoses and therapy may be correlated with treatment response and tolerance, duration, locations, and cellular expansion, source, and pathophysiological understanding. A vital and significant component of clinical medicine, cancer bioinformatics is also a vital instrument and method in the study of cancer [3]. The human body makes a large number of genes, proteins, and RNA that are regulated in space and time to work as a complicated network. This issue renders the traditional gene-by-gene approach unreliable and unable to provide a complete picture of cellular activity. With the use of microarray technology, gene expression across the entire genome can now be assessed in a single experiment [4].

Personalized treatment, prognosis, and diagnosis may all profit from multi-omics approaches. Sufficient bioinformatics capabilities are needed to meet the demands of precision oncology by organizing, integrating, and interpreting large amounts of complex data. Because of a stricter policy requirement, we analyze the unique requirements of bioinformatics approaches and tools that emerge in the context of clinical cancer, including the need for rapid, highly reproducible, and stable procedures. We describe the steps involved in creating a molecular tumor board, from the first analysis of raw genomic profile data to the automated report preparation. We also discuss the specific bioinformatics support needed for this procedure [5]. To mediate biological activity, DNA is transcribed into RNA, and RNA is translated into protein, according to the widely accepted fundamental concept of genetics. However, the human genome project shows to our amazement that just 1.5% of the human genome contains protein-coding genes [6-8]. Previously genome assemblers were only capable of successfully complete the creation of tiny bacterial genomes, but progression in data quality and quantity, together with more sophisticated design methods and processing resources, have made it possible to complete the assembly of more complex eukaryotic genomes [9]. Projects that aim to explain cancer from a global perspective are giving researchers the chance to have greater accessibility to data to integrate and analyze in fresh ways. The ultimate objective of cancer bioinformatics is the development of new treatment and diagnostic methods. Many tools were introduced and developed to know various problems starting from tumor heterogeneity to the analysis of gene mutation [10].

Bioinformatics in cancer diagnosis:

Early diagnosis of cancer causes improved prognosis, but at the same time, it is difficult to confirm the diagnosis at a very early stage due to



the lack of top-notch statistical models that take into consideration clinical triage and variations in aggressiveness. Research suggests that identification delay in cancer matters; the evidence for this is mounting, although it is difficult to quantify its effect on survival or fatality [11,12]. The substances present in tobacco are DNA-toxic agents that may have a significant impact on the development and spread of certain cancers [13,14]. Cancer bioinformatics plays a major part in the identification and authentication of biomarkers, specified to early identification, about clinical phenotype, and also measures and observes the prognosis of disease and outcome of treatment and predicts the enhancement of patient's life quality. The incorporation of knowledge on protein annotations, relations, and signaling pathways, and complex biomarkers and novel classes of biomarkers with protein-protein interlinkage were studied. One of the novel approaches is to use effective complex biomarkers to track and assess changes in network biomarkers at various stages and intervals throughout the onset of illnesses. Clinical informatics such as clinical manifestations, patient complaints, treatment histories, biochemical analyses, imaging, pathologies, and other data were anticipated to be connected [15]. The use of DNA microarrays and proteomics studies for large-scale gene expression research has advanced technology, thus elevating the significance of bioinformatics tools. In today's research, wet experimentation and the application of bioinformatics analytics go side by side [16]. Molecular profiling of tumor biopsies is becoming more crucial to both cancer research and the treatment of cancer. Diagnosis, prognosis, and individualized treatment improvements are possible with bioinformatics. Relevant bioinformatics approaches for managing, integrating, and analyzing big, network information are required to fulfill the word precision oncology. Bioinformatics techniques

and software that are developed in the context of oncology as a result of the strict monitoring environment and demand for quick, incredibly reproducible, and reliable methods are particularly needed. The molecular tumor board's plan and particular bioinformatics assistance that is needed, from the initial study of fresh molecular profiling information to the computerized making of the report, must be outlined. Numerous clinical studies and genomic tumor boards at specialist cancer centers and medical centers across the globe have employed similar approaches to various extents. Previous initiatives as well as modern ones ought to be examined to integrate tumor boards with certain other top pan-omics patient data, as well as the capability of clinical methodologies to convert molecular discoveries into advice on appropriate treatments. The method used to investigate the genetic basis of cancer is being revolutionized. Instead of concentrating on specific genes, researchers are now investigating important areas of the expressed genome. The quantity of data saved in the patient record and the amount of molecular data produced from the testing facility are growing at an incredible rate. To obtain a fresh understanding of the genetics of cancer, it is essential to find innovative approaches for combining these data. As a result, bioinformatics, the fusion of biology, information science, and computation, continues to become an essential part of cancer research [17].

Bioinformatics databases and tools

It is very important to collect useful and important data related to the study before using any technique. This process of collection of data is called data mining [18]. Oncomine is the name of the software system utilized in the data mining procedure. All free-of-charge cancer microarray data is meticulously curated, examined, and made accessible through the database and information mining technology known as Oncomine [19]. The tools of bioinformatics include the Database for



Annotation, Visualization and Integrated Discovery (DAVID), Gene Ontology (GO), Surveillance, Epidemiology, Results Program (SEER), and Gene Expression Profiling Iterative Analysis (GEPIA) [20]. Commonly used cancer databases that are freely accessible for cancer research described below sequentially.

A. Oncomine Database:

Databases are classified as follows: databases harboring gene/microRNA expression profiles, databases for copy number variations (CNVs), DNA mutation detection databases, epigenetic profiles databases, databases with integrative analyses, and databases with other data types [21]. Differential expression of TCGA (The Cancer Genome Atlas) samples for different cancer can be compared with normal complements through the ONCOMINE database (<https://www.ONCOMINE.org/resource/login.html>)[22-24]. The analysis of expression level are normally conducted considering a threshold parameter of; p-value: 1E-4, fold change: 2, gene ranking: 10%.

B. GEPIA2:

GEPIA stands for Gene Expression Profiling Interactive Analysis is a renowned server for differential gene expression, survival analysis, correlation analysis, similar gene detection etc. The expression profile for genes throughout different cancer and their complementary normal tissues are detected through the GEPIA2 website (<http://gepia2.cancer-pku.cn/>) [25].

C. GENT2:

A platform for exploring gene expression patterns across normal and tumor tissues (<http://gent2.appex.kr/gent2/>). This platform is designed to help in gene profile analysis, subtype analysis, meta-survival analysis [26].

D. GTEx:

Genotype-Tissue Expression (GTEx) (<https://gtexportal.org/home/>) is a platform for isoform expression, exon expression, junction

expression analysis for cancer genes. The GTEx (Genotype-Tissue Expression) data and TCGA data are generally matched by the ANOVA differential method and to detect the expression with default threshold settings through the GEPIA2 website [27].

E. Human Protein Atlas database:

Expression of protein in different cancer tissues and their counterpart normal tissue are retrieved as immunohisto-chemistry images retrieved Human Protein Atlas database (<https://www.proteinatlas.org/>) [28, 29].

F. UALCAN:

University of Alabama at Birmingham CANCER Data Analysis Portal is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data. Sample gene expression and promoter methylation analysis are conducted with the UALCAN website (<http://ualcan.path.uab.edu/index.html>) by comparing the TPM (transcript per million) and beta value, respectively [28, 29]. The TCGA database is used in UALCAN to analyzed mRNA transcription count of cancer tissues with their normal counterparts, tissues in different cancer stages, different ethnicities and gender.

G. cBioPortal:

The number and location of the mutations in the peptide sequence are detected using cBioPortal (<https://www.cbioportal.org/>) [30, 31]. Frequency of alteration (Mutation, Amplification, Deep Deletion and Multiple Alterations) are investigated using the cBioPortal web.

H. PrognoScan:

PrognoScan database helps to identify gene expression and patient prognosis of the cancer. PrognoScan is utilized in examining the effect on overall survival (OS), relapse-free survival (RFS), and diseases-free survival (DFS) by multivariate and univariate investigation of gene expression [32].

I. Kaplan–Meier plotter:



The Kaplan Meier plotter (<https://kmplot.com/analysis/>) is capable of assessing the correlation between the expression of all genes (mRNA, miRNA, protein, & DNA) and survival in 35k+ samples from 21 tumor types. Applied statistical tools include Cox proportional hazards regression and the computation of the False Discovery Rate. With 18,000 analyses per day, the KM-plotter is a worldwide reference for the discovery and validation of survival biomarkers. It is mainly used in survival analysis of cancer patients in association with gene expression [32].

J. The Cancer Genom Atlas:

The Cancer Genome Atlas (TCGA), a landmark cancer genomics program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. This joint effort between NCI and the National Human Genome Research Institute began in 2006, bringing together researchers from diverse disciplines and multiple institutions. Different cancer databases use this database as a source of data such as UALCAN, GDC portal, etc [33].

K. Enrichr:

Enrichment analysis is a popular method for analyzing gene sets generated by genome-wide experiments. Enrichr (<https://maayanlab.cloud/Enrichr/>) currently contains a large collection of diverse gene set libraries available for analysis and download. In total, Enrichr currently contains 73582803 annotated gene sets from 227 gene set libraries. New features have been added to Enrichr including the ability to submit fuzzy sets, upload BED files, improved application programming interface and visualization of the results as clustergrams. Overall, Enrichr is a comprehensive resource for curated gene sets and a search engine that accumulates biological knowledge for further biological discoveries. [34].

L. UCSC Xena:

UCSC Xena (<https://xenabrowser.net/>) allows users to explore functional genomic data sets for correlations between genomic and/or phenotypic variables. The UCSC Xena platform (<http://xena.ucsc.edu/>) allows biologists and bioinformaticians to securely analyze and visualize their private functional genomics data in the context of public genomic and clinical data sets. The Xena platform consists of a set of federated data hubs and the Xena browser, which integrates across hubs, providing one location to analyze and visualize all data. Our expanding public Xena Data Hubs currently hosts 1400+ data sets from more than 35 cancer types, as well as Pan-Cancer data sets. This data hub serves seminal cancer genomics and functional genomics data set to the scientific community, including the latest TCGA, TARGET, ICGC, and GTEx data sets. This server helps to analyze data types including somatic and germline SNPs, INDELS, large structural variants, CNV, gene-, transcript-, exon- protein-, miRNA-expression, DNA methylation, phenotypes, clinical data, subtype classifications and genomic biomarkers [35].

M. Cancer Cell Line Encyclopedia (CCLE):

CCLE (<https://sites.broadinstitute.org/ccle/>) provides public access to genomic data, analysis, and visualization for over 1,000 cancer cell lines. It contains cell line data, mutation data, gene data, proteomics data, RNA expression data, Epigenetic modification data, metabolomics data [36].

N. recount3 (uniformly processed RNA-seq):

recount3 (<https://rna.recount.bio/>) is an online resource consisting of RNA-seq gene, exon, and exon-exon junction counts as well as coverage bigWig files for 8,679 and 10,088 different studies for human and mouse respectively. By taking care of several pre-processing steps and combining many datasets into one easily-



accessible website, this server makes finding and analyzing RNA-seq data considerably more straightforward [37].

O. ArrayExpress:

ArrayExpress

(<https://www.ebi.ac.uk/biostudies/arrayexpress>) is a functional genomics data collection that stores data from high-throughput functional genomics experiments, and provides data for reuse to the research community. In line with community guidelines, a study typically contains metadata such as detailed sample annotations, protocols, processed data and raw data. Raw sequence reads from high-throughput sequencing studies are brokered to the European Nucleotide Archive (ENA), and links are provided to download the sequence reads from ENA [38].

P. GDC Portal:

Genomic Data Common Data Portal (<https://portal.gdc.cancer.gov/>) is a harmonized datasets. It is a repository and computational platform for cancer researchers who need to understand cancer, its clinical progression, and response to therapy. This is a High-quality datasets spanning 44,637 cases from cancer genomic studies such as TCGA, Human Cancer Models Initiative (HCMI), Foundation Medicine Inc.(FMI), Clinical Proteomic Tumor Analysis Consortium (CPTAC) [39]. Currently it has 83

projects, 69 primary sites, 44637 cases, 1019317 files, 22534 genes, 2934380 mutations.

Q. TIMER2:

TIMER (<http://timer.comp-genomics.org/timer/>) is a comprehensive resource for systematical analysis of immune infiltrates across diverse cancer types. This version of webserver provides immune infiltrates' abundances estimated by multiple immune deconvolution methods, and allows users to generate high-quality figures dynamically to explore tumor immunological, clinical and genomic features comprehensively. Instead of just using one algorithm, TIMER2.0 (<http://timer.cistrome.org/>) provides more robust estimation of immune infiltration levels for The Cancer Genome Atlas (TCGA) or user-provided tumor profiles using six state-of-the-art algorithms. TIMER2.0 provides four modules for investigating the associations between immune infiltrates and genetic or clinical features, and four modules for exploring cancer-related associations in the TCGA cohorts. Each module can generate a functional heatmap table, enabling the user to easily identify significant associations in multiple cancer types simultaneously. Overall, the TIMER2.0 web server provides comprehensive analysis and visualization functions of tumor infiltrating immune cells [40]. Table 1

Table 1 List of the database and server used in Cancer Bioinformatics

| Database | Data Type | Analysis Type | Dataset | Website Link |
|----------|-----------------|----------------------------|---|---|
| ONCOMINE | Gene expression | Cancer vs. normal analysis | 1. Broad Tumor scape 2. Cell Line Panel 3. DNA Copy Number 4. Multi-Cancer Panel 5. Normal Tissue Panel 6. | (https://www.ONCOMINE.org/resource/login.html) |

| | | | TCGA Datasets | |
|-------------------------|--|---|----------------------------------|---|
| GEPIA2 | Gene expression | Gene expression profile across all tumor samples and paired normal tissues | TCGA/GTEX dataset | (http://gepia2.cancer-pku.cn/) |
| GENT2 | Gene expression | Tissue-wide gene expression profile across paired tissues | Public gene expression data sets | (http://gent2.appex.kr/gent2/) |
| UALCAN | Transcript per million Transcript per million | RNA molecule was compared across different categories | TCGA | (http://ualcan.path.uab.edu/index.html) |
| UALCAN | Promoter DNA methylation | Promoter DNA methylation was compared across different categories using beta value | TCGA | (http://ualcan.path.uab.edu/index.html) |
| cBioPortal | Genomic alteration Mutation Mutation and copy number alterations | Frequency of mutation, amplification, deep deletion and multiple alterations across specific cancer was analyzed Mutation across TAP1 protein was determined for specific cancer Mutation and copy number alteration was identified in TAP1 mRNA expression for specific cancer | TCGA | (http://www.cbioportal.org/) |
| The Human Protein Atlas | Protein expression | Protein expression data was compared | The tissue atlas | (http://www.proteinatlas.org/) |

| | | | | |
|--|---------------------------------------|---|-----------------------------------|--|
| | | across different categories using immunohistochemistry | | |
| PrognScan | Gene expression and patient prognosis | Gene expression was analyzed comparing patient prognosis such as overall survival (OS), disease-free survival (DFS) and relapse-free survival for cancer and normal patient | Public cancer microarray datasets | (http://dna00.bio.kyutech.ac.jp/PrognScan/) (https://kmplot.com/analysis/) |
| R2: Genomics Analysis and Visualization Platform | Positively co-expressed genes | Identification and visualization of positively co-expressed genes of different cancers | TCGA | (https://hgserver1.amc.nl/) (http://bioinformatics.psb.ugent.be/webtools/Venn) |
| Enricher | Positively co-expressed genes | Signaling pathway and gene ontology analysis of positively co-expressed genes in KEGG human pathways 2019, Panther 2016, GO Biological Process, GO Molecular Function (2018) and GO Cellular Component ranking by p-value | TCGA | (http://amp.pharm.mssm.edu/Enrichr/enrich#) |

Bioinformatics Approach and steps in cancer detection and prognosis

The servers above are generally used for cancer research via computational approach. The steps that are taken into consideration during the data

collection, data analysis, data interpretation are described below:

A. The Gene expression in different kinds of cancer:

The first and foremost important things needed for biomarker identification is to analyze gene



expression and comparison their expression between normal and tumor cells. Oncomine, GEPIA, GENT2 are normally used for this. In this case during statistical analysis, p value considered <0.05 as significant.

B. The pattern of target gene expression in various cancer:

More than 30 cancers are till now have been documented. Gene expression patterns are most of the cases different but sometimes they have some sort similarities regarding gene expression patterns. For example, gene expression pattern of TAP1 in various cancer shows difference. Using the ONCOMINE database to observe the gene expression and fold changes, it was proceeded by considering four cancer types: breast, lung, liver, and ovarian cancers. The expression was seen to be upregulated in breast, liver, ovary, and lung cancers. The GEPIA2 tool was used to look into the expression of the TAP1 gene, where the expression levels for LIHC and OV cancers were significantly higher than the normal tissues. The expression in the primary tumor and the normal was compared using the TCGA database in the UALCAN tool. A significant overexpression for the TAP1 expression was seen in the primary tumor in comparison with the normal in BRCA, LIHC, and LUAD [41].

C. Expression analysis of Target gene with clinical characteristics:

UALCAN online database is used to analyze gene expression with clinical characteristics. As example, the expression of the TAP1 gene in normal tissue was compared with tissues in patients with different clinical outcomes for breast, liver, lung, and ovarian cancers. Overexpression of TAP1 gene in cancer patient compared to normal tissue was most in stage 2 for LUAD ($p = 3.37e-09$) and LIHC ($p = 2.97e-08$) cancer, stage 2 ($p = 1.62e-12$) and stage 4 ($p = 1.08e-03$) were among the highest for BRCA. Stage 3 ($p < 1e-12$) had the least level of

expression for TAP1 in BRCA, and stage 4 ($p = 5.96e-03$) showed the lowest level of expression for LIHC and stage 1 ($p = 1.04e-10$) and stage 4 ($p = 1.33e-02$) were among the lowest for LUAD. There was no significant TAP1 expression in ovarian cancer. All stages, for BRCA, LIHC, and LUAD, in comparison with the normal tissues had a significant elevated mRNA expression. Statistically significant overexpression of the TAP1 gene in cancer patients from different ethnicity compared to normal tissues were seen in BRCA, LIHC, and LUAD. TAP1 gene expression levels were significantly higher in females having breast cancer than normal counterparts. TAP1 expression had a significant difference for both, female vs normal, and male vs normal, for LIHC and LUAD [42].

D. Promoter methylation of gene with clinical characteristics:

The level of target gene promoter methylation with different clinical characteristics can be done using the UALCAN online database. In case of TAP1 gene, CpG probes were used to identify a correlation between the expression levels and promoter methylation. The promoter methylation of the TAP1 gene in normal tissue was compared with tissues in patients with different clinical outcomes for breast, liver, lung, and ovarian cancer. Promoter methylation was increased significantly in tumor samples than normal tissues for BRCA ($p = 4.00e-03$), LIHC ($p = 2.36e-02$), and LUAD ($p = 5.99e-03$). Compared to normal tissue, patients having stage 1 ($p = 3.25e-02$) and stage 2 ($p = 4.17e-04$) BRCA showed a significant upregulation in beta value, patients with LIHC had the highest promoter methylation in stage 1 ($p = 2.32e-02$), patients having stage 1 and 3 LUAD had a significant rise in promoter methylation than normal tissue (Compared to normal tissue, Asian and African American patients had increased promoter methylation in BRCA, and Caucasian patients had increased methylation for

LIHC and LUAD. In breast, liver, and lung cancer, TAP1 promoter methylation was significantly unchanged for females compared to normal tissue. However, the analysis suggests no specific relation of TAP1 DNA methylation and clinico-pathological subtypes [42].

E. Mutant mRNA, gene mutations, and copy number alterations of Gene:

By utilizing the cBioPortal database, genetic alteration of gene in different cancers can be studied. For example, generated database queried to observe the genetic mutation of TAP1 in 7710 specimens from 12 cases of BRCA, LIHC, LUAD and OV cancers. Of the total queried samples, there was a 2% alteration in the gene set or pathways with the somatic mutation frequency of 0.3%. Considering multiple sample studies, in total 21 mutations, including 9 duplications, were reported for the TAP1 gene area. We observed between 1 and 808 amino acids in the TAP1 pro-peptide and TAP1 domain for the query. Amidst those mutations, 19 were missense, and 2 truncating mutations were identified thoroughly. Breast adenocarcinoma and lung cancer had the highest level of mutations found in them, and the mutations laid among a hotspot in R547C/H. Two mutations were identified in the R547C/H. The site contained mutations, such as missense mutations, that were discovered in 8 breast adenocarcinoma specimens, 3 lung cancer expressed missense mutations. For ovarian cancer datasets, alteration frequency was found highest (>6%) among four cancer types. Consequently, it generated the expression of TAP1 mRNA (RNA Seq V2) among 12 cases of cancers by utilizing the cBioPortal. Breast cancer, 8 (7 mis-sense and 1 truncating) cases, had the highest level of mutation in mRNA expression, and subsequently, the liver cancer was next in mutation with 4 affected [42].

F. Prognostic investigation for the TAP1 expression among cancer patients:

Different categories of cancer need to be considered for the prognosis of gene mRNA expression and summarization of the data using the prognostic databases with Cox p-value of a significance of ($p < 0.05$). For example, to analyze the interaction as for the expression of TAP1 with the ratio of survival for breast, lung, liver, and ovarian cancer patients, the PrognoScan database was used where the high and low TAP1 expressing groups were defined by log-rank test. In case of breast cancer, GSE9195 and GSE1456-GPL96 sets provided data that patients with decreased expression of TAP1 gene ($n = 57$ and $n = 139$ respectively) had significant higher relapse-free survival in comparison with the ones with a greater expression of TAP1 ($n = 20$, for both). The results found for the lung cancer dataset had the same consequences as the breast cancer results, for OS and RFS. Here the GSE31210 and GSE31210 datasets of lung cancer exhibited that the low expression ($n = 180$, for both) of TAP1 mRNA exhibited a significantly higher OS in comparison with the higher TAP1 mRNA expressed ($n = 24$, for both cases) group. On the other hand, the analysis of dataset GSE26712 and GSE26712 of PrognoScan showed significantly lower OS and DFS of ovarian cancer, for the lower TAP1 mRNA expression ($n = 111$ and $n = 149$, respectively) in contrast with the higher levels of counterparts ($n = 74$ and $n = 36$, respectively). For in lower TAP1 expression group ($n = 203$ and $n = 228$ respectively) on KM plotter for liver cancer high survival ratio was reported in case of OS and RFS, respectively, compared to the higher expression of the counterparts ($n = 161$ and $n = 88$, respectively). Primarily, the data-focused that indifferent to the single ovarian cancer difference in the expression, high TAP1 expression is in a positive correlation with the low survival rate in breast, lung, and liver cancers [42].

G. Analyses of pathway and gene ontologies of co-expressed genes of gene:

Lastly, it needed to figure out the genes that positively correlated with the target. As example TAP1 gene has some co-expressed genes in BRCA, LIHC, LUAD, and OV cancer that was detected by using the R2 genomics analysis and visualization platform. The correlated genes were used in Venn Draw to draw a Venn diagram giving us the common correlated genes in BRCA (6058 genes), LIHC (3773 genes), LUAD (4012 genes), and OV cancer (2901 genes). Then it was extracted the positively correlated common genes to conduct an ontology investigation. We used the Enrichr software in order to understand which signaling pathways were influenced by the positively co-expressed genes and the TAP1 gene in BRCA, LIHC, LUAD, and OV. In the pathway analysis of the KEGG database, We saw that the most correlated 10 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of TAP1 and the gene which are in positive correlation with TAP1, were primarily associated with cytokine-cytokine receptor pathway, chemokine signaling pathway, hematopoietic cell lineage, primary immunodeficiency, rheumatoid arthritis, cell adhesion molecules, Chagas disease, toll-like receptor signaling pathway, Salmonella infection, and human immunodeficiency virus-1 infection. Cytokine-cytokine receptor pathway and chemokine signaling pathway being the most significant pathways to be influenced [42].

DISCUSSION:

A. Bioinformatics in Lung Cancer:

At present, the diagnosis and treatment of NSCLC is still far from satisfactory, and the number of this case is still rising year by year. It is necessary to investigate the pathogenesis and biomarker of NSCLC to provide effective treatment. Great progress has been made on the mechanism of initiation and development of NSCLC. Many experiments including vitro tumor cell lines,

animal tumor models, and patients' tumor model have been done, however, NSCLC demands more comprehensive analysis because the progress of lung cancer is a multi-stage and multi-cause process. Fortunately, with the development of human genome sequencing, the high throughput and associated tumor database developed and were more available to get. The integration of data by bioinformatics analyses from multiple datasets has become a vital source of data for studies of lung cancer [43].

B. Bioinformatics in Breast Cancer:

Using serial analysis of gene expression (SAGE), cell-type-specific cell surface indicators, and magnetic resonance imaging, Allinen et al. characterized the full transcriptome of every cell type making up healthy breast tissue as well as in situ and metastatic breast carcinomas. Beads were utilized for the quick isolation of each step. Their findings imply that all cell groups undergo modifications, but only cancer epithelial cells have genetic changes identified [44].

C. Bioinformatics in Liver Cancer:

After lung cancer, liver cancer accounts for the majority of cancer-related fatalities. Using a mouse hepatoblast model and RNAi, Sawey et al. carried out a forward genetic screening under the direction of human hepatocellular cancer amplification data. They discovered that the selected susceptibility to FGF19 inhibition was caused by overexpression. Since CCND1 and FGF19 are both equally significant driver genes of the 11q13.3 amplicons in liver cancer, 11q13.3 amplifications may serve as a useful biomarker for individuals who are expected to respond favorably to anti-FGF19 treatment [45].

D. Bioinformatics in Brain Tumors:

By employing this categorization to separate medulloblastomas from other histologically comparable brain tumors, Pomeroy et al. were able to anticipate the treatment activity of



medulloblastomas [46]. Moreover, the molecular profile demonstrated that medulloblastomas and primitive neuroectodermal tumors (PNETS), two forms of brain cancers frequently regarded as single entities, are physiologically separate from one another. The medulloblastoma gene expression pattern revealed unanticipated participation of the sonic hedgehog signaling pathway and suggested cerebellar granule cells as their cell of origin. Bredel et al. also applied genomic network knowledge to the investigation of critical activities and addressing glioma genesis to employ gene transcription profiling in the biology view of human gliomas [47]. Several cancer investigators have used microarray technology to study multiple myeloma. The morphological uniformity of multiple myeloma was verified by Claudio et al. [48]. Locati et al. used self-organizing map techniques to organize publicly accessible HPV+ cancer information and derived gene signatures related to three different subgroups of the illness as a substitute strategy and to truly comprehend how gene groups may connect with prognosis [49]. A 10-gene proposed methodology of relapse duration and prognosis in epithelial ovarian cancer was discovered by Lu et al. using a support vector machine learning algorithm to examine the information from the cancer cell line encyclopedia. This model was verified on two different pieces of information [50].

E. Bioinformatics in Oral Cancer Detection:

Oral cancer is one of the most commonly seen neoplasms in the head and neck region and has a poor prognosis, and among oral cancer, oral squamous cell carcinoma (OSCC) is the most common [51,52]. Kumar et al. compared the top-ranked genes with the genes corresponding to strongly enriched GO keywords relevant to oral cancer. A total of 39 prospective oral cancer target genes were identified. Initial analysis of research and experimental data revealed 29 genes to be

associated with OSCC. They proposed a function for the chosen candidate genes in the invasion and metastasis in OSCC following a thorough pathway analysis. Using immunohistochemistry (IHC), they further verified their hypotheses and discovered that in the OSCC specimens, FLNA was elevated, whereas ARRB1 and HTT were downregulated [52]. Nakashima et al. found that the miR-1290 expression level in the plasma of oral squamous cell carcinoma patients is lower than in healthy individuals. However, circulating miR-1290 status has been proposed as a promising biomarker for evaluating both overall survival and clinical response to chemo radiotherapy in patients with oral squamous cell carcinoma [53]. Differentially expressed genes, hub proteins, and pathways demonstrated a strong correlation with OSCC development. Thorough investigation using bioinformatics is necessary for understanding the underlying process of OSCC advancement. Important genes and pathways may serve as OSCC therapy management goals [54].

F. Bioinformatics in Ovarian Cancer:

Ovarian cancer is the most prevalent and main cause of female death globally among numerous gynecological cancers. The pathophysiology and underlying causes of illness development remain unexplained despite substantial investigation. Different non-coding RNAs have been recognized as key regulators in the development of ovarian cancer. Beg et al. highlighted the significance of several ncRNAs, which have a strong promise as a therapeutic strategy for the treatment of ovarian cancer [55].

G. Bioinformatics in Leukemia:

This study screened DEGs through the transcriptome sequencing results of AML samples. The GO enrichment analysis illustrated that the DEGs of AML were significantly enriched in functional items such as positive regulation of cell promotion, protein, and immune



response. These items are closely related to the occurrence and function of tumors. The key process of malignant tumor proliferation is positive regulation of cell proliferation, and proteolysis and immune response inhibition are involved in tumor proliferation and invasion. The KEGG pathway enrichment analysis indicated that AML DEGs were significantly enriched in p53 signal pathway, TNF signal pathway, HIF-1 signal pathway, and other signal pathways. These pathways are common pathways for the progression of malignant tumors and are involved in regulating gastric cancer, bladder cancer, lung cancer, liver cancer, leukemia, and other malignant tumors. The results of GO enrichment analysis and KEGG pathway enrichment analysis showed that the DEGs of AML screened of this study were representative, which may be the key pathogenic genes of AML, participating in the regulation of tumor cell proliferation and invasion as well as in the occurrence and progression of AML [56].

H. Bioinformatics in Prostate Cancer:

Prostate cancer (PCa) is one of the pervasive carcinoma occurring in men and a large health burden worldwide [57]. The lethality of PCa is due to the lack of treatment options that can produce a lasting response at the genetic and cellular biological levels. About 15% of patients with PCa are diagnosed with high-risk disease [58]. Therefore, through utilizing univariate Cox and iterative lasso Cox regression analyses, a 3-gene (KCNK3, AK5, and ARHGEF38) risk signature model in PCAs was constructed. The ROC curves further approved the accuracy of our risk model. It is reported that KCNK3 influenced physiological processes, ranging from vascular tone to metabolic diet through inflammation [59]. Also, KCNK3 was correlated with prolonged survival after surgery in colorectal cancer [60]. AK5 was reported as a new prognosis marker that promotes autophagy and proliferation in human

gastric cancer [61]. Interestingly, ARHGEF3 was proved to be an oncogene and may be a novel biomarker for predicting invasive PCa [61, 62].

CONCLUSION:

Although these tools provide great convenience for prognostic biomarker development, several key aspects of these tools remain elusive. Differences in datasets collected and split points may result in significantly different results. So, it is essential to collect dataset and their source of these web servers and find excluding TCGA data and there are significant differences in other data sources. This may be one of the reasons why the analysis results of different tools are not completely consistent. In the future, efforts should be made in data optimization, prognostic tools should be improved to be able to predict multi-gene markers, select optimal cut-off computation, use hierarchical clustering and consider complex multi-omics networks of interactions. In addition, more molecular subtypes and clinical information including tumor tissue image and treatment data should be collected and mined to identify more meaningful prognostic markers through more detailed subtype analysis.

REFERENCES

1. Umar A: Applications of bioinformatics in cancer detection: a lexicon of bioinformatics terms. *Ann N Y Acad Sci.* 2004, 1020:263-76. 10.1196/annals.1310.021
2. Tao Z, Shi A, Li R, Wang Y, Wang X, Zhao J : Microarray bioinformatics in cancer- a review. *J BUON.* 2017, 22:838-43.
3. Wu D, Rice CM, Wang X: Cancer bioinformatics: a new approach to systems clinical medicine. *BMC Bioinformatics.* 2012, 13:71. 10.1186/1471-2105-13-71
4. Brenner C: Applications of bioinformatics in cancer. *Cancers (Basel).* 2019, 11:10.3390/cancers11111630
5. Singer J, Irmisch A, Ruscheweyh HJ, et al.: Bioinformatics for precision oncology. *Brief*



- Bioinform. 2019, 20:778-88. 10.1093/bib/bbx143
6. Hauptman N, Glavač D: Long non-coding RNA in cancer. *Int J Mol Sci.* 2013, 14:4655-69. 10.3390/ijms14034655
 7. Lander ES, Linton LM, Birren B, et al.: Initial sequencing and analysis of the human genome . *Nature.* 2001, 409:860-921. 10.1038/35057062
 8. Stein LD: Human genome: end of the beginning. *Nature.* 2004, 431:915-6. 10.1038/431915a
 9. Rubin G, Vedsted P, Emery J: Improving cancer outcomes: better access to diagnostics in primary care could be critical. *Br J Gen Pract.* 2011, 61:317-8. 10.3399/bjgp11X572283
 10. de Matos MR, Posa I, Carvalho FS, Morais VA, Grosso AR, de Almeida SF: A systematic pan-cancer analysis of genetic heterogeneity reveals associations with epigenetic modifiers. *Cancers (Basel).* 2019, 11:10.3390/cancers11030391
 11. Lee HC, Lai K, Lorenc MT, Imelfort M, Duran C, Edwards D. Bioinformatics tools and databases for analysis of next-generation sequence data. *Brief Funct Genomics.* 2012 Jan;11(1):12-24. doi: 10.1093/bfgp/elr037. Epub 2011 Dec 19. PMID: 22184335.
 12. Lalnunsanga J, Lalruatzela, Lalduhawma J, Chhabra KG, Madhu PP: Determining the risk factor of oral cancer on chewing smokeless tobacco kharra: a case study. *Eur J Cancer Clin Oncol.* 2021, 3:001-002.
 13. Lohe V, Kadu R: Evaluation of age of earning and starting of adverse habit in patients visiting to dental hospital. *J Datta Meghe Inst Med Sci Univ.* 2021, 16:437-40. 10.4103/jdmimsu.jdmimsu_129_19
 14. Gadbail AR, Chaudhary M, Gawande M, et al.: Oral squamous cell carcinoma in the background of oral submucous fibrosis is a distinct clinicopathological entity with better prognosis. *J Oral Pathol Med.* 2017, 46:448-53. 10.1111/jop.12553
 15. Wang X: Role of clinical bioinformatics in the development of network-based biomarkers . *J Clin Bioinforma.* 2011, 1:28. 10.1186/2043-9113-1-28
 16. Yigitoglu B, Uctepe E, Yigitoglu R, Gunduz E, Gunduz M: Bioinformatics in breast cancer research. A concise review of molecular pathology of breast cancer. Gunduz M (ed): IntechOpen, 2014. 175-85. 10.5772/59519
 17. Hanauer DA, Rhodes DR, Sinha-Kumar C, Chinnaiyan AM: Bioinformatics approaches in the study of cancer. *Curr Mol Med.* 2007, 7:133-41. 10.2174/156652407779940431
 18. Reza MS, Siddika MA, Rahman K, Adhikary J: Role of bioinformatics in cancer diagnosis, prognosis and therapies. *IJBTCM.* 2023, 2:01-10.
 19. Ushasri K, Prasad AR, Reddy JK, Saravana S: Significance of data mining in bioinformatics . *Int J Eng Res Technol.* 2014, 1:86-8. 10.17577/IJERTCONV2IS15004
 20. Valand JH, Twine D, Kyomukamaa M, et al.: Role of bioinformatics in cancer diagnosis . *Research Square.* 2022, 1-15. 10.21203/rs.3.rs-1299906/v1
 21. Charoentong P, Angelova M, Efremova M, Gallasch R, Hackl H, Galon J, Trajanoski Z: Bioinformatics for cancer immunology and immunotherapy. *Cancer Immunol Immunother.* 2012, 61:1885-903. 10.1007/s00262-012-1354-x
 22. Rhodes DR et al (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6(1):1–6
 23. Samad A, Jafar T, Rafi JH (2020) Identification of angiotensin converting enzyme 2 (ACE2) protein as the potential biomarker in SARS-CoV-2 infection-related

- lung cancer using computational analyses. *Genomics* 112(6):4912–4923
24. Samad A et al (2020) Computational assessment of MCM2 transcriptional expression and identification of the prognostic biomarker for human breast cancer. *Heliyon* 6(10)
25. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 45(W1):W98–W10
26. Shin G, Kang T-W, Yang S, Baek S-J, Jeong Y-S, Kim S-Y (2011) GENT: gene expression database of normal and tumor tissues. *Cancer Inform* 10:CIN. S7226
27. Chandrashekar DS et al (2017) UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19(8):649–658
28. Uhlén M et al (2015) Tissue-based map of the human proteome. *Science* 347(6220):1260419
29. Rafi JH et al (2021) High expression of bone morphogenetic protein 1 (BMP1) is associated with a poor survival rate in human gastric cancer, a dataset approaches. *Genomics* 113(1, Part 2):1141–1154
30. E. Cerami et al. 2012 “The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data,” ed: AACR.
31. Gao J et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269):pH
32. Mizuno H, Kitada K, Nakai K, Sarai A (2009) PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genet* 2(1):18.
33. Han, Namshik (2018). The Cancer Genome Atlas (TCGA) RNA-seq meta-analysis. *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.5851743.v1>
34. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016 Jul 8;44(W1):W90-7. doi: 10.1093/nar/gkw377. Epub 2016 May 3. PMID: 27141961; PMCID: PMC4987924.
35. Goldman, Mary & Craft, Brian & Zhu, Jingchun & Haussler, David. (2017). Abstract 2584: The UCSC Xena system for cancer genomics data visualization and interpretation. *Cancer Research.* 77. 2584-2584. 10.1158/1538-7445.AM2017-2584.
36. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012 Mar 28;483(7391):603-7. doi: 10.1038/nature11003. Erratum in: *Nature.* 2012 Dec 13;492(7428):290. Erratum in: *Nature.* 2019 Jan;565(7738):E5-E6. doi: 10.1038/s41586-018-0722-x. PMID: 22460905; PMCID: PMC3320027.
37. Wilks, C., Zheng, S.C., Chen, F.Y. et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome Biol* 22, 323 (2021). <https://doi.org/10.1186/s13059-021-02533-6>.
38. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lusk M, Mani R, Rayner T, Sharma A, Williams E, Sarkans U, Brazma A. ArrayExpress--a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D747-50. doi: 10.1093/nar/gkl995. Epub 2006 Nov 28. PMID: 17132828; PMCID: PMC1716725.



39. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. *Blood*. 2017 Jul 27;130(4):453-459. doi: 10.1182/blood-2017-03-735654. Epub 2017 Jun 9. PMID: 28600341; PMCID: PMC5533202.
40. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020 Jul 2;48(W1):W509-W514. doi: 10.1093/nar/gkaa407. PMID: 32442275; PMCID: PMC7319575.
41. Zheng H, Zhang G, Zhang L, Wang Q, Li H, Han Y, Xie L, Yan Z, Li Y, An Y, Dong H, Zhu W and Guo X (2020) Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis. *Front. Oncol*. 10:68. doi: 10.3389/fonc.2020.00068.
42. Tabassum, A., Samdani, M.N., Dhali, T. et al. Transporter associated with antigen processing 1 (TAP1) expression and prognostic analysis in breast, lung, liver, and ovarian cancer. *J Mol Med* 99, 1293–1309 (2021). <https://doi.org/10.1007/s00109-021-02088-w>.
43. Ye X, Gao Q, Wu J, Zhou L, Tao M. Identification of significant genes in non-small cell lung cancer by bioinformatics analyses. *Transl Cancer Res*. 2020 Jul;9(7):4330-4340. doi: 10.21037/tcr-19-2596. PMID: 35117799; PMCID: PMC8799091.
44. E. Cerami et al. 2012 “The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data,” ed: AACR.
45. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013 Apr 2;6(269):pl1. doi: 10.1126/scisignal.2004088. PMID: 23550210; PMCID: PMC4160307.
46. Mizuno H, Kitada K, Nakai K, Sarai A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics*. 2009 Apr 24;2:18. doi: 10.1186/1755-8794-2-18. PMID: 19393097; PMCID: PMC2689870.
47. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W90-7. doi: 10.1093/nar/gkw377. Epub 2016 May 3. PMID: 27141961; PMCID: PMC4987924.
48. Leclerc M, Mezquita L, Guillebot De Nerville G, Tihy I, Malenica I, Chouaib S, Mami-Chouaib F. Recent Advances in Lung Cancer Immunotherapy: Input of T-Cell Epitopes Associated With Impaired Peptide Processing. *Front Immunol*. 2019 Jul 3;10:1505. doi: 10.3389/fimmu.2019.01505. PMID: 31333652; PMCID: PMC6616108.
49. Raghavan M. Immunodeficiency due to defective antigen processing: the molecular basis for type 1 bare lymphocyte syndrome. *J Clin Invest*. 1999 Mar;103(5):595-6. doi: 10.1172/JCI6455. PMID: 10074475; PMCID: PMC408134.
50. Keusekotten, Kirstin, Ralf M. Leonhardt, Sarah Ehses and Michael R. Knittler. “Biogenesis of Functional Antigenic Peptide Transporter TAP Requires Assembly of Pre-existing TAP1 with Newly Synthesized TAP2*.” *Journal of Biological Chemistry* 281 (2006): 17545 - 17551.
51. Mou MA, Keya NA, Islam M, Hossain MJ, Al Habib MS, Alam R, Rana S, Samad A, Ahammad F. Validation of CSN1S1

- transcriptional expression, promoter methylation, and prognostic power in breast cancer using independent datasets. *Biochem Biophys Rep.* 2020 Dec 2;24:100867. doi: 10.1016/j.bbrep.2020.100867. PMID: 33381666; PMCID: PMC7767798.
52. Raufaste-Cazavieille V, Santiago R, Droit A. Multi-omics analysis: Paving the path toward achieving precision medicine in cancer treatment and immuno-oncology. *Front MolBiosci.* 2022 Oct 11;9:962743. doi:10.3389/fmolb.2022.962743. PMID: 36304921; PMCID: PMC9595279.
53. Karim MA, Samad A, Adhikari UK, Kader MA, Kabir MM, Islam MA, Hasan MN. A Multi-Omics Analysis of Bone Morphogenetic Protein 5 (BMP5) mRNA Expression and Clinical Prognostic Outcomes in Different Cancers Using Bioinformatics Approaches. *Biomedicines.* 2020 Jan 21;8(2):19. doi: 10.3390/biomedicines8020019. PMID: 31973134; PMCID: PMC7168281.
54. Barman UD et al (2020) Clinicopathological and prognostic significance of GPC3 in human breast cancer and its 3D structure prediction. *Network Model Analys Health Inform Bioinform* 9(1):24.
55. Saha, Subbroto & Kader, M Ashraful & Samad, Abdus & Biswas, Krishno & Rahman, Md Aatur & Parvez, Md. Anowar Khasru & Rahman, Shahedur. (2020). Prognostic and clinico-pathological significance of BIN1 in breast cancer. 100327. 10.1016/j.imu.2020.100327.
56. Chen W, Liu D, Wang G, Pan Y, Wang S, Tang R. Screening diagnostic markers for acute myeloid leukemia based on bioinformatics analysis. *Transl Cancer Res.* 2022 Jun;11(6):1722-1729. doi: 10.21037/tcr-22-1257. PMID: 35836534; PMCID: PMC9273715.
57. Hou Y, Hu J, Zhou L, Liu L, Chen K, Yang X. Integrative Analysis of Methylation and Copy Number Variations of Prostate Adenocarcinoma Based on Weighted Gene Co-expression Network Analysis. *Front Oncol.* 2021 Apr 1;11:647253. doi: 10.3389/fonc.2021.647253. PMID: 33869043; PMCID: PMC8047072.
58. M, Hoque., SN, Saha., T, Akram., Nanogels Based Drug Delivery System: A Promising Therapeutic Strategy, *Int. J. in Pharm. Sci.*, 2023, Vol 1, Issue 11, 103-111. <https://doi.org/10.5281/zenodo.1007086>
59. West JD, Austin ED, Rizzi EM, Yan L, Tanjore H, Crabtree AL, Moore CS, Muthian G, Carrier EJ, Jacobson DA, Hamid R, Kendall PL, Majka S, Rathinasabapathy A. KCNK3 Mutation Causes Altered Immune Function in Pulmonary Arterial Hypertension Patients and Mouse Models. *Int J Mol Sci.* 2021 May 9;22(9):5014. doi: 10.3390/ijms22095014. PMID: 34065088; PMCID: PMC8126011.
60. Cavalieri D, Dolara P, Mini E, Luceri C, Castagnini C, Toti S, Maciag K, De Filippo C, Nobili S, Morganti M, Napoli C, Tonini G, Baccini M, Biggeri A, Tonelli F, Valanzano R, Orlando C, Gelmini S, Cianchi F, Messerini L, Luzzatto L. Analysis of gene expression profiles reveals novel correlations with the clinical course of colorectal cancer. *Oncol Res.* 2007;16(11):535-48. doi: 10.3727/096504007783438376. PMID: 18306933.
61. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 1999 Jan 1;27(1):29-34. doi: 10.1093/nar/27.1.29. PMID: 9847135; PMCID: PMC148090.
62. Liang X, Wang Y, Pei L, Tan X, Dong C. Identification of Prostate Cancer Risk

Genetics Biomarkers Based on Intergraded
Bioinformatics Analysis. *Front Surg.* 2022
Mar 17;9:856446. doi:
10.3389/fsurg.2022.856446. PMID:
35372462; PMCID: PMC8967941.

HOW TO CITE: Md Solayman Hossain, Farha Matin
Juliana, Bioinformatics In Early Detection And
Prognosis Of Cancer, *Int. J. of Pharm. Sci.*, 2024, Vol 2,
Issue 7, 144-161.
<https://doi.org/10.5281/zenodo.12610166>

