Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Review Article **DOI: 10.26479/2026.1203.01**

A LITERATURE REVIEW OF THE IDENTIFICATION OF CANCER BIOMARKERS THROUGH BIOINFORMATICS APPROACHES

Vivekananda[1]*, Sujata Kumari[2]

1. Department of Zoology, Patna University, Patna, Bihar, India.

2. Department of Zoology, Magadh Mahila College, Patna, Bihar, India.

ABSTRACT: Cancer remains one of the leading causes of mortality worldwide, presenting a significant challenge to global health systems. Despite advancements in treatment, early detection remains critical to reducing mortality rates and improving patient outcomes. Traditional diagnostic methods frequently fail to detect cancer at early stages, underscoring the urgent need for effective biomarkers that can aid in early diagnosis, prognosis, and treatment monitoring. Biomarkers—biological molecules altered in the presence of cancer—offer a promising solution due to their ability to reflect cancerous changes at molecular levels. This review explores the identification of cancer biomarkers using integrated bioinformatics tools and databases. Bioinformatics approaches allow for high-throughput screening and analysis of differentially expressed genes, proteins, and metabolites, facilitating the discovery of novel biomarkers. Key tools and databases such as Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), and STRING provide critical insights into the molecular underpinnings of various cancer types. The review highlights biomarkers linked to the five most prevalent cancer types (breast, lung, colorectal, liver and gastric) globally, emphasizing their roles in early detection and targeted therapy. Through a detailed examination of existing literature, this review suggests that integrating bioinformatics approaches into biomarker discovery can revolutionize cancer diagnosis and treatment. Future research should focus on validating these biomarkers in clinical settings, enabling the development of precise, sensitive, and effective diagnostic technologies. By leveraging these advancements, healthcare systems can improve early detection rates and ultimately reduce cancer-related mortality.

Keywords: Cancer biomarkers, cancer therapy, computational biology, differentially expressed genes.

Article History: Received: May 08, 2025; Revised: May 14, 2026; Accepted: May 18, 2026.

Corresponding Author: Dr. Vivekananda* Ph.D.
Department of Zoology, Patna University, Patna, Bihar, India
Email Address: acevivek2223@gmail.com

1. INTRODUCTION

Cancer is the leading cause of death worldwide. It accounts for one in four deaths (22.8%) globally from non-communicable diseases. It also causes significant social and financial losses, collectively termed the "cancer burden". The number of new cases of cancer is expected to rise to 29.9 million and the number of cancer-related deaths to 15.3 million by 2040[1]. The global burden of cancer is a significant public health challenge, with increasing incidence and mortality rates observed worldwide. According to the latest estimates from the International Agency for Research on Cancer (IARC), in 2022, there were approximately 20 million new cancer cases and 9.7 million cancer-related deaths globally. These figures highlight a growing need for comprehensive cancer care services, particularly in underserved populations where access to treatment is limited [2]. The most common cancers are breast cancer, lung cancer, colon cancer, skin cancer, leukemia, bladder cancer, non-Hodgkin lymphoma, pancreatic cancer and liver cancer[3]. In developed vs developing countries, overall incidences of cancer were 2- to 3-fold greater for both sexes in developed countries. Female breast and cervical cancer death rates, on the other hand, were much higher in developing nations than in developed countries[4]. India is ranked third in the world in terms of the number of cancer cases. According to the National Cancer Registry Program statistics, over 13 lakh persons in India are diagnosed with cancer each year. The Indian Council of Medical Research (ICMR) estimates that there will be a 12% rise in cancer cases in India by the next five years[5]. Identifying cancer at an initial stage, often before symptoms appear, can significantly enhance the treatment outcomes and survival rates for many types of cancer. Early detection of cancer can be done through biomarkers. They are biological molecule that are found in blood, other body fluids, or tissues that is a sign of a normal condition or disease condition. Cancer biomarkers are produced either by cancerous tissue or by the body reactions to cancer during cancer development and progression. They have important and promising applications in cancer screening, early diagnosis, prognosis prediction, recurrence detection, and therapeutic efficacy monitoring[6].

Cancer cell exhibits various genetic changes as compared to normal cell. Expression of multiple genes altered in many cancer tissues. These changes can be detected in majority of patients with a particular type of tumor, these changes can be used as potential biomarkers for detection of cancer[7]. Substantial advances in cancer biology have resulted in the discovery of various cancer biomarker that are linked to cancer progression and development. These can be identified using transcriptomics, metabolomics, genomics and proteomics. Living cells have a finite life span, and their DNA

transcribes into RNA, which upon translation produces proteins effecting numerous physiological and metabolic processes required by the body. Any change in these mechanisms, such as a mutation in DNA, causes disruption which leads to a disease. The detection of mutations in DNA can be used to predict Cancer risk[8]. Consequently, measurement of RNA, protein, and metabolite expression levels can provide important information about illness progression and profiling. In this review, we aim to discuss the biomarkers identification using integrated bioinformatics analysis. There are more than 200 types of cancer reported, however in this review, we will discuss about the various biomarkers associated with five most prevalent cancer types in the world, which can be exploited in designing of sensitive and effective diagnostic technology for early detection of cancer. These biomarkers include proteins that are down-regulated or up-regulated during carcinogenesis, circulating nucleic acids or cells, metabolites, and the recently discovered biomarkers found using quantitative proteomics. Our particular emphasis will be on novel biomarkers that are expected to impact the cancer screening program, ultimately increasing early-stage cancer detection and decreasing mortality. Besides, a brief insight on integrated bioinformatics analysis tools and databases have also been discussed.

2. MATERIALS AND METHODS

2.1 Integrated bioinformatics analysis tools and databases

Various integrated bioinformatics databases have been utilized for the identification of prognostic biomarkers in the treatment of various kinds of cancer. Some of which have been enlisted in Table 1 along with database links. The biomarkers associated with different types of Cancers identified with the help of integrated bioinformatics tools depicted in Figure 1.

Table 1. Bioinformatics tools and databases used for identification of biomarkers.

S.No.	Databases	Bioinformatics tools	Link/url
1	Gene Expression Omnibus (GEO)	GEO2R	https://www.ncbi.nlm.nih.gov/geo/
2	The Cancer Genome Atlas (TCGA)	cBioPortal	https://www.cbioportal.org/
3	STRING Database	STRING	https://string-db.org/
4	UniProt	BLAST	https://www.uniprot.org/
5	KEGG (Kyoto Encyclopedia of Genes and Genomes)	KEGG Mapper	https://www.kegg.jp/

2.2 Microarray and RNASeq data collection

The microarray data collection is done using the GEO database which refers to Gene Expression Omnibus. It is easily accessed via online medium using http:// www.ncbi.nlm.nih.gov/geo/link. The

GEO database is basically being used to obtain high-throughput gene expression profiles of PTC (Papillary, Thyroid Carcinoma) and normal thyroid tissues. Independent datasets are chosen, and they are all based on the specified platforms, including the relevant tissues. Various microarray datasets have been collected using the GEO database and then processed with bioinformatics to discover hub genes. New technologies have emerged for the analysis of gene expression and for the identification of cancer biomarkers. One such technology is RNASeq technology which is the most up to date technology to analyze gene expression. With the use of NGS (Next generation genome sequencer) the gene expression profile analysis carried out. The first stage in the process is to convert the population of RNA to be sequenced into complementary DNA (cDNA) fragments which are present in biological sample. This is accomplished using reverse transcription, allowing the RNA to be used in an NGS procedure. After that, the cDNA is fragmented, and adapters are attached to each fragment's end. The functional elements present on adapters which allowed sequencing. The cDNA library is evaluated by NGS after amplification, size selection, cleanup, and quality verification, yielding short sequences that correspond to all or part of the fragment from which it was formed. The extent to which the library is sequenced is determined by the intended use of the output data. After completing the RNA sequencing technology workflow, the data can be matched to a reference genome if one is available or built from scratch to provide an RNA sequence map that encompasses the transcriptome.

2.3 Screening of DEGs

The GEO2R program, which could be easily accessed via http://www.ncbi.nlm.nih.gov/geo/geo2r/link, is used for the detection of these differentially expressed genes which are known as DEGs. Further, R package Limma has been utilized to screen out these DEGs.

2.4 Enrichment analysis via GO and KEGG pathway

Followed by the screening of DEGs, the enrichment analysis using GO and KEGG pathway is performed using the database for Annotation, Visualization and Integrated Discovery, commonly known as DAVID database (http://david.abcc.ncifcrf.gov/). This process includes biological processes, cellular components, molecular function and KEGG pathway analysis. Further, the GO plot package of R could be used to display the results of analysis and the pathway analysis results can also be analyzed using the clueGO plug-ins of cytoscape software 3.7.2[9].

2.5 Construction of the PPI network and analysis of the module

After the enrichment analysis, the PPI network is being built upon using the STRING[10] (http://www.bork.embl-heidelberg.de/STRING/) database which refers to Search Tool for the Retrieval of Interacting Genes/Proteins, to uncover DEG associations based on minimum prescribed interaction scores. Followed by this, using the Cytoscape (http://www.cytoscape.org/) database, the PPI network is then analyzed and visualized. Additionally, MCODE is also one such bioinformatics

tool utilized to screen the PPI network's main module.

2.6 Survival analysis and validation of hub gene expression

The Cancer Genome Atlas (https://tcga-data.nci.nih.gov/tcga/) is utilized to examine the association between gene expression and survival of patients. RNA expression data from hundreds of samples from the TCGA and GTEx projects was analyzed using the Gene Expression Profiling Interactive Analysis tool (GEPIA) (http://gepia2021.cancer-pku.cn/). Additionally, Human Protein Atlas, and Kaplan-Meier plotter tool databases could also be used to look at the translational and protein expression levels, as well as survival statistics, of DEGs. Apart from this, miRWalk[11] and TargetScan, were used to predict the corresponding change in the expression level of microRNAs in DEGs and the expression profiles were evaluated using OncomiR[12]. Finally, RT-qPCR is used to confirm the expression of new DEGs. Hence, the constructed biomarkers could be considered potential candidates for various kinds of Cancers.

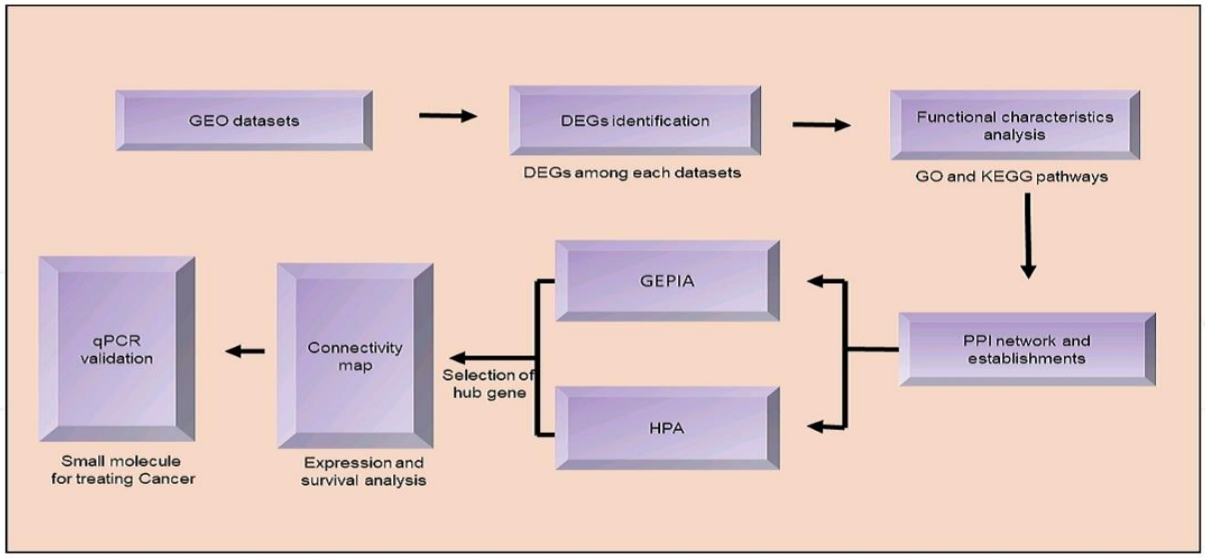

Figure 1: Workflow for identifying potential small molecules to treat cancer: starting from GEO datasets, DEGs are identified and analyzed for functional characteristics through GO and KEGG pathways. Hub genes are selected using GEPIA and HPA tools, followed by PPI network construction. Connectivity maps and expression/survival analyses lead to qPCR validation of small molecule candidates.

3. Biomarkers associated with different types of cancer

Biomarkers play a crucial role in cancer diagnosis, prognosis, and treatment, providing insights into the underlying mechanisms of various cancers. Advanced bioinformatics tools facilitate the identification and validation of these biomarkers, enabling researchers to discover potential candidates for targeted therapies. They analyze large datasets from genomic, transcriptomic, and proteomic studies to uncover patterns associated with different cancers. For instance, biomarkers such as KRAS mutations in colorectal cancer predict poor responses to certain therapies, while HER2 overexpression in breast cancer indicates responsiveness to targeted treatments like trastuzumab[13], [14].

Utilizing comprehensive databases allows researchers to correlate specific biomarkers with clinical outcomes; for example, PD-L1 expression is critical for predicting responses to immune checkpoint inhibitors[15]. The strategic application of bioinformatics not only aids in understanding cancer mechanisms but also enhances therapeutic decision-making, ultimately improving patient outcomes through personalized medicine approaches[16].

3.1 Prostate Cancer

As one of the leading healthcare concerns worldwide, prostate cancer has become one of the most prevalent malignancies of adult males with an incidence of typically 100-150 per 100,000 men (~0.01%) [17]. Although numerous progresses have been made to uncover the molecular mechanisms of prostate cancer development and progression, the outcomes remain obscure and the inconsistencies among different published studies are obvious. The main reason for such phenomenon is generally thought to be the complexity and heterogeneity nature of prostate cancer[18]. Thus, it's of vital significance to perform further research in this regard to validate and update the information acquired. In a study by Zihao He et al., screened a total of 252 (186 up-regulated and 66 down-regulated) DEGs from the unused GEO dataset GSE103512. The GO/KEGG pathway enrichment analyses were implemented thereafter. Up-regulated DEGs were mainly enriched in metabolic pathways, FA metabolism, and PPAR signaling pathway, whereas down-regulated DEGs were mostly associated with protein digestion and absorption. Kallikrein-related peptidase 3, cadherin1(CDH1), Kallikrein-related peptidase 2 (KLK2), fork head box A 1(FOXA1), and epithelial cell adhesion molecule (EPCAM) were identified as hub genes from the PPI network. CDH1, FOXA1, and EPCAM were validated by other relevant gene expression omnibus datasets[19].

In another study by Rui Wang et al., the top five significantly overexpressed mRNA (AMACR, PPP1R14b, PCA3, DLX1, and RPL22L1) and the top five significantly under expressed mRNA (DUOX1, EFS, GSTP1, S100A16, and NCRNA00087) were selected for further validation in prostate cancer patients and healthy donors by qRT-PCR. The results showed that AMACR, DLX1, PCA3, DUOX1, and GSTP1 mRNA were stably amplified in plasma. Additionally, DLX1, PCA3, DUOX1, and GSTP1 mRNA expression was significantly different between prostate cancer circulating free mRNA samples and healthy donors. These mRNAs may be useful biomarkers for prostate cancer diagnosis[20].

In another study, Farhad Kosari et al., described a process for the identification of genes that can report on the aggressiveness of prostate tumors and thereby add to the information provided by current pathologic analysis. Expression profiling data from over 100 laser capture microdissection derived samples from non-neoplastic epithelium; Gleason patterns 3, 4, and 5 and node metastasis prostate cancer were used to identify genes at abnormally high levels in only some tumors. These variably overexpressed genes were stratified by their association with aggressive phenotypes and

were subsequently filtered to exclude genes with redundant expression patterns. Selected genes were validated in a case-control study in which cases (systemic progression within 5 years) and controls (no systemic progression at 7 years of follow-up) were matched for all clinical and pathologic criteria from time of prostatectomy (n = 175). The most prominent candidates were SSTR1and genes related to proliferation, includingTOP2A. The process described by Farhad Kosari et al., have identified genes that add information not available from current clinical measures and can improve the prognosis of prostate cancer[21].

Similarly, in another study by Teppei Iwata et al., identified that a subset of prostate cancer displayed a poor clinical outcome. Therefore, identifying this poor prognostic subset within clinically aggressive groups (defined as a Gleason score (GS) ≧8) and developing effective treatments were essential for improving prostate cancer survival. Here, they performed a bioinformatics analysis of a TCGA dataset (GS ≧8) to identify pathways up-regulated in a prostate cancer cohort with short survival. They identified pathways involving up-regulation of GRB2. Overexpression of GRB2 was linked to shorter survival in the TCGA dataset, a finding validated by histological examination of biopsy samples taken from the patients for diagnostic purposes. Thus, GRB2 is a novel biomarker that predicts shorter survival of patients with aggressive prostate cancer (GS ≧8)[22].

In another study Chipampe Patricia Lombe et al., obtained metastatic prostate cancer associated microRNA array profiles from the GSE28029 dataset in the GEO database. MicroRNA target prediction was done using the databases, TargetScanHuman, miRDB and DIANA microT, six target genes (FOXC1, CDKN1A, BIRC2, CTNND1, ELK1 and LRP8) were found to be common among the three different databases. Their study suggested that CDKN1A, FOXC1 and BIRC2 might be core genes for prostate cancer that play an important role in its diagnosis, development and progression[8].

Table 2. List of prostate cancer biomarkers

S. No.	Biomarkers Identified	Reference Number
1	KLK2, CDHI, FOXA1, and EPCAM	[19]
2	DLX1, PCA3, and DUOX1	[20]
3	SSTR1, TOP2A	[21]
4	GRB2	[22]
5	CDKN1A, FOXC1 and BIRC2	[8]

3.2 Lung cancer

Lung cancer is one of the most common causes of the cancer-related death around the globe. Despite great attempts to enhance treatment approaches in previous decades, the clinical outcome of traditional therapies such as surgery, radiation, and chemotherapy remains poor when compared to

other major forms of cancer such as colon, prostate, and breast cancers. The challenges in making an early-stage diagnosis of lung cancer and the high recurrence rate after curative treatments are the main reasons for the lack of improvement in prognosis [28]. To improve the clinical result of lung cancer treatments, it is critical to identify and validate diagnostic and prognostic biomarkers.

Lung adenocarcinoma (LUAD), which accounts for 60% of non-small-cell lung cancers, is poorly diagnosed and has a low average 5-year survival rate (approximately 20%). It remains the leading cause of cancer-related deaths worldwide. Studies on long noncoding RNAs (lncRNAs) in LUAD-related competing endogenous RNA (ceRNA) networks are limited.

In a study by Jili Hou and Cheng Yao, the researchers aimed to identify novel prognostic biomarkers for LUAD using bioinformatics tools and data analysis. They systemically integrated differentially expressed genes and clinically significant modules using weighted correlation network analysis. They performed a functional analysis of the collected candidate genes and explored three LUAD-related genes (VWF, PECAM1, and COL1A1) associated with the overall survival rates of patients with LUAD. Based on Cox proportional hazards analysis of candidate mRNAs and lncRNAs together with differentially expressed microRNAs, they constructed ceRNA networks, obtained 12 lncRNAs in the ceRNA networks, and revealed seven novel lncRNAs AC021016.2, AC079630.1, AC116407.1, AC125807.2, AF131215.5, LINC01936, and RHOXF1-AS1.

These lncRNAs were found to be associated with overall survival rates and are suitable for the prediction of prognosis by Kaplan-Meier survival and receiver operating characteristic curve analyses. In particular, three lncRNAs—AF131215.5, AC125807.2, and LINC01936—showed an independent prognostic value of overall survival for patients with LUAD. They evaluated the diagnostic capabilities of seven lncRNAs for patients with LUAD using principal component analysis and the Gene Set Variation Analysis index. lncRNAs and crucial genes could be effectively used for distinguishing LUAD tumors from normal tissues in the Gene Expression Omnibus profile. In particular, AC021016.2 showed a significant prognostic value in the validation dataset. Their findings revealed the significance of exploring lncRNAs in cancer-related ceRNAs using bioinformatics strategies[23].

In another study by Bumjin Kim et al., identified candidate biomarkers for lung cancers through bioinformatics analysis of the public SAGE and EST data and validated their potential using clinical specimens. CBLC and CYP24A1 are two particularly promising biomarkers for non–small-cell lung cancer. Other genes (ALDH3A1, AKR1B10, and LOC147166) seem to have fair potential as well. It was interesting to note the origin of biomarker genes. Two genes (CYP24A1 and S100P) were derived from SAGE data and others (CBLC, ALDH3A1, AKR1B10, and LOC147166) were from the EST data. This implied different ranges of coverage for the two data sets and the benefits of using both types of data. Their study also showed that candidates from meta-analysis of the public expression data should be carefully tested through validation using clinical samples. One of the

major strengths of this study was the use of multiple clinical samples. Strong statistical support was thus possible although additional clinical samples should be used for further validation down the road. In addition, they tested only 20 genes in this study with several hundreds of candidates examined. Biochemical studies for promising biomarkers were necessary as well to examine the potential of the candidate genes as drug targets. As additional expression data become available, it would also be interesting to see if combinations of several differentially regulated genes could function with more sensitivity and specificity in the diagnosis and prognosis of lung cancers[24].

Similarly, in another study by Bai Dai et al., non-small-cell lung cancer (NSCLC) accounts for >85% of lung cancers, and its incidence is increasing. they explored expression differences between NSCLC and normal cells and predicted potential target sites for detection and diagnosis of NSCLC. Three microarray datasets from the Gene Expression Omnibus database were analyzed using GEO2R. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis were conducted. Then, the String database, Cytoscape, and MCODE plug-in were used to construct a protein–protein interaction (PPI) network and screen hub genes. Overall and disease-free survival of hub genes were analyzed using Kaplan-Meier curves, and the relationship between expression patterns of target genes and tumor grades were analyzed and validated. Gene set enrichment analysis and receiver operating characteristic curves were used to verify enrichment pathways and diagnostic performance of hub genes. In total, 293 differentially expressed genes were identified and mainly enriched in cell cycle, ECM–receptor interaction, and malaria. In the PPI network, 36 hub genes were identified, of which 6 were found to play significant roles in carcinogenesis of NSCLC: CDC20, ECT2, KIF20A, MKI67, TPX2, and TYMS. The identified target genes can be used as biomarkers for the detection and diagnosis of NSCLC. The Differentially Expressed Genes (DEGs) for lung cancer were isolated from Gene Expression Omnibus (GEO) database using R software tool GEO2R. A total of 407 DEGs (254 up-regulated and 153 down-regulated) from non-treatment studies and 547 DEGs (133 up-regulated and 414 down-regulated) from treatment studies were isolated. Two Cytoscape apps, namely, CytoHubba and MCODE, were used for identifying biomarker genes from functional networks developed using DEG genes. This distinct of biomarkers genes– one from non-treatment studies and the other from treatment studies, each set containing 16 genes. Survival analysis results show that most non-treatment biomarker genes have prognostic capability by indicating low-expression groups have higher chance of survival compare to high-expression groups. Whereas, most treatment biomarkers have prognostic capability by indicating high-expression groups have higher chance of survival compare to low-expression groups. This study developed a computational framework to discover biomarker genes for lung cancer using gene expression profiles from GEO database. Two different types of studies– non-treatment and treatment– are considered for experiment. A total of 32 biomarker genes- 16 from non-treatment studies and 16 from treatment studies- were discovered in this study. The results show that most of

the non-treatment biomarker genes (12 out of 16) have prognostic capability by indicating that low-expression groups have higher chance of survival compared to high-expression groups. On the other hand, most of the treatment biomarker genes (11 out of 16) have prognostic capability by indicating that high-expression groups have higher chance of survival compare to low expression groups. The opposite prognostic characteristics of biomarker genes discovered from non-treatment and treatment studies are expected since in non-treatment studies, controls are healthy samples and cases are cancer patients; whereas, in treatment studies, controls are cancer cell lines without treatment and cases are cancer cell lines with treatment. Most of the biomarkers in non-treatment studies (11 out of 16) were up-regulated while most of the biomarkers in treatment studies (14 out of 16) were down-regulated. The biomarker genes identified from non-treatment studies play vital role in tumor progression and metastasis. These biomarker genes are associated with cell cycle and consistent with their role in preventing genomic instabilities[25].

In another study, the expression of CDCA3 in squamous and non-squamous NSCLC was investigated by Mark N. Adams et al., using bioinformatics, Western blot analysis of matched tumor and normal tissue, and immunohistochemistry of a tissue microarray. The function of CDCA3 in NSCLC was determined by using several in vitro assays with small interfering RNA depleting CDCA3 in a panel of three immortalized human bronchial epithelial cell (HBEC) lines and seven NSCLC cell lines. In this study, cell division cycle associated 3 gene (CDCA3) transcripts were identified as highly increased in NSCLC versus in nonmalignant tissue, with high levels of CDCA3 being associated with poor patient prognosis. CDCA3 protein was also increased in NSCLC tissue and expression was limited to tumor cells. CDCA3 expression was similarly increased in a panel of NSCLC cell lines compared with in three HBEC lines. Although depletion of CDCA3 in the HBEC lines did not affect cellular proliferation, depletion of CDCA3 expression markedly reduced the proliferation of all NSCLC cell lines. CDCA3 depletion caused a defective G2/M-phase cell cycle progression, upregulation of p21 independent of p53, and induction of cellular senescence. Our findings highlight CDCA3 as a prognostic factor and potential novel therapeutic target in NSCLC through inhibition of tumor growth and promotion of tumor senescence[26].

Similarly, in another study by Meng Wang et al., through bioinformatics analysis identified replication factor C 5 (RFC5) as a potential novel oncogene in lung cancer. RFC5 functions as a clamp loader and is involved in DNA replication and repair. Analysis of public databases and reverse transcription-quantitative polymerase chain reaction indicated that RFC5 was significantly increased in tumor tissues compared with adjacent normal tissues. A high RFC5 expression was observed to be associated with more aggressive malignant clinic-pathological features, including higher T stage, more advanced regional lymph node metastasis and a higher probability of relapse. Notably, there were notable differences in overall survival (OS), first progression and post-progression survival between the high RFC5 expression group and low RFC5 expression group.

Univariate and multivariate Cox regression analyses indicated that RFC5 was an independent risk factor that was associated with poorer OS and disease-free survival. According to GSEA, several gene sets that are associated with cell cycle and DNA damage were enriched in the RFC5 overexpression group, which indicated that RFC5 might promote the proliferation of lung cancer cells. Our finding indicated that RFC5 might be a novel prognostic biomarker of lung cancer, and it might be serve as a potential diagnosis and therapy target for lung cancer in the future[27].

In another study by Xingyuan Liu et al., The gene expression profile GSE18842 was downloaded from the Gene Expression Omnibus database in this prospective study, which consisted of 46 tumors and 45 controls. After screening differentially expressed genes (DEGs), they conducted functional enrichment analysis and KEGG analysis with up-regulated differentially expressed genes (uDEGs) and down-regulated differentially expressed genes (dDEGs), respectively. Protein–protein interaction (PPI) networks among DEGs and corresponding coding protein complexes, constructed using the STRING database, were analyzed using Cytoscape. Kaplan-Meier method was used to verify survival associated with hub genes. The GEPIA webserver was used to plot the gene expression level heat map of hub genes between NSCLC and adjacent lung tissues in the TCGA databases. They identified 368 DEGs (168 uDEGs and 200 dDEGs) in NSCLC samples relative to control samples after gene integration. We established a PPI network for the DEGs, which had 249 nodes and 1472 edges protein pairs. Ten undefined hub genes with the highest connectivity degree (CDK1, UBE2C, AURKA, CCNA2, CDC20, CCNB1, TOP2A, ASPM, MAD2L1, and KIF11) were verified by survival analysis, and 9 of them were associated with poorer overall survival in NSCLC. The expression reliability of hub genes was verified by use of the GEPIA web tool. The results suggested that UBE2C, AURKA, CCNA2, CDC20, CCNB1, TOP2A, ASPM, MAD2L1, and KIF11 were inherent key biomarkers for diagnosis and prognosis, while KEGG analysis results showed the mitotic cell cycle pathway is a probable signaling pathway contributing to NSCLC progression. These genes could be promising biomarkers for diagnosis and provide a new approach for developing targeted therapeutic NSCLC drugs[28].

Table 3. List of lung cancer biomarkers

S. No.	Biomarkers identified	References
1	AF131215.5, AC125807.2, and LINC01936	[23]
2	CBLC, CYP24A1, ALDH3A1, AKR1B10, S100P, PLUNC, and LOC147166	[24].
3	CDC20, ECT2, KIF20A, MKI67, TPX2, and TYMS.	[25].
4	BUB3, CCNB1, CCNB2, CDC20, CDCA8, CDK1, CENPF, CENPI, KIF18A, KNTC1,	[29]

	MAD2L1, NDC80, NUP37, PCNA, RAD21, and ZWINT	
5	CDCA3	[26]
6	RFC5	[27]
7	UBE2C, AURKA, CCNA2, CDC20, CCNB1, TOP2A, ASPM, MAD2L1, and KIF11	[28].

3.3 Gastric cancer

Despite a substantial drop in incidence and death in North America and most Western European countries in recent decades, gastric cancer (GC) remains the fifth most prevalent malignancy worldwide and poses a serious medical burden, particularly in Eastern Asia. The poor 5-year survival rate in gastric cancer (GC) is largely due to the fact that most patients are diagnosed at an advanced stage, often with metastatic disease, thereby missing the opportunity for curative resection. Substantial progress has been made in comprehending the epidemiology, pathophysiology, and molecular mechanisms of GC, as well as in implementing new therapy alternatives like as targeted and immune-based therapies, not all patients react to molecularly targeted medications developed for specific biomarkers. Hence, due to molecular complexity, poor prognosis, and significant reoccurrence of GC, new diagnostic and prognostic biomarkers are urgently needed. Microarray and high-throughput sequencing technologies have advanced in recent years, allowing researchers to decipher important genetic or epigenetic changes in carcinogenesis and discover promising biomarkers for cancer diagnosis, treatment, and prognosis. Nevertheless, integrated bioinformatics methods have been used in cancer research to overcome limited or inconsistent results due to the use of different technology platforms or a small sample size, and a large range of valuable biological information has been revealed. Hence, here we have reviewed a few studies to ensure the role of biomarker identification associated to gastric cancer using integrated bioinformatics analysis tools. In a study by Fengyong Luo et al., they have analyzed Docking protein 5 (DOK5), a member of the docking protein group of membrane proteins and was adapter protein involved in signal transduction. Nevertheless, the role of DOK5 expression in the prognosis of gastric cancer (GC) remained unclear. Methods. In this study, clinical prognostic parameters and survival data related to DOK5, in patients with GC, were analyzed using bioinformatics analysis comprising Oncomine and TIMER, UALCAN database, Kaplan-Meier plotter, GEPIA, GSEA, DAVID, and cBioPortal websites. Results. In their study, GC contained various DOK5 expressions, which forecasted poor survival outcomes. Moreover, their research showed that high DOK5 could predict high-level infiltration of several GC immune cells, as evidenced by M1, TAM, M2, B cell, and T cell failure. Hence, DOK5 might become a new gastric cancer biomarker and therapeutic target. In the following analysis, in order to explore the prognostic value of DOK5 in GC, more clinical trials were needed to validate their results. Through multiple database verifications, DOK5 was found to be part of the pathogenic

genes for GC. Thus, it could change the formation and progression of tumors by acting on human immunity[30]. In another study by Da-Guang Wang et al., Genechips of 10 GC tissues and 10 gastric mucosa (GM, para-carcinoma tissue, normal control) tissues were generated using an exon array of Affymetrix containing 30,000 genes. The differentially expressed genes (DEGs) between GC tissues and normal control were identified by the Limma package and analyzed by hierarchical clustering analysis. Gene ontology (GO) and pathway enrichment analyses were performed for investigating the functions of DEGs. Receiver operating characteristics (ROC) analysis was performed to measure the effects of biomarker candidates for diagnosis of GC. Results: Totals of 896 up-regulated and 60 down-regulated DEGs were identified to be differentially expressed between GC samples and normal control. Hierarchical clustering analysis showed that DEGs were highly differentially expressed and most DEGs were up-regulated. The most significantly enriched GO-BP term was revealed to be mitotic cell cycle and the most significantly enriched pathway was cell cycle. The intersection analysis showed that most significant DEGs were cyclin B1 (CCNB1) and cyclin B2 (CCNB2). The sensitivities and specificities of CCNB1 and CCNB2 were both high ($p<0.0001$). Areas under the ROC curve for CCNB1 and CCNB2 were both greater than 0.9 ($p<0.0001$). Conclusions: CCNB1 and CCNB2, which were involved in cell cycle, played significant roles in the progression and development of GC and these genes may be potential biomarkers for diagnosis and prognosis of GC[31]. To examine potential therapeutic targets for GC, four Gene Expression Omnibus (GEO) datasets Wei Wang et al., downloaded and screened for differentially expressed genes (DEGs). Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were subsequently performed to study the function and pathway enrichment of the identified DEGs. A protein-protein interaction (PPI) network was constructed. The CytoHubba plugin of Cytoscape was used to calculate the degree of connectivity of proteins in the PPI network, and the two genes with the highest degree of connectivity were selected for further analysis. Additionally, the two DEGs with the largest and smallest log Fold Change values were selected. These six key genes were further examined using Oncomine and the Kaplan-Meier plotter platform. A total of 99 up-regulated and 172 down-regulated genes common to all four GEO datasets were screened. The DEGs were primarily enriched in the Biological Process terms: ‘extracellular matrix organization’, ‘collagen catabolic process’ and ‘cell adhesion’. These three KEGG pathways were significantly enriched in the categories: ‘ECM-receptor interaction’, ‘protein digestion and absorption’, and ‘focal adhesion’. Based on Oncomine, expression of ATP4A and ATP4B were down-regulated in GC, whereas expression of the other genes were all up-regulated. The Kaplan-Meier plotter platform confirmed that up-regulated expression of the identified key genes was significantly associated with worse overall survival of patients with GC. The results of the present study suggested that FN1, COL1A1, INHBA and CST1 may be potential biomarkers and therapeutic targets for GC. Additional studies are required to explore the potential value of ATP4A and ATP4B in the treatment of GC[32]. Another

study by Xinkui Liu et al., aimed to identify potential key genes associated with the pathogenesis and prognosis of GC. Differentially expressed genes between GC and normal gastric tissue samples were screened by an integrated analysis of multiple gene expression profile datasets. Key genes related to the pathogenesis and prognosis of GC were identified by employing protein–protein interaction network and Cox proportional hazards model analyses. We identified nine hub genes (TOP2A, COL1A1, COL1A2, NDC80, COL3A1, CDKN3, CEP55, TPX2, and TIMP1) which might be tightly correlated with the pathogenesis of GC. A prognostic gene signature consisted of CST2, AADAC, SERPINE1, COL8A1, SMPD3, ASPN, ITGBL1, MAP7D2, and PLEKHS1 was constructed with a good performance in predicting overall survivals. The findings of this study would provide some directive significance for further investigating the diagnostic and prognostic biomarkers to facilitate the molecular targeting therapy of GC. In conclusion, with the employment of multiple gene expression profile datasets and integrated bioinformatics analysis, we identified nine hub genes which might be involved in the pathogenesis of GC. Besides, a nine-gene signature which might act as a potential prognostic biomarker in patients with GC was constructed, and the prognostic model presented a good performance in predicting 1-, 3-, and 5-year OSs. These findings would provide some directive significance for the future prognosis prediction and molecular targeting therapy of GC. However, further experimental studies are urgently demanded to validate their results because the study was performed based on data analysis[33]. Similarly in another study by Xinyu Chong et al., differentially expressed genes (DEGs) were analyzed using GEO2R from GSE54129 and GSE13911 of the Gene Expression Omnibus (GEO). Then, gene enrichment analysis, protein-protein interaction (PPI) network construction, and topological analysis were performed on the DEGs by the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, STRING, and Cytoscape. Finally, we performed survival analysis of key genes through the Kaplan-Meier plotter. A total of 1034 DEGs were identified in GC. GO and KEGG results showed that DEGs mainly enriched in plasma membrane, cell adhesion, and PI3K-Akt signaling pathway. Subsequently, the PPI network with 44 nodes and 333 edges was constructed, and 18 candidate genes in the network were focused on by centrality analysis and module analysis. Furthermore, data showed that high expressions of fibronectin 1(FN1), the tissue inhibitor of metalloproteinases 1 (TIMP1), secreted phosphoprotein 1 (SPP1), apolipoprotein E (APOE), and versican (VCAN) were related to poor overall survivals in GC patients. In summary, this study suggests that FN1, TIMP1, SPP1, APOE, and VCAN may act as the key genes in GC[34].

Table 4. List of gastric cancer biomarker

S. No.	Biomarkers Identified	References
1	DOK5	[30]
2	CCNB1 and CCNB2	[32]
3	FN1, COL1A1, INHBA and CST1	[32]
4	TOP2A, COL1A1, COL1A2, NDC80, COL3A1, CDKN3, CEP55, TPX2, and TIMP1	[28]
5	FN1, TIMP1, SPP1, APOE, and VCAN	[34]

3.4. Liver cancer

Liver cancer is among the most frequent malignancies in the world, and it is the second leading cause of cancer death[35]. Due to advances in detection and therapy, the prognosis for patients with people with liver cancer remains poor. Most patients are already in severe stages of symptoms and miss the opportunity to undertake radical resection due to the lack of distinct clinical signs in the early stages[36]. As a result, understanding the pathophysiology of liver cancer aids in early detection, treatment selection, scheduling of follow-up appointments, and prognosis evaluation, all of which can help patients with liver cancer live longer[37]. MicroRNAs (miRNAs) are improperly expressed in a range of tumors and are linked to the pathogenesis of cancers, including liver cancer, according to growing evidence. As tumor suppressor genes or oncogenes, miRNAs play a role in the development of liver cancer[38]. As a result, more research into miRNA expression patterns and consequences could lead to the discovery of new diagnostic or therapeutic targets for liver cancer. Hence, here in this subsection of this chapter we have reviewed certain research which provide a potential aspect toward identification of biomarkers associated with cancer in relevance to liver utilizing integrated bioinformatics analysis. Hepatitis B virus (HBV) infection has long been known as a major risk factor for hepatocellular carcinoma (HCC), accounting for at least half of all HCC cases worldwide. Yet, the underlying molecular mechanism of HBV-associated HCC is still unknown[39]. In an investigation led by Thong Ba Nguyen et al., investigated the hub genes and the potential molecular pathways through which these genes contribute to liver cancer onset and development. The weighted gene co-expression network analysis (WCGNA) was performed on the main data attained from the GEO (Gene Expression Omnibus) database. The Cancer Genome Atlas (TCGA) dataset was used to evaluate the association between prognosis and these hub genes. The expression of genes from the black module was found to be significantly related to liver cancer. Based on the results of protein protein interaction, gene co-expression network, and survival analyses, DNA topoisomerase II alpha (TOP2A), ribonucleotide reductase regulatory subunit M2 (RRM2), never in mitosis-related kinase 2 (NEK2), cyclin-dependent kinase 1 (CDK1), and cyclin B1 (CCNB1) were identified as the hub genes. Gene Ontology and Kyoto Encyclopedia of Genes

and Genomes pathway enrichment analyses showed that the differentially expressed genes (DEGs) were enriched in the immune associated pathways. These hub genes were further screened and validated using statistical and functional analyses. Additionally, the TOP2A, RRM2, NEK2, CDK1, and CCNB1 proteins were overexpressed in tumor liver tissues as compared to normal liver tissues according to the Human Protein Atlas database and previous studies. Our results suggest the potential use of TOP2A, RRM2, NEK2, CDK1, and CCNB1 as prognostic biomarkers in liver cancer[40]. In another study by Lin Li et al., to identify the candidate genes in the carcinogenesis and progression of HCC, microarray datasets GSE19665, GSE33006 and GSE41804 were downloaded from Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) were identified, and function enrichment analyses were performed. The protein-protein interaction network (PPI) was constructed, and the module analysis was performed using STRING and Cytoscape. A total of 273 DEGs were identified, consisting of 189 down-regulated genes and 84 up-regulated genes. The enriched functions and pathways of the DEGs include protein activation cascade, complement activation, carbohydrate binding, complement and coagulation cascades, mitotic cell cycle and oocyte meiosis. Sixteen hub genes were identified, and biological process analysis revealed that these genes were mainly enriched in cell division, cell cycle and nuclear division. Survival analysis showed that BUB1, CDC20, KIF20A, RACGAP1 and CEP55 may be involved in the carcinogenesis, invasion or recurrence of HCC. In conclusion, DEGs and hub genes identified in the present study help us understand the molecular mechanisms underlying the carcinogenesis and progression of HCC and provide candidate targets for diagnosis and treatment of HCC[41]. Similarly, in another study by Zengyuan Zhou et al, they obtained expression profiles of GSE121248, GSE45267 and GSE84402 from the Gene Expression Omnibus (GEO), including 132 HCC and 90 noncancerous liver tissues. Differentially expressed genes (DEGs) between HCC and noncancerous samples were identified by GEO2 R and Venn diagrams. In total, 109 DEGs were identified in these datasets, including 24 up-regulated genes and 85 down-regulated genes. Subsequently, Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) preliminary analyses of the DEGs were performed using DAVID. The protein protein interaction (PPI) network of the DEGs was constructed with the Search Tool for the Retrieval of Interacting Genes (STRING) and visualized in Cytoscape. Module analysis of the PPI network was performed using MCODE to get hub genes. Moreover, the influence of the hub genes on overall survival was determined with Kaplan–Meier plotter. All hub genes were analyzed by Gene Expression Profiling Interactive Analysis (GEPIA) and KEGG. Overall, the hub genes DTL, CDK1, CCNB1, RACGAP1, ECT2, NEK2, BUB1B, PBK, TOP2A, ASPM, HMMR, RRM2, CDKN3, PRC1, and ANLN were up-regulated in HCC, and the survival rate was lower for HCC with increased expression of these hub genes. CCNB1, CDK1, and RRM2 were enriched in the p53 signaling pathway, and CCNB1, CDK1, and BUB1B were enriched in the cell cycle. In brief, they

screened 15 hub genes and pathways to identify potential prognostic markers for HCC treatment. However, the specific occurrence and development of HCC with expression of the hub genes should be verified in vivo and in vitro[42]. In another study Qifan Zhang et al., aimed to identify novel prognostic biomarkers in hepatocellular carcinoma. Integrated proteomics and bioinformatics analysis were performed to investigate the expression landscape of prognostic biomarkers in 24 paired HCC patients. As a result, eight key genes related to prognosis, including ACADS, HSD17B13, PON3, AMDHD1, CYP2C8, CYP4A11, SLC27A5, CYP2E1, were identified by comparing the weighted gene co-expression network analysis (WGCNA), proteomic differentially expressed genes (DEGs), proteomic turquoise module, The Cancer Genome Atlas (TCGA) cohort DEGs of HCC. Furthermore, we trained and validated eight pivotal genes integrating these independent clinical variables into a nomogram with superior accuracy in predicting progression events, and their lower expression was associated with a higher stage/risk score. The Gene Set Enrichment Analysis (GSEA) further revealed that these key genes showed enrichment in the HCC regulatory pathway. All in all, we found that these eight genes might be the novel potential prognostic biomarkers for HCC and also provide promising insights into the pathogenesis of HCC at the molecular level[43]. In a study by Ye Cheng Wang et al., they presented a new method for getting 6 key genes, aiming to diagnose and treat the early liver cancer. They firstly analyzed the different expression microarrays based on TCGA database, and a total of 1564 differentially expressed genes were obtained, of which 1400 were up-regulated and 164 were down-regulated. Furthermore, these differentially expressed genes were studied by using GO and KEGG enrichment analysis, a PPI network was constructed based on the STRING database, and 15 hub genes were obtained. Finally, 15 hub genes were verified by applying the survival analysis method of Oncomine database, and 6 key genes were ultimately identified, including PLK1, CDC20, CCNB2, BUB1, MAD2L1 and CCNA2. The robustness analysis of four independent data sets verifies the accuracy of the key gene's classification of the data set. Although there are complicated differences between cancer and normal cells in gene functions, cancer cells could be differentiated in case that a group of special genes expresses abnormally. Here they presented a new method to identify the 6 key genes for diagnosis and treatment of early liver cancer, and these key genes can help us understand the pathogenesis of liver cancer more deeply[44].

Table-5. Identified biomarkers highlighting key molecular markers associated with various biological processes in liver cancer

S. No	Biomarkers Identified	References
1	TOP2A, RRM2, NEK2, CDK1, and CCNB1	[40]
2	BUB1, CDC20, KIF20A, RACGAP1 and CEP55	[41]
3	DTL, CDK1, CCNB1, RACGAP1, ECT2, NEK2, BUB1B, PBK, TOP2A, ASPM, HMMR, RRM2, CDKN3, PRC1, and ANLN	[42].
4	ACADS, HSD17B13, PON3, AMDHD1, CYP2C8, CYP4A11, SLC27A5, CYP2E1	[43]
5	PLK1, CDC20, CCNB2, BUB1, MAD2L1 and CCNA2	[44]

3.5 Breast cancer

Breast cancer is becoming more common worldwide, and it is now considered a serious health concern. Asia has recently emerged as a high-risk location for breast cancer, ranking first among female malignant tumors[45]. Breast cancer therapy has improved recently because of constant efforts and advances in contemporary medicine, and the death rate of breast cancer has decreased dramatically. Recurrence and metastasis of breast cancer, on the other hand, have remained unaddressed and have become the most difficult clinical difficulties[46]. To better understand the functions of tumor-related genes and the roles of tumor cell signaling pathways, researchers are turning to genetic studies. Together bioinformatics and system biology are strong multidisciplinary topics that combine biological information collecting, storage, processing, and distribution, summarize life sciences and computer science, and collect and analyze genetic data[47]. Hence, here in this chapter we have reviewed a few studies led by researchers to identify most prevalent biomarkers associated with breast cancer utilizing integrated bioinformatics approaches.

In an investigation by Hatem Zayed, they analyzed that Breast cancer (BC) is one of the leading causes of cancer-related death among women worldwide, and claudin- low breast cancer (CLBC) is a subtype of BC that remains poorly described. This study aimed to identify up-regulated genes and significant pathways involved in CLBC. The SUM159 cell line is derived from human CLBC tissue; the GSE50697 dataset contains three replicates of SUM159 cells treated with pBabe puro miR-203 and three replicates of control SUM159 cells (pBabe puro). The data were normalized and up-regulated, and down-regulated genes were identified based on the logFC values. Gene Ontology (GO) and pathway analysis identified the most significant pathways and genes involved in CLBC

pathogenesis. A total of 156 significant genes were identified (69 up-regulated genes and 64 down-regulated genes). The up-regulated genes were the focus of this study, from the pathway analysis, the senescence-associated secretory phenotype, which involves the CXCL8, IL1A, and IL6 genes, was found to be mapped through more than one pathway (WikiPathways and Reactome). From the refined GO analysis, using MetaCore, Cortellis solution software, the IL-13 signaling pathway was identified; this pathway includes the IL6, CXCL8, VEGF-C, NRG1, and EREG genes, which were mapped as hub genes in several pathogenesis pathways. From the survival analysis, high levels of IL6, CXCL8, and EREG were related to high survival rates, and low levels of VEGFC and NRG1 were related to high survival rates. The IL6 and CXCL8 genes were the most significant and the most highly represented in the GO and refined GO analyses. This study sheds light on the molecular pathology of CLBC and might provide a potential biomarker for the treatment of CLBC[48]. In another study by Jun-Li Deng et al., they aimed to identify potential pathogenic, and prognostic differentially expressed genes (DEGs) in breast adenocarcinoma through bioinformatics analysis of public datasets. Four datasets (GSE21422, GSE29431, GSE42568, and GSE61304) from Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) dataset were used for the bioinformatics analysis. DEGs were identified using LIMMA Package of R. The GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analyses were conducted through FunRich. The protein-protein interaction (PPI) network of the DEGs was established through STRING (Search Tool for the Retrieval of Interacting Genes database) website, visualized by Cytoscape and further analyzed by Molecular Complex Detection (MCODE). UALCAN and Kaplan–Meier (KM) plotter was employed to analyze the expression levels and prognostic values of hub genes. The expression levels of the hub genes were also validated in clinical samples from breast cancer patients. In addition, the gene-drug interaction network was constructed using Comparative Toxico genomics Database (CTD). In total, 203 up-regulated and 118 down-regulated DEGs were identified. Mitotic cell cycle and epithelial-to-mesenchymal transition pathway were the major enriched pathways for the up-regulated and down-regulated genes, respectively. The PPI network was constructed with 314 nodes and 1,810 interactions, and two significant modules are selected. The most significant enriched pathway in module 1 was the mitotic cell cycle. Moreover, six hub genes were selected and validated in clinical sample for further analysis owing to the high degree of connectivity, including CDK1, CCNA2, TOP2A, CCNB1, KIF11, and MELK, and they were all correlated to worse overall survival (OS) in breast cancer. These results revealed that mitotic cell cycle and epithelial-to-mesenchymal transition pathway could be potential pathways accounting for the progression in breast cancer, and CDK1, CCNA2, TOP2A, CCNB1, KIF11, and MELK may be potential crucial genes. Further, it could be utilized as new biomarkers for prognosis and potential new targets for drug synthesis of breast cancer[49]. In this investigation carried by Ningning Wang et al., they pooled profile datasets from three cohorts to illuminate the underlying key genes and

pathways of BC. Expression profiles GSE42568, GSE45827, and GSE124646, including 244 BC tissues and 28 normal breast tissues, were integrated and analyzed. Differentially expressed genes (DEGs) were screened out based on these three datasets. Functional analysis including Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway were performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID). Moreover, Cytoscape with Search Tool for the Retrieval of Interacting Genes (STRING) and Molecular Complex Detection (MCODE) plugin were utilized to visualize protein protein interaction (PPI) of these DEGs. The module with the highest connectivity of gene interactions was selected for further analysis. All these hub genes had a significantly worse prognosis in BC by survival analysis. Additionally, four genes (CDK1, CDC20, AURKA, and MCM4) dramatically were enriched in oocyte meiosis and cell cycle pathways through re-analysis of DAVID. Moreover, the mRNA and protein levels of CDK1, CDC20, AURKA, and MCM4 were significantly increased in BC patients. In addition, knockdown of CDK1 and CDC20 by small interfering RNA remarkably suppressed cell migration and invasion in MCF-7 and MDA-MB-231 cells. In conclusion, our results suggested that CDK1, CDC20, AURKA, and MCM4 were reliable biomarkers of BC via bioinformatics analysis and experimental validation and may act as prospective targets for BC diagnosis and treatment[50]. Similarly, in another study Kai Huang et al demonstrated from recent studies that chaperonin-containing TCP1 subunit 6A (CCT6A) efficiently suppresses transforming growth factor-β-mediated metastasis by inhibiting the function of SMAD family member 2 in lung cancer. However, the functional significance of CCT6A in other types of cancer, including BC, remains to be investigated. Therefore, this study evaluated CCT6A expression in BC samples, and further analysed its association with survival, clinicopathological parameters and related signalling pathways using online datasets. The present study indicated that CCT6A expression was significantly higher in BC tissues compared with in surrounding noncancerous tissues at both mRNA and protein levels. Furthermore, increased CCT6A expression was significantly associated with poor survival, including overall survival, relapse-free survival, distant metastasis-free survival and post progression survival, in patients with BC. Pathway finder analysis indicated that CCT6A was significantly associated with the cell cycle, and its expression was significantly positively correlated with cyclin (CCN)B2 and CCNA2 expression. Taken together, to the best of knowledge, the present study was the first to indicate that CCT6A may serve a significant role in BC tumor progression[51]. In a study by Gang Chen et al., they aimed to screen key genes and potential prognostic biomarkers for BC using bioinformatics analysis. Total 58 normal tissues and 203 cancer tissues were collected from three Gene Expression Omnibus (GEO) gene expression profiles, and then the differential expressed genes (DEGs) were identified. Subsequently, the Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genome (KEGG) pathway were analyzed to investigate the biological function of DEGs. Additionally, hub genes were screened by constructing a protein–

protein interaction (PPI) network. Then, we explored the prognostic value and molecular mechanism of these hub genes using Kaplan–Meier (KM) curve and Gene Set Enrichment Analysis (GSEA). As a result, 42 up- regulated and 82 down-regulated DEGs were screened out from GEO datasets. Furthermore, 12 hub genes (FN1, AURKA, CCNB1, BUB1B, PRC1, TPX2, NUSAP1, TOP2A, KIF20A, KIF2C, RRM2, ASPM) with a high degree were identified initially, among which, 11 hub genes were significantly correlated with the prognosis of BC patients based on the Kaplan–Meier-plotter. GSEA reviewed that these hub genes correlated with KEGG_CELL_CYCLE and HALLMARK_P53_PATHWAY. In conclusion, this study identified 11 key genes as BC potential prognosis biomarkers based on integrated bioinformatics analysis. This finding will improve our knowledge of the BC progress and mechanisms. In conclusion, their study identified 13 hub genes and 3 non-hub genes (AURKA, CCNB1, BUB1B, PRC1, TPX2, NUSAP1, TOP2A, KIF20A, KIF2C, RRM2, ASPM, PPARG, non-hub genes are CDCA3, ZWINT and UBE2S) that might be involved in the progression of BC using multiple gene expression data sets and a series of comprehensive analyses of bioinformatics. These findings provide new insights into the diagnosis treatment of the BC, while the main limitation of this research is lacking experiment to verify the hub genes expression level in the BC tissues and function in BC cells[52]. Therefore, the further experimental studies are still needed to ensure our findings.

Table 6. List of identified breast cancer biomarker

S. No.	Biomarkers Identified	References
1	IL6, CXCL8, VEGF-C, NRG1, and EREG	[48]
2	CDK1, CCNA2, TOP2A, CCNB1, KIF11, and MELK	[49]
3	CDK1, CDC20, AURKA, and MCM4	[50]
4	CCT6A	[51]
5	AURKA, CCNB1, BUB1B, PRC1, TPX2, NUSAP1, TOP2A, KIF20A, KIF2C, RRM2, ASPM, PPARG, and CDCA3, ZWINT, UBE2S	[52]

3.6 Colorectal cancer

Colorectal cancer (CRC) is one of the top causes of death among cancer patients around the world. Older age, male sex, lifestyle, inflammatory bowel illness, and a previous personal history of CRC are all risk factors for the disease. A positive family history is also substantially linked to a higher lifetime relative risk of CRC diagnosis. CRC, on the other hand, is an indolent disease in its early stages, becoming symptomatic only when it evolves to more advanced stages. Numerous attempts have been made to develop adequate screening technologies, but they remain intrusive even now, resulting in reduced attainment rates among large community[53]. Recent breakthroughs in our understanding of the molecular underpinnings and cellular mechanisms of CRC have resulted in the

widespread use of molecular diagnostics in clinical practice. The patient's risk is stratified and therapy is decided based on the test results. Conversely, current research into biomarkers associated with colorectal cancer could usher in a new age in diagnosis, risk prediction, and treatment selection[54]. In a study performed by Yeonjoo Jung et al., aimed to identify marker genes of colorectal cancer (CRC) by combining bioinformatics analysis of gene expression data and validation experiments using patient samples and to examine the potential connection between validated markers and the established oncogenes such as c-Myc and K-ras. Publicly available data from GenBank and Oncomine were meta-analyzed leading to 34 candidate marker genes of CRC. Multiple case-matched normal and tumor tissues were examined by RT-PCR for differential expression, and 9 genes were validated as CRC biomarkers. Statistical analyses for correlation with major clinical parameters were carried out, and RNA interference was used to examine connection with major oncogenes. Results: We show with high confidence that 9 (ECT2, ETV4, DDX21, RAN, S100A11, RPS4X, HSPD1, CKS2, and C9orf140) of the 34 candidate genes are expressed at significantly elevated levels in CRC tissues compared to normal tissues. Furthermore, high-level expression of RPS4X was associated with non-mucous cancer cell type and that of ECT2 with lack of lymphatic invasion while upregulation of CKS2 was correlated with early tumor stage and lack of family history of CRC. We also demonstrate that RPS4X and DDX21 are regulatory targets of c-Myc and ETV4 is downstream to K-ras signaling. Conclusions: identified multiple novel biomarkers of CRC. Further analyses of their function and connection to signaling pathways may reveal potential value of these biomarkers in diagnosis, prognosis, and treatment of CRC[55]. In another study by Juan Chen et al., the expression profile of human multistage colorectal mucosa tissues, including healthy, adenoma, and adenocarcinoma samples was downloaded to identify critical genes and potential drugs in CRC. Methods: Expression profiles, GSE33113 and GSE44076, were integrated using bioinformatics methods. Differentially expressed genes (DEGs) were analyzed by R language. Functional enrichment analyses of the DEGs were performed using the Database for Annotation, visualization, and integrated discovery (DAVID) database. Then, the search tool for the retrieval of interacting genes (STRING) database and Cytoscape were used to construct a protein–protein interaction (PPI) network and identify hub genes. Subsequently, survival analysis was performed among the key genes using Gene Expression Profiling Interactive Analysis (GEPIA). Connectivity Map (CMap) was used to query potential drugs for CRC. Results: A total of 428 up-regulated genes and 751 down-regulated genes in CRC were identified. The functional changes of these DEGs were mainly associated with cell cycle, oocyte meiosis, DNA replication, p53 signaling pathway, and progesterone-mediated oocyte maturation. A PPI network was identified by STRING with 482 nodes and 2,368 edges. Survival analysis revealed that high mRNA expression of AURKA, CCNB1, CCNF, and EXO1 was significantly associated with longer overall survival. Moreover, CMap predicted a panel of small molecules as possible adjuvant drugs to treat CRC. Conclusion:

study found key dysregulated genes involved in CRC and potential drugs to combat it, which may provide novel insights and potential biomarkers for prognosis, as well as providing new CRC treatments[56]. In a similar study by Wenping Lian et al., they identified differentially expressed genes (DEGs) and lncRNAs (DELs) of CRC from The Cancer Genome Atlas (TCGA) database. Then they conducted the weighted gene co-expression network analysis (WGCNA) to investigate co-expression modules related with CRC metastasis. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, DEG-DEL co-expression network and survival analyses of significant modules were also conducted. Finally, the expressions of selected biomarkers were validated in cell lines by quantitative real-time PCR (qRT-PCR). 2032 DEGs and 487 DELs were involved the construction of WGCNA network, and greenyellow, turquoise and brown module were identified to have more significant correlation with CRC metastasis. GO and KEGG pathway analysis of these three modules have proven that the functions of DEGs were closely involved in many important processes in cancer pathogenesis. Through the DEG-DEL co-expression network, 12 DEGs and 2 DELs were considered as hub nodes. Besides, survival analysis showed that 30 DEGs were associated with the overall survival of CRC. Then 10 candidate biomarkers were chosen for validation and the expression of CA2, CHP2, SULT1B1, MOGAT2 and C1orf115 were significantly decreased in CRC cell lines when compared to normal human colonic epithelial cells, which were consistent with the results of differential expression analysis. Especially, low expression of SULT1B1, MOGAT2 and C1orf115 were closely correlated with poorer survival of CRC. Conclusion: This study identified 5 genes as new biomarkers affecting the metastasis of CRC. Besides, SULT1B1, MOGAT2 and C1orf115 might be implicated in the prognosis of CRC patients[57]. In another study by Ahmad Hammad et al., aimed to perform an integrated bioinformatics and machine learning analyses to explore novel biomarkers for CRC prognosis. In this study, acquired gene expression microarray data from Gene Expression Omnibus (GEO) database. The microarray expressions GSE103512 dataset was downloaded and integrated. Subsequently, differentially expressed genes (DEGs) were identified and functionally analyzed via Gene Ontology (GO) and Kyoto Enrichment of Genes and Genomes (KEGG). Furthermore, protein protein interaction (PPI) network analysis was conducted using the STRING database and Cytoscape software to identify hub genes. However, the hub genes were subjected to Support Vector Machine (SVM), Receiver operating characteristic curve (ROC) and survival analyses to explore their diagnostic values. Meanwhile, TCGA transcriptomics data in Gene Expression Profiling Interactive Analysis (GEPIA) database and the pathology data presented by in the human protein atlas (HPA) database were used to verify our transcriptomic analyses. A total of 105 DEGs were identified in this study. Functional enrichment analysis showed that these genes were significantly enriched in biological processes related to cancer progression. Thereafter, PPI network explored a total of 10 significant hub genes. The ROC curve was used to predict the potential application of

biomarkers in CRC diagnosis, with an area under ROC curve (AUC) of these genes exceeding 0.92 suggesting that this risk classifier can discriminate between CRC patients and normal controls. Moreover, the prognostic values of these hub genes were confirmed by survival analyses using different CRC patient cohorts. Our results demonstrated that these 10 differentially expressed hub genes (IGF1, MYH11, CLU, MYL9, CXCL12, LMOD1, C3, CNN1, FOS, and HIST1H2BO) could be used as potential biomarkers for CRC diagnosis[58]. Investigation performed by Zhen Sun et al., in which mRNA microarray datasets GSE113513, GSE21510, GSE44076, and GSE32323 were obtained from the Gene Expression Omnibus (GEO) and analyzed with bioinformatics to identify hub genes in CRC development. Differentially expressed genes (DEGs) were analyzed using the GEO2R tool[59]. Gene ontology (GO) and KEGG analyses were performed through the DAVID database. STRING database and Cytoscape software were used to construct a protein-protein interaction (PPI) network and identify key modules and hub genes. Survival analyses of the DEGs were performed on GEPIA database. The Connectivity Map database was used to screen potential drugs. A total of 865 DEGs were identified, including 374 up-regulated and 491 down-regulated genes. These DEGs were mainly associated with metabolic pathways, pathways in cancer, cell cycle and so on. The PPI network was identified with 863 nodes and 5817 edges. Survival analysis revealed that HMMR, PAICS, ETFDH, and SCG2 were significantly associated with overall survival of CRC patients. Blebbistatin and Sulconazole were identified as candidate drugs. In conclusion, their study found four hub genes involved in CRC, which may provide novel potential biomarkers for CRC prognosis, and two potential candidate drugs for CRC[59].

Table 7. List of biomarkers used Colorectal Cancer

S. No.	Biomarkers Identified	References
1	ECT2, ETV4, DDX21, RAN, S100A11, RPS4X, HSPD1, CKS2, and C9orf140	[55]
2	AURKA, CCNB1, CCNF, and EXO1	[56]
3	CA2, CHP2, SULT1B1, MOGAT2 and C1orf115	[57]
4	IGF1, MYH11, CLU, MYL9, CXCL12, LMOD1, C3, CNN1, FOS, and HIST1H2BO	[58]
5	HMMR, PAICS, ETFDH, and SCG2	[59]

3.7 Cervical Cancer

Cervical cancer is a malignant neoplasm originating from cells of the squamocolumnar junction of the uterine cervix[60]. It is the second most frequent kind of cancer and one of the major causes of

cancer-related deaths among women across the world, particularly when the disease is identified at an advanced stage. Cervical cancer can only occur if a person is infected with the "highly oncogenic" strain of Human Papillomavirus (HPV). When cervical cancer is diagnosed at a late stage, the average survival time is less than one year[61]. Therefore, it is crucial to develop effective screening tests that can provide early detection and prevention. In a study by Medi Kori et al., meta-analysis was performed on cervical cancer-associated transcriptome data and reporter biomolecules were identified at RNA (mRNA, miRNA), protein (receptor, transcription factor, etc.), and metabolite levels by the integration of gene expression profiles with genome-scale biomolecular networks. This approach revealed already-known biomarkers, tumor suppressors and oncogenes in cervical cancer as well as various receptors (e.g. ephrin receptors EPHA4, EPHA5, and EPHB2; endothelin receptors EDNRA and EDNRB; nuclear receptors NCOA3, NR2C1, andNR2C2), miRNAs(e.g., miR-192-5p, miR-193b-3p, andmiR-215-5p), transcription factors (particularly E2F4, ETS1, and CUTL1), other proteins (e.g., KAT2B, PARP1, CDK1, GSK3B,WNK1, and CRYAB),and metabolites (particularly, arachidonic acids) as novel biomarker candidates and potential therapeutic targets. The differential expression profiles of all reporter biomolecules were cross-validated in independent RNA-Seq and miRNA-Seq datasets, and the prognostic power of several reporter biomolecules, including KAT2B, PCNA, CD86, miR-192-5p and miR-215-5p was also demonstrated. In this study, we reported valuable data for further experimental and clinical efforts, because the proposed biomolecules have significant potential as systems biomarkers for screening or therapeutic purposes in cervical carcinoma[62]. Xiao fang Li et al., in another study aimed to explore the molecular mechanism of carcinogenesis and biomarkers for cervical cancer by integrated bioinformatics analysis. Employed RNA-sequencing details of 254 cervical squamous cell carcinomas and 3 normal samples from The Cancer Genome Atlas. To explore the distinct pathways, messenger RNA expression was submitted to a Gene Set Enrichment Analysis. Kyoto Encyclopedia of Genes and Genomes and protein–protein interaction network analysis of differentially expressed genes were performed. Then, they conducted pathway enrichment analysis for modules acquired in protein–protein interaction analysis and obtained a list of pathways in every module. After intersecting the results from the 3 approaches, they evaluated the survival rates of both mutual pathways and genes in the pathway, and 5 survival-related genes were obtained. Finally, Cox hazards ratio analysis of these 5 genes was performed. Expression levels of MCM2, MCM4, MCM5, PCNA, and RNASEH2A were significantly related to the OS of patients with cervical squamous cancer while the expression levels of the other 7 genes were uncorrelated with the survival rate of patients with cervical squamous cancer. On the other hand, high expression of MCM2, MCM4, MCM5, PCNA, and RNASEH2A could bring high rate of survival, which was consistent with the result of gene set survival curves we presented before. Together, high level of these 5 genes may be the promising prognostic factors to predict the better survival of cervical cancer. DNA

replication pathway (P < .001; 12 genes included) was suggested to have the strongest association with the prognosis of cervical squamous cancer. In total, 5 of the 12 genes, namely, mini chromosome maintenance 2, mini chromosome maintenance 4, mini chromosome maintenance 5, proliferating cell nuclear antigen, and ribonuclease H2 subunit A were significantly correlated with survival. Mini chromosome maintenance 5 was shown as an independent prognostic biomarker for patients with cervical cancer. This study identified a distinct pathway (DNA replication). Five genes which may be prognostic biomarkers and mini chromosome maintenance 5 were identified as independent prognostic biomarkers for patients with cervical cancer[63]. The aim of the study under consideration performed by Kejia Wu et al., was to investigate the key pathways and genes in the progression of cervical cancer. The gene expression profiles GSE7803 and GSE63514 were obtained from the Gene Expression Omnibus database. Differentially expressed genes (DEGs) were identified using GEO2R and the limma package, and Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted using the Database for Annotation, Visualization and Integrated Discovery. The hub genes were identified using Cytoscape and protein-protein interaction (PPI) networks were constructed using the STRING database. A total of 127 and 99 DEGs were identified in the pre-invasive and invasive stages of cervical cancer, respectively. GO enrichment analysis indicated that the DEGs in pre-invasive cervical cancer were primarily associated with the ‘protein binding’, ‘single-stranded DNA-dependent ATPase activity’, ‘DNA replication origin binding’ and ‘microtubule binding’ terms, whereas the DEGs in invasive cervical cancer were associated with the ‘extracellular matrix (ECM) structural constituent’, ‘heparin binding’ and ‘integrin binding’. KEGG enrichment analysis revealed that the pre-invasive DEGs were significantly enriched in the ‘cell cycle’, ‘DNA replication’ and ‘p53 signaling pathway’ terms, while the invasive DEGs were enriched in the ‘amoebiasis’, ‘focal adhesion’, ‘ECM-receptor interaction’ and ‘platelet activation’ terms. The PPI network identified 4 key genes (PCNA, CDK2, VEGFA and PIK3CA), which were hub genes for pre-invasive and invasive cervical cancer. In conclusion, bioinformatics analysis identified 4 key genes in cervical cancer progression (PCNA, CDK2, VEGFA and PIK3CA), which may be potential biomarkers for differentiating normal cervical epithelial tissue from cervical cancer[64]. Similarly, in another study by Yunan He et al., Gene set data from the NCBI-GEO database were used in this study to compare the differences of gene and protein levels between normal and cancer tissues through significant pathway selection and core gene signature analysis to screen potential clinical biomarkers of CESC. Subsequently, the molecular and protein levels of clinical samples were verified by quantitative transcription PCR, western blot and immunohistochemistry. Results. Three differentially expressed genes (RFC4, MCM2, TOP2A) were found to have a significant survival (P <0.05) and highly expressed in CESC tissues. Molecular biological verification using quantitative reverse transcribed PCR, western blotting and immunohistochemistry assays exhibited significant differences in the expression of

RFC4 between CESC and para-cancerous tissues (P <0.05). Conclusion. This study identified three potential biomarkers (RFC4, MCM2, TOP2A) of CESC which may be useful to clarify the underlying mechanisms of CESC and predict the prognosis of CESC patients[65]. In another study, Han Xue et al., aimed to analyse the microarray datasets using comprehensive bioinformatics tools and identified novel biomarkers associated with the prognosis of patients with cervical cancer. The differentially expressed genes (DEGs) from Gene Expression Omnibus (GEO) datasets including GSE138080, GSE113942 and GSE63514 were analysed using GEO2R tool. The functional enrichment analysis was performed using g:Profiler tool. The protein–protein interaction (PPI) network construction and hub genes identification were performed using the STRING database and Cytoscape software, respectively. The hub genes were subjected to expression and survival analysis in the cervical cancer. The EdU incorporation and Cell Counting Kit-8 assays were performed to evaluate the effects of hub gene knockdown on the proliferation of cervical cancer cells. A total of 89 overlapping DEGs (63 up-regulated and 26 down-regulated genes) were identified in the microarray datasets. The functional enrichment analysis indicated that the overlapping DEGs were mainly associated with “DNA replication” and “cell cycle”. Furthermore, the PPI network analysis revealed that the network contains 87 nodes and 309 edges. Sub-module analysis using the Molecular Complex Detection tool identified 21 hub genes from the PPI network. The expression levels of the 21 hub genes were all up-regulated in the cervical cancer tissues when compared to normal cervical tissues as analysed by GEPIA tool. The survival analysis showed that the low expression of cell division cycle 45 (CDC45), GINS complex subunit 2 (GINS2), minichromosome maintenance complex component 2 (MCM2) and proliferating cell nuclear antigen (PCNA) was significantly correlated with the shorter overall survival of patients with cervical cancer. Moreover, the protein expression levels of GINS2, MCM2 and PCNA, but not CDC45, were significantly up-regulated in the cervical cancer tissues when compared to normal cervical tissues. Finally, knockdown of MCM2 significantly suppressed the proliferation of HeLa and SiHa cells. In conclusion, we screened a total of 89 overlapping DEGs from the GEO datasets, and further analysis identified four hub genes (CDC45, GINS2, MCM2 and PCNA) that were likely associated with the prognosis of patients with cervical cancer. MCM2 knockdown repressed the cervical cancer cell proliferation. The current findings may provide novel insights into understanding the pathophysiology of cervical cancer and develop therapeutic targets for patients with cervical cancer[66].

Table 8. List of biomarkers used in cervical cancer therapy

S. No.	Biomarkers Identified	
1	BRCA1, ESR1, PCNA, and FGFR2	[62]
2	MCM2, MCM4, MCM5, PCNA, and RNASEH2A	[63]
3	PCNA, CDK2, VEGFA and PIK3CA	[64]
4	RFC4, MCM2, TOP2A	[65]
5	CDC45, GINS2, MCM2 and PCNA	[66]

4. Challenges and Future Outlook

The development of biomarkers for early detection cancer screening and therapy monitoring has biological as well as financial hurdles. Most existing cancer detection tools only detect late stage or fully grown cancer, not premalignant or early abnormalities that can be treated. Even though a screening test may detect cancer just at preclinical stage, it is not suitable for follow-up, and hence may miss micro metastases, limiting the benefits of early identification and treatment[67]. Additional barrier to the development of cancer biomarkers is the fact that cancer is a diverse illness, with several biologically distinct phenotypes that respond differently to treatments. Between cells of a single macroscopic tumor, the nature of its heterogeneity can be found. Biomarker development may be hampered by this variability[68]. As a result, developing biomarkers using genomic and proteomic methods could help to solve the variability challenges[69]. An even more issue is that pre-neoplastic lesions are far more common than aggressive malignancies in several organs, such as the prostate and colon. This examines whether any screening method should concentrate exclusively on early lesions or also take into account the tumor's behaviour. Over the past two decades, extensive understanding of cancer at the cellular and molecular levels has significantly improved, leading to substantial advancements in the characterization of human tumours. A shift toward the development of targeted therapies, the foundation of molecular diagnostics. Omics technology may serve as the foundation for the development of novel cancer biomarker and/or panels that have significant advantages over currently utilized biomarkers[70]. Omics has enhanced the number of potential biomarkers such as DNA, RNA, and other protein biomolecules that may be studied. The previous idea of single biomarker discovery has lately been supplanted by multi-biomarker discovery of a panel of genes or proteins, raising the question of whether heterogeneous and complex cancers can

have a single fingerprint[71]. Biomarkers in association with cancer are used in oncology and clinical practice for risk assessment, screening, and diagnosis in combination with other diagnostic methods, and most importantly for determining prognosis and treatment response and/or recurrence. Cancer biomarkers can also help with cancer diagnosis at the molecular level[72], [73]. Clinicians and researchers must have a thorough understanding of the molecular aspects, clinical utility, and reliability of biomarkers to determine whether a biomarker is clinically useful for patient care and whether additional evaluation is required before integration into routine care. Biomarkers, through simplifying the integration of therapies and diagnostics, have the potential to play a key role in the development of customized medicine.

5. CONCLUSION

Research in the field of cancer-specific biomarkers has provided a promising source of novel diagnostic tools. Various groups have reported that altered cancer associated biomarkers can be exploited to diagnose and monitor various cancers with greater sensitivity and specificity. Assessment of genomic and transcriptomic biomarkers found to be potentially very sensitive approaches for discriminating between cancerous and non-cancerous (benign) conditions. Besides, this one could detect cancers at a much earlier stage by quantitative analysis of potential biomarker associated with specific cancer. Given the possible diagnostic power of genomic, transcriptomic, proteomic, and metabolomic biomarkers, these are currently one of the most promising areas of research in the field of development of cancer prognostic and diagnostics devices.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals or humans were used for the studies that are based on this research.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

ACKNOWLEDGEMENT

Dr. Vivekananda spearheaded conceptualization, methodology, and draft-writing; and oversaw the project, contributed to data analysis and manuscript preparation, and focused on data curation. Dr. Sujata Kumari handled software and validation, enhancing the research's technical robustness and contributed to supervision. Both authors together made conceptual efforts, managed the project's administration, and participated in funding acquisition. The collective effort of all authors, acknowledging their specific contributions, culminated in this interdisciplinary study endeavour.

CONFLICT OF INTEREST

Not Applicable

REFERENCES

1. NCI, “Common Cancer Types - NCI.” Accessed: Oct. 14, 2024. [Online]. Available: https://www.cancer.gov/types/common-cancers
2. WHO_Cancer, “Global cancer burden growing, amidst mounting need for services.” Accessed: Jan. 15, 2025. [Online]. Available: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services
3. J. Rajput, “CANCER: A COMPREHENSIVE REVIEW,” International Journal of Research in Pharmacy and Chemistry, vol. 1, pp. 42–49, Feb. 2023.
4. F. A. Sloan and H. Gelband, “The Cancer Burden in Low- and Middle-Income Countries and How It Is Measured,” in Cancer Control Opportunities in Low- and Middle-Income Countries, National Academies Press (US), 2007. Accessed: Jan. 15, 2025. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK54028/
5. K. Sathishkumar, M. Chaturvedi, P. Das, S. Stephen, and P. Mathur, “Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India,” Indian J Med Res, vol. 156, no. 4–5, pp. 598–607, 2022.
6. A. Passaro et al., “Cancer Biomarkers - Emerging Trends and Clinical Implications for personalized treatment,” Cell, vol. 187, no. 7, pp. 1617–1635, Mar. 2024.
7. Z. Herceg and P. Hainaut, “Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis,” Mol Oncol, vol. 1, no. 1, pp. 26–41, Jun. 2007.
8. C. P. Lombe, M. Meyer, and A. Pretorius, “Bioinformatics Prediction and Analysis of MicroRNAs and Their Targets as Biomarkers for Prostate Cancer: A Preliminary Study,” Mol Biotechnol, vol. 64, no. 4, pp. 401–412, Apr. 2022.
9. M. Kohl, S. Wiese, and B. Warscheid, “Cytoscape: Software for Visualization and Analysis of Biological Networks,” in Data Mining in Proteomics, vol. 696, M. Hamacher, M. Eisenacher, and C. Stephan, Eds., in Methods in Molecular Biology, vol. 696. , Totowa, NJ: Humana Press, 2011, pp. 291–303.
10. D. Szklarczyk et al., “STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” Nucleic Acids Res, vol. 47, no. Database issue, pp. D607–D613, Jan. 2019.
11. C. Sticht, C. De La Torre, A. Parveen, and N. Gretz, “miRWalk: An online resource for prediction of microRNA binding sites,” PLoS ONE, vol. 13, no. 10, p. e0206239, Oct. 2018.
12. N. W. Wong, Y. Chen, S. Chen, and X. Wang, “OncomiR: an online resource for exploring pan-cancer microRNA dysregulation,” Bioinformatics, vol. 34, no. 4, pp. 713–715, Feb. 2018.
13. A. Vaziri-Moghadam and M. Foroughmand-Araabi, “Integrating machine learning and bioinformatics approaches for identifying novel diagnostic gene biomarkers in colorectal cancer

| Scientific Reports." Accessed: Jan. 15, 2025. [Online]. Available: https://www.nature.com/articles/s41598-024-75438-6#citeas

14. L. Xie, L. Wang, W. Zhu, J. Zhao, and X. Guo, "Editorial: Bioinformatics tools (and web server) for cancer biomarker development, volume II," Front. Genet., vol. 13, Oct. 2022.
15. L. Xie, L. Wang, W. Zhu, J. Zhao, and X. Guo, "Editorial: Bioinformatics Tools (and Web Server) for Cancer Biomarker Development," Front Oncol, vol. 10, p. 599085, Oct. 2020.
16. J. Huang, L. Mao, Q. Lei, and A.-Y. Guo, "Bioinformatics tools and resources for cancer and application," Chinese Medical Journal, vol. 137, no. 17, p. 2052, Sep. 2024.
17. L. Wang, B. Lu, M. He, Y. Wang, Z. Wang, and L. Du, "Prostate Cancer Incidence and Mortality: Global Status and Temporal Trends in 89 Countries From 2000 to 2019," Front Public Health, vol. 10, p. 811044, Feb. 2022.
18. S. Maekawa, R. Takata, and W. Obara, "Molecular Mechanisms of Prostate Cancer Development in the Precision Medicine Era: A Comprehensive Review," Cancers, vol. 16, no. 3, Art. no. 3, Jan. 2024, doi: 10.3390/cancers16030523.
19. Z. He, X. Duan, and G. Zeng, "Identification of potential biomarkers and pivotal biological pathways for prostate cancer using bioinformatics analysis methods," PeerJ, vol. 7, p. e7872, Oct. 2019.
20. R. Wang, Y. Wu, J. Yu, G. Yang, H. Yi, and B. Xu, "Plasma Messenger RNAs Identified Through Bioinformatics Analysis are Novel, Non-Invasive Prostate Cancer Biomarkers," Onco Targets Ther, vol. 13, pp. 541–548, 2020.
21. F. Kosari et al., "Identification of prognostic biomarkers for prostate cancer," Clin Cancer Res, vol. 14, no. 6, pp. 1734–1743, Mar. 2008.
22. T. Iwata et al., "A new bioinformatics approach identifies overexpression of GRB2 as a poor prognostic biomarker for prostate cancer," Sci Rep, vol. 11, p. 5696, Mar. 2021.
23. J. Hou and C. Yao, "Potential Prognostic Biomarkers of Lung Adenocarcinoma Based on Bioinformatics Analysis," BioMed Research International, vol. 2021, no. 1, p. 8859996, Jan. 2021.
24. B. Kim et al., "Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data," Cancer Res, vol. 67, no. 15, pp. 7431–7438, Aug. 2007.
25. B. Dai, L. Ren, X. Han, and D. Liu, "Bioinformatics analysis reveals 6 key biomarkers associated with non-small-cell lung cancer," J Int Med Res, vol. 48, no. 3, p. 0300060519887637, Mar. 2020.
26. M. N. Adams et al., "Expression of CDCA3 Is a Prognostic Biomarker and Potential Therapeutic Target in Non–Small Cell Lung Cancer," Journal of Thoracic Oncology, vol. 12, no. 7, pp. 1071–1084, Jul. 2017.

27. M. Wang et al., “Identification of RFC5 as a novel potential prognostic biomarker in lung cancer through bioinformatics analysis,” Oncol Lett, Jul. 2018.
28. X. Liu, X. Liu, J. Li, and F. Ren, “Identification and Integrated Analysis of Key Biomarkers for Diagnosis and Prognosis of Non-Small Cell Lung Cancer,” Med Sci Monit, vol. 25, pp. 9280–9289, Dec. 2019.
29. M. Maharjan, R. B. Tanvir, K. Chowdhury, W. Duan, and A. M. Mondal, “Computational identification of biomarker genes for lung cancer considering treatment and non-treatment studies,” BMC Bioinformatics, vol. 21, no. S9, p. 218, Dec. 2020.
30. F. Luo et al., “DOK5 as a Prognostic Biomarker of Gastric Cancer Immunoinvasion: A Bioinformatics Analysis,” BioMed Research International, vol. 2022, no. 1, p. 9914778, Jan. 2022.
31. D.-G. Wang, G. Chen, X.-Y. Wen, D. Wang, Z.-H. Cheng, and S.-Q. Sun, “Identification of Biomarkers for Diagnosis of Gastric Cancer by Bioinformatics,” Asian Pacific Journal of Cancer Prevention, vol. 16, no. 4, pp. 1361–1365, Mar. 2015.
32. W. Wang, Y. He, Q. Zhao, X. Zhao, and Z. Li, “Identification of potential key genes in gastric cancer using bioinformatics analysis,” biom rep, Feb. 2020.
33. X. Liu et al., “Identification of Potential Key Genes Associated With the Pathogenesis and Prognosis of Gastric Cancer Based on Integrated Bioinformatics Analysis,” Front Genet, vol. 9, p. 265, 2018.
34. X. Chong, R. Peng, Y. Sun, L. Zhang, and Z. Zhang, “Identification of Key Genes in Gastric Cancer by Bioinformatics Analysis,” BioMed Research International, vol. 2020, no. 1, p. 7658230, Jan. 2020.
35. J. H. Oh and D. W. Jun, “The latest global burden of liver cancer: A past and present threat,” Clin Mol Hepatol, vol. 29, no. 2, pp. 355–357, Apr. 2023.
36. D. Anwanwan, S. K. Singh, S. Singh, V. Saikam, and R. Singh, “Challenges in liver cancer and possible treatment approaches,” Biochim Biophys Acta Rev Cancer, vol. 1873, no. 1, p. 188314, Jan. 2020.
37. J. Bruix, K.-H. Han, G. Gores, J. M. Llovet, and V. Mazzaferro, “Liver cancer: Approaching a personalized care,” J Hepatol, vol. 62, no. 1 0, pp. S144–S156, Apr. 2015.
38. T.-Y. Ha, “MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease,” Immune Netw, vol. 11, no. 3, pp. 135–154, Jun. 2011.
39. A. Chakrabortty, D. J. Patton, B. F. Smith, and P. Agarwal, “miRNAs: Potential as Biomarkers and Therapeutic Targets for Cancer,” Genes (Basel), vol. 14, no. 7, p. 1375, Jun. 2023.
40. T. B. Nguyen, D. N. Do, T. Nguyen-Thanh, V. B. Tatipamula, and H. T. Nguyen, “Identification of Five Hub Genes as Key Prognostic Biomarkers in Liver Cancer via Integrated Bioinformatics Analysis,” Biology (Basel), vol. 10, no. 10, p. 957, Sep. 2021.

41. L. Li, Q. Lei, S. Zhang, L. Kong, and B. Qin, “Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatics analysis,” Oncology Reports, vol. 38, no. 5, pp. 2607–2618, Nov. 2017.
42. Z. Zhou et al., “Screening Hub Genes as Prognostic Biomarkers of Hepatocellular Carcinoma by Bioinformatics Analysis,” Cell Transplant, vol. 28, no. 1_suppl, pp. 76S-86S, Dec. 2019.
43. Q. Zhang et al., “Integrated Proteomics and Bioinformatics to Identify Potential Prognostic Biomarkers in Hepatocellular Carcinoma,” CMAR, vol. Volume 13, pp. 2307–2317, Mar. 2021.
44. Y.-C. Wang, Z.-B. Tian, and X.-Q. Tang, “Bioinformatics screening of biomarkers related to liver cancer,” BMC Bioinformatics, vol. 22, no. S3, p. 521, May 2021.
45. Y. X. Lim, Z. L. Lim, P. J. Ho, and J. Li, “Breast Cancer in Asia: Incidence, Mortality, Early Detection, Mammography Programs, and Risk-Based Screening Initiatives,” Cancers (Basel), vol. 14, no. 17, p. 4218, Aug. 2022.
46. G. W. Sledge, E. P. Mamounas, G. N. Hortobagyi, H. J. Burstein, P. J. Goodwin, and A. C. Wolff, “Past, Present, and Future Challenges in Breast Cancer Treatment,” J Clin Oncol, vol. 32, no. 19, pp. 1979–1986, Jul. 2014.
47. A. Fawaz, A. Ferraresi, and C. Isidoro, “Systems Biology in Cancer Diagnosis Integrating Omics Technologies and Artificial Intelligence to Support Physician Decision Making,” J Pers Med, vol. 13, no. 11, p. 1590, Nov. 2023.
48. H. Zayed, “The identification of highly up-regulated genes in claudin-low breast cancer through an integrative bioinformatics approach,” Comput Biol Med, vol. 127, p. 103806, Dec. 2020.
49. J.-L. Deng, Y.-H. Xu, and G. Wang, “Identification of Potential Crucial Genes and Key Pathways in Breast Cancer Using Bioinformatics Analysis,” Front Genet, vol. 10, p. 695, 2019.
50. N. Wang, H. Zhang, D. Li, C. Jiang, H. Zhao, and Y. Teng, “Identification of novel biomarkers in breast cancer via integrated bioinformatics analysis and experimental validation,” Bioengineered, vol. 12, no. 2, pp. 12431–12446, Dec. 2021.
51. K. Huang, Y. Zeng, Y. Xie, L. Huang, and Y. Wu, “Bioinformatics analysis of the prognostic value of CCT6A and associated signalling pathways in breast cancer,” Mol Med Rep, vol. 19, no. 5, pp. 4344–4352, May 2019.
52. G. Chen et al., “Identification of candidate biomarkers correlated with poor prognosis of breast cancer based on bioinformatics analysis,” Bioengineered, vol. 12, no. 1, pp. 5149–5161, Dec. 2021.
53. T. Sawicki, M. Ruszkowska, A. Danielewicz, E. Niedźwiedzka, T. Arłukowicz, and K. E. Przybyłowicz, “A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms and Diagnosis,” Cancers (Basel), vol. 13, no. 9, p. 2025, Apr. 2021.
54. A. R. Sepulveda et al., “Molecular Biomarkers for the Evaluation of Colorectal Cancer,” The Journal of Molecular Diagnostics, vol. 19, no. 2, pp. 187–225, Mar. 2017.

55. Y. Jung et al., “Clinical Validation of Colorectal Cancer Biomarkers Identified from Bioinformatics Analysis of Public Expression Data,” Clinical Cancer Research, vol. 17, no. 4, pp. 700–709, Feb. 2011.
56. J. Chen, Z. Wang, X. Shen, X. Cui, and Y. Guo, “Identification of novel biomarkers and small molecule drugs in human colorectal cancer by microarray and bioinformatics analysis,” Mol Genet Genomic Med, vol. 7, no. 7, p. e00713, Jul. 2019.
57. W. Lian et al., “Identification of novel biomarkers affecting the metastasis of colorectal cancer through bioinformatics analysis and validation through qRT-PCR,” Cancer Cell Int, vol. 20, p. 105, 2020.
58. A. Hammad, M. Elshaer, and X. Tang, “Identification of potential biomarkers with colorectal cancer based on bioinformatics analysis and machine learning,” Math Biosci Eng, vol. 18, no. 6, pp. 8997–9015, Oct. 2021.
59. Z. Sun, C. Liu, and S. Y. Cheng, “Identification of four novel prognosis biomarkers and potential therapeutic drugs for human colorectal cancer by bioinformatics analysis,” J Biomed Res, vol. 35, no. 1, pp. 21–35, Jan. 2021.
60. Jain and Limaiem, “Cervical Squamous Cell Carcinoma - StatPearls - NCBI Bookshelf.” Accessed: Jan. 30, 2025.
61. T. Šarenac and M. Mikov, “Cervical Cancer, Different Treatments and Importance of Bile Acids as Therapeutic Agents in This Disease,” Front Pharmacol, vol. 10, p. 484, Jun. 2019.
62. M. Kori and K. Yalcin Arga, “Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective,” PLoS One, vol. 13, no. 7, p. e0200717, 2018.
63. X. Li et al., “Identification of Significant Gene Signatures and Prognostic Biomarkers for Patients With Cervical Cancer by Integrated Bioinformatics Methods,” Technol Cancer Res Treat, vol. 17, p. 1533033818767455, Jan. 2018.
64. K. Wu, Y. Yi, F. Liu, W. Wu, Y. Chen, and W. Zhang, “Identification of key pathways and genes in the progression of cervical cancer using bioinformatics analysis,” Oncol Lett, May 2018.
65. Y. He, S. Hu, J. Zhong, A. Cheng, and N. Shan, “Identification of significant genes signatures and prognostic biomarkers in cervical squamous carcinoma via bioinformatics data,” PeerJ, vol. 8, p. e10386, Dec. 2020, doi: 10.7717/peerj.10386.
66. H. Xue, Z. Sun, W. Wu, D. Du, and S. Liao, “Identification of Hub Genes as Potential Prognostic Biomarkers in Cervical Cancer Using Comprehensive Bioinformatics Analysis and Validation Studies,” Cancer Management and Research, vol. 13, pp. 117–131, Jan. 2021.

67. B. K. Prasanth, S. Alkhowaiter, G. Sawarkar, B. D. Dharshini, and A. R. Baskaran, “Unlocking Early Cancer Detection: Exploring Biomarkers, Circulating DNA, and Innovative Technological Approaches,” Cureus, vol. 15, no. 12, p. e51090.
68. A. Parihar, S. Malviya, and R. Khan, “Identification of Biomarkers Associated with Cancer Using Integrated Bioinformatics Analysis,” in Biomedical Engineering, vol. 12, G. Kais and Y. Hamdi, Eds., IntechOpen, 2022.
69. S. Das, M. K. Dey, R. Devireddy, and M. R. Gartia, “Biomarkers in Cancer Detection, Diagnosis, and Prognosis,” Sensors, vol. 24, no. 1, p. 37, Dec. 2023.
70. I. Ullah et al., “Multi-Omics Approaches in Colorectal Cancer Screening and Diagnosis, Recent Updates and Future Perspectives,” Cancers, vol. 14, no. 22, Art. no. 22, Jan. 2022.
71. F. Dilnawaz and A. K. Behura, Artificial Intelligence Based Cancer Nanomedicine: Diagnostics, Therapeutics and Bioethics. Bentham Science Publishers, 2022.
72. P. Maruvada, W. Wang, P. D. Wagner, and S. Srivastava, “Biomarkers in Molecular Medicine: Cancer Detection and Diagnosis,” BioTechniques, vol. 38, no. sup4, pp. 9–15, Apr. 2005.
73. S. Sethi, S. Ali, P. A. Philip, and F. H. Sarkar, “Clinical Advances in Molecular Biomarkers for Cancer Diagnosis and Therapy,” International Journal of Molecular Sciences, vol. 14, no. 7, Art. no. 7, Jul. 2013.