We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Identification of Biomarkers Associated with Cancer Using Integrated Bioinformatic Analysis

Arpana Parihar, Shivani Malviya and Raju Khan

Abstract

Among the leading cause of death cancer ranked in top position. Early diagnosis of cancer holds promise for reduced mortality rate and speedy recovery. The cancer associated molecules being altered in terms of under/over expression when compared to normal cells and thus could act as biomarkers for therapeutic designing and drug repurposing. The information about the known cancer associated biomarkers can be exploited for targeting of cancer specifically in terms of selective personalized medicine designing. This chapter deals with various types of biomarkers associated with different types of cancer and their identification using integrated bioinformatic analysis. Besides, a brief insight on integrated bioinformatics analysis tools and databases have also been discussed.

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

1. Introduction

Cancer is the dreadful disease in which cells divide uncontrollably and, at a later stage, begin attacking neighboring tissues. Hereditary mutations, toxin exposure, radiation exposure, alcohol usage, smoking, and radical lifestyle changes are all known to cause cancer. Early detection of cancer results in good therapy. The traditional diagnostic procedures of X-ray, CT-scan, and tissue biopsy are unable to detect it at an early stage, resulting in a delay in treatment that has resulted in the death of several people globally due to cancer [1, 2]. Substantial advances in cancer biology have resulted in the discovery of various biomolecules that are especially linked to cancer progression and development, and therefore referred to as "biomarkers." Biomarkers are basically alterations which are cellular, biochemical, and molecular changes that can be used to identify or monitor a normal, abnormal, or just a biological process. They are utilized to test and evaluate pathogenic processes, normal biological processes, and the pharmacological response to a treatment intervention objectively. Biomarkers could be classified based on their chemical nature and functionality that can be identified using transcriptomics, metabolomics, genomics and proteomics (**Figure 1**) [3, 4].
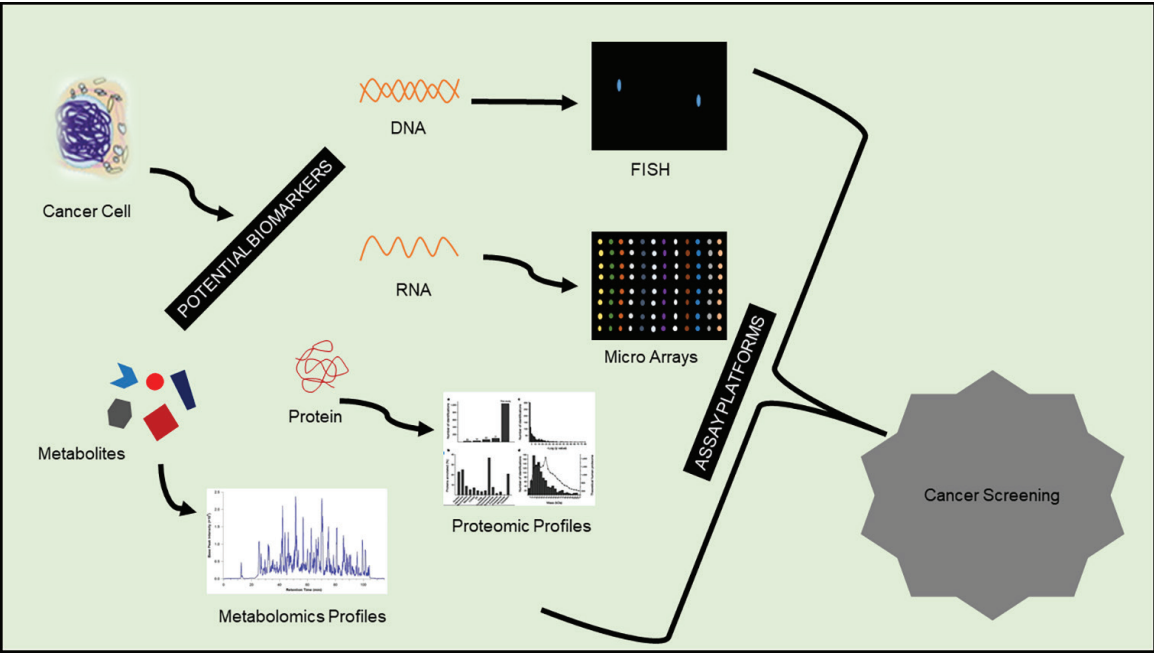
Figure 1.
Analysis of potential biomarkers using different integrated bioinformatics analysis assay platforms such as DNA based from FISH assay platform, RNA based biomarkers from micro arrays, protein based biomarkers from proteomic profiles and metabolites based on biomarkers from metabolomics profiles which led to screening of various kinds of cancer resulting in identification of potential candidate genes for prognostic therapeutic approach.

Usually, living cells have a finite life span, and their genome deoxy ribonucleic acid (DNA) transcribes into ribonucleic acid (RNA), which upon translation results in the creation of proteins that participate in 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 dreadful disease namely, Cancer. The detection of mutations in DNA can be used to predict Cancer risk [5]. Consequently, measurement of RNA, protein, and metabolite expression levels can provide important information about illness progression and profiling. There are more than 200 types of cancer reported, however in this chapter, we gathered and presented information about various biomarkers associated with top 5 types of cancer in the world, which can be exploited in designing of sensitive and effective diagnostic technology for early detection of cancer. Basically, various types of biomarkers associated with different types of cancer and their identification using integrated bioinformatic analysis will be discussed. Besides, a brief insight on integrated bioinformatics analysis tools and databases have also been discussed.

2. Biomarkers associated with different types of cancer

Biomarkers have been generally known to play crucial role in the association with different cancer resulting in therapeutic aspects. These could be constructed with the help of advanced integrated bioinformatics analysis tools which could provide an ease to identify biomarkers which could be treated as potential candidates to treat diversities of Cancer. We have listed biomarkers associated with various types of cancer using integrated bioinformatics approaches in **Table 1**. The mechanistic insight regarding how the databases can be utilized to extract and identify various biomarkers associated with respective cancers have been depicted in **Figure 2**.

S. No.	Type of cancer	Biomarkers identified	Investigators	References
1	Lung Cancer	TOP2A, CCNB1, CCNA2, UBE2C, KIF20A, and IL-6	Ni et al., 2018	[6]
2		CDC20, ECT2, KIF20A, MKI67, TPX2, and TYMS	Dai et al., 2020	[7]
3		DDX5, DDX11, DDX55 and DDX56	Cui et al., 2021	[8]
4		NDC80, BUB1B, PLK1, CDC20, and MAD2L1	Liao et al., 2019	[9]
5		UBE2T, UNF2, CDKN3, ANLN, CCNB2, and CKAP2L	Tu et al., 2019	[10]
6		UBE2C, AURKA, CCNA2, CDC20, CCNB1, TOP2A, ASPM, MAD2L1, and KIF11	Liu et al., 2020	[11]
7	Gastric Cancer	CST2, AADAC, SERPINE1, COL8A1, SMPD3, ASPN, ITGBL1, MAP7D2, and PLEKHS1	Liu et al., 2018	[12]
8		FN1, COL1A1, INHBA, and CST1	Wang et al., 2020	[13]
9		COL1A2	Rong et al., 2018	[14]
10		LINC01018, LOC553137, MIR4435-2HG, and TTTY14	Miao et al., 2017	[15]
11		UCA1, HOTTIP, and HMGA1P4	Zang et al., 2019	[16]
12	Liver Cancer	PBK, ASPM, NDC80, AURKA, TPX2, KIF2C, and centromere protein F	Ji et al., 2020	[17]
13		miR1055p, miR7675p, miR12665p, miR47465p, miR500a3p, miR11803p, and miR1395p	Shen et al., 2020	[18]
14		BUB1, CCNB2, CDC20, CDK1, KIF20A, KIF2C, RACGAP1 and CEP55	Li et al., 2017	[19]
15	Breast Cancer	TXN, ANXA2, TPM4, LOXL2, TPRN, ADCY6, TUBA1C, and CMIP	Wang et al., 2019	[20]
16		ADH1A, IGSF10, and the 14 microRNAs	Wu et al., 2021	[21]
17		TPX2, KIF2C, CDCA8, BUB1B, and CCNA2	Cai et al., 2019	[22]
18		CDC45, PLK1, BUB1B, CDC20, AURKA and MAD2L1	Wu et al., 2020	[23]
19	Colorectal Cancer	SLC4A4, NFE2L3, GLDN, PCOLCE2, TIMP1, CCL28, SCGB2A1, AXIN2, and MMP1	Chen et al., 2019	[24]
20		BLACAT1	Dai et al., 2017	[25]
21		HMMR, PAICS, ETFDH, and SCG2	Sun et al., 2021	[26]
22		hsa-miR-183-5p, hsa-miR-21-5p, hsa-miR-195-5p and hsa-miR-497-5p	Falzone et al., 2018	[27]

Table 1.
Biomarkers identified by using integrated bioinformatics tools, associated with various types of cancer such as lung cancer, gastric cancer, colorectal cancer.
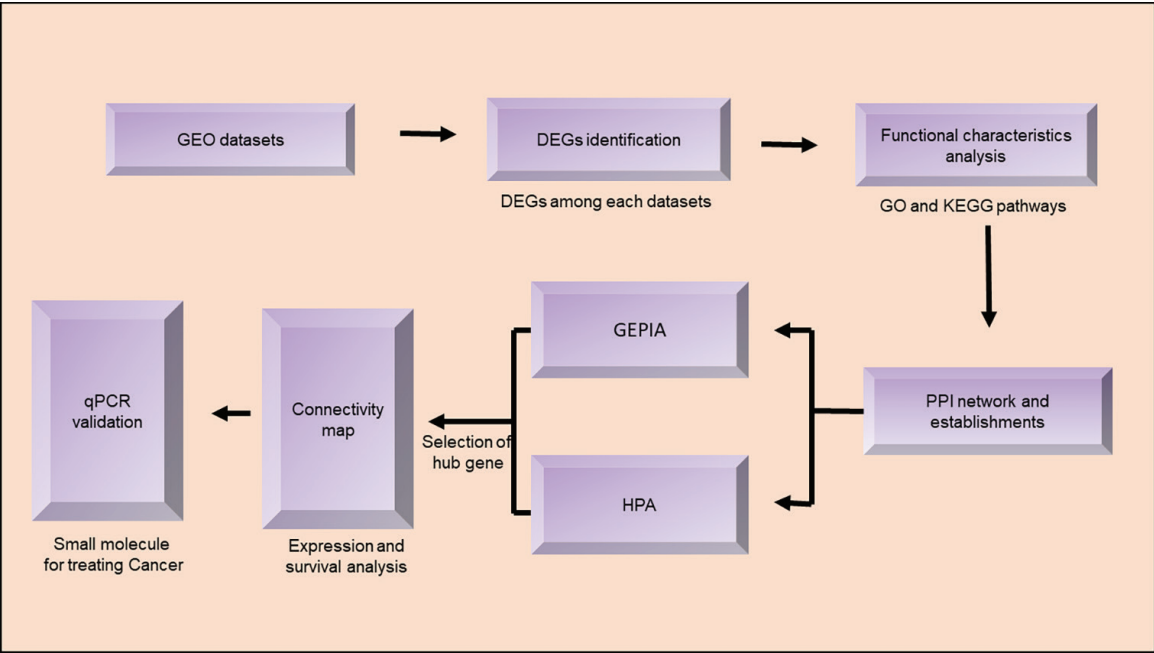
Figure 2.
The schematic representation of extraction of datasets from the GEO database then the identification of DEGs followed by its functional analysis and subsequent qPCR validation leading to identification of small molecule known as biomarker for treating Cancer.

2.1 Lung cancer

Lung cancer is the most common 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. Therefore, here in this section of chapter we have reviewed studies led by certain researchers for identification of the lung cancer biomarkers using integrated bioinformatics analysis. There are mainly 2 types of the lung cancer. In 80–85% cases, the type of lung cancer is non-small cell lung cancer (NSCLC). The main subtypes of which are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. These subtypes generally begin from different types of the lung cells that are grouped together as NSCLC and their treatment and prognoses are almost similar. The other type is small cell lung cancer (SCLC) and around 10–15% of all lung cancers are SCLC and it is sometimes called oat cell cancer. SCLC grows and spread faster than NSCLC.

In a study by Ni et al., four GEO datasets GSE18842, GSE19804, GSE43458, and GSE62113, were extracted form Gene Expression Omnibus (GEO) database into which the limma package was used to assess differentially expressed genes (DEGs) between NSCLC and normal samples, and the RobustRankAggreg (RRA) programme was used to undertake gene integration. Furthermore, they established the protein–protein interaction (PPI) network of these DEGs using the Search Tool for the Retrieval of Interacting Genes database (STRING), Cytoscape, and Molecular Complex Detection (MCODE). Funrich (http://www.funrich.org)and OmicShare (https://www.omicshare.com/tools/)were also conducted to ensure functional

enrichment and pathway enrichment analysis for DEGs. Besides this, they used the gene Expression Profiling Interactive Analysis (GEPIA) and Kaplan Meier-plotter (KM) online datasets to analyze the expressions and prognostic values of top genes. Hence, it led to the identification of a total of 249 DEGs including 113 upregulated and 136 downregulated after gene integration. Followed by this, they established a PPI network with 166 nodes and 1784 protein pairings resulting in TOP2A, CCNB1, CCNA2, UBE2C, KIF20A, and IL-6 to be considered as possible important genes, whereas they further added, the mitotic cell cycle pathway to play a crucial role in NSCLC advancement resulting in its employment as a novel biomarker for NSCLC diagnosis and to guide synthesis medication [6].

In another study by Dai et al., 6 key biomarkers associated with non- small cell lung cancer in which GEO2R were analyzed to examine three microarray datasets from the Gene Expression Omnibus collection along with the enrichment analysis which was performed using Gene Ontology and the Kyoto Encyclopedia of Genes and Genomes. Further, the String database, Cytoscape, and the MCODE plug-in were then used to build a PPI network and screen hub genes using the String database, Cytoscape, and the MCODE plug-in. Kaplan–Meier curves were used to examine overall and disease-free survival of hub genes, as well as the association between target gene expression patterns and tumor grades. To verify enrichment pathways and diagnostic effectiveness of hub genes, researchers performed gene set enrichment analysis and receiver operating characteristic curves. A total of 293 differentially expressed genes were discovered, with cell cycle, ECM–receptor interaction, and malaria being the most prevalent. The PPI network identified 36 hub genes, six of which were reported to have important roles in NSCLC (non- small cell lung cancer) carcinogenesis: CDC20, ECT2, KIF20A, MKI67, TPX2, and TYMS. The target genes discovered can be employed as potential biomarkers to identify and diagnose non-small cell lung cancer as per their investigations [7].

Similarly, in another study by Cui et al., they used integrated bioinformatic analysis of multivariate large-scale databases to assess the potential of DEAD/H box helicases as prognostic indicators and therapeutic targets in lung cancer. They were able to discover four biomarkers with the most significant changes after analyzing the survival and differential expression of these helicases. The unfavorable prognostic factors DDX11, DDX55, and DDX56, as well as the good prognosis factor DDX5, were discovered. MYC signaling is adversely linked with DDX5 gene expression, but favorably associated with DDX11, DDX55, and DDX56 gene expression, according to pathway enrichment analysis led by them. Low mutation levels of TP53 and MUC16, the two most frequently mutated genes in lung cancer, are related with high expression levels of the DDX5 gene. High levels of DDX11, DDX55, and DDX56 gene expression, on the other hand, were linked to high levels of TP53 and MUC16 mutation. The levels of DDX5 gene expression in tumor-infiltrated CD8 + T and B cells are positively correlated, but the other three DEAD box helicases are negatively correlated. Furthermore, while each DDX has a unique miRNA signature, the DDX5-associated miRNA profile is distinct from the miRNA profiles of DDX11, DDX55, and DDX56. The discovery of these four DDX helicases as biomarkers could be considered useful for lung cancer prognostication and targeted treatment development [8].

In another study by Liao et al., they have identified candidate genes associated with the pathogenesis of small cell lung cancer analyzed using integrated bioinformatics tools. GSE60052, GSE43346, GSE15240, and GSE6044 were the four datasets that they downloaded from the Gene Expression Omnibus. R software was used to examine the differentially expressed genes (DEGs) between the SCLC and normal

samples. For each dataset, the limma software was utilized. The DEGs from the four datasets were combined using the RobustRankAggreg package. FunRich software and R software were used to conduct functional and route enrichment analyses using the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes databases, accordingly. The DEGs' protein–protein interaction (PPI) network was also built using the STRING database and the Cytoscape software. Molecular Complex Detection in Cytoscape software was used to find hub genes and important modules. Ultimately, the Oncomine online database was used to assess the expression values of hub genes. Following the integration of the four datasets, 412 DEGs were discovered, comprising 146 upregulated genes and 266 downregulated genes. The increased DEGs were mostly involved in cell division, cell cycle, and microtubule binding. The complement and coagulation cascades, the cytokine-mediated signaling pathway, and protein binding were all heavily represented among the downregulated DEGs. Based on a subset of the PPI network, eight hub genes and one major module connected to the cell cycle pathway were discovered. Eventually, in comparison to normal tissue, five hub genes were shown to be substantially expressed in SCLC tissue. The cell cycle route may be the one that is most closely linked to SCLC pathophysiology. As a result, follow-up studies in the diagnosis and therapy of SCLC should focus on NDC80, BUB1B, PLK1, CDC20, and MAD2L1 [9].

In another similar study by Tu et al., GEO2R was used to search the mRNA microarray datasets GSE19188, GSE33532, and GSE44077 for differentially expressed genes (DEGs). The DEGs were analyzed for functional and pathway enrichment using the DAVID database. STRING was used to create a protein–protein interaction (PPI) network, which was then displayed in Cytoscape. MCODE was used to analyze the PPI network's modules. The Kaplan Meier-plotter was used to analyze the overall survival (OS) of genes from MCODE. Total of 221 DEGs were found, with words linked to cell division, cell proliferation, and signal transduction being the most abundant. A PPI network with 221 nodes and 739 edges was created. The PPI network revealed a substantial module containing 27 genes. UBE2T, UNF2, CDKN3, ANLN, CCNB2, and CKAP2L all have high expression levels and have been linked to a poor prognosis in NSCLC patients. Protein binding, ATP binding, cell cycle, and the p53 signaling pathway were among the enriched functions and pathways. DEGs in non- small cell lung cancer (NSCLC) have the potential to be useful targets for diagnosing and treating the disease [10].

In another study by Liu et al., in this prospective investigation, which included 46 tumors and 45 controls, the gene expression profile GSE18842 was acquired from the Gene Expression Omnibus database. They used functional enrichment analysis and KEGG analysis using upregulated differentially expressed genes (uDEGs) and downregulated differentially expressed genes (dDEGs), respectively, after screening differentially expressed genes (DEGs). The STRING database was used to create protein–protein interaction (PPI) networks between DEGs and their corresponding coding protein complexes, which were then examined using Cytoscape. The Kaplan–Meier approach was used to confirm the survival of hub genes. In the TCGA database, the gene expression level heat map of hub genes between NSCLC and neighboring lung tissues was plotted using the GEPIA webserver. After gene integration, they found 368 DEGs (168 uDEGs and 200 dDEGs) in NSCLC samples compared to control samples. They built a PPI network for the DEGs with 249 nodes and 1472 protein pairings on the edges. Survival study confirmed that ten undefined hub genes with the highest connectivity degree (CDK1, UBE2C, AURKA, CCNA2, CDC20, CCNB1, TOP2A, ASPM, MAD2L1, and KIF11) were related with lower overall survival in

NSCLC. The GEPIA web tool was used to verify the expression dependability of hub genes. The findings suggested that UBE2C, AURKA, CCNA2, CDC20, CCNB1, TOP2A, ASPM, MAD2L1, and KIF11 are inherent critical biomarkers for diagnosis and prognosis, and that the mitotic cell cycle pathway is a likely signaling pathway contributing to NSCLC progression, according to KEGG analysis. Such genes could be useful diagnostic biomarkers, as well as a new strategy to designing targeted NSCLC treatments [11].

2.2 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 [29, 30]. The fact that most patients are discovered at an advanced stage, even with metastatic illnesses, and thus miss out on the potential for a curative resection, accounts for the poor 5-year survival in GC [31, 32]. 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 [32, 33]. Hence, due to molecular complexity, poor prognosis, and significant reoccurrence of GC, new diagnostic and prognostic biomarkers are urgently needed [34, 35]. 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 [36]. 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 [37–39].

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 Liu et al., they have considered TOP2A, COL1A1, COL1A2, NDC80, COL3A1, CDKN3, CEP55, TPX2, and TIMP1 which are nine hub genes that may be linked to the etiology of GC. Hence, CST2, AADAC, SERPINE1, COL8A1, SMPD3, ASPN, ITGBL1, MAP7D2, and PLEKHS1 were used to construct a prognostic gene signature that performed well in predicting overall survival. An integrated analysis of several gene expression profile datasets was used by them to find differentially expressed genes between GC and normal gastric tissue samples. Furthermore, protein–protein interaction network and Cox proportional hazards model studies were used to identify key genes related to the pathophysiology and prognosis of GC resulting in their constructed gene signature to be considered as a potential candidate for the biomarker to facilitate the molecular targeting therapy of GC [12].

In a study by Wang et al., they discovered promising biomarkers that could be used to diagnose GC patients. Four Gene Expression Omnibus (GEO) datasets were obtained and examined for differentially expressed genes to look for possible treatment targets for GC (DEGs). The function and pathway enrichment of the discovered DEGs were then investigated using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. A network of protein–protein interactions (PPI) was created. The degree of connection of proteins in the PPI network was calculated using the CytoHubba plugin of Cytoscape, and the two genes with the highest degree

of connectivity were chosen for further investigation. The two DEGs with the highest and lowest log Fold Change values were also chosen. Oncomine and the KaplanMeier plotter platform were used to investigate these six important genes further. A total of 99 genes that were upregulated and 172 genes that were downregulated across all four GEO datasets were examined. The Biological Process phrases 'extracellular matrix organization,' 'collagen catabolic process,' and 'cell adhesion' were primarily enriched in the DEGs. The categories 'ECMreceptor interaction,' 'protein digestion and absorption,' and 'focal adhesion' were considerably enriched in these three KEGG pathways. According to Oncomine, ATP4A and ATP4B expression were downregulated in GC, while all other genes were increased. Upregulated expression of the identified important genes was substantially associated with worse overall survival of GC patients, according to the KaplanMeier plotter platform. The current findings imply that FN1, COL1A1, INHBA, and CST1 could be used as gastric cancer biomarkers and treatment targets. Additional research is needed to determine the role of ATP4A and ATP4B in the treatment of gastric cancer [13].

In another study by Rong et al., their research outlines an integrated bioinformatics approach to identifying molecular biomarkers for stomach cancer in cancer tissues of patients. In large gastric cancer cohorts, they reported distinct expression genes from Gene Expression Ominus (GEO). Their findings found that 433 genes in human stomach cancer have significantly distinct expression patterns. Bioinformatic studies and co-expression network design were used to confirm the different gene expression profiles in gastric cancer. They identified collagen type I alpha 2 (COL1A2), which encodes the pro-alpha2 chain of type I collagen whose triple helix comprises two alpha1 chains and one alpha2 chain, as the key gene in a 37-gene network that modulates cell motility by interacting with the cytoskeleton, based on the co-expression network and top-ranked genes. Immunohistochemistry on human gastric cancer tissue was also used to investigate the predictive function of COL1A2. When compared to normal gastric tissues, COL1A2 was substantially expressed in human gastric cancer. The level of COL1A2 expression was found to be substantially related to histological type and lymph node status after statistical analysis. There were no links found between COL1A2 expression and age, lymph node count, tumor size, or clinical stage. Finally, the unique bioinformatics used in this study led to the discovery of improved diagnostic biomarkers for human stomach cancer, which could aid future research into the crucial change that occurs during the disease's course [14].

In another study, the goal of their research is to find an lncRNA-related signature that can be used to assess the overall survival of 379 GC patients from The Cancer Genome Atlas (TCGA) database. The univariate and multivariate Cox proportional hazards regression models were used to assess the correlations between survival outcome and the expression of lncRNAs. Overall survival was found to be substantially linked with four lncRNAs (LINC01018, LOC553137, MIR4435-2HG, and TTTY14). These four lncRNAs were combined to form a prognostic signature. There was a strong favorable link between overall survival and GC patients with low-risk scores (P = 0.001). Subsequent research found that the predictive usefulness of this four-lncRNA pattern was unaffected by clinical characteristics. These four lncRNAs were linked to many tumor molecular pathways, according to gene set enrichment analysis. Based on bioinformatics analysis, their research suggests that this unique lncRNA expression pattern could be a helpful diagnostic of prognosis for GC patients [15].

The researchers wanted to see if there were any long noncoding RNAs (lncRNAs) that were linked to the pathophysiology and prognosis of GC. The Gene Expression Omnibus (GEO) database was used to retrieve raw noncoding RNA microarray

data (GSE53137, GSE70880, and GSE99417). After gene reannotation and batch normalization, an integrated analysis of various gene expression profiles was used to screen for differentially expressed genes between GC and neighboring normal stomach tissue samples. The Cancer Genome Atlas (TCGA) database validated the presence of differentially expressed genes. To identify hub lncRNAs and explore possible biomarkers related to GC diagnosis and prognosis, researchers used a competitive endogenous RNA (ceRNA) network, Gene Ontology term, and Kyoto Encyclopedia of Genes and Genomes pathway, as well as survival analysis. After intersections of differential genes between the GEO and TCGA databases, a total of 246 integrated differential genes were identified, including 15 lncRNAs and 241 messenger RNAs (mRNAs). Three lncRNAs (UCA1, HOTTIP, and HMGA1P4), 26 microRNAs (miRNAs), and 72 mRNAs make up the ceRNA network. Three lncRNAs controlled the cell cycle and cellular senescence, according to functional analyses. The survival rate of HMGA1P4 was statistically connected to the total survival rate, according to a survival analysis. They discovered that HMGA1P4, a miR-301b/miR-508 target, regulates CCNA2 in the GC and is implicated in cell cycle and senescence. Ultimately, three lncRNAs' expression levels were shown to be elevated in GC tissues. As a result, three lncRNAs, UCA1, HOTTIP, and HMGA1P4, may play a role in GC development, and their possible functions may be linked to GC prognosis [16].

2.3 Liver cancer

Liver cancer is among the most frequent malignancies in the world, and it is the second largest cause of cancer death [40, 41]. Due to advances in detection and therapy, people with liver cancer still have a terrible prognosis. 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. 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 [42]. 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. 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 researches 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. Hence, in an investigation led by Ji et al., they retrieved three microarray datasets from the Gene Expression Omnibus (GEO) collection, including 170 tumoral samples and 181 adjacent normal tissues from the liver of patients with HBV-related HCC which were subjected to integrated analysis of differentially expressed genes (DEGs). Following that, the protein–protein interaction network (PPI) and function and pathway enrichment analyses were carried out. The expression profiles and survival analyses of the ten hub genes selected from the PPI network were carried out. Overall, 329 DEGs were discovered in which 67 were upregulated and 262 were downregulated. PDZ-binding kinase (PBK), abnormal spindle microtubule

assembly (ASPM), nuclear division cycle 80 (NDC80), aurora kinase A (AURKA), targeting protein for xenopus kinesin-like protein 2 (TPX2), kinesin family member 2C (KIF2C), and centromere protein F were among the ten DEGs with the highest degree of connectivity (CENPF). Overexpression levels of KIF2C and TPX2 were linked to both poor overall survival and relapse-free survival in a Kaplan–Meier study. Therefore, the hub genes identified in this investigation could be useful in the diagnosis, prognosis, and treatment of HBV-related HCC. Furthermore, their research identifies a number of important biological components (e.g., extracellular exosomes) and signaling pathways that are involved in the progression of HCC caused by HBV, providing a more thorough understanding of the mechanisms underlying HBV-related HCC [17].

In another study by Shen et al., they created nine co-expression modules and discovered that in liver cancer, miR1055p, miR7675p, miR12665p, miR47465p, miR500a3p, miR11803p, and miR1395p were differentially expressed. These miR-NAs were found to have a strong link to the prognosis of patients with liver cancer. MiR1055p and miR1395p may be considered separate prognostic variables among them. As a result, seven miRNAs could be used as predictive indicators in the case of liver cancer [18].

In another study by Li et al., The GSE19665, GSE33006, and GSE41804 microarray datasets were obtained from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were found and function enrichment analyses were carried out. STRING and Cytoscape were used to create the protein–protein interaction network (PPI) and perform module analysis. There were a total of 273 DEGs found, with 189 downregulated genes and 84 upregulated genes. Protein activation, complement activation, carbohydrate binding, complement and coagulation cascades, mitotic cell cycle, and oocyte meiosis are among the DEGs' enhanced activities and pathways. A biological process study found that these genes were primarily abundant in cell division, cell cycle, and nuclear division. BUB1, CDC20, KIF20A, RACGAP1 and CEP55 were found to be involved in the carcinogenesis, invasion, and recurrence of HCC in a survival analysis. Finally, the DEGs and hub genes discovered in this work contribute to our understanding of the molecular pathways underlying HCC carcinogenesis and development, as well as providing candidate targets for HCC diagnosis and treatment [19].

2.4 Breast cancer

Breast cancer is becoming more common over the world, and it is now considered a serious disease among women. Asia has recently emerged as a high-risk location for breast cancer, ranking first among female malignant tumors [43, 44]. Breast cancer therapy has improved recently as a result 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 [43, 45]. 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 [46, 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 Wang et al. they have analyzed gene expression profiles of GSE48213 using Gene Expression Omnibus database. Further, validation was done using RNA-seq data and clinical information on breast cancer from The Cancer Genome Atlas. In their study, they identified the gene co-expression network which revealed four modules, one of which was found to be strongly linked with patient survival time. They found that the black module which was found to be basal, was made up of 28 genes; the dark red module which was found to be claudin-low, was made up of 18 genes; the brown module which was found to be luminal, was made up of nine genes; and the midnight blue module was made up of seven genes which was investigated to be nonmalignant. Due to a considerable difference in survival time between the two groups, these modules were clustered into two groups. Hence, TXN and ANXA2 in the nonmalignant module, TPM4 and LOXL2 in the luminal module, TPRN and ADCY6 in the claudin-low module, and TUBA1C and CMIP in the basal module were identified by them as the genes with the highest betweenness, implying that they play a central role in information transfer in the network. Therefore, TXN, ANXA2, TPM4, LOXL2, TPRN, ADCY6, TUBA1C, and CMIP are eight hub genes that have been identified and validated by them as being linked to breast cancer progression and poor prognosis to be considered [20].

In another study by Wu et al., Differentially expressed genes (DEGs) in breast cancer were discovered using three data sets from the GEO database. The functional roles of the DEGs were determined using Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes pathway studies. They also used the Gene Expression Profiling Interactive Analysis (GEPIA), Oncomine, Human Protein Atlas, and Kaplan Meier plotter tool databases to look at the translational and protein expression levels, as well as survival statistics, of DEGs in patients with breast cancer. Using miRWalk and TargetScan, the corresponding change in the expression level of microRNAs in DEGs was predicted, and the expression profiles were evaluated using OncomiR. Finally, RT-qPCR was used to confirm the expression of new DEGs in Chinese breast cancer tissues. ADH1A, IGSF10, and the 14 microRNAs have all been identified as promising new biomarkers for breast cancer diagnosis, therapy, and prognosis [21].

In another study by Cai et al., the Gene Expression Omnibus (GEO) database was used to obtain GSE102484 gene expression profiles. The most potent gene modules related with the metastatic risk of breast cancer were found using weighted gene co-expression network analysis (WGCNA), which yielded a total of 12 modules. 21 network hub genes (MM > 0.90) were kept for further analysis in the most significant module ($R2 = 0.68$). The biomarkers with the greatest interactions in gene modules were then investigated further using protein–protein interaction (PPI) networks. Five hub genes (TPX2, KIF2C, CDCA8, BUB1B, and CCNA2) were identified as important genes associated with breast cancer progression by the PPI networks. Furthermore, using data from The Cancer Genome Atlas (TCGA) and the Kaplan–Meier (KM) Plotter, the predictive value and differential expression of these genes were confirmed. The mRNA expression levels of these five hub genes have excellent diagnostic value for breast cancer and surrounding tissues, according to a Receiver Operating Characteristic (ROC) curve study. Furthermore, KM Plotter revealed that these five hub genes were substantially related with lower distant metastasis-free survival (DMFS) in the patient group. Five hub genes (TPX2, KIF2C, CDCA8, BUB1B, and CCNA2) linked to the likelihood of distant metastasis were extracted for future study and could be employed as biomarkers to predict breast cancer distant metastasis [22].

In another study by Wu et al., there were a total of 215 DEGs found, with 105 upregulated genes and 110 downregulated genes. The enriched keywords and pathways were primarily linked to cell cycle, proliferation, drug metabolism, and oncogenesis, according to GO and KEGG analyses. Cell Division Cycle 45 (CDC45), Polo Like Kinase 1 (PLK1), BUB1 Mitotic Checkpoint Serine/Threonine Kinase B (BUB1B), Cell Division Cycle 20 (CDC20), Aurora Kinase A (AURKA), and Mitotic Arrest Deficient 2 Like 1 were identified as hub genes from the PPI network (MAD2L1). These hub genes' resilience was confirmed by survival analysis and expression validation tests [23].

2.5 Colorectal cancer

CRC (colorectal cancer) 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 [48]. Recent breakthroughs in our understanding of the molecular underpinnings and cellular mechanisms of CRC have resulted in the widespread use of particular 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. Here, we have reviewed a few investigations led to ensure its attainment using integrated bioinformatics analysis [49].

In an investigation led by Chen et al., they analyzed 207 common DEGs in colorectal cancer using the integrated GEO and TCGA databases into which they constructed a PPI network consists of 70 nodes and 170 edges and identified 10 top hub genes. A prognostic gene signature which includes SLC4A4, NFE2L3, GLDN, PCOLCE2, TIMP1, CCL28, SCGB2A1, AXIN2, and MMP1 was constructed by them which revealed overall survival in patients suffering from CRC. Hence, it could be considered as a good potential candidate for further treatments [24].

In a study by Dai et al., they discovered nine differentially expressed lncRNAs and their putative mRNA targets using integrated data mining. They evaluated key pathways and GO words that are associated to the up-regulated and down-regulated transcripts, respectively, after a series of bioinformatics investigations. Meanwhile, qRT-PCR was used to validate the nine lncRNAs in 30 matched tissues and cell lines, and the results were largely compatible with the microarray data. They also looked for nine lncRNAs in the blood of 30 CRC patients with tissue matching, 30 non-cancer patients, and 30 healthy people. Finally, they discovered that BLACAT1 was important for CRC diagnosis. Between CRC patients and healthy controls, the area under the curve (AUC), sensitivity, and specificity were 0.858 (95% CI: 0.765–0.951), 83.3%, and 76.7%, respectively. Furthermore, BLACAT1 exhibited a particular utility in distinguishing CRC from non-cancer disorders. The findings suggest that significantly elevated lncRNAs as well as associated potential target transcripts could be used as therapeutic targets in CRC patients. Conversely, the lncRNA BLACAT1 could be a new supplemental biomarker for CRC detection [25].

In another study by Sun et al., The Gene Expression Omnibus (GEO) mRNA microarray datasets GSE113513, GSE21510, GSE44076, and GSE32323 were collected

and processed with bioinformatics to discover hub genes in CRC development. The GEO2R tool was used to look for differentially expressed genes (DEGs). The DAVID database was used to conduct gene ontology (GO) and KEGG studies. To build a protein–protein interaction (PPI) network and identify essential modules and hub genes, researchers employed the STRING database and Cytoscape software. The DEGs' survival studies were done using the GEPIA database. Potential medications were screened using the Connectivity Map database. There were a total of 865 DEGs found, with 374 upregulated and 491 downregulated genes. These DEGs were mostly linked to metabolic pathways, cancer pathways, cell cycle pathways, and so on. With 863 nodes and 5817 edges, the PPI network was discovered. HMMR, PAICS, ETFDH, and SCG2 were found to be strongly linked with overall survival of CRC patients in a survival analysis. Blebbistatin and sulconazole have also been discovered as potential treatments [26].

Falzone et al. used the mirDIP gene target analysis in a sample of 19 differentially expressed miRNAs to determine the interaction between miRNAs and the most changed genes in CRC. DIANA-mirPath prediction analysis was used to identify miRNAs that can activate or inhibit genes and pathways involved in colorectal cancer development. As a whole, these studies found that the up-regulated hsa-miR-183-5p and hsa-miR-21-5p, as well as the down-regulated hsa-miR-195-5p and hsa-miR-497-5p, were linked to colorectal cancer development via interactions with the Mismatch Repair pathway and the Wnt, RAS, MAPK, PI3K, TGF-, and p53 signaling pathways [27].

3. 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 2** along with database links. The biomarkers associated with different types of Cancers identified with the help of integrated bioinformatics tools depicted in **Figure 3**.

3.1 Microarray and RNASeq data collection

The microarray data collection is done using the GEO database which refers to Gene Expression Omnibus. It could be 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. As per our review of various studies which are aforementioned in this chapter, various microarray datasets have been collected using the GEO database and then processed with bioinformatics to discover hub genes. Several 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 nowadays considered to be the most up to date technology to analyze gene expression. Into this technology, 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 is present in biological sample (a cDNA library). This is accomplished using reverse transcription, allowing the RNA

S. No.	Name of database	Link/URL
1	Gene Expression Omnibus (GEO)	http://www.ncbi.nlm.nih.gov/geo/
2	GEO2R	http://www.ncbi.nlm.nih.gov/geo/geo2r/
3	DAVID	http://david.abcc.ncifcrf.gov/
4	STRING	http://www.bork.embl-heidelberg.de/STRING/
5	Cytoscape	http://www.cytoscape.org/
6	GEPIA	http://gepia2021.cancer-pku.cn/
7	TGCA	https://tcga-data.nci.nih.gov/tcga/
8	Kaplan–Meier (KM) Plotter	http://kmplot.com/analysis/
9	DIANA-mirPath	http://www.microrna.gr/miRPathv3
10	mirDIP	http://ophid.utoronto.ca/mirDIP
11	GOplot	http://cran.r-project.org/web/packages/GOplot
12	clueGO	http://apps.cytoscape.org/apps/cluego
13	MCODE	http://baderlab.org/Software/MCODE
14	GTEx	https://gtexportal.org
15	Oncomine	http://www.oncomine.org/resource/login.html
16	Human Protein Atlas	www.proteinatlas.org
17	miRWalk	http://mirwalk.uni-hd.de/
18	TargetScan	www.targetscan.org
19	OncomiR	http://www.oncomir.org/oncomir/search_target_miR.html

Table 2.
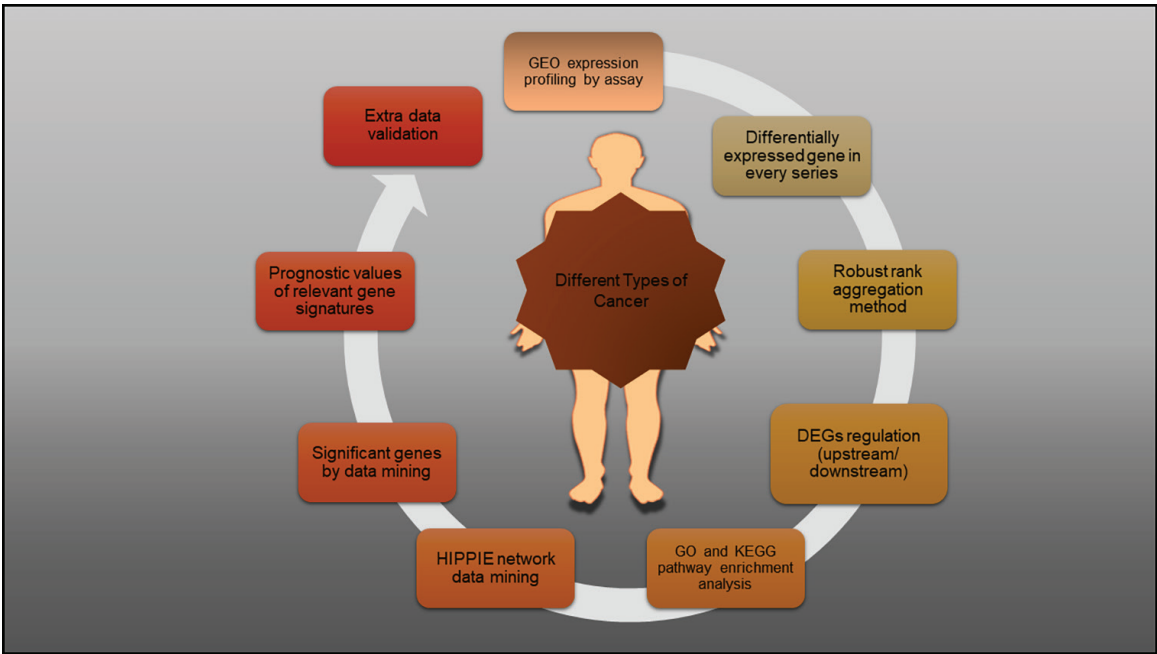*List of databases used for data mining.*

Figure 3.
Mechanistic insight of extraction, construction and identification of biomarkers associated with different kinds of cancers with the help of integrated bioinformatics tools.

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 adopters which allowed sequencing. The cDNA library is evaluated by NGS after amplification, size selection, clean-up, 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. Sequencing can be done in one of two ways: single-end or paired-end. Single-read sequencing is a less expensive and faster method of sequencing cDNA fragments from only one end (approximately 1% of the cost of Sanger sequencing). While paired-end approaches are more expensive since they sequence from both ends, but they provide advantages in post-sequencing data reconstruction. 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. A bioinformatics workflow is developed to discover various alternative biomarkers via LC- MS/MS technique (liquid chromatography coupled tandem mass spectrometry). Further, open Mass spectrometry Search Algorithm is used against the customized alternative splicing database along with the preferred cancer plasma proteome for the identification of respective biomarker [50, 51].

3.2 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 is been utilized to screen out these DEGs.

3.3 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 GOplot 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. [52].

3.4 Construction of the PPI network and analysis of the module

After the enrichment analysis, the PPI network is being built upon using the STRING (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.

3.5 Survival analysis and validation of hub gene expression

At last. The Cancer Genome Atlas (https://tcga-data.nci.nih.gov/tcga/), was utilized to examine the association between important gene expression and survival

of patients with PTC (Papillary thyroid carcinoma). 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 Oncomine, 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 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. Finally, RT-qPCR has been used to confirm the expression of new DEGs. Hence, the constructed biomarkers could be treated as potential candidates for various kinds of Cancers.

4. Challenges and future outlook

The development of biomarkers for early detection cancer screening and therapy monitoring has biological as well as financial hurdles. The majority of existing cancer detection tools only detect late stage or fully grown cancer, not premalignant or early abnormalities that can be resected and treated. Despite the fact that 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 [53]. 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. As a result, developing biomarkers using genomic and proteomic methods could help to solve the variability challenges [3]. 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 [54]. This addresses the possibility of whether any screening strategy should focus solely on early lesions or should additionally consider the tumor's behavior. In the last two decades, detailed and comprehensive knowledge of cancer at the cellular and molecular levels has increased dramatically and exponentially, resulting in significant improvements in the characterization of human tumors, which has catalyzed a shift toward the development of targeted therapies, the foundation of molecular diagnostics [55, 56]. 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. 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.

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. Clinicians and researchers must have a thorough understanding of the molecular aspects, clinical utility, and reliability of biomarkers in order to determine whether or not a biomarker is clinically useful for patient care and whether or not 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. Conclusions

Research in the field of cancer-specific biomarkers have 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 bio-markers found to be potentially very sensitive approaches for discriminating between cancerous 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.

Author details

Arpana Parihar[1*], Shivani Malviya[2] and Raju Khan[1,3]

1 Microfluidics and MEMS Centre, CSIR-Advanced Materials and Processes Research Institute (AMPRI), Bhopal, India

2 Department of Biochemistry and Genetics, Barkatullah University, Bhopal, Madhya Pradesh, India

3 Academy of Scientific and Innovative Research (AcSIR), CSIR-AMPRI, Bhopal, India

*Address all correspondence to: arpana_parihar@yahoo.com

IntechOpen

References

[1] Bertram JS. The molecular biology of cancer. Molecular Aspects of Medicine. 2000;**21**(6):167-223

[2] Zitvogel L, Tesniere A, Kroemer G. Cancer despite immunosurveillance: Immunoselection and immunosubversion. Nature Reviews Immunology. 2006;**6**(10): 715-727

[3] Wagner PD, Verma M, Srivastava S. Challenges for biomarkers in cancer detection. Annals of the New York Academy of Sciences. 2004;**1022**(1):9-16

[4] Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nature Reviews Cancer. 2005;**5**(11):845-856

[5] Ames BN, Gold LS, Willett WC. The causes and prevention of cancer. Proceedings of the National Academy of Sciences. 1995;**92**(12):5258-5265

[6] Ni M, Liu X, Wu J, Zhang D, Tian J, Wang T, et al. Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. Frontiers in Genetics. 2018;**9**. DOI: 10.3389/fgene. 2018.00469

[7] Dai B, Ren L-Q, Han X-Y, Liu D-J. Bioinformatics analysis reveals 6 key biomarkers associated with non-small-cell lung cancer. The Journal of International Medical Research. 2020;**48**(3):1-14. DOI: 10.1177/ 0300060519887637

[8] Cui Y, Hunt A, Li Z, Birkin E, Lane J, Ruge F, et al. Lead DEAD/H box helicase biomarkers with the therapeutic potential identified by integrated bioinformatic approaches in lung cancer. Computational and Structural Biotechnology Journal. 2021;**19**:261-278

[9] Liao Y, Yin G, Wang X, Zhong P, Fan X, Huang C. Identification of candidate genes associated with the pathogenesis of small cell lung cancer via integrated bioinformatics analysis. Oncology Letters. 2019;**18**(4):3723-3733

[10] Tu H, Wu M, Huang W, Wang L. Screening of potential biomarkers and their predictive value in early stage non-small cell lung cancer: A bioinformatics analysis. Translational Lung Cancer Research. 2019;**8**(6): 797-807

[11] Zhang J, Li D, Zhang Y, Ding Z, Zheng Y, Chen S, et al. Integrative analysis of mRNA and miRNA expression profiles reveals seven potential diagnostic biomarkers for non-small cell lung cancer. Oncology Reports. 2020;**43**(1):99-112

[12] Liu X, Wu J, Zhang D, Bing Z, Tian J, Ni M, et al. Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis. Frontiers in Genetics. 2018;**9**. DOI: 10.3389/fgene.2018.00265

[13] Wang W, He Y, Zhao Q, Zhao X, Li Z. Identification of potential key genes in gastric cancer using bioinformatics analysis. Biomedical Reports. 2020;**12**(4): 178-192

[14] Rong L, Huang W, Tian S, Chi X, Zhao P, Liu F. COL1A2 is a novel biomarker to improve clinical prediction in human gastric cancer: Integrating bioinformatics and meta-analysis. Pathology Oncology Research. 2018; **24**(1):129-134

[15] Miao Y, Sui J, Xu S-Y, Liang G-Y, Pu Y-P, Yin L-H. Comprehensive analysis of a novel four-lncRNA signature as a prognostic biomarker for human gastric cancer. Oncotarget. 2017;**8**(43): 75007-75024

[16] Zhang X, Zhang W, Jiang Y, Liu K, Ran L, Song F. Identification of functional lncRNAs in gastric cancer by integrative analysis of GEO and TCGA data. Journal of Cellular Biochemistry. 2019;**120**(10):17898-17911

[17] Ji Y, Yin Y, Zhang W. Integrated bioinformatic analysis identifies networks and promising biomarkers for hepatitis B virus-related hepatocellular carcinoma. International Journal of Genomics. 2020;**2020**:2061024

[18] Shen B, Li K, Zhang Y. Identification of modules and novel prognostic biomarkers in liver cancer through integrated bioinformatics analysis. FEBS Open Bio. 2020;**10**(11):2388-2403

[19] Li L, Lei Q, Zhang S, Kong L, Qin B. Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis. Oncology Reports. 2017;**38**(5):2607-2618

[20] Wang CCN, Li CY, Cai J-H, Sheu PC-Y, Tsai JJP, Wu M-Y, et al. Identification of prognostic candidate genes in breast cancer by integrated bioinformatic analysis. Journal of Clinical Medicine. 2019;**8**(8):1160

[21] Wu M, Li Q, Wang H. Identification of novel biomarkers associated with the prognosis and potential pathogenesis of breast cancer via integrated bioinformatics analysis. Technology in Cancer Research & Treatment. 2021;**20**:1-16. DOI: 10.1177/1533033821992081

[22] Cai Y, Mei J, Xiao Z, Xu B, Jiang X, Zhang Y, et al. Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. Hereditas. 2019;**156**(1). DOI: 10.1186/s41065-019-0096-6

[23] Wu J, Lv Q, Huang H, Zhu M, Meng D. Screening and identification of key biomarkers in inflammatory breast cancer through integrated bioinformatic analyses. Genetic Testing and Molecular Biomarkers. 2020;**24**(8):484-491

[24] Chen L, Lu D, Sun K, Xu Y, Hu P, Li X, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. Gene. 2019;**692**:119-125

[25] Dai M, Chen X, Mo S, Li J, Huang Z, Huang S, et al. Meta-signature LncRNAs serve as novel biomarkers for colorectal cancer: Integrated bioinformatics analysis, experimental validation and diagnostic evaluation. Scientific Reports. 2017;**7**(1). DOI: 10.1038/srep46572

[26] Sun Z, Liu C, Cheng SY. Identification of four novel prognosis biomarkers and potential therapeutic drugs for human colorectal cancer by bioinformatics analysis. Journal of Biomedical Research. 2021;**35**(1):21

[27] Falzone L, Scola L, Zanghì A, Biondi A, Di Cataldo A, Libra M, et al. Integrated analysis of colorectal cancer microRNA datasets: Identification of microRNAs associated with tumor development. Aging (Albany NY). 2018;**10**(5):1000-1014

[28] van Zandwijk N. New methods for early diagnosis of lung cancer. Lung Cancer. 2002;**38**(1):S9-S11

[29] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: a Cancer Journal for Clinicians. 2015;**65**(2):87-108

[30] Ferro A, Peleteiro B, Malvezzi M, Bosetti C, Bertuccio P, Levi F, et al. Worldwide trends in gastric cancer mortality (1980-2011), with predictions to 2015, and incidence by subtype. European Journal of Cancer. 2014;**50**(7): 1330-1344

[31] Li R, Zhuang C, Jiang S, Du N, Zhao W, Tu L, et al. ITGBL1 predicts a poor prognosis and correlates EMT phenotype in gastric cancer. Journal of Cancer. 2017;**8**(18):3764-3773

[32] Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. Lancet. 2016;**388**(10060): 2654-2664

[33] Chau I. Checkpoint inhibition: An ATTRACTION in advanced gastric cancer? Lancet. 2017;**390**(10111): 2418-2419

[34] Kang M-H, Choi H, Oshima M, Cheong J-H, Kim S, Lee JH, et al. Author Correction: Estrogen-related receptor gamma functions as a tumor suppressor in gastric cancer. Nature Communications. 2018;**9**(1):3599

[35] Wang Z, Chen G, Wang Q, Lu W, Xu M. Identification and validation of a prognostic 9-genes expression signature for gastric cancer. Oncotarget. 2017; **8**(43):73826-73836

[36] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adeno-carcinoma. Nature. 2014;**513**(7517): 202-209

[37] Sun M, Song H, Wang S, Zhang C, Zheng L, Chen F, et al. Integrated analysis identifies microRNA-195 as a suppressor of Hippo-YAP pathway in colorectal cancer. Journal of Hematology & Oncology. 2017;**10**(1). DOI: 10.1186/s13045-017-0445-8

[38] Sun C, Yuan Q, Wu D, Meng X, Wang B. Identification of core genes and outcome in gastric cancer using bioinformatics analysis. Oncotarget. 2017;**8**(41):70271-70280

[39] Song E, Song W, Ren M, Xing L, Ni W, Li Y, et al. Identification of potential crucial genes associated with carcinogenesis of clear cell renal cell carcinoma. Journal of Cellular Biochemistry. 2018;**119**(7):5163-5174

[40] Wu J, Zhu P, Lu T, Du Y, Wang Y, He L, et al. The long non-coding RNA LncHDAC2 drives the self-renewal of liver cancer stem cells via activation of Hedgehog signaling. Journal of Hepatology. 2019;**70**(5):918-929

[41] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA: A Cancer Journal for Clinicians. 2018;**68**(1):7-30

[42] Bai Y, Long J, Liu Z, Lin J, Huang H, Wang D, et al. Comprehensive analysis of a ceRNA network reveals potential prognostic cytoplasmic lncRNAs involved in HCC progression. Journal of Cellular Physiology. 2019;**234**(10):18837-18848

[43] Bhikoo R, Srinivasa S, Yu T-C, Moss D, Hill AG. Systematic review of breast cancer biology in developing countries (part 2): Asian subcontinent and South East Asia. Cancers (Basel). 2011;**3**(2):2382-2401

[44] Waks AG, Winer EP. Breast cancer treatment: A review. JAMA. 2019; **321**(3):288-300 19323

[45] Tang J, Kong D, Cui Q, Wang K, Zhang D, Gong Y, et al. Prognostic genes of breast cancer identified by gene co-expression network analysis. Frontiers in Oncology. 2018;**8**:374

[46] Yuan L, Chen L, Qian K, Qian G, Wu CL, Wang X, et al. Coexpression

network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). Genomics Data. 2017;**14**:132-140

[47] Li J-J, Zhou D, Qiu W, Shi Y, Yang J-J, Chen S, et al. Application of weighted gene co-expression network analysis for data from paired design. Scientific Reports. 2018;**8**(1):1-8. DOI: 10.1038/ s41598-017-18705-z

[48] Kolligs FT. Diagnostics and epidemiology of colorectal cancer. Visceral Medicine. 2016;**32**(3):158-164

[49] Coppedè F. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. World Journal of Gastroenterology. 2014;**20**(4):943

[50] Zhang F, Deng CK, Wang M, Deng B, Barber R, Huang G. Identification of novel alternative splicing biomarkers for breast cancer with LC/MS/MS and RNA-Seq. BMC Bioinformatics. 2020;**21**(9):1-7

[51] Yamada A, Yu P, Lin W, Okugawa Y, Boland CR, Goel A. A RNA-Sequencing approach for the identification of novel long non-coding RNA biomarkers in colorectal cancer. Scientific Reports. 2018;**8**(1):1

[52] Yang J, Wang F, Zhong S, Chen B. Identification of hub genes with prognostic values in multiple myeloma by bioinformatics analysis. Hematology. 2021;**26**(1):453-459

[53] Pollak MN, Foulkes WD. Challenges to cancer control by screening. Nature Reviews Cancer. 2003;**3**(4):297-303

[54] Neugut AI, Jacobson JS, Rella VA. Prevalence and incidence of colorectal adenomas and cancer in asymptomatic persons. Gastrointestinal Endoscopy Clinics of North America. 1997;**7**(3): 387-399

[55] Diamandis EP. Towards identification of true cancer biomarkers. BMC Medicine. 2014;**12**(1):1-4

[56] Kulasingam V, Pavlou MP, Diamandis EP. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. Nature Reviews Cancer. 2010;**10**(5): 371-378