Research Article

Computational Analysis of Specific MicroRNA Biomarkers for Noninvasive Early Cancer Detection

Tianci Song,[1] Yanchun Liang,[1,2] Zhongbo Cao,[1] Wei Du,[1] and Ying Li[1]

[1]*College of Computer Science and Technology, Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun 130012, China*
[2]*Zhuhai Laboratory of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Jilin University, Zhuhai 519041, China*

Correspondence should be addressed to Wei Du; dwtyfx@gmail.com and Ying Li; liying@jlu.edu.cn

Cancer is a complex disease residing in various tissues of human body, accompanied with many abnormalities and mutations in genomes, transcriptome, and epigenome. Early detection plays a crucial role in extending survival time of all major cancer types. Recent advances in microarray and sequencing techniques have given more support to identifying effective biomarkers for early detection of cancer. MicroRNAs (miRNAs) are more and more frequently used as candidates for biomarkers in cancer related studies due to their regulation of target gene expression. In this paper, the comparative analysis is used to discover miRNA expression patterns in cancer versus normal samples on early stage of eight prevalent cancer types. Our work focuses on the specific miRNAs biomarkers identification and function analysis. Several identified miRNA biomarkers in this paper are matched well with those reported in existing researches, and most of them could serve as potential candidate indicators for clinical early diagnosis applications.

1. Introduction

Cancer is a highly complex disease that contains many abnormalities and mutations in the genomes, transcriptome, and epigenome. These abnormalities play important roles in the cancer cell growth [1]. Cancer is one of the leading causes of death all over the world. According to the world cancer report of 2014, there are about 14 million new cases and more than 8 million deaths related to cancer in 2012. The number will continue to rise if there are no effective prediction and treatment of cancer. It is expected that the number of new cases will rise to 22 million over the next two decades [2].

Early detection of cancer is undoubtedly important. During the development of cancer, some genes or proteins, associated with genetic mutations, transcription, or epigenetic alterations, could be detected from the tissue of cancer or inflammation when being compared to the normal tissue. These genes or proteins could give a quantitative measurement to the severity of a cancer. There have been many genes or proteins as biomarkers used in the clinical detection of cancer including alpha-fetoprotein (AFP) for liver cancer [3], BRCA1 and BRCA2 for breast and ovarian cancer [4], prostate specific antigen (PSA) for prostate cancer [5], and epidermal growth factor receptor (EGFR) for non-small-cell lung carcinoma [6].

In recent years, more and more researches on microRNA (miRNA) biomarkers have been published. miRNAs are small noncoding RNA molecules that contain 21–24 nucleotides. They play important roles in the posttranscriptional regulation of target gene expression [7]. The miRNA biomarker identification has been extensively studied in recent years. High-throughput microarray and sequencing techniques are widely used for transcriptome analysis. We can acquire lots of transcriptome information for various kinds of cancers on gene expression level from public databases, such as Gene Expression Omnibus (GEO) [8], Stanford Microarray Database (SMD) [9], Oncomine [10], and the Cancer Genome Atlas (TCGA) [5]. Dysregulation of miRNA expression is important for cancer development through various

mechanisms including deletions, amplifications, or epigenetic silencing [11]. The circulating miRNAs are suggested to be effective indicators of disease. They are important for clinical applications such as disease diagnostics, monitoring therapeutic effect, and predicting recurrence in cancer patients [12]. Circulating miRNAs are widely used as biomarkers for various human cancers, such as prostate cancer and breast cancer [13–16]. Some gene or miRNA biomarkers are identified using statistical methods to the high-throughput omics data. MiRNAs are widely used as biomarkers for human cancers. Resnick et al. acquired 21 differentially expressed miRNAs through comparing microarray data between 28 patients with ovarian cancer and 15 normal samples. Five differentially expressed miRNAs with overexpression and 3 miRNAs with underexpression in patients were evaluated through real-time PCR [17]. Xie et al. did the research on aberrant miRNAs used as biomarkers for the diagnosis of non-small-cell lung cancer (NSCLC). Their result indicated that mir-21 was detected with high expression in sputum specimen of patient [18]. Du Rieu et al. did the study on whether abnormal miRNA production for noninvasive precursor pancreatic intraepithelial neoplasia (PanIN) can be used as a potential early biomarker of pancreatic ductal adenocarcinoma (PDAC). They indicated that mir-21 was aberrantly expressed in early development of PanIN and it was worthy for further study as a biomarker for early detection of PDAC [19]. Habbe et al. did the research on the aberrant expressed miRNAs in intraductal papillary mucinous neoplasms (IPMNs). They showed that aberrant miRNA expression was an early event for pancreatic cancer, and miR-155 was worthy of further study as a biomarker for IPMNs in clinical samples [20].

Although several miRNA biomarkers have been reported in the above researches, only one or very few miRNAs are identified in each experiment which is involved in a single cancer type. The accuracy and specificity of some miRNAs are not so promising. In this paper, we analyze the miRNA expression patterns comparatively in cancer versus normal samples on early stage of eight prevalent cancer types. The datasets used in our paper are all RNA-Seq data from TCGA database to make a comprehensive identification of miRNA biomarkers for various cancers. We focus on identifying the specific miRNAs biomarkers including four aspects: (a) detecting differentially expressed miRNAs for each cancer type; (b) detecting specific differentially expressed miRNAs; (c) detecting specific miRNAs biomarkers; and (d) analyzing function and pathway of these miRNAs. Several identified miRNA biomarkers have been reported in existing researches and most of them could serve as potential candidate biomarkers for clinical early diagnosis.

2. Materials and Methods

2.1. miRNA Expression Datasets. The original miRNA expression data based on miRNA-Seq for eight prevalent cancer types are downloaded from the TCGA database [21]. The eight cancer types include prostate, thyroid, breast, head and neck, kidney, stomach, lung, and liver cancer. For each cancer type, we select the cancer samples which have corresponding

TABLE 1: The detailed information of cancer miRNA expression datasets.

Cancer type	Abbreviation	Number of normal samples	Number of cancer samples
Breast	BRCA	87	157
Colorectal	CRAD[1]	22	96
Kidney	KIDC[2]	105	350
Liver	LIHC	50	192
Lung	LUNG[3]	91	459
Prostate	PRAD	52	149
Stomach	STAD	45	62
Thyroid	THCA	59	325
All		511	1790

[1]CRAD contains COAD (colon adenocarcinoma) and READ (rectum adenocarcinoma).
[2]KIDC contains KIRC (kidney renal clear cell carcinoma) and KIRP (kidney renal papillary cell carcinoma).
[3]LUNG contains LUSC (lung squamous cell carcinoma) and LUAD (lung adenocarcinoma).

samples from their normal tissues as the paired samples to identify cancer related miRNA biomarkers of different cancer types. Biomarkers can ensure that the prediction results could be well generalized to clinical research and utility by using these paired samples. Detailed information of the datasets used in this paper is listed in Table 1. In this paper, we only use the samples of cancer from pathologic stage I to detect miRNA biomarkers for cancers of early stage. The pathologic stage is collected from clinical patient information of TCGA database. The value of "reads per million miRNA mapped" is used as expression value of each miRNA.

2.2. Identification of Differentially Expressed miRNAs. For each cancer type, firstly, Wilcoxon signed-rank test is used to identify differentially expressed miRNAs between cancer samples and normal samples. Then, FDR controlling method is used to eliminate false discovery rate of the result by Wilcoxon signed-rank test. After the above processes, an improved fold change method is applied to identify the differential expressed miRNAs between cancer samples and normal samples.

In this paper, we use quantiles to calculate the fold change of each miRNA. A quantile is a cut point which divides a set of observational data into equal sized groups. So, the q-quantiles indicate the value that partitions a set of finite data into q groups of equal sizes. The number of q-quantiles of value is $q - 1$. For a variable X, the kth q-quantile ($0 < k < q$) Q_k can be estimated using the following formula:

$$Q_k = x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor) \left(x_{\lfloor h \rfloor + 1} - x_{\lfloor h \rfloor} \right), \tag{1}$$

where $h = (N - 1)(k/q) + 1$ and N is the sample size. The set of q-quantiles is defined as follows:

$$SQ \subseteq \left\{ Q_1, \ldots, Q_k, \ldots, Q_{q-1} \right\}. \tag{2}$$

Let $E \in \mathfrak{R}^{m \times s}$ be the transcript measurements in a miRNA expression matrix with m miRNAs and s samples. Here,

variable X is $E[i,:]$ $(0 < i \leq m)$ and the set of q-quantiles of miRNA i can be calculated using formulae (1) and (2). For miRNA i on normal samples and cancer samples, we can obtain two sets of q-quantiles for normal and cancer samples as NQ_i and CQ_i, respectively. The fold change value of the miRNA i of kth q-quantile can be calculated using the following formula:

$$FC_{ik} = \begin{cases} \dfrac{CQ_{ik}}{NQ_{ik}} - 1 & (CQ_{ik} \geq NQ_{ik}) \\[2mm] 1 - \dfrac{NQ_{ik}}{CQ_{ik}} & (CQ_{ik} < NQ_{ik}), \end{cases} \tag{3}$$

where CQ_{ik} and NQ_{ik} are the expression values of miRNA i of kth q-quantile in cancer and normal samples, respectively. And then, the original fold change OFC_i of miRNA i, which is calculated by the average of FC_{ik} across all the samples, is defined as follows:

$$OFC_i = \frac{1}{q-1} \sum_{j=1}^{q-1} FC_{ik}, \tag{4}$$

where q is the selected number of quantiles.

In this paper, the impact of standard deviation is also introduced to the original fold change. The improved fold change denoted by IFC_i of miRNA i is shown as follows:

$$IFC_i = \frac{OFC_i}{2\sqrt{\sigma(CQ_i)\,\sigma(NQ_i)}}, \tag{5}$$

where σ is the standard deviation of sets of q-quantiles. In formula (5), not only the mean value of fold change but also the influence of variance across different samples is considered, which is a more effective and robust statistical analysis.

In this paper, q is set to 100 to calculate the quantiles. And then, we select differentially expressed miRNAs for each individual cancer type using the following rules: (1) the improved fold change (IFC) is less than -0.5 or greater than 0.5; and (2) q-value by FDR controlling for Wilcoxon signed-rank test is less than 0.05.

2.3. Identification of Specific Differentially Expressed miR-NAs.

The specific miRNA biomarkers for each cancer are selected, which can be used as discriminators for other cancers. We select specific differentially expressed miRNAs based on the differential expressed miRNA results from last section for each cancer. Let the improved fold change of miRNA i in cancer C_1 be $IFC_i^{C_1}$ and in other cancers $C_2, C_3, \ldots, C_k$ $(k = 1, 2, \ldots, 8)$ be $IFC_i^{C_2}, IFC_i^{C_3}, \ldots, IFC_i^{C_k}$. The miRNA i is regarded as a specific miRNA biomarker for cancer C_1, if and only if miRNA i complied with the following formula:

$$\left| IFC_i^{C_1} \right| > t_2,$$
$$\sum_{j=2}^{k} \left(\left| IFC_i^{C_1} - IFC_i^{C_j} \right| \right) \times \delta_i^{C_j} > t_1 \times t_2 \times (k-1), \tag{6}$$

where k is the number of cancer types and $k = 8$ in this paper. t_1 means the threshold of the ratio of different cancer types (the number of cancer types being considered); here we set t_1 to be 0.75. t_2 means the threshold of the improved fold change (IFC value); here we set t_2 to be 0.5. $\delta_i^{C_j}$ is a factor to filter the miRNAs with low expression value and is defined as follows:

$$\delta_i^{C_j} = 1, \quad \left(\overline{X}_i^{C_1} \geq \overline{X}_i^{C_j} \right),$$
$$\delta_i^{C_j} = 0, \quad \left(\overline{X}_i^{C_1} < \overline{X}_i^{C_j} \right), \tag{7}$$

where $X_i^{C_j}$ is a vector which means the expression value of miRNA i on cancer C_j and $\overline{X}_i^{C_j}$ is the average of $X_i^{C_j}$.

For example, if miRNA i is upexpressed in cancer C_1 (the first inequality in (6) is satisfied, which is $|IFC_i^{C_1}| > t_2$) and downexpressed or changed very little in other cancer types (the second inequality in (6) is satisfied, which is $\sum_{j=2}^{k}(|IFC_i^{C_1} - IFC_i^{C_j}|) \times \delta_i^{C_j} > t_1 \times t_2 \times (k-1)$), then this miRNA is a specific differentially expressed miRNA for cancer C_1. It is a good discriminator for distinguishing cancer types between C_1 and other C_j $(j = 2, 3, \ldots, 8)$.

2.4. Identification of Specific miRNA Biomarkers.

After obtaining the specific differentially expressed miRNAs for each single cancer type, we further select the circulating and upregulated miRNAs as biomarkers. The information of extracellular circulating miRNAs is downloaded from miRandola database [22]. Based on the source of extracellular miRNAs, the miRNAs in miRandola database are divided into four categories: Ago2, exosome, HDL, and circulating. In this paper, we only select the circulating miRNAs which are source of plasma and serum in miRandola database as candidate miRNA biomarkers.

For improving the sensitivity and specificity, we identify the combined miRNA biomarkers based on the specific single miRNA biomarkers. The rules of selecting combined biomarkers are considered here in order to identify k-miRNA discriminators. These k-miRNA discriminators are used as combined biomarkers for multiple cancer types, specific biomarkers for cancer types with similar survival rates (high survival rate, medium survival rate, and low survival rate), and a specific cancer type.

A computational process of finding the k-miRNA $(k = 1, 2, 3, 4, 5)$ combination biomarkers is proposed in this paper to give a best distinction among the different cancer groups using a linear classifier [23]. Linear Discriminant Analysis (LDA) [23] is employed to evaluate the performance of k-miRNA combination biomarkers for multiple cancer types, cancers of similar survival rates, and a single cancer type. The overall accuracy is defined as the fraction of the total number of true positives and true negatives and the number of all the samples in the following:

$$OA = \frac{(TP + TN)}{N}, \tag{8}$$

where TP is the number of true positives, TN is the number of true negatives, and N is the total number of samples.

TABLE 2: The differentially expressed miRNA in eight cancer types.

Cancer types	Upregulation	Downregulation	All	Ratio of up/down
BRCA	32	47	79	0.681
CRAD	82	154	236	0.533
KIDC	33	59	92	0.559
LIHC	33	67	100	0.493
LUNG	46	18	64	2.556
PRAD	71	3	74	23.667
STAD	77	14	91	5.500
THCA	19	48	67	0.396

The performance of all the k-miRNA combinations is evaluated using Leave-One-Out Cross Validation (LOOCV) method [24]. In each LOOCV step, one single sample is randomly chosen as the validation data, and the remaining samples are chosen as the training data to build the classifier using LDA. The OA value is calculated as the accuracy in this LOOCV step. This process is iterated until all of samples are selected as the validation data. Due to the computational complexity, the number of k combinations is set to be 1, 2, 3, 4, and 5 in this paper.

2.5. Function and Pathway Analysis. The functional annotation and pathway analysis are conducted using miEAA, a miRNA Enrichment Analysis and Annotation tool [25]. miEAA is a web-based system, which offers miRNA set enrichment analysis similar to Gene Set Enrichment Analysis (GSEA) [26]. The tool also provides rich functionality in terms of miRNA categories and contains over 14,000 miRNA sets, including pathways, diseases, organs, and target genes. In this paper, we perform the functional annotation and pathway analysis on specific upexpressed miRNAs for each cancer type among multiple cancer types, respectively.

3. Results and Discussion

The differentially expressed miRNAs in each cancer type are identified in order to study the specific roles of miRNAs involved in different cancer types. Due to the various alterations accumulated in the development of the oncogenesis, different cancers may have their specific miRNAs having differential expression. These miRNAs may be involved in some biological processes in the formation and progression of cancers. Similar to genes, several specific miRNAs are identified and used as targets of the prevention and diagnosis of cancers.

3.1. Differentially Expressed miRNAs. In this part, we use the strategy in the mentioned methods to identify the differentially expressed miRNAs in eight cancer types. We detect the upregulated and downregulated miRNAs, respectively, most of which may be involved in some important biology processes or pathways. The differentially expressed miRNAs of eight cancer types are summarized in Table 2. There are 79 differentially expressed miRNAs in breast cancer, among which 32 miRNAs are upregulated and 47 miRNAs are down-regulated. There are 236 differentially expressed miRNAs in

colorectal cancer, among which 82 miRNAs are upregulated and 154 miRNAs are downregulated. In the lung cancer results, 64 miRNAs are found to be differentially expressed, among which 46 miRNAs are upregulated and 18 miRNAs are downregulated. Interestingly, the number of differentially expressed miRNAs of different cancer types varies a lot. One reasonable explanation is that these miRNAs may reflect different regulatory mechanisms in different cancers and different cancers may accumulate different set of miRNAs.

Another interesting observation is that the number of upregulated and downregulated miRNAs in each cancer differs too much; that is, the ratio of upregulated and downregulated miRNAs differs a lot. In Table 2, we can see that there are more upregulated miRNAs than downregulated miRNAs in prostate and stomach cancers. Differently, the downregulated miRNAs in thyroid and liver cancers account for a major portion among the differently expressed miRNAs. It may possibly indicate the unique and similar characteristics of different cancer types.

3.2. Specific Differentially Expressed miRNAs. A detailed statistical analysis on specific differentially expressed miRNAs for single cancer type is conducted. The results are summarized in Table 3. There are 51 specific differentially expressed miRNAs in breast cancer, among which 21 miRNAs are upregulated and 30 miRNAs are downregulated. There are 180 specific differentially expressed miRNAs in colorectal cancer, among which 62 miRNAs are upregulated and 118 miRNAs are downregulated. In lung cancer, 46 miRNAs are found as specific differentially expressed miRNAs, among which 33 miRNAs are upregulated and 13 miRNAs are downregulated. Similarly, while the majority of differentially expressed circulating miRNAs in prostate and stomach cancers are upregulated, in kidney and liver cancers, the majority of such miRNAs are downregulated. The details of the specific differentially expressed miRNAs in eight cancers are illustrated in Table 3.

3.3. Specific miRNAs Biomarkers. The identification of specific miRNA biomarkers for each cancer is also performed in this part. For each cancer type, we use the strategy mentioned in Materials and Methods to find the miRNAs that have larger upregulation in each cancer type and have high expression value, and meanwhile the miRNAs have different changes or expression values in other cancer types. Such miRNAs

TABLE 3: The specific differentially expressed miRNAs in eight cancer types.

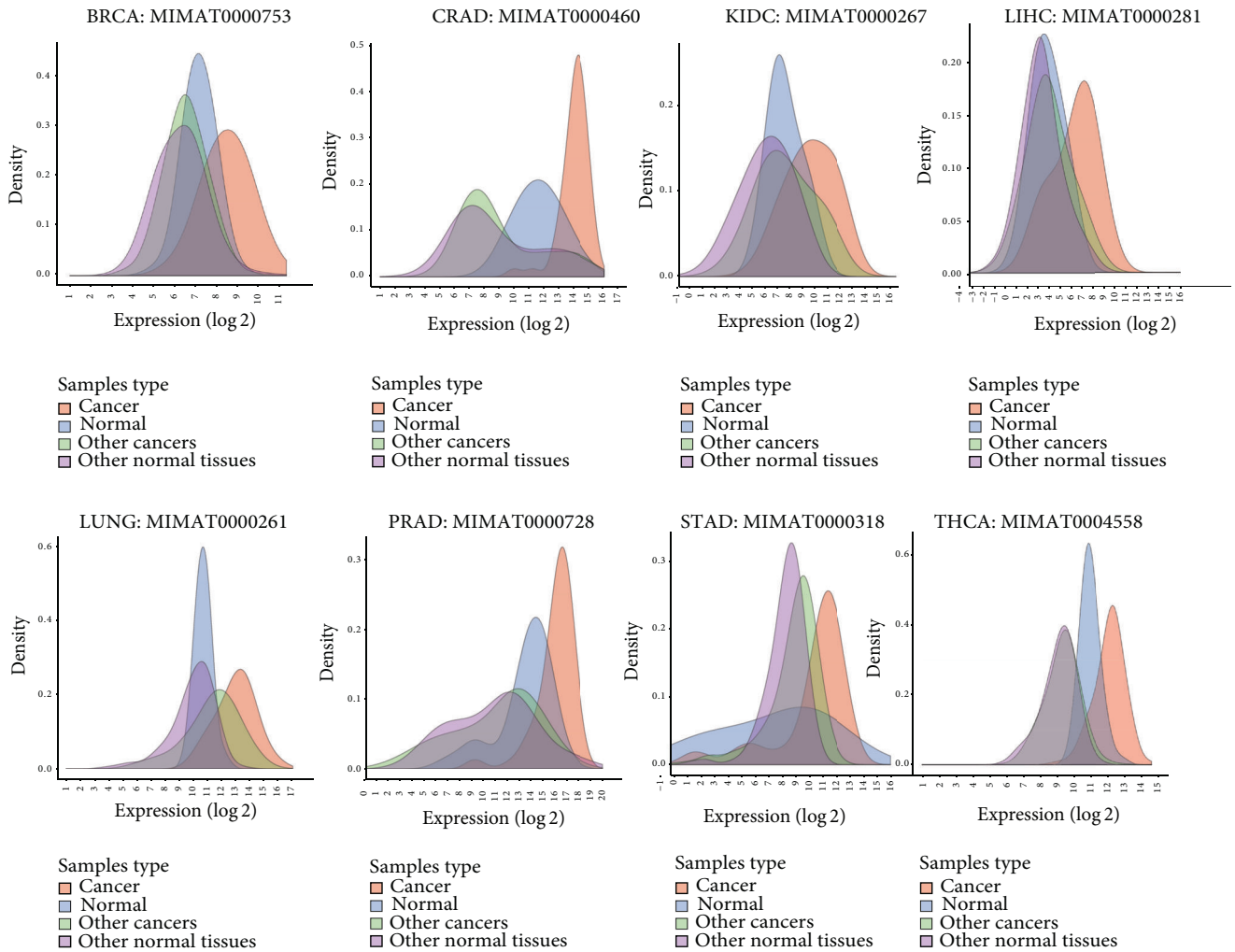| Cancer types	Upregulation	Downregulation	All	Ratio of up/down
BRCA	21	30	51	0.700
CRAD	62	118	180	0.525
KIDC	11	24	35	0.458
LIHC	19	41	60	0.463
LUNG	33	13	46	2.538
PRAD	19	2	21	9.500
STAD	51	8	59	6.375
THCA	15	25	40	0.600

FIGURE 1: Specific differentially expressed miRNAs in eight cancer types.

can make a better distinction between one cancer and other cancer types.

The detailed information is illustrated in Figure 1. MiRNA MIMAT0000753 is taken as an example (the first one in Figure 1). It shows the expression distribution of miRNA MIMAT0000753 in breast cancer, normal breast, other cancers, and other normal tissues. MIMAT0000753 gets a fold change value of nearly 0.92 in breast cancer, which indicates its upregulation obviously. The approximate mean expression values are 470 and 160 in breast cancer and normal samples, respectively. The approximate mean expression values are 125 and 107 in other cancer and normal samples, respectively. So MIMAT0000753 gets a better performance in prostate cancer compared to other cancers. It could be used to distinguish prostate cancer from other cancer types.

In some cases, using one single gene or miRNA is not enough to distinguish the specific cancer. The identification of the combination of some miRNAs is an important step

TABLE 4: Five-miRNA biomarker for eight cancers.

Cancer types	Markers	Accuracy 1	Accuracy 2	Mean
BRCA	MIMAT(0000076 + 0000259 + 0000434 + 0000753 + 0003218)	0.7702	0.7954	0.7828
CRAD	MIMAT(0000067 + 0000070 + 0000253 + 0000438 + 0004598)	0.8300	0.8657	0.8478
KIDC	MIMAT(0000068 + 0000076 + 0000255 + 0000267 + 0004494)	0.7473	0.8000	0.7736
LIHC	MIMAT(0000075 + 0000255 + 0000441 + 0002871 + 0002888)	0.7021	0.7726	0.7373
LUNG	MIMAT(0000100 + 0000261 + 0000432 + 0001080 + 0004494)	0.7311	0.7815	0.7563
PRAD	MIMAT(0000074 + 0000083 + 0000432 + 0000728 + 0001413)	0.8060	0.8366	0.8213
STAD	MIMAT(0000071 + 0000076 + 0000280 + 0000318 + 0004703)	0.7853	0.8029	0.7941
THCA	MIMAT(0000256 + 0000257 + 0000278 + 0000279 + 0004558)	0.6876	0.8805	0.7841

of differentially expressed miRNA analysis. In order to find the better combinations of miRNAs as biomarkers which can be used to distinguish the specific cancer, k-miRNA ($k = 1, 2, 3, 4, 5$) combinations of differentially expressed miRNAs are selected in this section. By identifying different miRNA combinations which have similar expression patterns in single cancer type but have different expression patterns in other cancers, we explore useful information of different and various mechanisms about carcinogenesis with single cancer type. In each k-miRNA combination, we calculate the classification capability in two aspects. One is the measurement of the k-miRNA in a single cancer dataset, which calculates the classification capability between cancer samples and the paired normal samples of the same cancer type. The other is the measurement of the k-miRNA in the datasets of multiple cancer types, which calculates the classification capability between one cancer type and the other cancer types using the fold change values of the cancer samples and the normal samples. Table 4 gives the results of distinguishing ability of the top five-miRNA specific biomarker combinations in eight cancers, respectively.

In Table 4, accuracy 1 is the results obtained by classifier between cancer samples and their corresponding paired normal samples; and accuracy 2 is the results obtained by classifier between fold change values from each cancer type and fold change values from other cancer types. These miRNAs could be used as biomarkers for corresponding cancer, and simultaneously they are found to be effective discriminators for distinguishing each cancer from other cancer types. For example, in breast cancer the top five-miRNA of MIMAT(0000076 + 0000259 + 0000434 + 0000753 + 0003218) gets the accuracy of 77.02% in single dataset evaluation and accuracy of 79.54% in multiple cancer datasets evaluation process, which are the best results among other combinations. The mean value can reach 78.28%, which indicates that the five-miRNA combination could be good biomarkers for breast cancer.

As noted, several miRNAs in each combination have been reported to be related to different cancers. For BRCA, MIMAT0000076 (hsa-miR-21) is reported to be overexpressed in human breast cancer and associated with clinical stage, metastasis, and prognosis [27]. MIMAT0000259 (hsa-miR-182) and MIMAT0003218 (hsa-miR-92b) are found to be overexpressed in human breast tumor [28, 29]. MIMAT0000434 (hsa-miR-142) is reported to relate to

the regulation of tumorigenicity of human breast cancer through the canonical WNT signaling pathway [30], and MIMAT0000753 (hsa-miR-342) regulates BRCA1 expression through modulation of ID4 in breast cancer [31].

For CRAD, MIMAT0000070 (hsa-miR-17) is found overexpressed in colon cancer and the expression is correlated with low survival rate in colorectal cancer patients [32], and MIMAT0000253 (hsa-miR-10a) is reported to be overexpressed in human colon cancer [33]. MIMAT0000067 (hsa-let-7g) is associated with tumorigenesis in colon cancer [34], and MIMAT0000438 (hsa-mir-152) is associated with cell proliferation, migration, and invasion in human cancer [35]. MIMAT0004598 (hsa-miR-141) is used as a novel biomarker for metastatic colon cancer [36].

For KIDC, MIMAT0000068 (hsa-miR-15a) is reported to be upregulated in urine of patients with renal cell carcinoma and undetectable in oncocytoma, other tumors, and urinary tract inflammation [37]. MIMAT0000076 (hsa-mir-21), MIMAT0004494 (hsa-miR-21-3p), and MIMAT0000267 (hsa-miR-210) are upregulated in renal cell carcinoma [37]. MIMAT0000255 (hsa-miR-34a) is reported as a specific oncogenic miRNA and shown to experience hypermethylation in kidney cancer cells [38].

For LIHC, MIMAT0000075 (hsa-miR-20a) is upregulated in liver cancer and correlated with hepatitis C virus-mediated liver disease progression [39]. MIMAT0000255 (hsa-miR-34a) and MIMAT 0002888 (hsa-miR-220) are reported as a potential therapeutic target in human cancer [40, 41]. MIMAT0000441 (hsa-miR-9) is upregulated in liver cancer tissues [42], and MIMAT0002871 (hsa-miR-500) is related to cancer survival [43].

For LUNG, MIMAT0000261 (hsa-miR-183) is reported to be overexpressed in lung cancer cell and suggested that high levels of CO_2 increase these miRNA levels and it is related to mitochondrial oxygen consumption, ATP production, and cell proliferation [44]. MIMAT0000100 (hsa-miR-29b) is found to mediate NF-κB signaling in KRAS-induced non-small-cell lung cancers (NSCLC) [45]. MIMAT0000432 (hsa-miR-141) and MIMAT0001080 (hsa-miR-196b) are upregulated in lung cancer [46]. MIMAT0004494 (hsa-miR-21-3p) is reported as biomarker to stratify for the subtype of lung cancer [47].

For PRAD, MIMAT0000074 (hsa-miR-19b) is reported to be related to prostate cancer and promoted prostate cell proliferation by targeting PTEN, PI3K/Akt pathway, and cyclin

TABLE 5: Enriched pathway information of eight cancer types.

Category	Term	Hint ratio
KEGG	Apoptosis (hsa04210)	1.000
Wiki Pathways	miRNAs involved in DDR (WP1545)	1.000
Wiki Pathways	Leptin signaling pathway (WP2034)	1.000
KEGG	Adipocytokine signaling pathway (hsa04920)	0.875
Wiki Pathways	Estrogen signaling pathway (WP712)	0.875
Wiki Pathways	Insulin signaling (WP481)	0.875
Wiki Pathways	DNA damage response (WP707)	0.875
Wiki Pathways	IL 4 signaling pathway (WP395)	0.875
KEGG	Wnt signaling pathway (hsa04310)	0.875
KEGG	Endocytosis (hsa04144)	0.875

D1 [48]. MIMAT0000083 (hsa-mir-26b) is overexpressed in prostate cancer [49]. MIMAT0000432 (hsa-miR-141) is reported as a promising biomarker for prostate cancer [50]. MIMAT0000728 (hsa-miR-375) plays a dual role in prostate carcinogenesis [51], and the expression of MIMAT0001413 (hsa-miR-20b) is associated with progression of prostate cancer cells [52].

For STAD, MIMAT0000280 (hsa-miR-223) has significantly higher expression and plays a role in regulating cellular apoptosis, proliferation, and invasion in gastric cancer [53]. MIMAT0000071 (hsa-miR-17-3p) and MIMAT0000318 (hsa-miR-200b) are reported to be overexpressed in gastric cancer patients [54, 55]. MIMAT0000076 (hsa-miR-21) is used as potential diagnostic and prognostic biomarkers for gastric cancer [56], and MIMAT0004703 (hsa-miR-335-3p) is applied as prognostic signature in gastric cancer [57].

For THCA, MIMAT0000256 (hsa-miR-181a) is overexpressed in thyroid carcinoma and has been proposed to play a role in the pathogenesis, development, progression, metastasis, prognosis, and therapeutic response to chemo- and radiotherapy [58]. MIMAT0000257 (hsa-miR-181b) is reported as a key regulator of the oncogenic process in cancer [59]. MIMAT0000278 (hsa-miR-221), MIMAT0000279 (hsa-miR-222), and MIMAT0004558 (hsa-miR-181a-2-3p) are upregulated in thyroid carcinoma [58].

3.4. Function and Pathway Analysis. In this section, the function annotation and pathway enrichment are performed using miEAA database. Firstly, we analyze the enriched pathways on specific upexpressed miRNAs for each cancer type among multiple cancer types, respectively. Then, we select the enriched pathways in most cancer types and sort them by the hint ratio. Table 5 gives top enriched pathway information with top ranks according to their hint ratio by miEAA [25]. The category means the source of pathway databases, the term means the enriched pathways, and the hint ratio means the enrichment score of miRNAs. From Table 5, we can see that the top enriched pathways are derived from Kyoto Encyclopedia of Genes and Genomes (KEGG) [60] and Wiki Pathways [61].

The first enriched pathway is apoptosis, which is described by its morphological characteristics and contributed to the high rate of cell loss in malignant tumors [62]. There are two enriched pathways, miRNAs involved in DDR and DNA damage response, which are related to DNA damage. There is an incontrovertible link between DNA damage and neoplastic phenotype in cancer [63]. Another interesting pathway is endocytosis, which entails selective packaging of cell-surface proteins and can be modified in cancer [64]. There are six signaling pathways in the enrichment results. Cellular signaling pathways are interconnected to form complex signaling networks and altered in cancer cells representing a major intellectual challenge [65].

4. Conclusions

Early detection of cancer is a very important and necessary way for cancer prevention. The biomarkers, as effective indicators to distinguish between cancer and normal samples or among different groups of cancer samples, are more and more used in cancer mechanism studies and clinical detection. The research of miRNA biomarkers is attracting more attention. But some of the biomarkers lack specificity and there are no effective common biomarkers for multiple cancer types. In this paper, we mainly focus on the identification of specific miRNAs and the common miRNA biomarkers. The miRNA expression data from RNA-Seq for eight cancer types are obtained from TCGA database, and the circulating miRNA information is collected from miRandola database.

For each cancer type, we apply Wilcoxon signed-rank test, which eliminates false discovery rate by using FDR controlling method, and improved fold change (FC) to identify miRNAs which have differential expressions between cancer samples and their corresponding normal samples. Then, the specific miRNA biomarkers for each cancer are further selected to act as discriminators for other cancers. After obtaining the specific differentially expressed miRNAs for each single cancer type, we further select the circulating and upregulated miRNAs as biomarkers. Several identified miRNA biomarkers in this paper are matched well with those reported in existing literatures and most of them could be taken as potential candidate indicators for clinical early diagnosis applications.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] M. Hejmadi, *Introduction to Cancer Biology*, Ventus Publishing ApS, 2010.

[2] B. W. Stewart and C. P. Wild, *World Cancer Report*, IARC, 2014.

[3] O. Arrieta, B. Cacho, D. Morales-Espinosa, A. Ruelas-Villavicencio, D. Flores-Estrada, and N. Hernández-Pedro, "The progressive elevation of alpha fetoprotein for the diagnosis of hepatocellular carcinoma in patients with liver cirrhosis," *BMC Cancer*, vol. 7, article no. 28, 2007.

[4] N. Petrucelli, M. B. Daly, G. L. Feldman et al., "BRCA1 and BRCA2 hereditary breast and ovarian cancer," in *GeneReviews®*, R. A. Pagon, M. P. Adam, H. H. Ardinger et al., Eds., University of Washington, Seattle, Wash, USA, 1993.

[5] M. C. Benson, Ihn Seong Whang, A. Pantuck et al., "Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer," *Journal of Urology*, vol. 147, no. 3, pp. 815–816, 1992.

[6] M. Verma and U. Manne, "Genetic and epigenetic biomarkers in cancer diagnosis and identifying high risk populations," *Critical Reviews in Oncology/Hematology*, vol. 60, no. 1, pp. 9–18, 2006.

[7] Y.-S. Huang, Y. Dai, X.-F. Yu et al., "Microarray analysis of microRNA expression in hepatocellular carcinoma and nontumorous tissues without viral hepatitis," *Journal of Gastroenterology and Hepatology*, vol. 23, no. 1, pp. 87–94, 2008.

[8] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.

[9] G. Sherlock, T. Hernandez-Boussard, A. Kasarskis et al., "The stanford microarray database," *Nucleic Acids Research*, vol. 29, no. 1, pp. 152–155, 2001.

[10] D. R. Rhodes, J. Yu, K. Shanker et al., "ONCOMINE: a cancer microarray database and integrated data-mining platform," *Neoplasia*, vol. 6, no. 1, pp. 1–6, 2004.

[11] N. Kosaka, H. Iguchi, and T. Ochiya, "Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis," *Cancer Science*, vol. 101, no. 10, pp. 2087–2092, 2010.

[12] R. Ma, T. Jiang, and X. Kang, "Circulating microRNAs in cancer: origin, function and application," *Journal of Experimental & Clinical Cancer Research*, vol. 31, article 38, 2012.

[13] N. Sapre and L. A. Selth, "Circulating MicroRNAs as biomarkers of prostate cancer: the state of play," *Prostate Cancer*, vol. 2013, 10 pages, 2013.

[14] M. G. Schrauder, R. Strick, R. Schulz-Wendtland et al., "Circulating micro-RNAs as potential blood-based markers for early stage breast cancer detection," *PLoS ONE*, vol. 7, no. 1, Article ID e29770, 2012.

[15] Y.-H. Taguchi and Y. Murakami, "Principal component analysis based feature extraction approach to identify circulating microRNA biomarkers," *PLOS ONE*, vol. 8, no. 6, Article ID e66714, 2013.

[16] K. Zen and C.-Y. Zhang, "Circulating MicroRNAs: a novel class of biomarkers to diagnose and monitor human cancers," *Medicinal Research Reviews*, vol. 32, no. 2, pp. 326–348, 2012.

[17] K. E. Resnick, H. Alder, J. P. Hagan, D. L. Richardson, C. M. Croce, and D. E. Cohn, "The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform," *Gynecologic Oncology*, vol. 112, no. 1, pp. 55–59, 2009.

[18] Y. Xie, N. W. Todd, Z. Liu et al., "Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer," *Lung Cancer*, vol. 67, no. 2, pp. 170–176, 2010.

[19] M. C. Du Rieu, J. Torrisani, J. Selves et al., "MicroRNA-21 is induced early in pancreatic ductal adenocarcinoma precursor lesions," *Clinical Chemistry*, vol. 56, no. 4, pp. 603–612, 2010.

[20] N. Habbe, J.-B. M. Koorstra, J. T. Mendell et al., "MicroRNA miR-155 is a biomarker of early pancreatic neoplasia," *Cancer Biology & Therapy*, vol. 8, no. 4, pp. 340–346, 2009.

[21] The Cancer Genome Atlas, http://cancergenome.nih.gov/.

[22] F. Russo, S. Di Bella, G. Nigita et al., "miRandola: extracellular circulating MicroRNAs database," *PLoS ONE*, vol. 7, no. 10, Article ID e47786, 2012.

[23] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, NY, USA, 1992.

[24] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 2*, pp. 1137–1143, Morgan Kaufmann Publishers, Montreal, Canada, August 1995.

[25] C. Backes, Q. T. Khaleeq, E. Meese, and A. Keller, "miEAA: microRNA enrichment analysis and annotation," *Nucleic Acids Research*, vol. 44, no. 1, pp. W110–W116, 2016.

[26] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[27] L.-X. Yan, X.-F. Huang, Q. Shao et al., "MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis," *RNA*, vol. 14, no. 11, pp. 2348–2360, 2008.

[28] Q. Wei, R. Lei, and G. Hu, "Roles of miR-182 in sensory organ development and cancer," *Thoracic Cancer*, vol. 6, no. 1, pp. 2–9, 2015.

[29] F. Peng, Y. Zhang, R. Wang et al., "Identification of differentially expressed miRNAs in individual breast cancer patient and application in personalized medicine," *Oncogenesis*, vol. 5, article e194, 2016.

[30] T. Isobe, S. Hisamori, D. J. Hogan et al., "miR-142 regulates the tumorigenicity of human breast cancer stem cells through the canonical WNT signaling pathway," *eLife*, vol. 3, Article ID e01977, 2014.

[31] E. Crippa, L. Lusa, L. De Cecco et al., "miR-342 regulates BRCA1 expression through modulation of ID4 in breast cancer," *PLoS ONE*, vol. 9, no. 1, Article ID e87039, 2014.

[32] E. Mogilyansky and I. Rigoutsos, "The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and

disease," *Cell Death and Differentiation*, vol. 20, no. 12, pp. 1603–1614, 2013.

[33] H. Mazeh, I. Mizrahi, N. Ilyayev et al., "The diagnostic and prognostic role of microrna in colorectal cancer—a comprehensive review," *Journal of Cancer*, vol. 4, no. 3, pp. 281–295, 2013.

[34] G. Nakajima, K. Hayashi, Y. Xi et al., "Non-coding microRNAs hsa-let-7g and hsa-miR-181b are associated with chemoresponse to S-1 in colon cancer," *Cancer Genomics and Proteomics*, vol. 3, no. 5, pp. 317–324, 2006.

[35] X. Liu, J. Li, F. Qin, and S. Dai, "MiR-152 as a tumor suppressor microRNA: target recognition and regulation in cancer," *Oncology Letters*, vol. 11, no. 6, pp. 3911–3916, 2016.

[36] A. Mohammadi, B. Mansoori, and B. Baradaran, "The role of microRNAs in colorectal cancer," *Biomedicine & Pharmacotherapy*, vol. 84, pp. 705–713, 2016.

[37] M. Li, Y. Wang, Y. Song et al., "MicroRNAs in renal cell carcinoma: a systematic review of clinical implications (Review)," *Oncology Reports*, vol. 33, no. 4, pp. 1571–1578, 2015.

[38] Q. Wei, Q.-S. Mi, and Z. Dong, "The regulation and function of micrornas in kidney diseases," *IUBMB Life*, vol. 65, no. 7, pp. 602–615, 2013.

[39] T. K. Motawi, O. G. Shaker, S. A. El-Maraghy, and M. A. Senousy, "Serum MicroRNAs as potential biomarkers for early diagnosis of hepatitis C virus-related hepatocellular carcinoma in egyptian patients," *PLOS ONE*, vol. 10, no. 9, Article ID e0137706, 2015.

[40] X. J. Li, R. J. Ren, and J. H. Tang, "MicroRNA-34a: a potential therapeutic target in human cancer," *Cell Death and Disease*, vol. 5, no. 7, 2014.

[41] H.-C. He, Z.-D. Han, Q.-S. Dai et al., "Global analysis of the differentially expressed miRNAs of prostate cancer in Chinese patients," *BMC Genomics*, vol. 14, article 757, 2013.

[42] L. Cai and X. Cai, "Up-regulation of miR-9 expression predicate advanced clinicopathological features and poor prognosis in patients with hepatocellular carcinoma," *Diagnostic Pathology*, vol. 9, no. 1, article 1000, 2014.

[43] M. L. Slattery, J. S. Herrick, L. E. Mullany et al., "An evaluation and replication of miRNAs with disease stage and colorectal cancer-specific mortality," *International Journal of Cancer*, vol. 137, no. 2, pp. 428–438, 2015.

[44] S. Dambal, M. Shah, B. Mihelich, and L. Nonn, "The microRNA-183 cluster: the family that plays together stays together," *Nucleic Acids Research*, vol. 43, no. 15, pp. 7173–7188, 2015.

[45] S. Langsch, U. Baumgartner, S. Haemmig et al., "miR-29b mediates NF-κB signaling in KRAS-induced non–small cell lung cancers," *Cancer Research*, vol. 76, no. 14, pp. 4160–4169, 2016.

[46] P. Leidinger, A. Keller, and E. Meese, "MicroRNAs—important molecules in lung cancer research," *Frontiers in Genetics*, vol. 2, article 104, 2012.

[47] V. Del Vescovo, M. Grasso, M. Barbareschi, and M. A. Denti, "MicroRNAs as lung cancer biomarkers," *World Journal of Clinical Oncology*, vol. 5, no. 4, pp. 604–620, 2014.

[48] H. Song, Y. Liu, J. Pan, and S. T. Zhao, "Expression profile analysis reveals putative prostate cancer-related microRNAs," *Genetics and Molecular Research*, vol. 12, no. 4, pp. 4934–4943, 2013.

[49] K. Erdmann, K. Kaulke, C. Thomae et al., "Elevated expression of prostate cancer-associated genes is linked to down-regulation of microRNAs," *BMC Cancer*, vol. 14, no. 1, article no. 82, 2014.

[50] A. Sita-Lumsden, D. A. Dart, J. Waxman, and C. L. Bevan, "Circulating microRNAs as potential new biomarkers for prostate cancer," *British Journal of Cancer*, vol. 108, no. 10, pp. 1925–1930, 2013.

[51] P. Costa-Pinheiro, J. Ramalho-Carvalho, F. Q. Vieira et al., "MicroRNA-375 plays a dual role in prostate carcinogenesis," *Clinical Epigenetics*, vol. 7, article no. 42, 2015.

[52] R. Ottman, C. Nguyen, R. Lorch, and R. Chakrabarti, "MicroRNA expressions associated with progression of prostate cancer cells to antiandrogen therapy resistance," *Molecular Cancer*, vol. 13, no. 1, article 1, 2014.

[53] J. Li, Y. Guo, X. Liang et al., "MicroRNA-223 functions as an oncogene in human gastric cancer by targeting FBXW7/hCdc4," *Journal of Cancer Research and Clinical Oncology*, vol. 138, no. 5, pp. 763–774, 2012.

[54] D. Park, S. C. Lee, J. W. Park, S. Y. Cho, and H. K. Kim, "Overexpression of miR-17 in gastric cancer is correlated with proliferation-associated oncogene amplification," *Pathology International*, vol. 64, no. 7, pp. 309–314, 2014.

[55] L. Chang, F. Guo, B. Huo, Y. Lv, Y. Wang, and W. Liu, "Expression and clinical significance of the microRNA-200 family in gastric cancer," *Oncology Letters*, vol. 9, no. 5, pp. 2317–2324, 2015.

[56] D. Wang, Z. Fan, F. Liu, and J. Zuo, "Hsa-miR-21 and Hsa-miR-29 in tissue as potential diagnostic and prognostic biomarkers for gastric cancer," *Cellular Physiology and Biochemistry*, vol. 37, no. 4, pp. 1454–1462, 2015.

[57] Z. Yan, Y. Xiong, W. Xu et al., "Identification of hsa-miR-335 as a prognostic signature in gastric cancer," *PLoS ONE*, vol. 7, no. 7, Article ID e40037, 2012.

[58] H. He, K. Jazdzewski, W. Li et al., "The role of microRNA genes in papillary thyroid carcinoma," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 52, pp. 19075–19080, 2005.

[59] J. Liu, W. Shi, C. Wu, J. Ju, and J. Jiang, "miR-181b as a key regulator of the oncogenic process and its clinical implications in cancer (Review)," *Biomedical Reports*, vol. 2, no. 1, pp. 7–11, 2014.

[60] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research*, vol. 45, no. 1, pp. D353–D361, 2017.

[61] M. Kutmon, A. Riutta, N. Nunes et al., "WikiPathways: capturing the full diversity of pathway knowledge," *Nucleic Acids Research*, vol. 44, no. 1, pp. D488–D494, 2016.

[62] F. K. Guerra-Vladusic, G. Scott, V. Weaver et al., "Constitutive expression of Heregulin induces apoptosis in an erbB-2 overexpressing breast cancer cell line SKBr-3," *International Journal of Oncology*, vol. 15, no. 5, pp. 883–892, 1999.

[63] C. J. Lord and A. Ashworth, "The DNA damage response and cancer therapy," *Nature*, vol. 481, no. 7381, pp. 287–294, 2012.

[64] I. Mellman and Y. Yarden, "Endocytosis and cancer," *Cold Spring Harbor perspectives in biology*, vol. 5, no. 12, p. a016949, 2013.

[65] G. S. Martin, "Cell signaling and cancer," *Cancer Cell*, vol. 4, no. 3, pp. 167–174, 2003.