Hindawi
Journal of Oncology
Volume 2020, Article ID 8458190, 9 pages
https://doi.org/10.1155/2020/8458190

Research Article

Identification of Monotonically Differentially Expressed Genes across Pathologic Stages for Cancers

Suyan Tian,[1] **Chi Wang**,[2,3] **Mingbo Tang**,[4] **Jialin Li**,[4] **and Wei Liu**[4]

[1]*Division of Clinical Research, The First Hospital of Jilin University, 1 Xinmin Street, Changchun 130021, Jilin, China*
[2]*Department of Internal Medicine, College of Medicine, University of Kentucky, 800 Rose Street, Lexington 40536, KY, USA*
[3]*Markey Cancer Center, University of Kentucky, 800 Rose Street, Lexington 40536, KY, USA*
[4]*Department of Thoracic Surgery, The First Hospital of Jilin University, 1 Xinmin Street, Changchun 130021, Jilin, China*

Correspondence should be addressed to Suyan Tian; windytian@hotmail.com and Wei Liu; drweiliu@outlook.com

Received 29 April 2020; Revised 17 October 2020; Accepted 28 October 2020; Published 12 November 2020

Academic Editor: Manu Kanjoormana Aryan

Given the fact that cancer is a multistage progression process resulting from genetic sequence mutations, the genes whose expression values increase or decrease monotonically across pathologic stages are potentially involved in tumor progression. This may provide insightful clues about how human cancers advance, thereby facilitating more personalized treatments. By replacing the expression values of genes with their GeneRanks, we propose a procedure capable of identifying monotonically differentially expressed genes (MEGs) as the disease advances. Using three real-world gene expression data that cover three distinct cancer types—colon, esophageal, and lung cancers—the proposed procedure has demonstrated excellent performance in detecting the potential MEGs. To conclude, the proposed procedure can detect MEGs across pathologic stages of cancers very efficiently and is thus highly recommended.

1. Introduction

Since cancer is a multistage progression process that results from genetic sequence mutations, the patterns of gene expression values differ as tumors develop. "Monotonic" genes, whose expression levels increase or decrease monotonically as the disease advances, are highly likely to be involved in the tumor progression. Therefore, they may provide insightful clues about how these complex diseases initiate and advance and have potential to facilitate personalized treatments. Thus, the roles they play in cancers are of critical importance.

Feature selection is one of the fundamental tasks in the area of machine learning. Generally speaking, the primary objective of feature selection is to identify an optimal subset of genes associated with the phenotype(s) of interest. Therefore, identification of genes presenting some specific expression change patterns over pathologic stages is essentially a process of feature selection. So far, only a few feature selection algorithms that are capable of identifying monotonically differentially expressed genes (MEGs) across time points/stages have been proposed.

The MFSelector method proposed by Wang et al. [1] and the pipeline to analyze longitudinal big data proposed by Carey et al. [2] are two such methods. Briefly, the MFSelector method selects K-1 (where K is the number of stages/time points under consideration) discriminating lines to separate the stages or time points apart and calculates a statistic (the number of misclassified subjects according to the K-1 discriminating lines), and the p value determines whether the specific gene is monotonically differentially expressed. In the analysis pipeline by Carey et al. [2], the functional principal component analysis is used to fit a smooth curve for longitudinal expression values of each gene. Then, modified F-tests are carried out to screen the genes according to the corresponding p values of F-statistics. Clustering is conducted to group the statistically significant genes according to whether they are co-expressed. These resulting clusters are

called as gene response modules in which the monotonically increasing and the monotonically decreasing patterns over time are the primary patterns of concern.

The MFSelector method and the longitudinal data analysis pipeline are conventional feature selection methods and do not take pathway information into account. Many studies have demonstrated that more advanced feature selection methods in which pathway information is incorporated as a priori to guide the process of feature selection outperform those classic feature selection methods in terms of predictive capacity, model stability, and biological implication. Such advanced feature selection algorithms are referred to as pathway-based feature selection algorithms [3]. As we mentioned in our previous studies [3, 4], the weighting strategy is the simplest way to account for pathway information and, as long as the estimation of those weights is accurate enough, the strategy can have an excellent performance and in many cases outperforms the competitive methods.

In this study, we replaced the original gene expression values with the weighted expression values generated by the GeneRank method [5] and suggested a procedure to identify MEGs. The method was evaluated using three sets of real-world gene expression data and the results were compared with the conventional method using original gene expression values and the MFSelector method [1].

2. Materials and Methods

2.1. Experimental Data

2.1.1. Non-Small-Cell Lung Cancer (NSCLC). The raw data of the NSCLC studies we used are stored on the Gene Expression Omibus (GEO: https://www.ncbi.nlm.nih.gov/geo/) repository under accession numbers GSE37745 [6] and GSE50081 [7] and are publicly assessable. The chips of these two experiments were all profiled on the Affymetrix HG-U133 Plus 2.0 platform. All patients in these two cohorts were adjuvant treatment naïve with their survival time available. In our previous study [8], we gave a detailed description on this data set. Briefly, of 104 patients in the dataset, 17 were stage IA patients, 57 stage IB patients, 5 stage IIA patients, and 25 stage IIB patients.

2.1.2. Colon Cancer (CC). The accession number on the GEO repository for the colon cancer data considered in this study is GSE62932 [9]. The chips of this experiment were hybridized on the Affymetrix HG-U133 Plus 2.0 platform as well. The data include 4 normal controls, 12 stage I patients, 17 stage II patients, 20 stage III patients, and 15 stage IV patients, for a total of 68 subjects in this study.

2.1.3. Esophageal Cancer (EC). The RNA-Seq data of the Cancer Genome Atlas Data Portal Esophageal Carcinoma (ESCA) cohort were downloaded from Genomic Data Commons (https://gdc.cancer.gov/). Patients with no clinical information on their pathologic stage were excluded, leaving 145 patients to be considered in the downstream analysis. Among them were 17 stage I patients, 70 stage II patients, 50 stage III patients, and 8 stage IV patients.

2.2. Preprocessing Procedures. Raw data (.cel files) of the three microarray datasets were downloaded from the GEO repository. The expression values were obtained using the fRMA algorithm [10] and were normalized using quantile normalization. For the NSCLC data, after the summary expression values were obtained, the COMBAT algorithm [11] was used to eliminate or alleviate the potential batch effects existing among different experiments.

For the RNA-Seq data of esophageal cancer, FPKM was downloaded from Genomic Data Commons (https://gdc.ca-ncer.gov/). The gene expression values were obtained by adding ones to FPKM counts and then having them log 2 transformed.

2.3. Pathway Information. The interaction/connection information among genes was retrieved from the Human Protein Reference Database (HPRD) [12], and the adjacency matrix was made on the basis of these gene-to-gene interactions. There were 9,672 protein-coding genes annotated in the HPRD database, Release 9 (http://www.hprd.org/).

2.4. Statistical Methods

2.4.1. GeneRank. Briefly, the GeneRank r for gene i is solved by

$$\left(I_p - dW D^{-1}\right)r = (1-d)\exp_i. \quad (1)$$

In this equation, I_p is a $p\times p$ identity matrix. Here, p is the number of genes under consideration; W stands for the adjacency matrix of genes and records how they interplay with one another, if the value in its kj cell is 1 then gene k and gene j are connected, and the value is zero otherwise. D is a $p\times p$ matrix, with its diagonal elements recording the degrees of freedom for these p genes and off-diagonal elements are zeroes. The degree of freedom is the number of genes to which a specific gene k ($k = 1,2, \ldots, p$) is connected; expi stands for gene expression values for sample i ($i = 1,2, \ldots, n$), and d is a damping or tuning parameter, balancing off the influence of the expression values and the pathway topological information within the network on the final rankings. The rankings can be completely determined either by the expression values when d equals to 0 or by the network structure when d equals to 1. The value of d is set at 0.5 by default.

2.5. Monotonic Expression Pattern Identification. The procedure we propose consists of three steps. First, the GeneRank of each gene for each subject is generated. Second, upon those GeneRanks that may be regarded as the weighted expression values of genes, the Kruskal-Wallis tests are carried out. Genes with adjusted p values less than a predetermined threshold (here, a grid of values are considered, i.e., 0.05, 0.1, 0.15, and 0.2) are deemed to be statistically differentially expressed genes. Among those differentially

expressed genes, different expressed patterns such as a U-shaped relationship or a spike at a single stage are possible but not of interest. Thus, the following equations are further exploited to distinguish monotonic expression patterns from other patterns:

$$(\overline{\exp}_{i0} \leq)\overline{\exp}_{i1} \leq \overline{\exp}_{i2} \leq \overline{\exp}_{i3} \leq \overline{\exp}_{i4}, \quad (2)$$

$$(\overline{\exp}_{i0} \geq)\overline{\exp}_{i1} \geq \overline{\exp}_{i2} \geq \overline{\exp}_{i3} \geq \overline{\exp}_{i4}, \quad (3)$$

for the monotonically increasing (MI) expressed genes and the monotonically decreasing (MD) expressed genes, respectively. Here, $\overline{\exp}_{i0}$ stands for the mean expression value of gene i in the normal control group. Notably, it is put inside parentheses to emphasize that not all studies have included controls. Furthermore, $\overline{\exp}_{ik}$ stands for the mean expression value of gene i among the patients at pathologic stage k ($k = 1$, 2, 3, and 4). Specifically, for the colon cancer and esophageal cancer studies, it corresponds to stages I II, III, or IV, and for the NSCLC study, it corresponds to stages IA, IB, IIA, or IIB.

2.6. Kruskal-Wallis Test. Kruskal-Wallis tests are carried out to determine if any differences in expression values exist among different stages, and then the differentially expressed genes presenting monotonic expression patterns are selected by equations (2) and (3). So, the only difference between this procedure and the procedure we propose is that the conventional one uses original expression values, whereas the proposed one uses weighted expression values generated by the GeneRank method.

2.7. MFSelector. Another method capable of identifying MEGs is the MFSelector method [1] in which a new statistic, the DEtotal (total discriminating error) score is introduced, and the corresponding adjusted p value which corrected for the multiple comparisons problem of the DEtotal score is calculated using permutation tests. Using the monotonically increasing scenario to illustrate the MFSelector method is described briefly as follows.

For a monotonically increasing expressed gene, it is naturally expected that subjects in early stages have smaller expression values compared to the subjects in later stages. First, $n1$ ($n1$ is the number of patients at stage I) discriminating lines may be drawn at the expression value of each stage I patient. The stage I patients above this line and the patients at higher levels below this line are misclassified, and the number of misclassified patients is counted. The final discriminating line to separate stage I from the higher levels corresponds to the line with the least misclassified number. This step is repeated for $K-1$ times to discriminate the patients at the first k ($k = 1,2, \ldots, K-1$, where K is the total number of stages) stages from the remaining patients, resulting in $K-1$ discriminating lines. If a gene has $K-1$ distinct discriminating lines and the lines for a later stage are above the lines for an earlier level, the expression change pattern of this gene has a perfect monotonically increasing expression tendency. Then, the DEtotal score is the sum of misclassified numbers for the $K-1$ segmentations, and a p value/q value of the DEtotal score is calculated using permutation tests (the patient's labels are perturbed) to determine whether or not this specific gene's increasing expression is statistically significant.

3. Results

3.1. Identification of MEGs. Colorectal cancer (CC), also known as colon cancer, is the second most common cancer in females and the third in males [13]. The molecular mechanisms of colon cancer have not yet been fully elucidated [14]. Likewise, the underlying mechanisms for esophageal cancer have not been unraveled, but the incidence and mortality rates are lower compared to colon cancer. For both sexes combined, lung cancer is the most commonly diagnosed cancer and the leading cause of cancer death [13]. Even though much more research is done on lung cancer compared to colon and esophageal cancers, complete deciphering of its etiology and progression has not yet been achieved.

Since the colon and esophagus both belong to the gastrointestinal tract, they may share more similarities regarding gene expression compared to lung cancer; thus, the MEGs for colon and esophageal cancers are expected to have more overlap with each other than with lung cancer. On the other hand, the platforms of the colon cancer and esophageal cancer studies differ, while the platforms for the NSCLC study and colon cancer study are identical, even though the origin of NSCLC is the respiratory system rather than the gastrointestinal tract. Moreover, esophagus and lung are located inside the thoracic cavity and the colon is inside the abdominal cavity. With these similarities and dissimilarities, the three data sets may disclose many interesting patterns. Utilizing the GeneRank method [5] as a building block, we propose a procedure that enables identification of monotonically expressed genes (MEGs) in this study, with the objective of revealing underlying molecular mechanisms for these three cancers.

The numbers of selected monotonically expressed genes over stages for these three types of cancers using the proposed procedure are given in Table 1, with the significant levels set at 0.05, 0.1, 0.15, and 0.2, respectively. In addition, the number of MEGs by the conventional Kruskal-Wallis method using the unweighted expression values and the MFSelector method (described briefly in the Methods section) are given in Table 1. Compared to the proposed method, both the Kruskal-Wallis method and the MFSelector method are too conservative, especially when using the MFSelector method, as no genes were identified as MEGs for any of the three studies for any of the significant levels considered. Therefore, the proposed procedure is decidedly more statistically powerful in detecting potential MEGs.

Interestingly, we observed the following tendency—for both esophageal cancer and NSCLC, the number of monotonically increasing genes is larger than that of monotonically decreasing genes. On the other hand, the opposite case is true for colon cancer, which may imply that more potential tumor suppressor genes are off balance for

Table 1: Monotonically differentially expressed genes.

Study		0.05	0.1	0.15	0.2
Colon cancer ($n = 68$)	MI	1 (0/0)	37 (32/0)	78 (81/0)	114 (127/0)
4 stages (I, II, III, and IV) and controls	MD	31 (0/0)	171 (97/0)	245 (157/0)	278 (204/0)
Esophageal cancer ($n = 145$)	MI	0 (0/0)	119 (0/0)	304 (0/0)	456 (25/0)
4 stages (I, II, III, and IV)	MD	0 (0/0)	13 (0/0)	32 (0/0)	54 (3/0)
NSCLC ($n = 104$)	MI	0 (0/0)	102 (0/0)	266 (0/0)	342 (0/0)
4 stages (IA, IB, IIA, and IIB)	MD	0 (0/0)	3 (0/0)	7 (0/0)	20 (0/0)

MI: monotonically increasing expressed genes; MD: monotonically decreasing expressed genes; NSCLC: non-small-cell lung cancer. (x/xx): x is the number of MEGs identified by the conventional Kruskal-Wallis method and xx is the number of MEGs identified by the MFSelector method. For example, for the NSCLC application at the significance level of 0.1, the (0/0) entity after 102 means both the conventional method and the MFSelection method identified 0 MI genes.

colon cancer, whereas more potential oncogenes are off to boost tumor progression for both esophageal cancer and NSCLC. Further investigation is warranted.

With the cutoff for adjusted p value set at 0.1, a Venn diagram of identified MEGs for the three studies using the proposed procedure is shown in Figure 1. It is observed that one overlap, i.e., COMMD7 (COMM domain containing 7) existed in the MEGs for colon and esophageal cancer, whereas the other overlap, i.e., HAND2 (heart and neural crest derivatives Expressing 2) existed in the MEGs for colon cancer and NSCLC.

A comparison between the proposed procedure and the conventional method for the colon cancer study was also made. The results are presented in Figure 2. The Venn diagrams stratified by the expression direction show that the overlap of MEGs by these two methods is substantial. The resulting weighted expression values balance between gene expression values and their importance (i.e., the degree of connectivity) in the gene-to-gene interaction network, thus MEGs identified by the proposed procedure alone tend to be essential genes in the network. Using the conventional method, these genes would be left out due to their subtle expression levels. For each cancer type, three MEGs were randomly selected, and violin plots representing their expression distributions stratified by pathologic stage are shown in Figure 3. Basically, no too extreme values are detected in the expression levels of the nine genes.

The enriched gene ontology (GO) terms [15] and KEGG pathways [16] by the MEGs were explored using the String software, stratified by each study. For NSCLC, there are 63 enriched GO biological process terms, 8 GO molecular function terms, 26 GO cellular component terms, and 0 KEGG pathways, respectively. For esophageal and colon cancers, the numbers of enriched GO terms by identified MEGs are 94 and 275 biological process terms, 8 and 48 molecular function terms, 58 and 49 cellular component terms, and 4 and 12 KEGG pathways, respectively. The Venn diagrams of overlapping GO terms and KEGG pathways are shown in Figure 4. Overall, at the gene set/pathway level, the overlap rate is higher than it is at the individual gene level, as expected.

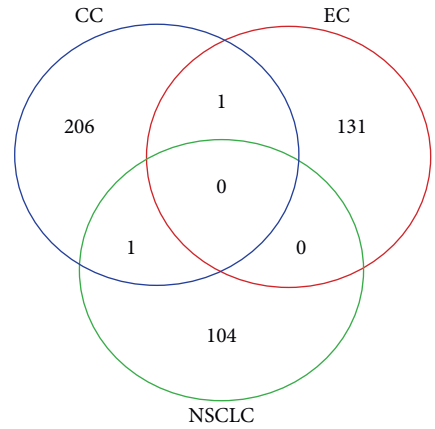

Figure 1: Venn diagram of the identified MEGs for colon cancer, esophageal cancer, and non-small-cell lung cancer studies. CC: colon cancer; EC: esophageal cancer; NSCLC: non-small-cell lung cancer. MEGs: monotonically expressed genes.

3.2. Biological Relevance

3.2.1. Overlapping MEGs. A recent study [17] claimed that COMMD7 overexpression positively correlated with histological differentiation and tumor node metastasis (TNM) stage of pancreatic ductal adenocarcinoma (PDAC), and PDAC patents with higher COMMD7 expression tended to have poorer overall survival rates. Also, COMMD7 has been reported to be upregulated in hepatocellular carcinoma (HCC) and promote HCC cell proliferation [18]. Even though in the literature we cannot find any studies suggesting COMMD7 is explicitly associated with esophageal or colon cancer, the proposed method identified it as a monotonic increasing gene for both EC and CC cohorts, consistent with the results of the two abovementioned studies and supporting the thought that COMMD7 is an oncogene. In contrast, another recent study [19] showed that HAND2 was hypermethylated and downregulated in colon cancer, while another study [20] demonstrated that HAND2 was overexpressed in the lung squamous cell carcinoma. In the present study, HAND2 was identified as a monotonically

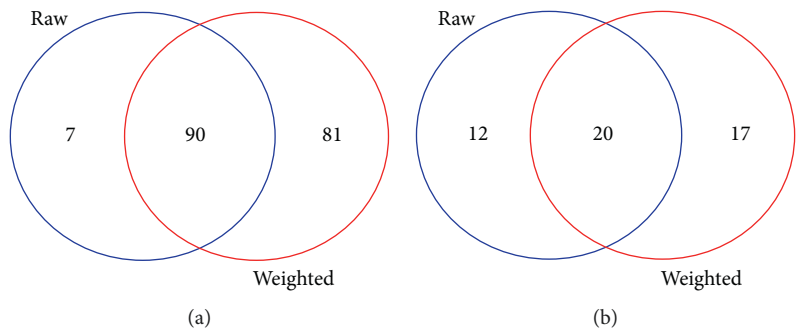

Figure 2: Comparison of the identified MEGs for colon cancer by the conventional method and the proposed procedure. (a) For the monotonically decreasing genes. (b). For the monotonically increasing genes. The significance level is set at 0.1. MI: monotonically increasing; MD: monotonically decreasing; raw: the MEGs identified by the conventional method upon the original expression profiles; weighted: the MEGs identified by the proposed method upon the weighted expression profiles; MEGs: monotonically expressed genes.

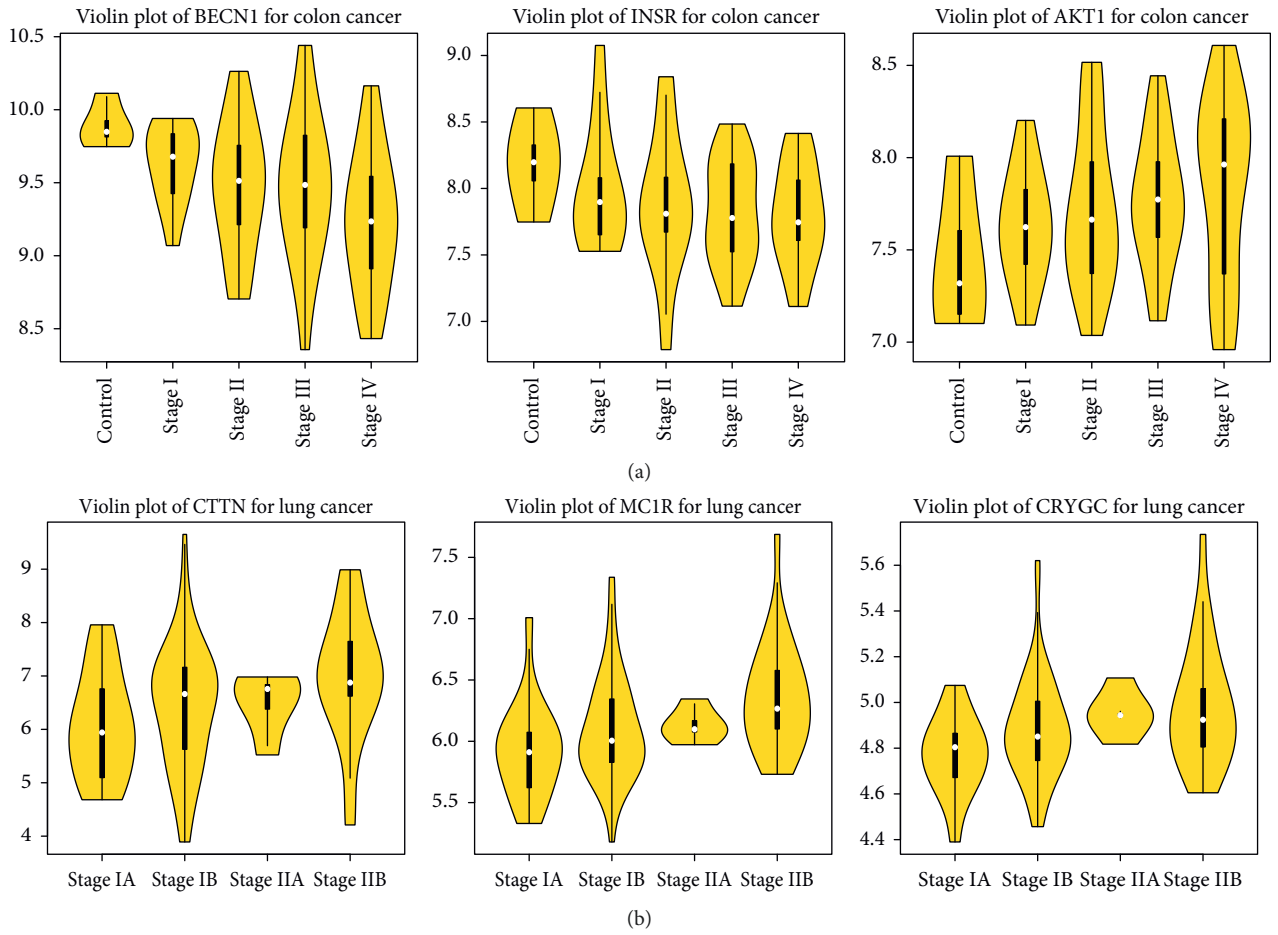

Figure 3: Continued.

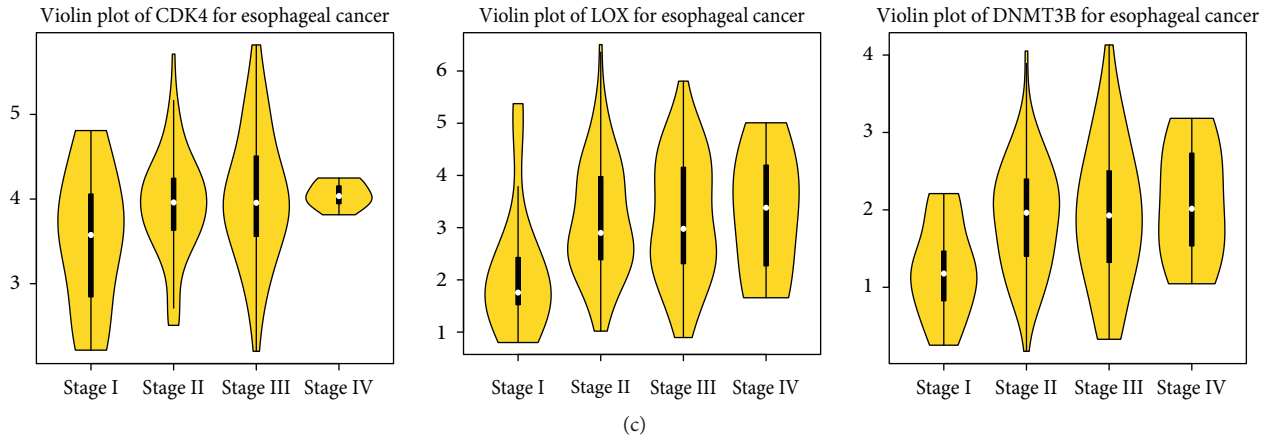

Figure 3: Violin plots of three randomly selected MEGs for each cancer type. (a) Colon cancer. (b) Esophageal cancer. (c) Non-small-cell lung cancer. CC: colon cancer; EC: esophageal cancer; NSCLC: non-small-cell lung cancer. MEGs: monotonically expressed genes.

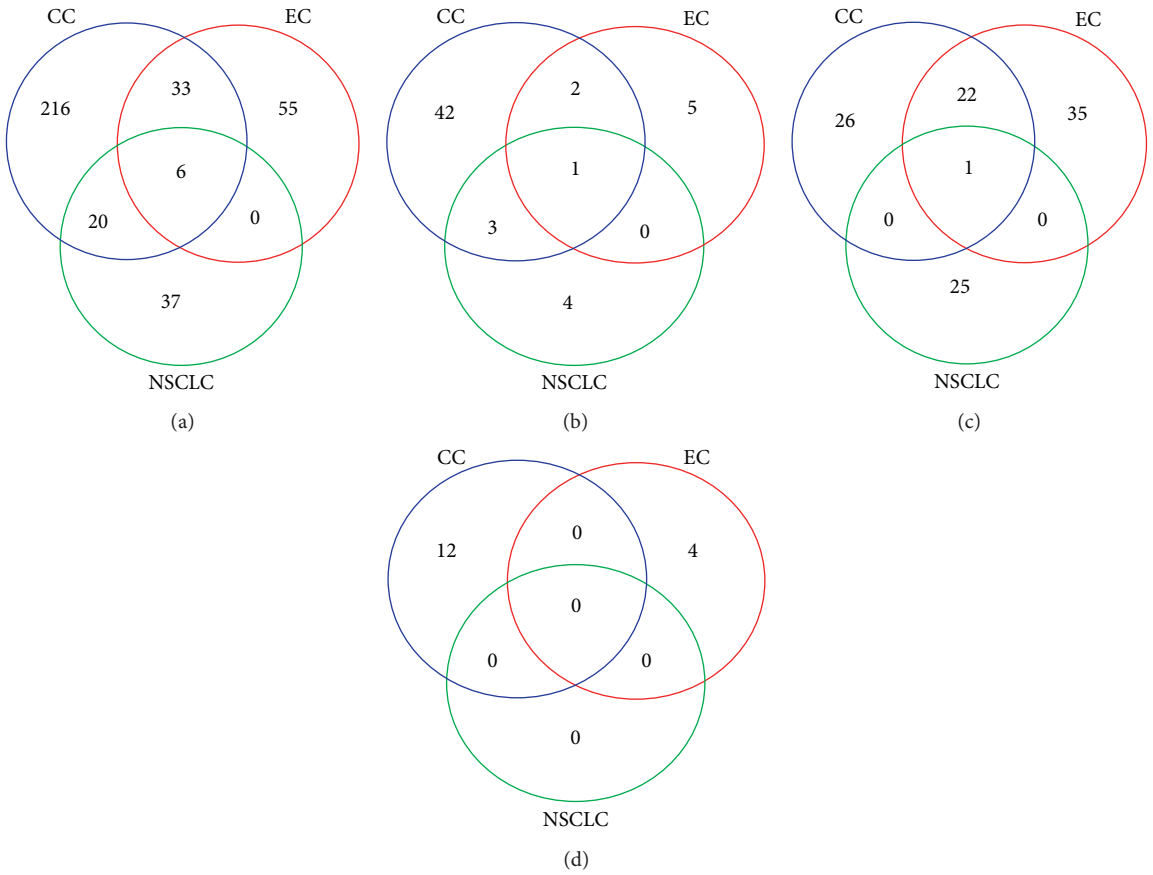

Figure 4: Venn diagrams of the enriched GO terms and KEGG pathways by the MEGs for colon cancer, esophageal cancer, and non-small-cell lung cancer studies. (a) GO biological process terms. (b) GO molecular function terms. (c) GO cellular component terms. (d) KEGG pathways. CC: colon cancer; EC: esophageal cancer; NSCLC: non-small-cell lung cancer, BP: biological process; MF: molecular function; CC: cellular component.

decreasing gene in colon cancer while a monotonically increasing gene in NSCLC, which is basically consistent with the results of the two previous studies.

3.2.2. Type-Specific MEGs. MEGs that are specific for one cancer type, meaning the genes were identified as the MEGs by only a single study, are referred to as type-specific MEGs.

According to the GeneCards database, all these genes are related to cancer either directly or indirectly. Some of them have been demonstrated to associate with these three cancer types under investigation by experimental means. For example, AKT1 (AKT serine/threonine kinase 1) has been demonstrated to play crucial roles in the development, progression, and drug resistance of colon cancer [21, 22]. In addition, Zhao et al. [23] showed that MiR-124 was significantly downregulated in NSCLC patients, and miR-124 negatively regulates AKT1. As far as esophageal cancer is concerned, the expression level of AKT1 has been reported to be significantly elevated in tumor tissue of patients with esophageal squamous cell carcinoma [24].

For esophageal cancer, the GeneCards database ranks CDK4 (cyclin-dependent kinase 4), DNMT3B (DNA methyltransferase 3 beta), and MAGEA4 (MAGE family member A4) as the top three relevant genes. Among the 24 MEGs that are directly related to esophageal cancer, a majority of them are associated with either NSCLC or colon cancer. For example, LOX (lysyl oxidase) has been shown to be overexpressed in lung cancer, and inhibition of LOX activity decreases the number of lung metastases [25].

For NSCLC, the GeneCards database indicates that 23 MEGs are directly related to lung cancer. ESR2 (estrogen receptor 2), CHKA (choline kinase alpha), and CRYGC (crystallin gamma C) are identified as the top three NSCLC-specific MEGs. ESR2 and CHKA are also associated with colon and esophageal cancers, while CRYGC is only directly related to colon cancer according to the GeneCards database. Even though type-specific genes were only identified as MEGs by a single study, many of them were correlated with the other two cancer types.

3.2.3. Oncogenes or Tumor Suppressor Genes? For the top MEGs with good biological relevance (i.e., the genes have a confidence score of >5 in the GeneCards database), whether the certain genes are oncogenes or tumor suppressor genes was investigated by searching the PubMed database and the TSGene 2.0 [26] database which records tumor suppressor genes for about 10 cancer types including colon cancer, lung adenocarcinoma, and lung squamous cell carcinoma.

For the top colon cancer MEGs, the consistent tumor suppressor genes included MAP2K4, MAPK10, RUNX3, WNK2 (the four genes were identified by the TSGene 2.0 database), BECN1 [27], FASN [28], NAT1 [29], and NR3C2 [30]. For monotonically increasing genes, the consistent ones include CCKBR, BMP4 [31, 32], and SLC29A1 [33] which were determined to be oncogenes by previous studies. In contrast, RB1 that was regarded as a tumor suppressor gene [34] in gastrointestinal stromal tumors and was identified as a monotonically increasing gene, while three oncogenes including NFE2L2 [35], ABL1 [36], and LASP1 [37] were identified as monotonically decreasing genes. Of note, AKT1 is indicated as a tumor suppressor gene by the TSGene 2.0 database [26], but many previous studies (e.g., [38]) report it as an oncogene, as does the present study.

Three lung cancer MEGs had a confidence score of >5: ESR2 and CRYGC (monotonically increasing) and CHKA (monotonically decreasing). A meta analysis [39] found no association between ESR2 expression level and the prognosis of NSCLC patients, and thus whether it is an oncogene or not remains controversial. For CHKA, previous studies present contradicting results, for example, [40] indicated its expression was lower, while [41] mentioned it was overexpressed in lung cancer. No literature about the expression status of CRYGC in lung cancer was found. In addition, there are some inconsistencies between our work, the literature, and the TSGene 2.0 database. Specifically, CDH4 [42], SFRP1 [43], and ERF [44] which are indicated to be tumor suppressor genes by the TSGene 2.0 database and have support from the literature as well; however, our method identified them as monotonically increasing genes. These genes may be false positives by our approach. Lastly, the roles of several genes play remain controversial. Namely, NFATC2 is indicated to be a tumor suppressor by the TSGene 2.0 database. However, Xiao's study [45] suggested high expression associated with poor tumor differentiation and poor survival. Similarly, a recent study [46] showed that the expression of AHNAK was upregulated in tumor samples, while the TSGene 2.0 database deems it as a tumor suppressor gene. The present study identified these two genes as monotonically increasing genes, being consistent with the previous studies.

Lastly, for the three monotonically increasing genes with good biological relevance, CDK4, DNMT3B, and MAGEA4 to esophageal cancer, previous studies [47, 48] suggested the last two genes as oncogenes for esophageal cancer while another study [49] suggested CDK4 was underexpressed in the tumor samples of esophageal cancer. The heterogeneity of study population, experimental techniques and personnel, and so on may explain the inconsistencies and contradictions to some extent. Further investigation on the roles that identified MEGs may play is highly desirable, especially for ones that are newly discovered by the proposed procedure.

4. Conclusions

After replacing the original expression values of genes with their GeneRanks [5], we defined a procedure capable of identifying genes with monotonically changed expression patterns across the pathologic stages of cancers. Using three real-world datasets, we show that the proposed method is superior to the conventional Kruskal-Wallis test and the MFSelector method [1]. Furthermore, the MEGs we identified are highly associated with the development and prognosis of cancer.

This procedure should be applicable to not only mRNA data but also many other data types such as lncRNA (long noncoding RNA) data. For the noncoding RNAs, there is no canonical knowledge base, such as STRING [50] and HPRD [12], to record how they interact. Furthermore, given the mechanism of how LncRNAs impact on a biological process by acting as a miRNA sponge, via the strategy of competing endogenous RNAs (ceRNAs) [51], the lncRNA-miRNA-mRNA interaction network may be more desirable. To address this shortage, statistical methods such as the WGCNA method [52] may be utilized to construct a data-

driven gene-to-gene interaction network, upon which the importance of specific lncRNAs and their expression patterns over pathologic stages can be inferred.

To conclude, the gain of efficiency in detecting MEGs using the proposed procedure is nontrivial; therefore, it is highly recommended.

Data Availability

Three microarray data (accession numbers: GSE37745, GSE50081, and GSE62932) were downloaded from the Gene Expression omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/), and the RNA-Seq data for the ESCA cohort were downloaded from the Cancer Genome Atlas data portal (https://tcga-data.nci.nih.gov/tcga/). They are all free to be downloaded.

Conflicts of Interest

The authors declare that there are no conflicts interests.

Authors' Contributions

ST and WL conceived and designed the study. ST, MT, and CW analyzed the data. MT, JL, CW, and ST interpreted the data analysis and results. ST, CW, WL, MT, and JL wrote the paper. All the authors reviewed and approved the final manuscript.

Acknowledgments

The Markey Cancer Center's Research Communications Office assisted with manuscript preparation. This study was supported by the Finance Department of Jilin Province (No. 2018SCZWSZX-018) and the Education Department of Jilin Province (No. JJKH20190032KJ).

References

[1] H.-W. Wang, H.-J. Sun, T.-Y. Chang et al., "Discovering monotonic stemness marker genes from time-series stem cell microarray data," *BMC Genomics*, vol. 16, no. 2, p. S2, 2015.

[2] M. Carey, J. C. Ramírez, S. Wu, and H. Wu, "A big data pipeline: identifying dynamic gene regulatory networks from time-course gene expression omnibusdata with applications to influenza infection," *Statistical Methods in Medical Research*, vol. 27, no. 7, pp. 1930–1955, 2018.

[3] S. Tian, H. H. Chang, and C. Wang, "Weighted-SAMGSR: combining significance analysis of microarray-gene set reduction algorithm with pathway topology-based weights to select relevant genes," *Biology Direct*, vol. 11, p. 50, 2016.

[4] A. Zhang and S. Tian, "Classification of early-stage non-small cell lung cancer by weighing gene expression profiles with connectivity information," *Biometrical Journal*, vol. 60, no. 3, pp. 537–546, 2018.

[5] J. L. Morrison, R. Breitling, D. J. Higham, and D. R. Gilbert, "GeneRank: using search engine technology for the analysis of microarray experiments," *BMC Bioinformatics*, vol. 6, no. 1, p. 233, 2005.

[6] S. Rousseaux, A. Debernardi, B. Jacquiau et al., "Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers," *Science Translational Medicine*, vol. 5, Article ID 186ra66, 2013.

[7] S. D. Der, J. Sykes, M. Pintilie et al., "Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients," *Journal of Thoracic Oncology*, vol. 9, no. 1, pp. 59–64, 2014.

[8] S. Tian, "Identification of monotonically differentially expressed genes for non-small cell lung cancer," *BMC Bioinformatics*, vol. 20, p. 177, 2019.

[9] X. Chen, N. G. Deane, K. B. Lewis et al., "Comparison of nanostring nCounter® data on FFPE colon cancer samples and affymetrix microarray data on matched frozen tissues," *PLoS One*, vol. 11, Article ID e0153784, 2016.

[10] M. N. McCall, B. M. Bolstad, and R. A. Irizarry, "Frozen robust multiarray analysis (fRMA)," *Biostatistics*, vol. 11, no. 2, pp. 242–253, 2010.

[11] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical Bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.

[12] T. S. Keshava Prasad, R. Goel, K. Kandasamy et al., "Human protein reference database-2009 update," *Nucleic Acids Research*, vol. 37, pp. D767–D772, 2009.

[13] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[14] G. Shi, Y. Wang, C. Zhang et al., "Identification of genes involved in the four stages of colorectal cancer: gene expression profiling," *Molecular and Cellular Probes*, vol. 37, pp. 39–47, 2018.

[15] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[16] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 27, no. 1, pp. 29–34, 1999.

[17] N. You, J. Li, Z. Gong et al., "COMMD7 functions as molecular target in pancreatic ductal adenocarcinoma," *Molecular Carcinogenesis*, vol. 56, no. 2, pp. 607–624, 2017.

[18] N. You, J. Li, X. Huang et al., "COMMD7 promotes hepatocellular carcinoma through regulating CXCL10," *Biomedicine & Pharmacotherapy*, vol. 88, pp. 653–657, 2017.

[19] Y. Yang, F.-H. Chu, W.-R. Xu et al., "Identification of regulatory role of DNA methylation in colon cancer gene expression via systematic bioinformatics analysis," *Medicine*, vol. 96, no. 47, p. e8487, 2017.

[20] P. Ranganathan, A. Agrawal, R. Bhushan et al., "Expression profiling of genes regulated by TGF-beta: differential regulation in normal and tumour cells," *BMC Genomics*, vol. 8, no. 1, p. 98, 2007.

[21] L. Yuan, K. Zhang, M.-M. Zhou et al., "Jiedu sangen decoction reverses epithelial-to-mesenchymal transition and inhibits invasion and metastasis of colon cancer via AKT/GSK-3β signaling pathway," *Journal of Cancer*, vol. 10, no. 25, pp. 6439–6456, 2019.

[22] G. B. Park, J. Y. Jeong, and D. Kim, "GLUT5 regulation by AKT1/3-miR-125b-5p downregulation induces migratory activity and drug resistance in TLR-modified colorectal cancer cells," *Carcinogenesis*, vol. 41, no. 10, pp. 1329–1340, 2020.

[23] X. Zhao, C. Lu, W. Chu et al., "MicroRNA-124 suppresses proliferation and glycolysis in non-small cell lung cancer cells by targeting AKT-GLUT1/HKII," *Tumor Biology*, vol. 39, Article ID 1010428317706215, 2017.

[24] K. Takahashi, M. Miyashita, H. Makino et al., "Expression of Akt and Mdm2 in human esophageal squamous cell carcinoma," *Experimental and Molecular Pathology*, vol. 87, no. 1, pp. 42–47, 2009.

[25] R. Gong, W. Lin, A. Gao et al., "Forkhead box C1 promotes metastasis and invasion of non-small cell lung cancer by binding directly to the lysyl oxidase promoter," *Cancer Science*, vol. 110, no. 12, pp. 3663–3676, 2019.

[26] M. Zhao, P. Kim, R. Mitra, J. Zhao, and Z. Zhao, "TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1023–D1031, 2016.

[27] F. Hu, G. Li, C. Huang et al., "The autophagy-independent role of BECN1 in colorectal cancer metastasis through regulating STAT3 signaling pathway activation," *Cell Death & Disease*, vol. 11, p. 304, 2020.

[28] A. Schcolnik-Cabrera, G. Dominguez-Gómez, A. Chávez-Blanco et al., "Erratum: a combination of inhibitors of glycolysis, glutaminolysis and de novo fatty acid synthesis decrease the expression of chemokines in human colon cancer cells," *Oncology Letters*, vol. 19, p. 2071, 2020.

[29] C. Shi, L. Y. Xie, Y. P. Tang et al., "Hypermethylation of N-acetyltransferase 1 is a prognostic biomarker in colon adenocarcinoma," *Frontiers in Genetics*, vol. 10, p. 1097, 2019.

[30] M. Yu, H. L. Yu, Q. H. Li, L. Zhang, and Y. X. Chen, "miR-4709 overexpression facilitates cancer proliferation and invasion via downregulating NR3C2 and is an unfavorable prognosis factor in colon adenocarcinoma," *Journal of Biochemical and Molecular Toxicology*, vol. 33, Article ID e22411, 2019.

[31] P. Hu, S. Zhang, S.-y. Lu et al., "An efficient scheme for purification of a novel recombinant immunotoxin, rCCK8PE38, for anti-tumour experiments," *Biomedical Chromatography*, vol. 32, no. 6, p. e4197, 2018.

[32] G. S. Karagiannis, H. Afaloniati, E. Karamanavi, T. Poutahidis, and K. Angelopoulou, "BMP pathway suppression is an early event in inflammation-driven colon neoplasmatogenesis of uPA-deficient mice," *Tumor Biology*, vol. 37, no. 2, pp. 2243–2255, 2016.

[33] A. V. Snezhkina, G. S. Krasnov, A. R. Zaretsky et al., "Differential expression of alternatively spliced transcripts related to energy metabolism in colorectal cancer," *BMC Genomics*, vol. 17, p. 1011, 2016.

[34] K. Ohshima, K. Fujiya, T. Nagashima et al., "Driver gene alterations and activated signaling pathways toward malignant progression of gastrointestinal stromal tumors," *Cancer Science*, vol. 110, no. 12, pp. 3821–3833, 2019.

[35] M. N. Uddin, M. Li, and X. Wang, "Identification of transcriptional signatures of colon tumor stroma by a meta-analysis," *Journal of Oncology*, vol. 2019, Article ID 8752862, 2019.

[36] Y. Liu, J. Cao, Y. N. Zhu et al., "C1222C deletion in exon 8 of ABL1 is involved in carcinogenesis and cell cycle control of colorectal cancer through IRS1/PI3K/akt pathway," *Frontiers in Oncology*, vol. 10, p. 1385, 2020.

[37] W. Wang, G. Ji, X. Xiao et al., "Epigenetically regulated miR-145 suppresses colon cancer invasion and metastasis by targeting LASP1," *Oncotarget*, vol. 7, no. 42, pp. 68674–68687, 2016.

[38] M. Rostami-Nejad, S. Rezaei Tavirani, V. Mansouri, S. Jahani-Sherafat, and H. Moravvej Farshi, "Gene expression profile analysis of colon cancer grade II into grade III transition by using system biology," *Gastroenterology and Hepatology from Bed to Bench*, vol. 12, pp. 60–66, 2019.

[39] B. M. Herman, S. Charlap, and W. H. Frishman, "Nitrates in congestive heart failure," *Medical Clinics of North America*, vol. 73, no. 2, pp. 361–371, 1989.

[40] O. Kowalczuk, T. Burzykowski, W. E. Niklinska, M. Kozlowski, L. Chyczewski, and J. Niklinski, "CXCL5 as a potential novel prognostic factor in early stage non-small cell lung cancer: results of a study of expression levels of 23 genes," *Tumor Biology*, vol. 35, no. 5, pp. 4619–4628, 2014.

[41] J. C. Lacal and J. M. Campos, "Preclinical characterization of RSM-932A, a novel anticancer drug targeting the human choline kinase alpha, an enzyme involved in increased lipid metabolism of cancer cells," *Molecular Cancer Therapeutics*, vol. 14, no. 1, pp. 31–39, 2015.

[42] Z. Li, D. Su, L. Ying, G. Yu, and W. Mao, "Study on expression of CDH4 in lung cancer," *World Journal of Surgical Oncology*, vol. 15, p. 26, 2017.

[43] S. H. Cho, I. Y. Kuo, P. F. Lu et al., "Rab37 mediates exocytosis of secreted frizzled-related protein 1 to inhibit Wnt signaling and thus suppress lung cancer stemness," *Cell Death & Disease*, vol. 9, p. 868, 2018.

[44] Y.-T. Chou, H.-H. Lin, Y.-C. Lien et al., "EGFR promotes lung tumorigenesis by activating miR-7 through a Ras/ERK/Myc pathway that targets the Ets2 transcriptional repressor ERF," *Cancer Research*, vol. 70, no. 21, pp. 8822–8831, 2010.

[45] Z. J. Xiao, J. Liu, S. Q. Wang et al., "NFATc2 enhances tumor-initiating phenotypes through the NFATc2/SOX2/ALDH axis in lung adenocarcinoma," *eLife*, vol. 6, 2017.

[46] S. Zhang, Y. Lu, L. Qi, H. Wang, Z. Wang, and Z. Cai, "AHNAK2 is associated with poor prognosis and cell migration in lung adenocarcinoma," *BioMed Research International*, vol. 2020, Article ID 8571932, 2020.

[47] J.-F. Su, F. Zhao, Z.-W. Gao et al., "piR-823 demonstrates tumor oncogenic activity in esophageal squamous cell carcinoma through DNA methylation induction via DNA methyltransferase 3B," *Pathology - Research and Practice*, vol. 216, no. 4, p. 152848, 2020.

[48] M. Ishihara, S. Kageyama, Y. Miyahara et al., "MAGE-A4, NY-ESO-1 and SAGE mRNA expression rates and co-expression relationships in solid tumours," *BMC Cancer*, vol. 20, p. 606, 2020.

[49] L. Cao, T. Hu, H. Lu, and D. Peng, "N-MYC downstream regulated gene 4 (NDRG4), a frequent downregulated gene through DNA hypermethylation, plays a tumor suppressive role in esophageal adenocarcinoma," *Cancers (Basel)*, vol. 12, 2020.

[50] A. Franceschini, D. Szklarczyk, S Frankild et al., "STRING v9.1: protein-protein interaction networks, with increased coverage and integration," *Nucleic Acids Research*, vol. 41, pp. D808–D815, 2013.

[51] L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi, "A ceRNA hypothesis: the rosetta stone of a hidden RNA language?" *Cell*, vol. 146, no. 3, pp. 353–358, 2011.

[52] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, p. 559, 2008.