IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 7, 2021, accepted January 19, 2021, date of publication January 27, 2021, date of current version February 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3054162

Toward Identifying Key Gene Group in the Occurrence and Development of Lung Adenocarcinoma

MING GU[1,2], TIAN-FANG ZHAO[1], (Student Member, IEEE), WEI LIU[3], XIAO-KUN WU[1], AND LIANG YANG[4,5,6]

[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China
[2]School of Basic Medical Science, Southern Medical University, Guangzhou 510515, China
[3]School of Foreign Languages, Central South University, Changsha 410083, China
[4]School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China
[5]State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences (CAS), Beijing 100864, China
[6]Hebei Province Key Laboratory of Big Data Calculation, Hebei University of Technology, Tianjin 300401, China

Corresponding authors: Tian-Fang Zhao (tianfang09@foxmail.com) and Liang Yang (yangliang@vip.qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61972442, in part by the Key Research and Development Project of Hebei Province of China under Grant 20350802D, in part by the Natural Science Foundation of Hebei Province of China under Grant F2020202040, and in part by the Natural Science Foundation of Tianjin of China under Grant 20JCYBJC00650.

ABSTRACT Lung adenocarcinoma (LUAD) has become the most common pathological type of lung cancer in recent years. However, the molecular mechanism of LUAD remains unclear. To reveal the laws in the occurrence and development of LUAD, this paper first collects 676 differentially expressed genes related to LUAD and the corresponding proteins encoded by the genes, resulting in a complex protein-protein interaction (PPI) network. The traditional analysis methods focus on the gene interaction relationships or the important genes separately, ignoring the fact that some important target genes may take effect jointly. In contrast, this paper adopts a new analysis method named the overlapping gene analysis method to screen out the closely interacted and important genes. Thereinto, the hub genes are first discovered according to the node importance index, and then the PPI network is divided into multiple communities. The overlapping genes are some genes belonging to both the hub genes and the genes in the same community, regarded as gene groups. Through experiments, 10 genes are identified as a gene group. Survival analyses to the gene group show that only PLK1, CCNB1, and CDK1 participate in the prognosis of LUAD, and relate to the pathological stage of LUAD. These demonstrate that the three genes are extremely important in the cell cycle, which is confirmed by the following enrichment analyses. Besides, some significant correlations are observed among the three genes, which provide a clue that the three genes may cooperate in the occurrence and development of LUAD. This finding provides possible clinical targets for the diagnosis and treatment of LUAD.

INDEX TERMS Complex protein network, hub gene analysis, community analysis, cell cycle, tumor and cancer.

I. INTRODUCTION

In recent years, the global incidence rate and mortality rate of lung cancer have been ranking the first among all cancers. Lung cancer, with its incidence accounting for 11.6% of the total cancer incidence, has the highest mortality rate in males [1], [2]. Lung adenocarcinoma (LUAD) is one type of non-small-cell carcinoma. Its incidence rate and mortality rate are higher than those of squamous cell carcinoma [3], [4]. It has been a crucial task to study the pathogenesis of LUDA and develop effective targeted drugs for curing it.

LUAD, or most cancers, is a kind of intractable disease with low cure rates. As cancers involve complex pathogenic factors and huge molecular networks, it is difficult to find molecular targets to achieve the expected therapeutic effect [5]. There have been many studies focusing on cancer research, including but not limited to cancer metabolic reprogramming, cell autophagy, cell apoptosis, stem cells,

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

cell cycle, angiogenesis [6], [7], cell adhesion, immunity and vaccines, microenvironment, epigenetic regulation, and even intestinal flora [8]. Many studies about cancers have analyzed the specific mechanisms of tumorigenesis, and most cancer researchers hope to develop new drugs for curing or even conquering a certain type of tumor through the exploration of mechanisms. Empirically, gene targets are required in the development of most cancer-targeted drugs, which are usually one specific gene or multiple gene combinations. Finding these gene targets contributes to the carcinogenesis and the targeted treatment. For LUAD, it is still a challenge to discover effective target genes, reflected by its low curing rate and few effective targeted drugs.

In recent years, network analysis on PPI networks has raised the interest of researchers, such as the node centrality method for identifying key proteins [9] and the MCODE method for mining functional modules in the PPI network [10]. Many novel network model and analysis methods are also proposed [11]–[15]. These methods have the potentials to be applied to PPI network analysis. However, most existing studies on PPI networks focus on the role of a single gene or the impact of a gene on the upstream and downstream genes, while rare attention is paid to the close relationships between multiple targeted genes which may play important role in the occurrence of cancer. For example, the interaction between tumor suppressor gene p53 and the potential tumor target SET can affect tumor growth [16]. Therefore, it is necessary to explore new network analysis methods to find multiple potential target genomes.

With the above considerations, the main contributions of this paper are two aspects:

1) This study adopts a novel overlapping gene analysis method. This method is used to find overlapping genes between network communities and hub genes, whose major principle is as follows. Firstly, the hub genes in the network are selected according to the importance index of nodes, where the top N nodes are selected to represent the most important N nodes. Then, the PPI network is divided into multiple communities, where the nodes in the same community are of great relevance and the nodes in the different communities have sparse connections. The overlapping genes are screened out from the hub genes and the genes in the same community. Experiments show that, for lung adenocarcinoma, this method can retrieve closely related and important target gene groups, instead of analyzing isolated target genes.

2) This study has found a possible targeted gene group for the occurrence and development of LUAD from the perspective of bioinformatics. Although the importance of LUAD is self-evident, the specific mechanism of the cell cycle in the occurrence and development of LUAD is still not very clear. With the help of network analysis, this paper deeply dissects the possible mechanism in the occurrence and development of LUAD. It is found that three critical genes (PLK1, CCNB1, and CDK1) are likely to cooperate to promote the occurrence and development of LUAD by regulating the cell cycle of LUAD cells. So far, to the best of our knowledge, there are still few studies to jointly study the effects of the interactions among the three genes. Our experiments provide a useful clue: it may be valuable to take the three genes as a whole in studying the occurrence and development of LUAD.

The rest of this paper is organized as follows. Section II introduces related work about lung adenocarcinoma. Section III provides the datasets and analysis tools. Section IV provides bioinformatics experiments and analysis results. Section V discusses the results in detail. This paper is concluded in Section VI.

II. RELATED WORK

Current researches about the molecular targeting of LUAD is abundant but far from enough. For the one hand, existing studies about LUAD are associated with mutations of several genes [17], including epidermal growth factor receptor (EGFR) [18]–[22], B-Raf proto-oncogene (BRAF) [20], [23], mitogen-activated protein kinase 1 (MAP2K1) [24], erb-b2 receptor tyrosine kinase 2 (HER2) [25], [26], serine/threonine kinase 11 (LKB1) [27], [28], etc. Also, some studies have shown that changes in the cell cycle [29]–[31], like the abnormal expression of cell cycle-related molecules of topoisomerase (DNA) II alpha (TOP2A), cyclin-dependent kinase 1 (CDK1) may affect the occurrence and development of LUAD [32], [33]. Meanwhile, LUAD has been reported to be associated with abnormal signal pathways [34], [35], such as the Ras / Raf / MEK / ERK classic tumor signaling pathway [36], [37]. Surgery and postoperative chemotherapy are the gold standards for the treatment of advanced and metastatic non-small cell lung cancer, but as the effective rate of the current first-line treatment is no more than 20-30% [38], it is still essential to discover the molecular mechanism in the occurrence and development of LUAD.

For the other, though some marketed drugs have shown their effectiveness in curing LUAD, such as gefitinib and erlotinib that target EGFR, the shortcomings of these drugs are also being noticed by many researchers in recent years. For example, gefitinib as a kind of molecular targeting drug can cause less damage to human body and achieve good curative effect in the clinic treatment of LUAD. However, patients often inevitably develop resistance within months to years [39], and the adverse reactions of gefitinib have been paid more and more attention by researchers. The occurrence of adverse reactions may involve multiple organs and systems, among which skin and accessories have the highest incidence of adverse reactions. Respiratory adverse reactions have the highest mortality rate. Serious adverse reactions can also occur in the circulatory and urinary systems [40], [41]. In a word, the emergence of drug resistance and adverse reactions require researchers to explore new drug targets, which further verify the complexity and importance of cancer research.

Fortunately, some researchers have noticed the above questions and focused on finding new molecular targets. Nowadays, bioinformatics-based targeted molecules

prediction of LUAD are principally targeted at genes, miRNA, lncRNA [42], circRNA, and epigenetic regulation of genome [43]–[46]. Although there may be overlap between different predictions, the level, number, and accuracy of predicted target molecules are different, and the results of analyses are also different. These predicted target molecules are not all effective, but as long as LUAD remains incurable, bioinformatics studies on LUAD are necessary. The major reason is that, from the determination of targeting molecules to pharmaceuticals, it requires many experiments, such as molecular experiments, cell experiments, animal experiments, and clinical trials, to verify the effectiveness of target molecules. These processes require the participation of many manpower and the investment of huge financial resources. Based on this, scientists should try their best to understand the mechanisms of target molecules before experiments, so that the valuable time and social wealth can be saved as much as possible.

III. DATASETS AND ANALYSIS METHODS

A. DATASETS

The dataset GSE2514 and the dataset of lung adenocarcinoma (LUAD)-related genes from the Gene Expression Profiling Interactive Analysis (GEPIA) database are utilized in this paper.

The dataset GSE2514, downloaded from Gene Expression Omnibus (GEO) database [47] (http://www.ncbi.nlm.nih.gov/geo), contains 20 tumor samples and 19 normal samples. It is based on the platform of GPL8300 (Affymetrix Human Genome U95 Version 2 Array).

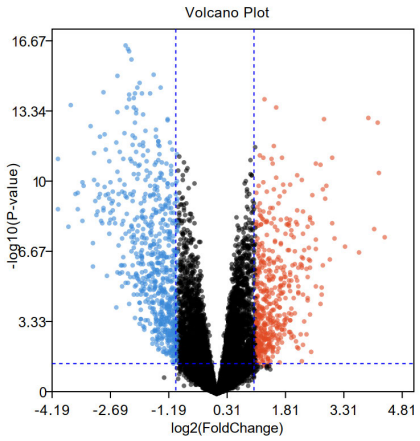


FIGURE 1. Volcano plot of GSE2514.

Fig. 1 describes the screening conditions for the dataset GSE2514. where each point is related to a gene in our dataset. In the x-axis, log_2 *(Fold Change)* is the base-2 logarithm of fold changes between tumor samples and normal samples. In the y-axis, $-log_{10}(p - value)$ represents the negative of the base-10 logarithm of the corrected P-value of the gene in the microarray experiment. The qualified genes have the following characteristics: the expression of the gene differs by more than 2 times in tumor samples and normal samples, and the microarray experiment of this gene shows that *adj. p-value* < 0.05. Therefore, in the figure, the red points represent that the expression of the genes in tumor samples is more than twice that in normal samples (up-regulated), and the points in blue represent that the expression of the genes in normal samples is more than twice that in tumor samples (down-regulated). The black dots represent genes that do not meet the screening conditions in the dataset GSE2514, for their expression differences between tumor samples and normal samples are less than 2-fold, or whose *adj. p-value* is larger than 0.05.

The dataset of GEPIA, which is downloaded from GEPIA (http://gepia.cancer-pku.cn/detail.php). It contains 483 LUAD samples and 347 normal samples. LIMMA is used to distinguish tumor vs paired normal samples. As recommended in the GEPIA database, the cutoff criteria of $|\log_2 FC| > 1$, $q - value < 0.01$ and $adj.p - value < 0.05$ are considered statistically significant.

B. ANALYSIS TOOLS

1) IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES (DEGs)

The dataset GSE2514 is extracted from GEO2R. GEO2R is an interactive web tool that enables users to compare two or more groups of samples in a GEO series so that the genes differentially expressed across experimental conditions can be identified. GEO2R performs comparisons on original submitter-supplied processed data tables using the GEO query and limma R packages from the Bioconductor project. The data are downloaded for further screening. $adj.p - value < 0.05$ and $|\log_2 FC| > 1$ are considered statistically significant. The overlapping DEGs among GSE2514 and GEPIA are produced by FunRich. So is the Venn diagram.

2) PROTEIN-PROTEIN INTERACTION (PPI) NETWORK CONSTRUCTION AND ANALYSIS

The PPI of DEGs is constructed by STRING (version 11.0, http://string-db.org) database [48], including 676 nodes and 5099 edges (Fig.2). *Combined score*>0.4 is considered statistically significant. Cytoscape (version 3.6.1) is an open-source software platform for visualizing complex networks and integrating these with any type of attribute data [49].

3) GO AND KEGG PATHWAY ENRICHMENT ANALYSIS

The Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8, https://david.ncifcrf.gov/) provides a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind a large list of genes [50], [51]. To further characterize the biological process (BP) and the pathway of DEGs, DAVID is used for Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. $p - value < 0.05$ is considered statistically significant.

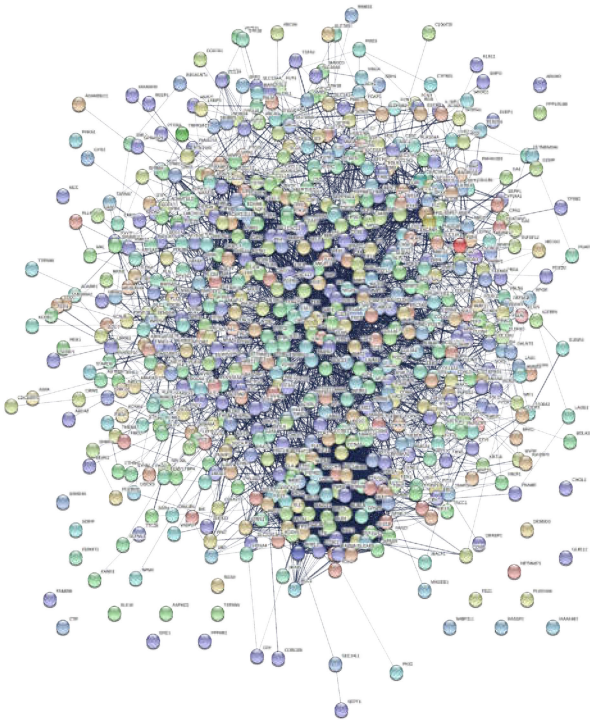


FIGURE 2. PPI network of DEGs.

4) GEPIA DATABASE VALIDATION

The overlapping genes are obtained among the top 20 hub genes and the most significant community. The GEPIA database is used for survival analysis and pathological stage analysis of the overlapping genes. The cutoff criteria are set to default. In the overall survival, disease-free survival (RFS), and pathological stage analysis, $p\text{-}value<0.05$ is considered statistically significant. Then, the pairwise correlation of the screened genes is analyzed. Pearson is selected for the Correlation Coefficient. The non-log scale for the calculation and the log-scale axis for visualization are used. The above analyses are based on the RNA sequencing expression data of the TCGA database and the GTEx database.

C. THE OVERLAPPING GENE ANALYSIS METHOD

As the PPI network is established, the degree of nodes in the network can be calculated out. The top N protein nodes with the highest degree are obtained as the hub nodes. Then, the most significant community in the PPI network is discovered based on the method of vertex weighting by local neighborhood density and outward traversal from a locally dense seed protein. Because of the correspondence between protein and gene, the overlapping genes between the top N hub genes and the most significant community can be selected out, termed as a gene group. Then, survival analyses and tumor pathological stage analyses are performed on the gene group to further select the genes with significant survival differences. Finally, the correlation analyses are performed on the obtained genes, the genes with strong correlations which are regarded as a key gene group are selected as possible clinical targets for the diagnosis and treatment. The concrete operation steps are as follows.

Step 1. Dividing the whole PPI network into communities by the MCODE algorithm [52] and selecting the most significant community. As shown in Fig. 4 (a), the most significant community includes 49 nodes and 1138 edges. MCODE has advantages over other graph clustering methods on two aspects: having a directional mode that allows fine-tuning of target clusters without considering the rest of the network and allowing the check of cluster interconnectivity of PPI networks. Parameter selection: Degree Cutoff: 2, Cluster Finding: Haircut, Node Density Cutoff: 0.1, Node Score Cutoff: 0.2, K-Core: 2, Max Depth: 100. Selecting the top 20 genes with the highest degree as the hub genes, as shown in Fig. 4(b). The number of overlapping genes of the most significant community and the hub genes is shown in Fig. 4(c).

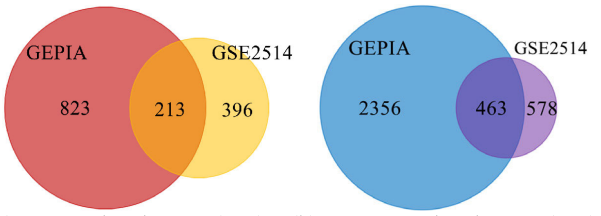
Step 2. Using the GEPIA database to conduct survival analyses and pathological stage analyses on the gene group, which is the overlapping genes between the top 20 hub genes and the most significant community.

Step 3. Selecting the genes with statistical differences in the cell cycle through survival analyses. Then, the correlation analyses are applied to the selected genes, which result in a key gene group that may provide clinical targets for the diagnosis and treatment.

IV. EXPERIMENTS

A. IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES (DEGs)

The dataset of GSE2514 contains 396 up-regulated genes and 578 down-regulated genes. The dataset of GEPIA contains 823 up-regulated genes and 2356 down-regulated genes. DEGs as the overlapping part between GSE2514 and GEPIA contain 213 up-regulated genes (Fig. 3(a)) and 463 down-regulated genes (Fig. 3(b)).

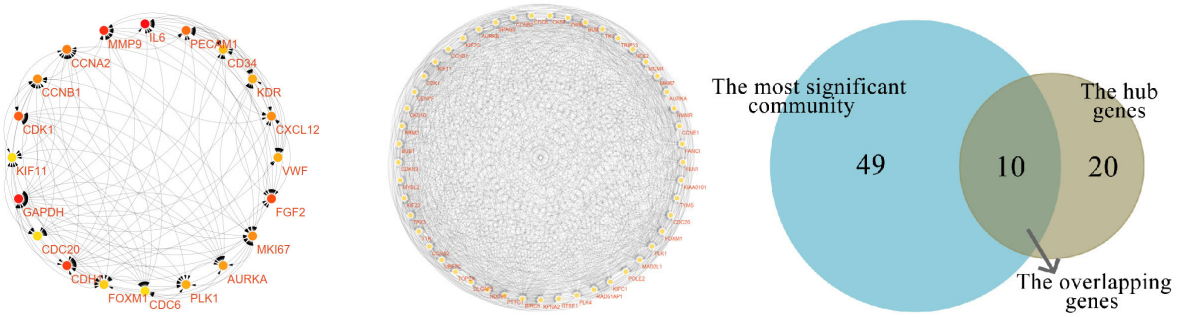


(a) Up-regulated genes (213) (b) Down-regulated genes (463)

FIGURE 3. The differentially expressed genes (DEGs).

B. GO AND KEGG PATHWAY ENRICHMENT ANALYSIS

An online analysis tool DAVID is utilized to analyze the biological process and pathway of the top 20 hub genes and the most significant community. Fig. 5(a)-(b) describes the biological process for the top 20 hub genes and the most significant community, respectively. The y-axis represents the biological process, whose two values shown in the x-axis represent the number of genes that participated in the biological process (referring to the bars in orange), and



(a) The top 20 hub genes (b) The most significant community (c) The overlapping part

FIGURE 4. PPI networks of the most significant community and the hub genes.

the negative of the base-10 logarithm of the corresponding p-value (referring to the bars in blue), respectively. Combing Fig. 5(a) and Fig. 5(b), it is found that most of the biological processes in which these two parts of genes are overlapped with each other and are closely related to the cell cycle. Fig. 5 (c)-(d) describe the KEGG pathway analyses of the top 20 hub genes and the most significant community in lung adenocarcinoma (LUAD) samples, respectively. The x-axis represents the pathway, whose two values shown in the y-axis represent the number of genes participated in the pathway (referring to the dots in orange), and the negative of the base-10 logarithm of the corresponding p-value (referring to the dots in blue), respectively. The two panels show that the pathway involving the most genes and the most significant pathway is the cell cycle.

The cell cycle refers to the whole process that a cell undergoes from the beginning of one division to the end of the next division, which is divided into two stages: interphase and division. The interphase is further divided into three stages, namely, the prophase of DNA synthesis (G1 phase), the stage of DNA synthesis (S phase), and the anaphase of DNA synthesis (G2 phase). A conclusion can be drawn from Fig. 5 that the abnormal cell cycle is an important factor in the occurrence and development of LUAD.

C. OVERALL SURVIVAL ANALYSIS AND DISEASE-FREE SURVIVAL ANALYSIS

The overall survival and disease-free survival are the main efficacy indicators to determine the prognosis of the tumor. After checking the overall survival and disease-free survival of the 10 overlapped genes, only three genes (PLK1, CCNB1, and CDK1) show statistical differences.

As shown in Fig. 6, the x-axis represents time and the y-axis represents percent survival. The red lines represent that the LUAD patients are with high expression of the corresponding gene, and the blue lines represent that the patients are with low expression of the corresponding gene. Fig. 6(a)-(c) demonstrate the overall survival of PLK1, CCNB1, and CDK1, respectively. Overall survival is defined as the time from randomization to death from any cause (the last follow-up time is for patients who are lost to follow-up; those who are still alive at the end of the study are the end of follow-up). Fig. 6 (d)-(f) demonstrate the disease-free survival of PLK1, CCNB1, and CDK1, respectively. Disease-free survival period denotes the time interval from randomization to the first tumor recurrence or metastasis or death of a subject for any reason. The results show that for the three genes, the survival rates with high expression are lower than those with low expression in both the overall survival period and disease-free survival. Therefore, PLK1, CCNB1, and CDK1 have the potentials to be significant biomarkers, prognostic indicators, or potential drug targets.

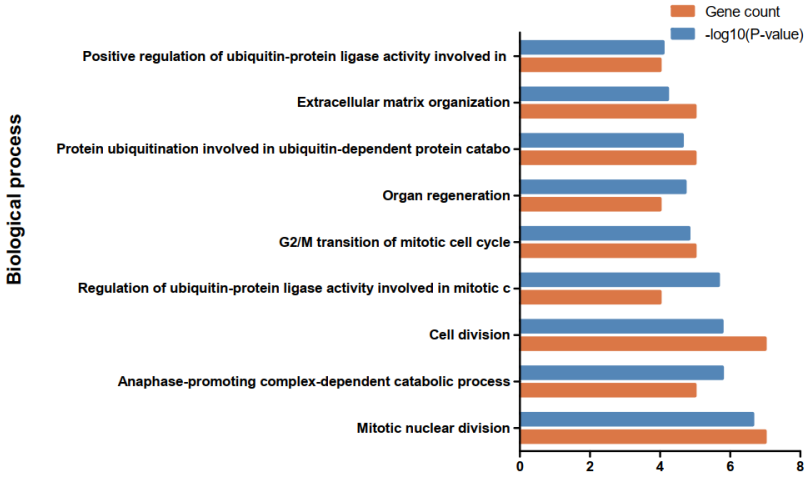
D. PATHOLOGICAL STAGE ANALYSIS AND CORRELATION ANALYSIS

The tumor pathological stages of PLK1, CCNB1, and CDK1 are analyzed with results shown in Fig. 7. There are significant differences in all results. The x-axis is the pathological stage grade of LUAD, and the y-axis is the relative expression of the corresponding gene. As is shown in the figure, the relative expression of the three genes is increased as the increasing pathological stage grade of LUAD. Clinical data shows that the higher the pathological stage grade of LUAD, the lower the cure rate. This confirms our previous conclusion that these three genes may play an important role in the occurrence and development of LUAD.

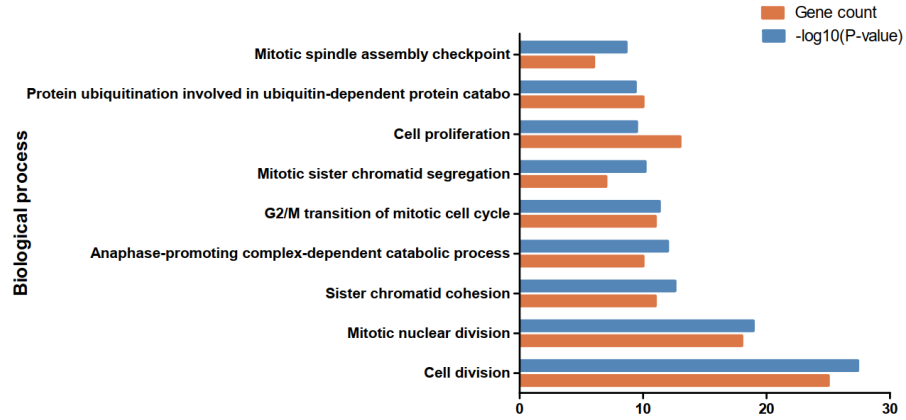
The correlation among PLK1, CCNB1, and CDK1 is analyzed in pairs, with results shown in Fig. 8. Normal samples and LUAD samples in the TCGA database and normal samples in the GETx database are selected, shown as black dots in the figure. The R values are 0.86 for PLK1 and CCNB1, 0.8 for CDK1 and CCNB1, 0.78 for CDK1 and PLK1, all of which are high values of correlation. Since the three genes are well-known genes in the cell cycle, it can be speculated that the three genes may cooperate to promote the occurrence and development of LUAD by regulating the cell cycle of LUAD cells.

V. DISCUSSION

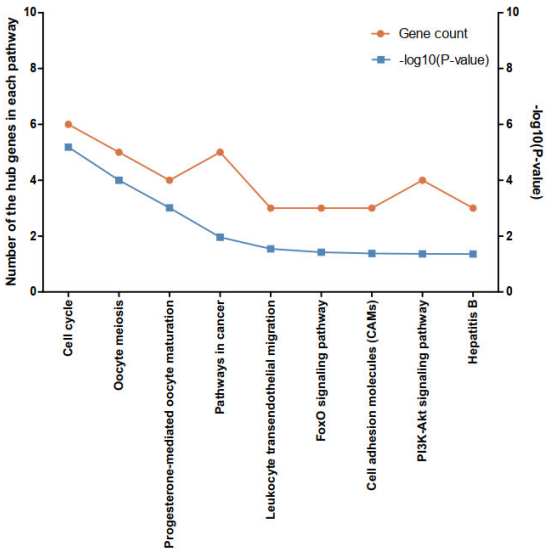
Lung adenocarcinoma (LUAD), the most common subtype of lung cancer, with its high incidence rate and high mortality rate, is a serious threat to human health [3], [4]. The occurrence and development of LUAD, from atypical adenomatous hyperplasia to adenocarcinoma in situ, to micro-invasive



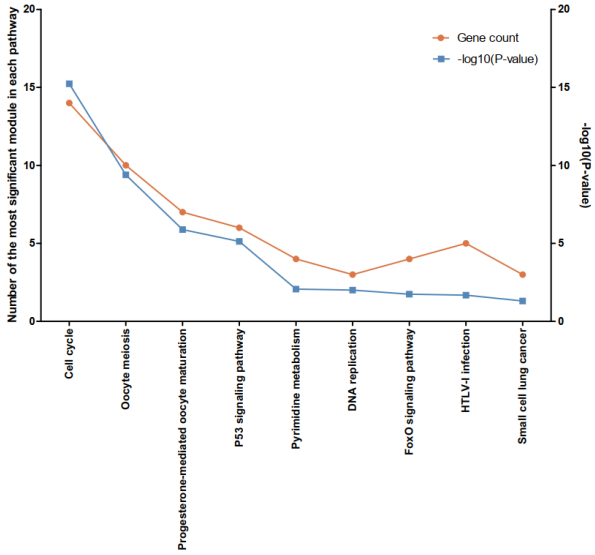
(a) Biological process for the top 20 hub genes



(b) Biological process for the most significant community



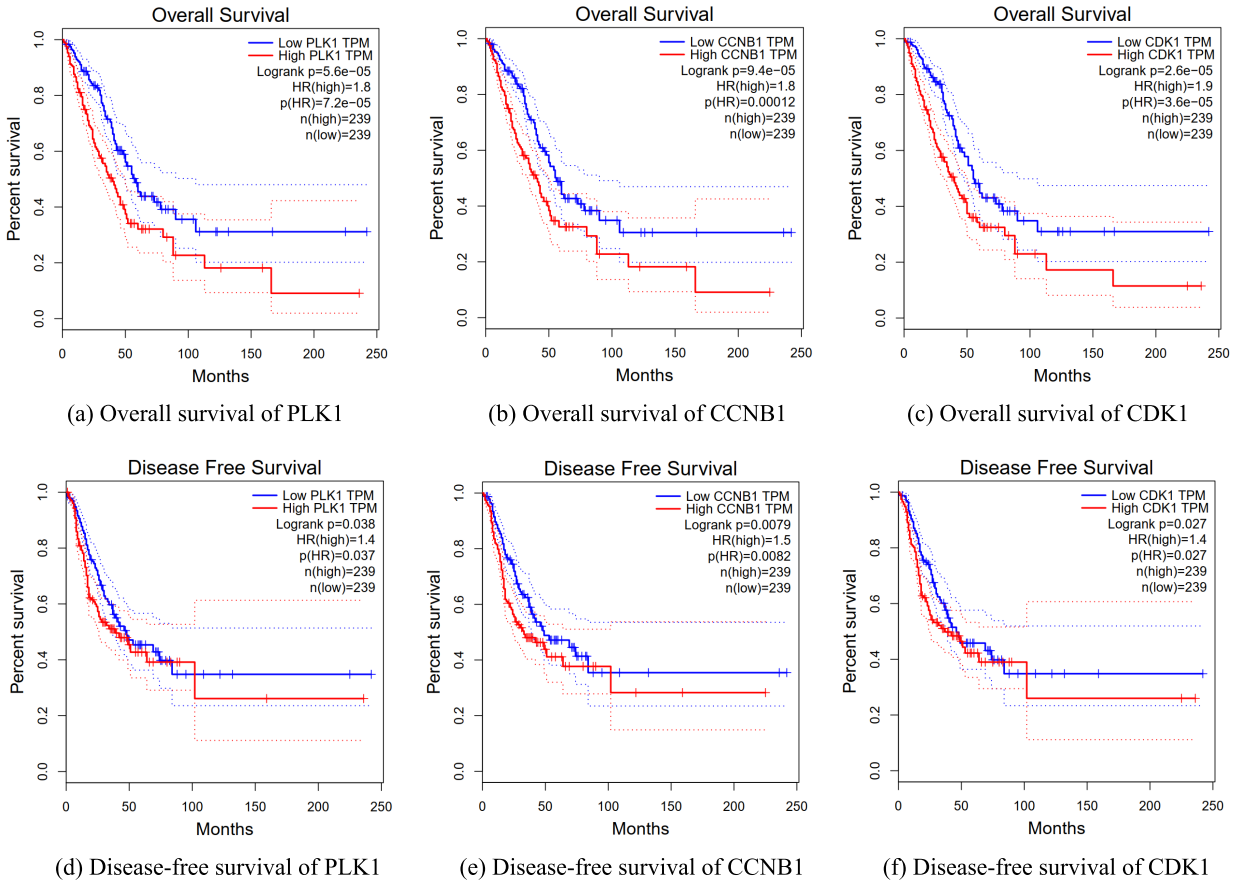
(c) KEGG pathway analysis of the top 20 hub genes



(d) KEGG pathway analysis of the most significant community

FIGURE 5. **GO and KEGG pathway enrichment analyses of the top 20 hub genes and the most significant community.**

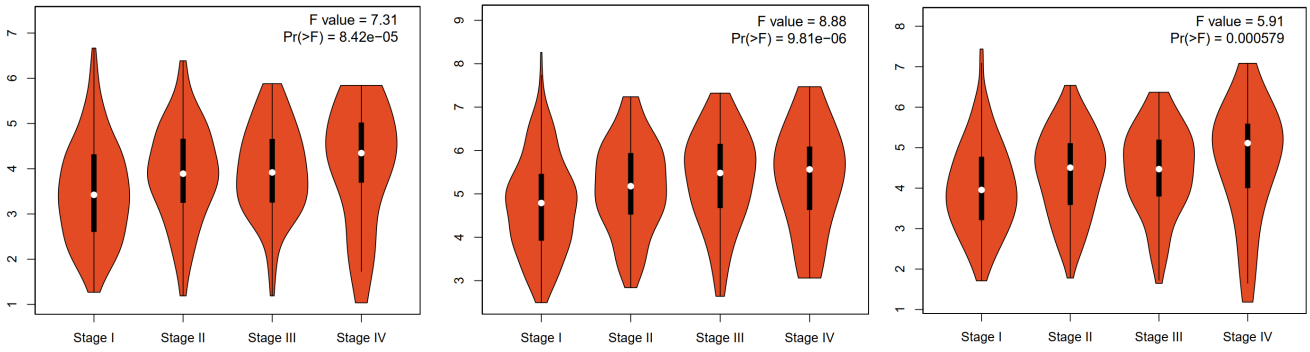
adenocarcinoma, and then to squamous infiltrating adenocarcinoma [53], is a complex biological process. Many molecules play important roles in this process, but the exact mechanism is still unclear. Previous studies have shown that epidermal growth factor receptor (EGFR) is highly correlated with targeted therapy for LUAD [18]–[20], and its mutation



(a) Overall survival of PLK1 (b) Overall survival of CCNB1 (c) Overall survival of CDK1

(d) Disease-free survival of PLK1 (e) Disease-free survival of CCNB1 (f) Disease-free survival of CDK1

FIGURE 6. Overall survival analyses and disease-free survival analysis of PLK1, CCNB1, and CDK1.



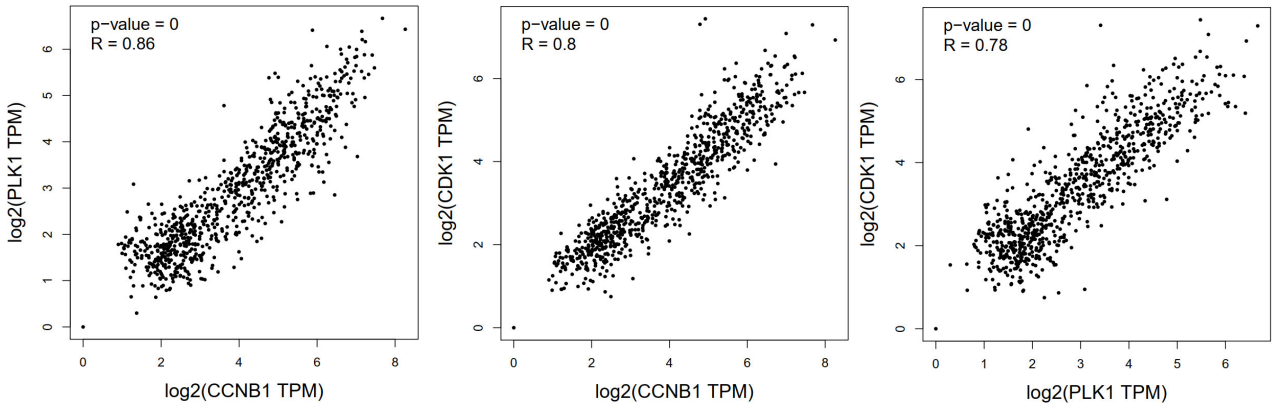
(a) PLK1 (b) CCNB1 (c) CDK1

FIGURE 7. Pathological stage analyses of PLK1, CCNB1, and CDK1.

rate decreases with the decrease of differentiation degree of non-small cell lung cancer [2]. The main mutations of EGFR are deletion mutation in frame No. 19 and L858R mutation in frame No. 21 [21], [22]. At present, EFGR tyrosine kinase inhibitors such as gefitinib and erlotinib have a fair anti-tumor effect in the treatment of LUAD, and further reveals the potential values of targeting key molecules in the tumorigenesis and development of LUAD.

Surgical resection of cancer tissue is the preferred treatment way of LUAD [38]. However, due to the clinical characteristics such as rapid progress, strong invasion, and high lethality, most patients rely on comprehensive treatments when diagnosed, including chemotherapy, radiochemotherapy, and molecular targeted therapy. Therefore, it is an urgent need to find potential markers to develop an efficient diagnosis and treatment of LUAD. With the rapid development of microarray technology, it has been proved to be an effective method to identify new biomarkers in complex diseases.

In our research, two mRNA microarray datasets are obtained and analyzed differentially expressed genes (DEGs) between LUAD samples and normal samples by strictly controlling the screening conditions, and identified a total



(a) CCNB1 and PLK1 (b) CCNB1 and CDK1 (c) PLK1 and CDK1

FIGURE 8. **Correlation analyses of PLK1, CCNB1, and CDK1.**

of 213 up-regulated genes and 463 down-regulated genes. Then, the protein-protein interaction (PPI) network of DEGs is constructed using the STRING database, which contains 676 nodes and 5099 edges. The hub genes of the PPI network of DEGs are screened, among which the top 20 hub genes are selected for further analysis. The result of enrichment analysis shows that hub genes are mainly concentrated in the cell cycle. Then the most significant community is obtained by community analysis of the PPI network of DEGs and the enrichment analysis on the community is carried out. Results show that genes in the most significant community are also mainly concentrated in the cell cycle. It is noticed that both the top 20 hub genes and the most significant community seem to be closely related to the cell cycle. Interestingly, 10 of the top 20 hub genes also appeared in the most significant community. These 10 genes are polo-like kinase 1 (PLK1), cyclin B1 (CCNB1), cyclin-dependent kinase 1 (CDK1), aurora kinase A (AURKA), cell division cycle 6 (CDC6), kinesin family member 11 (KIF11), a marker of proliferation Ki-67 (MKI67), forkhead box M1 (FOXM1), cyclin A2 (CCNA2), and cell division cycle 20 (CDC20), most of which play an important role in the cell cycle. Survival analyses are taken to these 10 genes by Gene Expression Profiling Interactive Analysis (GEPIA) database, and it is shown that only PLK1, CCNB1, and CDK1 are statistically different in disease-free survival and overall survival, indicating that they are connected with tumor prognosis. To explore whether PLK1, CCNB1, and CDK1 are related to the pathological stage of LUAD, pathological stage analyses are taken to them. The three genes show a statistical difference in tumor pathological stage, whose relative expression is increased as the increase of the pathological stage grade of LUAD. This indicates that PLK1, CCNB1, CDK1 play essential roles in the occurrence and development of LUAD. More interestingly, PLK1, CCNB1, and CDK1 are indispensably involved in the cell cycle. The correlation among PLK1, CCNB1, and CDK1 is analyzed pairwise, with results shown the R values are 0.86 for PLK1 and CCNB1, 0.8 for CDK1 and CCNB1, 0.78 for CDK1 and PLK1, all of which are high values of correlation. All kinds of evidence prove that PLK1, CCNB1, and CDK1 are key genes in LUAD, which possibly promote the occurrence and development of LUAD by regulating the cell cycle of LUAD cells.

With the understanding of the cell cycle, more and more attention has been paid to the relationship between cell cycle disorder and tumors [54]. The cell cycle is the most important process in cell life activities. Many cell cycle events can be completed by regulating factors in different phases. The regulation mechanism is mainly based on PPI, and the main process is a series of cascade reactions caused by signal transmission. Studies have shown that almost all tumorigenesis is related to the abnormally stunted growth differentiation and the apoptosis caused by the disorder of the cell cycle regulation mechanism [55], [56].

PLK1 belongs to the mitotic serine/threonine kinase family and is highly conserved in eukaryotes. It helps maintain the stability of the cell genome. Therefore, the change of PLK1 may lead to an increase in the mutation rate of defective cells and the development of tumors. The expression of PLK1 starts at the end of the S stage and reaches its peak when CCNB1 and cell division cycle 25C (CDC25C) are phosphorylated, and then cells begin mitosis [57]. The specific binding of PLK1 Polo box binding domain (PBD) to CDC25C gives rise to the phosphorylation of CDC25C and then activates the CDK1 / CCNB1 complex [58], finally promotes mitosis by phosphorylation and activation of enzymes related to mitosis [59]. PLK1 is overexpressed in a variety of such human tumors as nasopharyngeal carcinoma [60], ovarian cancer, prostate cancer, gastric cancer, breast cancer, esophageal cancer [61], leukemia, and lung cancer [62], and is often used as an indicator of poor prognosis. It has been found that PLK1 plays an important part in the PDK1/c-Myc pathway, maintaining the growth and differentiation of tumor cells [63], and it is pointed out that inhibiting PLK1 overexpression can curb tumor recurrence and metastasis [64].

CCNB1 is one of the important members of the cyclin family and plays an important regulatory role in G2 / M phase. CCNB1 starts to synthesize at the end of S and early G2, reaches its peak at the M phase, and then decreases rapidly. As a regulatory subunit, CCNB1 binds to CDK1 to form the MPF complex, promoting the transition from G2 to M phase [65], [66], and is input into the nucleus before the nuclear membrane ruptures. Previous studies have concluded that CCNB1 is overexpressed in many tumor cells. As is mentioned above, overexpression of CCNB1 may lead to dysfunction of MPF phosphorylation regulation, thus leading to malignant transformation of cells [67], [68].

CDK1 is a member of the serine / threonine-protein kinase family, and CDK1 has kinase activity only when it binds with cyclin. Also, it is a key factor in the regulatory network of cell cycle and gamete meiosis maturation [69]. As mentioned above, CDK1 plays a key role in cell cycle transition from G2 to M. Overexpression of CDK1 also leads to cell cycle disorder, and often causes malignant cell transformation and tumor formation. It has been reported that CDK1 is overexpressed in many kinds of tumors, such as colorectal cancer [70], lung cancer [71], oral squamous cell carcinoma [72], etc. High expression of CDK1 often indicates poor prognosis of patients.

These studies reflect that PLK1, CCNB1, and CDK1 play significant roles in the occurrence and development of various types of tumors, including the LUAD we studied, which further confirm our results. Moreover, it is found that PLK1, CCNB1, and CDK1, as key genes in LUAD, not only have a significant difference in overall survival, disease-free survival, and pathological stage analysis but also have an inalienable correlation with each other, which may cast new ideas for future research. It remains a question that how is the relationship between genes and how they cooperatively promote the occurrence and development of LUAD.

VI. CONCLUSION

Lung adenocarcinoma (LUAD) has not been conquered yet, and its target molecules are still unclear. This study aims to identify differentially expressed genes (DEGs) that are possibly involved in the occurrence and development of LUAD, so. To achieve this goal, 676 DEGs are identified and 10 overlapping genes are obtained from the most significant community and the top 20 hub genes. Through survival analysis and pathological stage analysis, it is found that PLK1, CCNB1, and CDK1 are key genes in the occurrence and development of LUAD, and can be used as biomarkers of LUAD. The correlation analysis found that they are very likely to cooperate to promote the occurrence and development of LUAD, but the delicate mechanism still needs further research.

Since PLK1, CCNB1 and, CDK1 are all extremely important genes in the cell cycle, some important work in the future deserves attention, e. g., how they interact with each other, how they promote the occurrence and development of LUAD by regulating the cell cycle of LUAD cells, and what the specific mechanism is.

LIST OF ABBREVIATIONS

LUAD	Lung adenocarcinoma, a type of lung cancer.
Gene	A specific sequence of nucleotides in DNA or RNA, usually located on a chromosome.
Gene group	A set of genes that cooperatively take effect.
Targeted genes	Certain abnormally expressed genes with diagnostic or therapeutic value in diseases, usually related to the occurrence, development, and prognosis of the disease.
Protein	A product encoded by a gene, which is the basic unit of the structure and function of an organism.
PPI network	Protein-protein interaction network, the physical contacts between proteins in the cell, reflecting the relationships between the corresponding genes.
Cell cycle	The series of events happening in a cell, resulting in the duplication of DNA and division of cellular materials to produce two child cells.
GEO database	A database of Gene Expression Omnibus.
GSE2514 dataset	A dataset that is download from GEO.
GEPIA database	A database for cancer big data visualization and cancer big data analysis.
STRING database	A database for searching protein-protein interactions.
TCGA database	Currently the largest cancer gene information database.
GETx database	A database that provides gene expression data of healthy organs.
DEGs	Differentially expressed genes, the genes with differences in expression between lung adenocarcinoma samples and normal samples.
MCODE	Molecular COmplex DEtection, the community detection algorithm.
Survival analysis	A method that studies the relationship between the expression of a certain gene and the survival time of cancer patients.
Overall survival analysis	A type of survival analysis that studies the relationship between the expression of a certain gene and the overall survival time of cancer patients.
Disease-free survival analysis	A type of survival analysis that studies the relationship between the expression of a certain gene and the disease-free survival time of cancer patients.

Enrichment analysis	A bioinformatics method for obtaining the biological process and pathway by clustering the genes based on a priori knowledge map.
Pathological stage analysis	A method that analyzes the relationship between the expression of a certain gene and tumor pathological stage.
Correlation analysis	A method that analyzes the relationship between the expressions of two different genes.
EGFR	The abbreviation of the gene ''Epidermal Growth Factor Receptor''
PLK1	The abbreviation of the gene ''Polo-like kinase 1''
CCNB1	The abbreviation of the gene ''Cyclin B1''
CDK1	The abbreviation of the gene ''Cyclin-dependent kinase 1''

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, ''Cancer statistics, 2018,'' *CA: A Cancer J. for Clinicians*, vol. 68, no. 1, pp. 7–30, Jan. 2018.

[2] 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,'' *A Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[3] W. D. Travis, E. Brambilla, A. P. Burke, A. Marx, and A. G. Nicholson, ''Introduction to the 2015 world health organization classification of tumors of the lung, pleura, thymus, and heart,'' *J. Thoracic Oncol.*, vol. 10, no. 9, pp. 1240–1242, Sep. 2015.

[4] L. A. Torre, R. L. Siegel, and A. Jemal, ''Lung cancer statistics,'' in *Lung Cancer and Personalized Medicine*. 2016, pp. 1–19.

[5] X. Wu, L. Chen, and X. Wang, ''Network biomarkers, interaction networks and dynamical network biomarkers in respiratory diseases,'' *Clin. Transl. Med.*, vol. 3, no. 1, pp. 1–7, Dec. 2014.

[6] J. Zhou, S. Dong, D. Tang, and X. Wu, ''A vascular invasive tumor growth optimization algorithm for multi-objective optimization,'' *IEEE Access*, vol. 8, pp. 29467–29488, 2020.

[7] M. Dorraki, A. Fouladzadeh, A. Allison, C. S. Bonder, and D. Abbott, ''Angiogenic networks in tumors-insights via mathematical modeling,'' *IEEE Access*, vol. 8, pp. 43215–43228, 2020.

[8] R. Villéger, A. Lopès, G. Carrier, J. Veziant, E. Billard, N. Barnich, J. Gagnière, E. Vazeille, and M. Bonnet, ''Intestinal microbiota: A novel target to improve anti-tumor treatment?'' *Int. J. Mol. Sci.*, vol. 20, no. 18, p. 4584, Sep. 2019.

[9] W. Lou, J. Chen, B. Ding, D. Chen, H. Zheng, D. Jiang, L. Xu, C. Bao, G. Cao, and W. Fan, ''Identification of invasion-metastasis-associated microRNAs in hepatocellular carcinoma based on bioinformatic analysis and experimental validation,'' *J. Transl. Med.*, vol. 16, no. 1, p. 266, Dec. 2018.

[10] L. Li, Q. Lei, S. Zhang, L. Kong, and B. Qin, ''Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis,'' *Oncol. Rep.*, vol. 38, no. 5, pp. 2607–2618, Nov. 2017.

[11] Y. Wang, Z. Bu, H. Yang, H.-J. Li, and J. Cao, ''An effective and scalable overlapping community detection approach: Integrating social identity model and game theory,'' *Appl. Math. Comput.*, vol. 390, Feb. 2021, Art. no. 125601.

[12] Z. Bu, Y. Wang, H.-J. Li, J. Jiang, Z. Wu, and J. Cao, ''Link prediction in temporal networks: Integrating survival analysis and game theory,'' *Inf. Sci.*, vol. 498, pp. 41–61, Sep. 2019.

[13] J. Cao, Z. Bu, Y. Wang, H. Yang, J. Jiang, and H.-J. Li, ''Detecting prosumer-community groups in smart grids from the multiagent perspective,'' *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 8, pp. 1652–1664, Aug. 2019.

[14] Z. Bu, H.-J. Li, C. Zhang, J. Cao, A. Li, Y. Shi, ''Graph K-means based on leader identification, dynamic game, and opinion dynamics,'' *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 7, pp. 1348–1361, Jul. 2019.

[15] A. Song, Y. Liu, Z. Wu, M. Zhai, and J. Luo, ''A local random walk model for complex networks based on discriminative feature combinations,'' *Expert Syst. Appl.*, vol. 118, pp. 329–339, Mar. 2019.

[16] D. Wang, N. Kon, G. Lasso, L. Jiang, W. Leng, W.-G. Zhu, J. Qin, B. Honig, and W. Gu, ''The p53-SET interplays reveal a new mode of acetylation-dependent regulation,'' *Nature*, vol. 538, no. 7623, p. 118, 2016.

[17] J. J.-K. Lee, S. Park, H. Park, S. Kim, J. Lee, J. Lee, J. Youk, K. Yi, Y. An, I. K. Park, C. H. Kang, D. H. Chung, T. M. Kim, Y. K. Jeon, D. Hong, P. J. Park, Y. S. Ju, and Y. T. Kim, ''Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma,'' *Cell*, vol. 177, no. 7, pp. 1842–1857, 2019.

[18] B. Han, S. Tjulandin, K. Hagiwara, N. Normanno, L. Wulandari, K. Laktionov, A. Hudoyo, Y. He, Y.-P. Zhang, M.-Z. Wang, C. Y. Liu, M. Ratcliffe, R. McCormack, and M. Reck, ''EGFR mutation prevalence in asia-pacific and russian patients with advanced NSCLC of adenocarcinoma and non-adenocarcinoma histology: The IGNITE study,'' *Lung Cancer*, vol. 113, pp. 37–44, Nov. 2017.

[19] F. Hsu, A. De Caluwe, D. Anderson, A. Nichol, T. Toriumi, and C. Ho, ''EGFR mutation status on brain metastases from non-small cell lung cancer,'' *Lung Cancer*, vol. 96, pp. 101–107, Jun. 2016.

[20] H. Ding, Z. Zhuang, J. Xie, H. Huang, Z. Tao, and Z. Liu, ''Durable clinical response of advanced lung adenocarcinoma harboring EGFR-19del/T790M/BRAF^{V600E} mutations after treating with osimertinib and dabrafenib plus trametinib: A case report,'' *OncoTargets Therapy*, vol. 13, p. 7933, Aug. 2020.

[21] J. L. Marks, S. Broderick, Q. Zhou, D. Chitale, A. R. Li, M. F. Zakowski, M. G. Kris, V. W. Rusch, C. G. Azzoli, V. E. Seshan, M. Ladanyi, and W. Pao, ''Prognostic and therapeutic implications of EGFR and KRAS mutations in resected lung adenocarcinoma,'' *J. Thoracic Oncol.*, vol. 3, no. 2, pp. 111–116, Feb. 2008.

[22] J.-Y. Yu, S.-F. Yu, S.-H. Wang, H. Bai, J. Zhao, T.-T. An, J.-C. Duan, and J. Wang, ''Clinical outcomes of EGFR-TKI treatment and genetic heterogeneity in lung adenocarcinoma patients with EGFR mutations on exons 19 and 21,'' *Chin. J. Cancer*, vol. 35, no. 1, p. 30, Dec. 2016.

[23] P. Nieto, C. Ambrogio, L. Esteban-Burgos, G. Gómez-López, M. T. Blasco, Z. Yao, R. Marais, N. Rosen, R. Chiarle, D. G. Pisano, M. Barbacid, and D. Santamaría, ''A BRAF kinase-inactive mutant induces lung adenocarcinoma,'' *Nature*, vol. 548, no. 7666, pp. 239–243, Aug. 2017.

[24] M. E. Arcila, A. Drilon, B. E. Sylvester, C. M. Lovly, L. Borsu, B. Reva, M. G. Kris, D. B. Solit, and M. Ladanyi, ''MAP2K1 (MEK1) mutations define a distinct subset of lung adenocarcinoma associated with smoking,'' *Clin. Cancer Res.*, vol. 21, no. 8, pp. 1935–1943, 2015.

[25] B. T. Li, D. S. Ross, D. L. Aisner, J. E. Chaft, M. Hsu, S. L. Kako, M. G. Kris, M. Varella-Garcia, and M. E. Arcila, ''HER2 amplification and HER2 mutation are distinct molecular targets in lung cancers,'' *J. Thoracic Oncol.*, vol. 11, no. 3, pp. 414–419, Mar. 2016.

[26] M. G. Kris, D. R. Camidge, G. Giaccone, T. Hida, B. T. Li, J. O'Connell, I. Taylor, H. Zhang, M. E. Arcila, Z. Goldberg, and P. A. Jänne, ''Targeting HER2 aberrations as actionable drivers in lung cancers: Phase II trial of the pan-HER tyrosine kinase inhibitor dacomitinib in patients with HER2-mutant or amplified tumors,'' *Ann. Oncol.*, vol. 26, no. 7, pp. 1421–1427, Jul. 2015.

[27] L. Bonanno, E. Zulato, A. Pavan, I. Attili, G. Pasello, P. Conte, and S. Indraccolo, ''LKB1 and tumor metabolism: The interplay of immune and angiogenic microenvironment in lung cancer,'' *Int. J. Mol. Sci.*, vol. 20, no. 8, p. 1874, Apr. 2019.

[28] A. Galan-Cobo, P. Sitthideatphaiboon, X. Qu, A. Poteete, M. A. Pisegna, P. Tong, P.-H. Chen, L. K. Boroughs, M. L. M. Rodriguez, W. Zhang, F. Parlati, J. Wang, V. Gandhi, F. Skoulidis, R. J. DeBerardinis, J. D. Minna, and J. V. Heymach, ''LKB1 and KEAP1/NRF2 pathways cooperatively promote metabolic reprogramming with enhanced glutamine dependence in KRAS-mutant lung adenocarcinoma,'' *Cancer Res.*, vol. 79, no. 13, pp. 3251–3267, Jul. 2019.

[29] J. Luo, K. Liu, Y. Yao, Q. Sun, X. Zheng, B. Zhu, Q. Zhang, L. Xu, Y. Shen, and B. Ren, ''DMBX1 promotes tumor proliferation and regulates cell cycle progression via repressing OTX2-mediated transcription of p21 in lung adenocarcinoma cell,'' *Cancer Lett.*, vol. 453, pp. 45–56, Jul. 2019.

[30] B. Chen, B. Zhang, L. Xia, J. Zhang, Y. Chen, Q. Hu, and C. Zhu, ''Knockdown of eukaryotic translation initiation factor 4E suppresses cell growth and invasion, and induces apoptosis and cell cycle arrest in a human lung adenocarcinoma cell line,'' *Mol. Med. Rep.*, vol. 12, no. 6, pp. 7971–7978, Dec. 2015, doi: 10.3892/mmr.2015.4468.

[31] J. Zhang, Y. Gao, C. Ma, and Y. Wang, "Murrayanine induces cell cycle arrest, oxidative stress, and inhibition of phosphorylated p38 expression in A549 lung adenocarcinoma cells," *Med. Sci. Monit. Int. Med. J. Express Clin. Res.*, vol. 25, pp. 2002–2008, Mar. 2019.
[32] W. Guo, S. Sun, L. Guo, P. Song, X. Xue, H. Zhang, G. Zhang, Z. Wang, B. Qiu, F. Tan, Q. Xue, Y. Gao, S. Gao, and J. He, "Elevated TOP2A and UBE2C expressions correlate with poor prognosis in patients with surgically resected lung adenocarcinoma: A study based on immunohistochemical analysis and bioinformatics," *J. Cancer Res. Clin. Oncol.*, vol. 146, pp. 821–841, Feb. 2020.
[33] J. Chu, S. Wu, and D. Xing, "Survivin mediates self-protection through ROS/cdc25c/CDK1 signaling pathway during tumor cell apoptosis induced by high fluence low-power laser irradiation," *Cancer Lett.*, vol. 297, no. 2, pp. 207–219, Nov. 2010.
[34] H. Guan, T. Zhu, S. Wu, S. Liu, B. Liu, J. Wu, J. Cai, X. Zhu, X. Zhang, M. Zeng, J. Li, E. Song, and M. Li, "Long noncoding RNA LINC00673-v4 promotes aggressiveness of lung adenocarcinoma via activating WNT/β-catenin signaling," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 28, pp. 14019–14028, Jul. 2019.
[35] N. Yin, Y. Liu, A. Khoor, X. Wang, E. A. Thompson, M. Leitges, V. Justilien, C. Weems, N. R. Murray, and A. P. Fields, "Protein kinase C_ι and Wnt/β-catenin signaling: Alternative pathways to Kras/Trp53-driven lung adenocarcinoma," *Cancer Cell*, vol. 36, no. 2, pp. 156–167, 2019.
[36] S. J. Heidorn, C. Milagre, S. Whittaker, A. Nourry, I. Niculescu-Duvas, N. Dhomen, J. Hussain, J. S. Reis-Filho, C. J. Springer, C. Pritchard, and R. Marais, "Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF," *Cell*, vol. 140, no. 2, pp. 209–221, Jan. 2010.
[37] G. Hrustanovic and T. G. Bivona, "RAS signaling in ALK fusion lung cancer," *Small GTPases*, vol. 7, no. 1, pp. 32–33, Jan. 2016.
[38] M. Reck, D. F. Heigener, T. Mok, J.-C. Soria, and K. F. Rabe, "Management of non-small-cell lung cancer: Recent developments," *Lancet*, vol. 382, no. 9893, pp. 709–719, Aug. 2013.
[39] J. H. Cho, Y.-M. You, Y. I. Yeom, D. C. Lee, B.-K. Kim, M. Won, B. C. Cho, M. Kang, S. Park, S.-J. Yang, J. S. Kim, J.-A. Kim, and K. C. Park, "RNF25 promotes gefitinib resistance in EGFR-mutant NSCLC cells by inducing NF-κKB-mediated ERK reactivation," *Cell Death Disease*, vol. 9, no. 6, pp. 1–13, Jun. 2018.
[40] H. Sun, D. Xue, and Y.-Q. Li, "A severe dermatologic adverse effect related with gefitinib: Case report and review of the literature," *J. Cancer Res. Therapeutics*, vol. 9, no. 5, p. 110, 2013.
[41] Q. Liu, H. Bai, M. Shi, and J. Zhang, "Erlotinib and gefitinib treatments of the lung cancer in an elderly patient result in gastrointestinal bleeding," *Pakistan J. Med. Sci.*, vol. 29, no. 5, p. 1278, Aug. 2013.
[42] B. Wang and J. Zhang, "Logistic regression analysis for LncRNA-disease association prediction based on random forest and clinical stage data," *IEEE Access*, vol. 8, pp. 35004–35017, 2020.
[43] H.-M. Zhang, B. Yang, H.-F. Chen, X.-H. Chi, Y.-B. Xi, X.-M. Chen, B. Guo, P.-F. He, and X.-C. Lu, "Prognosis-related miRNA bioinformatics screening of lung adenocarcinoma and its clinical significance," *Chin. J. Appl. Physiol.*, vol. 34, no. 6, pp. 530–535, 2018.
[44] M. Qiu, W. Xia, R. Chen, S. Wang, Y. Xu, Z. Ma, W. Xu, E. Zhang, J. Wang, T. Fang, J. Hu, G. Dong, R. Yin, J. Wang, and L. Xu, "The circular RNA circPRKCI promotes tumor growth in lung adenocarcinoma," *Cancer Res.*, vol. 78, no. 11, pp. 2839–2851, Jun. 2018.
[45] Y.-H. Zhang, T. Zeng, X. Pan, W. Guo, Z. Gan, Y. Zhang, T. Huang, and Y.-D. Cai, "Screening dys-methylation genes and rules for cancer diagnosis by using the pan-cancer study," *IEEE Access*, vol. 8, pp. 489–501, 2020.
[46] Z. Si, H. Yu, and Z. Ma, "Learning deep features for DNA methylation data analysis," *IEEE Access*, vol. 4, pp. 2732–2737, 2016.
[47] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 207–210, Jan. 2002.
[48] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. V. Mering, "STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, Jan. 2019.
[49] P. Shannon, A. Markiel, O. Ozier, N. S Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003.
[50] B. T. Sherman and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, p. 44, 2009.
[51] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Res.*, vol. 37, no. 1, pp. 1–13, Jan. 2009.
[52] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinf.*, vol. 4, no. 2, pp. 1–27, 2003.
[53] K. Inamura, "Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: Tumor initiation and progression," *Int. J. Mol. Sci.*, vol. 19, no. 4, p. 1259, Apr. 2018.
[54] M. Ingham and G. K. Schwartz, "Cell-cycle therapeutics come of age," *J. Clin. Oncol.*, vol. 35, no. 25, p. 2949, 2017.
[55] K. Engeland, "Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM," *Cell Death Differentiation*, vol. 25, no. 1, pp. 114–132, Jan. 2018.
[56] P. Icard, L. Fournel, Z. Wu, M. Alifano, and H. Lincet, "Interconnection between metabolism and cell cycle in cancer," *Trends Biochem. Sci.*, vol. 44, no. 6, pp. 490–501, Jun. 2019.
[57] C. McInnes and M. D. Wyatt, "PLK1 as an oncology target: Current status and future potential," *Drug Discovery Today*, vol. 16, nos. 13–14, pp. 619–625, Jul. 2011.
[58] C. Sánchez-Huertas and J. Lüders, "The augmin connection in the geometry of microtubule networks," *Current Biol.*, vol. 25, no. 7, pp. R294–R299, Mar. 2015.
[59] R. F. Lera, G. K. Potts, A. Suzuki, J. M. Johnson, E. D. Salmon, J. J. Coon, and M. E. Burkard, "Decoding polo-like kinase 1 signaling along the kinetochore–centromere axis," *Nature Chem. Biol.*, vol. 12, no. 6, pp. 411–418, Jun. 2016.
[60] W. Shi, N. M. Alajez, C. Bastianutto, A. B. Y. Hui, J. D. Mocanu, E. Ito, P. Busson, K.-W. Lo, R. Ng, J. Waldron, B. O'Sullivan, and F.-F. Liu, "Significance of Plk1 regulation by miR-100 in human nasopharyngeal cancer," *Int. J. Cancer*, vol. 126, no. 9, pp. 2036–2048, 2010.
[61] Y.-B. Feng, D.-C. Lin, Z.-Z. Shi, X.-C. Wang, X.-M. Shen, Y. Zhang, X.-L. Du, M.-L. Luo, X. Xu, Y.-L. Han, Y. Cai, Z.-Q. Zhang, Q.-M. Zhan, and M.-R. Wang, "Overexpression of PLK1 is associated with poor survival by inhibiting apoptosis via enhancement of survivin level in esophageal squamous cell carcinoma," *Int. J. Cancer*, vol. 124, no. 3, pp. 578–588, Feb. 2009.
[62] Z.-X. Wang, D. Xue, Z.-L. Liu, B.-B. Lu, H.-B. Bian, X. Pan, and Y.-M. Yin, "Overexpression of polo-like kinase 1 and its clinical significance in human non-small cell lung cancer," *Int. J. Biochem. Cell Biol.*, vol. 44, no. 1, pp. 200–210, Jan. 2012.
[63] J. Tan, Z. Li, P. L. Lee, P. Guan, M. Y. Aau, S. T. Lee, M. Feng, C. Z. Lim, E. Y. J. Lee, Z. N. Wee, Y. C. Lim, R. K. M. Karuturi, and Q. Yu, "PDK1 signaling toward PLK1–MYC activation confers oncogenic transformation, tumor-initiating cell activation, and resistance to mTOR-targeted therapy," *Cancer Discovery*, vol. 3, no. 10, pp. 1156–1171, Oct. 2013.
[64] K. Hu, J. H. Law, A. Fotovati, and S. E. Dunn, "Small interfering RNA library screen identified polo-like kinase-1 (PLK1) as a potential therapeutic target for breast cancer that uniquely eliminates tumor-initiating cells," *Breast Cancer Res.*, vol. 14, no. 1, pp. 1–15, Feb. 2012.
[65] L. A. Porter, G. Singh, and J. M. Lee, "Abundance of cyclin B1 regulates ?-radiation–induced apoptosis," *Blood, J. Amer. Soc. Hematol.*, vol. 95, no. 8, pp. 2645–2650, 2000.
[66] U. Asghar, A. K. Witkiewicz, N. C. Turner, and E. S. Knudsen, "The history and future of targeting cyclin-dependent kinases in cancer therapy," *Nature Rev. Drug Discovery*, vol. 14, no. 2, pp. 130–146, Feb. 2015.
[67] H. Kao, J. A. Marto, T. K. Hoffmann, J. Shabanowitz, S. D. Finkelstein, T. L. Whiteside, D. F. Hunt, and O. J. Finn, "Identification of cyclin b1 as a shared human epithelial tumor-associated antigen recognized by t cells," *J. Experim. Med.*, vol. 194, no. 9, pp. 1313–1324, Nov. 2001.
[68] A. M. Egloff, L. A. Vella, and O. J. Finn, "Cyclin b1 and other cyclins as tumor antigens in immunosurveillance and immunotherapy of cancer: Figure 1," *Cancer Res.*, vol. 66, no. 1, pp. 6–9, Jan. 2006.
[69] W. Li, J. Liu, W. Fu, X. Zheng, L. Ren, S. Liu, J. Wang, T. Ji, and G. Du, "3-O-acetyl-11-keto-β-boswellic acid exerts anti-tumor effects in glioblastoma by arresting cell cycle at G2/M phase," *J. Exp. Clin. Cancer Res.*, vol. 37, no. 1, p. 132, Dec. 2018.
[70] W.-W. Sung, Y.-M. Lin, P.-R. Wu, H.-H. Yen, H.-W. Lai, T.-C. Su, R.-H. Huang, C.-K. Wen, C.-Y. Chen, C.-J. Chen, and K.-T. Yeh, "High nuclear/cytoplasmic ratio of Cdk1 expression predicts poor prognosis in colorectal cancer patients," *BMC Cancer*, vol. 14, no. 1, pp. 1–7, Dec. 2014.

[71] C. Zhang, A. G. Elkahloun, M. Robertson, J. J. Gills, J. Tsurutani, J. H. Shih, J. Fukuoka, M. C. Hollander, C. C. Harris, W. D. Travis, J. Jen, and P. A. Dennis, "Loss of cytoplasmic CDK1 predicts poor survival in human lung cancer and confers chemotherapeutic resistance," *PLoS ONE*, vol. 6, no. 8, Aug. 2011, Art. no. e23849.
[72] X. Chen, F. Zhang, Q. Chen, Y. Wang, Y. Wang, J. He, and J. Zhou, "The clinical significance of CDK1 expression in oral squamous cell carcinoma," *Medicina Oral Patología Oral y Cirugia Bucal*, vol. 20, no. 1, p. e7, 2015.

MING GU currently pursuing the degree in basic medical science with the School of Basic Medical Science, Southern Medical University, China. His current research interests include computational biology, bioinformatics, systems biology, and oncology.

TIAN-FANG ZHAO (Student Member, IEEE) received the master's degree from the Dalian University of Technology, China, in 2017. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, China. Her current research interests include complex network theory, social data analytics, network communication, and collective intelligence.

WEI LIU currently pursuing the degree in english with the School of Foreign Languages, Central South University, China. His current interests include academic thesis writing, grammatical analysis, and contextual analysis.

XIAO-KUN WU received the Ph.D. degree from Shanghai University, China, in 2010. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, China. She is also a Professor with the School of Journalism and Communication, South China University of Technology. She has authored three academic books and more than 30 research articles in her research areas. Her research interests include online collective actions, communication models, and mass communications.

LIANG YANG received the B.E. and M.E. degrees in computational mathematics from Nankai University, Tianjin, China, and the Ph.D. degree in computer science from the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China. He is currently an Associate Professor with the School of Artificial Intelligence, Hebei University of Technology, Tianjin. His current research interests include community detection, semi-supervised learning, low-rank modeling, and deep learning.