Key genes involved with prognosis were identified in lung adenocarcinoma by integrated bioinformatics analysis

Hao Song [a,1], Junfeng Wu [a,1], Wang Liu [b,1], Kaier Cai [a], Zhilong Xie [a], Yingao Liu [a], Jiandi Huang [c], Siyuan Gan [c,**], Yinghuan Xiong [d,*], Yanqin Sun [c,***]

[a] *The Affiliated Hospital of Guangdong Medical University, Zhanjiang 524001, China*
[b] *Department of Respiratory, The Second Affilated Hospital of Guangdong Medical University, Zhanjiang 524001, China*
[c] *Department of Pathology, Guangdong Medical University, Dongguan 523808/Zhanjiang 524001, China*
[d] *Biological Sample Bank, The Affilated Hospital of Guangdong Medical University, Zhanjiang 524001, China*

ABSTRACT

Objective: By screening the core genes in lung adenocarcinoma (LUAD) with bioinformatics, our study evaluated its prognosis value and role in infiltration process of immune cells.
Methods: Using GEO database, we screened 5 gene chips, including GSE11072, GSE32863, GSE43458, GSE115002, and GSE116959. Then, we obtained the corresponding differentially expressed genes by analyzed 5 gene chips online by GEO2R (P < 0.05, |logFC| > 1). Then, through DAVID online platform, Cytoscape 3.6.1 software and PPI network analysis, the network was visualized and obtain the final core genes. Next, we plan to use the GEPIA, UALCAN, Kaplan–Meier plotter and Time 2.0 database for corresponding analysis. The GEPIA database was used to verify the expression of core genes in LUAD and normal lung tissues, and survival analysis was used to evaluate the value of core genes in the prognosis of LUAD patients. UALCAN was used to verify the expression of the LUAD core gene and promoter methylation status, and the predictive value of core genes was evaluated in LUAD patients by the Kaplan–Meier plotter online tool. Then, we used the Time 2.0 database to identify the relationship to immune infiltration in LUAD. Finally, we used the human protein atlas (HPA) database for online immunohistochemical analysis of the expressed proteins.
Results: The expression of CCNB2 and CDC20 in LUAD were higher than those in normal lung tissues, their increased expression was negatively correlated with the overall survival rate of LUAD, and they were involved in cell cycle signal transduction, oocyte meiosis signal transduction as well as the infiltration process of immune cells in LUAD. The expression proteins of CCNB2 and CDC20 were also different in lung cancer tissue and normal lung tissue. Therefore, CCNB2 and CDC20 were identified as the vital core genes.
Conclusion: CCNB2 and CDC20 are essential genes that may constitute prognostic biomarkers in LUAD, they also participate the immune infiltration process and protein expression process of LUAD, and might provides basis for clinical anti-tumor drug research.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: gansiyuan@gdmu.edu.cn (S. Gan), xyh821215@163.com (Y. Xiong), sunyanqin@gdmu.edu.cn (Y. Sun).
[1] These authors contributed equally to this work.

1. Introduction

Globally, lung cancer incidence and mortality rates are often the highest among malignant tumor types. LUAD, the most crucial pathological type, accounts for 40–50% of all lung cancers [3]. Therefore, exploring the carcinogenic mechanism of LUAD and studying prevention and treatment countermeasures are of great significance in improving people's health. Our urgent task is to

---

List of Abbreviations

LUAD	Lung Adenocarcinoma
GEO	Gene Expression Omnibus
DEGs	Differentially Expressed Genes
DAVID	Database for Annotation, Visualization and Integrated Discovery
PPI	Protein-Protein Interaction
MCODE	Molecular Complex Detection
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Gene and Genome
BPs	Biological Processes
MFs	Molecular Functions
CCs	Cell Components
STRING	Search Tool for the Retrieval of Interacting Genes
GEPIA	Gene Expression Profiling Interactive Analysis
LNM	Lymph Node Metastasis
HPA	Human Protein Atlas

---

provide active and effective prevention and treatment intervention measures for LUAD [4]. Many studies have found that the development of human tumors is related to the abnormal expression of various functional genes. Similarly, the occurrence and development of LUAD is a process of multi-gene regulation, with the participation of a variety of critical genes [5,6]. Therefore, we need to deeply mine many tumor-based databases and build a prediction model based on multiple genes to make it possible to understand the urgent prognosis direction of patients.

Bioinformatics is a subject that studies the collection, processing, storage, dissemination, analysis and interpretation of biological information. It reveals the biological laws endowed by a large amount of complex biological data through the comprehensive utilization of various subject technologies. This study used the above related technologies to screen the differentially expressed genes (DEGs) related to lung adenocarcinoma. After obtaining the original microarray dataset from the Gene Expression Omnibus (GEO) database and analyzing five groups of data with online tools, such as GEO2R, Venn and DAVID (database for annotation, visualization and integrated discovery), protein–protein interaction (PPI) analysis was performed to determine the core genes, which were then reintroduced into DAVID. After visualizing the network with Cytoscape 3.6.1 software, it was concluded that the final key core genes were cyclin B2 (CCNB2) and cell division cycle 20 (CDC20). For the obtained key core genes, we used GEPIA, UALCAN and Kaplan–Meier plots to verify the expression, prognosis and survival of the genes in LUAD patients to verify them as key core genes. We further analyzed the involvement of core genes in immune cell infiltration in LUAD through TIMER2.0 database, and HPA database (https://www.proteinatlas.org/) was used to observe the protein expression of core genes between lung cancer and normal tissue, with a view to providing reference value for targeted drug therapy of LUAD.

2. Methods and materials

2.1. Data acquisition

From the GEO database (http://www.ncbi.nlm.gov/gds/), 5 datasets (GSE10072, GSE32863, GSE43458, GSE115002 and GSE116959) were screened by searching "lung adenocarcinoma". The GSE10072, GSE32863, GSE43458, GSE115002 and GSE116959 datasets contained 49 normal tissues and 58 LUAD tissues, 58 normal tissues and 58 LUAD tissues, 30 normal tissues and 80 LUAD tissues, 52 normal tissues and 52 LUAD tissues, and 11 normal tissues and 57 LUAD tissues, respectively. The platforms of the five datasets were GPl6259, GPl6884, GPl6244, GPl13497 and GPl17077.

2.2. Data processing

Five data groups were analyzed online by GEO2R, and through screening, $P < 0.05$ and $|logFC| > 1$, and the corresponding differentially expressed genes (DEGs) of the 5 datasets were obtained. The obtained data were analyzed by the Venn online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) to obtain the intersection of differentially expressed genes of the five datasets.

2.3. GO and KEGG pathway enrichment analysis

Using the DAVID online platform (https://david.ncifcrf.gov/), the differentially expressed genes obtained above were analyzed by Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The accepted biological processes (BPs), molecular function (MFs), cell components (CCs), and tumor action-related pathways were visually analyzed, and $P < 0.05$, FDR <0.05 was selected as the inclusion criterion.

2.4. PPI network construction and screening of core genes

The differentially expressed genes were stored in the STRING online database (http://string-db.org/), and PPI network analysis was performed using Cytoscape 3.6.1 (http://www.cytoscape.org/). The software visualizes the network and then screens the genes on the MCODE (degree cutoff = 2, max depth = 100, k-core = 2, node score cutoff = 0.2) plug-in. Then, DAVID analysis was performed again to obtain the final core genes through gene screening.

2.5. Expression verification of core genes and survival analysis of LUAD patients

Excel 2016 software was used to generate a heatmap to visually analyze the gene expression of LUAD tissue and normal lung tissue. Then, the GEPIA database was used (http://gepia.cancer-pku.cn/index.html/) to verify the expression of core genes in LUAD and normal lung tissues and evaluate the value of core genes in the prognosis of LUAD patients with survival analysis. In addition, the expression of core genes and the methylation status of LUAD genes were verified using UALCAN. In addition, the Kaplan–Meier plotter online tool (http://kmplot.com/analysis/) was used to evaluate the prognostic value of core genes in LUAD patients.

2.6. Correlation analysis between expression of core genes and tumor immune cell infiltration

After obtaining the core gene, The TIMER 2.0 database (http://timers.comp-genomics.org) was used associated the core gene with immunogenic infiltrating cells common in LUAD, Thermal mapping was used to visually analyze the core gene and immune cell infiltration, and to validate the correlation between LUAD core gene expression and immune cell infiltration.

2.7. Online analysis of differences in protein expression of core genes between normal tissues and lung cancer

For further verification, the HPA database was used to verify the correlation between protein expression of core genes in normal and lung cancer tissues through observation of pathological sections and differences in protein expression of related core genes in normal lung tissues, LUAD and LUSC.

3. Results

3.1. Data identification of LUAD differentially expressed genes

We obtained 664, 1267, 899, 5829, and 2258 differentially expressed genes after five microarray screens of GSE10072, GSE32863, GSE43458, GSE115002, and GSE116959, respectively. A Venn diagram was drawn to obtain the intersection of the five datasets, and we obtained 188 differentially expressed genes (see Table 1), including 39 upregulated genes and 149 downregulated genes (see Fig. 1A and B).

3.2. GO and KEGG pathway analyses of LUAD DEGs

One hundred eighty-eight differentially expressed genes were analyzed by DAVID software. The GO analysis results ($P < 0.05$, see

Table 1
Differential gene intersection of five data sets obtained from Venn map.

| DEGS	Gene Name
Up-regulated	*TPX2 IGF2BP3 COL1A1 TMPRSS4 SPP1 TYMS MELK PLPP2 TIMP1 MMP12 SLC2A1 ECT2 MMP11 CRABP2 CCNB2 PRC1 CEACAM5 COMP TNFRSF21 SULF1 AGR2 SPINK1 HMGB3 DSP GOLM1 LGSN SFN CDC20 KDELR3 COL3A1 ASPM TCN1 SLC7A5 NQO1 TOP2A THBS2 COL10A1 CENPF MMP9*
Down-regulated	*ABCA3 EDN1 PAPSS2 A2M HIGD1B ANXA3 TEK GHR MAOA SLCO2A1 LAMP3 BCHE STARD13 ID1 SPTBN1 TSPAN7 ACADL AHNAK NDNF FHL1 KLF4 TCF21 NEDD4L SFTPD OLR1 AGER HOXA5 CD93 MSR1 WIF1 TPPP3 RAMP3 CLIC5 PROS1 MME ICAM2 PDK4 EFEMP1 WFS1 CACNA2D2 ADARB1 LRRC32 SASH1 SCGB1A1 CA4 PLA2G1B SESN1 LYVE1 P3H2 PHACTR2 FEZ1 CPA3 PID1 TBX3 PDZD2 ADIRF HBB GNG11 FLRT3 SLC39A8 SFTPC TGFBR3 LPL CAT SOSTDC1 GPC3 LMCD1 MARCO C4BPA DCN EPAS1 SLIT2 CX3CR1 PCOLCE2 HSD17B6 CYP4B1 SMAD6 SLC6A4 TMEM204 TIMP3 DACH1 EML1 HBEGF CRYAB CD36 PTPRB ADGRL2 AQP4 TNNC1 IL1RL1 ABCA8 AOC3 TMEM47 ZBTB16 DES EDNRB MYH10 MYH11 SLPI RGCC VWF VSIG4 HPGD WASF3 HEG1 ANGPT1 SEMA6A TMEM100 DUOX1 FERMT2 DPT MFAP4 PECAM1 CDO1 ANOS1 STXBP6 S1PR1 FOS LIMCH1 SPOCK2 LDB2 CALCRL RECK CAV1 JAM2 SPARCL1 CA2 METTL7A ITM2A ARHGAP44 FAM189A2 LRRN3 IL7R FMO2 IL33 CTNNAL1 PGC ADH1B FABP4 FCN3 CRTAC1 FAM107A GPM6B SEMA5A RAMP2 VIPR1 CLDN18 OLFML1 FBLN5*
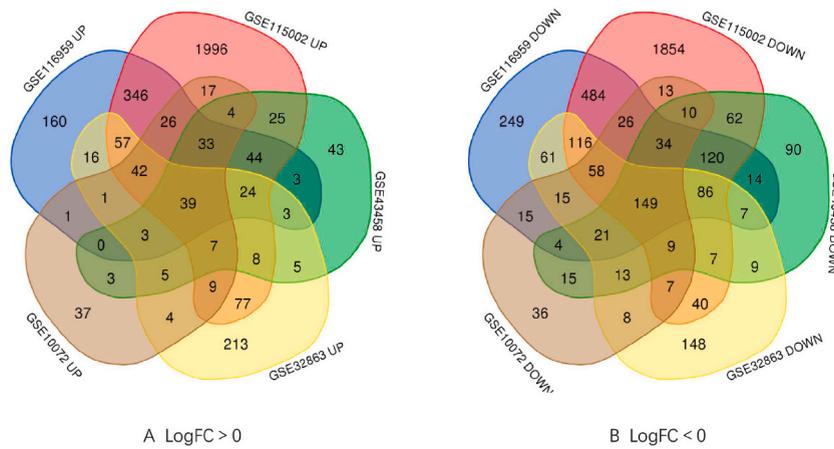
Fig. 1. Upregulated and downregulated genes of five datasets obtained from Venn diagram.

Table 2) are shown.

(1) In BPs, the upregulated DEGs were enriched in extracellular matrix organization, extracellular matrix disassembly, skeletal system development, skin development, collagen fibril organization, and extracellular matrix organization, and the down-regulated DEGs were mainly enriched in cell adhesion, angiogenesis, receptor internalization, cellular response to hormone stimulus, positive regulation of gene expression and response to glucocorticoid pathway.

(2) In CCs, the upregulated DEGs were particularly enriched in extracellular space, spindle, midbody, extracellular matrix, collagen trimer, and endoplasmic reticulum lumen, and the downregulated DEGs were particularly enriched in cell surface, extracellular space, plasma membrane, extracellular region, cell surface, integral component of plasma membrane and integral component of membrane.

(3) In MFs, the upregulated DEGs were particularly enriched in extracellular matrix structural constituent. In contrast, the downregulated DEGs were particularly enriched in beta-amyloid binding, adrenomedullin receptor activity, calcium ion binding, identical protein binding and protein binding.

Table 2
GO the analysis results of 188 differential genes by DAVID software ($P < 0.05$, FDR <0.05).

Expression	Category	Term	Count	%	P-Value	FDR
Up- regulated	GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	8	20.51282	6.45E-07	3.11E-04
	GOTERM_BP_DIRECT	GO:0030199~collagen fibril organization	5	12.82051	5.24E-05	0.011368
	GOTERM_BP_DIRECT	GO:0043588~skin development	4	10.25641	7.08E-05	0.011368
	GOTERM_BP_DIRECT	GO:0001501~skeletal system development	5	12.82051	1.35E-04	0.016296
	GOTERM_BP_DIRECT	GO:0022617~extracellular matrix disassembly	4	10.25641	3.07E-04	0.029591
	GOTERM_CC_DIRECT	GO:0031012~extracellular matrix	7	17.94872	7.76E-06	8.61E-04
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	14	35.89744	2.67E-05	0.001483
	GOTERM_CC_DIRECT	GO:0005819~spindle	5	12.82051	1.52E-04	0.005641
	GOTERM_CC_DIRECT	GO:0005788~endoplasmic reticulum lumen	6	15.38462	2.63E-04	0.007305
	GOTERM_CC_DIRECT	GO:0030496~midbody	5	12.82051	3.65E-04	0.008108
	GOTERM_CC_DIRECT	GO:0005581~collagen trimer	4	10.25641	7.13E-04	0.013187
	GOTERM_MF_DIRECT	extracellular matrix structural constituent	5	12.8	1.70E-04	2.20E-02
Down- regulated	GOTERM_BP_DIRECT	angiogenesis	13	8.7	4.20E-07	5.20E-04
	GOTERM_BP_DIRECT	response to glucocorticoid	8	5.4	6.80E-07	5.20E-04
	GOTERM_BP_DIRECT	cell adhesion	17	11.4	4.70E-06	2.40E-03
	GOTERM_BP_DIRECT	receptor internalization	6	4	2.50E-05	9.60E-03
	GOTERM_BP_DIRECT	positive regulation of gene expression	15	10.1	4.00E-05	1.20E-02
	GOTERM_BP_DIRECT	cellular response to hormone stimulus	5	3.4	6.60E-05	1.70E-02
	GOTERM_CC_DIRECT	cell surface	25	16.8	2.40E-11	5.90E-09
	GOTERM_CC_DIRECT	extracellular space	42	28.2	1.80E-10	2.20E-08
	GOTERM_CC_DIRECT	extracellular region	44	29.5	3.20E-10	2.50E-08
	GOTERM_CC_DIRECT	plasma membrane	70	47	4.30E-09	2.60E-07
	GOTERM_CC_DIRECT	integral component of plasma membrane	31	20.8	1.80E-07	8.80E-06
	GOTERM_CC_DIRECT	integral component of membrane	62	41.6	9.50E-05	2.90E-03
	GOTERM_MF_DIRECT	beta-amyloid binding	8	5.4	3.30E-06	1.20E-03
	GOTERM_MF_DIRECT	adrenomedullin receptor activity	3	2	1.70E-04	3.10E-03
	GOTERM_MF_DIRECT	identical protein binding	27	18.1	3.70E-04	4.30E-02
	GOTERM_MF_DIRECT	calcium ion binding	16	10.7	4.80E-04	4.30E-02
	GOTERM_MF_DIRECT	protein binding	113	75.8	5.80E-04	4.30E-02

According to KEGG analysis (P < 0.05, see Table 3), the upregulated DEGs were significantly enriched in ECM receptor interaction, adhesion, the PI3K Akt signaling pathway, protein digestion and absorption, and the cell cycle, while the downregulated DEGs were significantly enriched in the PPAR signaling pathway, complement, and coagulation cascade.

3.3. PPI and core gene analysis

Based on the above results, 158 differentially expressed genes (123 upregulated and 35 downregulated genes) were used to construct the PPI network. The differential gene PPI network had 158 nodes and 448 edges. However, 30 of the 188 differentially expressed genes were not included in the network (in Fig. 2A). Then, using the Cytoscape MCODE plug-in for further analysis, ten central nodes were identified as core genes in 158 nodes, including ten upregulated genes (CCNB2, ASPM, CDC20, PRC1, TOP2A, TPX2, ECT2, CENPF, TYMS, and MELK) (in Fig. 2B).

The KEGG pathway enrichment analysis of the 10 core genes identified by DAVID (P < 0.05, see Table 4) showed that they were mainly enriched in cell cycle signal transduction and oocyte meiosis signal transduction but not for human T-cell leukemia virus type 1 infection signal transduction (P > 0.05). Moreover, the genes CDC20 and CCNB2 both participate in cell cycle signal transduction and oocyte meiosis signal transduction, and the significance of their involvement in the cell cycle signal transduction pathway is greater (P < 0.05, see Fig. 3), so it was determined that the vital core genes are CDC20 and CCNB2.

3.4. Verification and visualization of core genes

Excel 2016 software was used to analyze two microarray datasets, GSE32863 and GSE115002, to verify the reliability and accuracy of the two core differentially expressed genes (CDC20 and CCNB2) in LUAD. The cluster heatmap showed that there were significant differences between the two core genes in the LUAD tissue group and the normal lung tissue group (see Fig. 4A and B).

3.5. GEPIA analysis of core genes

The two core genes, CDC20 and CCNB2, were put into the GEPIA database to verify their expression, and this database includes 483 LUAD tissues and 347 normal tissues. The results showed that the expression levels of CDC20 and CCNB2 in LUAD tissue samples were significantly higher than those in normal lung tissues (P < 0.01, see Fig. 5A and B). Survival analysis showed that the increased expression of CDC20 and CCNB2 was negatively correlated with the overall survival rate of LUAD (P < 0.05, see Fig. 6A and B).

3.6. UALAN analysis of core genes

UALAN online software was used to detect the expression of the two core genes in LUAD. The expression levels of CDC20 and CCNB2 in LUAD tissue were significantly higher than those in normal lung tissue (P < 0.01, see Fig. 7A and B). The expression levels of the two core genes in LUAD were evaluated according to lymph node metastasis (LNM) status. In the N0 phase, the expression levels of CDC20 and CCNB2 in LUAD tissue were higher than those in normal lung tissue (P < 0.01, see Fig. 8A and B). At the same time, in N1 phase, the expression levels of CDC20 and CCNB2 in LUAD tissue were higher than those in N0 phase LUAD tissue (P < 0.01, see Fig. 8A and B). Compared with normal lung tissue, the methylation levels of the CDC20 and CCNB2 promoters in LUAD tissue were lower (P < 0.01, see Fig. 9A and B).

Table 3
KEGG analysis results of 188 differential genes by DAVID software ($P < 0.05$).

| Expression	Pathway ID	Term	Count	%	P-Value	Genes
Up- regulated	hsa04512	ECM-receptor interaction	4	10.3	1.20E-03	COMP,COL1A1,SPP1,THBS2
	hsa04510	Focal adhesion	4	10.3	1.30E-02	COMP,COL1A1,SPP1,THBS2
	hsa05415	Diabetic cardiomyopathy	4	10.3	1.30E-02	COL1A1,COL3A1,MMP9,SLC2A1
	hsa04974	Protein digestion and absorption	3	7.7	2.60E-02	COL1A1,COL3A1,COL10A1
	hsa04110	Cell cycle	3	7.7	3.80E-02	CDC20,CCNB2,SFN
	hsa04926	Relaxin signaling pathway	3	7.7	4.00E-02	COL1A1,COL3A1,MMP9
	hsa05165	Human papillomavirus infection	4	10.3	4.60E-02	COMP,COL1A1,SPP1,THBS2
Down- regulated	hsa04270	Vascular smooth muscle contraction	7	4.7	2.90E-03	CALCRL,EDN1,MYH10,MYH11,PLA2G1B,RAMP2, RAMP3
	hsa03320	PPAR signaling pathway	5	3.4	8.10E-03	CD36,ACADL, FABP4,LPL,OLR1
	hsa04610	Complement and coagulation cascades	5	3.4	1.20E-02	VSIG4,A2M,C4BPA,PROS1,VWF
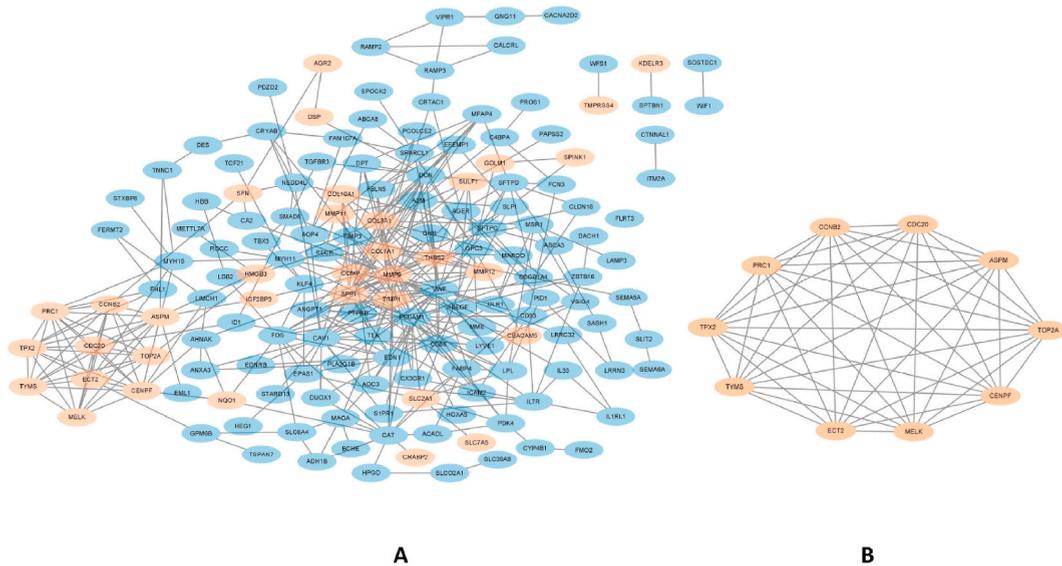
Fig. 2. Differences between the gene network of PPI analysis and ten genes of a central node.

Table 4
Results of KEGG pathway enrichment analysis of 10 core genes by David ($P < 0.05$).

Pathway ID	Name	Count	%	P-Value	Genes
hsa04110	Cell cycle	2	20	4.60E-02	CCNB2, CDC20
hsa04114	Oocyte meiosis	2	20	4.80E-02	CCNB2, CDC20

3.7. Kaplan–Meier analysis of core genes

Kaplan–Meier plotter is an online tool that can analyze the relationship between genes and the survival of tumor patients. We used this tool to evaluate the predictive value of core genes in LUAD patients. The results showed that the survival rate of patients with high expression of the CDC20 and CCNB2 genes was lower comparatively (P < 0.05, see Fig. 10A and B). The results showed that the expression levels of CDC20 and CCNB2 had a significant impact on the survival time of LUAD patients.

3.8. Correlation of the two core genes signatures with immune cell infiltration

We used TIMER 2.0 database to evaluate the immune cell infiltration in LUAD patients. The correlation of the two genes signatures with immune cell infiltration levels in LUAD. The two gene signatures were correlated with common lymphoid progenitor, macrophage M0/M1/monocyte cells, MDSC cell, CD4$^+$ memory T cell, CD4$^+$ Th1/Th2 cell levels. Infiltration levels were all significantly and positively correlated (P < 0.05, see Fig. 11).

3.9. Analysis of protein expression of two core genes

We used the HPA database to assess and validate differences in protein expression between core gene-normal lung tissue and lung adenocarcinoma and lung squamous cell carcinoma. Pathological sections showed that the two core genes had certain differences in protein expression in normal lung tissues, lung adenocarcinoma and lung squamous cell carcinoma tissues (see Fig. 12), further verifying that our core genes play a certain role in lung adenocarcinoma tissues.

4. Discussion

At present, although targeted therapy has achieved fruitful results in the treatment of lung cancer, it is still necessary for us to study the therapeutic targets of drugs. Therefore, we search for key core genes through the analysis of biological information [7]. Through bioinformatics analysis, we screened 39 upregulated DEGs and 149 downregulated DEGs from 200 normal lung tissues and 307 LUAD samples by analyzing five datasets, GSE11072, GSE32863, GSE43458, GSE115002, and GSE116959. Then, through DAVID's GO analysis, KEGG analysis, the Cytoscape MCODE plug-in, and other comprehensive bioinformatics analyses, it was determined that the vital core genes are CDC20 and CCNB2. Then, the cluster heatmap showed that there were significant differences in the expression levels of the two core genes between the LUAD tissue group and normal lung tissue group. At the same time, GEPIA data analysis
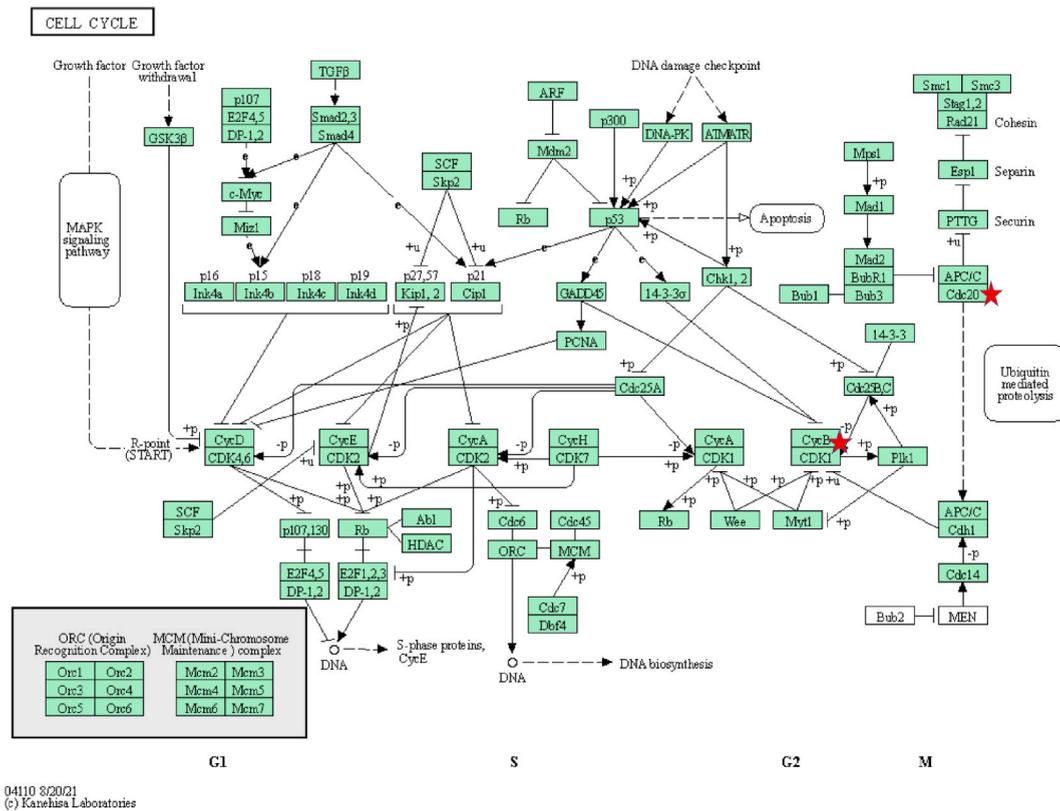
Fig. 3. KEGG pathway enrichment analysis of CDC20 and CCNB2 in the cell cycle (P < 0.05).

confirmed that the expression levels of CDC20 and CCNB2 in LUAD tissue were significantly higher than those in normal lung tissue, and their high expression was negatively correlated with the prognosis of patients. In UALAN, the expression levels of CDC20 and CCNB2 in LUAD were different from those in adjacent normal lung tissues, and the high expression of CDC20 and CCNB2 was related to lymph node metastasis in LUAD tumor tissues. Kaplan–Meier analysis showed that the expression levels of the CDC20 and CCNB2 genes had a great impact on the prognosis of patients. Finally, TIMER2.0 database and the HPA database were used to further observe the immune cell infiltration level and protein expression level of core genes in lung adenocarcinoma tissues, which further verified the role of core genes in lung cancer.

P53 is an essential regulatory molecule in the cell cycle pathway. Responses to various pressures induce apoptosis, DNA repair, and cell cycle arrest and reduce mammalian tumor cell formation [8]. As a tumor suppressor protein, P53 plays an irreplaceable role in the occurrence and development of tumors. Studies have shown that cyclin B1 and B2 mRNA levels decrease after induction of wild-type P53 but do not decrease after induction of P53 DNA binding-deficient mutants, indicating that the expression of cyclin B can be inhibited through P53-dependent transcription during cell damage [9].

CCNB2 is a class B cyclin family member. It is a protein expressed, involved, and decomposed in cell cycle specificity or phase. Its abnormal expression can affect the cell cycle and cell proliferation. In the cell cycle pathway, CCNB2 stably binds to cyclin-dependent kinase 1 (CDK1), and once bound, CDK1 acts as a regulatory subunit and plays a specific role in regulating the cell cycle through its phosphorylation and dephosphorylation. The complex formed by CCNB2 and CDK1 can orderly control the G2/M transition in the cell cycle and promote the progression of mitosis [10]. Relevant experiments have proven that KPNA2 is a new marker for predicting the prognosis of patients in liver cancer tissues, and knockdown of KPNA2 can downregulate CCNB2 and CDK1, inhibit cell proliferation and induce cell cycle arrest at the G2/M phase [11,12]. Many studies at home and abroad have shown that the overexpression of CCNB2 can be detected in various human tumors, including lung cancer, and is closely related to the poor prognosis of the disease [13–18], which confirms the conclusion of our bioinformatics analysis.

Cell division cycle 20 (CDC20) also plays a vital role in the occurrence and development of tumors. CDC20 is also a cell cycle-related protein and an essential protein in the regulation of cell division activities [19]. In mitosis, CDC20 binds and activates the ubiquitin ligase activity of a macromolecular machine called the anaphase-promoting complex/loop (APC/C), which degrades securin and cyclin B, thus promoting the progression of anaphase and mitosis [20]. It was found that knockdown of CDC20 has a significant effect on tumor proliferation and treatment response. For example, after the expression of CDC20 was decreased, the growth of prostate cancer decreased significantly. Meanwhile, for the treatment of prostate cancer, a study found that the combination of siCDC20 with docetaxel achieved a better anticancer effect than docetaxel alone, indicating that the knockdown of CDC20 may also enhance the anticancer effect of docetaxel [21]. In cholangiocarcinoma, the CDC20 inhibitor dinaciclib has better antitumor activity

A   GSE32863
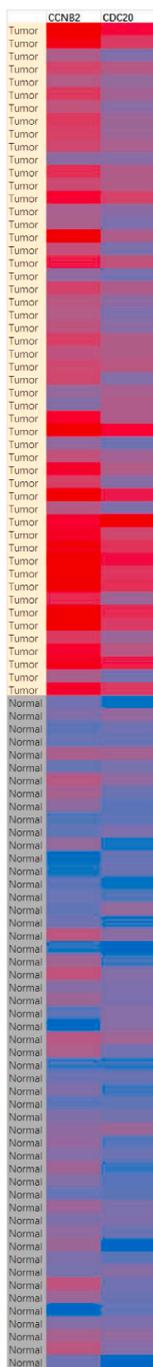
B   GSE115002

(caption on next page)

Fig. 4. In GSE32863 and GSE115002, the core differential genes CDC20 and CCNB1 of LUAD tissue and normal lung tissue were screened and verified by Excel 2016 software, and their reliability was confirmed by cluster heatmap (earthy yellow: LUAD tissue, gray: normal lung tissue) (red: high expression, blue: low expression). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
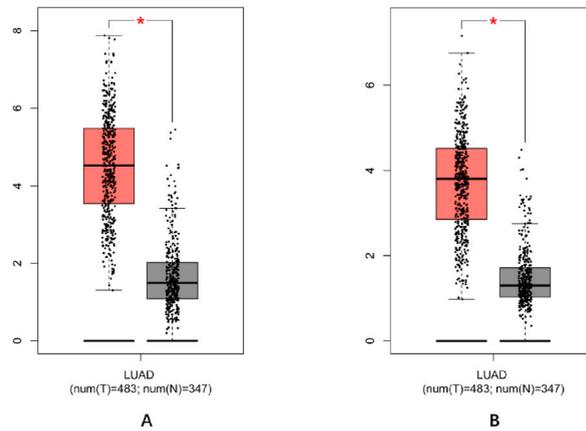
Fig. 5. Expression of CDC20 and CCNB2 in normal lung and LUAD tissues (Fig. A: CDC20, Fig. B: CCNB2).
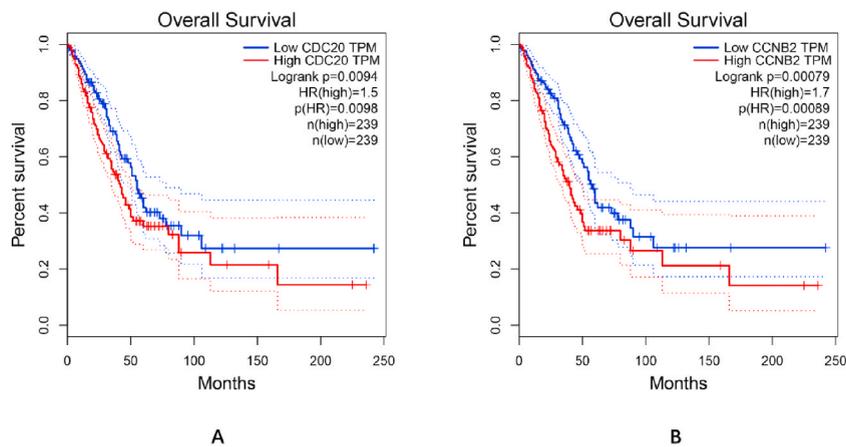
Fig. 6. Relationship between increased expression of CDC20 and CCNB2 and survival rate of lung cancer (Fig. A: CDC20, Fig. B: CCNB2).
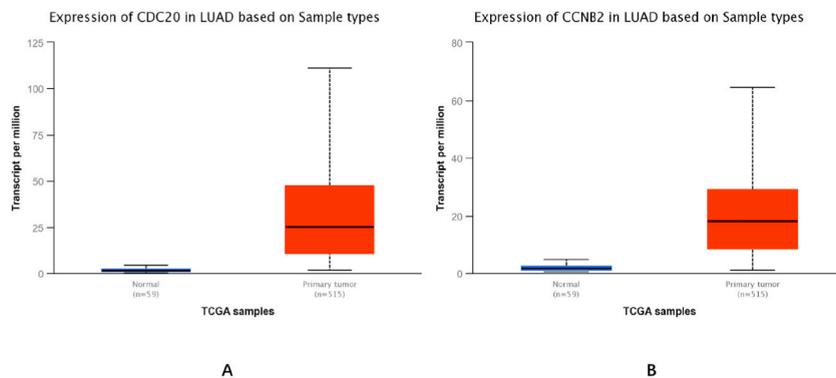

Fig. 7. Expression of CDC20 and CCNB2 in normal lung and LUAD tissues analyzed by UALCAN (Fig. A: CDC20, Fig. B: CCNB2) (P < 0.01).
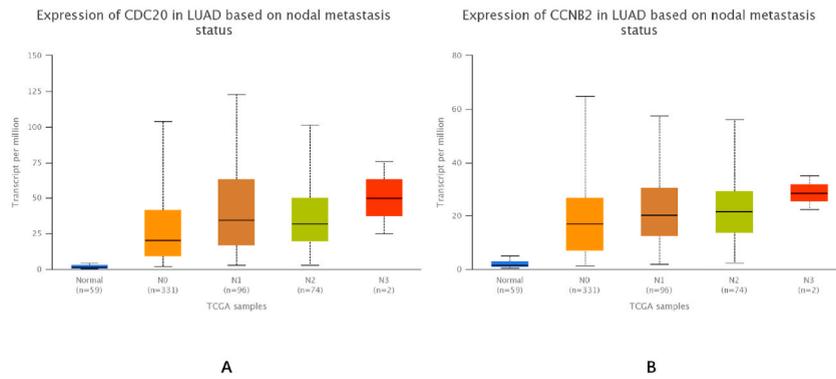
Fig. 8. Expression of the CDC20 and CCNB2 genes and LNM status of LUAD (Fig. A: CDC20, Fig. B: CCNB2) (P < 0.01).
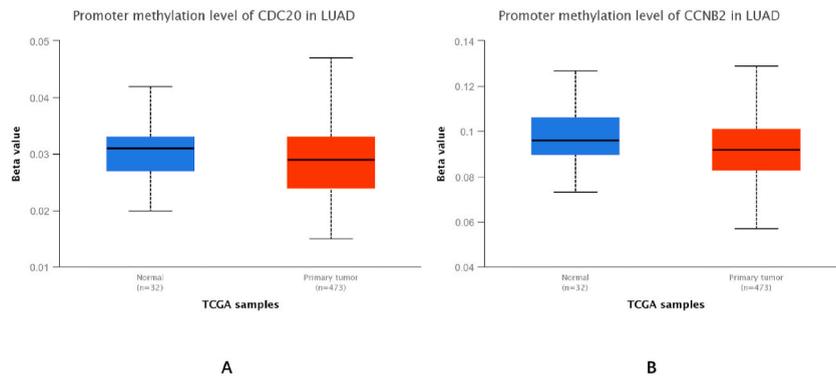

Fig. 9. Levels of CDC20, CCNB2, and promoter methylation in normal lung and LUAD tissues (Fig. A: CDC20, Fig. B: CCNB2) (P < 0.01).
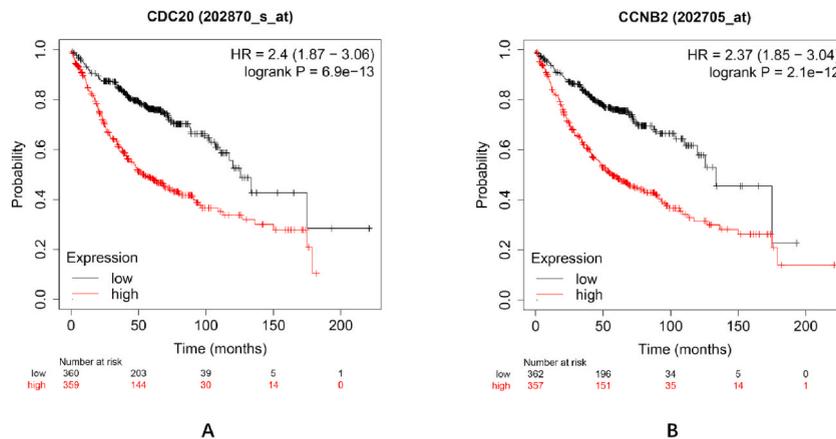

Fig. 10. Kaplan–Meier analysis shows the prognostic value of CDC20 and CCNB2 in LUAD patients (Fig. A: CDC20, Fig. B: CCNB2).

than the standard chemotherapy drug (gemcitabine) [22]. Therefore, for the treatment of LUAD, knocking down the expression of CDC20 may be a good direction for medication development, but more clinical data are needed for verification.

Abnormal expression of CDC20 can affect the operation of the cell cycle and lead to mitotic errors, which can promote the proliferation of tumor cells and inhibit the apoptosis of tumor cells [23]. Studies have shown that CDC20 is overexpressed in skin squamous cell carcinoma tissues and cell lines. Knockdown of CDC20 can inhibit cell growth and migration and promote cell cycle arrest and cell apoptosis in skin squamous cell carcinoma [24]. Abnormal expression of CDC20 may also play a role in the occurrence and progression of pancreatic ductal adenocarcinoma and is a specific direction of study for disease progression, prognosis, and therapeutic targets [25]. This shows that CDC20 has an essential effect on many tumor tissues and cells. Recent studies have shown that compared with normal lung tissues, CDC20 is highly expressed in LUAD tissues, and its high expression is significantly related to the
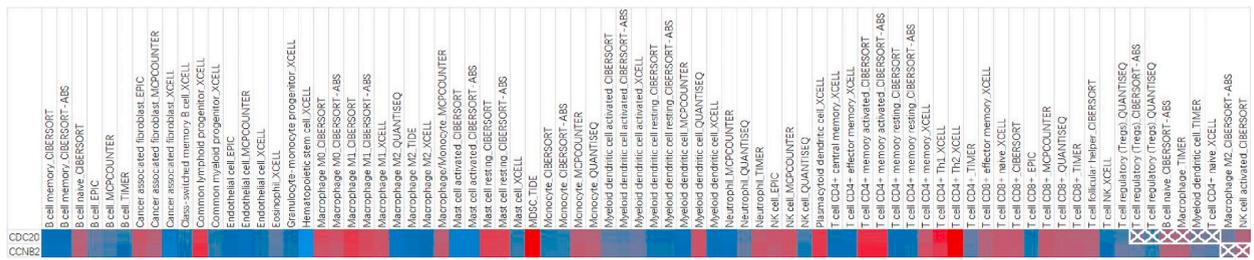
Fig. 11. TIMER 2.0 database shows the relationship between the expression of CDC20 and CCNB2 and the infiltration of immune cells in LUAD (P < 0.05).
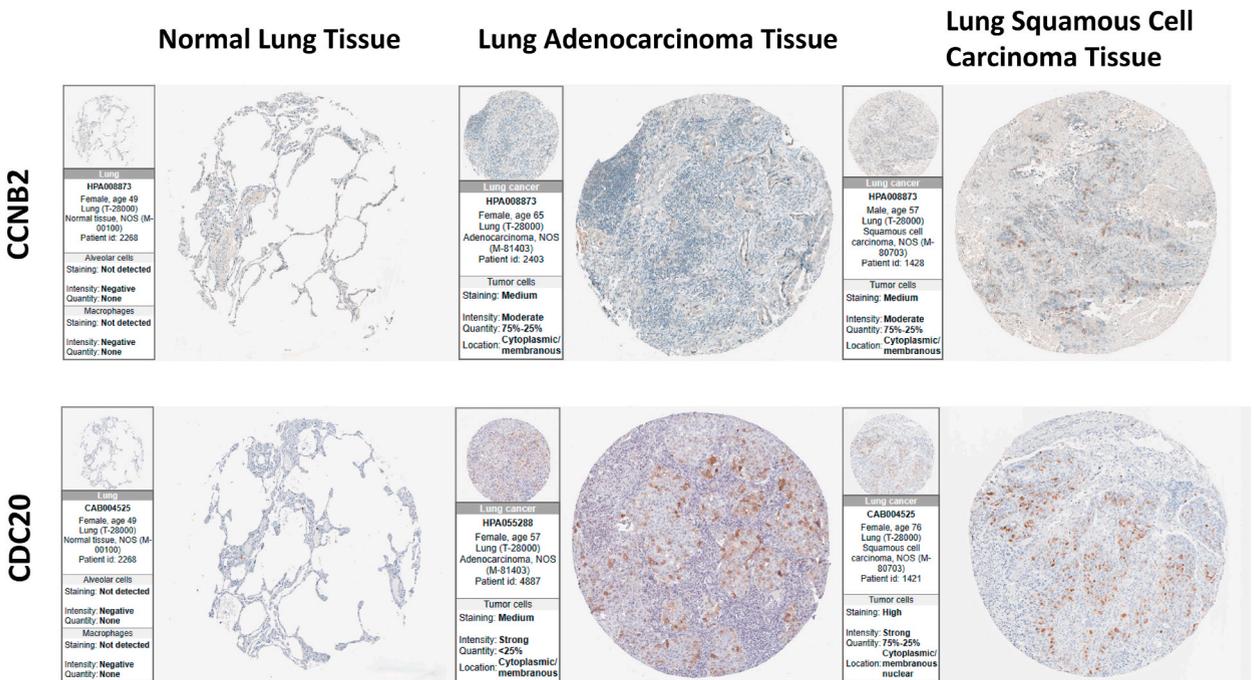

Fig. 12. The HPA database shows the protein expression of CCNB2 and CDC20 is different in normal tissues and lung cancer tissues.

poor prognosis of LUAD. CDC20 has been proposed as a new target for LUAD treatment and for prognosis evaluation of LUAD and may be a promising molecular target for chemotherapy [26].

At the same time, based on some current biological information analysis, in some studies of non-small cell lung cancer, CDC 20 and CCNB2 are related to the survival rate of patients to a certain extent. Meanwhile, in cell experiments, the expression level of CCNB2 is also high in lung cancer [27–29]. In addition, as a heterogeneous disease. Based on TCGA and GEO databases, CDC20 and CCNB2 may be the pathogenic genes of young-onset non-small cell lung cancer [30]. More importantly, based on genome-wide association analysis of GWAS, many key genes associated with lung adenocarcinoma were identified through co-expression networks, including CCNB2, which plays a key role in the cell cycle [31]. The above results further confirmed that our core genes play a certain role in the occurrence, development and prognosis of lung adenocarcinoma, and are expected to be targeted therapeutic targets for lung cancer in the future.

It is known that immune cell infiltration in tumors is closely related to the survival of patients [32]. We analyzed the correlation between CCNB2 and CDC20 gene expression and immune cell infiltration levels using TIMER2.0 data and found that the two gene signatures were significantly and positively correlated with several immune cell infiltration levels. This finding suggested that these gene signatures may influence LUAD tumor progression by regulating the tumor micro-environment.

5. Conclusion

In conclusion, we screened the core genes CCNB2 and CDC20 that play an essential role in LUAD progression through comprehensive bioinformatics methods, and CCNB2 and CDC20 are related to the poor prognosis of LUAD. They can be used as prognostic biomarkers for LUAD patients and may become new therapeutic targets for LUAD. However, more research and clinical data are

needed to verify the effectiveness of these predictors.

Funding

Guangdong Basic and Applied Basic Research Foundation (2021B1515140067, 2022A1515012190), Guangdong Medical Science and Technology Research Foundation (A2022019), National Innovation and Entrepreneurship Training Program for College Students (202110571008, 202110571011, 202210571036), Guangdong Medical University Research Fund (GDMUM201818, GDMUM2019036), Zhanjiang City Science and Technology Research Project (2021B01043), "Clinical medicine $+$" cooperation project of Affiliated Hospital of Guangdong Medical University, Zhanjiang Science and Technology Development Special Fund Competitive Allocation Project (2021A05107, 2022A01018).

Ethical approval

Not applicable.

Consent to Participate

All authors contributed in this project.

Consent for publication

All authors are agreed.

Author contribution statement

Hao Song: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper. Kaier Cai: Junfeng Wu: Analyzed and interpreted the data. Yingao Liu: Wang Liu: Performed the experiments. Jiandi Huang: Conceived and designed the experiments. Siyuan Gan: Conceived and designed the experiments; Analyzed and interpreted the data. Yinghuan Xiong: Zhilong Xie: Contributed reagents, materials, analysis tools or data. Yanqin Sun: Analyzed and interpreted the data; Wrote the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[3] L.A. Torre, R.L. Siegel, A.L.A. Jem, Lung cancer statistics, Adv. Exp. Med. Biol. 893 (2016) 1–19.

[4] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020:GLOBOCAN estimates of incidence and mortality worldwidefor 36 cancers in 185 countries, Ca - Cancer J. Clin. 71 (3) (2021) 209–249.

[5] E.N. Kontomanolis, A. Koutras, A. Syllaios, D. Schizas, A. Mastoraki, N. Garmpis, M. Diakosavvas, K. Angelou, G. Tsatsaris, A. Pagkalos, T. Ntounis, Z. Fasoulakis, Role of oncogenes and tumor-suppressor genes in carcinogenesis: a review, Anticancer Res. 40 (11) (2020 Nov) 6009–6015.

[6] F. Li, C. Huang, L. Qiu, P. Li, J. Shi, G. Zhang, Comprehensive analysis of immune-related metabolic genes in lung adenocarcinoma, Front. Endocrinol. 13 (2022 Jul 8), 894754.

[7] F. Wu, L. Wang, C. Zhou, Lung cancer in China: current and prospect, Curr. Opin. Oncol. 33 (1) (2021 Jan) 40–46.

[8] W.R. Taylor, G.R. Stark, Regulation of the G2/M transition by p53, Oncogene 20 (15) (2001) 1803–1815.

[9] K. Krause, M. Wasner, W. Reinhard, et al., The tumour suppressor protein p53 can repress transcription of cyclin B, Nucleic Acids Res. 28 (22) (2000) 4410–4418.

[10] Y. Zou, S. Ruan, L. Jin, et al., CDK1, CCNB1, and CCNB2 are prognostic biomarkers and correlated with immune infiltration in hepatocellular carcinoma, Med. Sci. Monit. 26 (2020), e925289.

[11] K. Vermeulen, D.R. Van Bockstaele, Z.N. Berneman, The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer, Cell Prolif. 36 (3) (2003) 131–149.

[12] C.L. Gao, G.W. Wang, G.Q. Yang, et al., Karyopherin subunit-α 2 expression accelerates cell cycle progression by upregulating CCNB2 and CDK1 in hepatocellular carcinoma, Oncol. Lett. 15 (3) (2018) 2815–2820.

[13] S. Wu, R. Su, H. Jia, Cyclin B2 (CCNB2) stimulates the proliferation of triple-negative breast cancer (TNBC) cells in vitro and in vivo, Dis. Markers 2021 (2021), 5511041.

[14] R. Li, X. Jiang, Y. Zhang, et al., Cyclin B2 overexpression in human hepatocellular carcinoma is associated with poor prognosis, Arch. Med. Res. 50 (1) (2019) 10–17.

[15] M.S. Wu, Q.Y. Ma, D.D. Liu, et al., CDC20 and its downstream genes: potential prognosis factors of osteosarcoma, Int. J. Clin. Oncol. 24 (11) (2019) 1479–1489.

[16] D. Stav, I. Bar, J. Sandbank, Usefulness of CDK5RAP3, CCNB2, and RAGE genes for the diagnosis of lung adenocarcinoma, Int. J. Biol. Markers 22 (2) (2007) 108–113.

[17] Jie Wang, Da-xing Chen, Ning Guo, et al., Cyclin B2 overexpression is a poor prognostic biomarker in NSCLC patients, China J. Mod. Med. 27 (17) (2017) 45–49.

[18] X. Wang, H. Xiao, D. Wu, D. Zhang, Z. Zhang, miR-335-5p regulates cell cycle and metastasis in lung adenocarcinoma by targeting CCNB2, OncoTargets Ther. 13 (2020 Jun 30) 6255–6263.

[19] K. Li, Y. Mao, L. Lu, et al., Silencing of CDC20 suppresses metastatic castration-resistant prostate cancer growth and enhances chemosensitivity to docetaxel, Int. J. Oncol. 49 (4) (2016) 1679–1685.

[20] I. Yuan, I. Leontiou, P. Amin, et al., Generation of a spindle checkpoint arrest from synthetic signaling assemblies, Curr. Biol. 27 (1) (2017) 137–143.

[21] P. Sungwan, W. Lert-Itthiporn, A. Silsirivanit, et al., Bioinformatics analysis identified CDC20 as a potential drug target for cholangiocarcinoma, PeerJ 9 (2021), e11067.

[22] Y. Kim, J.W. Choi, J.H. Lee, et al., MAD2 and CDC20 are upregulated in high-grade squamous intraepithelial lesions and squamous cell carcinomas of the uterine cervix, Int. J. Gynecol. Pathol. 33 (5) (2014) 517–523.

[23] Z. Chu, X. Zhang, Q. Li, et al., CDC20 contributes to the development of human cutaneous squamous cell carcinoma through the Wnt/β-catenin signaling pathway, Int. J. Oncol. 54 (5) (2019) 1534–1544.

[24] D.Z. Chang, Y. Ma, B. Ji, et al., Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression, J. Hematol. Oncol. 5 (2012 Apr 4) 15.

[25] W.T. Liu, Y. Wang, J. Zhang, et al., A novel strategy of integrated microarray analysis identifies CENPA, CDK1 and CDC20 as a cluster of diagnostic biomarkers in lung adenocarcinoma, Cancer Lett. 425 (2018 Jul 1) 43–53.

[26] Y. Gao, B. Zhang, Y. Wang, et al., Cdc20 inhibitor apcin inhibits the growth and invasion of osteosarcoma cells, Oncol. Rep. 40 (2) (2018) 841–848.

[27] K. Gong, H. Zhou, H. Liu, T. Xie, Y. Luo, H. Guo, J. Chen, Z. Tan, Y. Yang, L. Xie, Identification and integrate analysis of key biomarkers for diagnosis and prognosis of non-small cell lung cancer based on bioinformatics analysis, Technol. Cancer Res. Treat. 20 (2021 Jan-Dec), 15330338211060202.

[28] Y. Zeng, N. Li, R. Chen, W. Liu, T. Chen, J. Zhu, M. Zeng, J. Cheng, J. Huang, Screening of hub genes associated with prognosis in non-small cell lung cancer by integrated bioinformatics analysis, Transl. Cancer Res. 9 (11) (2020 Nov) 7149–7164.

[29] Y. Shen, X. Tang, X. Zhou, Y. Yi, Y. Qiu, J. Xu, X. Tian, Screening of key prognosis genes of lung adenocarcinoma based on expression analysis on TCGA database, J Oncol 2022 (2022 Dec 26), 4435092.

[30] X. Han, P. Ren, S. Ma, Bioinformatics analysis reveals three key genes and four survival genes associated with youth-onset NSCLC, Open Med. 17 (1) (2022 Jul 6) 1123–1133.

[31] G. Bidkhori, Z. Narimani, S. Hosseini Ashtiani, A. Moeini, A. Nowzari-Dalini, A. Masoudi-Nejad, Reconstruction of an integrated genome-scale co-expression network reveals key modules involved in lung adenocarcinoma, PLoS One 8 (7) (2013 Jul 11), e67552.

[32] X.S. Liu, L.M. Zhou, L.L. Yuan, Y. Gao, X.Y. Kui, X.Y. Liu, Z.J. Pei, NPM1 is a prognostic biomarker involved in immune infiltration of lung adenocarcinoma and associated with m6A modification and glycolysis, Front. Immunol. 12 (2021 Jul 16), 724741. Erratum in: Front Immunol. 2021 Aug 31;12:751004. PMID: 34335635; PMCID: PMC8324208. Statements And Declarations Data Availability All databases (GEO, DAVID, Venn diagram, STRING, GEPIA database, TIMER 2.0 database, the HPA database) are freely available on the web. The Kaplan–Meier plotter also helped us assess the core genes CCNB2 and CDC20 that play an essential role in LUAD progression.