# *POLE2* Serves as a Prognostic Biomarker and Is Associated with Immune Infiltration in Squamous Cell Lung Cancer

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

B 1 **Zhen Wu***
CDF 2 **Yue-Ming Wang***
CDF 1 **Yu Dai**
A 1,2 **Liang-An Chen**

1 Respiratory Department, Chinese People's Liberation Army (PLA) General Hospital, Beijing, P.R. China
2 School of Medicine, Nankai University, Beijing, P.R. China

Corresponding Author:
Source of support:

* Zhen Wu and Yue-Ming Wang contributed equally to this work
Liang-An Chen, e-mail: chenliangan301@163.com
Departmental sources

Background:
Squamous cell lung cancer is the main cause of cancer-associated mortality. The discovery of promising prognostic biomarkers for predicting the survival of patients with squamous cell lung cancer remains a challenge.

Material/Methods:
Gene expression profiles of GSE33479 and GSE51855, including 42 squamous cell lung cancer tissues and 17 normal tissues, from the GEO database were assessed to find common differentially expressed genes (DEGs) via the GEO2R online tool and Venn diagram software. Then, gene ontology (GO) and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analyses were conducted. The key protein-protein interaction (PPI) network within those common DEGs was subsequently illustrated through a combination of Search Tool for Retrieval of Interacting Genes (STRING) and Cytoscape software. Finally, core genes associated with survival and levels of immune infiltration were demonstrated by the Kaplan-Meier plotter and Tumor Immune Estimation Resource (TIMER) online database, respectively.

Results:
In total, 483 DEGs were involved, including 216 upregulated genes enriched in "cell division", "DNA replication", and "DNA repair pathway" and 267 downregulated genes enriched in "cell adhesion", "oxidation-reduction process", and "cell-cell signaling". The 75 core genes were selected by Molecular Complex Detection applied in Cytoscape. Four genes – MND1, FOXM1, CDC6, and POLE2 – were found to be significantly associated with survival. Further analysis of the KEEG pathway and TIMER database revealed that only POLE2 was enriched in "DNA replication" and its higher expression was negatively associated with survival and immune infiltration.

Conclusions:
Higher expression of POLE2 is a prognosis-related biomarker for worse survival and is negatively associated with immune infiltration in squamous cell lung cancer.

MeSH Keywords:
**Biological Markers • Carcinoma, Non-Small-Cell Lung • Prognosis**

Full-text PDF:
https://www.medscimonit.com/abstract/index/idArt/921430

📄 2641   ▦ 4   📊 6   📖 41

## Background

Lung cancer is the main cause of cancer-associated mortality worldwide, and 85% of cases are non-small cell lung cancer (NSCLC) [1,2]. With the development of molecular-targeted therapy, overall survival time and prognosis of NSCLC have increased dramatically [3,4]. As a main component of NSCLC, patients with squamous cell lung cancer have relatively worse overall survival compared to those with lung adenocarcinoma due to clinicopathological differences, such as more patients with a smoking history and fewer patients with driver mutation [5,6]. The therapeutic value of immunotherapy in squamous cell lung cancer is crucial to improve prognosis [7,8]; however, reliable biomarkers for prediction of squamous cell lung cancer prognosis are lacking. Therefore, further identification of promising biomarkers for prognosis is required to improve the treatment efficacy and determine the underlying mechanism in squamous cell lung cancer.

Gene chip technology has been used to detect differentially expressed genes (DEGs) in different groups of patients for decades and has proven to be a reliable diagnostic and prognostic tool for patients with cancer [9–11]. This independent data, which is stored in public databases, has allowed researchers to examine potential mechanisms for diagnosis and therapy. For example, using data from The Cancer Genome Atlas (TCGA), IGF2BP1 has been demonstrated to be more commonly mutated in adenocarcinoma compared with squamous cell lung cancer [12–14]. Recent bioinformatics studies on lung cancer have determined the underlying mechanisms [15–17]. For example, it has been reported that gene alterations were different in lung adenocarcinoma *vs.* squamous cell lung cancer patients, as PPP3CA, DOT1L, and FTSJD1 were found to be altered in lung adenocarcinoma, while RASA1 was indicated to be mutated. However, these studies mainly focused on the different genomic profiles between lung cancer and healthy controls or between lung adenocarcinoma and squamous cell lung cancer. The biomarkers for prognosis of squamous cell lung cancer require further investigation [18–20].

We selected GSE33479 and GSE51855 from the Gene Expression Omnibus (GEO) database. The serial numbers starting with GSE are the accession numbers of the GEO database. GEO2R was subsequently applied to analyze the DEGs in these 2 datasets. Venn diagram software was then used to collect the common DEGs in GSE33479 and GSE51855. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) was used to evaluate these common DEGs in KEGG pathways. Subsequently, protein-protein interaction (PPI) network and Cytotype Molecular Complex Detection (MCODE) were used for analysis of the key DEGs and for identifying core genes. The Kaplan-Meier plotter database was used to identify key genes with significant prognostic information, and the Gene

Expression Profiling Interactive Analysis (GEPIA) was used to validate the key differences in gene expression between squamous cell lung cancer tissues and normal lung tissues. As a result, only 4 DEGs were qualified to be associated with patient survival. KEGG pathway enrichment was reanalyzed in these 4 genes, and only CDC6 and POLE2 were found to be enriched in cell cycle and pyrimidine metabolism. Finally, the association of CDC6 and POLE2 with infiltration of immune cells in squamous cell lung cancer microenvironments was analyzed using TIMER. We found that only POLE2 was associated with immune infiltrates and prognosis in squamous cell lung cancer. In conclusion, POLE2 is a prognosis-related biomarker that is associated with infiltration of immune cells in squamous cell lung cancer.

## Material and Methods

### Microarray data analysis

NCBI-GEO is a public database providing open and free microarray gene profiles. Gene expression profiles of GSE33479 and GSE51855 in squamous cell lung cancer and normal lung tissues were downloaded from the GEO database. Microarray data of GSE33479 and GSE51855 were all assessed in GP6480 Platforms, which consisted of 14 squamous cell lung cancer tissues and 13 normal lung tissues and 28 squamous cell lung cancer tissues and 4 normal lung tissues, respectively.

### DEGs analysis

DEGs between squamous cell lung cancer specimens and normal lung specimens were analyzed using online GEO2R tools [21] with adjusted P<0.05 and |logFC| >2. DEGs with log FC >0 were viewed as upregulated genes, while those with log FC <0 were regarded as downregulated genes. The common DEGs among these 2 datasets were subsequently identified with the Venn software online tool.

### Analysis of gene ontology (GO) and pathway enrichment

The unique biological significance of the high-throughput transcriptome was examined via GO analysis [22], with biological process, cellular component, and molecular function involved, and biological pathways included together, using the online bioinformatics tool DAVID [23] (P<0.05). The KEGG database was also used to identify the biological pathways and diseases [24].

### Analysis of PPI network and module

The search tool STRING was used to evaluate PPI information for the network of common DEGs [25]. Cytoscape was subsequently used to analyze any potential correlations between

**Table 1.** All 483 commonly differentially expressed genes (DEGs) were collected from 3 profile datasets, including 216 upregulated genes and 267 downregulated genes in the squamous cell lung cancer tissues compared to normal tissues.

| DEGs | Genes names |
|---|---|
| Upregulated | KRT16P2, ARHGAP11A, DEPDC1B, IGF2BP3, HOXD11, HRG, HOXC13, PKP1, NETO2, CALB1, CXCL13, HOXA13, LY6K, GPNMB, AKR1B10, GINS1, C16orf59, DUXAP8, ADAMDEC1, LOC399815, ANLN, BIRC5, CHRNA5, GJB6, KLK8, FOXM1, SPRR2D, CDK1, HIST1H3B, CHEK1, NEIL3, KIF11, PTHLH, KIF18B, ADAM23, AIM1L, POU4F1, CDC6, SAPCD2, CENPU, DSC2, MASTL, SMC4, KIF14, KIAA1549L, MAD2L1, NMRAL1P1, ARNTL2, CSTA, CARD14, PITX2, DEFB4A, RGS20, CENPL, ANKRD22, LINC01296, HOXA10, ARHGEF39, SPP1, SMCO2, CA9, LYPD3, MCM10, FAM83C, PPP2R2C, CASC9, FAM72D, TICRR, SOST, CENPA, TYMS, CDCA7, GRHL3, MELK, SFN, IL1RAP, ERC2, PRDM13, CENPN, IL36G, ZWINT, NDC80, SLC7A11, SGO1, GREM1, SPRR1A, PGLYRP4, KRT34, OIP5, CCNA2, IGFL2, ABCA12, BUB1, POLQ, NUF2, PBK, COCH, KRT6A, GNGT1, SKA1, MMP1, MMP12, CDCA5, UBE2T, CKS2, KIF23, DEPDC1, MMP11, FAM83A, CPA6, PLPP4, SKA3, TK1, DIAPH3, HOXC13-AS, TMEM158, HOXA11-AS, SPRR2C, MMP3, DSC3, CENPE, ZIC5, ASPM, PRSS2, ANXA10, KNL1, ATAD2, BRIP1, MND1, HOXD10, CCNB2, KRT6C, POLE2, MELTF, PIF1, CDT1, LRP8, FOXL2, POU6F2, CEP55, LINC01615, AUNIP, RRM2, KRT16, TOP2A, ERO1A, WDR72, ABCC5, FANCI, PLAC1, RAD54L, CCNE2, CDC45, S100A2, C6orf10, SPC25, CNTNAP2, KIF18A, WISP1, KIF15, KRT16P3, DSG3, BUB1B, LINC00659, DLGAP5, RAD51AP1, ZFP64, ESCO2, KRTAP4-1, IL36RN, MKI67, CLSPN, DTL, PTTG3P, ERVMER34-1, CDCA2, RAD51, AKR1C1, HMMR, HIST1H3D, SPRR3, CCDC150, C12orf56, CENPI, FBXO45, ERCC6L, JAKMIP3, AIM2, ORC6, S100A7, HOXC8, COL10A1, DEFB103B, KRT6B, COL11A1, LCAL1, PRAME, CTHRC1, UHRF1, FAM111B, KIAA0101, LHX2, KRT79, TTK, CDKN3, SH2D5, NCAPG, NEK2, GINS2, CENPF, NUSAP1, KRT75, ALPK2, SPRR1B, CDCA8 |
| Downregulated | CHRDL1, ANGPTL5, UMODL1-AS1, MAATS1, WISP2, ITIH5, RGN, LGR6, FAT4, SLC22A4, NOSTRIN, SYNPO2, FLJ26850, PRR29, CYP2F1, NXPH3, CNN1, SAXO2, PLN, ANO5, LINC00261, CFAP206, HIF3A, CNTN4, C11orf88, SLC16A12, DACT3, PRDM16, GDF10, SEMA3E, ATOH8, TSPAN8, PPP1R9A, TNXB, PDZRN3, ANGPTL7, RSPH4A, EFHC1, LOC286191, STMND1, SMAD9, PIGR, RIIAD1, FOLR1, RCAN2, CAPN9, ITGA8, KRT222, SPATA4, SCRG1, SRD5A2, ZNF540, DUSP26, PTGFR, FHL1, C10orf107, C16orf71, DNALI1, PGM5-AS1, LEPR, PPP1R1B, RIC3, ABCA8, PPARGC1A, MFSD4A, ZFPM2, BHMT2, TMEM132C, FBLN1, TCF21, JPH2, RSPH1, LMOD1, ZBTB16, ELN, ADRA1B, DES, ASPA, CFAP43, MS4A8, EDNRB, RSPO1, UBXN10, HRCT1, ENKUR, SH3BGRL2, SPAG6, FAM81B, ARMC4, C1QTNF7, TRIM63, TMEM229A, ABI3BP, NRG2, WDR78, RUNDC3B, WIF1, C1QTNF2, SNCA, ROPN1L, ALDH7A1, CDHR3, PRUNE2, ERICH3, CRISP2, PIH1D3, NTRK3, C7, KHDRBS2, MORN5, NKX2-1, DYNLRB2, CSH1, C9orf135, PDK4, GPM6A, ACTN2, FILIP1, MAGI2-AS3, PDE1A, CPED1, FGF14, RERGL, PIFO, ANGPTL1, TMEM100, SLC22A3, HSD17B13, CAPN8, GREM2, ARHGAP20, CFAP70, NR3C2, EMCN, DPT, RGS5, ARMC3, GPIHBP1, EPHA3, MFAP4, ADGB, DRC3, DNAH7, OGN, SCARA5, PRG4, ATP1A2, ERICH2, DRC1, ADGRB3, CDO1, SDPR, ANOS1, SORBS1, PCDH20, ABCA6, SFTA3, LIFR, SGCA, EFHB, FGF9, PLCL1, LIMCH1, LRRC74B, AKAP12, C11orf97, MEOX2, SNTN, CXCL12, ADGRG2, APOD, ZNF880, PDE5A, NECAB1, MYOC, FMO5, PEG3-AS1, JAM2, DUSP27, KIAA1456, SPARCL1, SYNE1, RGS22, PDZRN4, TTC29, LINC01082, ZBBX, NME5, KCNA5, CACNB2, METTL7A, RHOJ, PGR, CCDC173, SRPX, SSTR1, FAM189A2, GPRASP1, C6orf118, SCN7A, SPATS1, PCDH9, CEP126, CAPSL, FLRT3, ZMYND12, LMO3, C9orf24, RBMS3, ACTG2, GNA14, SHANK2, SPAG8, TMEM213, IL33, C1orf194, ADH1B, ARHGAP6, KLF15, ME3, MYRIP, LINC00968, KCNB1, FAXDC2, CASC1, FAM107A, KCTD16, CDH19, LDLRAD1, CCBE1, DCDC5, PGM5, EYA4, FXYD1, PPP1R3C, ADAMTS9-AS2, ASB5, BMPR1B, SEMA5A, PI16, DAW1, SP5, LOC100506725, FAM47E, EFCAB1, CASQ2, SLIT2, HSPB7, WFDC1, FLJ30901, SCN2B, MACROD2, MAMDC2, ETV1, ADAMTSL3, ROR1, AR, MAPK4, CCDC178, GALNT16, LINC00472, DIO3OS, RNASE4, CYP4B1, PTGIS, WDR63 |

these DEGs [26]. Also, MCODE, a Cytoscape plugin, was applied to identify the modules of the PPI network, and only those with nodes >3 and MCODE score >3 were presented.

**Kaplan-Meier plotter database analysis**

The Kaplan-Meier plotter is a widely-used tool for evaluating effects of various genes on survival based on TCGA and EGA databases [27]. These genes and their respective significant associations with survival were further verified using GEPIA to investigate data on gene expression based on thousands of samples in TCGA.

**Tumor Immune Estimation Resource (TIMER) database analysis**

TIMER (*https://cistrome.shinyapps.io/timer/*) is a valuable tool for comprehensive and systematic exploration of relationships between gene expression and immune infiltrations, including level of CD4+ T cells, CD8+ T cells, B cells, macrophages, dendritic cells, and neutrophils in multiple types of tumors [28]. TIMER uses a de-convolution and statistical method previously used to assess the enrichment of tumor-infiltrating immune cells (TIICs) from gene expression profiles. POLE2 expression in squamous cell lung cancer was assessed for possible associations with immune infiltration, including CD4+ T cells, CD8+ T cells, B cells, macrophages, dendritic cells, and neutrophils.
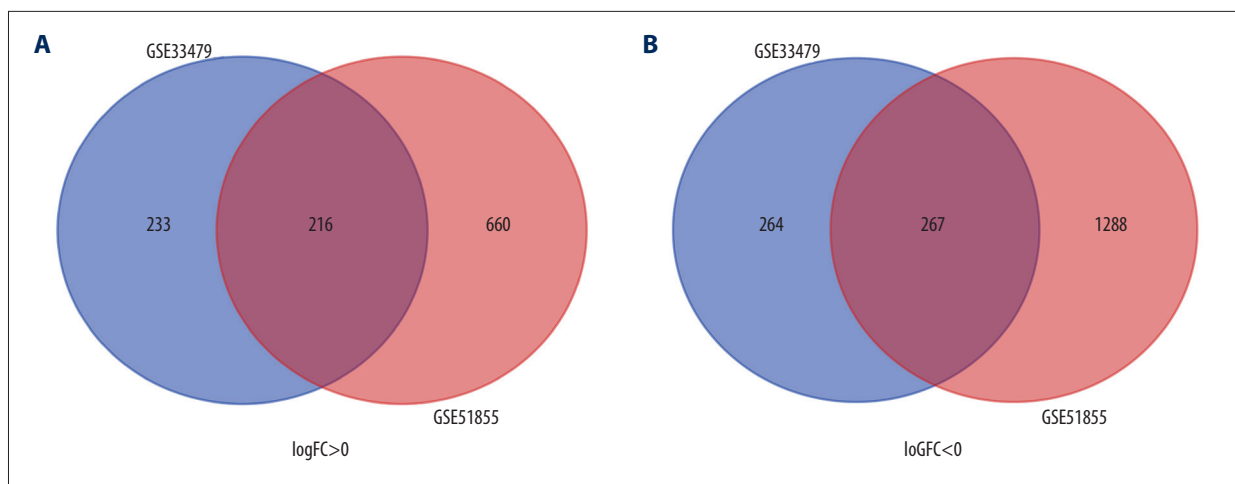
**Figure 1.** Identification of 483 common DEGs in 2 databasets (GSE33479 and GSE51855) via Venn diagrams online tool. And the different colors represent different datasets. (**A**) There were 216 up-regulated DEGs in 2 datasets (logFC >0). (**B**) There were 264 dwon-regulated DEGs in 2 datasets (logFC <0).

## Statistical analysis

Survival curves were produced with Kaplan-Meier plots and are presented with P values and HR as determined by the log-rank test. Biomarkers predicting squamous cell lung cancer prognosis were selected by survival analysis via Kaplan-Meier plotter. The strength of the relationship of gene expression with level of immune infiltration was evaluated via Pearson's correlation using absolute values as follows: 0.80–1.0 is "very strong", 0.60–0.79 is "strong", 0.40–0.59 is "moderate", 0.20–0.39 is "weak", and 0.00–0.19 is "very weak" [29]. $P<0.05$ was considered to indicate a statistically significant difference. Finally, biomarkers associated with both squamous cell lung cancer survival and immune infiltration level were selected.

## Results

### Identification of DEGs in squamous cell lung cancer

The study included 42 squamous cell lung cancer tissues and 17 normal lung tissues. A total of 1142 and 2897 DEGs from GSE33479 and GSE51855, respectively, were extracted via GEO2R online tools. The Venn diagram software was subsequently used to examine common DEGs in these 2 datasets. A total of 483 common DEGs, including 216 upregulated genes (logFC >0) and 267 downregulated genes (logFC <0) in squamous cell lung cancer tissues, were included (Table 1, Figure 1).

### Analysis of GO and KEGG pathway in squamous cell lung cancer

A total of 216 upregulated genes and 267 downregulated genes were investigated in the DAVID online website. The results of

GO analysis showed that common upregulated DEGs were gathered in "cell division", "sister chromatid cohesion", "mitotic nuclear division", "chromosome segregation", "DNA replication", and "DNA repair". Common downregulated DEGs were enriched in "cell adhesion", "oxidation-reduction process", "cell-cell signaling", and "muscle contraction". For CC, upregulated DEGs were enriched in "nucleus", "nucleoplasm", cytosol", "extracellular region", and "extracellular space". Downregulated DEGs were enriched in "extracellular region", "extracellular space", and "intracellular and proteinaceous extracellular matrix". For MF, common upregulated DEGs were enriched in "protein binding", "ATP binding", "DNA binding", "chromatin binding", "calcium ion binding", and "sequence-specific DNA binding", while downregulated GEDs were enriched in "calcium ion binding", "receptor binding", "heparin-binding", and "protein dimerization activity" (Table 2).

All DEGs were explored by KEGG pathway and DAVID, and results showed that common upregulated and downregulated genes were mainly involved in "cell cycle", "p53 signaling pathway", "oocyte meiosis", "progesterone-mediated oocyte maturation", "Fanconi anemia pathway", "calcium signaling pathway", and "pyrimidine metabolism" (Table 3).

### Analysis of PPI network and modular evaluation

All of the upregulated and downregulated DEGs were imported into the STRING online website. In total, 381 DEGs were submitted into the PPI online tool, including 381 nodes and 3586 edges, consisting of 201 upregulated and 180 downregulated genes (Figure 2A). Cytoscape MCODE was used to further analyze the central nodes, and the results showed 75 nodes, including 73 upregulated and 2 downregulated genes, were evaluated in those 381 nodes (Figure 2B).

**Table 2.** Gene ontology analysis of differentially expressed genes in squamous cell lung cancer.

| Expression | Category | Term | Count | % | p Value | FDR |
|---|---|---|---|---|---|---|
| Upregulated | GOTERM_BP_DIRECT | GO: 0051301 ~ cell division | 33 | 15.28 | 2.4E-21 | 2.2E-18 |
| | GOTERM_BP_DIRECT | GO: 0007067 ~ mitotic nuclear division | 28 | 12.96 | 1.1E-19 | 3.2E-17 |
| | GOTERM_BP_DIRECT | GO: 0007062 ~ sister chromatid cohesion | 22 | 10.19 | 2.4E-21 | 2.2E-18 |
| | GOTERM_BP_DIRECT | GO: 0007059 ~ chromosome segregation | 15 | 6.94 | 1.2E-14 | 2.7E-12 |
| | GOTERM_BP_DIRECT | GO: 0006260 ~ DNA replication | 15 | 6.94 | 1.2E-9 | 2.3E-7 |
| | GOTERM_BP_DIRECT | GO: 0006281 ~ DNA repair | 14 | 6.48 | 1.5E-6 | 1.8E-4 |
| | GOTERM_CC_DIRECT | GO: 0005634 ~ nucleus | 93 | 43.06 | 1.0E-7 | 3.7E-6 |
| | GOTERM_CC_DIRECT | GO: 0005654 ~ nucleoplasm | 61 | 28.24 | 2.2E-8 | 1.0E-6 |
| | GOTERM_CC_DIRECT | GO: 0005737 ~ cytoplasm | 75 | 34.72 | 2.2E-3 | 1.9E-2 |
| | GOTERM_CC_DIRECT | GO: 0005829 ~ cytosol | 53 | 24.54 | 1.6E-3 | 1.7E-2 |
| | GOTERM_CC_DIRECT | GO: 0005576 ~ extracellular region | 28 | 12.96 | 1.1E-2 | 7.5E-2 |
| | GOTERM_CC_DIRECT | GO: 0043565 ~ extracellular space | 21 | 9.72 | 7.9E-2 | 3.2E-1 |
| | GOTERM_MF_DIRECT | GO: 0005515 ~ protein binding | 126 | 58.33 | 3.4E-6 | 4.5E-4 |
| | GOTERM_MF_DIRECT | GO: 0005524 ~ ATP binding | 32 | 14.81 | 3.0E-4 | 1.6E-2 |
| | GOTERM_MF_DIRECT | GO: 0003677 ~ DNA binding | 25 | 11.57 | 9.1E-2 | 6.8E-1 |
| | GOTERM_MF_DIRECT | GO: 0003682 ~ chromatin binding | 18 | 8.33 | 1.2E-6 | 3.1E-4 |
| | GOTERM_MF_DIRECT | GO: 0005509 ~ calcium ion binding | 14 | 6.48 | 4.7E-2 | 5.3E-1 |
| | GOTERM_MF_DIRECT | GO: 0043565 ~ sequence-specific DNA binding | 12 | 5.56 | 2.5E-2 | 3.8E-1 |
| Downregulated | GOTERM_BP_DIRECT | GO: 0007155 ~ cell adhesion | 14 | 5.24 | 3.5E-3 | 7.5E-1 |
| | GOTERM_BP_DIRECT | GO: 0055114 ~ oxidation-reduction process | 12 | 4.49 | 9.7E-2 | 9.4E-1 |
| | GOTERM_BP_DIRECT | GO: 0007267 ~ cell-cell signaling | 9 | 3.37 | 1.1E-2 | 7.4E-1 |
| | GOTERM_BP_DIRECT | GO: 0006936 ~ muscle contraction | 8 | 3.00 | 2.9E-4 | 2.9E-1 |
| | GOTERM_CC_DIRECT | GO: 0005576 ~ extracellular region | 37 | 13.86 | 2.7E-4 | 2.9E-2 |
| | GOTERM_CC_DIRECT | GO: 0043565 ~ extracellular space | 29 | 10.86 | 3.9E-3 | 1.2E-1 |
| | GOTERM_CC_DIRECT | GO: 0005622 ~ intracellular | 26 | 9.74 | 2.1E-2 | 3.0E-1 |
| | GOTERM_CC_DIRECT | GO: 0005578 ~ proteinaceous extracellular matrix | 14 | 5.24 | 2.6E-5 | 5.8E-3 |
| | GOTERM_MF_DIRECT | GO: 0005509 ~ calcium ion binding | 23 | 8.61 | 3.2E-5 | 1.2E-2 |
| | GOTERM_MF_DIRECT | GO: 0005102 ~ receptor binding | 9 | 3.37 | 5.4E-2 | 7.8E-1 |
| | GOTERM_MF_DIRECT | GO: 0008201 ~ heparin binding | 10 | 3.74 | 1.1E-4 | 2.0E-2 |
| | GOTERM_MF_DIRECT | GO: 0046983 ~ protein dimerization activity | 6 | 2.24 | 3.1E-2 | 8.6E-1 |

**Table 3.** KEEG pathway analysis.

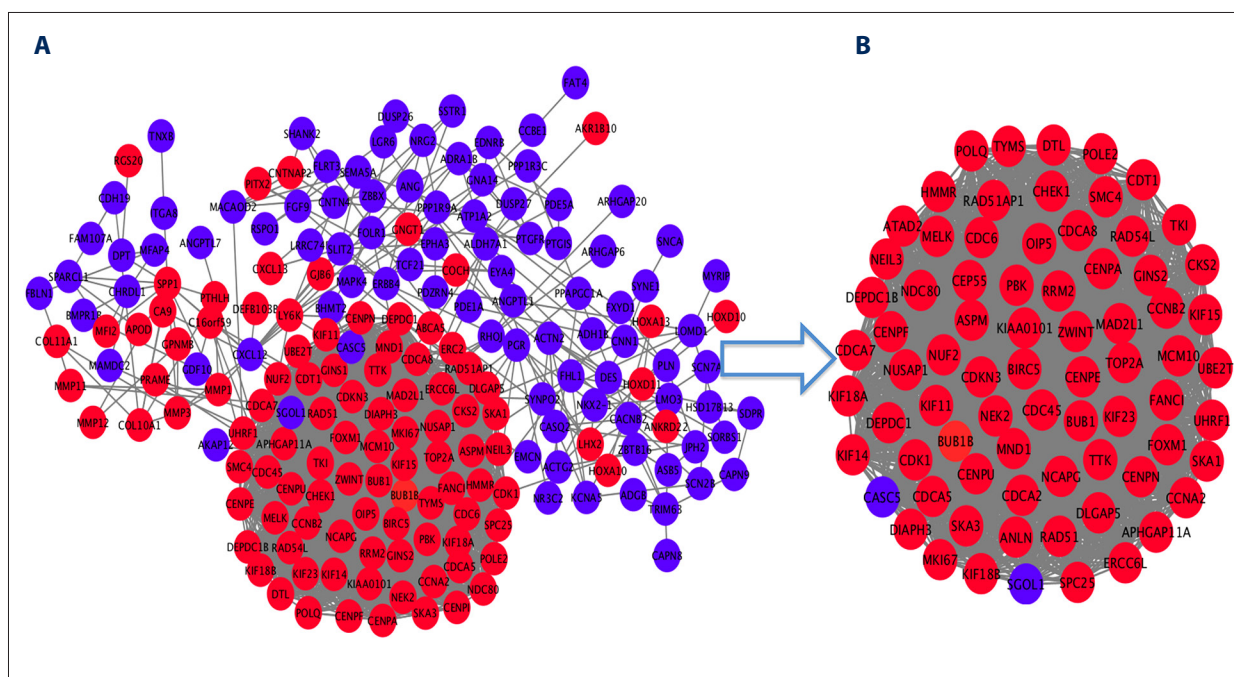| Pathway ID | Name | Count | % | P value |
|---|---|---|---|---|
| hsa04110 | Cell cycle | 13 | 6.01 | 2.0E-10 |
| hsa04115 | p53 signaling pathway | 6 | 2.78 | 2.0E-4 |
| hsa04114 | Oocyte meiosis | 6 | 2.78 | 2.0E-3 |
| hsa04914 | Progesterone-mediated oocyte maturation | 5 | 2.31 | 5.3E-3 |
| has03460 | Fanconi anemia pathway | 4 | 1.85 | 9.0E-3 |
| hsa04020 | Calcium signaling pathway | 6 | 2.78 | 3.7E-2 |
| has00240 | Pyrimidine metabolism | 4 | 1.85 | 4.9E-2 |



**Figure 2.** PPI network of common DEGs analyzed by STRING and Cytoscape. (**A**) There were 381 DEGs in the PPI network. The nodes represent proteins, while the edges represented the interaction of proteins. Red circle represented the up-regulated DEGs and blue circles reptesented down-regulated DEGs. (**B**) Module analysis via MCODE in Cytoscope software.

### Analysis of key genes by GEPIA and Kaplan-Meier plotter

The Kaplan-Meier plotter website was used to analyze the survival data for these 75 key genes, showing that only 4 genes can significantly predict survival for squamous cell lung cancer (P<0.05; Figure 3; Table 4), while 71 genes indicated no significant difference in the prediction of survival rate. Among these 4 genes, only CDC6 high expression predicted a better survival outcome compared with the other genes, in which low expression was associated with worse survival. Finally, GEPIA was used to examine the expression level of the 4 genes between patients with squamous cell lung cancer and healthy individuals. Results indicated that all these 4 genes showed high expression in squamous cell lung cancer compared to normal lung samples (P<0.05; Figure 4).

### Re-analysis of the 4 core genes via KEGG pathway enrichment

To further explore the underlying pathway for these 4 genes, the enrichment of KEGG pathway was re-assessed using DAVID (P<0.05) and the KEGG online website tool. Results indicated that only CDC6 was enriched in "cell cycle", and POLE2 was enriched in "base excision repair" and "DNA replication".

### Analysis of correlation of POLE2 and CDC6 with immune infiltration level in squamous cell lung cancer via TIMER database

Tumor-infiltrating lymphocytes have been reported to be an independent predictor of survival in multiple types of
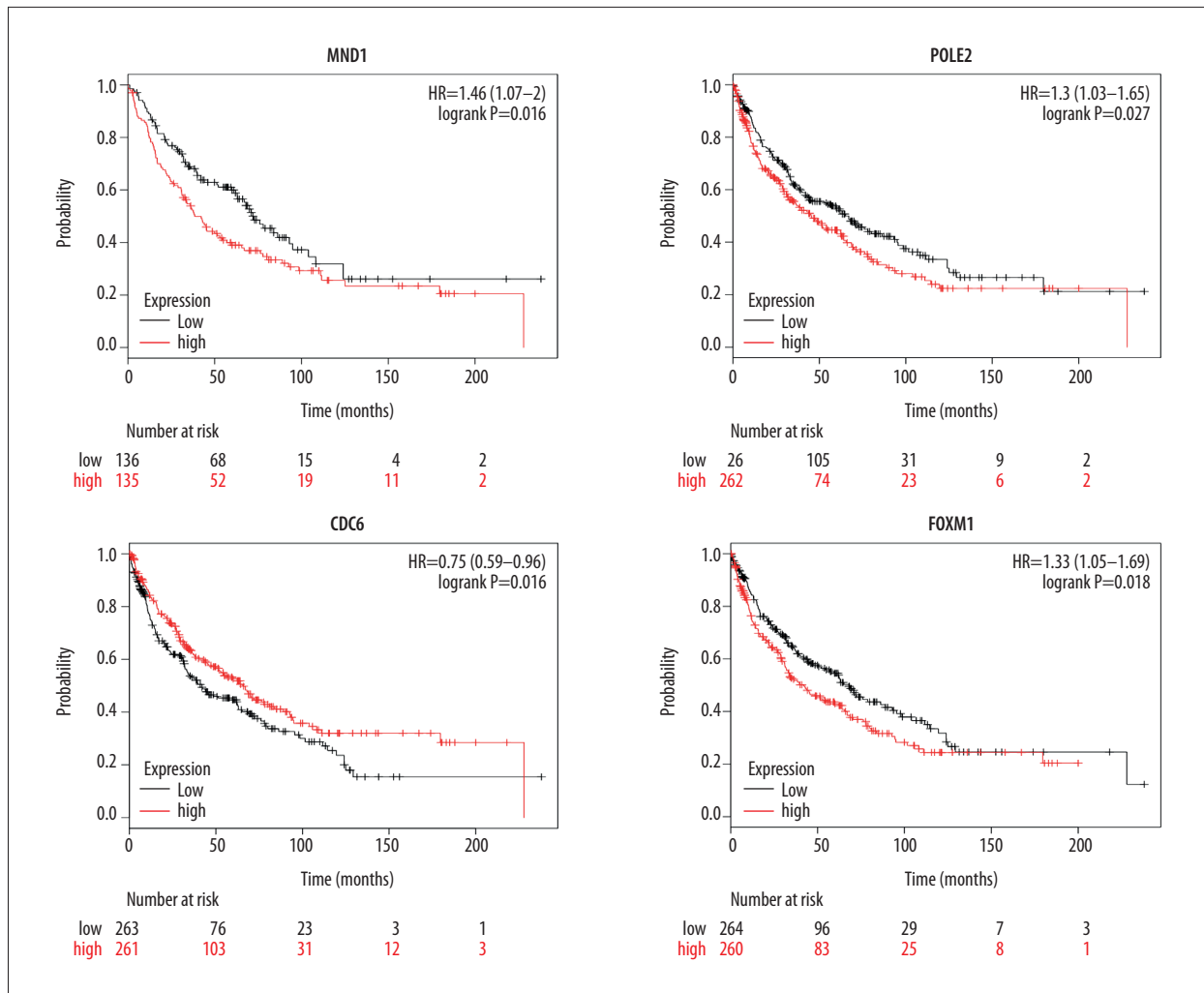
This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

e921430-6

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Figure 3.** The prognostic value of the 75 key genes. Kaplan-Meier plotter were applied to evaluate the prognostic information of 75 core genes and only 3 of 75 genes had a significantly worse survival rate and 1 of 75 genes had a significant positive survival rate (P<0.05).

**Table 4.** The prognostic information of the 75 key candidate genes.

| Category | Genes |
|---|---|
| Genes predicting significantly worse survival (P<0.05) | MND1, POLE2, CDC6, FOXM1 |
| Genes predicting significantly worse survival (P>0.05) | KIF23, SKA3, SPC25, ATAD2, CDC45, CASC5, POLQ, CDK1, CCNA2, BIRC5, CDCA2, TTK, NDC80, ASPM, GINS2, CDKN3, TK1, CDT1, OIP5, DEPDC1, CCNB2, RAD54L, CEP55, NCAPG, DIAPH3, SKA1, NEIL3, CENPE, CDCA7, TYMS, KIF15, ANLN, CDCA5, CHEK1, NUF2, KIF18B, SGOL1, CKS2, MAD2L1, NUSAP1, BUB1, UHRF1, CENPN, HMMR, RAD51, KIF14, BUB1B, CDCA8, RRM2, ZWINT, DEPDC1B, KIF18A, FANCI, CENPU, NEK2, SMC4, PBK, ERCC6L, MKI67, MELK, KIF11, CENPF, MCM10, UBE2T, CENPA, RAD51AP1, KIAA0101, DLGAP5, DTL, ARHGAP11A, TOP2A |
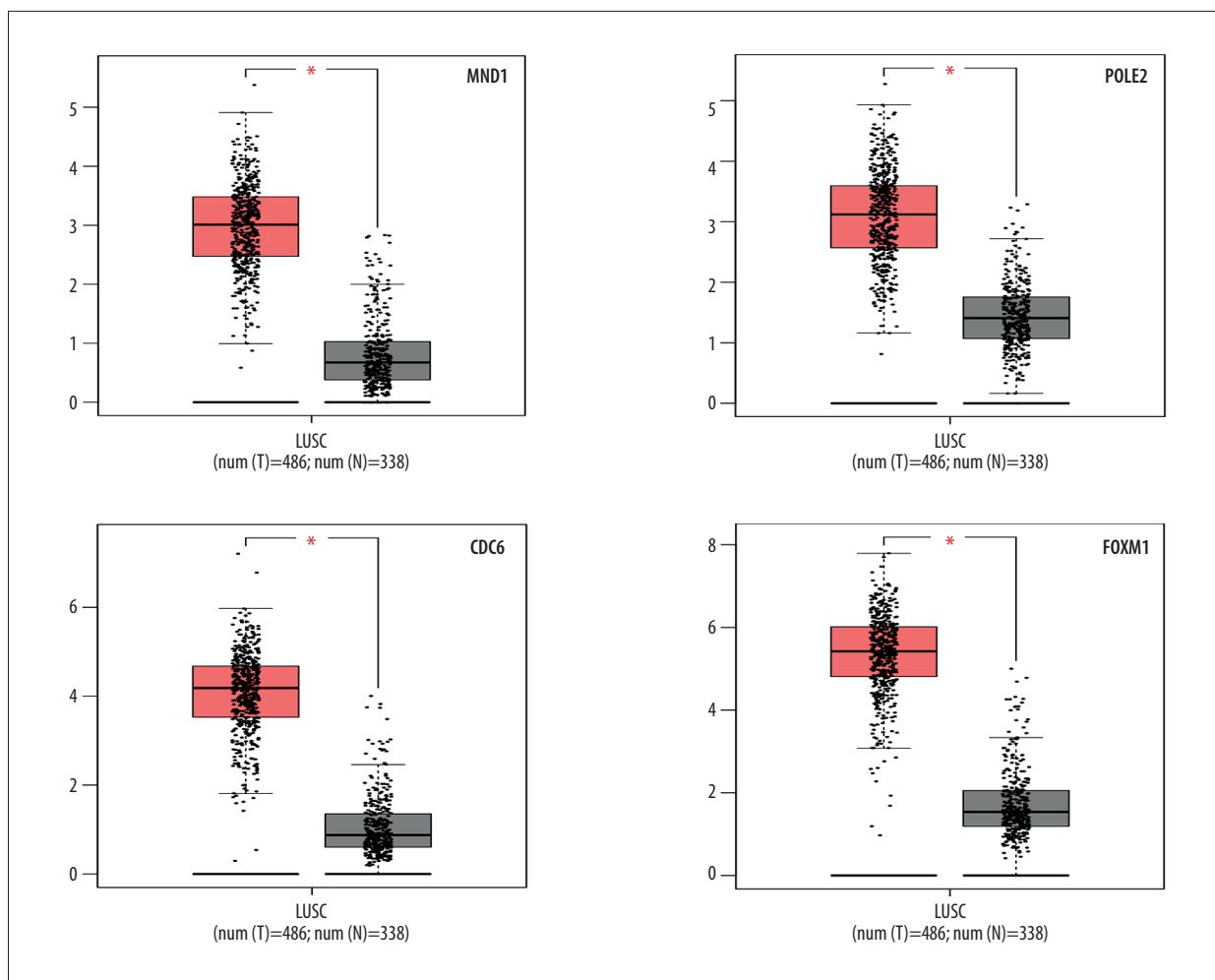
**Figure 4.** Expression of 4 genes related with survival in LUSC patients compared with healthy people. Expression of 4 genes which were related with prognosis were analyzed by GEPIA website and all of those had significant expression level in LUSC compared to normal one (* P<0.05). Red color represents tumor tissue and grey color represents normal tissue.

cancer [30,31]. Intense lymphocytic infiltration was demonstrated to be a favorable prognostic biomarker for survival in NSCLC [32]. To determine whether POLE2 and CDC6 expression levels were associated with immune infiltration levels in squamous cell lung cancer, TIMER was applied to assess the correlations of these 2 genes with immune infiltration levels. The results indicated that POLE2 expression was positively correlated with tumor purity and negatively correlated with CD4+ T cells (r=−0.229; P=4.31×10$^{-7}$), while there was only a very weak correlation with B cells (r=−0.136; P=3.11×10$^{-3}$), CD8+ T cells (r=−0.095; P=3.85×10$^{-2}$), macrophage cells (r=−0.168; P=2.19×10$^{-4}$), neutrophils (r=−0.099; P=3.13×10$^{-2}$), and dendritic cells (r=−0.14; P=2.19×10$^{-3}$) in squamous cell lung cancer (Figure 5). Similarly, CDC6 expression was positively correlated with tumor purity and negatively correlated with CD4+ T cells (r=−0.231; P=3.70×10$^{-7}$), macrophage cells (r=−0.340; P=2.47×10$^{-3}$), neutrophils (r=−0.200; P=1.05×10$^{-5}$), and dendritic cells (r=−0.220; P=1.38×10$^{-6}$), and very weakly correlated

with B cells (r=−0.12; P=8.87×10$^{-3}$) and CD8+ T cells (r=−0.104; P=2.34×10$^{-2}$) in squamous cell lung cancer (Figure 6). POLE2 and CDC6 expression exhibited no significant correlations with tumor purity and infiltration levels of immune cells, including CD4+ T cells, CD8+ T cells, B cells, macrophages, neutrophils, and dendritic cells in SKCM, which was used for the negative control group (Figures 5, 6). These findings suggest that POLE2 serves a core role in immune infiltrating in squamous cell lung cancer, especially those of CD4+ T cells. CDC6 expression was also correlated with immune infiltration in squamous cell lung cancer, particularly for CD4+ T cells, macrophage cells, neutrophils, and dendritic cells.

## Discussion

Due to lack of effective targeted therapy, squamous cell lung cancer has a worse prognosis than lung adenocarcinoma.
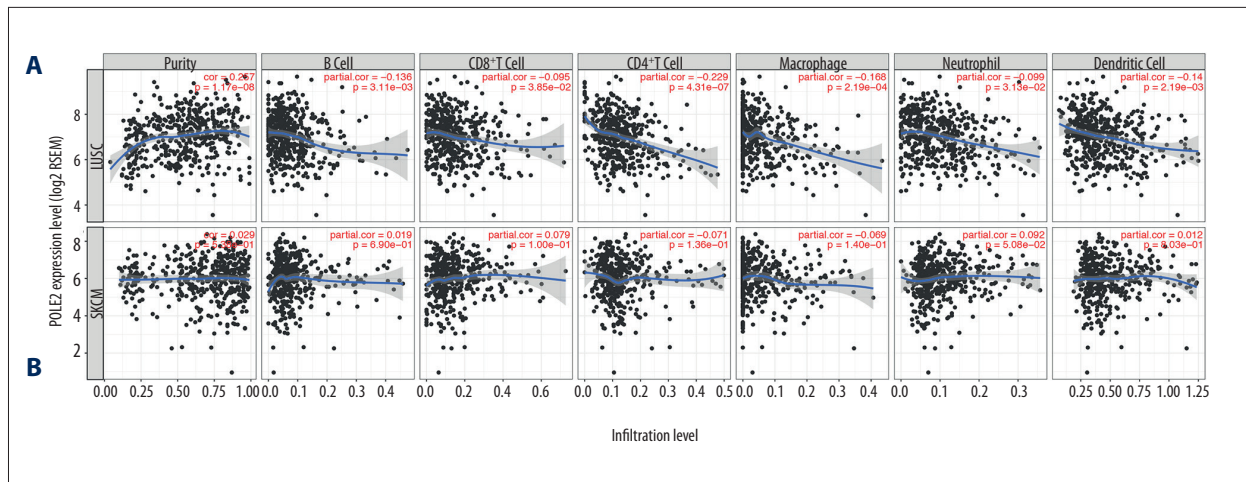
**Figure 5.** Correlation of POLE2 expression with immune infiltration level in LUSC (lung squamous cell carcinoma), and SKCM (skin cutaneous melanoma). (**A**) POLE2 expression was related to tumor purity significantly positively and had negative correlations with infiltrating levels of CD4+ T cells, very weak correlations with B cell, CD8+ T cells, macrophages, neutrophils, and dendritic cells in LUSC. (**B**) POLE2 expression has no significant correlation with tumor purity and the infiltrating levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells in SKCM.
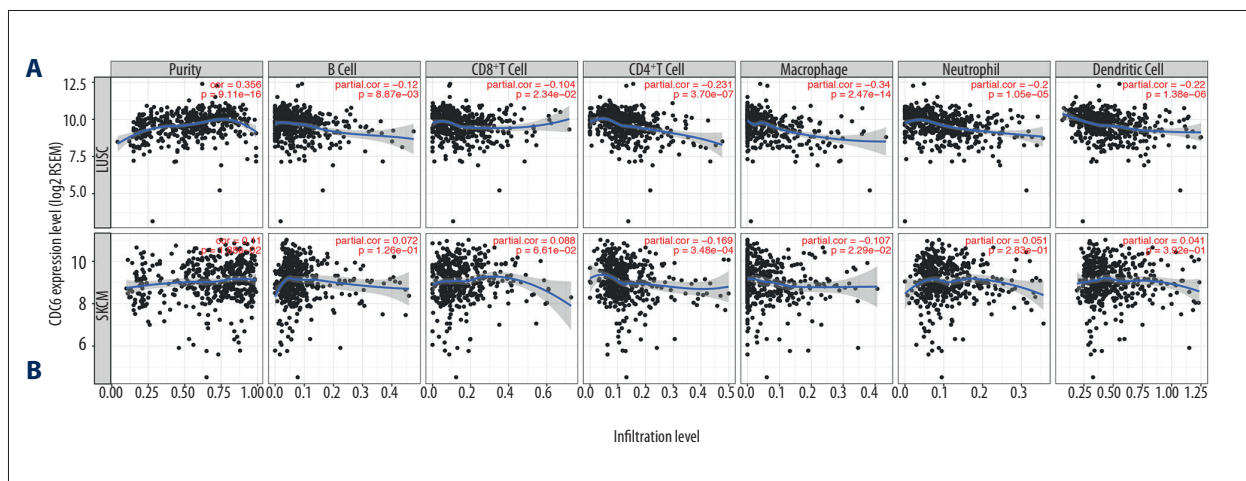


**Figure 6.** Correlation of CDC6 expression with immune infiltration level in LUSC (lung squamous cell carcinoma), and SKCM (skin cutaneous melanoma). (**A**) Results showed the expression of CDC6 had significant positive correlations with tumor purity and significant negative correlation with CD4+ T cells, macrophage cell, neutrophil cell and dendritic cell, and had very weak correlation with B cell and CD8+ T cell in LUSC. (**B**) CDC6 expression has no significant correlation with tumor purity and the infiltrating levels of B cells, CD8+ T cells, CD4+ T cell, macrophages, neutrophils, and dendritic cells in SKCM.

For example, the mutation in epidermal growth factor receptor and gene fusions in anaplastic lymphoma kinase are common in lung adenocarcinoma but rare in squamous cell lung cancer [33]. Novel effective biomarkers are needed to better predict the survival of squamous cell lung cancer patients. To identify promising prognostic biomarkers in squamous cell lung cancer, comprehensive bioinformatics analysis was applied based on 2 profile datasets: GSE33479 and GSE51855. In total, 42 squamous cell lung cancer tumor tissues and 17 normal tissues were included in the present study. Following a series of analyses via GE02R, GO and KEGG pathway, PPI, and

GEPIA, only POLE2 and CDC6 were associated with the prognosis of squamous cell lung cancer.

In this study, CDC6 was found to be associated with survival in squamous cell lung cancer, and its high expression predicted better survival (Figure 3). As a regulator of the cell cycle, CDC6 can regulate DNA replication and cell cycle in lung cancer [34]. High expression of CDC6 has been demonstrated to be associated with worse survival in multiple cancers, including lung adenocarcinoma and breast cancer [35–37]. However, the association of CDC6 with survival found in this study was

This work is licensed under Creative Common Attribution-
NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

e921430-9

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System]
[ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS]

not consistent with the previous studies mentioned above. Furthermore, we found that CDC6 expression was negatively correlated with immune infiltration in B cells and CD4+ T cells (Figure 6), which has been reported to be positively correlated with survival in lung cancer [38]. The correlation of CDC6 with immune infiltration level found in the TIMER database suggests that high expression of CDC6 should predict worse survival in squamous cell lung cancer (Figure 6). Although this study investigated the correlation of CDC6 expression with survival in squamous cell lung cancer, we could not determine whether high expression predicts better or worse survival in squamous cell lung cancer.

High DNA expression of POLE2 was significantly associated with decreased survival of patients with squamous cell lung cancer (Figure 3). To the best of our knowledge, there are no reports on the exact role of POLE2 in squamous cell lung cancer. Li et al. reported that POLE2-knockdown inhibited proliferation and apoptosis of lung adenocarcinoma cells in A549 and NCIH1299 cells [39]. However, the underlying mechanism of POLE2 and its role in the initiation and progression of squamous cell lung cancer remains unknown. Research has demonstrated that the POLE mutation is associated with a high mutational burden and elevated expression of some immune checkpoint genes, suggesting that POLE mutations are good biomarkers for immune therapy in endometrial cancer [40]. In this study, POLE2 expression was found to be correlated with immune infiltration level and negatively correlated with CD4+ T cells in squamous cell lung cancer. As reported in previous research, tumor-infiltrating lymphocytes, especially CD4+ T cells with a high CD4 level, predicted better overall survival and survival rate in head and neck squamous cell carcinoma [41]. This study also demonstrated that high POLE2 expression was negatively correlated with tumor-infiltrating lymphocytes. According to previous research and consistent with this study, a poor level of immune infiltration indicated a worse survival rate, and high POLE2 expression was negatively correlated with worse survival in squamous cell lung cancer.

The present study has certain limitations. First, all of the analyzed data were from a public database and the sample size was only 59. Second, complete clinical information is required to better assess the association of POLE2 with prognosis of squamous cell lung cancer. Finally, further validation of the association of POLE2 with squamous cell lung cancer prognosis is required in studies with larger samples to understand the underlying mechanism.

## Conclusions

In conclusion, we performed comprehensive bioinformatics analysis of various gene expression profiles of squamous cell lung cancer compared with normal tissue. All of the DEGs participate in a variety of pathways, including tumor initiation, progression, and immune infiltration of squamous cell lung cancer. Only POLE2 was found to be a key gene and to be simultaneously associated with survival and immune infiltration level in squamous cell lung cancer. Further studies are required to identify the function of POLE2 in squamous cell lung cancer.

### Conflict of interests

None.

## References:

1. Howlader N NA, Krapcho M, Miller D et al., (eds): SEER Cancer Statistics Review, 1975–2014, National Cancer Institute. Bethesda, MD, 2017

2. Torre LA, Bray F, Siegel RL et al: Global cancer statistics, 2012. Cancer J Clin, 2015; 65: 87–108

3. Zhou C, Wu YL, Chen G et al: Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): A multicentre, open-label, randomised, phase 3 study. Lancet Oncol, 2011; 12: 735–42

4. Mok TS, Wu YL, Thongprasert S et al: Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med, 2009; 361: 947–57

5. Fukui T, Taniguchi T, Kawaguchi K et al: Comparisons of the clinicopathological features and survival outcomes between lung cancer patients with adenocarcinoma and squamous cell carcinoma. Gen Thorac Cardiovasc Surg, 2015; 63: 507–13

6. Abel S, Hasan S, White R et al: Stereotactic ablative radiotherapy (SABR) in early stage non-small cell lung cancer: Comparing survival outcomes in adenocarcinoma and squamous cell carcinoma. Lung Cancer, 2019; 128: 127–33

7. Krefting F, Basara N, Schutte W et al: Clinical experience of immunotherapy treatment: Efficacy and toxicity analysis of the compassionate use program of Nivolumab in patients with advanced squamous cell non-small cell lung cancer. Oncol Res Treat, 2019; 42: 243–55

8. Zhang Y, Zhou H, Zhang L: Which is the optimal immunotherapy for advanced squamous non-small-cell lung cancer in combination with chemotherapy: Anti-PD-1 or anti-PD-L1? J Immunother Cancer, 2018; 6: 135

9. AEGIS Study Team: Shared gene expression alterations in nasal and bronchial epithelium for lung cancer detection. J Natl Cancer Inst, 2017; 109(7)

10. Sato T, Kaneda A, Tsuji S et al: PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. Sci Rep, 2013; 3: 1911

11. Wang Q, Gan H, Chen C et al: Identification and validation of a 44-gene expression signature for the classification of renal cell carcinomas. J Exp Clin Cancer Res, 2017; 36: 176

12. Sites ER, Smolarek TA, Martin LJ et al: Analysis of copy number variants in 11 pairs of monozygotic twins with neurofibromatosis type 1. Am J Med Genet A, 2017; 173: 647–53

13. Kim DS, Watkinson JC: Gene chip expression analysis in head and neck cancer. Clin Otolaryngol Allied Sci, 2002; 27: 296–303

14. Lin EW, Karakasheva TA, Lee DJ et al: Comparative transcriptomes of adenocarcinomas and squamous cell carcinomas reveal molecular similarities that span classical anatomic boundaries. PLoS Genet, 2017; 13: e1006938

15. Liu X, Liu X, Li J et al: Identification and integrated analysis of key biomarkers for diagnosis and prognosis of non-small cell lung cancer. Med Sci Monit, 2019; 25: 9280–89

16. Jing H, Qu X, Liu L et al: A novel long noncoding RNA (lncRNA), LL22NC03-N64E9.1, promotes the proliferation of lung cancer cells and is a potential prognostic molecular biomarker for lung cancer. Med Sci Monit, 2018; 24: 4317–23

17. Sun L, Chen Y, Su Q et al: Increased plasma miRNA-30a as a biomarker for non-small cell lung cancer. Med Sci Monit, 2016; 22: 647–55

18. Campbell JD, Alexandrov A, Kim J et al: Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet, 2016; 48: 607–16

19. Jung YJ, Katilius E, Ostroff RM et al: Development of a protein biomarker panel to detect non-small-cell lung cancer in Korea. Clin Lung Cancer, 2017; 18: e99–107

20. Li Y, Gu J, Xu F, Zhu Q et al: Transcriptomic and functional network features of lung squamous cell carcinoma through integrative analysis of GEO and TCGA data. Sci Rep, 2018; 8: 15834

21. Davis S, Meltzer PS: GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics, 2007; 23: 1846–47

22. Ashburner M, Ball CA, Blake JA et al: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000; 25: 25–29

23. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009; 4: 44–57

24. Altermann E, Klaenhammer TR: PathwayVoyager: Pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics, 2005; 6: 60

25. Szklarczyk D, Franceschini A, Wyder S et al: STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res, 2015; 43: D447–52

26. Shannon P, Markiel A, Ozier O et al: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 2003; 13: 2498–504

27. Hou GX, Liu P, Yang J, Wen S: Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan-Meier plotter. PLoS One, 2017; 12: e0174515

28. Li T, Fan J, Wang B et al: TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res, 2017; 77: e108–10

29. Pan JH, Zhou H, Cooper L et al: LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers. Front Immunol, 2019; 10: 6

30. Hu W, Sun R, Chen L et al: Prognostic significance of resident CD103(+) CD8(+)T cells in human colorectal cancer tissues. Acta Histochem, 2019; 121: 657–63

31. Djenidi F, Adam J, Goubar A et al: CD8+CD103+ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. J Immunol (Baltimore, Md.: 1950), 2015; 194: 3475–86

32. Brambilla E, Le Teuff G, Marguet S et al: Prognostic effect of tumor lymphocytic infiltration in resectable non-small-cell lung cancer. J Clin Oncol, 2016; 34: 1223–30

33. Dearden S, Stevens J, Wu YL, Blowers D: Mutation incidence and coincidence in non small-cell lung cancer: Meta-analyses by ethnicity and histology (mutMap). Ann Oncol, 2013; 24: 2371–76

34. Zhang X, Xiao D, Wang Z et al: MicroRNA-26a/b regulate DNA replication licensing, tumorigenesis, and prognosis by targeting CDC6 in lung cancer. Mol Cancer Res, 2014; 12: 1535–46

35. Mahadevappa R, Neves H, Yuen SM et al: The prognostic significance of Cdc6 and Cdt1 in breast cancer. Sci Rep, 2017; 7: 985

36. Yu C, Chen F, Jiang J et al: Screening key genes and signaling pathways in colorectal cancer by integrated bioinformatics analysis. Mol Med Rep, 2019; 20: 1259–69

37. Li S, Xuan Y, Gao B et al: Identification of an eight-gene prognostic signature for lung adenocarcinoma. Cancer Manag Res, 2018; 10: 3383–92

38. Liu X, Wu S, Yang Y et al: The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. Biomed Pharmacother, 2017; 95: 55–61

39. Li J, Wang J, Yu J et al: Knockdown of POLE2 expression suppresses lung adenocarcinoma cell malignant phenotypes *in vitro*. Oncol Rep, 2018; 40: 2477–86

40. Mehnert JM, Panda A, Zhong H et al: Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. J Clin Invest, 2016; 126: 2334–40

41. Nguyen N, Bellile E, Thomas D et al: Tumor infiltrating lymphocytes and survival in patients with head and neck squamous cell carcinoma. Head Neck, 2016; 38: 1074–84