Clinics in Oncology

Research Article
Published: 02 Oct, 2023

TMPRSS2, the SARS-CoV-2 Motivator as a Negative Prognostic Biomarker Decreased in Lung Adenocarcinoma

*Yao J[1,2] and Wang H[1,2]**

[1]*Shaoxing Keqiao Women & Children' Hospital, Zhejiang, China*

[2]*The Children's Hospital, Zhejiang University School of Medicine, Keqiao Branch, Zhejiang, China*

Abstract

Background: Angiotensin-Converting Enzyme (*ACE2*), Transmembrane Protease Serine 2 (*TMPRSS2*), Cathepsin L (*CTSL*) and *FURIN* are key factors to SARS-CoV-2 infection. We aimed to evaluate the differential expression of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* and the association between these four genes and prognosis in NSCLC, and further to explore their susceptibilities to SARS-CoV-2.

Methods: A total of 1026 Non-Small Cell Lung Cancer (NSCLC) patients in The Cancer Genome Atlas (TCGA) were enrolled to investigate the association between gene expression of *ACE2*, *TMPRSS2*, *CTSL* as well as *FURIN* and the overall survival. Then, 920 NSCLC patients from Gene Expression Omnibus (GEO) were analyzed to validate the corresponding genes for prognostic analysis utilizing meta-analysis. Kaplan-Meier curves were also plotted to verify the prognostic value of the respective genes. In addition, we analyzed the correlation between DNA methylation and immune infiltration with gene expression. Ultimately, GSE157057 and GSE163547 datasets were applied to elucidate the changes of *TMPRSS2* expression in lung or lung adenocarcinoma cells after SARS-CoV2 infection.

Results: *TMPRSS2* expression was lower in both LUAD and LUSC tissues compared to normal tissues. DNA methylation level of *TMPRSS2* were statistically higher in LUAD and LUSC (LUAD: P=1.62E-12; LUSC: P<1.00E-12). Meta-analysis showed that *TMPRSS2* continuous gene expression was significantly correlated with overall survival in LUAD (HR=0.83, 95% CI: 0.59-0.95), which was also verified in Kaplan-Meier plotter database (HR=0.47 (0.36-0.60), log rank P=1.40E-09). The expression of *TMPRSS2* in LUAD was positively correlated with the level of immune infiltration of B cell (r=0.242, P=6.66E-08), $CD4^+$ T cells (r=0.244, P=5.51E-08), macrophage (r=0.109, P=1.62E-02) and dendritic cells (r=0.159, P=4.40E-04).

Conclusion: *TMPRSS2*, correlated with immune infiltration, may be a tumor suppressor gene and a prognostic marker in LUAD. Due to the decreased expression of *TMPRSS2*, LUAD tissues may be more resistant to SARS-CoV-2 infection.

Keywords: *TMPRSS2*; Lung adenocarcinoma; Prognosis; SARS-CoV-2

OPEN ACCESS

Correspondence:
Hailong Wang, Shaoxing Keqiao Women & Children' Hospital, Zhejiang, China

Received Date: *07 Sep 2023*
Accepted Date: *26 Sep 2023*
Published Date: *02 Oct 2023*

Citation:
Yao J, Wang H. TMPRSS2, the SARS-CoV-2 Motivator as a Negative Prognostic Biomarker Decreased in Lung Adenocarcinoma. Clin Oncol. 2023; 8: 2018.

ISSN: 2474-1663

Abbreviations

SRAS-CoV-2: The Severe Acute Respiratory Syndrome Coronavirus 2; *ACE2*: Angiotensin-Converting Enzyme 2; *TMPRSS2*: Transmembrane Protease Serine 2; *CTSL*: Cathepsin L; NSCLC: Non-Small Cell Lung Cancer; LUSC: Lung Squamous Cell Carcinoma; LUAD: Lung Adenocarcinoma; TCGA: The Cancer Genome Atlas; GTEx: Genotype-Tissue Expression; GEO: Gene Expression Omnibus

Introduction

The novel Coronavirus Disease (COVID-19) has become a global epidemic since December 2019 [1]. Lung is the main target organ of COVID-19 [2,3]. And lung cancer has been the leading cause of cancer deaths worldwide for decades. Non-Small Cell Lung Cancer (NSCLC) accounts for about 85% of lung cancers, including Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC) [4]. Several epidemiological studies have shown that lung cancer patients are more susceptible to COVID-19 than healthy individuals [5-7].

It is now well established from many studies that host cell proteases like Angiotensin-Converting Enzyme 2 (*ACE2*), Transmembrane Protease Serine 2 (*TMPRSS2*) [8,9], Cathepsin L (*CTSL*) [10] and

FURIN [11], may promote the SARS-CoV-2 infection of host cells. Also, *ACE2* and *TMPRSS2* were aberrantly expressed in many tumors such as colon adenocarcinoma and pancreatic adenocarcinoma [12-14]. Notably, the expression of *TMPRSS2* in tumor tissues of lung cancer patients was decreased compared with normal tissues [14]. *CTSL* plays an important role in the growth of tumor cell and may regulate resistance to chemotherapy of tumor cells [15]. Also, the mRNA level and protein level of *FURIN* seems to be associated with the aggressiveness of NSCLC, which could be a potential marker and therapeutic target for lung carcinomas [16]. Depending on recent studies, *ACE2*, *TMPRSS2*, *CTSL*, and *FURIN* are key factors to COVID-19 and lung cancer. It is of great significance to conduct comprehensive research on the susceptibility of lung cancer tissues to SARS-CoV-2 at the molecular level of these four genes.

In this study, we investigated the relationship between gene expression of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* and the prognosis of NSCLC using univariate cox model. As a result, we found *TMPRSS2* and *CTSL* had significant association with the overall survival in LUAD. Moreover, 13 independent external validation sets from Gene Expression Omnibus (GEO) were collected to validate the corresponding genes for prognostic analysis by meta-analysis. It was ultimately verified that *TMPRSS2* is a prognostic marker for LUAD. Furthermore, we conducted DNA methylation analysis to better characterize the biological mechanisms of gene expression underlying cancer development. Our findings revealed the important role of *TMPRSS2* and provided a potential mechanism related to methylation and immune infiltration in LUAD. It also illustrated the possible sensitivity of LUAD patients to SARS-CoV-2 and the impact on the prognosis of COVID-19 patients.

Method

Study population and datasets

Gene expression data and clinical information of LUAD and LUSC were downloaded from The Cancer Genome Atlas (TCGA) database website (https://cancergenome.nih.gov/), aiming to explore the association between gene expression and survival outcome of NSCLC. Besides, we also downloaded DNA methylation data from the TCGA database to identify the potential underlying biological mechanisms.

A total of 1026 NSCLC patients (age >18 years old) were selected as participants in the present study. Only subjects with survival information would be included in the survival analysis. Clinical covariates, including the overall survival outcome, race, gender, age, and tumor stage. For validation, we collected 7 datasets of 920LUAD and LUSC patients from Gene Expression Omnibus (GEO), including GSE3141 [17], GSE8894 [18], GSE29013 [19], GSE30219 [20], GSE31210 [21], GSE37745 [22] and GSE37745 [23]. These data were profiled using Affymetrix Human Genome U133A Array (GPL96), Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) and Illumina HiSeq 2500 (GPL16791).

In addition, GSE17400, GSE157057 and GSE163547 databases were applied to elucidate the changes of *TMPRSS2* expression in lung or lung adenocarcinoma cells after SARS-CoV-2 infection. The GSE17400 dataset containing 9 infected samples and 9 control samples was based on the GPL570 platform (HG-U133_Plus_2). In the GSE157057 dataset, 4 Human lung organoids were grown in standard culture infected with SARS-CoV-2 on the GPL24676 (Illumina NovaSeq 6000) platform. In the GSE163547 dataset, 8 H522 human lung adenocarcinoma cells were infected with SARS-CoV-2 at varying multiplicities of infection and samples were processed for RNA-seq at 4, 24, 48, 72 and 96 h post-infection by Illumina NextSeq 500 (GPL18573).

GEPIA2 and UALCAN analysis

We utilized Gene Expression Profiling Interactive Analysis 2 (GEPIA2) [24] (http://gepia2.cancer-pku.cn/) web server to investigate the differential expression of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* between tumor tissue and normal counterparts of LUAD and LUSC subjects from public databases, including TCGA and Genotype-Tissue Expression (GTEx).

Additionally, survival analysis of these four genes in LUAD and LUSC was performed using Cox regression model, and DNA methylation analysis of these four genes in NSCLC from TCGA using UALCAN (http://ualcan.path.uab.edu/) [25] tool was also conducted.

Validation of the prognostic effect of four genes

To validate the prognostic significance of the four genes, 13 datasets containing LUAD and LUSC was attained from Gene Expression Omnibus (GEO). The univariate cox hazard model was applied to evaluate overall survival of subjects carrying specific genes in these datasets.

After computing the HRs and 95% confidence intervals, we conducted a meta-analysis to improve the statistical power of the results. The heterogeneity across multiple datasets was assessed by the Q test (I^2 statistics). If no obvious heterogeneity (I^2<50%, P>0.05), a fixed-effects model would be chosen for meta-analysis. Otherwise, a random effects model would be selected. And eight datasets (GSE3141, GSE19188, GSE29013, GSE30219, GSE31210, GSE37745 and TCGA) of lung cancer were combined to analyze in Kaplan-Meier Plotter (http://kmplot.com) [26].

Regarding the different microarray platforms, Human Genome U133A Array (GPL96), Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) and Human Genome U133A 2.0 (GPL3921) were included due to having 22,277 probe sets in common [26].

In addition, a multivariate Cox proportional hazards model was further used to confirm the association between each gene and overall survival with adjustment for age and gender.

Correlation analysis between gene expression and immune infiltration

Tumor Immune Estimation Resource (TIMER; cistrome.shinyapps.io/timer) [27] was used to investigate molecular characterization of tumor-immune interactions. We evaluated correlation between gene expression and six abundances of various immune cells (CD4+ T cells, CD8+ T cells, B cells, neutrophils, determine the dendritic cells and macrophages by TIMER algorithm). Furthermore, we estimated the association between the corresponding gene and T cells (general), CD8+ T cells, B cells, CD8+ T cells, neutrophils, monocytes, Tumor-Associated Macrophages (TAM), M1 cells, M2 cells, dendritic cells, NK cells, Th1 cells, Th2 cells as well as Treg cells. These genetic markers of immune cells have been well illustrated in the previous study [28].

Gene network and enrichment analysis of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN*

To analyze the potential interactions of these genes, we performed a gene network analysis by GENEMANIA website [29]. Also, we

conducted pathway enrichment analysis with Metascape [30] software to further explore the biological mechanisms related to the identified gene. Gene Ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) were used as reference. We used Benjamini-Hochberg method for multiple correction and False Discovery Rate (FDR) <0.05 was set as a cut-off threshold.

Statistical analysis

Firstly, we performed descriptive statistical analysis of LUAD and LUSC in TCGA. Continuous variables were described as mean ± Standard Deviation (SD) and categorized variables were expressed in frequency (n) and proportion (%). Subjects were categorized into the low-expression group and high-expression group according to the median value of the identified genes.

We utilized Kaplan-Meier curves to display the prognostic significance of gene expression and DNA methylation levels. Log-rank test was used to evaluate the difference between survival curves.

To identify whether DNA methylation influences cancer development through regulating the expression of target genes, we detected the association between DNA methylation and gene expression using Pearson's correlation (r) method.

P value <0.05 was considered statistically significant. P values were two-sided. All statistical analyses were performed using the R version 3.6.3 (R Foundation), unless otherwise specified.

Results

DNA methylation and gene expression of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN*

Utilizing GPEIA2, we firstly explored the expression of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* in LUAD and LUSC from TCGA and GTEx database. Under Threshold of Log2(fold change)>1 and P-value <0.05, we found that LUAD and LUSC tissues both have lower expression of *TMPRSS2* than normal tissues (Figure 1A).

Then, UALCAN tool were used to examine the DNA methylation levels of these four genes. The results showed that DNA methylation levels of all genes were statistically different between normal and cancer tissues. Among them, the DNA methylation level of *TMPRSS2* existed a AAA statistically higher methylation level in LUAD and LUSC (LUAD: P=1.62E-12; LUSC: P<1.00E-12). And clearly showed the distribution of 16 CpG sites of *TMPRSS2* in LUAD and LUSC.

We applied the median expression value of these four genes to divide LUAD and LUSC patients from TCGA into low or high expression groups for survival analysis using GEPIA2. The Kaplan-Meier diagram shows that in LUAD patients, high expression level of *TMPRSS2* is significantly associated with longer overall survival (Figure 2A, HR=0.74, log-rank P=4.60E-02). We also found that the high-expression group of *CTSL* had significantly shorter survival time than the low-expression group (Figure 2A, LUAD: HR=1.60, log-rank P=2.70E-03; LUSC: HR=1.40, log-rank P=7.70E-03).

At the same time, we also performed survival analysis in TCGA. The results of *TMPRSS2* and *CTSL* in TCGA were significantly consistent with those of GEPIA2 regardless of adjustment for age, race, sex, age, and tumor stage (Table 1).

Validation of prognostic effect of *TMPRSS2* and *CTSL* in LUAD

To verify the prognostic effect of the above genes in LUAD and LUSC, we collected 13 independent validation GEO datasets for

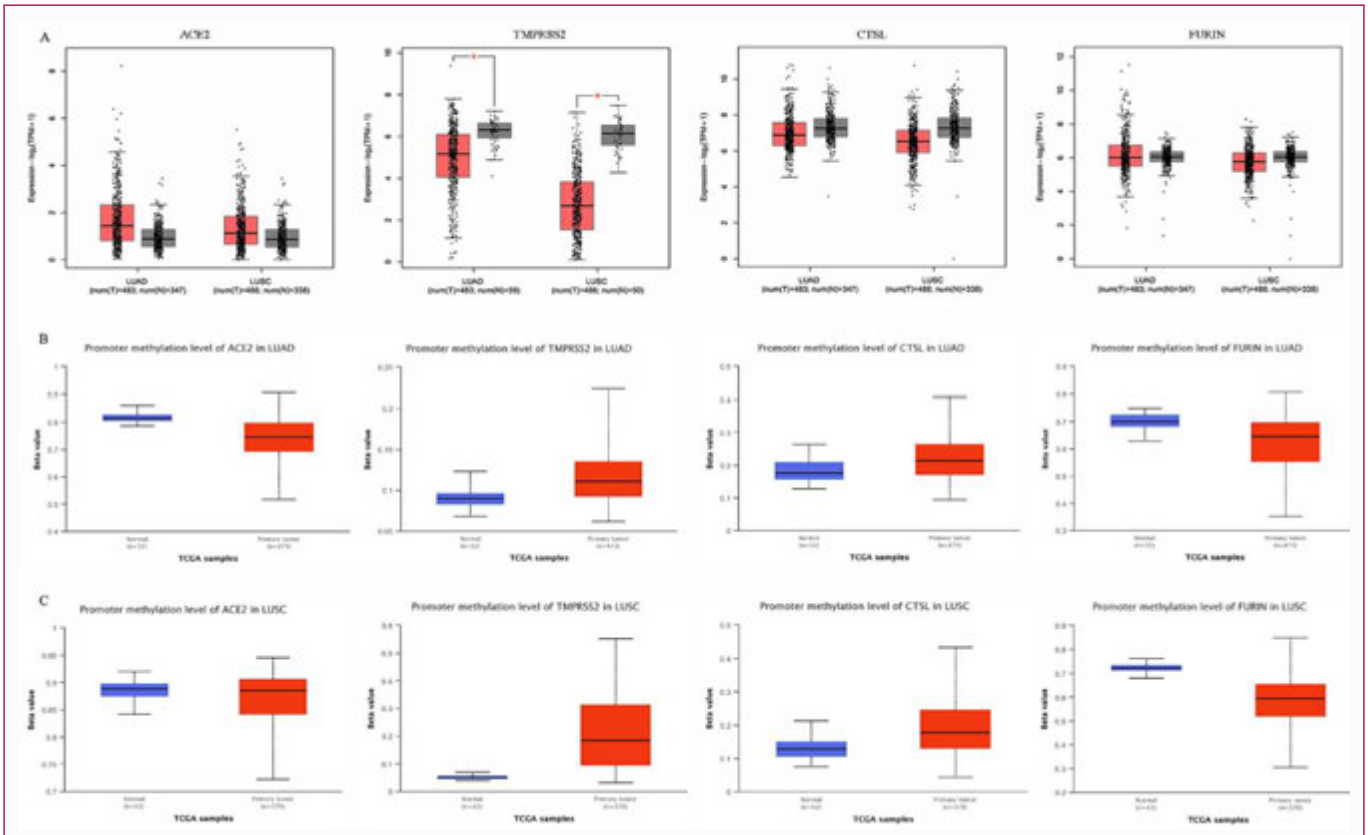

Figure 1: (A) The expression of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* in LUAD and LUSC using GPEIA2. The methylation level of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* in (B) LUAD and (C) LUSC demonstrating by UALCAN.

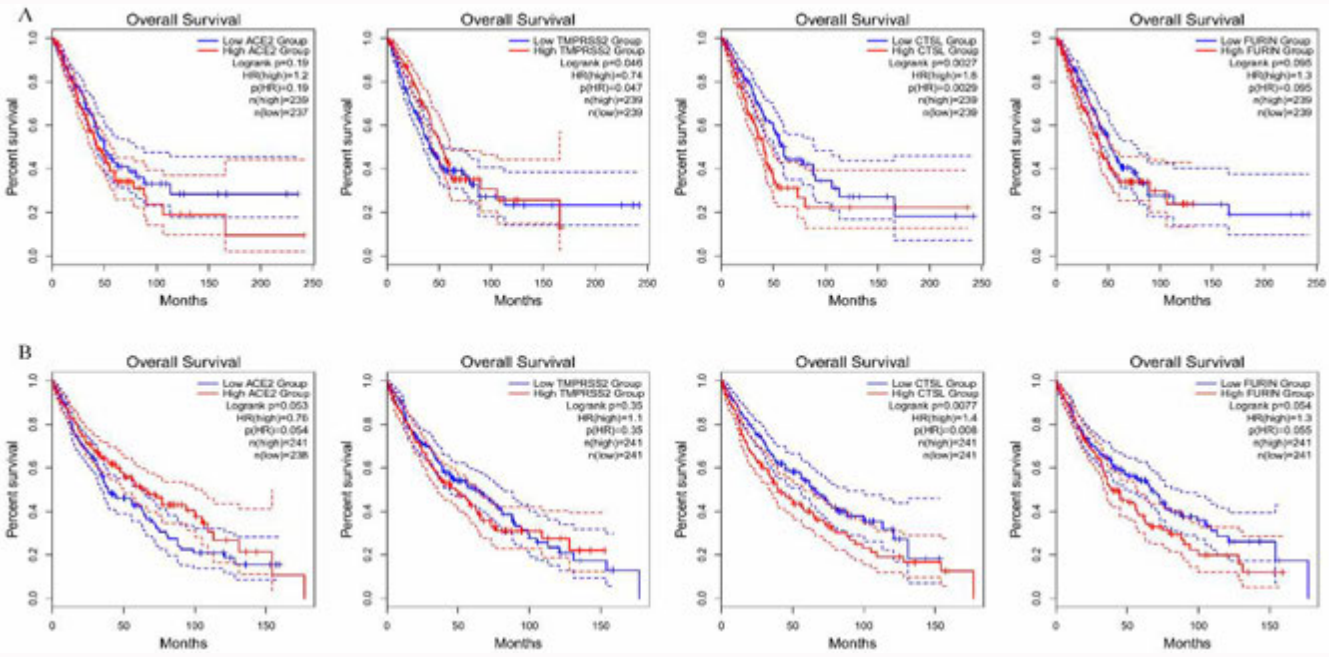

Figure 2: Prognostic gene and patient survival in TCGA dataset using GEPIA2. Kaplan-Meier survival analyses of patients categorized into low-expression and high-expression groups for (A) LUAD and (B) LUSC. P values were calculated using the log-rank test.

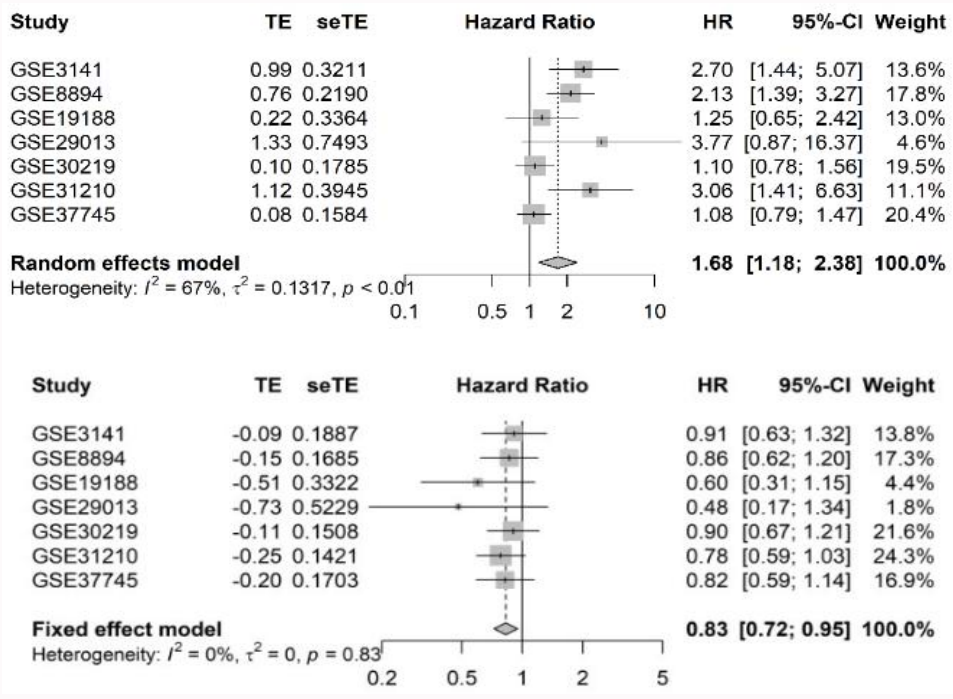

Figure 3: Forest plot of (A) *TMPRSS2* and (B) *CTSL* in 7 GEO datasets of LUAD.

meta-analysis. Meta-analysis in LUAD patients revealed that there was a significant correlation between *TMPRSS2* continuous gene expression and overall survival (HR=0.83, 95% CI: 0.59-0.95, Figure 3A), suggesting *TMPRSS2* as a tumor suppressor gene. No significant heterogeneity was observed in these 7 datasets of LUAD (I^2=0%, P=0.83, Figure 3A). Meta-analysis also showed that the lower *CTSL* expression in LUAD patients had a better survival (HR=1.68, 95% CI: 1.18-2.38, Figure 3B), but there existed heterogeneity in these 7 data sets of LUAD (I^2=67%, P<0.01, Figure 3B). However, Meta-analysis of *ACE2* and *FURIN* in lung cancer indicated no statistical difference between gene expression and survival.

In addition, we employed a Kaplan-Meier plotter to further verify our results, taking the differences between different platforms and the batch effect between different datasets into account. We ultimately found higher expression levels of *TMPRSS2* significantly associated with longer overall survival in LUAD (HR=0.47 (0.36-0.60), log rank P=1.40E-09, Figure 4A), while there was no statistical correlation

Table 1: Results of survival analysis using Cox regression in TCGA.

		No adjustment		Adjustment*	
Tumor	Variable	HR (95% CI)	P-Value	HR (95% CI)	P-Value
LUAD	*ACE2*	0.96 (0.84-1.10)	5.74E-01	0.94 (0.82-1.09)	4.17E-01
	TMPRSS2	0.88 (0.80-0.98)	1.50E-02	0.87 (0.79-0.96)	6.98E-03
	CTSL	1.61 (1.33-1.95)	7.10E-07	1.65 (1.36-2.00)	4.88E-07
	FURIN	1.23 (1.06-1.43)	6.37E-03	1.23 (1.05-1.43)	8.50E-03
LUSC	*ACE2*	0.96 (0.81-1.13)	6.16E-01	0.97 (0.82-1.14)	6.92E-01
	TMPRSS2	1.06 (0.96-1.16)	2.49E-01	1.04 (0.95-1.15)	4.16E-01
	CTSL	1.18 (0.99-1.42)	7.03E-02	1.27 (1.04-1.55)	1.89E-02
	FURIN	1.10 (0.88-1.37)	3.90E-01	1.06 (0.85-1.33)	5.85E-01

*We adjusted the models with age, race, gender and stage

Table 2: Survival analyses of different DNA methylation CpGs in LUAD and LUSC.

Gene symbol	CpGs	HR (95% CI)	P value
LUAD			
ACE2	cg16734967	6.98 (1.49-32.64)	1.35E-02
TMPRSS2	cg13489049	29.64 (5.37-163.66)	1.01E-04
CTSL	cg14190128	0 (0-0.06)	9.92E-05
CTSL	cg21177940	0.09 (0.01-0.86)	3.68E-02
FURIN	cg00758797	0.04 (0-0.33)	3.31E-03
FURIN	cg14231966	0.05 (0-0.72)	2.69E-02
LUSC			
ACE2	cg16734967	0.24 (0.07-0.85)	2.69E-02
ACE2	cg08559914	0.13 (0.02-0.71)	1.78E-02
TMPRSS2	cg14982276	0.45 (0.22-0.95)	3.58E-02
TMPRSS2	cg24901042	0.23 (0.07-0.75)	1.45E-02
TMPRSS2	cg26309194	0.37 (0.16-0.89)	2.53E-02
TMPRSS2	cg16371860	0.13 (0.03-0.65)	1.26E-02
CTSL	cg13985445	0.21 (0.05-0.89)	3.46E-02

between the expression of *CTSL* and overall survival in LUAD patients.

Notably, the meta-analysis of *CTSL* was heterogeneous, so the results may be unstable. Taken together, we finally concluded that *TMPRSS2* is a prognostic marker of LUAD.

Functional exploration of *TMPRSS2* methylation in LUAD development

For further exploring the prognostic effects of the methylation levels of these four genes in LUAD and LUSC, we obtained the CpG sites data of the four genes from TCGA and conducted survival analysis using univariate cox model. We found 6CpGs significantly associated with overall survival in LUAD and 7CpGs in LUSC (Table 2).

Further, the association between cg13489049 and *TMPRSS2* expression was evaluated using the TCGA dataset as cg13489049 maps to *TMPRSS2*. Correlation analysis showed that DNA methylation of cg13489049 and expression of *TMPRSS2* were negatively correlated (r= -0.44, P=3.80E-23, Figure 5A), indicating that cg13489049 cis-regulates *TMPRSS2* expression. And survival analysis suggested that higher DNA methylation level of cg13489049 prolonged survival in LUAD (HR=30.50, 95% CI: 5.46-170.60, P= 9.96E-05, Figure 5B).

TMPRSS2 expression correlated with tumor immune infiltration and cell type markers

Previous studies have shown that tumor invasion was associated with the prognosis of lung adenocarcinoma [31,32]. Therefore, we performed association analysis between *TMPRSS2* and immune infiltration in LUAD. The expression of *TMPRSS2* in LUAD was positively correlated with the immune infiltration level of B cell (r=0.242, P=6.66E-08), CD4+ T cells (r=0.244, P=5.51E-08), macrophage (r=0.109, P=1.62E-02) and dendritic cells (r=0.159, P=4.40E-04) (Figure 6).

We further explored the relationship between *TMPRSS2* expression and the cell type markers of different immune cells in LUAD. The cell type markers of B-cells, CD8+ T cells, neutrophils, macrophages, dendritic cells NK cells, Th1 cells, Treg cells and monocyte were analyzed by TIMER database. Correlation analysis indicated that *TMPRSS2* in LUAD was positively associated with MS4A1 in B cells, SIGLEC5 as well as CSF3R in Neutrophils and CD84 in Macrophages.

After adjustment for tumor purity and age, these correlations remain statistically stable (Table 3). The results above further confirmed that the expression of *TMPRSS2* in LUAD is related to immune infiltration.

TMPRSS2 expression analysis in SARS-CoV-2 infected human lung organoids and adenocarcinoma cells

To further investigate the changes of *TMPRSS2* in lung tissue and lung adenocarcinoma infected with SARS-CoV-2, we analyzed three datasets about SARS-CoV or SARS-CoV-2 infection in lung tissue.

A previous study showed that *TMPRSS2* expression was significantly different in SARS-CoV infected cells and mock-infected Calu-3 cells (P=4.65E-45, Figure 7A) in GSE17400, suggesting that SARS-CoV induces cytokines secreted by epithelial CALU-3 cells that may alter host inflammation and T-cell responses [6]. In a recent study on human lung organoid for modeling infection and disease conditions (GSE157057), human lung organoids were grown in standard culture and infected with SARS-CoV-2, where *TMPRSS2* expression was statistically different between infected and control group (P=0.04, Figure 7B). The studies above indicated that *TMPRSS2* expression may decrease when normal lung tissue is infected with SARS-CoV or SARS-CoV-2. Another recently released dataset GSE163547 conducted transcriptome analysis of SARS-CoV-2 on infected H522 human lung adenocarcinoma cells, we found no significant change in *TMPRSS2* expression (P=0.21, Figure 7C), indicating that lung adenocarcinoma cells with the lower expression

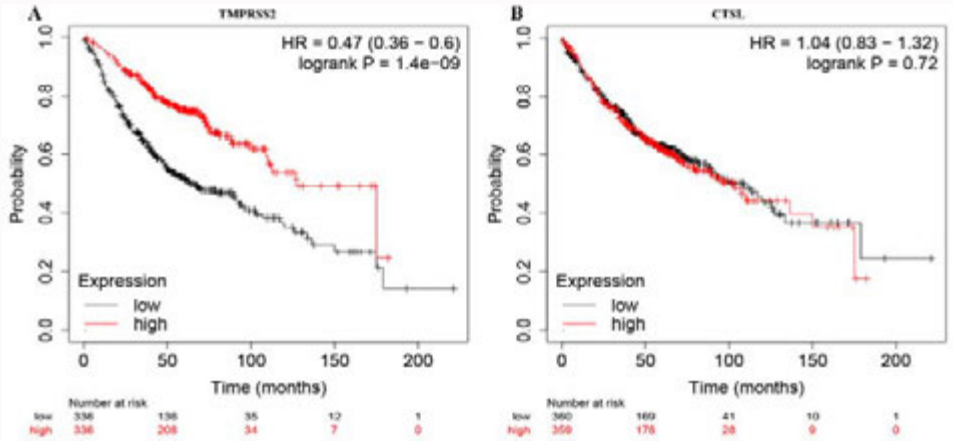

Figure 4: Kaplan-Meier survival curves of different level expression at (A) *TMPRSS2* and (B) *CTSL* using KM Plotter tool. Patients were categorized into low- and high-expression groups using median value of gene. P value was calculated using Cox regression model, and HR indicates hazard ratio.

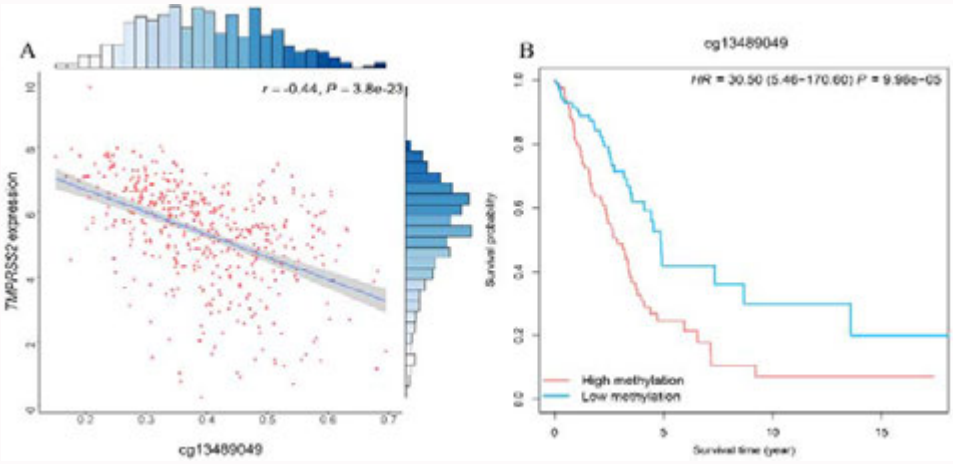

Figure 5: (A) Association between DNA methylation of cg13489049 and expression of corresponding gene *TMPRSS2*. (B) Kaplan–Meier survival curves of high and low methylation level of cg13489049 in LUAD.

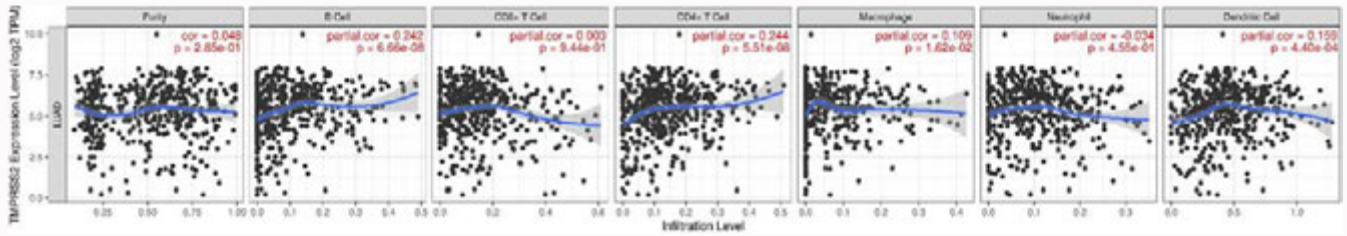

Figure 6: Correlation between *TMPRSS2* expression and immune infiltration in LUAD in TIMER database. *TMPRSS2* expression was positively correlated with B cell, CD4+T cell, macrophage and dendritic cell.

of *TMPRSS2* may be insensitive to SARS-CoV-2.

Gene network and enrichment analysis of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN*

In the gene network analysis, *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* were identified to be hub genes, while a new hub gene *NOTCH3* was also identified with most connectivity degrees in physical interactions (Figure 8). All related 24 genes in the network analysis. We then conducted enrichment analysis of these genes and identified pathways related to SARS-CoV-2 (WP4883) and inflammatory response (GO:0050727).

Discussion

Many studies have used bioinformatic approaches to explore the association between COVID-19 related genes and various cancers. For example, PERIKLIS et al. evaluated the differential expression of *ACE2*, *TMPRSS2* and *CTSL* in tumor patients [33]. However, most of the researchers mainly focused on pan-cancer. In the present study, we targeted NSCLC for the great impairment of lung function by COVID-19. Utilizing GEPIA2 and UALCAN, we mainly analyzed gene expression and DNA methylation levels of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* in LUAD and LUSC patients. Meanwhile, we

Table 3: Correlation analysis between *ACE2* and immune cell type markers in TIMER database.

Cell type	Gene symbol	None		Purity		Age	
		r	P value	r	P value	r	P value
B cells	FCRL2	0.083	6.01E-02	0.117	9.20E-03	0.073	1.10E-01
	CD19	0.047	2.85E-01	0.088	5.06E-02	0.035	4.37E-01
	MS4A1	0.179	4.32E-05	0.238	9.31E-08	0.168	2.03E-04
CD8+ T cells	CD8A	-0.154	4.47E-04	-0.132	3.22E-03	-0.15	9.28E-04
	CD8B	-0.156	3.94E-04	-0.134	2.77E-03	-0.15	9.52E-04
Neutrophils	FCGR3B	-0.09	4.23E-02	-0.092	4.06E-02	-0.083	6.77E-02
	CEACAM3	0.053	2.33E-01	0.061	1.80E-01	0.053	2.49E-01
	SIGLEC5	0.115	9.00E-03	0.147	1.08E-03	0.109	1.68E-02
	FPR1	-0.043	3.35E-01	-0.031	4.89E-01	-0.043	3.40E-01
	CSF3R	0.283	6.57E-11	0.301	8.58E-12	0.294	4.72E-11
	S100A12	-0.244	1.95E-08	-0.238	8.52E-08	-0.24	9.26E-08
Macrophages	CD68	-0.026	5.57E-01	0	9.92E-01	-0.039	3.94E-01
	CD84	0.099	2.40E-02	0.127	4.65E-03	0.094	3.87E-02
	CD163	-0.036	4.15E-01	-0.013	7.70E-01	-0.043	3.49E-01
	MS4A4A	0.004	9.24E-01	0.031	4.86E-01	-0.005	9.14E-01
Dendritic cells	CD209	-0.067	1.31E-01	-0.041	3.59E-01	-0.062	1.75E-01
NK cells	KIR3DL3	-0.234	7.47E-08	-0.243	4.38E-08	-0.227	4.71E-07
	NCR1	-0.18	4.19E-05	-0.174	1.08E-04	-0.178	8.73E-05
Th1 cells	TBX21	-0.037	4.01E-01	-0.011	8.07E-01	-0.043	3.46E-01
Treg	FOXP3	0.017	7.00E-01	0.046	3.10E-01	0.02	6.56E-01
	CCR8	0.067	1.28E-01	0.096	3.29E-02	0.069	1.28E-01
Monocyte	C3AR1	0.016	7.16E-01	0.043	3.36E-01	0.006	9.02E-01
	CD86	-0.002	9.66E-01	0.03	5.12E-01	-0.006	8.87E-01

Note: NK Cells: Natural Killer Cells; Th1 Cells: Type I Helper T Cells; Treg: Regulatory T Cells; COR: r Value of Spearman's Correlation; Purity, correlation adjusted by purity; Age correlation adjusted by age

investigated the prognostic value of the above genes in LUAD and LUSC patients. Then *TMPRSS2* was verified to be a promising prognostic marker for LUAD. Furthermore, we conducted a correlation analysis of *TMPRSS2* gene expression levels and immune infiltration, indicating that *TMPRSS2* expression in LUAD is related to immune infiltration.

TMPRSS2 activates the synergistic effect of SARS-2-S invading cells and *ACE2* protein, thus promoting the progression of new coronary pneumonia, which has led to considerable studies. *TMPRSS2* is a crucial factor for the SARS-CoV-2 entry [8,9]. Kong et al. reported that *TMPRSS2* was downregulated in lung cancers, which is consistent with our results (Figure 1A) [6]. Survival analysis showed that LUAD patients with lower expression of *TMPRSS2* had better survival. Recent studies also reported that lung cancer patients might be more resistant to SARS-CoV-2 infection due to the reduced expression of *TMPRSS2*, which was a SARS-CoV-2 internalization protease [34]. Furthermore, *ACE2* and *TMPRSS2* were co-expressed in lung cancer tissues and were positively correlated in LUAD (r=0.27, P=4.7e-09, Figure S), which may lead to less sensitivity to SARS-CoV-2 [6]. The high expression level of *ACE2* may indicate a greater susceptibility to SARS-CoV-2 infection, while *TMPRSS2* plays a supporting role.

In addition, *TMPRSS2* might be a tumor suppressor gene for its low expression in LUAD [6], and the mechanism has not been reported yet. *TMPRSS2* belongs to the type II class of serine transmembrane proteases and was considered to play a dominant role in the initiation of Epithelial-Mesenchymal Transition (EMT) in prostate cancer. It was reported that intact *TMPRSS2* could activate MET whereas this could be reversed with a c-MET inhibitor. One motif nominated by a chemical library screen corresponded to a precursor form of Hepatocyte Growth Factor (HGF) showed the crucial role of *TMPRSS2* in HGF/c-MET signaling. Also, compared to *TMPRSS2* null, a transcriptomic signature of *TMPRSS2* suggested an EMT signature with high expression of CXCL12/CXCR4 as well as HGF signatures and the increase of the EMT marker N-Cadherin. Thus, *TMPRSS2* could enhance EMT signaling *via* c-MET activation [35]. When compared to a spectrum of other human tissues, *TMPRSS2* was highly expressed in normal prostate tissue [36]. Similarly, biological mechanisms might exist in lung adenocarcinoma as well.

Detection of DNA hypermethylation of abnormal promoter CpG island could be used as an assessment of cancer risk and the emergence of potential biomarkers for early detection, prognosis, and prediction of treatment response [37,38]. DNA methylation biomarkers could be applied to the molecular prognosis of patients with potentially curable non-small cell lung cancers [39]. We found lower methylation of *TMPRSS2* cg02268510 may promote higher *TMPRSS2* expression, leading to poor prognosis in LUAD patients. Furthermore, we found that *TMPRSS2* expression was correlated with tumor immune infiltration and cell type markers in LUAD. These

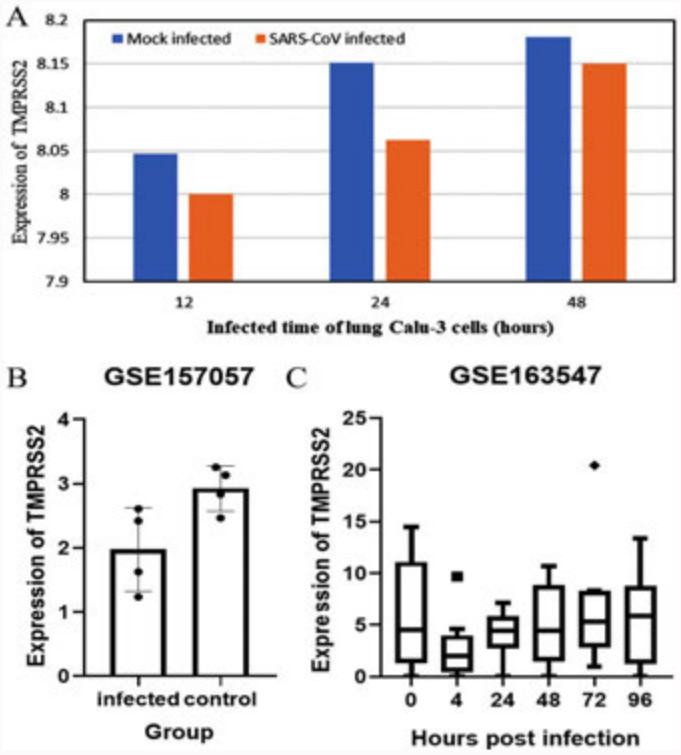

Figure 7: *TMPRSS2* genes expression (A) in human airway bronchial epithelial cells (Calu-3, a non-small-cell lung cancer cell line) infected with SARS-CoV (GSE17400), (B) in Human lung organoids infected with SARS-CoV-2 (GSE157057), and (C) in H522 human lung adenocarcinoma cells infected with SARS-CoV-2 at 4, 24, 48, 72 as well as 96 h post-infection (GSE163547).

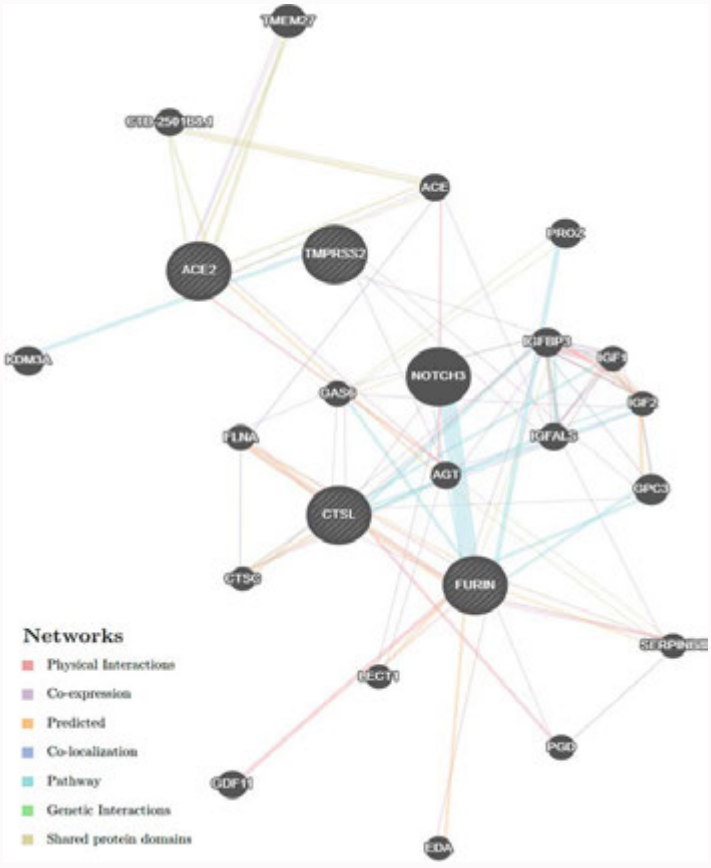

Figure 8: Gene network analysis of *ACE2*, *TMPRSS2*, *CTSL* and *FURIN*. The size of nodes represents the connectivity degree of each node, and the color of lines represents different cluster identification of networks.

results further supported the prognostic role of *TMPRSS2* in LUAD. Since our prognostic study of *TMPRSS2* in LUAD is a comprehensive bioinformatics analysis, further functional studies and validation in a larger clinical cohort are necessary.

In the gene network analysis, *ACE2*, *TMPRSS2*, *CTSL* and *FURIN* were identified to be hub genes along with another gene NOTCH3. Notably, the identified genes were also enriched in the SARS-CoV-2 and Angiotensin-converting enzyme 2 receptor: Molecular mechanisms pathway (WP4883) and regulation of inflammatory response pathway (GO:0050727). These results indicated the association between identified genes and COVID-19.

Admittedly, there are some limitations in our study. First, although we conducted a comprehensive bioinformatics analysis to identify *TMPRSS2* as a promising prognostic marker for LUAD, further experimental validation was still needed. Second, a recent study had revealed that head and neck cancer patients were not susceptible to SARS-CoV-2 due to the downregulation of *TMPRSS2*. We found similar characteristics in LUAD by *TMPRSS2* expression analysis in SARS-CoV-2 infected human lung organoids and adenocarcinoma cells, but this conclusion is conservative and needs validation in a larger population and experimental verification.

Conclusion

In conclusion, *TMPRSS2* may be a tumor suppressor gene and a prognostic marker in LUAD, which was correlated with immune infiltration. Due to the decreased expression of *TMPRSS2*, LUAD tissues may be more resistant to SARS-CoV-2 infection.

Acknowledgement

We thank the study participants and research staff for their contributions and commitment to this study.

References

1. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: A systematic review and meta-analysis. BMJ. 2011;342:d671.
2. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel Coronavirus (2019-nCoV). Radiology. 2020;295(1):202-7.
3. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner society: Glossary of terms for thoracic imaging. Radiology. 2008;246(3):697-722.
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-49.
5. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: A nationwide analysis in China. Lancet Oncol. 2020;21(3):335-7.
6. Kong Q, Xiang Z, Wu Y, Gu Y, Guo J, Geng F. Analysis of the susceptibility of lung cancer patients to SARS-CoV-2 infection. Mol cancer. 2020;19(1):80.
7. Pruis MA, Hussain B, Bakker M, Hoek RAS, Miedema JR, Dingemans AC, et al. A lung cancer patient with dyspnea: Diagnostic difficulties during the COVID-19 pandemic. Cancer Cell. 2020;38(5):611-12.
8. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181(2):271-80. e8.
9. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11(1):1620.
10. Smieszek SP, Przychodzen BP, Polymeropoulos MH. Amantadine disrupts lysosomal gene expression: A hypothesis for COVID-19 treatment. Int J Antimicrob Agents. 2020;55(6):106004.
11. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 2020;181(2):281-92.e6.
12. Chai P, Yu J, Ge S, Jia R, Fan X. Genetic alteration, RNA expression, and DNA methylation profiling of Coronavirus Disease 2019 (COVID-19) receptor ACE2 in malignancies: A pan-cancer analysis. J Hematol Oncol. 2020;13(1):43.
13. Liu Y, Qu HQ, Qu J, Tian L, Hakonarson H. Expression pattern of the SARS-CoV-2 entry genes ACE2 and TMPRSS2 in the respiratory tract. Viruses. 2020;12(10):1174.
14. Sacconi A, Donzelli S, Pulito C, Ferrero S, Spinella F, Morrone A, et al. TMPRSS2, a SARS-CoV-2 internalization protease is downregulated in head and neck cancer patients. J Exp Clin Cancer Res. 2020;39(1):200.
15. Zhao YF, Han ML, Xiong YJ, Wang L, Fei Y, Shen X, et al. A miRNA-200c/cathepsin L feedback loop determines paclitaxel resistance in human lung cancer A549 cells in vitro through regulating epithelial-mesenchymal transition. Acta Pharmacol Sin. 2018;39(6):1034-47.
16. Jaaks P, Bernasconi M. The proprotein convertase Furin in tumor progression. Int J Cancer. 2017;141(4):654-63.
17. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006;439(7074):353-7.
18. Lee ES, Son DS, Kim SH, Lee J, Jo J, Han J, et al. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. Clin Cancer Res. 2008;14(22):7397-404.
19. Xie Y, Xiao G, Coombes KR, Behrens C, Solis LM, Raso G, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. Clin Cancer Res. 2011;17(17):5705-14.
20. Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci Transl Med. 2013;5(186):186ra66.
21. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. PLoS One. 2012;7(9):e43923.
22. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, et al. Biomarker discovery in non-small cell lung cancer: Integrating gene expression profiling, meta-analysis, and tissue microarray validation. Clin Cancer Res. 2013;19(1):194-204.
23. Lohr M, Hellwig B, Edlund K, Mattsson JS, Botling J, Schmidt M, et al. Identification of sample annotation errors in gene expression datasets. Arch Toxicol. 2015;89(12):2265-72.
24. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47:W556-60.
25. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017;19(8):649-58.
26. Győrffy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis

software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One. 2013;8(12):e82241.

27. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res. 2017;77(21):e108-10.

28. Yang J, Li H, Hu S, Zhou Y. ACE2 correlated with immune infiltration serves as a prognostic biomarker in endometrial carcinoma and renal papillary cell carcinoma: Implication for COVID-19. Aging (Albany NY). 2020;12(8):6518-35.

29. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. Nucleic Acids Res. 2018;46(W1):W60-4.

30. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523-3.

31. Luo C, Lei M, Zhang Y, Zhang Q, Li L, Lian J, et al. Systematic construction and validation of an immune prognostic model for lung adenocarcinoma. J Cell Mol Med. 2020;24(2):1233-44.

32. Kadara H, Choi M, Zhang J, Parra ER, Rodriguez-Canales J, Gaffney SG, et al. Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. Ann Oncol. 2017;28(1):75-82.

33. Katopodis P, Anikin V, Randeva HS, Spandidos DA, Chatha K, Kyrou I, et al. Pan-cancer analysis of transmembrane protease serine 2 and cathepsin L that mediate cellular SARS-CoV-2 infection leading to COVID-19. Int J Oncol. 2020;57(2):533-9.

34. Iwata-Yoshikawa N, Okamura T, Shimizu Y, Hasegawa H, Takeda M, Nagata N. TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. J Virol. 2019;93(6):e01815-18.

35. Lucas JM, Heinlein C, Kim T, Hernandez SA, Malik MS, True LD, et al. The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumor microenvironment and promotes prostate cancer metastasis. Cancer Discov. 2014;4(11):1310-25.

36. Lin B, Ferguson C, White JT, Wang S, Vessella R, True LD, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res. 1999;59(17):4180-4.

37. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3(4):253-66.

38. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer. 2011;11(10):726-34.

39. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med. 2008;358(11):1118-28.