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

Yao J1,2 and Wang H1,2*

1Shaoxing Keqiao Women & Children’s Hospital, Zhejiang, China
2The 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 cells (r=0.244, P=5.51E-08), CD4+ T cells (r=0.234, 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

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
samples was based on the GPL570 platform (HG-U133_Plus_2). and Illumina HiSeq 2500 (GPL16791). (GPL96), Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) data were profiled using Affymetrix Human Genome U133A Array [20], GSE31210 [21], GSE37745 [22] and GSE37745 [23]. These and LUSC patients from Gene Expression Omnibus (GEO), and tumor stage. For validation, we collected 7 datasets of 920LUAD covariates, including the overall survival outcome, race, gender, age, as participants in the present study. Only subjects with survival mechanisms.

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 (F statistics). If no obvious heterogeneity (I2<50%, P>0.05), a fixed-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
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 $P=1.40\times10^{-9}$, Figure 4A), while there was no statistical correlation
Table 2: Survival analyses of different DNA methylation CpGs in LUAD and LUSC.

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

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

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

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Variable</th>
<th>No adjustment</th>
<th>Adjustment*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUAD</td>
<td>ACE2</td>
<td>5.74E-01</td>
<td>4.17E-01</td>
</tr>
<tr>
<td></td>
<td>TMPRSS2</td>
<td>1.50E-02</td>
<td>6.98E-03</td>
</tr>
<tr>
<td></td>
<td>CTSL</td>
<td>7.10E-07</td>
<td>4.88E-07</td>
</tr>
<tr>
<td></td>
<td>FURIN</td>
<td>6.37E-03</td>
<td>8.50E-03</td>
</tr>
<tr>
<td>LUSC</td>
<td>ACE2</td>
<td>6.16E-01</td>
<td>6.92E-01</td>
</tr>
<tr>
<td></td>
<td>TMPRSS2</td>
<td>2.49E-01</td>
<td>4.16E-01</td>
</tr>
<tr>
<td></td>
<td>CTSL</td>
<td>7.03E-02</td>
<td>1.89E-02</td>
</tr>
<tr>
<td></td>
<td>FURIN</td>
<td>3.90E-01</td>
<td>5.85E-01</td>
</tr>
</tbody>
</table>

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

**TMPPRSS2 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 TMPPRSS2 and immune infiltration in LUAD. The expression of TMPPRSS2 in LUAD was positively correlated with the immune infiltration level of B cell (r=0.242, P=6.66E-08), CD8+ 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 TMPPRSS2 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, T1h cells, Treg cells and monocyte were analyzed by TIMR database. Correlation analysis indicated that TMPPRSS2 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 TMPPRSS2 in LUAD is related to immune infiltration.

**TMPPRSS2 expression analysis in SARS-CoV-2 infected human lung organoids and adenocarcinoma cells**

To further investigate the changes of TMPPRSS2 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 TMPPRSS2 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 TMPPRSS2 expression was statistically different between infected and control group (P=0.04, Figure 7B). The studies above indicated that TMPPRSS2 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 TMPPRSS2 expression (P=0.21, Figure 7C), indicating that lung adenocarcinoma cells with the lower expression...
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
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 showed that LUAD patients with lower expression of TMPRSS2 had better survival. Recent studies also reported that lung cancer patients...
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**


26. Györffy B, Surowiak P, Budczies J, Lánczyk A. Online survival analysis...


