RESEARCH ARTICLE

Structural variations and expression profiles of the SARS-CoV-2 host invasion genes in lung cancer

Rahsan Ilikci Sagkan[1] [ID] | Dilara Fatma Akin-Bali[2] [ID]

[1]Department of Medical Biology, School of Medicine, Usak University, Usak, Turkey

[2]Department of Medical Biology, Faculty of Medicine, Nigde Omer Halisdemir University, Nigde, Turkey

Correspondence
Rahşan Ilikci Sagkan, Department of Medical Biology, School of Medicine, Usak University, Usak, Turkey.
Email: rahsan.ilikci@usak.edu.tr

Abstract

Recent days have seen growing evidence of cancer's susceptibility to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and of the effect of genomic differences on the virus' entrance genes in lung cancer. Genetic confirmation of the hypotheses regarding gene expression and mutation pattern of target genes, including angiotensin-converting enzyme-2 (ACE2), transmembrane serine protease 2 (TMPRSS2), basigin (CD147/BSG) and paired basic amino acid cleaving enzyme (FURIN/PCSK3), as well as correlation analysis, was done in relation to lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC) using in silico analysis. Not only were gene expression and mutation patterns detected, but also there were correlation and survival analysis between ACE2 and other target genes expression levels. The total genetic anomaly carrying rate of target genes, including ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3, was determined as 8.1% and 21 mutations were detected, with 7 of these mutations having pathogenic features. p.H34N on the RBD binding residues for SARS-CoV-2 was determined in our LUAD patient group. According to gene expression analysis results, though the TMPRSS2 level was statistically significantly decreased in the LUSC patient group compared to healthy control, the ACE2 level was determined to be high in LUAD and LUSC groups. There were no meaningful differences in the expression of CD147 and FURIN genes. The challenge for today is building the assessment of genomic susceptibility to COVID-19 in lung cancer, requiring detailed experimental laboratory studies, in addition to in silico analyses, as a way of assessing the mechanism of novel virus invasion that can be used in the development of effective SARS-CoV-2 therapy.

KEYWORDS

ACE2, CD147, Covid-19, FURIN, lung cancer, severe acute respiratory syndrome coronavirus 2, TMPRSS2

1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an RNA virus, which means that the genetic material of this virus contains positive-sense, single-stranded RNA. The novel coronavirus SARS-CoV-2 leads to coronavirus disease 2019 (synonym: COVID-19). Replication and its function are similar to other related coronaviruses that infect mammals, including humans.[1,2] The molecular mechanism of interaction between new coronavirus and mammalian host cells has not been fully understood. Numerous researcher groups have been focusing on the COVID-19 pandemic in recent days. There are two mechanisms

of the entire process of infection. The main target molecule of spike (S) protein of SARS-CoV-2 is angiotensin-converting enzyme-2 (ACE2) as a receptor on mammalian host cells.[3-5] The serine protease transmembrane serine protease 2 (TMPRSS2), another human protein, helps activate the coronavirus S protein to allow it to enter the cell.[6-8] In addition, the FURIN/PCSK3 cleavage of spike protein helps in SARS-CoV-2 interaction with ACE2 receptor.[9,10] FURIN/PCSK3 is one of the important proteases which facilitates viral invasion. The combination of binding and activation allows the virus to enter the host cell.[9,10] The other mechanism is S protein, which also binds to alternative or the additive molecule CD147/BSG, which is a novel receptor glycoprotein of the immunoglobulin superfamily, and which acts as the mediating the viral invasion.[11]

Cancer patients are at a risk for the development of several infections, including the prevalent COVID-19 pandemic in recent days.[12] The lungs are the main target of novel coronaviruses. Patients with lung cancer increase the susceptibility to infection.[8] ACE2 is commonly used molecule in the entrance of new viruses to host eukaryotic cells.[3,4,7,9,13] Transformed cells can change the surface molecules and signaling pathway enzymes and affect the susceptibility of virus infection. However, ACE2 is a powerful renin-angiotensin system (RAS) negative regulator that is crucial for maintaining the homeostasis of RAS, and the ACE2 is localized on the Xp22 chromosome and contains 18 exons.[3,4,14,15]

Lung cancer is the most common form of cancer and about 2 million people were diagnosed with lung cancer, consisting of 11.6% or 18.1 million of new cancer cases worldwide in 2018.[16] The most common form of lung cancer is non–small-cell lung cancer divided into three subtypes being adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma.[16,17] Infection depends on the type of mutation cancer has and how bad general health is. Lung tumors may block normal mucus drainage, which can lead to infection, and further, other types of entrance molecules that have been changed by cancer can be more prone to COVID-19 infection. This virus targets certain surface molecules involving ACE2 and CD147/BSG, which are surface of cancer cells. These target molecules help the new coronaviruses attack them.[3-8,11,13-15] However, specific studies are needed for investigating possible correlations among ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes in lung cancer for the assessment of SARS-CoV-2 infection susceptibility. In this way, we first designed our in silico analyses regarding COVID-19-related genes (ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3) expression, as well as mutation analyses, for revealing the susceptibility of COVID-19 in lung cancer subtypes (lung squamous carcinoma [LUSC] and lung adenocarcinoma [LUAD]). Secondly, the correlation of target genes was detected in two lung cancer subtypes. Additionally, survival rates vary widely. For this reason, we also determined the association between target genes and survival rates. Kaplan–Meier plotter was also used to study the prognostic relationship between target genes and lung cancer subtypes.

This study aims to investigate the potential of computational tools in increasing expressions and mutations of concerned invasion proteins, as well as their impact, which may be helpful for assessing possible COVID-19 susceptibility in patients with LUSC and LUAD. The possible susceptibility of cancers to SARS-CoV-2 and the prognosis of COVID-19 patients with LUSC and LUAD are also demonstrated.

2 | MATERIALS AND METHODS

2.1 | Data collection

The lung cancer data set was obtained from The Cancer Genome Atlas (TCGA) database. Demographic, clinical and genetic data regarding the patient group are summarized in Table 1.

2.2 | Mutation analysis

The cBio Cancer Genomics Portal (http://cbioportal.org) is an open accessed platform containing TCGA data set; it is a collaboration between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), is the large-scale cancer genomics projects that allow interactive research of multiple cancer genomic data sets, and provides access to the data of more than 5000 tumor samples from various cancer studies.[18] The LUAD and LUSC were chosen as the type of cancers of interest on the web interface to examine mutations in ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes in LUAD and LUSC patients presented in the portal. The selected TCGA data set comprised the genome sequencing data of 1097 LUAD and LUSC patients. To that end, we used an algorithm to analyze the mutation distribution of specific protein functional regions using the *OncoPrint*, *Cancer Type Summary*, and *Mutation* tools by the interface. These tools provide an overview of genomic alterations in particular genes affecting particular individual samples.

2.3 | Impact analysis of detected mutations

To determine the possible pathogenicity of the detected mutations in ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes, we used the scores provided by the Polymorphism Phenotyping v2 (PolyPhen-2), Screening for Non-Acceptable Polymorphisms (SNAP), and the Catalog of Somatic Mutations in Cancer (COSMIC) databases.

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2),[19] which can be accessed via a web server, helps in estimating the impact of possible mutations on the stability and function of human proteins using the structural and comparative evolutionary analyses of the amino acid substitution of these possible mutations. PolyPhen-2 calculates the functional definition of single nucleotide polymorphisms (SNPs), maps coding SNPs to gene transcripts, extracts protein sequence annotations, and structurally attributes and builds conservation profiles. The program estimates the probability of the missense mutation being damaging based on a combination of all these features and ensures both a qualitative

TABLE 1 Demographic, clinical and genetic data of patients with LUAD and LUSC

Characteristic	Patient data (n: 1097)
Gender	
Male/female/NA	615/411/69
Diagnosis age, y	71 (range, 33-90)
Sample type	
Primary	1095
Recurrent	2
Overall survival status	
Living	605
Deceased	396
NA	96
Metastasis stage code	
M0	767 (70%)
MX	219 (20%)
M1	23 (2.1%)
M1a	3 (0.3%)
M1b	6 (0.5%)
NA	79 (7.2%)
Alteration frequency in LUSC	Case (frequency %)
ACE2 mutation	3 (0.6%)
ACE2 amplification	6 (1.2%)
ACE2 deep deletion	8 (1.59%)
TMPRSS2 mutation	1 (0.2%)
TMPRSS2 amplification	2 (0.4%)
TMPRSS2 deep deletion	6 (1.2%)
CD147/BSG mutation	0
CD147/BSG amplification	2 (0.4%)
CD147/BSG deep deletion	4 (0.8%)
FURIN mutation	1 (0.2%)
FURIN amplification	11 (2.19%)
FURIN deep deletion	0
Alteration frequency in LUAD	
ACE2 mutation	6 (1.16%)
ACE2 amplification	1 (0.19%)
ACE2 deep deletion	5 (0.97%)
TMPRSS2 mutation	1 (0.19%)
TMPRSS2 amplification	3 (0.58%)
TMPRSS2 deep deletion	1 (0.19%)
CD147/BSG mutation	1 (0.19%)
CD147/BSG amplification	2 (0.39%)
CD147/BSG deep deletion	8 (1.55%)
FURIN mutation	5 (0.97%)
FURIN amplification	5 (0.97%)
FURIN deep deletion	1 (0.19%)
Tumor stage code in individuals with mutation (n)	T1 (1)
	T1a (1)
	T1b (4)
	T2 (8)
	T2a (3)

(Continues)

TABLE 1 (Continued)

Characteristic	Patient data (n: 1097)
	T4 (1)
Average smoking years in individuals with mutations	36.7

Abbreviations: ACE2, angiotensin-converting enzyme-2; BSG, basigin; LUAD, lung adenocarcinoma; LUSC, lung squamous carcinoma; MX, distant metastasis cannot be assessed; M0, no distant metastasis; M1, distant metastasis; M1a, distant metastasis to lung on opposite side of the primary tumor, pleural lymph nodes or malignant or pericardial effusion; M1b, distant metastasis; NA, not applicable.

prediction (*probably damaging, possibly damaging, benign, or unknown*) with a score.

SNAP[20] (https://www.rostlab.org/services/SNAP/) is a machine learning device called "neural network." It distinguishes between effect and neutral variants/non-synonymous SNPs by taking a variety of sequence and variant features into account. It includes evolutionary constraints, structural features, and protein annotation information. The most important single characteristic for SNAP prediction is conservation in a family of related proteins as reflected by PSIC scores. If the values are between −100-0 and 0-100, the mutations are considered neutral and effected, respectively, in the SNAP software.

Finally, the score given by the COSMIC[21] (https://cancer.sanger.ac.uk/cosmic) database was used to predict and verify the pathogenic effect of detected mutations. Evolutionary conservation region analysis was carried out to determine whether the mutations which were detected affect the critical amino acid codons for the target proteins. The evolutionary conservation analyses of the detected mutant amino acids were evaluated among different species via the "Multiple sequence alignment" tool in the PolyPhen-2 software.

2.4 | Gene expression analysis

Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index.html) is an advanced interactive network supporting normal and tumor tissue sample gene expression profiling and interactive analyses.[22] GEPIA offers customizable features such as differentially expressed tumor/normal analysis from the TCGA and the Genotype-Tissue Expression databases. It is an interactive web server advanced to provide several customizable analyses such as differential gene expression, profiling by cancer types or pathological stages, survival analysis, similar gene detection, correlation analysis, and dimensionality reduction.

The gene expression profiles of ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes were analyzed as box plot graphs created by the GEPIA database using data from 483 LUAD and 486 LUSC patients and healthy tissue samples obtained from the server. Furthermore, the correlation analyses between the expression levels of the ACE2 gene and other target genes were performed

using the software. The *P* values were automatically calculated by the software in both analyses, and *P* value .05 was established statistically significant. The survival analyses of the target genes (ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3) according to their varying gene expression fluctuations were performed using the web interface.

2.5 | Statistical analysis

All statistical analyses were carried out on the GEPIA database. Kaplan–Meier survival curves indicated that the overall survival. Low and high expression groups were compared using the log-rank test. Pearson test was performed for correlation analyses using the online database. *P* < .05 was established as statistically significant.

3 | RESULTS

3.1 | Results of mutation analysis

In our study, genome sequencing data of 1097 LUAD and LUSC cancer patients were selected and analyzed from cBioPortal portal to detect genetic changes in ACE2, TMPRSS2, CD147/BSG, and FURIN/ PCSK3 genes in LUAD and LUSC patient samples. It was determined that 8.1% of LUAD and LUSC patients who made up our study group had at least one genetic change (missense, nonsense, inframe mutation, gene amplification, and deep deletion) in the target genes. When we looked separately at the frequency of genetic anomaly transport in cancer types, it was found to be 7.4% for LUAD and 9.8% for LUSC. A total of 21 mutations (17 missense, 2 nonsense, 1 splice site, and 1 deletion) were detected for 4 target genes. Detailed information about the detected mutations is given in Table 2. ACE2 is the gene that carries the most genetic change in both types of cancer (Figure 1A-D). ACE2 is a membrane-bound ACE homolog and functions as a carboxypeptidase. The SARS-CoV-2 spike (S) protein uses the ACE2 receptor for host cell entry, as in SARS and MERS coronaviruses. The full length (805 amino acid) ACE2 consists of an N-terminal peptidase domain (PD) and a C-terminal Collectrin-like domain (CLD) that ends with a single transmembrane helix and an intracellular segment of approximately 40 amino acid residues.[4] A total of nine mutations (eight missense, one splice site), deep deletion, and gene amplification were detected in the ACE2 gene. It is known that the receptor-binding domain (RBD) on the SARS-CoV-2 spike protein and ACE2 has a total of 18 amino acid (Q24, T27, K31 H34, E37, D38, Y41, Q42, L45, L79, M82, Y83, N90, Q325, E329, N330, K353, and G354) residues on ACE2, and the H34 codon between these codons appears to be p.H34N with a missense mutation in the LUAD cancer group. Apart from this, when we consider functionally mutations, it is likely to cause an anomaly in ACE2 gene expression since the p.X233_splice mutation seen in the LUSC patient group is located in the splice region that is 100% protected between species in the evolutionary process. The other target gene,

TMPRSS2, is a type II transmembrane serine protease consisting of 492 amino acids expressed on the cell surface and, thus, ideally positioned to regulate cell–cell and cell–matrix interactions.[23] The binding of the S protein to ACE2 triggers a conformational change in the S protein of the coronavirus by allowing proteolytic digestion by host cell proteases (TMPRSS2), thereby infecting the cell by entering the viral RNA into the cell. In our study, a total of two mutations (one missense, one nonsense), gene amplification, and deep deletion were detected in the TMPRSS2 gene. The two nucleotides detected are on the transmembrane domain, in particular the p.G19* mutation may cause truncated protein formation as the TMPRSS2 polypeptide is used early in the codon 19 amino acid.

Another molecule is CD147 used by SARS-CoV-2 to enter host cells. CD147/BSG is a transmembrane glycoprotein of the immunoglobulin superfamily that plays a role in tumor development, plasmodium infestation, and virus infection.[11] In our study, one missense mutation p.F275V was detected in CD147/BSG gene in the LUAD patient group. Recently, SARS-CoV-2 has been reported to contain four residues (Pro681, Arg682, Arg683, and Ala684) as a potential cleavage site for the FURIN/PCSK3 protease of the S protein. Therefore, our fourth target gene in our study is FURIN/PCSK3. It includes nine nucleotide exchanges (seven missense, one deletion, one nonsense mutation) as well as gene amplification and deep deletion. All of the mutations detected except p.L428V missense change were detected in the LUAD patient group. Also, the p.E112* mutation may cause the FURIN/PCSK3 polypeptide to terminate at the codon 11 amino acid, causing truncated protein to form. In our study, the schematic representation of mutations detected in target genes on protein domains was summarized in Figure 2.

3.2 | Results of impact analysis of detected mutations

First, this analysis was carried out by including the missense mutations that we explained in detail in the results of mutation analysis according to our pathogenicity analysis results with the Poly-Phen2 Database Program. It was determined that 7 of 17 missense mutations given in Table 1 may have a pathogenic feature (*Probably damaging*) because the pathogenic score is close to 1. However, as a result of the second analysis made using the SNAP program, four of the same missense mutations were identified as being affected by the score between 0 and 100 (Figure 3A-D). In addition, a comparative analysis of amino acid sequences affected by missense mutations detected between different species was performed using the "Multiple sequence alignment" option in the Poly-Phen2 program. As a result of this analysis, it was determined that p.R219P, p.I256M, p.L320F, p.D693N, p.Q183K p.E230D, p.G277W, p.L420L, p.L428V, and p.Y495Q mutations were conserved due to evolutionarily critical conserved amino acids. In Figure 3, only missense mutations of ACE-2 were included. All of the conjectural pathogenic features and evolutionary conservation analyzes, which were conducted by using the Poly-Phen2 software, have been presented elaborately in Figure 3A-D.

TABLE 2 Mutations of the ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes in LUAD and LUSC patients

No	Gene	Nucleotide alteration	Accession Number	Alteration type	Localization	AA position	Previously determined disease/browser	Clinical significance			COSMIC prediction
								Poly-Phen2 (score)	SNAP (score)		
M-1	ACE2	c.100C>A	COSV53027349	Missense mutation	Peptidase domain	p.H34N	Lung adenocarcinoma	Benign (0.00)	Neutral (−83)	Neutral (score 0.02)
M-2	ACE2	c.1471G>T	COSV53026505	Missense mutation	Peptidase domain	p.G147V	Lung squamous cell carcinoma	Probably damaging (0.96)	Neutral (−11)	Neutral (score 0.12)
M-3	ACE2	c.631G>T	COSV53024979	Missense mutation	Peptidase domain	p.G211W	Lung adenocarcinoma	Benign (0.001)	Neutral (−21)	Neutral (score 0.14)
M-4	ACE2	c.656G>C	COSV53027016	Missense mutation	Peptidase domain	p.R219P	Lung adenocarcinoma	Probably damaging (0.98)	Effect (63)	Pathogenic (score 0.98)
M-5	ACE2	c.696G>A	rs771188220	Splice site mutation	Peptidase domain	p.X233_splice	Lung squamous cell carcinoma	N/A	N/A	N/A
M-6	ACE2	c.768C>G	COSV53026275	Missense mutation	Peptidase domain	p.I256M	Lung adenocarcinoma	Benign (0.38)	Neutral (−65)	Pathogenic (score 0.96)
M-7	ACE2	c.958C>T	COSV53026813	Missense mutation	Peptidase domain	p.L320F	Lung adenocarcinoma	Probably damaging (1.00)	Neutral (−22)	Pathogenic (score 0.99)
M-8	ACE2	c.1471G>T	COSV53026505	Missense mutation	Peptidase domain	p.V491L	Lung squamous cell carcinoma	Benign (0.001)	Neutral (−90)	Neutral (score 0.12)
M-9	ACE2	c.2077G>A	COSV99383055	Missense mutation	Transmembrane domain	p.D693N	Lung adenocarcinoma	Benign (0.001)	Neutral (−29)	Neutral (score 0.08)
M-10	TMPRSS2	c.55G>T	COSV59823689	Nonsense mutation	Transmembrane domain	p.G19*	Lung adenocarcinoma	N/A	N/A	Pathogenic (score 0.85)
M-11	TMPRSS2	c.83C>T	rs749665029	Missense mutation	Transmembrane domain	p.A28S	Lung squamous cell carcinoma	Benign (0.07)	Neutral (−97)	N/A
M-12	CD147/ BSG	c.825T>G	rs371073966	Missense mutation	I-set domain	p.F275V	Lung adenocarcinoma	Benign (0.019)	Neutral (−29)	N/A
M-13	PCSK3/ FURIN	c.248G>T	COSV51575105	Missense mutation	Pro-Segment	p.R83L	Lung adenocarcinoma	Benign (0.00)	Neutral (−53)	Pathogenic (score 0.87)
M-14	PCSK3/ FURIN	c.330G>T	N/A	Nonsense mutation	Pro-Segment	p.E112*	Lung adenocarcinoma	N/A	N/A	N/A
M-15	PCSK3/ FURIN	c.547C>A	COSV99184617	Missense mutation	Catalytic domain	p.Q183K	Lung adenocarcinoma	Benign (0.44)	Neutral (−66)	Pathogenic (score 0.98)
M-16	PCSK3/ FURIN	c.690G>T	COSV51575473	Missense mutation	Catalytic domain	p.E230D	Lung adenocarcinoma	Benign (0.001)	Neutral (−68)	Pathogenic (score 0.89)

(Continues)

TABLE 2 (Continued)

No	Gene	Nucleotide alteration	Accession Number	AA position	Previously determined disease/browser	Localization	Alteration type	Clinical significance		
								Poly-Phen2 (score)	SNAP (score)	COSMIC prediction
M-17	PCSK3/ FURIN	c.829G>T	COSV51579051	p.G277W	Lung adenocarcinoma	Catalytic domain	Missense mutation	Probably damaging (1.00)	Effect (71)	Pathogenic (score 0.99)
M-18	PCSK3/ FURIN	c.1260G>C	N/A	p.V420L	Lung adenocarcinoma	Catalytic domain	Missense mutation	Probably damaging (0.99)	Effect (35)	N/A
M-19	PCSK3/ FURIN	c.1282C>G	COSV51575092	p.L428V	Lung squamous cell carcinoma	Catalytic domain	Missense mutation	Probably damaging (0.99)	Neutral (−7)	Pathogenic (score 0.95)
M-20	PCSK3/ FURIN	c.1484A>G	COSV51576874	p.Y495C	Lung adenocarcinoma	P_domain	Missense mutation	Probably damaging (1.00)	Effect (56)	Pathogenic (score 0.97)
M-21	PCSK3/ FURIN	-	cBio web site	p.E783_R784del	Lung adenocarcinoma	Cytoplasmic domain	Deletion	N/A	NA	N/A

Abbreviations: AA, amino acid; M, mutation; N/A, not applicable.

3.3 | Results of gene expression analysis

The mRNA expression analysis was performed to determine whether 483 LUAD and 486 LUSC patients differ in ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 gene expression profiles compared to the healthy sample group. According to the results of our analysis, ACE2 and CD147/BSG gene expression levels are high in both cancer groups, although not statistically significant in healthy patient samples. However, in our LUSC patient group statistically significantly, TMPRSS2 gene expression was found to be lower in the patient group compared to the healthy sample group. There is no significant difference for FURIN/PCSK3 in both cancer groups (Figure 4A-D). In addition, the relationship among TMPRSS2, CD147/BSG, and FURIN/PCSK3 expression results and the expression profile of ACE2 were evaluated separately by Pearson's correlation test, and no correlation was detected (Figure 5). According to our results of survival analysis, the overall survival times of those with low TMPRSS2 expression in the LUAD cancer group are statistically significantly longer than those with high levels (Figure 6, p = .04).

4 | DISCUSSION

Since the first case detected in December 2019, a new coronavirus causing COVID-19 disease has led to more than 3 500 000 infections worldwide and about 245 000 deaths.[24] Deaths caused by this pandemic affected more people with many critical diseases, including cancer. Cancer patients have more susceptibility to infectious and varying responses to pathogen invasion to host cells regulated gene expression and mutation development in the related genes. Gene expression and mutation profile may show variability between individuals depending on various parameters such as therapy regimen, age, and immune system condition. Cancer patients may be at a higher risk of being more susceptible to COVID-19 infectious diseases. Recent studies have shown that ACE2 and cellular protease TMPRSS2 play a role in the entry of SARS-CoV-2 into lung cells.[6-8,23,25]

Our study is complementary to the lack of information in the literature to make a comparative evaluation of the mutation and gene expression profiles of ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes in LUAD and LUSC patient groups. The study of ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3's genomic and functional expression variants for potential susceptibility and/or resistance to coronavirus infection is not particularly available in the literature. First of all, mutation profiles of target genes, including ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 from the genome sequencing results of the LUAD and LUSC patient, which are accessible in TGCA data sets, were analyzed extensively. In our analysis results, it was determined that 8.1% of LUAD and LUSC patients had genetic abnormalities and ACE2 was the most mutating gene among these four target genes. However, when cancer groups are evaluated separately,
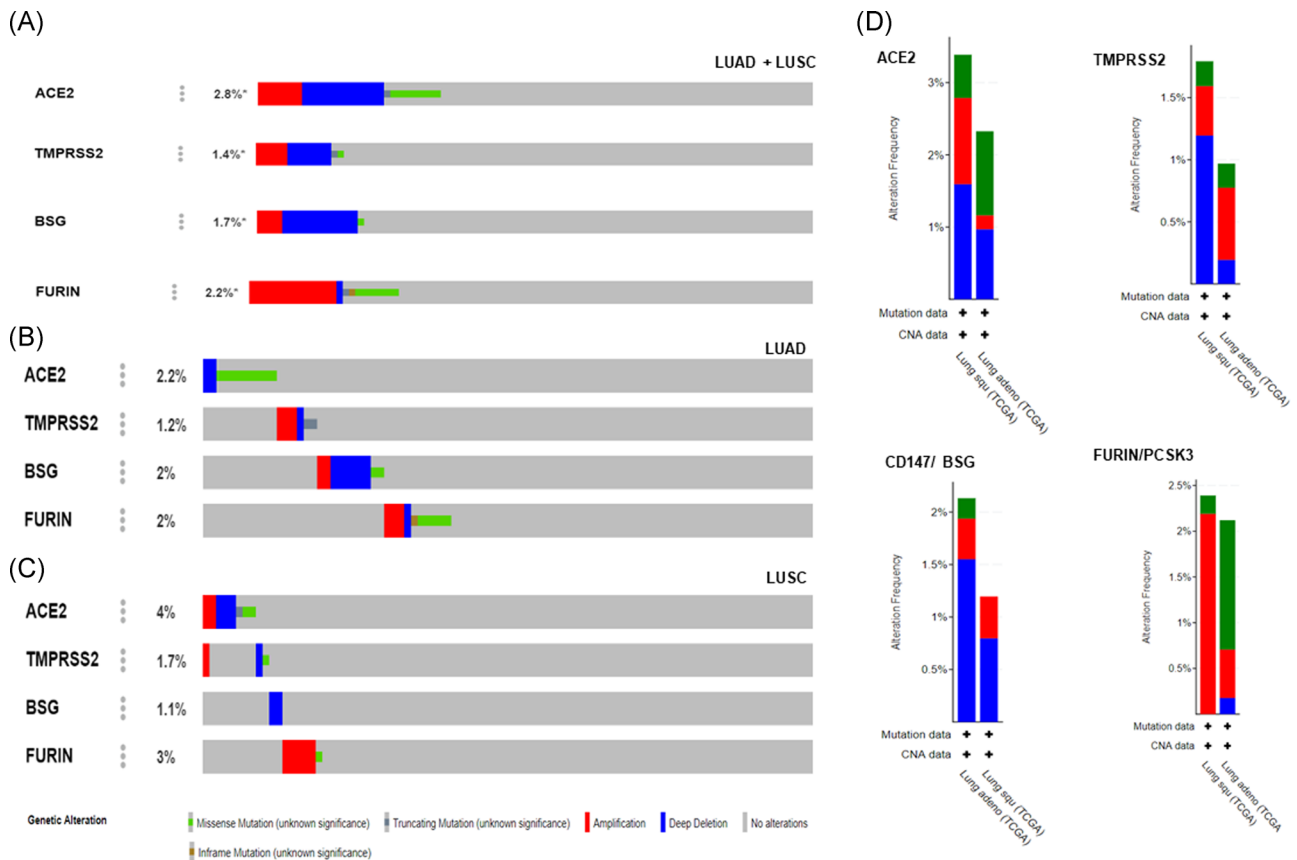
FIGURE 1 A-D, Distribution of genetic alterations in ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes in patients with LUAD and LUSC. ACE2, angiotensin-converting enzyme-2; BSG, basigin; LUAD, lung adenocarcinoma; LUSC, lung squamous carcinoma; TMPRSS2, transmembrane serine protease 2

our results show that the frequency of carrying genetic anomalies is higher for the LUSC group. We detected mutations in sequences encoding important domains of target genes in both types of cancer. In the LUSC patient group, the p.X233_splice site mutation in the ACE2 gene is able to prevent a functional transcript from occurring because it disrupts splicing activity, but these results must be confirmed by wet laboratory studies. In particular, missense mutations in ACE2 and FURIN/PCSK3 genes are likely to be capable of impairing the function of the protein to occur due to the fact that they affect critical amino acids that are conserved among species throughout the evolutionary process. In addition, deep deletions in all four target genes indicate that genes are likely to be homozygous deletions, and in this case, gene expression may be impressive. Gene amplification detected in ACE2 and FURIN/PCSK3 genes can cause uncontrolled and excessive gene expression. Hoffman et al reported that TMPRSS2 showed that SARS-CoV-2 is required for the interaction of S protein with ACE2 receptor and the entry and propagation of SARS-CoV-2 into the host cell. TMPRSS2 p.G19* nonsense mutation is truncating mutation, and we think that TMPRSS2 protein synthesis will cause deficient/immature enzyme formation with the formation of stop codon before the completion of this condition and this may lead to disruption in protein function. Particularly, genetic variants in ACE2

are thought to affect the interaction of RBD on the SARS-CoV-2 spike protein. It is known that a total of 18 amino acids (Q24, T27, K31 H34, E37, D38, Y41, Q42, L45, L79, M82, Y83, N90, Q325, E329, N330, K353, and G354) on ACE2 are the binding sites with RBD. Experimental and bioinformatics based studies are carried out on the fact that the mutations/variants in these localizations will be able to change the binding affinity of SARS-CoV-2. In our study, one missense mutation (p.F275V) was detected on the I-set domain on CD147/BSG, which is defined as a new binding receptor for the SARS-CoV-2 spike protein. It was demonstrated in the enhanced expression level of novel target CD147/BSG in inflammation and cancer progression by other researchers. Ulrich et al[11] reported that inhibition of CD147/BSG may have beneficial effects in the prevention of diabetic complications, involving severe acute respiratory syndrome triggers by COVID-19.

There are several reasons why the lung appears to be the most vulnerable target organ for this virus. The first is that the large surface area of the lung makes the lung quite susceptible to the inhaled viruses, but also a biological factor plays a role as a second cause. Zhao et al[26] showed that 83% of ACE2-expressing cells were alveolar epithelial type II cells (AECII), suggesting that these cells can be used as a reservoir for SARS-CoV-2 invasion. The expression of the ACE2 receptor outside of the lung is known to be found in
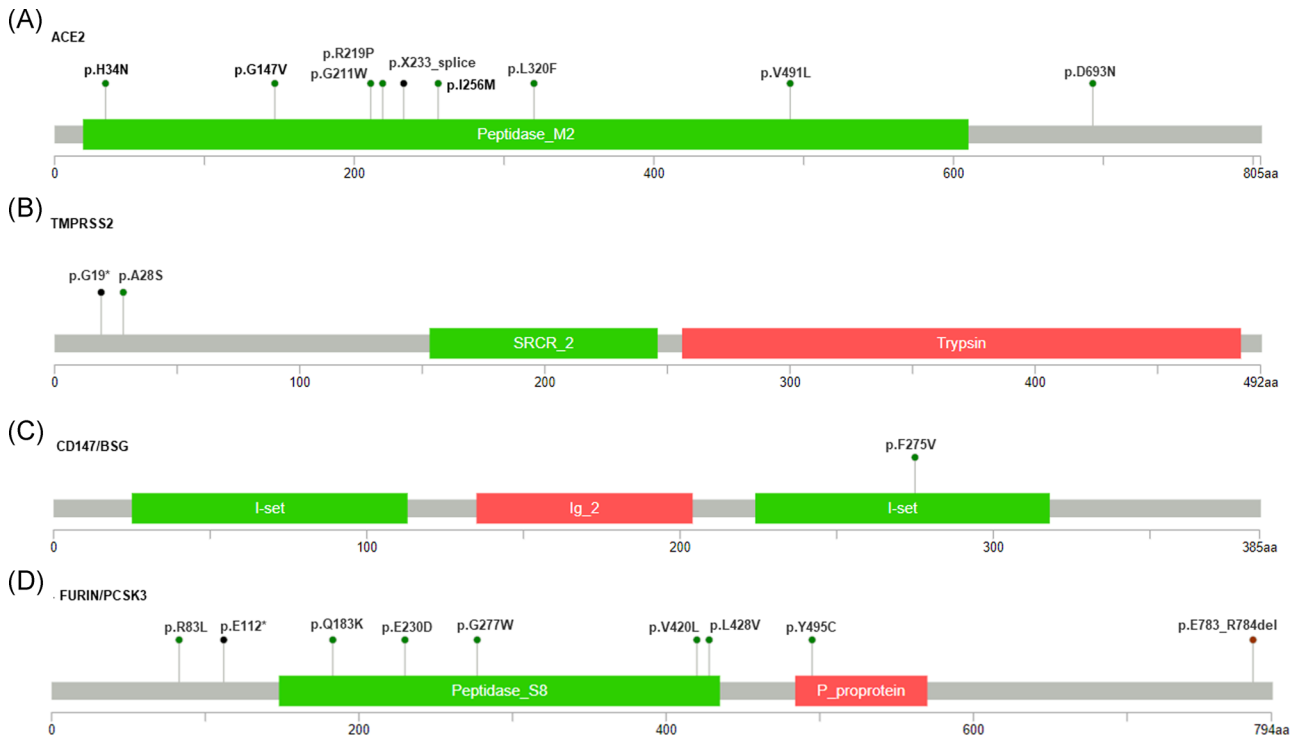
FIGURE 2 Schematic representation of domain architecture of the ACE2, TMPRSS2, CD147, and FURIN/PCSK3 proteins and mutations found in LUAD and LUSC. A, Human ACE2 is a polypeptide of 805 amino acids. B, Human TMPRSS2 is a polypeptide of 492 amino acids. C, Human CD147/BSG is a polypeptide of 385 amino acids. D, Human FURIN/PCSK3 is a polypeptide of 794 amino acids. ACE2, angiotensin-converting enzyme-2; BSG, basigin; LUAD, lung adenocarcinoma; LUSC, lung squamous carcinoma; TMPRSS2, transmembrane serine protease 2
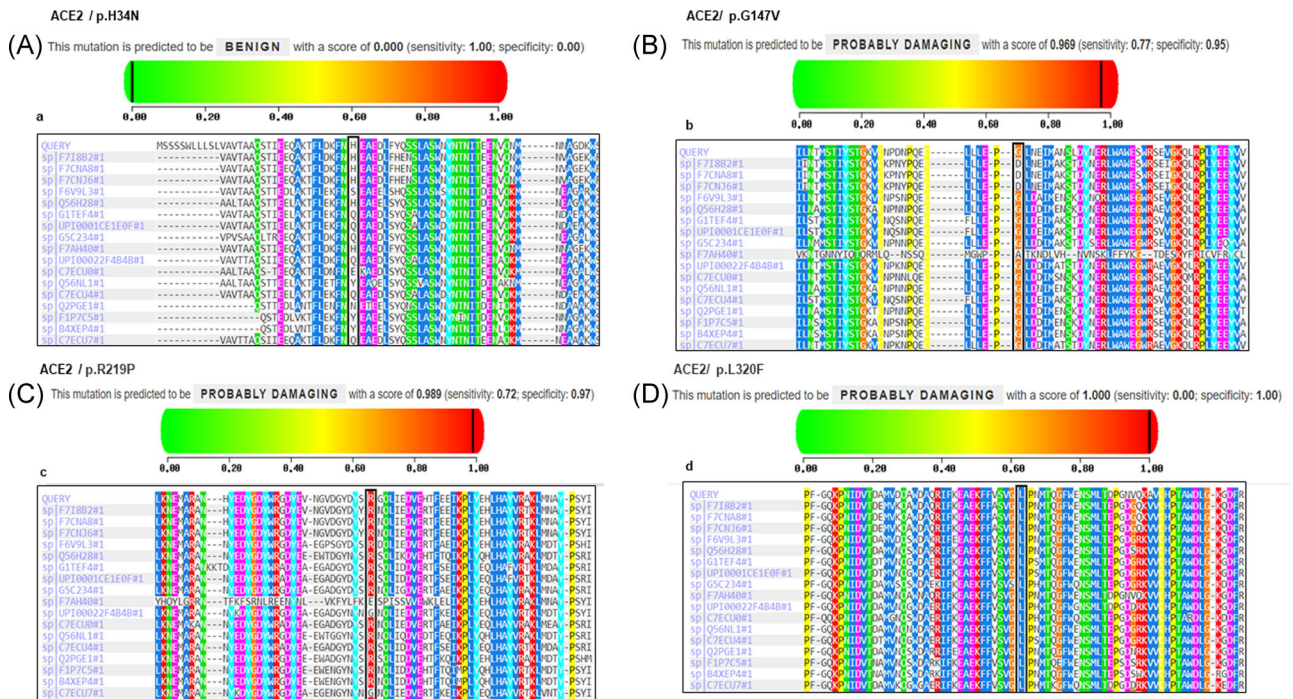
FIGURE 3 A-D, Estimation of possible functional effects of missense mutations in ACE2 gene with PolyPhen-2 program. The evolutionary conservation analyses of the mutated amino acids ACE2 gene in the present study. The mutated amino acids were demonstrated. The detected mutant amino acids were evaluated among different species (a-d). ACE2, angiotensin-converting enzyme-2
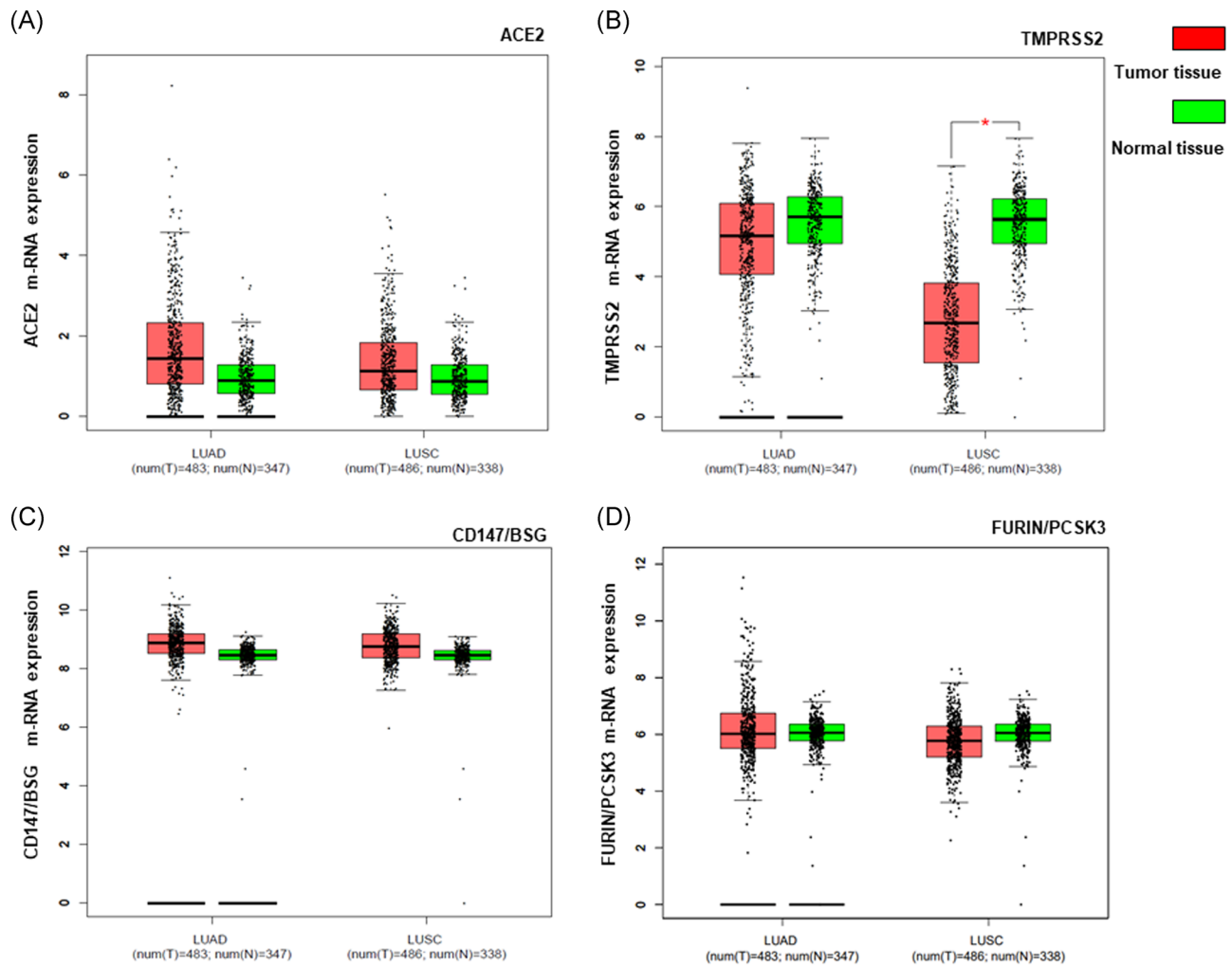
FIGURE 4 Comparative analysis of the tissue-specific differential expression of ACE2, TMPRSS2, CD147/BSG, and FURIN/PCSK3 genes in lung tissues using GEPIA. A, ACE2 gene expression in LUAD and LUSC patients. B, TMPRSS2 gene expression in LUAD and LUSC patients. C, CD147/BSG gene expression in LUAD and LUSC patients. D, FURIN/PCSK3 gene expression in LUAD and LUSC patients. ACE2, angiotensin-converting enzyme-2; BSG, basigin; GEPIA, Gene Expression Profiling Interactive Analysis; LUAD, lung adenocarcinoma; LUSC, lung squamous carcinoma; TMPRSS2, transmembrane serine protease 2. *$P < .05$

many extrapulmonary tissues, including the heart, kidney, endothelium, and intestine.[2,4,27] In our study, according to the expression profile analysis of ACE2, TMPRSS2, CD147/BSG and FURIN/PCSK3 genes in the LUAD and LUSC groups, it was determined that the expression of TMPRSS2 was significantly lower especially in the LUSC patient group, compared to the healthy group and the LUAD group. LUAD can provide the patient group with protective properties against SARS-CoV-2 invasion. ACE2 and CD147/BSG expressions show higher expression in both groups compared to the healthy group, but this enhanced expression is not statistically significant. Some research groups reported that ACE2 expression showed a positive correlation to SARS-CoV-2 infection in experimental studies.[6,7,25,28,29] The gene expression level of ACE2 may indicate that it is susceptible to SARS-CoV-2 infection and that TMPRSS2 plays a supporting role. Furthermore, we conducted a low TMPRSS2 expression resulted in significantly longer overall survival times in LUAD cancer patients.

Therefore, assessment of the TMPRSS2 gene downregulated expression only may be useful for predicting prognosis and susceptibility to COVID-19 in these patient groups.

5 | CONCLUSION

We concluded that increased expression of ACE2 and CD147/BSG and decreased expression of TMPRSS2 in tumor tissues result in more susceptibility to SARS-CoV-2 infection in COVID-19 patients with LUSC and LUAD. Moreover, tumor tissues infected with SARS-CoV-2 may undergo an increase in TMPRSS2, which may worsen the COVID-19 patients with LUSC and LUAD. The impact of genetic variations in lung cancer patients needs to be assessed to effectively reveal potential invasion genes for SARS-CoV-2 in COVID-19 susceptibility. The effect of mutation can be predicted using in silico, which may provide a better understanding of genetic

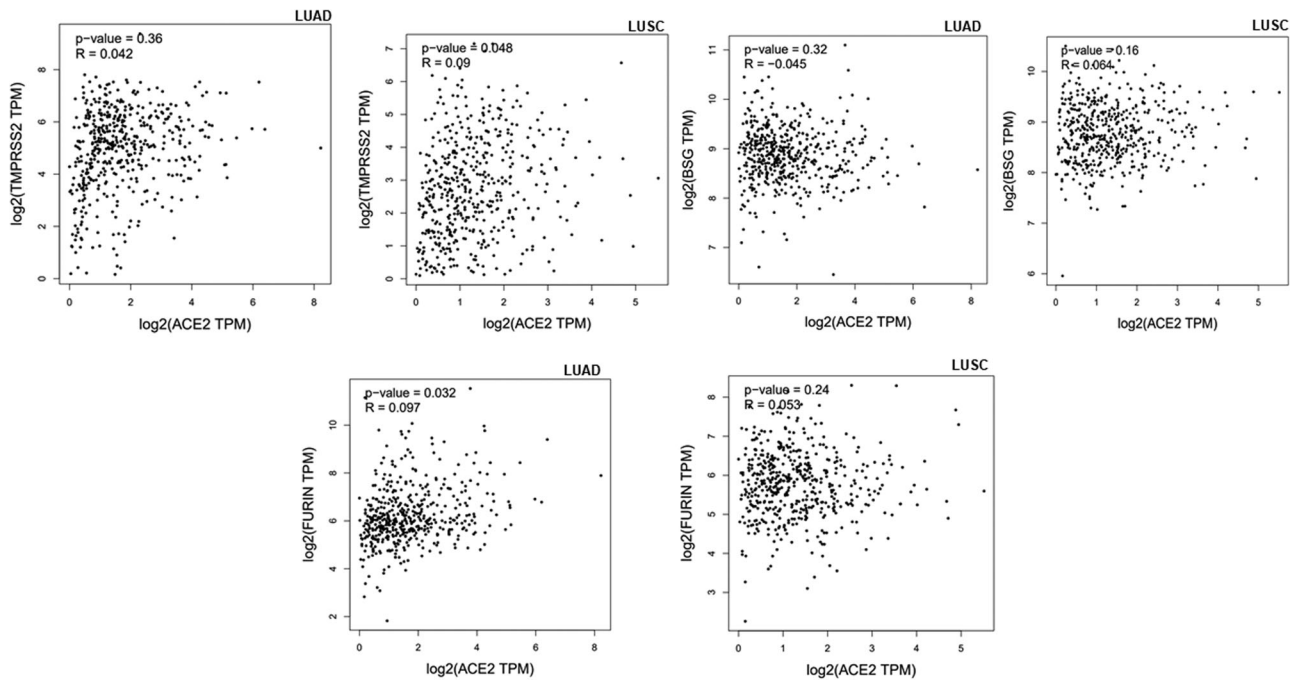**FIGURE 5** Pearson correlation analyses of TMPRSS2, CD147/BSG and FURIN/PCSK3 mRNA expressions with ACE2 expression in LUAD and LUSC patients. ACE2, angiotensin-converting enzyme-2; BSG, basigin; GEPIA, Gene Expression Profiling Interactive Analysis; LUAD, lung adenocarcinoma; LUSC, lung squamous carcinoma; TMPRSS2, transmembrane serine protease 2

variations in lung cancer disease for COVID-19 susceptibility. Advanced bioinformatics methods have facilitated the identification of various types of mutations in the target genes which might be SARS-CoV-2 associated, and enable us to study the structural and functional effect of mutations on target proteins.

Nonetheless, integrating expression and mutation patterns of virus invasion genes with defined correlation may prove highly informative, as highlighted by the successful application of in silico analysis to understand lung cancer subtypes variations in possible response to COVID-19.
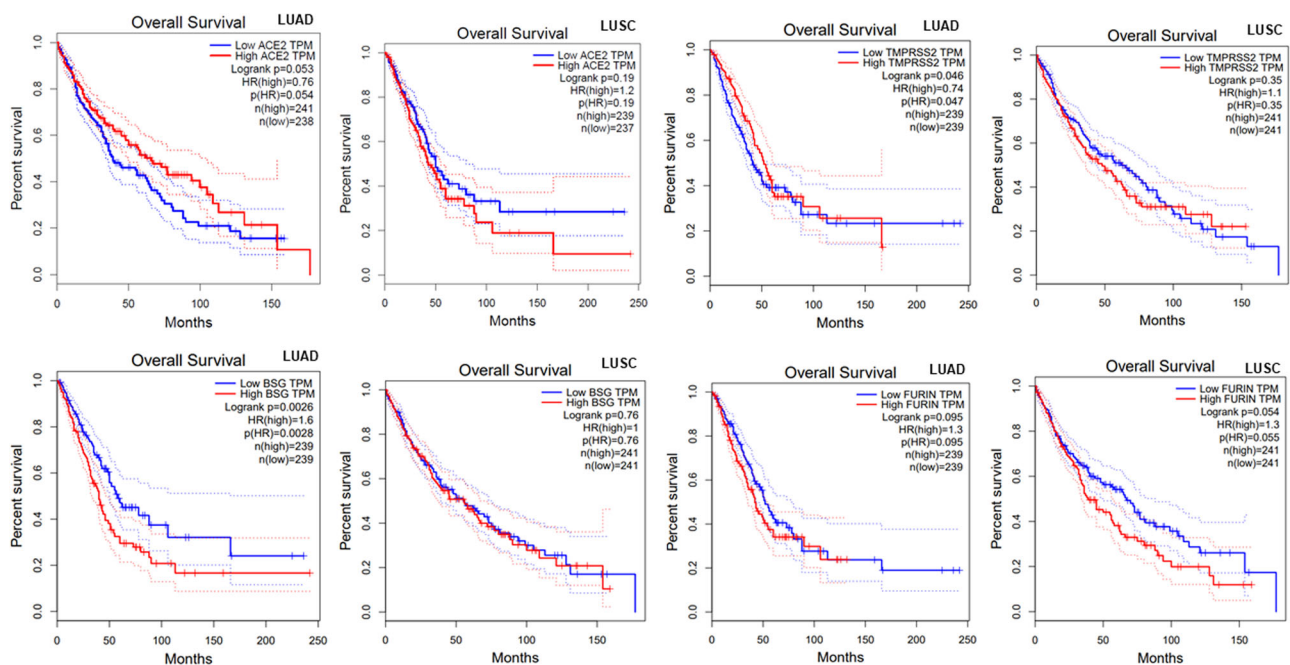

FIGURE 6 Kaplan–Meier analysis of overall survival of LUAD and LUSC patients according to different ACE2, TMPRSS2, CD147/BSG and FURIN/PCSK3 levels. ACE2, angiotensin-converting enzyme-2; BSG, basigin; LUAD, lung adenocarcinoma; LUSC, lung squamous carcinoma; TMPRSS2, transmembrane serine protease 2

ORCID

Rahsan Ilikci Sagkan http://orcid.org/0000-0003-3844-6158
Dilara Fatma Akin-Bali http://orcid.org/0000-0002-0903-0017

REFERENCES

1. Li H, Liu SM, Yu XH, Tang SL, Tang CK. Coronavirus disease 2019 (COVID-19): current status and future perspectives. *Int J Antimicrob Agents*. 2020;55:105951.
2. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol*. 2020 Apr;92(4):418-423.
3. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581:215-220.
4. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med*. 2020;46(4):586-590.
5. Kumar S, Maurya VK, Prasad AK, Bhatt MLB, Saxena SK. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *Virusdisease*. 2020;31(1):13-21.
6. Hoffmann M, Kleine-Weber H, Schroeder 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):271-280.e8.
7. Lukassen S, Chua RL, Trefzer T, et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J*. 2020:e105114.
8. 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.
9. Hasan A, Paray BA, Hussain A, et al. A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *J Biomol Struct Dyn*. 2020:1-9.
10. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res*. 2020;176:104742.
11. Ulrich H, Pillat MM. CD147 as a target for COVID-19 treatment: suggested effects of azithromycin and stem cell engagement. *Stem Cell Rev Rep*. 2020;16:434-440.
12. Moujaess E, Kourie HR, Ghosn M. Cancer patients and research during COVID-19 pandemic: a systematic review of current evidence. *Crit Rev Oncol Hematol*. 2020;150:102972.
13. Hussain M, Jabeen N, Raza F, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2spike protein. *J Med Virol*. 2020.
14. 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):281-292.e6.
15. Gheblawi M, Wang K, Viveiros A, et al. Angiotensin converting enzyme 2: SARS-CoV-2 receptor and regulator of the renin-angiotensin system. *Circ Res*. 2020;126:1456-1474.
16. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
17. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health*. 2019;22:85. pii: 8.
18. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.
19. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:7. Unit7.20.
20. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. 2007;35:3823-3835.
21. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47(D1):D941-D947.
22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98-W102.
23. Chen YW, Lee MS, Lucht A, et al. TMPRSS2, a serine protease expressed in the prostate on the apical surface of luminal epithelial cells and released into semen in prostasomes, is misregulated in prostate cancer cells. *Am J Pathol*. 2010;176(6):2986-2996.
24. https://www.who.int/emergencies/diseases/novel-coronavirus-2019
25. Li G, He X, Zhang L, et al. Assessing ACE2 expression patterns in lung tissues in the pathogenesis of COVID-19. *J Autoimmun*. 2020:102463.
26. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan COVID-19. *bioRxiv*. 2020.
27. Yan T, Xiao R, Lin G. Angiotensin-converting enzyme 2 in severe acute respiratory syndrome coronavirus and SARS-CoV-2: a double-edged sword? *FASEB J*. 2020;34:6017-6026.
28. Li MY, Li L, Zhang Y, Wang XS. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect Dis Poverty*. 2020;9(1):45.
29. Zhang Q, Lu S, Li T, et al. ACE2 inhibits breast cancer angiogenesis via suppressing the VEGFa/VEGFR2/ERK pathway. *J Exp Clin Cancer Res*. 2019;38(1):173.