

## Genomic, epigenomic, and immune subtype analysis of CTSL/B and SARS-CoV-2 receptor ACE2 in pan-cancer

Huimin Li<sup>1,2,\*</sup>, Longxiang Xie<sup>1,\*</sup>, Lin Chen<sup>1,\*</sup>, Lu Zhang<sup>1</sup>, Yali Han<sup>1</sup>, Zhongyi Yan<sup>1</sup>, Xiangqian Guo<sup>1</sup>

<sup>1</sup>Institute of Biomedical Informatics, Cell Signal Transduction Laboratory, Bioinformatics Center, Henan Provincial Engineering Center for Tumor Molecular Medicine, School of Basic Medical Sciences, Henan University, Kaifeng 475004, China

<sup>2</sup>Department of Histology and Embryology, Henan International Joint Laboratory for Nuclear Protein Regulation, School of Basic Medical Sciences, Henan University, Kaifeng 475004, China

\*Equal contribution

Correspondence to: Xiangqian Guo; email: [xqguo@henu.edu.cn](mailto:xqguo@henu.edu.cn)

Keywords: SARS-CoV-2, CTSL/B

Received: July 10, 2020

Accepted: September 21, 2020

Published: November 20, 2020

Copyright: © 2020 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

SARS-coronavirus 2 (SARS-CoV-2) has been spreading widely and posing an international challenge for both healthcare and society. At present, cancer has been identified as an individual risk factor for COVID-19. Angiotensin converting enzyme 2 (ACE2) and Cathepsin L/Cathepsin B (CTSL/B), which act as the receptor and entry-associated proteases of SARS-CoV-2 respectively, are pivotal for SARS-CoV-2 infection. To investigate the possible SARS-CoV-2 infection risk of pan-cancer, we analyzed the genetic alterations, RNA expression, DNA methylation, and the association with immune subtypes of ACE2 and CTSL/B with the prognosis in pan-cancer. Results showed the upregulation of CTSL/B and ACE2 in Pancreatic adenocarcinoma (PAAD) and Stomach adenocarcinoma (STAD) and demonstrated a positive correlation between copy number alteration (CNA) and gene expression for CTSB in PAAD and STAD. Hypomethylation and a negative correlation of gene expression and methylation for CTSB were detected in PAAD. In addition, ACE2 and CTSL/B are overexpressed in the IFN-gamma immune subtype of ovarian serous Cystadenocarcinoma (OV), Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), and Bladder urothelial carcinoma (BLCA). Our study presents a bioinformatics assessment for the potential risk of SARS-CoV-2 infection in pan-cancer.

### TO THE EDITOR

The worldwide spread of COVID-19 is posing a global challenge for both healthcare and society. The disease caused by SARS-CoV-2 has symptoms similar to those caused by SARS-CoV and MERS-CoV. To date, more than 10 million people worldwide have been diagnosed with COVID-19, which has caused over 550,000 deaths. Since COVID-19 started to spread in the world, patients with cancer were designated as a particularly vulnerable sub-group of the population [1]. At present, the COVID-19 and Cancer Consortium (CCC19), a consortium of over 120 cancer centers and other organizations that have

come together to collect data about patients with cancer who have been diagnosed with COVID-19, included patients with several types of malignancies (gastro-intestinal, 12%; breast cancer, 21%; thoracic cancers, 10%; hematological malignancies, 22%; and others, 19%) [1, 2]. SARS-Cov-2 entry into a cell is dependent upon the cellular membrane receptor ACE2 which binds the viral spike (S) protein, and the host cell proteases such as CTSL/B which cleave and activate the viral spike (S) protein. Coronaviruses utilize the viral spike (S) protein to bind to their cellular receptors. Such binding and cleavage of the S protein into two subunits named S1 and S2 by host proteases lead to fusion between the cell and

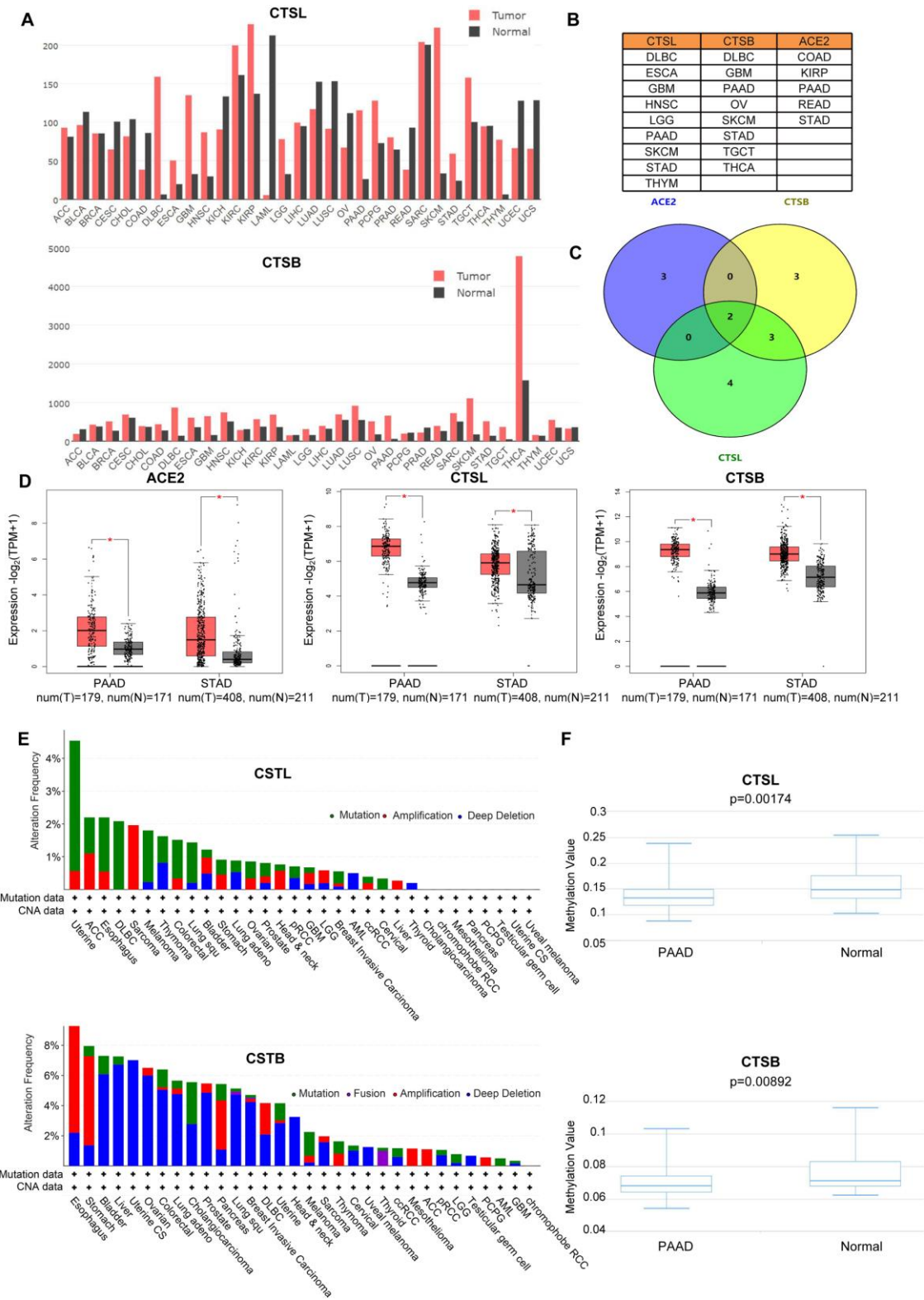
viral membranes and viral activation for viral infection. SARS-CoV-2 uses the ACE2 receptor for entry [3]. The S1/S2 cleavage site of SARS-CoV-2 is between the threonine and methionine at positions 696 and 697, and is identical to that of SARS-CoV which has been shown to be cleaved by CTSL [4]. SARS-CoV takes advantage of the endosomal cysteine proteases CTSL and CTSB. In healthy individuals those organs including lung, heart, kidney, bladder and oesophagus which express a high level of ACE2 seem to be more vulnerable to SARS-CoV-2 infection [5]. Hence, the expression and distribution of ACE2 and CTSL/B may explain the different susceptibility to SARS-Cov-2 infection for pan-cancer. To address this possibility, we performed a comprehensive analysis of CTSL/B and ACE2 expression by using genomic, transcriptomic and epigenomic data of pan-cancer so as to provide a reference for the assessment of SARS-CoV-2 infection in pan-cancer.

Based on the transcriptional level analysis in GEPIA [6], we found an elevation of CTSL in nine types of tumors including Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), Esophageal carcinoma (ESCA), Glioblastoma multiforme (GBM), Head and Neck squamous cell carcinoma (HNSC), LGG, PAAD, Skin Cutaneous Melanoma (SKCM), STAD and Thymoma (THYM) (Figure 1A), and an increase in CTSB in DLBC, GBM, OV, PAAD, SKCM, STAD, Testicular Germ Cell Tumors (TGCT) and Thyroid carcinoma (THCA) (Figure 1A, 1B). Notably, ACE2 is upregulated in Colon adenocarcinoma (COAD), Kidney renal papillary cell carcinoma (KIRP), PAAD, Rectum adenocarcinoma (READ) and STAD (Supplementary Figure 1A, 1B). Further analysis revealed that ACE2 and CTSL/B are all highly expressed in PAAD and STAD (Figure 1C, 1D). Oncomine analysis [7] verified the upregulation ( $p < 0.05$ , fold change  $> 1.5$ ) of CTSL/B in PAAD and STAD when compared to normal tissues (Supplementary Tables 1, 2). With respect to ACE2 in the Oncomine analysis, the small sample size likely contributes to the lack of significant over-expression in PAAD and STAD as compared to normal tissues.

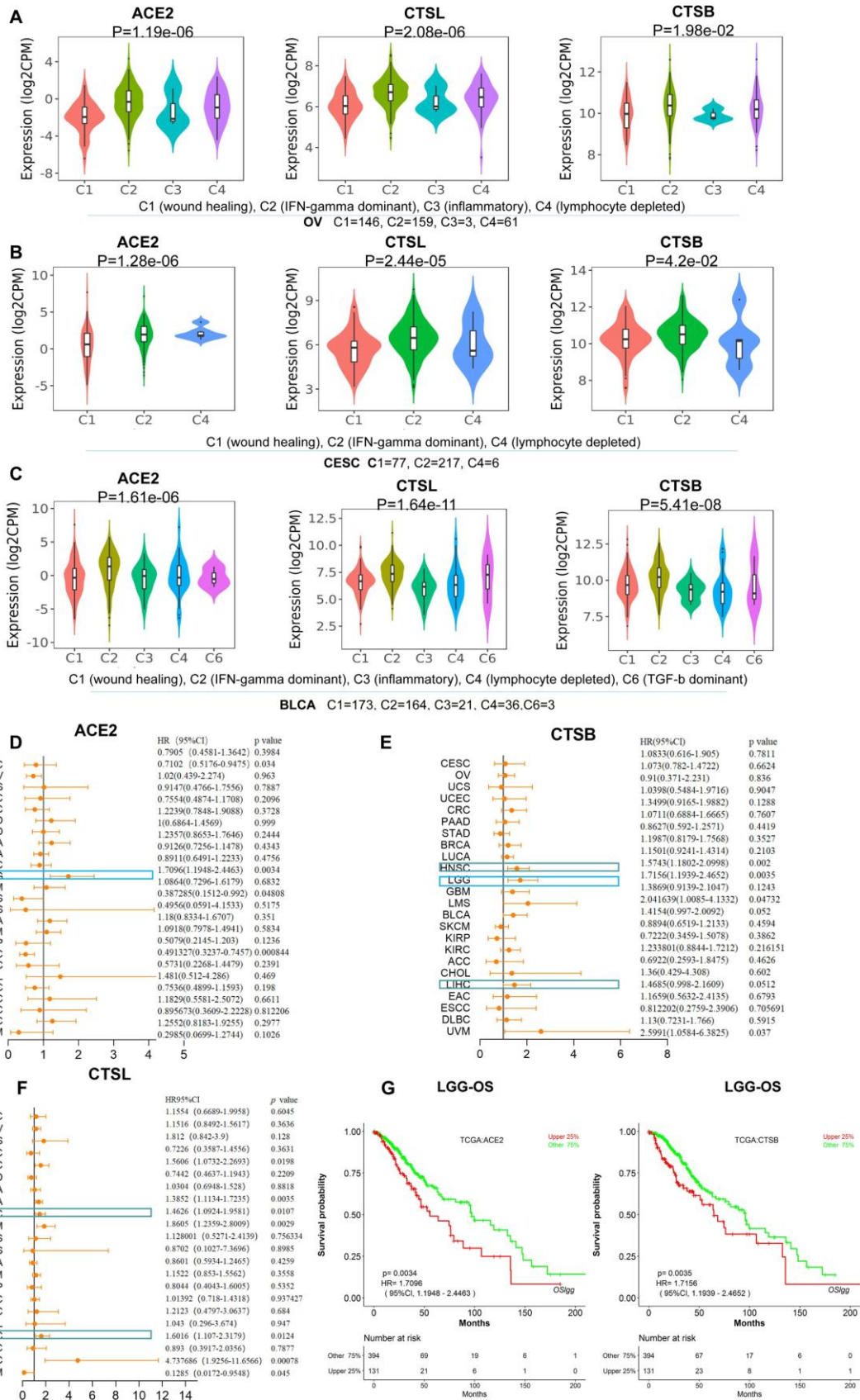
CNA, mutation and DNA methylation can prompt the variation of gene expression. CNA includes amplification, gain, diploidy, shallow deletion and deep deletion. c-BioPortal analysis [8] based on The Cancer Genome Atlas (TCGA) data indicates that genomic amplification increased the gene expression of CTSB in STAD and PAAD (Figure 1E and Supplementary Figure 2,  $R > 0.2$ ,  $p < 0.05$ ). For CTSL in STAD, its expression could be elevated by genomic amplification as well (Figure 1E). For ACE2, DNA mutation acts as the dominant factor in regulating gene expression in STAD

and PAAD (Supplementary Figure 1B). However, the correlation analysis between gene expression and CNA/mutation showed neither ACE2 nor CTSL expression is relevant to mutation or CNA (Supplementary Figures 3, 4). For other types of tumors, the most frequent DNA alteration of the CTSL gene is DNA mutation in Uterine Corpus Endometrial Carcinoma (UCEC), ESCA, DLBC and Melanoma (Figure 1E), and followed by amplification in Sarcoma (SARC), Adrenocortical carcinoma (ACC), UCEC and ESCA. For CTSB, deep deletion in Uterine Carcinosarcoma (UCS), Liver hepatocellular carcinoma (LIHC), BLCA and OV, is the most frequent DNA alteration, followed by amplification in ESCA, STAD, PAAD and DLBC. For ACE2, the most frequent DNA alteration is mutation in UCEC, UCS, STAD and MEL (Supplementary Figure 1B). To explore the influence of epigenetic modification on the mRNA levels for CTSL/B and ACE2 in PAAD and STAD, seven, fourteen and six CpG probes targeting the promoters of the CTSL, CTSB and ACE2 genes, respectively, using Methylation450k profiles were utilized to recapitulate the DNA methylation levels (Supplementary Table 3) in the DiseaseMeth database [9]; the results demonstrated that DNA methylation of CTSL/B is significantly decreased in PAAD but not in STAD when compared with normal tissues (Figure 1F and Supplementary Figure 5). However, the analysis based on DNMIIVD [10] showed that gene expression of CTSB rather than CTSL and ACE2 is negatively correlated with DNA methylation in PAAD and STAD (Supplementary Tables 4–6,  $R < -0.1$ ,  $p < 0.05$ ). Together, the analyses suggest that both DNA methylation and CNA could influence CTSB gene expression in PAAD, while only CNA can affect CTSB gene expression in STAD. For proteomic analysis, we analyzed the expression variation of ACE2 and CTSL/CTSB with TCGA proteomics data, which includes measuring the expression of 261 proteins; unfortunately, these three proteins (ACE2, CTSL and CTSB) were not included in the 261 proteins.

We further studied the involvement of ACE2 and CTSL/B in cancer immune subtypes including C1 (wound healing), C2 (IFN-gamma dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet) and C6 (TGF-b dominant) in pan-cancer (Supplementary Figure 6) according to TISIDB [11]. As compared with other types of tumors, BRCA, HNSC, OV, KIRC, CESC and BLCA present greater differential ACE2 expression. Combining the immune subtype analysis of ACE2 and CTSL/B in the six tumor types, we found that ACE2 and CTSL/B are most highly overexpressed in the IFN-gamma dominant immune subtype of OV, CESC and BLCA (Figure 2A–2C), yet not in BRCA, HNSC and KIRC (Supplementary Figure 7). The results indicated that the



**Figure 1. The analyses of gene variation and epigenetics of CTSL/B and ACE2 in pan-cancer. (A)** The distribution of differential expression of CTSL/B in 31 types of cancers and adjacent tissues based on TCGA and Genotype-Tissue Expression (GTEx) data (GEPiA). **(B)** The exhibition of tumors upregulating CTSL/B and ACE2, respectively. **(C)** The overlap of tumors overexpressing ACE2, CTSL/B using venny 2.0.2. **(D)** The significant difference analysis of CTSL/B in PAAD and STAD. **(E)** The mutation and CNA distribution of CTSL/B in pan-cancer (cBioportal). **(F)** The statistical difference graphs of DNA methylation for CTSB/L in PAAD (DiseaseMeth version 2.0).  $p < 0.05$  was regarded as significant.



**Figure 2. The identification of immune subtypes and prognosis for CTSL/B and ACE2 in pan-cancer. (A–C)** The distribution graphs of ACE2 and CTSL/B in six immune subtypes in OV, CESC and BLCA. ACE2 and CTSL/B in OV, CESC and BLCA were most highly expressed in the

C2 named IFN-gamma dominant subtype. (D–F) Forest maps analysis of overall survival for ACE2 and CTSL/B in pan-cancer (LOGpc). The blue boxes represent the overlap of tumors in which ACE2 and CTSB can predict adverse prognosis; the green boxes represent the overlap of tumors in which CTSL and CTSB can predict adverse prognosis. (G) The overall survival analyses of ACE2 and CTSB in LGG utilizing the LOGpc online tool based on TCGA data, with results showing that ACE2 and CTSB both predict poor prognosis.  $p < 0.05$  was regarded as significant;  $HR > 1$ ,  $p < 0.05$  represented poor prognosis.

IFN-gamma dominant immune subtype of OV, CESC and BLCA may provide more opportunities than other immune subtypes for SARS-CoV-2 infection.

To explore the prognostic values of CTSL/B and ACE2, we analyzed gene expression profiling and long-term follow-up data from TCGA using LOGpc online tools and discovered that upregulated CTSL and CTSB both predict poor prognosis in HNSC and LIHC (Supplementary Figure 8). Respectively, CTSL predicts poor overall survival (OS) in six types of tumors including Colorectal cancer (CRC), Lung cancer (LUCA), HNSC, GBM, LIHC and ESCC; overexpression of CTSB predicts unfavorable OS in four types of tumors including HNSC, LGG, Leiomyosarcoma (LMS) and LIHC (Figure 2D–2F). ACE2 predicts poor prognosis only in LGG in which CTSB also portends poor prognosis (Figure 2G). Although the three genes did not concurrently predict poor prognosis in a given type of tumor, the prognostic roles of two of the three pivotal factors for COVID-19 infection may be of value in predicting COVID-19 infection risk in cancers. Hence, LGG, HNSC and LIHC may indicate a greater risk for COVID-19 infection than other types of cancers, but needs further clinical validation. Yu-Jun Dai et al. also emphasized that LIHC patients with high expression level of ACE2 should be more cautious of virus infection. Both studies may provide potential clues for preventing infection of SARS-CoV-2 in cancers [5].

CTSL/B have been reported to be a vital regulators in proliferation, metastasis [12], invasion [13] and prognosis [14] in various types of cancers. For instance, Cathepsins B and L could drive the invasive growth of human melanoma cells [15], and overexpression of CTSL is a marker of invasion and metastasis in ovarian cancer [16], together manifesting a potential capacity for the two proteases CTSL/B as malignant phenotype markers for pan-cancer. In addition, the finding that ACE2 overexpression presents a poor prognosis in LGG, consistent with previous reports [17].

Regarding COVID patients with malignancies, LUAD patients with overexpressed ACE2 were found to have a higher incidence of COVID-19 [18]. Herein, we showed that both ACE2 and CTSL/B are upregulated in PAAD and STAD and are overexpressed in IFN-gamma immune subtypes of OV, CESC and BLCA, implying an infection risk for SARS-CoV-2 in PAAD and STAD,

and for IFN-gamma dominance in OV, CESC and BLCA. To clarify these results derived from database analyses, further verification in a large clinical cohort is essential.

## Abbreviations

ACC: Adrenocortical carcinoma; BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: Cholangio carcinoma; COAD: Colon adenocarcinoma; CRC: colorectal Cancer; DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; EAC: Esophageal Adenocarcinoma; ESCA: Esophageal carcinoma; ESCC: esophageal squamous cell carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LAML: Acute Myeloid Leukemia; LGG: Brain Lower Grade Glioma; LIHC: Liver hepatocellular carcinoma; LMS: Leiomyosarcoma; LUAD: Lung adenocarcinoma; LUCA: Lung cancer; LUSC: Lung squamous cell carcinoma; MESO: Mesothelioma; MFS: Myxofibrosarcoma; OV: Ovarian serous cystadenocarcinoma; OS: Osteosarcoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SARC: Sarcoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach Adenocarcinoma; TGCT: Testicular Germ Cell Tumors; THCA: Thyroid carcinoma; THYM: Thymoma; UCEC: Uterine Corpus Endometrial Carcinoma; UCS: Uterine Carcinosarcoma; UVM : Uveal Melanoma.

## AUTHOR CONTRIBUTIONS

XG and HL and LX designed the study; HL and LX performed the analysis and drafted the manuscript; LC and YH collected the data; LZ and ZY modified the manuscript. All the authors approved this manuscript.

## CONFLICTS OF INTEREST

This manuscript has not been published and is not under consideration for publication elsewhere. We have no conflicts of interest to disclose.

## FUNDING

This study was supported by National Natural Science Foundation of China (No.81602362), Program for Innovative Talents of Science and Technology in Henan Province (No. 18HASTIT048), Henan Province Scientific and Technology Research Project (202102310063).

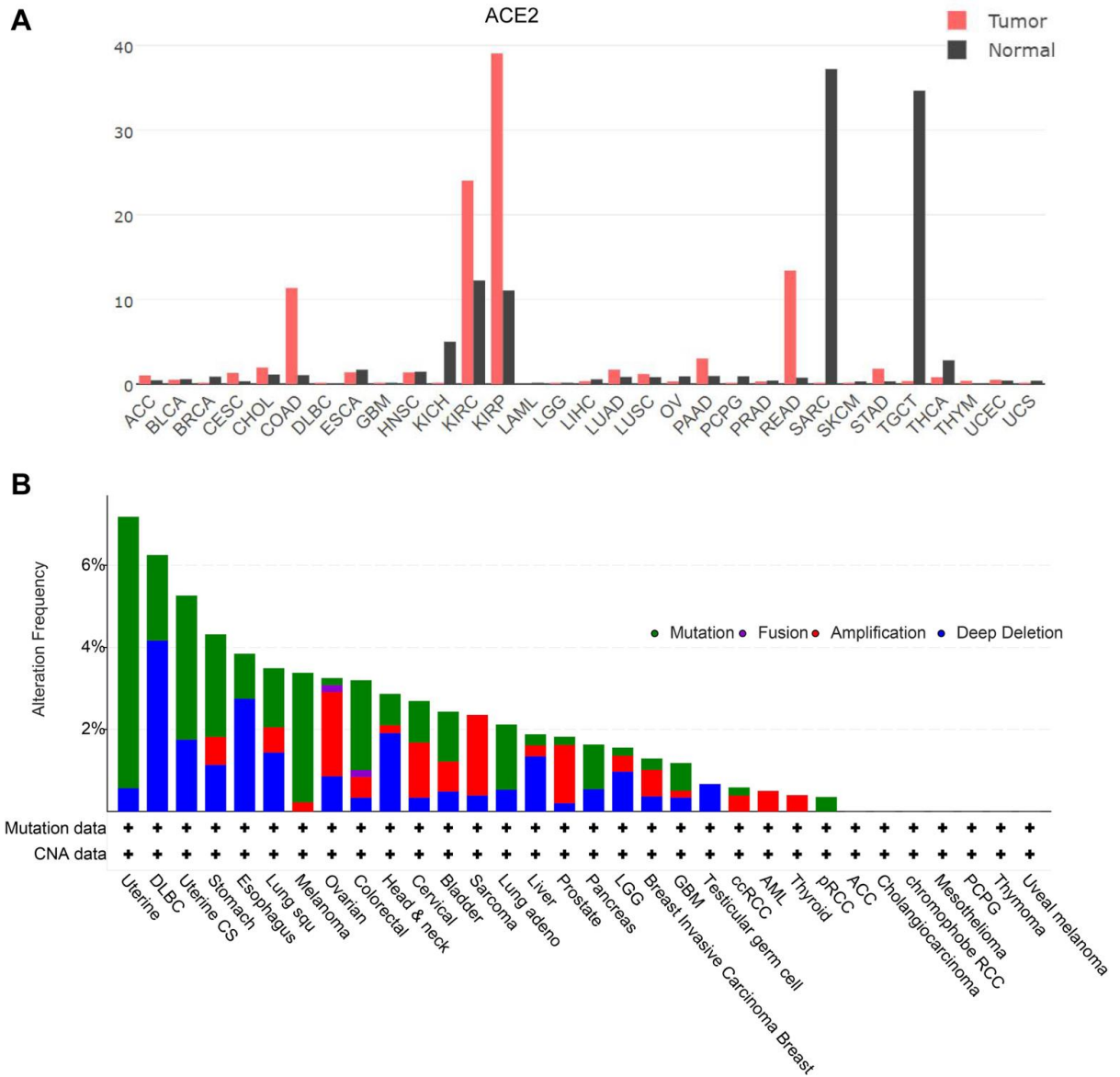
## REFERENCES

1. Curigliano G. Cancer patients and risk of mortality for COVID-19. *Cancer Cell*. 2020; 38:161–63. <https://doi.org/10.1016/j.ccell.2020.07.006> PMID:[32710820](https://pubmed.ncbi.nlm.nih.gov/32710820/)
2. Kuderer NM, Choueiri TK, Shah DP, Shyr Y, Rubinstein SM, Rivera DR, Shete S, Hsu CY, Desai A, de Lima Lopes G Jr, Grivas P, Painter CA, Peters S, et al, and COVID-19 and Cancer Consortium. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet*. 2020; 395:1907–18. [https://doi.org/10.1016/S0140-6736\(20\)31187-9](https://doi.org/10.1016/S0140-6736(20)31187-9) PMID:[32473681](https://pubmed.ncbi.nlm.nih.gov/32473681/)
3. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020; 581:215–20. <https://doi.org/10.1038/s41586-020-2180-5> PMID:[32225176](https://pubmed.ncbi.nlm.nih.gov/32225176/)
4. Smieszek SP, Przychodzen BP, Polymeropoulos MH. Amantadine disrupts lysosomal gene expression: a hypothesis for COVID19 treatment. *Int J Antimicrob Agents*. 2020; 55:106004. <https://doi.org/10.1016/j.ijantimicag.2020.106004> PMID:[32361028](https://pubmed.ncbi.nlm.nih.gov/32361028/)
5. Dai YJ, Hu F, Li H, Huang HY, Wang DW, Liang Y. A profiling analysis on the receptor ACE2 expression reveals the potential risk of different type of cancers vulnerable to SARS-CoV-2 infection. *Ann Transl Med*. 2020; 8:481. <https://doi.org/10.21037/atm.2020.03.61> PMID:[32395525](https://pubmed.ncbi.nlm.nih.gov/32395525/)
6. 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:W98–102. <https://doi.org/10.1093/nar/gkx247> PMID:[28407145](https://pubmed.ncbi.nlm.nih.gov/28407145/)
7. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6:1–6. [https://doi.org/10.1016/s1476-5586\(04\)80047-2](https://doi.org/10.1016/s1476-5586(04)80047-2) PMID:[15068665](https://pubmed.ncbi.nlm.nih.gov/15068665/)
8. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401–04. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID:[22588877](https://pubmed.ncbi.nlm.nih.gov/22588877/)
9. Xiong Y, Wei Y, Gu Y, Zhang S, Lyu J, Zhang B, Chen C, Zhu J, Wang Y, Liu H, Zhang Y. DiseaseMeth version 2.0: a major expansion and update of the human disease methylation database. *Nucleic Acids Res*. 2017; 45:D888–95. <https://doi.org/10.1093/nar/gkw1123> PMID:[27899673](https://pubmed.ncbi.nlm.nih.gov/27899673/)
10. Ding W, Chen J, Feng G, Chen G, Wu J, Guo Y, Ni X, Shi T. DNMIVD: DNA methylation interactive visualization database. *Nucleic Acids Res*. 2020; 48:D856–62. <https://doi.org/10.1093/nar/gkz830> PMID:[31598709](https://pubmed.ncbi.nlm.nih.gov/31598709/)
11. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SS, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics*. 2019; 35:4200–02. <https://doi.org/10.1093/bioinformatics/btz210> PMID:[30903160](https://pubmed.ncbi.nlm.nih.gov/30903160/)
12. Liu P, Kong L, Jin H, Wu Y, Tan X, Song B. Differential secretome of pancreatic cancer cells in serum-containing conditioned medium reveals CCT8 as a new biomarker of pancreatic cancer invasion and metastasis. *Cancer Cell Int*. 2019; 19:262. <https://doi.org/10.1186/s12935-019-0980-1> PMID:[31632196](https://pubmed.ncbi.nlm.nih.gov/31632196/)
13. Rafn B, Nielsen CF, Andersen SH, Szyniarowski P, Corcelle-Termeau E, Valo E, Fehrenbacher N, Olsen CJ, Daugaard M, Egebjerg C, Bøttzauw T, Kohonen P, Nylandsted J, et al. ErbB2-driven breast cancer cell invasion depends on a complex signaling network activating myeloid zinc finger-1-dependent cathepsin B expression. *Mol Cell*. 2012; 45:764–76. <https://doi.org/10.1016/j.molcel.2012.01.029> PMID:[22464443](https://pubmed.ncbi.nlm.nih.gov/22464443/)
14. Jain M, Bakhshi S, Shukla AA, Chauhan SS. Cathepsins B and L in peripheral blood mononuclear cells of pediatric acute myeloid leukemia: potential poor prognostic markers. *Ann Hematol*. 2010; 89:1223–32. <https://doi.org/10.1007/s00277-010-1012-3> PMID:[20567828](https://pubmed.ncbi.nlm.nih.gov/20567828/)
15. Yin M, Soikkeli J, Jahkola T, Virolainen S, Saksela O, Hölttä E. TGF- $\beta$  signaling, activated stromal fibroblasts, and cysteine cathepsins B and L drive the invasive

- growth of human melanoma cells. *Am J Pathol.* 2012; 181:2202–16.  
<https://doi.org/10.1016/j.ajpath.2012.08.027>  
PMID:[23063511](https://pubmed.ncbi.nlm.nih.gov/23063511/)
16. Sui H, Shi C, Yan Z, Wu M. Overexpression of Cathepsin L is associated with chemoresistance and invasion of epithelial ovarian cancer. *Oncotarget.* 2016; 7:45995–6001.  
<https://doi.org/10.18632/oncotarget.10276>  
PMID:[27351223](https://pubmed.ncbi.nlm.nih.gov/27351223/)
17. 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; 13:43.  
<https://doi.org/10.1186/s13045-020-00883-5>  
PMID:[32366279](https://pubmed.ncbi.nlm.nih.gov/32366279/)
18. 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; 19:80.  
<https://doi.org/10.1186/s12943-020-01209-2>  
PMID:[32345328](https://pubmed.ncbi.nlm.nih.gov/32345328/)

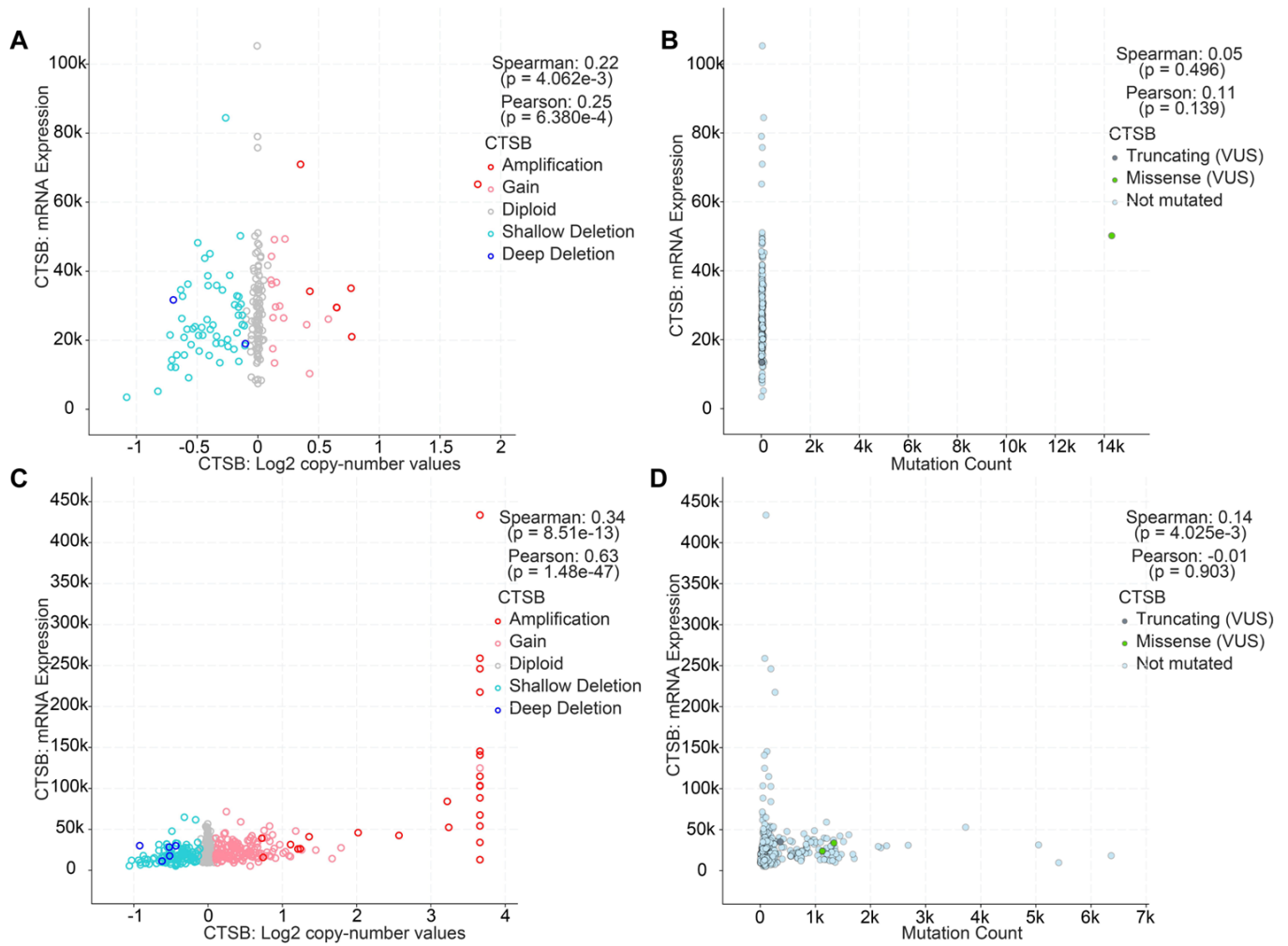
SUPPLEMENTARY MATERIALS

Supplementary Figures

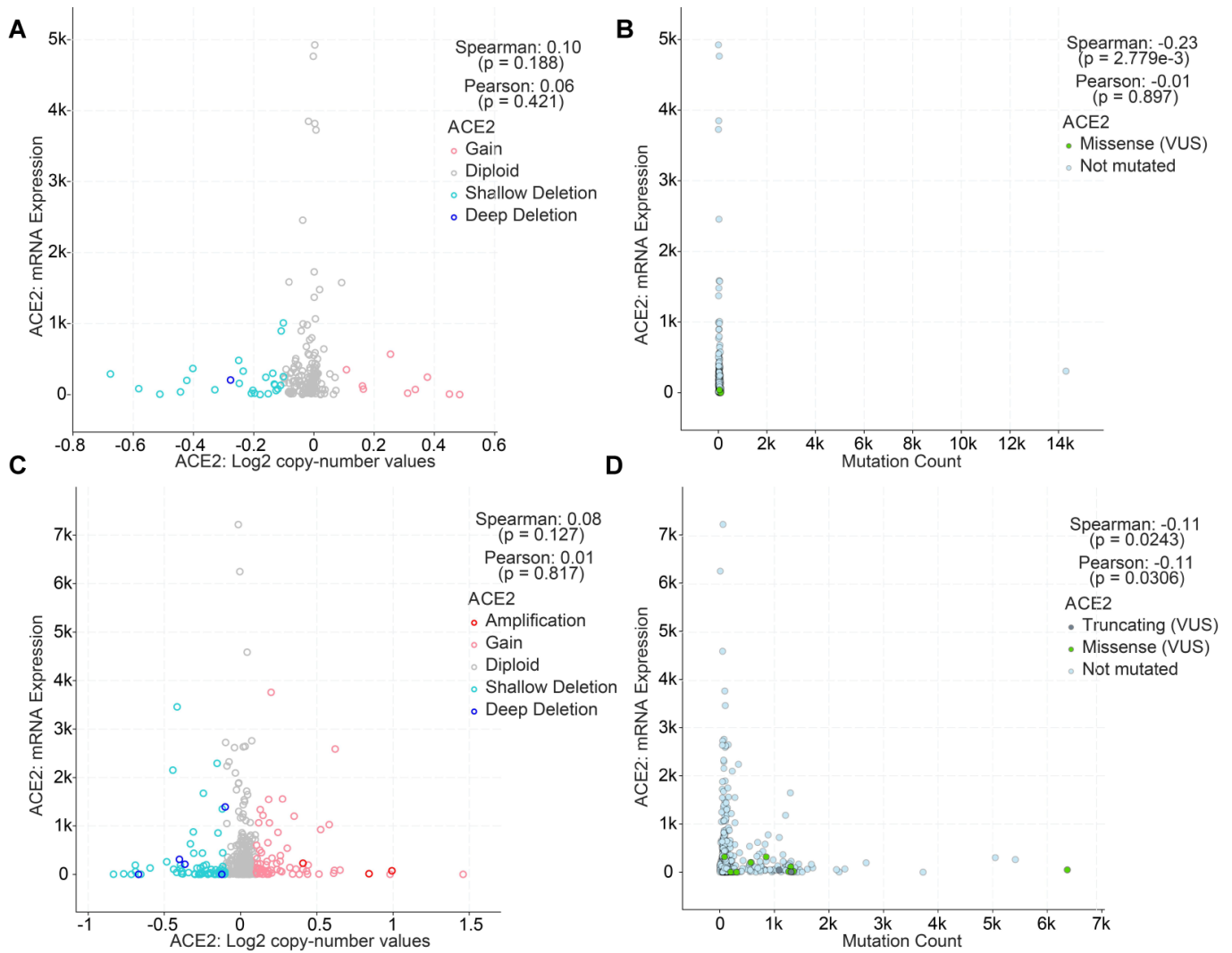


**Supplementary Figure 1.** (A) The distribution of different expression of ACE2 in 31 types of cancers and adjacent normal based on TCGA and GTex data (GEPIA). (B) The mutation and CNA distribution of ACE2 in pan-cancer.

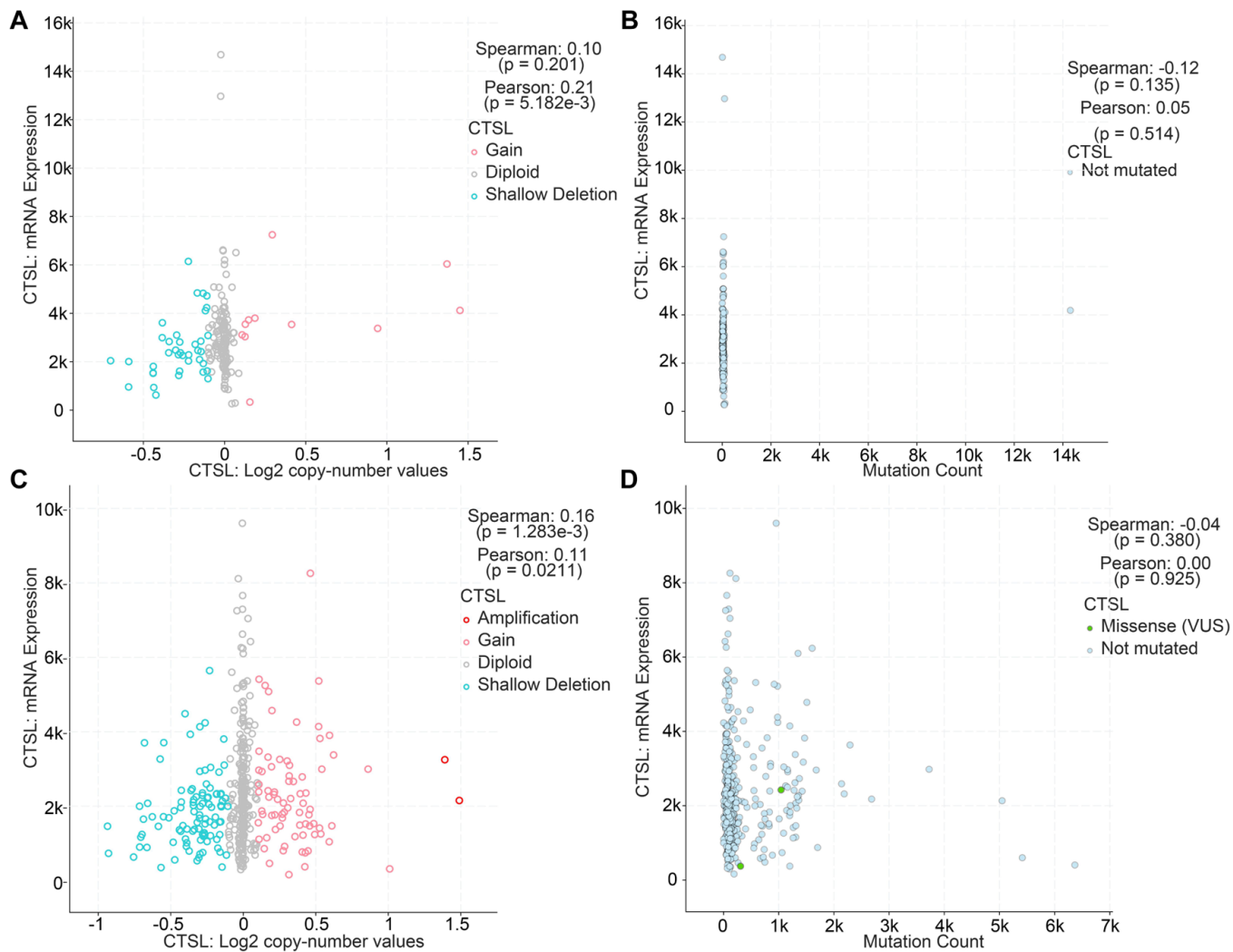




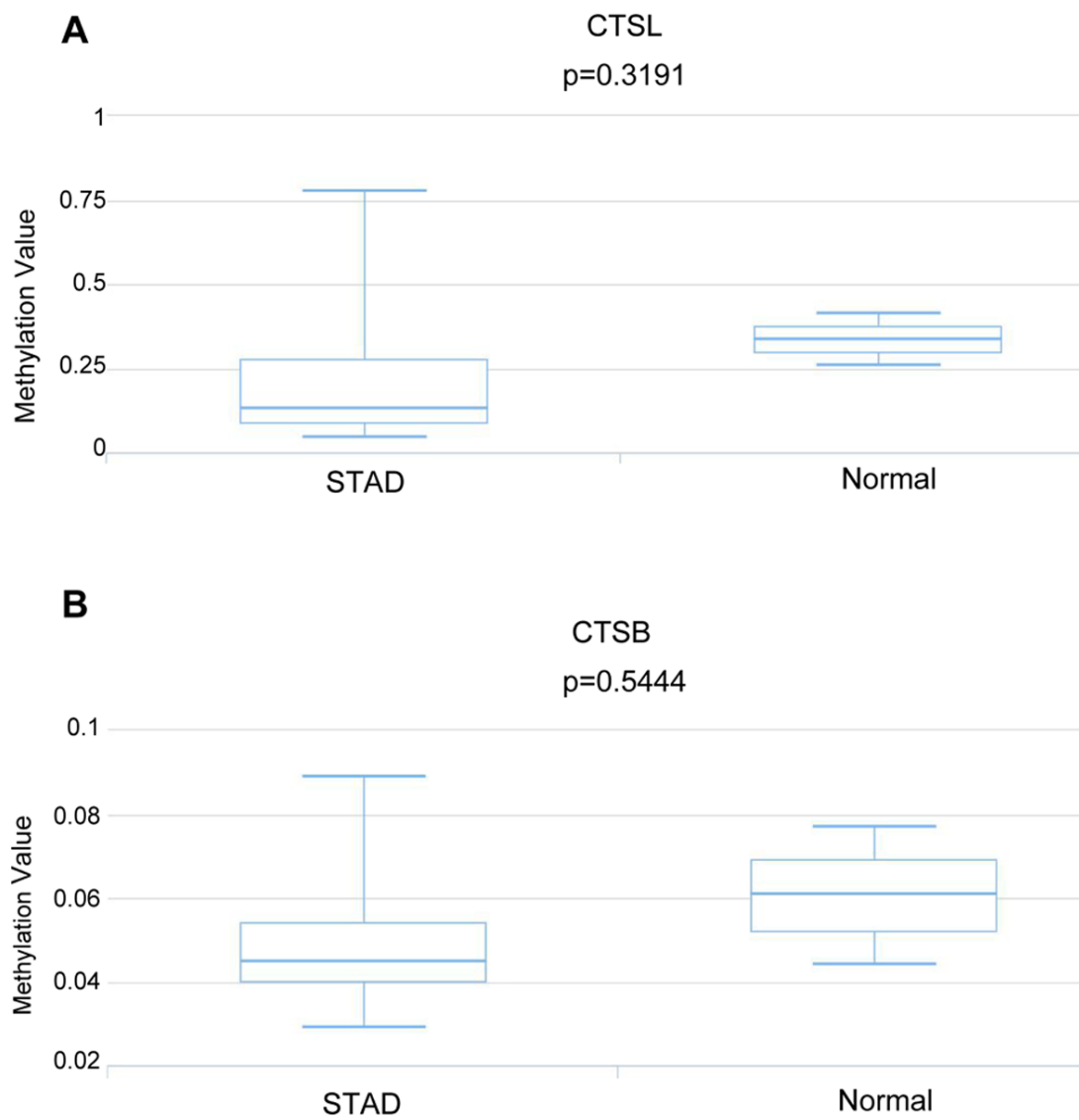
**Supplementary Figure 2.** The correlation analyses of Copy number values or Mutant count with mRNA expression for CTSB in PAAD (A, B) and STAD (C, D) (cBioportal).  $|R| > 0.1$ ,  $P < 0.05$  was regarded as having a correlation.



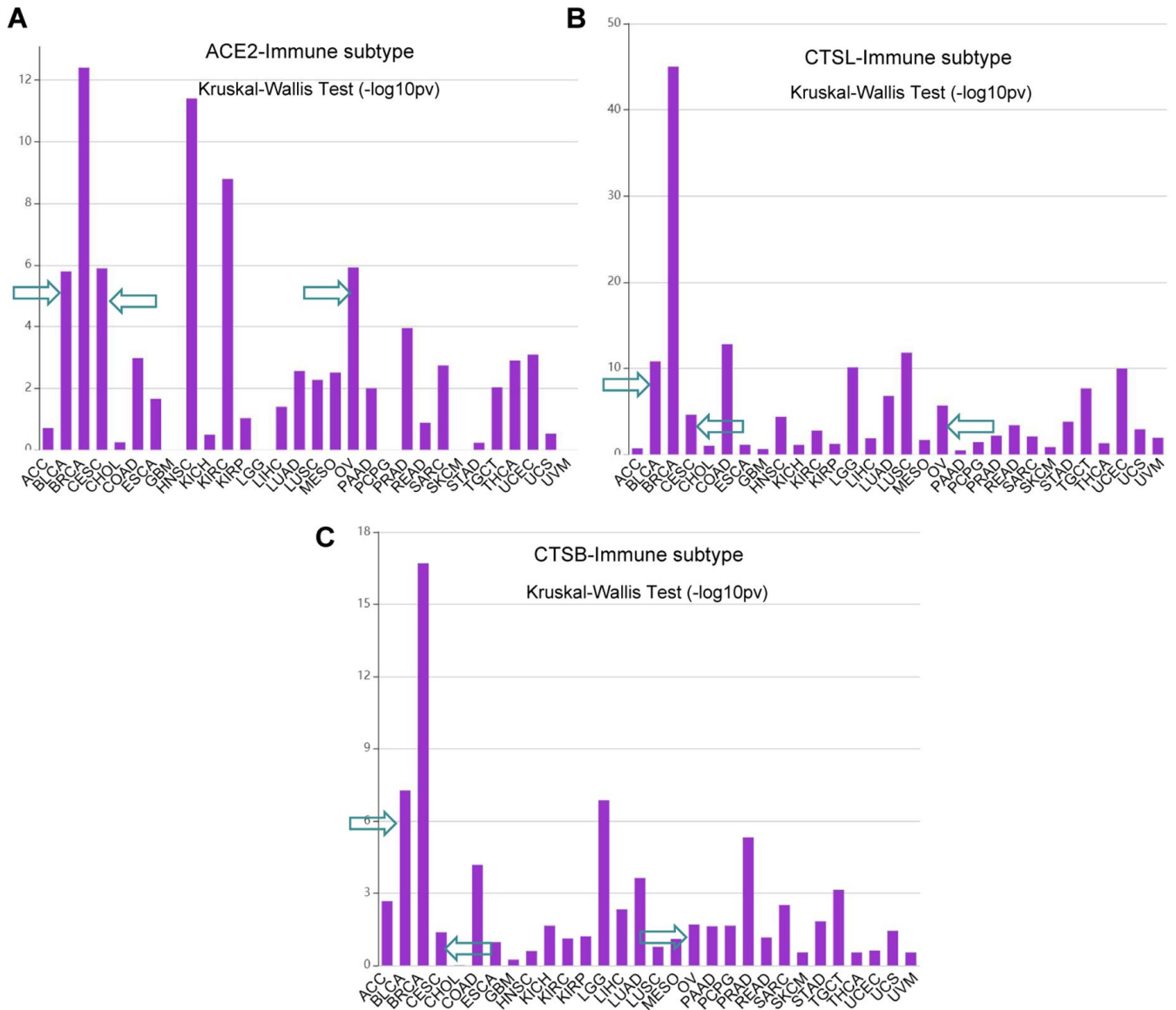
**Supplementary Figure 3.** The correlation analyses of Copy number values or Mutant count with mRNA expression for ACE2 in PAAD (A, B) and STAD (C, D) (cBioportal).  $|R| > 0.1$ ,  $P < 0.05$  was regarded as having a correlation.



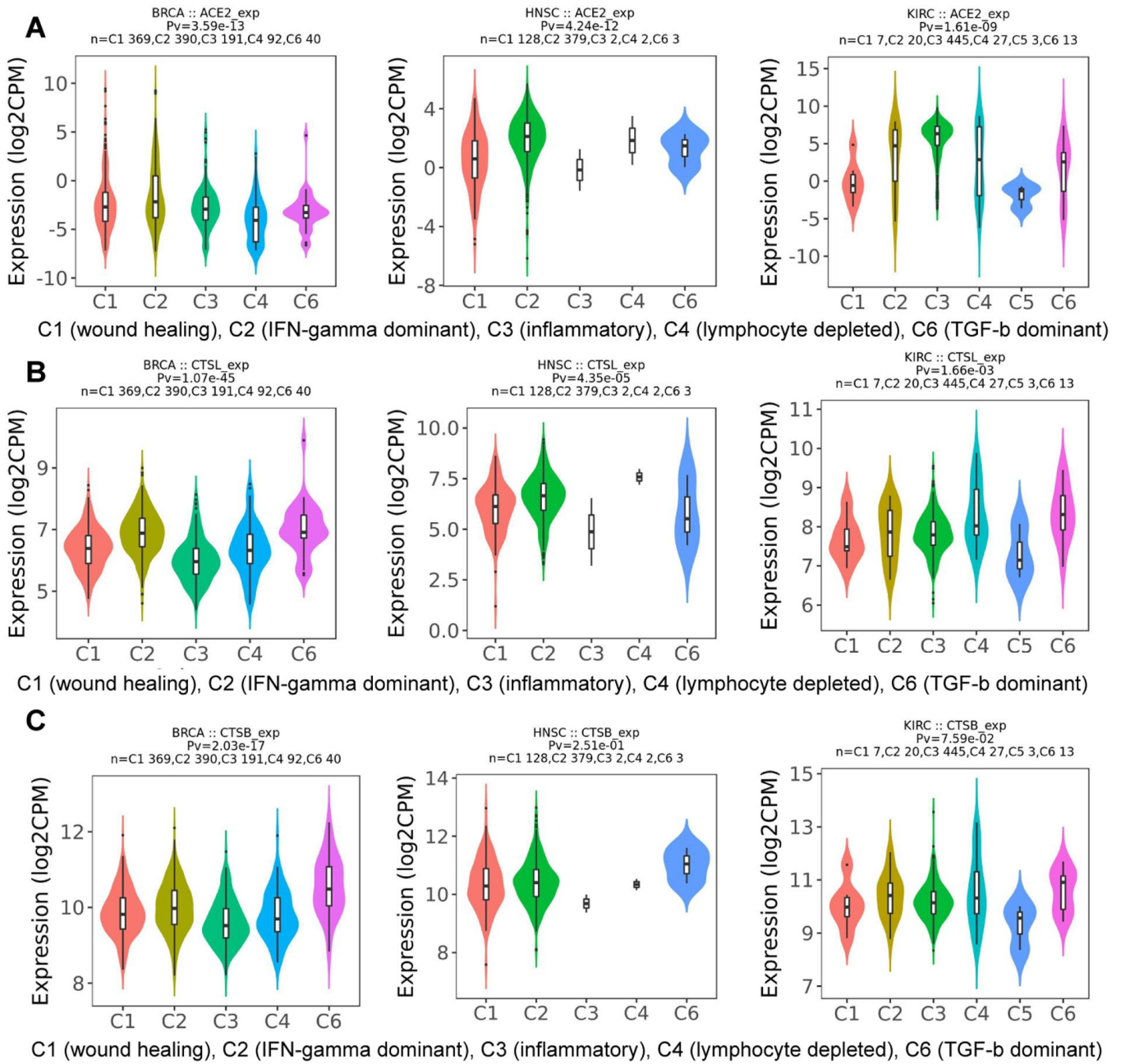
**Supplementary Figure 4.** The correlation analyses of Copy number values or Mutant count with mRNA expression for CTSL in PAAD (A, B) and STAD (C, D) (cBioportal).  $|R| > 0.1$ ,  $P < 0.05$  was regarded as having a correlation.



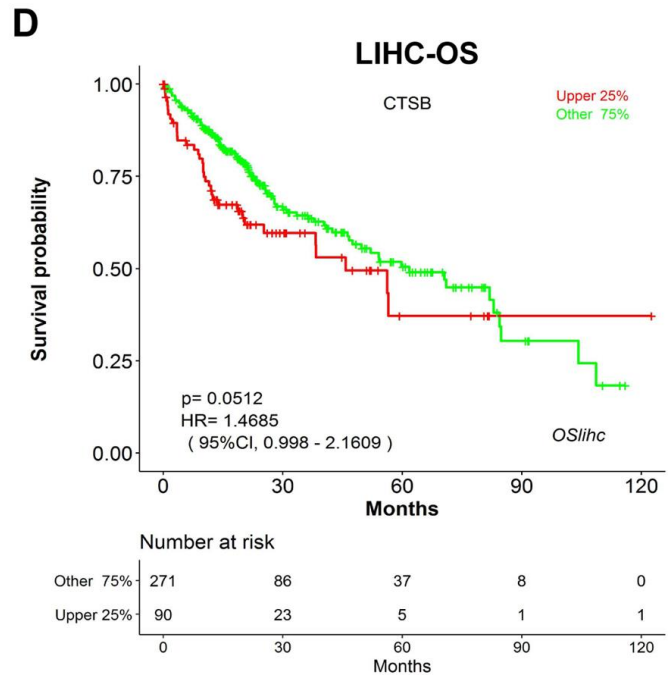
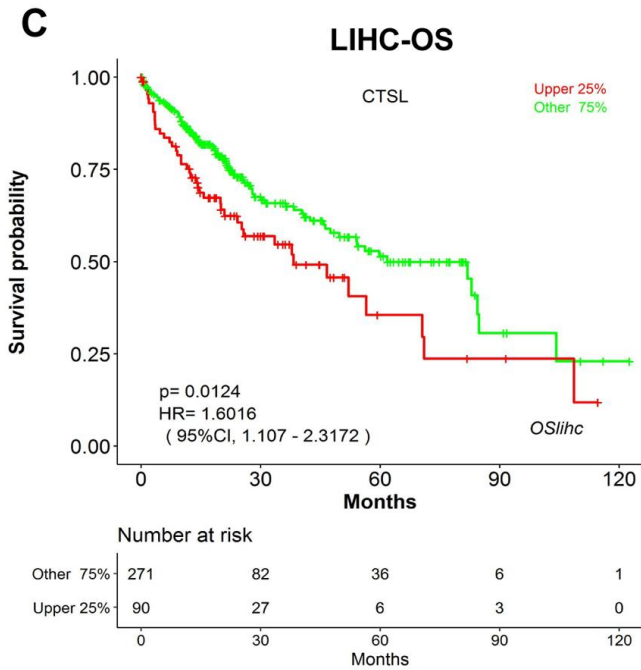
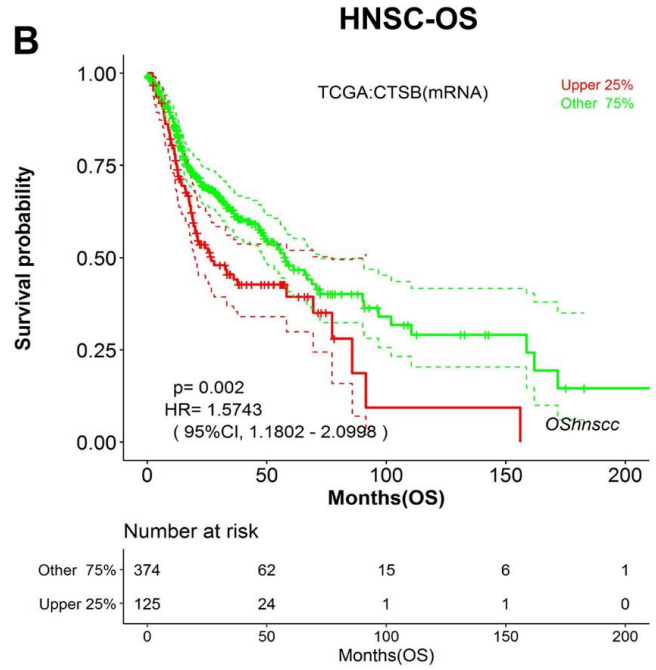
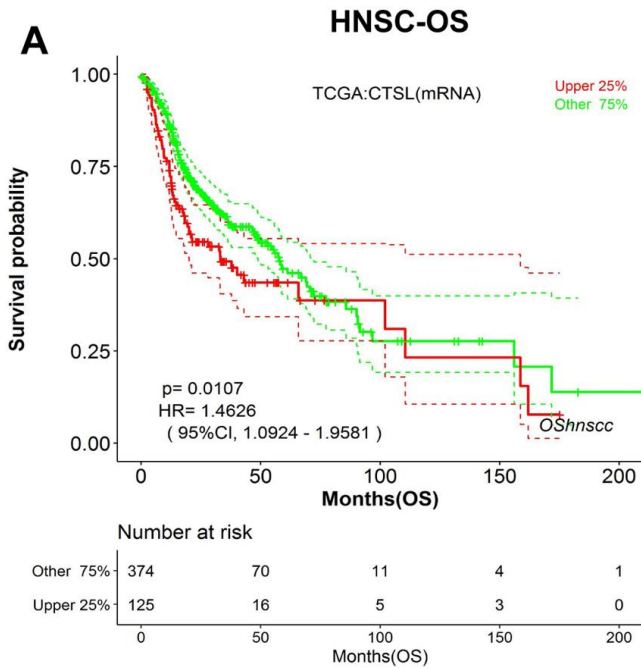
**Supplementary Figure 5.** The identification of DNA methylation for CTSL (A) and CTSB (B) in STAD (DiseaseMeth version 2.0).  $P < 0.05$  was regarded as significance.



**Supplementary Figure 6.** The distribution analyses of immune subtypes including C1 (wound healing), C2 (IFN-gamma dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet) and C6 (TGF-b dominant) for ACE2 (A), CTSL (B) and CTSB (C) in 30 types cancers via TISIDB web tool. The green arrows represent the difference analyses of ACE2, CTSL, and CTSB among the immune subtypes of OV, CESC, and BLCA.



**Supplementary Figure 7.** The distribution graphs of ACE2 (A) and CTSL (B) and CTSB (C) in different immune subtypes in BRCA, HNSC, and KIRC (TISIDB). C1: wound healing; C2: IFN-gamma dominant; C3: inflammatory; C4: lymphocyte depleted; C5: immunologically quiet; C6: TGF-b dominant.



**Supplementary Figure 8.** The OS (Overall survival) analyses of CTSL (A) and CTSB (B) in HNSC utilizing the LOGpc web tool based on TCGA and GEO data. The OS analyses of CTSL(C) and CTSB (D) in LIHC utilizing the LOGpc web tool based on TCGA and GEO data.

## Supplementary Tables

**Supplementary Table 1. The distribution of CTSL in PAAD, STAD and their normal tissues.**

Cancer types	Dataset	Tumor (Cases)	Normal (Cases)	Fold change	t-Test	P-value
PAAD	Badea Pancreas	Pancreatic Ductal Adenocarcinoma(39)	Pancreas(39)	3.124	8.078	1.40E-11
	Segara Pancreas	Pancreatic Carcinoma(11)	Pancreas(6)	3.233	3.966	7.32E-04
	Iacobuzio-Donahue Pancreas	Pancreatic Adenocarcinoma(12)	Pancreas(5)	2.811	4.173	0.004
	Grutzmann Pancreas	Pancreatic Ductal Adenocarcinoma(11)	Pancreas Duct(11)	1.791	1.667	0.056
	Pei Pancreas	Pancreatic Carcinoma(36)	Pancreas(16)	2.443	3.482	0.001
STAD	Chen Gastric	Gastric Intestinal Type Adenocarcinoma(64)	Gastric Mucosa(29)	2.424	13.942	9.87E-22
	Chen Gastric	Diffuse Gastric Adenocarcinoma(13)	Gastric Mucosa(29)	1.834	6.328	2.24E-06
	Chen Gastric	Gastric Mixed Adenocarcinoma(8)	Gastric Mucosa(29)	3.357	7.385	3.01E-05
	DErrico Gastric	Gastric Mixed Adenocarcinoma(4)	Gastric Mucosa(31)	2.847	16.872	3.40E-10
	DErrico Gastric	Gastric Intestinal Type Adenocarcinoma(26)	Gastric Mucosa(31)	3.837	10.142	2.67E-14
	DErrico Gastric	Diffuse Gastric Adenocarcinoma(6)	Gastric Mucosa(31)	2.378	4.596	1.00E-03
	Gui Gastric	Gastric Cancer(80)	Gastric Tissue(80)	1.594	6.047	5.15E-09
	Cho Gastric	Gastric Intestinal Type Adenocarcinoma(20)	Gastric Tissue(19)	1.797	5.201	5.44E-06
	Cho Gastric	Diffuse Gastric Adenocarcinoma(31)	Gastric Tissue(19)	1.776	6.056	1.05E-07
	Cho Gastric	Gastric Adenocarcinoma(4)	Gastric Tissue(19)	2.073	2.696	3.30E-02
	Wang Gastric	Gastric Cancer(12)	Gastric Tissue(12) Gastric Mucosa(3)	1.948	3.419	1.00E-03

PAAD: Pancreatic adenocarcinoma; STAD: Stomach Adenocarcinoma.



**Supplementary Table 2. The distribution of CTSB in PAAD, STAD and their normal tissues.**

Cancer types	Dataset	Tumor(Cases)	Normal(Cases)	Fold change	t-Test	P-value
PAAD	Logsdon Pancreas	Pancreatic Adenocarcinoma(10)	Pancreas(5)	1.614	2.225	0.022
	Badea Pancreas	Pancreatic Ductal Adenocarcinoma(39)	Pancreas(39)	1.975	6.077	3.70E-08
	Iacobuzio-Donahue Pancreas 2	Pancreatic Adenocarcinoma(12)	Pancreas(5)	1.797	1.727	0.069
	Pei Pancreas	Pancreatic Carcinoma(36)	Pancreas(16)	1.514	2.525	0.009
STAD	Chen Gastric	Gastric Mixed Adenocarcinoma(8)	Gastric Mucosa(29)	2.844	11.259	4.29E-09
	Chen Gastric	Gastric Intestinal Type Adenocarcinoma(67)	Gastric Mucosa(29)	2.717	14.424	6.38E-22
	Chen Gastric	Diffuse Gastric Adenocarcinoma(13)	Gastric Mucosa(29)	2.538	5.275	5.12E-05
	DErrico Gastric	Gastric Intestinal Type Adenocarcinoma(26)	Gastric Mucosa(31)	2.357	8.821	7.30E-11
	DErrico Gastric	Diffuse Gastric Adenocarcinoma(6)	Gastric Mucosa(31)	1.887	4.323	3.00E-03
	DErrico Gastric	Gastric Mixed Adenocarcinoma(4)	Gastric Mucosa(31)	2.772	3.866	1.40E-02
	Wang Gastric	Gastric Cancer(12)	Gastric Mucosa(12) Gastric Tissue(3)	2.273	3.988	5.18E-04
	Cho Gastric	Diffuse Gastric Adenocarcinoma(31)	Gastric Tissue(19)	1.793	3.673	3.07E-04

PAAD: Pancreatic adenocarcinoma; STAD: Stomach Adenocarcinoma.

**Supplementary Table 3. The CpG methylation probes located in promoters of CTSL/B and ACE2.**

Gene Symbol	CpG	Group	Relation To Island
CTSL	cg11154542	TSS1500	N_Shore
CTSL	cg11306701	5'UTR	Island
CTSL	cg13985445	TSS1500	N_Shore
CTSL	cg14236855	5'UTR;1stExon	Island
CTSL	cg14243623	5'UTR;1stExon	Island
CTSL	cg14557714	5'UTR	Island
CTSL	cg15242570	5'UTR;1stExon	Island
CTSB	cg00057317	5'UTR	Island
CTSB	cg00276098	TSS1500	S_Shore
CTSB	cg02637136	5'UTR;1stExon	Island
CTSB	cg04315434	TSS1500	Island
CTSB	cg08012294	5'UTR	N_Shore
CTSB	cg16624891	5'UTR	OpenSea
CTSB	cg17019285	5'UTR	Island
CTSB	cg18787975	TSS1500	S_Shore
CTSB	cg20539307	5'UTR	OpenSea
CTSB	cg21281001	5'UTR	Island
CTSB	cg21756720	5'UTR	N_Shore
CTSB	cg21919729	5'UTR	OpenSea
CTSB	cg22139271	5'UTR;1stExon	Island
CTSB	cg26407100	TSS1500	Island
ACE2	cg05748796	5'UTR	OpenSea
ACE2	cg08559914	TSS200	OpenSea
ACE2	cg16734967	5'UTR;1stExon	OpenSea
ACE2	cg18458833	TSS1500	OpenSea
ACE2	cg18877734	TSS1500	OpenSea
ACE2	cg21598868	TSS1500	OpenSea

**Supplementary Table 4. The correlation analysis of DNA methylation level of CTSL to its gene expression in pan-cancer.**

<b>Disease</b>	<b>GeneSymbol</b>	<b>Pearson R</b>	<b>Pearson p-value</b>
BLCA	CTSL	-0.00312611	0.948826
BRCA	CTSL	-0.027521	0.418322
CESC	CTSL	-0.366557	0
CHOL	CTSL	-0.239902	0.112444
COAD	CTSL	-0.0123726	0.82659
ESCA	CTSL	-0.252473	0.000863984
GBM	CTSL	-0.255165	0.04186
HNSC	CTSL	-0.301539	0
KIRC	CTSL	-0.0301101	0.578949
KIRP	CTSL	-0.0844174	0.146702
LIHC	CTSL	-0.0159312	0.74655
LUAD	CTSL	-0.0655713	0.153184
LUSC	CTSL	-0.148525	0.00384861
PAAD	CTSL	-0.0665734	0.371896
PCPG	CTSL	-0.0782841	0.289514
PRAD	CTSL	0.0319835	0.462059
READ	CTSL	0.0645515	0.521309
SARC	CTSL	0.0854136	0.167247
SKCM	CTSL	-0.340934	0
STAD	CTSL	-0.0153432	0.778991
THCA	CTSL	0.11606	0.00601102
THYM	CTSL	-0.142602	0.118683
UCEC	CTSL	0.0546759	0.240321

BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: Cholangio carcinoma; COAD: Colon adenocarcinoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LIHC: Liver hepatocellular carcinoma; LMS: Leiomyosarcoma; LUAD: Lung: adenocarcinoma; LUSC: Lung squamous cell carcinoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SARC: Sarcoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach Adenocarcinoma; THCA: Thyroid carcinoma; THYM: Thymoma; UCEC: Uterine Corpus Endometrial Carcinoma;  $0.1 < |R| < 0.3$  and  $P < 0.05$  was regarded as weak correlation;  $0.3 < |R| < 0.5$  and  $P < 0.05$  was regarded as moderate correlation;  $|R| > 0.5$  and  $P < 0.05$  was regarded as strong correlation.

**Supplementary Table 5. The correlation analysis of DNA methylation level of ACE2 to its gene expression in pan-cancer.**

<b>Disease</b>	<b>GeneSymbol</b>	<b>Pearson R</b>	<b>Pearson p value</b>
BLCA	ACE2	-0.22944	0.000001803
BRCA	ACE2	-0.150097	9.0638E-06
CESC	ACE2	0.0420417	0.461514
CHOL	ACE2	0.000887264	0.995385
COAD	ACE2	-0.594137	1.5447e-31
ESCA	ACE2	-0.123937	0.106303
GBM	ACE2	-0.00669388	0.958133
HNSC	ACE2	-0.114473	0.00885114
KIRC	ACE2	-0.0672159	0.215015
KIRP	ACE2	0.104754	0.0714444
LIHC	ACE2	-0.365046	0
LUAD	ACE2	-0.208215	4.6257E-06
LUSC	ACE2	-0.0820471	0.111732
PAAD	ACE2	0.214959	0.00356698
PCPG	ACE2	0.0938661	0.203775
PRAD	ACE2	-0.00895685	0.836858
READ	ACE2	-0.51804	2.89e-8
SARC	ACE2	0.100805	0.102859
SKCM	ACE2	0.0787119	0.0893085
STAD	ACE2	-0.0777361	0.154476
THCA	ACE2	0.156607	0.000201307
THYM	ACE2	0.0975874	0.286938

BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: Cholangio carcinoma; COAD: Colon adenocarcinoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LIHC: Liver hepatocellular carcinoma; LMS: Leiomyosarcoma; LUAD: Lung: adenocarcinoma; LUSC: Lung squamous cell carcinoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SARC: Sarcoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach Adenocarcinoma; THCA: Thyroid carcinoma; THYM: Thymoma; UCEC: Uterine Corpus Endometrial Carcinoma;  $0.1 < |R| < 0.3$  and  $P < 0.05$  was regarded as weak correlation;  $0.3 < |R| < 0.5$  and  $P < 0.05$  was regarded as moderate correlation;  $|R| > 0.5$  and  $P < 0.05$  was regarded as strong correlation.

**Supplementary Table 6. The correlation analysis of DNA methylation level of CTSB to its gene expression in pan-cancer.**

<b>Disease</b>	<b>GeneSymbol</b>	<b>Pearson R</b>	<b>Pearson p value</b>
BLCA	CTSB	-0.0862408	0.0760846
BRCA	CTSB	-0.101124	0.00287371
CESC	CTSB	0.0331376	0.561709
CHOL	CTSB	-0.334951	0.0245063
COAD	CTSB	-0.0965971	0.0864639
ESCA	CTSB	-0.283268	0.000173876
GBM	CTSB	0.00697352	0.956386
HNSC	CTSB	-0.0901373	0.0395278
KIRC	CTSB	0.114073	0.0349671
KIRP	CTSB	-0.125537	0.0305488
LIHC	CTSB	-0.183504	0.000173767
LUAD	CTSB	-0.0666532	0.146504
LUSC	CTSB	-0.161446	0.00166104
PAAD	CTSB	-0.187391	0.0113059
PCPG	CTSB	-0.114122	0.121919
PRAD	CTSB	-0.284738	0
READ	CTSB	-0.239448	0.0158798
SARC	CTSB	-0.0538639	0.384302
SKCM	CTSB	-0.194021	2.42374E-05
STAD	CTSB	-0.118964	0.028996
THCA	CTSB	-0.0641417	0.129853
THYM	CTSB	0.00115651	0.989955
UCEC	CTSB	0.11232	0.0156076

BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: Cholangio carcinoma; COAD: Colon adenocarcinoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LIHC: Liver hepatocellular carcinoma; LMS: Leiomyosarcoma; LUAD: Lung: adenocarcinoma; LUSC: Lung squamous cell carcinoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SARC: Sarcoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach Adenocarcinoma; THCA: Thyroid carcinoma; THYM: Thymoma; UCEC: Uterine Corpus Endometrial Carcinoma;  $0.1 < |R| < 0.3$  and  $P < 0.05$  was regarded as weak correlation;  $0.3 < |R| < 0.5$  and  $P < 0.05$  was regarded as moderate correlation;  $|R| > 0.5$  and  $P < 0.05$  was regarded as strong correlation.