



Cathepsin L in Lung Adenocarcinoma: Prognostic Significance and Immunotherapy Response Through a Multi Omics Perspective

Cancer Informatics
Volume 23: 1–14
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351241307492



Jianming Lu^{1*}, Jiaqi Liang^{2*}, Gang Xiao^{1,3*}, Zitao He¹, Guifang Yu⁴, Le Zhang⁵, Chao Cai⁶, Gao Yi² and Jianjiang Xie^{1,3}

¹Center for Medical Research on Innovation and Translation, Institute of Clinical Medicine, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China. ²Pneumology Department, Key Laboratory of Biological Targeting Diagnosis, Therapy and Rehabilitation of Guangdong Higher Education Institutes, The Fifth Affiliated Hospital, Guangzhou Medical University, Guangzhou, Guangdong, China. ³Department of Thoracic Surgery, Guangzhou First People's Hospital, The Second Affiliated Hospital of South China University of Technology, Guangzhou, Guangdong, China. ⁴Oncology Department, Key Laboratory of Biological Targeting Diagnosis, Therapy and Rehabilitation of Guangdong Higher Education Institutes, The Fifth Affiliated Hospital, Guangzhou Medical University, Guangzhou, Guangdong, China. ⁵Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, USA. ⁶Department of Urology, Guangdong Key Laboratory of Urology, Guangzhou Institute of Urology, Minimally Invasive Surgery Center, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China.

ABSTRACT

OBJECTIVES: Lung adenocarcinoma (LUAD), a predominant form of lung cancer, is characterized by a high rate of metastasis and recurrence, leading to a poor prognosis for LUAD patients. This study aimed to identify and rigorously validate a highly precise biomarker, Cathepsin L (CTSL), for the prognostic prediction of lung adenocarcinoma.

METHODS: We employed a multicenter and omics-based approach, analyzing RNA sequencing data and mutation information from public databases such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). The DepMap portal with Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) technology was used to assess the functional impact of CTSL. Immunohistochemistry (IHC) was conducted on a local cohort to validate the prognostic significance of CTSL at the protein expression level.

RESULTS: Our findings revealed a significant correlation between elevated CTSL expression and advanced disease stage in LUAD patients. Kaplan–Meier survival analysis and Cox regression modeling revealed that high CTSL expression is associated with poor overall survival. The *in vitro* studies corroborated these findings, revealing notable suppression of tumor proliferation following CTSL knockout in cell lines, particularly in LUAD. Functional enrichment revealed that CTSL activated pathways associated with tumor progression, such as angiogenesis and Transforming growth factor beta (TGF- β) signaling, and inhibited pathways such as apoptosis and DNA repair. Mutation analysis revealed distinct variations in the CTSL expression groups.

CONCLUSION: This study highlights the crucial role of CTSL as a prognostic biomarker in LUAD. This combined multicenter and omics-based analysis provides comprehensive insights into the biological role of CTSL, supporting its potential as a target for therapeutic intervention and a marker for prognosis in patients with LUAD.

KEYWORDS: Lung adenocarcinoma, Cathepsin L, prognostic biomarker, multi-center study, omics analysis, immunotherapy

RECEIVED: August 23, 2024. **ACCEPTED:** November 28, 2024.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The article is supported by Guangdong Basic and Applied Basic Research Foundation (2020A1515011290) and the Guangzhou Municipal Science and Technology Project (202201010053, 202201010787).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Jianjiang Xie, Department of Thoracic Surgery, Guangzhou First People's Hospital, The Second Affiliated Hospital of South China University of Technology, Panfu Road, Yuexiu District, Guangzhou 510180, China. Email: eyxiejianjiang@scut.edu.cn

Gao Yi, Pneumology Department, Key Laboratory of Biological Targeting Diagnosis, Therapy and Rehabilitation of Guangdong Higher Education Institutes, The Fifth Affiliated Hospital, Guangzhou Medical University, Gangwan Road, Huangpu District, Guangzhou 510700, China. Email: 15360802191@163.com.

Introduction

Consistently occupying a leading position in cancer statistics, lung cancer poses an ongoing and severe public health challenge.¹ Within this context, lung adenocarcinoma (LUAD) has emerged as a significant subtype, representing a substantial proportion of

lung cancer cases. Characterized by a high likelihood of distant metastasis and a mean overall survival (OS) of under 5 years, LUAD presents a formidable threat.² The limitations and side effects associated with traditional treatments such as surgery, radiotherapy, and chemotherapy further complicate the clinical management of this condition.³ Therefore, there is an urgent need for the development and implementation of reliable

*These authors have contributed equally to this work.



biomarkers to guide specific treatment plans for LUAD, with the aim of improving patient outcomes.

A defining feature of contemporary medicine in this century is the transformative understanding of the biological landscape of cancer.⁴ The advent of cancer biomarkers has revolutionized patient care, offering a tailored approach to treatment.⁵ This evolution has rendered the process of cancer malignant transformation more discernible, leading to the development of innovative methods for combating the disease.⁶ Advancements in high-throughput sequencing technologies have been instrumental in the identification of pivotal genes associated with LUAD.⁷ This technological progress opens new avenues for the discovery of groundbreaking biomarkers and therapeutic targets.

Cathepsin L (CTSL), a cysteine protease family member, is pivotal for lysosomal degradation of intracellular and endocytosed proteins, intracellular protein breakdown, and SARS-CoV-2 infection processes.^{8,9} CTSL upregulation is frequently observed in a variety of human cancers and contributes to oncogenesis, tumor progression, invasion, and poor patient prognosis.¹⁰⁻¹² Moreover, the significant association of CTSL with cancer-related osteolysis detrimentally impacts both the quality and length of life of cancer patients.¹³

Reports on CTSL in LUAD have been sporadic. Han and colleagues, in a preclinical study, demonstrated that EMT induced by CTSL upregulation was linked to chemotherapy resistance in A549 cells.¹⁴ Additionally, higher expression levels of CTSL are correlated with shorter OS in patients with lung cancer and may indicate a poor response to immunotherapy.¹⁵ To further our understanding of the role of CTSL in LUAD, our study is the first to utilize a comprehensive approach combining multicenter, multiomics data with our local cohort. This systematic and comprehensive methodology aims to explore the potential of CTSL as a biomarker in LUAD. Previous studies have been limited to single-omic analyses and lack validation with clinical samples from their respective centers. Our integrated approach provides a more holistic view of the implications of CTSL in LUAD.

Methods

Data processing

This research began with acquiring RNA sequencing (RNA-seq) data, mutation details, and clinical information for Lung Adenocarcinoma (TCGA-LUAD) and pan-cancer cohorts, sourced from the UCSC Xena database (<https://xenabrowser.net/datapages/>). To further substantiate our findings, LUAD datasets from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), namely GSE50081, GSE31210, and GSE68465, were collected. These datasets are listed in Supplemental Table S1. To convert Ensembl IDs to SYMBOL IDs within the RNA-seq datasets, we employed the R packages clusterProfiler and org.Hs.eg.db (version 3.17.0).¹⁶ For the microarray data, batch effects were adjusted through the ComBat function

within the sva R package.¹⁷ After obtaining these datasets, we retained the LUAD samples only. Additionally, only samples with both gene expression and follow-up information were included in subsequent analyses.

Statistical analysis

Prognostic analysis in this study involved calculating Hazard Ratios (HR) using Kaplan–Meier (KM) analysis, coupled with time-dependent Receiver Operating Characteristic (ROC) curve assessments. The “survival” R package was employed for KM analysis and Cox regression modeling. The “survivalROC” R package was used to determine the Area Under the ROC Curves (AUC).¹⁸ Specifically, Cox regression models were built using the coxph function from this R package to investigate the association between gene expression and clinical outcomes. The significance of the Kaplan–Meier survival curves was determined through the log-rank test.

In assessing the relationships between variables, Spearman correlation analysis was employed to calculate correlation coefficients. When comparing two continuous variables, we utilized either the Wilcoxon rank-sum test based on the distribution characteristics of the data.

Immunohistochemistry (IHC)

In this investigation, LUAD tissue microarray (TMA) were obtained from Shanghai Zhuoli Biotech Company. The TMA, identified as ZL-Luc601, contained 31 validated LUAD tissue samples, each accompanied by relevant clinical data. The acquisition of these samples was approved by the Ethics Committee of Shanghai Zhuoli Biotech Company (approval number SHLLS-BA-22101102), with written informed consent acquired from each participant for the use of their samples in this study.

For the immunohistochemistry protocol, detailed methodologies are outlined in our previously published paper,¹⁹ with a summary provided herein. Antigen retrieval was performed on tissue slices or TMAs using a citrate antigen retrieval solution. This was followed by blocking the endogenous peroxidase activity with a 1% H₂O₂ solution and applying nonimmune goat serum for blocking. Tissue slices were incubated overnight at 4°C with the primary antibody, followed by a 30-minute room temperature incubation with a biotinylated secondary antibody. This was succeeded by a 15-minute room temperature incubation with streptavidin-conjugated horseradish peroxidase (HRP). The HRP activity was visualized using diaminobenzidine tetrahydrochloride (DAB), and nuclear staining was conducted using haematoxylin. The scoring of immunoreactive cells was categorized as 0 (0%), 1 (1-10%), 2 (11-50%), and 3 (>50%), while staining intensity was visually scored and graded as 0 (negative), 1 (weak), 2 (moderate), and 3 (strong). The specific antibody employed in this study was the anti-CTSL antibody (Immunoway, YT5124).

Functional enrichment analysis

We explored the relationship between CTSL expression and the expression profiles of all mRNAs, utilizing the Spearman correlation method for analysis. To pinpoint functional enrichment and establish connections with Hallmark pathways, the “clusterProfiler”¹⁶ R package was employed. This method enabled us to detect and highlight biologically significant pathways that may be associated with CTSL expression.

Assessing CTSL dependency

Our investigation into the dependency of various tumor cell lines on CTSL was conducted through the Dependency Map (DepMap) portal, specifically leveraging CRISPR data (<https://depmap.org/portal/>). The analysis encompassed two main modules within the DepMap resource: “Perturbation Effects” and “Predictability.” These modules were critical in providing a comprehensive analysis of gene dependency and co-dependency. Key to this analysis was the utilization of the “Achilles_gene_effect.csv” and “sample_info.csv” files from the DepMap dataset. These files provided the necessary data for a thorough assessment of CTSL’s role across different tumor cell lines.

The landscape of mutation

In our genomic analysis, we quantified all base substitutions, insertions, and deletions within the coding regions of targeted genes. The top 10 genes exhibiting frequent mutations were identified as driver genes using the “maftools” R package.²⁰ Corresponding to the expression levels of CTSL, we delineated the mutation landscapes for these genes across two distinct sample groups, considering a P -value $< .05$ to indicate significant differences. To identify critically amplified and deleted genomic regions, we employed the Genomic Identification of Significant Targets in Cancer (GISTIC) 2.0 pipeline. This enabled us to construct a mutational landscape map for Copy Number Alterations (CNA), highlighting amplifications and deletions in the top 5 genes based on copy number variations. We further examined and compared the expression of each gene across groups categorized by varying CTSL expression levels. For visualization purposes, the “ComplexHeatmap” R package²¹ was utilized, providing a clear and comprehensive representation of these genomic differences.

Immunotherapy response evaluation

We utilized the online tool “BEST” (https://rookieutopia.com/app_direct/BEST/) to forecast patient responses to immune checkpoint blockade (ICB) therapy.²² The transcriptomic expression profiles were compared among patient groups with different levels of CTSL expression and varied immunotherapy outcomes. For the immunotherapy cohort, samples were stratified according to optimal CTSL expression cut-off values for KM survival analysis.

Results

Multicenter data indicate that CTSL is an unfavorable prognostic factor in LUAD

This study, outlined in Figure 1, involved the collection of multi-omic data from the TCGA and three large-sample LUAD datasets from the GEO, specifically from Japan (GSE31210), Canada (GSE50081), and the United States (GSE68465). Our analysis identified CTSL as a potential biomarker for LUAD. Subsequent validation through immunohistochemical analyses in a local cohort, multi-omics evaluations, and pan-cancer assessments confirmed that CTSL was an adverse prognostic factor in LUAD.

Initially, to assess the prognostic significance of CTSL in LUAD, KM survival analysis was conducted across multicenter cohorts. Considering various prognostic endpoints (OS, progression-free interval [PFI], disease-free survival [DFS], and progression-free survival [PFS]), groups with higher CTSL expression had poorer prognoses in LUAD patients than those with lower CTSL expression (Figure 2A–G, $P < .05$). Both univariate and multivariate Cox regression analyses were subsequently performed. The results indicated that the significance of CTSL in multivariate Cox regression was not statistically significant only in the GSE31210-OS (hazard ratio [HR] = 1.438, $P = .22$) and GSE50081-DFS (HR = 1.471, $P = .194$) datasets. In contrast, in the other datasets, higher CTSL expression levels were independently associated with increased risk in patients with LUAD (Figure 2H and I, $P < .05$).

IHC of CTSL in the local cohort

To substantiate the findings obtained from public datasets, we performed an IHC assay on our local cohort. This step was pivotal for confirming the prognostic significance of CTSL in LUAD at the protein expression level using a tissue microarray. Our findings, as depicted in Figure 3A, revealed a predominant cytoplasmic localization of CTSL. Notably, elevated CTSL expression was correlated with advanced disease stage (Figure 3A). KM survival analysis further revealed that high CTSL expression was indicative of an unfavorable prognosis in terms of OS (Figure 3B). Time-dependent ROC analysis, as shown in Figure 3C, revealed that the CTSL had a significant ability to predict third- and fifth-year survival (3-year AUC = 0.85, 5-year AUC = 0.91). This in-depth analysis highlights the crucial role of CTSL as a prognostic marker for survival outcomes in LUAD patients.

CTSL as a prognostic factor in multiple solid tumors

Our research initially highlighted the significant prognostic role of CTSL in LUAD. To explore whether CTSL’s predictive capability extends beyond LUAD, we performed a comprehensive pan-cancer analysis. Using the TCGA database, we

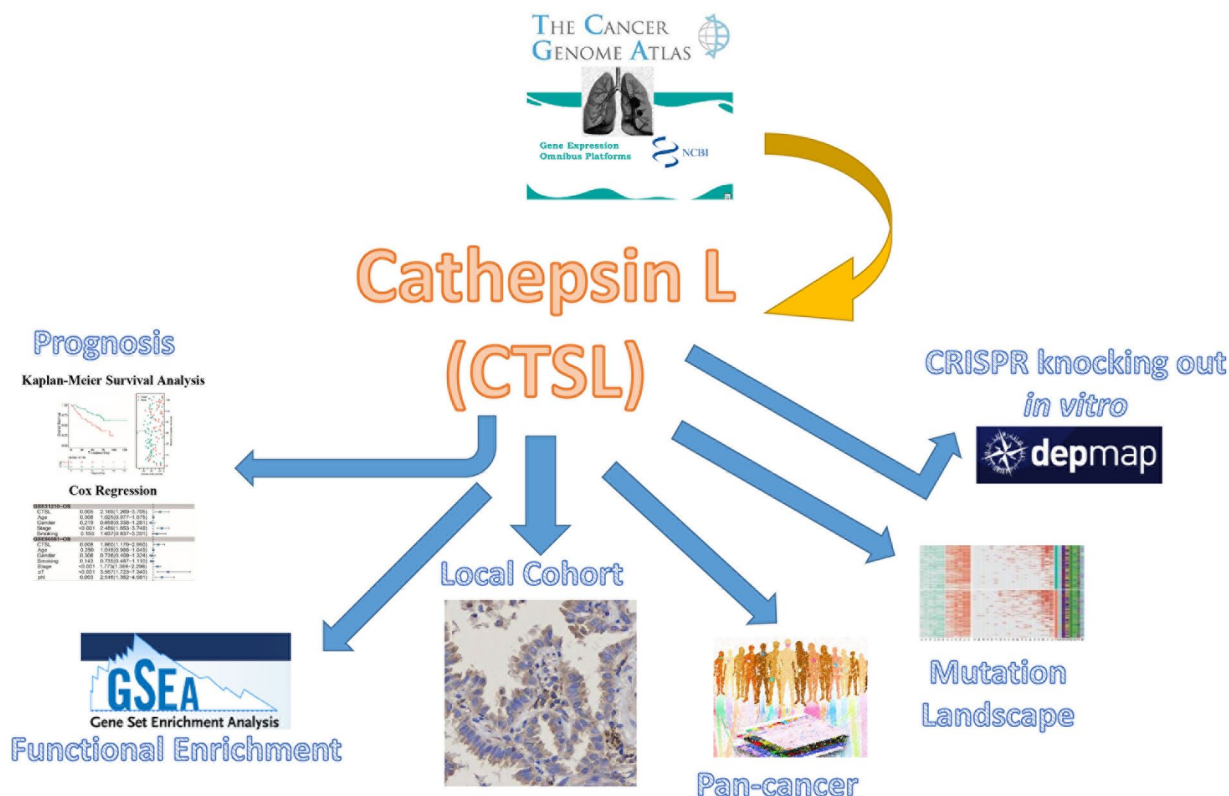


Figure 1. Flowchart.

examined 32 different cancer types. Our findings revealed increased CTSL expression in various cancers, including uveal melanoma (UVM), skin cutaneous melanoma (SKCM), kidney renal papillary cell carcinoma (KIRP), sarcoma (SARC), and diffuse large B-cell lymphoma (DLBC). Conversely, a reduction in CTSL expression was noted in Esophageal Carcinoma (ESCA), Rectum Adenocarcinoma (READ), Colon Adenocarcinoma (COAD), Stomach Adenocarcinoma (STAD), Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC) (Figure 4A). Univariate Cox regression analysis (Figure 4B) identified CTSL as a risk factor for several tumors, which may affect OS, PFI, or both, including glioblastoma multiforme (GBM), ESCA, head and neck squamous cell carcinoma (HNSC), low-grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung squamous cell carcinoma (LUSC), STAD, kidney chromophobe (KICH), and prostate adenocarcinoma (PRAD). Interestingly, CTSL appeared to act as a protective factor in UVM and kidney renal clear cell carcinoma (KIRC). These observations suggest that the role of CTSL is not only related to LUAD prognosis but also extends to other solid tumors, highlighting its broader predictive significance.

Knocking out CTSL inhibits LUAD cell viability

To elucidate the functional impact of CTSL in vitro, we used CRISPR/Cas9 datasets for the targeted knockout of CTSL in various cell lines. Analysis conducted with DepMap revealed

that negative scores on the gene effect scale (dependency scores) are indicative of reduced cell proliferation and survival subsequent to the depletion of a specific gene. As depicted in Figure 5A, the knockout of CTSL manifested differential effects across cell lines, with a majority displaying dependency scores below zero. This trend suggests a predominant inhibitory influence on cellular growth (Supplemental Table S3). In particular, the LUAD cell line NCIH1573 showed the most pronounced dependency (Figure 5B). NCIH1573 is a type of adherent cell that originates from a 35-year-old female who was diagnosed with stage 4 LUAD. It is derived from tumor cells of soft tissue prior to radiation therapy. Furthermore, Figure 5C shows that the top ten lung cancer cell lines exhibited the most significant sensitivity to CTSL depletion, underscoring their vulnerability to its absence.

Functional enrichment analysis of CTSL

Given the prognostic importance of CTSL in cancer, along with its in vitro oncogenic effects, our research aimed to explore its underlying biological functions. We employed GSEA focused on hallmark pathways to characterize the activity profile of CTSL. Figure 6 shows that CTSL predominantly activated several pathways critically associated with tumor progression, such as angiogenesis, hedgehog signaling, TGF-beta signaling, epithelial-mesenchymal transition, and hypoxia. Concurrently, CTSL was found to suppress a number of classical pathways, including apoptosis, mTORC signaling, DNA

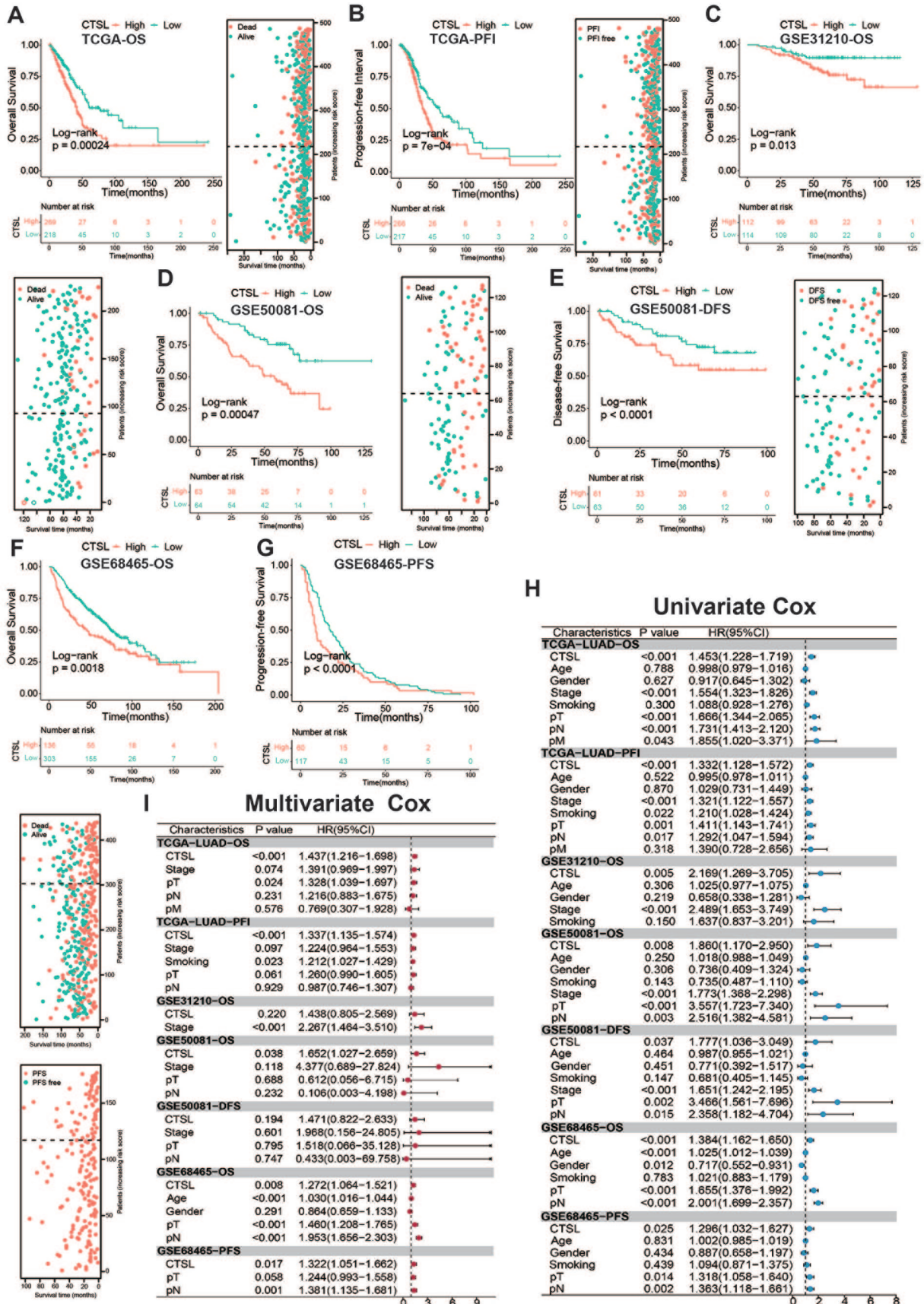


Figure 2. CTSL is an unfavorable biomarker in LUAD: (A–G) Kaplan–Meier survival analyses for CTSL in different LUAD datasets, (H) univariate Cox regression, and (I) multivariate Cox regression assessing CTSL in multi-center datasets.

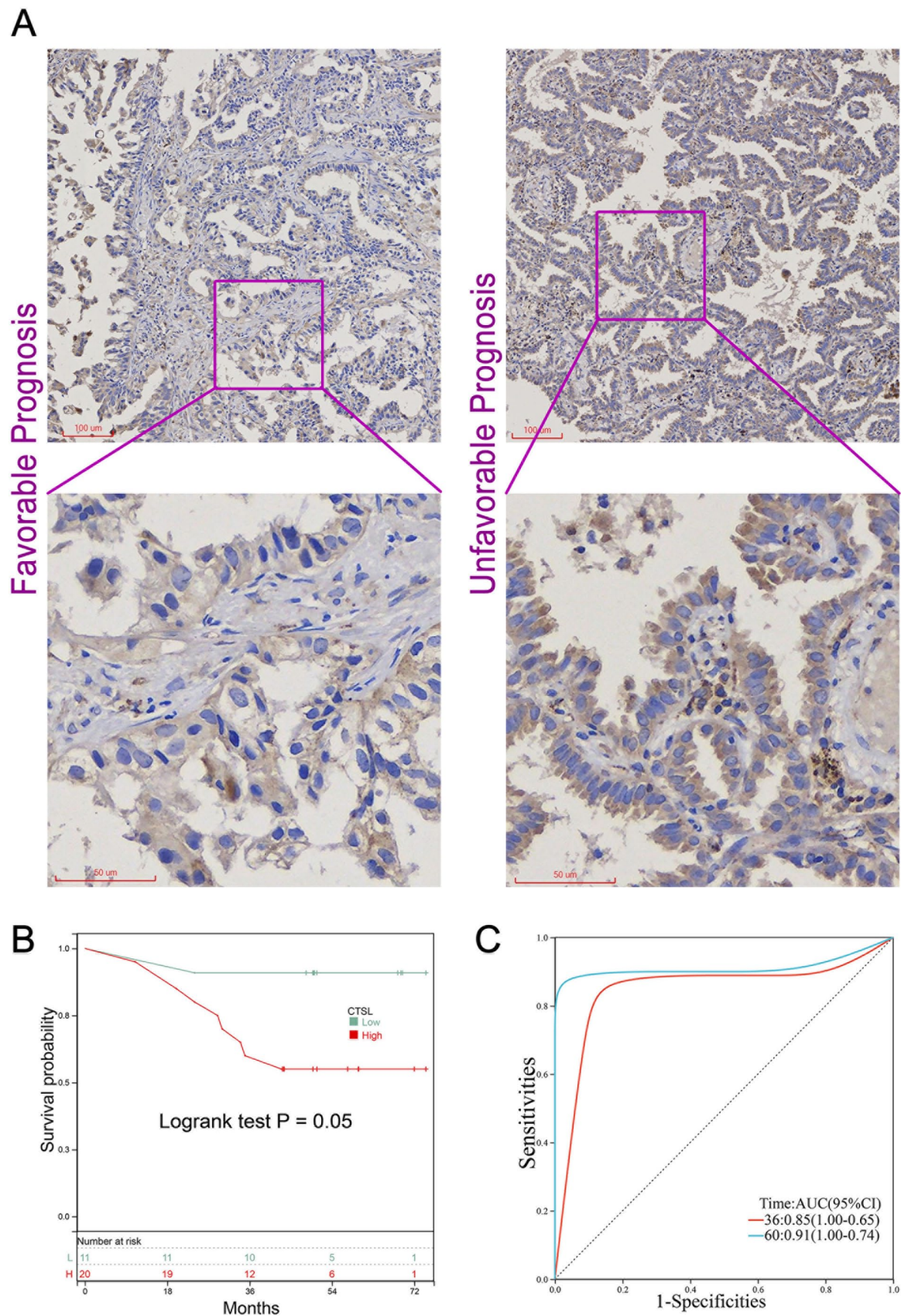
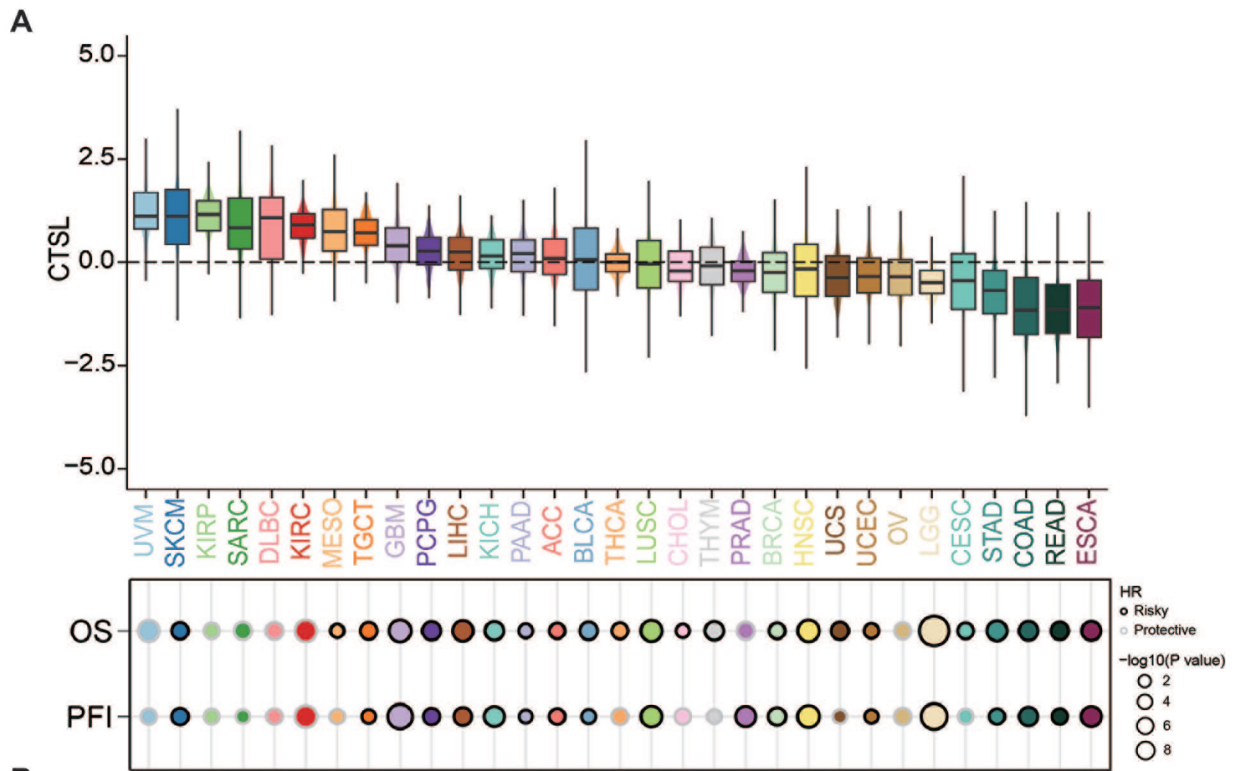


Figure 3. IHC of CTSL in local LUAD cohort: (A) representative staining images of CTSL in patients with varying prognoses within the LUAD cohort, (B) Kaplan–Meier survival curves depicting OS in LUAD patients, categorized into low ($n=11$) and high ($n=20$) CTSL expression groups, and (C) time-dependent receiver ROC analysis evaluating the predictive accuracy of CTSL for 3-year and 5-year OS in LUAD patients.



B Univariate Cox regression of CTSL in Pancancer(OS) Univariate Cox regression of CTSL in Pancancer(PFI)

Cancers	P Value	HR(95%CI)
ACC	0.538	1.155(0.730-1.829)
BLCA	0.183	1.084(0.962-1.222)
BRCA	0.470	1.075(0.884-1.307)
CESC	0.486	1.073(0.880-1.307)
CHOL	0.923	1.036(0.506-2.122)
COAD	0.083	1.217(0.975-1.518)
DLBC	0.240	0.666(0.338-1.312)
ESCA	0.073	1.215(0.982-1.502)
GBM	0.005	1.382(1.102-1.732)
HNSC	0.007	1.191(1.049-1.351)
KICH	0.101	2.221(0.855-5.771)
KIRC	0.004	0.672(0.511-0.884)
KIRP	0.530	0.884(0.602-1.299)
LGG	<0.001	2.347(1.732-3.179)
LIHC	0.021	1.365(1.049-1.776)
LUSC	0.011	1.186(1.040-1.353)
MESO	0.909	1.020(0.720-1.445)
OV	0.371	0.940(0.822-1.076)
PAAD	0.831	1.027(0.807-1.307)
PCPG	0.118	1.843(0.856-3.968)
PRAD	0.207	0.477(0.151-1.508)
READ	0.194	1.381(0.848-2.248)
SARC	0.201	0.882(0.727-1.069)
SKCM	0.353	1.054(0.943-1.178)
STAD	0.033	1.218(1.016-1.461)
TGCT	0.293	3.086(0.377-25.269)
THCA	0.437	1.600(0.489-5.235)
THYM	0.099	2.995(0.812-11.047)
UCEC	0.697	1.086(0.718-1.642)
UCS	0.216	1.302(0.857-1.978)
UVM	0.009	0.460(0.256-0.826)

Cancers	P Value	HR(95%CI)
ACC	0.506	1.124(0.797-1.586)
BLCA	0.823	1.014(0.900-1.142)
BRCA	0.204	1.142(0.930-1.402)
CESC	0.656	0.956(0.786-1.163)
CHOL	0.455	0.769(0.386-1.533)
COAD	0.106	1.179(0.966-1.439)
DLBC	0.501	0.810(0.439-1.495)
ESCA	0.030	1.251(1.022-1.532)
GBM	<0.001	1.496(1.212-1.845)
HNSC	0.005	1.209(1.060-1.379)
KICH	0.032	2.242(1.070-4.696)
KIRC	0.003	0.656(0.495-0.869)
KIRP	0.635	0.923(0.664-1.284)
LGG	<0.001	1.745(1.343-2.266)
LIHC	0.170	1.169(0.936-1.460)
LUSC	0.014	1.222(1.041-1.434)
MESO	0.713	0.933(0.645-1.350)
OV	0.279	0.935(0.829-1.055)
PAAD	0.964	1.005(0.793-1.274)
PCPG	0.365	1.309(0.731-2.345)
PRAD	0.025	1.658(1.065-2.580)
READ	0.536	1.153(0.734-1.811)
SARC	0.858	0.986(0.845-1.151)
SKCM	0.339	1.048(0.952-1.154)
STAD	0.578	1.055(0.874-1.272)
TGCT	0.890	1.036(0.627-1.712)
THCA	0.356	0.718(0.355-1.453)
THYM	0.810	0.923(0.483-1.767)
UCEC	0.911	1.021(0.713-1.461)
UCS	0.857	0.962(0.630-1.470)
UVM	0.533	0.851(0.511-1.415)

Figure 4. Pan-cancer analysis of CTSL: (A) expression levels of CTSL in TCGA pan-cancer and (B) univariate Cox regression assessing CTSL expression in PFI and univariate Cox regression assessing CTSL expression in OS across different tumor types.

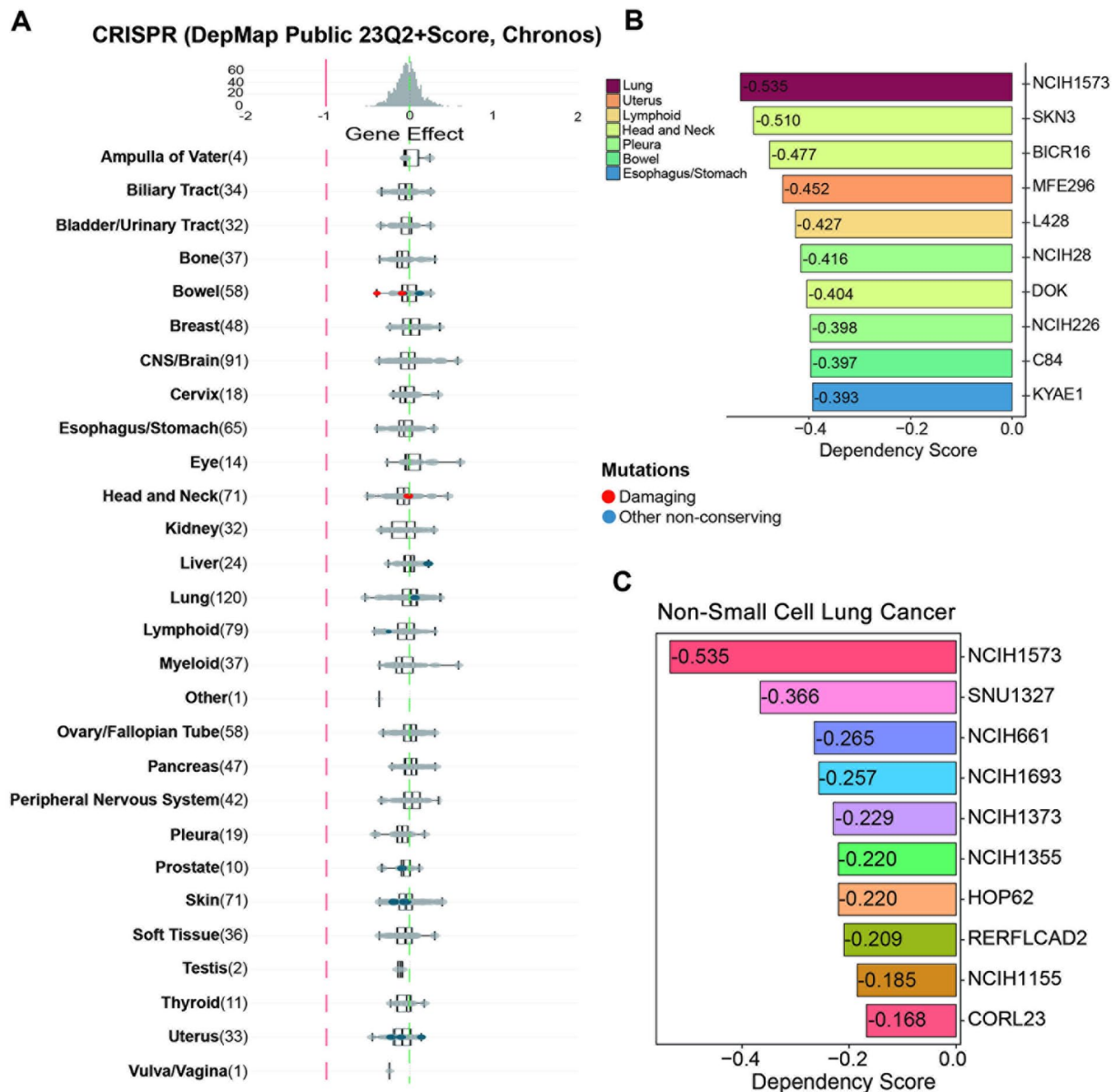


Figure 5. Knocking out CTSL displays negative dependency score: (A) knocking out CTSL by CRISPR/Cas9 in different human cancer cell lines, (B) identification of the top 10 cancer cell lines exhibiting heightened sensitivity to CTSL knockout, and (C) identification of the top 10 lung cancer cell lines exhibiting heightened sensitivity to CTSL knockout.

repair, and those related to the cell cycle. This dual role of CTSL, both in the activation and suppression of key pathways, offers a strategic direction for our subsequent mechanistic investigations into the role of CTSL in cancer biology.

Role of CTSL in the genomic landscape of LUAD mutations

While our preceding analyses focused primarily on the transcriptome, we explored the role of CTSL at the mutation level in LUAD (Figure 7). Within the TCGA-LUAD cohort, the top ten genes exhibiting high mutation rates in tumor samples included TP53, TTN, MUC16, APOB, ASXL3, HMCN1, CNTNAP5, PKHD1, NF1, and SCEP1 (Figure 7A). Notably, TP53, TTN, and MUC16 had

mutation frequencies of 49%, 45%, and 39%, respectively. Differential analysis based on CTSL expression levels revealed that genes such as TP53, APOB, and FRAS1 presented a relatively high mutation frequency in the high-CTSL-expressing group (Figure 7B). With respect to copy number variations (CNVs), the most commonly amplified genomic regions were at 1q22-Amp and 1q21.3-Amp, affecting up to 80% of the samples. Additionally, the most frequent CNV losses were 9q21.3-Del, 9q23-Del, and 13q12.11-Del, with frequencies of 60%, 59%, and 58%, respectively. However, no significant differences were observed between the high and low CTSL expression groups in terms of CNVs (Figure 7B). These findings suggest that CTSL may be more closely associated with gene mutations than with changes at the CNV level in LUAD.

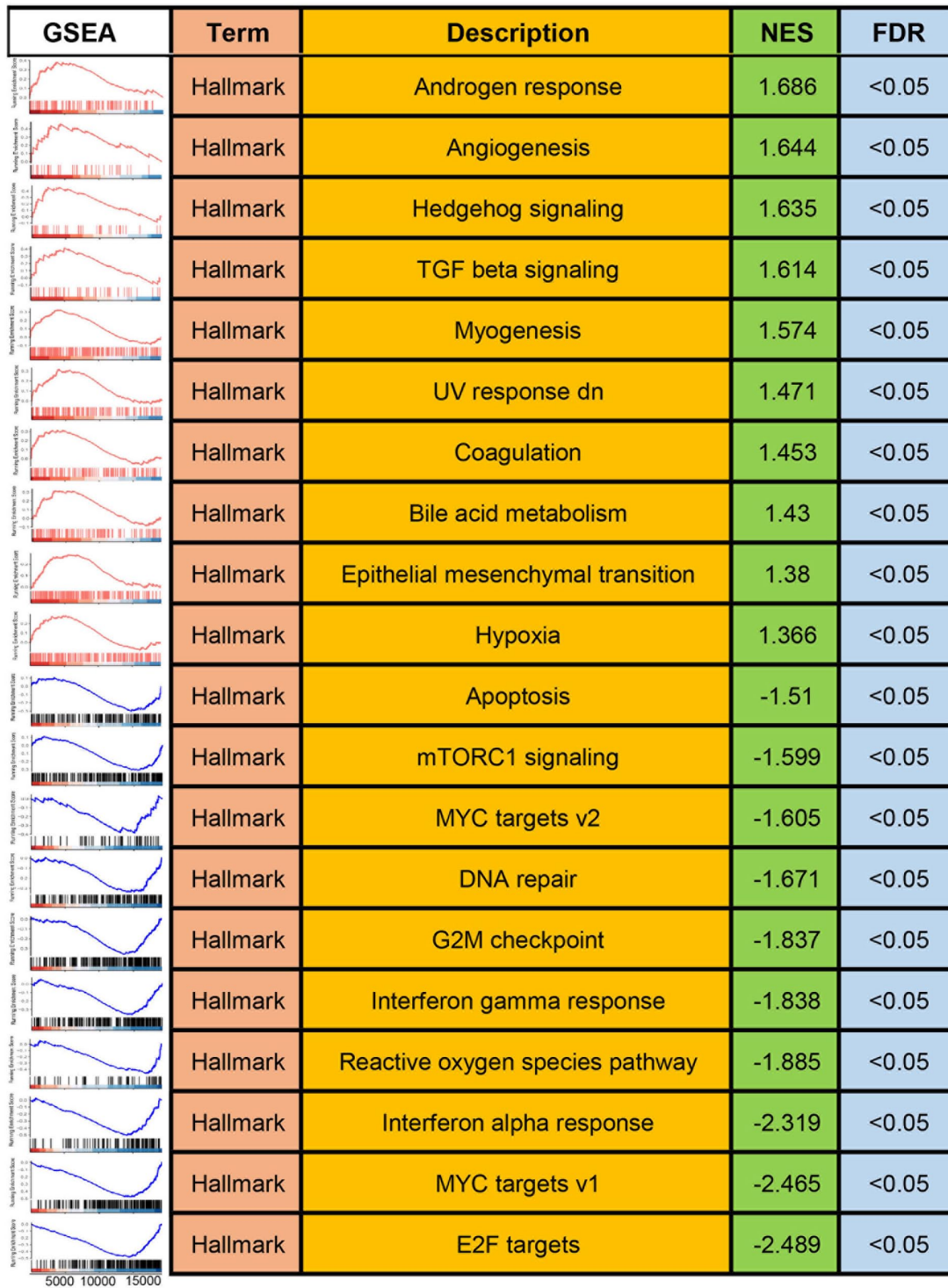


Figure 6. Functional enrichment analysis. Top 10 Hallmark pathways depicting activation and suppression by CTSL.

CTSL is associated with the immunotherapy response

Immune checkpoint inhibitors (ICIs) have become a pivotal treatment modality for LUAD. Following the identification of CTSL as a prognostic factor in LUAD, we explored its association with the immunotherapy response. In the anti-PD-1/

PD-L1 cohorts of Cho and Kim, both NSCLC patients, elevated CTSL expression was correlated with improved outcomes. Specifically, patients in the responder group presented increased CTSL expression levels, and the AUC for the ability of CTSL to predict treatment response exceeded 0.7 (Figure 8A and B). Furthermore, similar trends were observed in other cancer cohorts treated with immune checkpoint inhibitors;

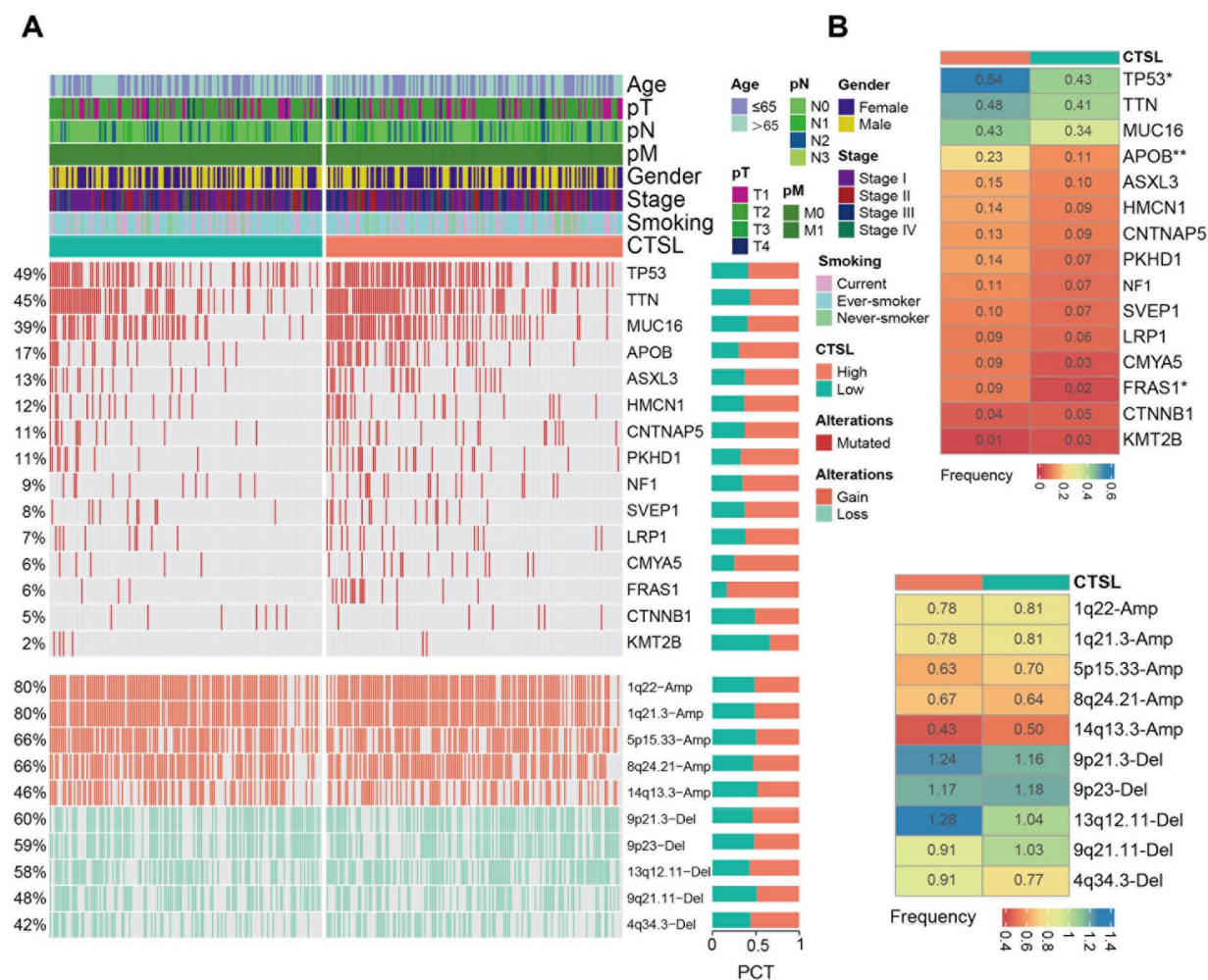


Figure 7. Mutation landscape: (A) heatmap presents an integrated genomic landscape of high and low CTSL expression and (B) mutational differences in FMGs of CTSL expression subgroups.

* $P < .05$. ** $P < .01$.

high CTSL expression was linked to better outcomes. Although the P values from the KM survival analysis did not reach statistical significance, there were notable differences in survival rates between the two groups (Figure 8C–E). These findings suggest that CTSL has prognostic value not only in postoperative LUAD patients but also as a potential biomarker for predicting response to immunotherapy.

Discussion

Lung cancer continues to be the leading cause of cancer-related deaths worldwide, despite significant advancements in diagnosis and treatment modalities.^{23,24} LUAD, the most common form of lung cancer, is characterized by substantial morphological heterogeneity and comprises tumor cells across a broad spectrum of histological subtypes.^{25,26} Through comprehensive bioinformatics analysis of big data, we identified candidate genes correlated with poor prognosis in patients with LUAD. This approach enables the prediction of target genes that may play crucial roles in the molecular progression of lung cancer, thus guiding the implementation of gene therapy in clinical settings. Capitalizing on these

extensive data repositories, our study identified CTSL as a promising biomarker for LUAD.

CTSL, a ubiquitously expressed lysosomal endopeptidase, functions primarily in the terminal degradation of intracellular and endocytosed proteins.^{24,25} Recent studies have highlighted the role of CTSL in the entry mechanism of SARS-CoV-2 into host cells, in which it cleaves the receptor-bound viral spike protein, facilitating viral activation and infection.²⁷ Furthermore, in the realm of cancer research, CTSL has been the subject of several studies. Its upregulation has been observed in various cancers, including ovarian, breast, prostate, lung, gastric, pancreatic, and colon cancers.^{10,11,14,28–30} Notably, increasing evidence suggests a correlation between CTSL expression and cancer grade and stage. Its potential role in advancing tumor progression and metastatic aggressiveness has also been proposed. Previous research has demonstrated a significant association between increased CTSL expression and reduced OS in patients with lung cancer and its contribution to chemotherapy resistance *in vitro*.^{12,31} Additionally, during tumor immunotherapy, elevated CTSL expression has been identified as an indicator of poor response.¹⁵ However, these findings

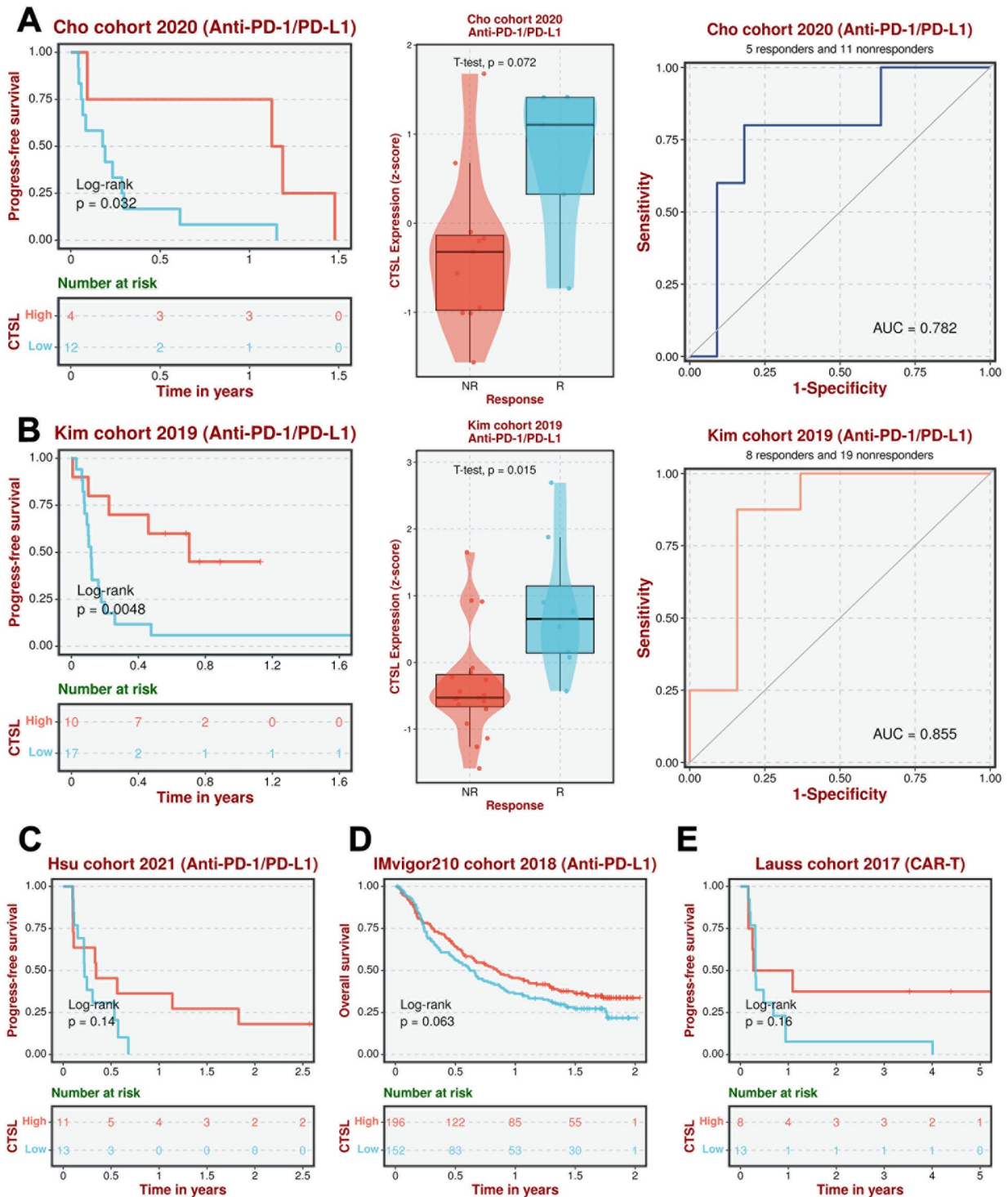


Figure 8. CTSL is a biomarker for therapy response in LUAD: (A and B) therapy response analysis of CTSL in LUAD anti-PD-1/PD-L1 cohorts and (C–E) therapy response analysis of CTSL in ICI cohorts.

have offered a somewhat one-dimensional view of CTSL in LUAD, lacking comprehensive validation through multiomics, multicenter, and local cohort validation. Our current research endeavors to fill this gap, providing a more holistic understanding of the role of CTSL in LUAD.

Our study revealed that CTSL serves as a detrimental prognostic factor not only in the TCGA-LUAD cohort, which predominantly consists of American patients, but also in Japanese,

Canadian, and our local Chinese cohorts. This uniformity in the role of CTSL across various ethnic backgrounds highlights its potential as a universal biomarker for LUAD prognosis. Interestingly, while CTSL serves as a negative prognostic factor in surgical intervention cohorts, it acts as a protective factor in immunotherapy cohorts. These results corroborate those of previous studies. This study revealed that CTSL is especially differentially expressed in myeloid cells. CTSL-related genes

are involved mostly in complement and coagulation cascades, immune response activation, cytokine-mediated signaling pathways, etc., according to functional analysis. Moreover, the CTSL high-expression group had higher StromalScore, ImmuneScore, and ESTIMATEScore, indicating enrichment of immune components and increased immune activity in the population.¹⁵ The interaction between PD-1 and PD-L1 leads to the suppression of T cells, causing tumor immune escape, which can be blocked by anti-PD-1/PD-L1.³² One hypothesis is that CTSL enhances the immune response, which is equivalent to enhancing the function of anti-PD-1/PD-L1. Moreover, high-CTSL patients tend to have a greater tumor mutation burden (TMB), and a previous study suggested that patients with a higher TMB might respond better to anti-PD1/PD-L1 therapy.³³ This observation suggests that, at an early stage, high CTSL expression predicts worse outcomes than low CTSL expression does; however, patients with high CTSL expression postsurgery might benefit more from prioritizing immunotherapy. This discovery could facilitate the development of more personalized treatment strategies for patients.

Moreover, CTSL also emerged as an indicator of poor prognosis in a range of solid tumors in our pancancer analysis. Acting as a cysteine protease, CTSL might be involved in a series of common pathways that are not confined to LUAD. We further explored the expression pattern of CTSL among various cancer types. Next, we explored the influence of CTSL on tumor cells. The knockout of CTSL not only affects lung cancer cell proliferation but also inhibits the growth of various cancer cell lines within the DepMap project. Remarkably, the dependency scores for the NCIH1573 LUAD cell line were among the highest. The functional enrichment analysis in our study suggested that the oncogenic effect of CTSL in LUAD may be attributed to the activation of pathways such as angiogenesis, Hedgehog signaling, TGF-beta signaling, epithelial-mesenchymal transition, and hypoxia. These pathways reveal that CTSL might affect cell proliferation, invasion and metastasis. The generation of blood vessels is also involved in tumor growth. In tumor cells, CTSL can be largely produced and augment the invasive/metastatic potential of cancer cells through direct degradative proteolysis of several components of the extracellular matrix (ECM) and basement membrane. In the nucleus, CTSL-processed CCAAT-displacement protein/cut homeobox (CDP/Cux) exhibits enhanced DNA-binding properties, which in turn confers a replicative and metastatic advantage. CDP/Cux promotes proliferation by accelerating the S phase of the cell cycle.¹³ These findings offer a comprehensive perspective on the role of CTSL in the context of LUAD.

Overall, we can view the comprehensive potential therapeutic implications of targeting CTSL in lung adenocarcinoma treatment. It is involved in several aspects of tumor development, such as invasion and proliferation, while it also acts in modulating the immune response and tumor microenvironment. Thus,

breakthroughs in both tumor biology and targeted therapy are needed. All these potential therapeutic implications may propel personalized therapy to distinguish CTSL levels among patients.

Limitation

It is important to acknowledge some limitations in our current study. For example, the analysis results derived from DepMap lack validation through in vivo and in vitro assays. Additionally, elucidation of the underlying mechanisms of CTSL in LUAD warrants further investigation in future research endeavors. In the future, further research on the molecular biology and immune profile of CTSL is needed. We hope to validate our conclusions clinically based on a large-scale study.

Conclusion

Overall, our results align with and build upon those of previous studies, establishing a foundation for further research to explore the broader implications of CTSL in different cancer types.

Abbreviations

AUC	area under the curve
CDP	CCAAT-displacement protein
CEC	cervical squamous cell carcinoma and endocervical adenocarcinoma
CNA	copy number alteration
CNV	copy number variation
COAD	colon adenocarcinoma
CRISPR	clustered regularly interspaced short palindromic repeats
CTSL	Cathepsin L
Cux	cut homeobox
DAB	diaminobenzidine tetrahydrochloride
DepMap	dependency map
DFS	disease free survival
DLBC	diffuse large B-cell lymphoma
ESCA	esophageal carcinoma
GBM	glioblastoma multiforme
GISTIC	Genomic identification of significant targets in cancer
HNSC	head and neck squamous cell carcinoma
HR	hazard ratio
HRP	horseradish peroxidase temperature
ICB	immune checkpoint blockade
ICI	immune checkpoint inhibitor
IHC	immunohistochemistry
KICH	kidney chromophobe
KIRC	kidney renal clear cell carcinoma
KIRP	kidney renal papillary carcinoma
KM analysis	Kaplan–Meier analysis
LGG	lower grade glioma
LIHC	liver hepatocellular carcinoma
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma

OS	overall survival
PFI	progression-free interval
PFS	progression-free survival
PRAD	prostate adenocarcinoma
READ	rectum adenocarcinoma
ROC	receiver operating characteristic
SARC	sarcoma
SKCM	skin cutaneous melanoma
STAD	stomach adenocarcinoma
TGF- β	transforming growth factor β
TMA	tissue microarray
TMB	tumor mutation burden
UVM	uveal melanoma

Acknowledgements

The authors would like to thank the researchers of the publicly available data. Our results published here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. The author would thank the TCGA Research Network of their work.

Author Contributions

Chao Cai, Jianjiang Xie and Gao Yi were instrumental in formulating the concept and design of this study. The task of gathering and analyzing the public datasets was carried out by Jianming Lu, Le Zhang, Jiaqi Liang, and Guifang Yu. The draft of the manuscript was prepared by Jianming Lu, Jiaqi Liang, and Gang Xiao. Additionally, the immunohistochemistry work was adeptly completed by Jianming Lu, Zitao He and Jiaqi Liang. All authors made significant contributions to the development of the article and have given their approval for the final version to be submitted.

Availability of Data and Materials

The datasets analyzed during the current study are available in public databases. This RNA sequencing (RNA-seq) data, mutation details, and clinical information for Lung Adenocarcinoma (TCGA-LUAD) and pan-cancer cohorts, sourced from the UCSC Xena database (<https://xenabrowser.net/datapages/>). To further substantiate our findings, LUAD datasets from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), namely GSE50081, GSE31210, and GSE68465, were collected. For any further inquiries, please contact the corresponding author.



Ethics Approval

LUAD tissue microarray (TMA) were obtained from Shanghai Zhuoli Biotech Company. The acquisition of these samples was approved by the Ethics Committee of Shanghai Zhuoli Biotech Company (approval number SHLLS-BA-22101102), with written informed consent acquired from each participant for the use of their samples in this study.

Consent to Participate

The written informed consent has been acquired from each participant.

ORCID iDs

Jianming Lu  <https://orcid.org/0000-0002-3794-641X>
Chao Cai  <https://orcid.org/0000-0001-7945-4301>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics. *CA Cancer J Clin.* 2023;73:17-48.
2. Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet.* 2018;391:1023-1075.
3. Miller M, Hanna N. Advances in systemic therapy for non-small cell lung cancer. *Br Med J.* 2021;375:2363.
4. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin.* 2016;66:75-88.
5. Huang H, Yang Y, Zhu Y, et al. Blood protein biomarkers in lung cancer. *Cancer Lett.* 2022;551:215886.
6. Wang M, Herbst RS, Boshoff C. Toward personalized treatment approaches for non-small-cell lung cancer. *Nat Med.* 2021;27:1345-1356.
7. Sholl LM. Next-generation sequencing from liquid biopsies in lung cancer patients: advances in comprehensive biomarker testing. *J Thorac Oncol.* 2017;12:1464-1466.
8. Zhao MM, Yang WL, Yang FY, et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduct Target Ther.* 2021;6:134.
9. Sun M, Ouzounian M, de Couto G, et al. Cathepsin-L ameliorates cardiac hypertrophy through activation of the autophagy-lysosomal dependent protein processing pathways. *J Am Heart Assoc.* 2013;2:e000191.
10. Cui K, Yao S, Liu B, et al. A novel high-risk subpopulation identified by CTSL and ZBTB7B in gastric cancer. *Br J Cancer.* 2022;127:1450-1460.
11. Sudhan DR, Pampo C, Rice L, Siemann DW. Cathepsin L inactivation leads to multimodal inhibition of prostate cancer cell dissemination in a preclinical bone metastasis model. *Int J Cancer.* 2016;138:2665-2677.
12. Zhao Y, Shen X, Zhu Y, et al. Cathepsin L-mediated resistance of paclitaxel and cisplatin is mediated by distinct regulatory mechanisms. *J Exp Clin Cancer Res.* 2019;38:333.
13. Sudhan DR, Siemann DW. Cathepsin L targeting in cancer treatment. *Pharmacol Ther.* 2015;155:105-116.
14. Han ML, Zhao YF, Tan CH, et al. Cathepsin L upregulation-induced EMT phenotype is associated with the acquisition of cisplatin or paclitaxel resistance in A549 cells. *Acta Pharmacol Sin.* 2016;37:1606-1622.
15. Huang L, Lou N, Xie T, Tang L, Han X, Shi Y. Identification of an antigen-presenting cells/T/NK cells-related gene signature to predict prognosis and CTSL to predict immunotherapeutic response for lung adenocarcinoma: an integrated analysis of bulk and single-cell RNA sequencing. *Cancer Immunol Immunother.* 2023;72:3259-3277.
16. Wu T, Hu E, Xu S, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb).* 2021;2:100141.
17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28:882-883.
18. Li R, Zhu J, Zhong WD, Jia Z. Comprehensive evaluation of machine learning models and gene expression signatures for prostate cancer prognosis using large population cohorts. *Cancer Res.* 2022;82:1832-1843.
19. Zhong C, Long Z, Yang T, et al. M6A-modified circRBM33 promotes prostate cancer progression via PDHA1-mediated mitochondrial respiration regulation and presents a potential target for ARSI therapy. *Int J Biol Sci.* 2023;19:1543-1563.
20. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28:1747-1756.
21. Gu Z, Hübschmann D. Make Interactive Complex Heatmaps in R. *Bioinformatics.* 2022;38:1460-1462.

22. Liu Z, Liu L, Weng S, et al. BEST: a web application for comprehensive biomarker exploration on large-scale data in solid tumors. *J Big Data*. 2023;10:165.
23. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA-A Cancer J Clin*. 2021;71:209-249.
24. Li C, Shao J, Li P, Feng J, Li J, Wang C. Circulating tumor DNA as liquid biopsy in lung cancer: biological characteristics and clinical integration. *Cancer Lett*. 2023;577:216365.
25. Lim Z-F, Ma PC. Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. *J Hematol Oncol*. 2019;12:134.
26. Zhan D, Zheng N, Zhao B, et al. Expanding individualized therapeutic options via genoproteomics. *Cancer Lett*. 2023;560:216123.
27. Zhao M-M, Zhu Y, Zhang L, et al. Novel cleavage sites identified in SARS-CoV-2 spike protein reveal mechanism for cathepsin L-facilitated viral infection and treatment strategies. *Cell Discovery*. 2022;8:53.
28. Zhang W, Wang S, Wang Q, Yang Z, Pan Z, Li L. Overexpression of cysteine cathepsin L is a marker of invasion and metastasis in ovarian cancer. *Oncol Rep*. 2014;31:1334-1342.
29. Singh N, Das P, Gupta S, et al. Plasma cathepsin L: a prognostic marker for pancreatic cancer. *World J Gastroenterol*. 2014;20:17532-17540.
30. Sigloch FC, Tholen M, Gomez-Auli A, Biniiossek ML, Reinheckel T, Schilling O. Proteomic analysis of lung metastases in a murine breast cancer model reveals divergent influence of CTSB and CTSL overexpression. *J Cancer*. 2017;8:4065-4074.
31. Cui F, Wang W, Wu D, He X, Wu J, Wang M. Overexpression of Cathepsin L is associated with gefitinib resistance in non-small cell lung cancer. *Clinic Translat Oncol*. 2016;18:722-727.
32. Niu M, Yi M, Li N, Luo S, Wu K. Predictive biomarkers of anti-PD-1/PD-L1 therapy in NSCLC. *Exp Hematol Oncol*. 2021;10:18.
33. Cristescu R, Mogg R, Ayers M, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science*. 2018;362:eaar3593.