Discover Oncology

Analysis

Appraising the causal role of cathepsins in genitourinary carcinoma: a two-sample mendelian randomization and prospective study based on 36,225 individuals

Qiyu Zhu[1] · Dingbang Liu[1,2] · Haoyang Liu[1] · Xiaohui Pan[3] · Yifu Shi[2] · Jingjing Guo[1] · Nanwei Tong[3] · Junru Chen[1] · Hao Zeng[1]

Received: 18 November 2024 / Accepted: 24 March 2025
Published online: 31 March 2025

Abstract

Background Cathepsin family proteases play an important role in the carcinogenesis of genitourinary carcinomas. However, the causality between serum cathepsin levels and genitourinary carcinomas remains uninvestigated.

Methods In this study, we conducted a two-sample Mendelian Randomization (MR) analysis exploring the causal association between different types of cathepsins and genitourinary carcinomas. Univariate, bidirectional and multivariate MR analyses were conducted based on the genome-wide association studies. Moreover, linkage disequilibrium score regression (LDSC) analysis, colocalization and transcriptomic analysis were also performed. 36,225 Individual data from UK biobank was utilized for further validation.

Results Our findings revealed seven causal associations following univariate analysis, in which five correlations were further validated in multivariate analysis. Cathepsin S (CTSS) was positively associated with papillary renal cell carcinoma (pRCC) [IVW: OR (95%CI) 1.444 (1.103–1.890), p: 8*10–3], and LDSC analysis indicated a genetic correlation between CTSS and pRCC [rg (SE): 0.559 (0.225); p: 0.013]. Other causal correlations included cathepsin B (CTSB), positively associated with testicular non-seminoma, and cathepsin L2 (CTSL2/CTSV), negatively associated with overall kidney cancer and pRCC. Transcriptomic analysis further validated the findings from MR analysis. In the UK biobank, CTSL2 was found to be negatively associated with the risk of cancer of the kidney [HR (95%CI) 0.567 (0.368, 0.873), p: 0.01].

Conclusions Cathepsins played an important role in urogenital carcinogenesis. Further large-scale studies are warranted for extended validation.

Keywords Cancer genetics · Adult oncology · Epidemiology · Urological tumours

Qiyu Zhu, Dingbang Liu, Haoyang Liu and Xiaohui Pan have co-first authors and contributed equally.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02219-6.

✉ Nanwei Tong, tongnw@scu.edu.cn; ✉ Junru Chen, jrchen@scu.edu.cn; ✉ Hao Zeng, kucaizeng@163.com | [1]Department of Urology, Institute of Urology, West China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, People's Republic of China. [2]West China School of Medicine, West China Hospital, Sichuan University, Chengdu 610041, People's Republic of China. [3]Department of Endocrinology and Metabolism, West China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, Sichuan, People's Republic of China.

Abbreviations

abf	Absolute value
AUC	Area under cumulative curve
BMI	Body mass index
ccRCC	Clear cell renal cell carcinoma
chr	Chromosome
chRCC	Chromophobe renal cell carcinoma
CPTAC	Clinical proteomic tumor analysis consortium
CTSB	Cathepsin B
CTSE	Cathepsin E
CTSF	Cathepsin F
CTSG	Cathepsin G
CTSH	Cathepsin H
CTSL2	Cathepsin L2
CTSO	Cathepsin O
CTSS	Cathepsin S
CTSZ	Cathepsin Z
DSS	Disease-specific survival
eaf	Effect allele frequency
GWAS	Genome-wide association study
HPA	The Human Protein Atlas
IVs	Instrument variables
IVW	Inverse variance-weighted
LDSC	Linkage disequilibrium score regression
MR	Mendelian randomization
OS	Overall survival
PFS	Progression-free survival
pos	Position
pRCC	Papillary renal cell carcinoma
ROC	Receiver operating characteristic
TCGA	The cancer genome atlas

1 Introduction

Approximately 597,160 newly diagnosed genitourinary cancer cases are expected to occur in 2024, with over 100,000 estimated deaths, accounting for 17% overall cancer-related death [1]. Prostate cancer is the mostly commonly diagnosed genitourinary cancer, followed by bladder cancer and kidney cancer.

Cathepsins play a vital role in the occurrence and development of various solid tumors, including lung cancer [2], colorectal cancer [3] and genitourinary carcinomas. They exert influence by promoting cell migration, regulating caspase-dependent apoptosis, and facilitating the digestion of recycled macromolecular [4–9]. Cathepsins can also impact the tumor homeostasis, thus potentially affecting the immune response and drug resistance process [8, 10, 11]. Recent study has revealed promoting effect of cathepsin B (CTSB) on the recurrence and survival outcomes of renal cell carcinoma mainly through down-regulation of its inhibitory factor, STFA [12]. Cathepsin K has also been revealed to promote proliferation of the prostate cancer through activating IL-17/CTSK/EMT axis and inducing increased M2 macrophage infiltration, while the expression of cathepsin D (CTSD), together with Thrombospondin 1, showed excellent discriminating ability between prostate cancer and benign prostatic hyperplasia [6, 13]. Furthermore, cathepsin L has been identified as having potential diagnostic value in cancers, typically testicular carcinoma and kidney cancer, due to its significant overexpression in tumor tissues [14]. All these findings suggested cathepsins played an unignorable role in carcinogenesis, progression and prognosis in patients with genitourinary carcinomas.

However, whether the cathepsins and genitourinary carcinogenesis were mutually influenced or causally effected remained largely unknown. Findings from existing studies are mostly based on small sample sizes, which could potentially lead to discrepancies across studies [4, 14].

Mendelian randomization (MR) analysis, based on large-scale genome-wide association studies (GWAS), can infer the causality between two traits (a risk factor and clinically relevant outcome) while minimizing the influence of acquired factors on the relationship by utilizing genetic variants as instruments variables (IVs) [15]. In this study, we performed bidirectional and multivariate MR analysis to investigate the causality between cathepsins and the carcinogenesis of genitourinary carcinomas, with multiple sensitivity analysis exploring the robustness of the results. We also performed (linkage disequilibrium score regression) LDSC analysis to explore potential genetic correlations and colocalization to assess whether the causalities were driven by a shared SNP. An observational study based on 36,225 participants from UK biobank with available data for cathepsins was performed to further validate the results.

2 Methods

This study was in accordance with the STROBE-MR guidelines (Supplementary Table 1) [16]. Three basic assumptions were necessary for the MR analysis: (1) the selected IVs associated with dietary habits; (2) the IVs are independent of any confounder; (3) the IVs influence the outcome only through exposure [17].

2.1 Data sources for exposures

Genetically predicted serum cathepsin levels were obtained from the INTERVAL study, comprising 3301 individuals of European ancestry, with SNPs being 10,265,284 for every GWAS data representing different cathepsins [18]. The INTERVAL study is a genomic bioresource containing 50,000 blood samples (150 µl) from healthy individuals from 25 assessment centers from UK [19]. Proteins were measured using SOMAscan assay and the genome-wide association tests were conducted using SNPTEST v2.5.2, with imputation uncertainty and covariates well addressed. The quality control and the missing data addressment for the blood samples from INTERVAL study was established and illustrated in previous study [20]. The trial was approved by the National Research Ethics Service (11/EE/0538), and the original data sources were accessible from the following link: (https://gwas.mrcieu.ac.uk). The selection criteria for cathepsin-related SNPs were as follows: (1) SNPs that were out of linkage disequilibrium ($r2 < 0.001$) within a window of 10,000 kb were eligible. (2) SNPs must be at the genome-wide significant level, with $p < 5*10^{-6}$. Linkage disequilibrium was estimated using the 1000 Genomes European Reference Panel locally [21]. The proportion of variance explained in the risk factor by the SNPs and the strength of the instrument were calculated with R^2 and F-statistic, respectively. An F-statistic less than 10 was considered indicative of a weak instrument and was excluded from the following analysis [22].

2.2 Genetic instruments for the genitourinary carcinomas

We obtained summary data for genetically predicted genitourinary carcinomas (prostate cancer; overall kidney cancer and its subtypes: papillary renal cell carcinoma (pRCC), chromophobe renal cell carcinoma (chRCC), clear cell renal cell carcinoma (ccRCC); overall testicular carcinoma and its subtypes: testicular non-seminoma, testicular seminoma; bladder cancer and pelvis cancer) from the most recently released FinnGen research project (https://www.finngen.fi/en) R10 [23], which contained summary data analyzed based on 412,181 Finnish biobank samples. The diagnostic criteria of the diseases were based on ICD-10. The selected instruments for MR analysis should not be significantly associated with the above-mentioned outcomes, with p-values required to be greater than $5*10^{-6}$. Detailed information about the cases and controls utilized for GWAS summary data calculation of the genitourinary carcinomas, along with other detailed information about the data sources, were displayed in Supplementary Table 2.

2.3 Statistical analysis for MR

Inverse-variance weighted (IVW) of random-effects model was used as the primary method to ascertain whether exposures had causal effects on the outcomes through selected instrumental variables (IVs). Additionally, MR-Egger, Weight Median, and MR-PRESSO were also utilized to further test the stability of the potential causal link and identify horizontal pleiotropy [24]. P value less than 0.05 indicated significant results. The study flowchart was displayed in Fig. 1. The p value less than 0.05 for the MR-Egger intercept and the global test of MR-PRESSO indicated pleiotropic effect. Scatter plots and funnel plots, along with leave-one-out analysis, were used to visualize potential outliers and evaluate the robustness of

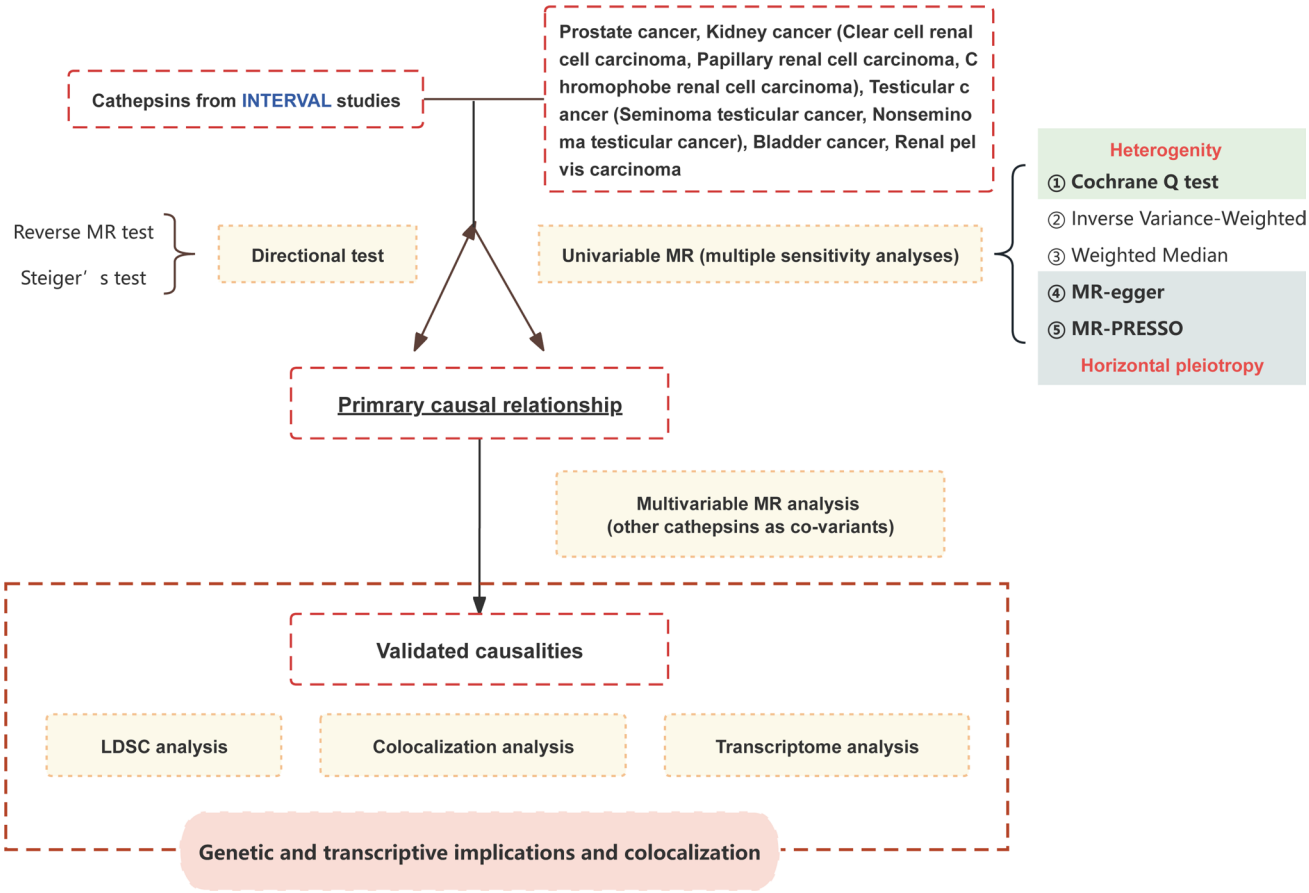

Fig. 1 Overview of the analytical plan of this mendelian randomization

the causal associations. The Steiger's test was also performed to examine the direction of the causal association between exposures and outcomes.

Multiple sensitivity analyses were conducted to create more robust results. Cochrane's Q test was performed to estimate heterogeneity, and a p-value less than 0.05 suggested significant heterogeneity. Multivariate MR was performed to exclude potential confounding effect from other types of cathepsins. Reverse MR was also conducted to explore any bidirectional causal links between exposures and outcomes. The genome-wide significant thresholds for genitourinary carcinomas in the reverse MR analysis were displayed in Supplementary Table 2.

LDSC analysis was carried out to estimate the liability-scale heritability (h^2) and examine the SNP-based genetic correlation (r_g) between cathepsins and genitourinary carcinomas. The regression was implemented based on pre-computed LD scores with reference to 1000 Genome European Data [25]. Colocalization analysis was performed based on GWAS summary data to identify whether a shared instrument existed between exposures and outcomes that were causally correlated. Colocalization was conducted based on chromosomal regions encoding cathepsins (e.g., CTSS, CTSH). There are four main hypotheses for posterior probability (PP.H0, PP.H1, PP.H2, PP.H3, PP.H4) of colocalization, in which the fourth hypothesis was selected as the evidence that the two traits were colocalized within the same genetic region (|PP.H4|> 0.5). Colocalization analysis was performed using "coloc" R package [26]. All statistical analyses were performed using R version (4.3.2).

2.4 Transcriptomic analysis

To compare the differences in expression profiles of specific genes between cancer samples and normal tissue samples, we downloaded and organized the transcriptome data from the Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov) database. The unpaired and paired sample data were visualized separately using R (version 4.3.2) and the R package "ggplot2", and the Wilcoxon test was used to verify the significance of differences. Based on the Clinical Proteomic

Tumor Analysis Consortium (CPTAC, https://pdc.cancer.gov/pdc/browse) data, we further validated the differences in proteomic expression of cathepsin-encoding gene between genitourinary carcinomas and normal tissues, if available. We further verified the differential expression of selected genes at protein level using the immunohistochemical results provided by The Human Protein Atlas (HPA, https://www.proteinatlas.org) [27]. Antibodies used for staining included HPA003524 and CAB017112.

2.5 Diagnostic and prognostic-related analysis

We fitted survival curves to the three prognostic types, including overall survival (OS), disease-specific survival (DSS), progress-free survival (PFS), based on TCGA data using the R packages "survival", "survminer", and "ggplot2". We then used the R package "pROC" and "ggplot2" to plot receiver operating characteristic (ROC) curves to assess the diagnostic value of specific genes for cancer [28]. We considered the results statistically significant if $p < 0.05$.

2.6 Individual data-based analysis from UK biobank

UK Biobank is a population-based cohort with over 500,000 participants aged 37 to 73, from 22 locations across England, Wales, and Scotland. Baseline data collected from 2006 to 2010 are linked to hospital and mortality records. Extensive sociodemographic, health behavior, and medical history information were obtained through touchscreen questionnaires and interviews. Physical measures and biological samples were collected by trained staff following standardized protocols. All participants gave written informed consent. Serum cathepsin level of 53,014 UK biobank participants were measured using the Olink platform by the UK Biobank Pharma Proteomics Project (UKB-PPP) [29]. We have excluded 14,001 participants without demographic data or covariates of concern (sex, age at recruitment, UK Biobank assessment center, education score, index of multiple deprivation, sleep duration, IPAQ score, smoking status, alcohol, overall health rating, body mass index (BMI), family history of cancer and ethnicity). We have also excluded the population with diagnosis of cancer before recruitment (n = 2758), leaving 36,225 individual data for final analysis.

As for the ascertainment of the outcomes, The UKB has developed a comprehensive dataset of "first occurrence" fields that map clinical codes from primary care visits, inpatient admissions, death records, and self-reported medical conditions to corresponding ICD-9 and ICD-10 codes. As per the most recent 2023 data update from UKB, the cutoff dates for inpatient data from the Hospital Episode Statistics for England (HES), Scottish Morbidity Record (SMR), and Patient Episode Database for Wales (PEDW) were respectively set as October 31, 2022, August 31, 2022, and May 31, 2022. Consequently, the analysis of genitourinary carcinoma was censored at the earliest of the following events: the first recorded occurrence of the genitourinary cancer of concern, the participant's death, or the cut-off date for inpatient data at the participant's corresponding hospital location.

3 Result

3.1 Univariate and bidirectional MR assessing the causal association between cathepsins and genitourinary carcinomas

Population from exposures and outcomes were both European ancestry and there were no sample overlaps between exposure and outcome population. We performed a two-sample MR analysis to explore the influence of nine types of cathepsins (CTSB, CTSE, CTSF, cathepsin G (CTSG), CTSH, cathepsin O (CTSO), CTSL2, CTSS, cathepsin Z (CTSZ)) on various genitourinary carcinomas, including prostate cancer, bladder cancer, overall kidney cancer (subtypes: ccRCC, pRCC, chRCC), renal pelvis cancer, and testicular carcinoma (subtypes: testicular seminoma, testicular non-seminoma). The associations revealed by IVW method are displayed in Fig. 2 and the significant associations are showed in Table 1.

Seven causal associations were identified, involving four cathepsins and six types of genitourinary carcinomas. CTSS [IVW: OR (95%CI) 1.486 (1.117–1.976), p: $6.50*10^{-3}$] showed significant positive correlation with pRCC, while CTSL2 [IVW: OR (95%CI) 0.504 (0.330–0.769), p: $1.50*10^{-3}$] was negatively associated with pRCC. Additionally, CTSS showed a weak negative causal effect on chRCC [IVW: OR (95%CI) 0.655 (0.431–0.996), p: 0.048]. CTSL2 [IVW: OR (95%CI) 0.868 (0.761–0.989), p:0.034], CTSH [IVW: OR (95%CI) 1.560 (1.111–2.192), p: 0.010], CTSB [IVW: OR (95%CI) 1.470 (1.050–2.058), p: 0.025], and CTSH [IVW: OR (95%CI) 0.932 (0.880–0.986), p: 0.015] showed positive correlations with overall kidney cancer, testicular seminoma, testicular non-seminoma, and prostate cancer, respectively. We conducted sensitivity analysis using multiple

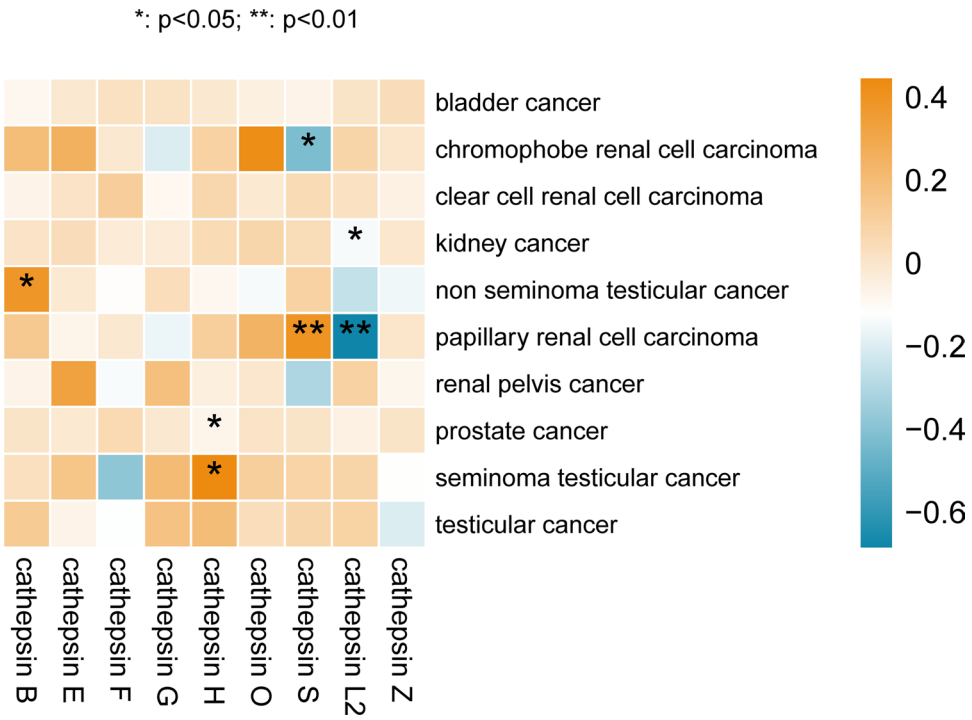

Fig. 2 Heatmap displaying the results of univariate mendelian randomization using inverse-variance weighted

Table 1 Univariate MR analysis revealed significant associations between cathepsins and genitourinary carcinoma

Outcome	Exposure	SNP	Inverse variance weighted		Weighted median		MR Egger	
			OR (95%CI)	P value	OR (95%CI)	P value	OR (95%CI)	P value
chRCC	CTSS	26	0.655 (0.431, 0.996)	0.048	0.770 (0.440, 1.347)	0.359	0.519 (0.260, 1.039)	0.076
Overall kidney cancer	CTSL2	13	0.868 (0.761, 0.989)	0.034	0.830 (0.697, 0.988)	0.036	0.909 (0.708, 1.169)	0.474
Testicular non-seminoma	CTSB	21	1.470 (1.050, 2.058)	0.025	1.159 (0.722, 1.862)	0.541	1.360 (0.661, 2.801)	0.414
pRCC	CTSS	26	1.486 (1.117, 1.976)	0.006	1.697 (1.163, 2.475)	0.006	1.345 (0.835, 2.167)	0.234
pRCC	CTSL2	13	0.504 (0.330, 0.769)	0.001	0.495 (0.270, 0.908)	0.023	0.596 (0.264. 1.344)	0.239
Prostate cancer	CTSH	23	0.931 (0.880, 0.986)	0.015	0.966 (0.900, 1.036)	0.328	0.943 (0.853, 1.043)	0.269
Testicular seminoma	CTSH	23	1.560 (1.111, 2.192)	0.010	1.187 (0.739, 1.908)	0.478	1.382 (0.763, 2.502)	0.298

MR mendelian randomization, *CTSB* cathepsin B, *CTSH* cathepsin H, *CTSS* cathepsin S, *CTSL2* cathepsin L2, *pRCC* papillary renal cell carcinoma, *chRCC* chromophobe renal cell carcinoma

methods (Weight Median, MR-Egger, MR-PRESSO) to ensure robustness of the above findings. The weighted median test yielded consistent results with IVW method. No horizontal pleiotropy was detected by MR Egger intercept and MR-PRESSO global test in all aforementioned associations between cathepsins and genitourinary carcinomas (Supplementary Table 3). Reverse univariate MR analyses revealed no existence of any reverse causality, indicating that the genitourinary carcinomas would not significantly affect the level of cathepsins (Supplementary Table 4). Detailed information about the SNPs utilized for both forward and reverse MR analysis was displayed in Supplementary Table 5 and Supplementary Table 6, respectively. Moreover, we visualized the outliers using funnel plots, scatter plots, and leave-out analysis plots, which indicated overall stability of the results (Supplementary Fig. 1, 2 and 3).

3.2 Multivariate MR for further validation of the causal roles of cathepsins on genitourinary carcinomas

We performed multivariate MR analysis, incorporating all nine cathepsins as covariates, to assess the robustness of the previously identified causal effects of cathepsins on genitourinary carcinomas (Fig. 3). The results showed that higher level of CTSS remained associated with an increased risk of the occurrence of pRCC after adjustment for the other eight types of cathepsins [IVW: OR (95%CI) 1.444 (1.103–1.890), p: $7.50*10^{-3}$]. CTSL2 showed a negative association

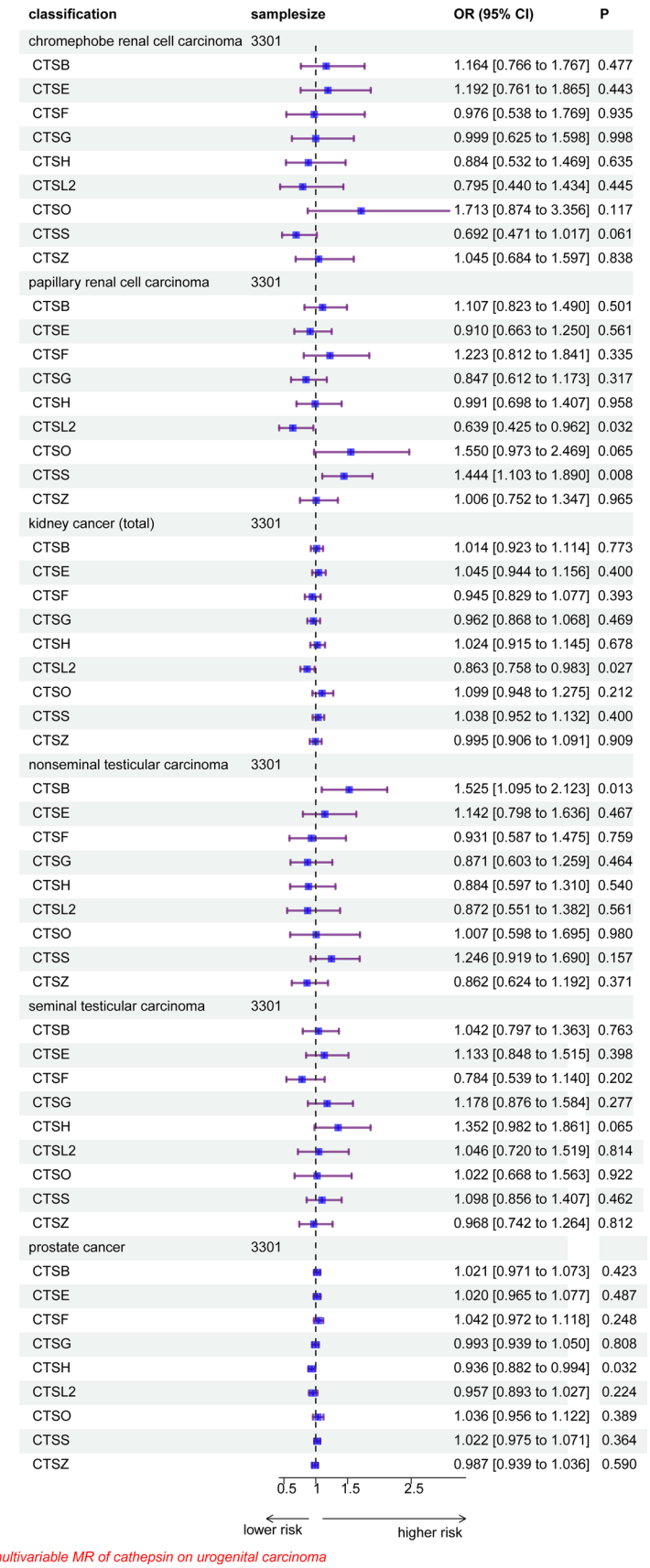

Fig. 3 Results of multivariate Mendelian Randomization using nine types of cathepsins as covariates exploring the causal associations among cathepsins and genitourinary cancers. *CTSB* cathepsin B, *CTSE* cathepsin E, *CTSF* cathepsin F, *CTSG* cathepsin G, *CTSH* cathepsin H, *CTSO* cathepsin O, *CTSS* cathepsin S, *CTSL2* cathepsin L2, *CTSZ* cathepsin Z, *ccRCC* clear cell renal cell carcinoma, *pRCC* papillary renal cell carcinoma, *chRCC* chromophobe renal cell carcinoma

with overall kidney cancer [IVW: OR (95%CI): 0.863 (0.758–0.983), p: 0.027] as well as pRCC [IVW: OR (95%CI) 0.639 (0.425–962), p: 0.032] in multivariate MR analysis. Additionally, CTSH showed a negative association with prostate cancer [IVW: OR (95%CI) 0.936 (0.882–0.994), p: 0.032], while CTSB showed a positive association with testicular non-seminoma [IVW: OR (95%CI) 1.525 (1.095–2.123), p: 0.013]. No significant causal associations were observed between CTSS and chRCC, as well as between CTSH and testicular seminoma in multivariate MR analysis. To further confirm the absence of horizontal pleiotropy and the robustness of the associations, we performed MR-Egger, Weighted-Median, and MR-Lasso, and the results are listed in the Supplementary Table 7.

3.3 Exploring genetic correlations between cathepsins and genitourinary carcinomas and colocalization

We further performed LDSC to investigate the genetic correlations between cathepsins and genitourinary carcinomas (Table 2). The results showed that CTSS exhibited a significant correlation with pRCC [rg (SE): 0.559 (0.225); p: 0.013]. No other genetic associations were observed between cathepsins and genitourinary carcinomas. Additionally, we conducted colocalization analysis to identify potential shared variants between genetically predicted cathepsins and genitourinary cancers. Genetic windows for colocalization around genes encoding cathepsins are displayed in Table 3. No colocalization was found (Table 3).

3.4 Transcriptomic analysis of cathepsins in genitourinary and potential diagnostic and prognostic values

Based on data from TCGA, we first compared the expression levels of cathepsin-encoding genes in genitourinary tumor samples and adjacent normal samples. The results suggested that CTSS was significantly overexpressed in pRCC tissues compared with normal tissues. CTSV, encoding cathepsin L2, was significantly lower expressed in overall kidney cancer and pRCC compared with normal tissues. CTSH gene expression levels were significantly lower in prostate cancer than in normal tissues (all $p < 0.001$, Supplementary Fig. 4a, b, 5a, b, c). Testicular non-seminoma showed higher CTSB expression levels compared with normal tissue (Supplementary Fig. 4c). Both unpaired and paired analyses showed consistent results. The transcriptomic differential expression results were consistent with the causal associations identified in MR analysis. Besides, we demonstrated that proteomic expression of CTSV was significantly lower in overall kidney cancer samples ($p < 0.0001$, Supplementary Fig. 5d). Immunohistochemical results further validated the low expression of CTSV in kidney cancer and CTSH in prostate cancer compared with normal tissues (Supplementary Fig. 4d, 5e).

ROC curves and area under the curve (AUC) suggested the accuracy of different genes encoding cathepsins in predicting the corresponding cancers: CTSH had moderate diagnostic accuracy in prostate cancer (AUC = 0.656, 95%CI 0.576–0.736, Supplementary Fig. 4i); CTSS (AUC = 0.777, 95%CI 0.699–0.854, Supplementary Fig. 4 h) and CTSV (AUC = 0.783, 95%CI 0.670–0.896, Supplementary Fig. 5i) had considerable diagnostic accuracy in pRCC; CTSV had relatively high diagnostic accuracy in overall kidney cancer (AUC = 0.865, 95%CI 0.831–0.898, Supplementary Fig. 5j); CTSB had extremely high diagnostic accuracy in testicular non-seminoma (AUC = 0.994, 95%CI 0.988–1.000, Supplementary Fig. 4j). Survival analyses suggested that none of these genes were associated with survival outcomes (Supplementary Fig. 4d, e, f, g, 5f, g, h).

3.5 Observational analysis based on individuals from UK biobank

To further validate the results drawn from MR analysis, we explored the associations between cathepsins and the risk of genitourinary carcinomas using clinical data of 36,225 individuals from UK biobank. The baseline characteristics of the cohort were displayed in Supplementary Table 8. The multivariable Cox regression results were showed in Table 4. We found that during a median follow up of 13.6 years, 112 participants developed cancer of the kidney. After adjusting for sex, age at recruitment, UK Biobank assessment center, education score, index of multiple deprivation, sleep duration, IPAQ score, smoking status, alcohol, overall health rating, body mass index (BMI), family history of cancer and ethnicity, CTSL2 was significantly associated with decreased risk of kidney cancer [HR (95%CI) 0.567 (0.368, 0.873), p = 0.01]. As for the testicular carcinoma, only one incidence has occurred during the entire follow-up, so we did not include it for further analysis.

Table 2 SNP-based heritabilities of cathepsins holding positive MR associations with the urogenital carcinomas and the genetic associations

Heritability			Lambda_GC	Heritability			Lambda_GC	Genetic correlation	
Trait 1	h2 (SE)	P value		Trait 2	h2 (SE)	P value		rg (SE)	P
Cathepsin B	0.386(0.325)	0.235	1.001	Testicular non-seminoma	0.006 (0.004)	0.095	1.008	0.039 (0.502)	0.938
Cathepsin H	0.123 (0.142)	0.388	1.009	Prostate cancer	0.05 (0.007)	2.35E−13	1.184	0.054 (0.007)	0.977
Cathepsin H	0.123 (0.142)	0.388	1.009	Testicular seminoma	0.01 (0.004)	7.00E−03	1.005	0.501 (0.379)	0.186
Cathepsin S	0.403 (0.216)	0.063	1.011	chRCC	0.001 (0.001)	0.425	0.989	− 0.026 (0.449)	0.954
Cathepsin S	0.403 (0.216)	0.063	1.011	pRCC	0.004 (0.001)	3.00E−03	1.022	0.559 (0.225)	0.013
Cathepsin L2	0.005 (0.002)	3.00E-03	1.038	Overall kidney cancer	0.005 (0.002)	3.00E−03	1.038	− 0.340 (0.675)	0.614
Cathepsin L2	0.005 (0.002)	3.00E-03	1.038	pRCC	0.004 (0.001)	3.00E−03	1.022	0.164 (0.699)	0.814

MR mendelian randomization, *CTSB* cathepsin B, *CTSH* cathepsin H, *CTSS* cathepsin S; *CTSL2* cathepsin L2, *pRCC* papillary renal cell carcinoma, *chRCC* chromophobe renal cell carcinoma

Table 3 Colocalization of genetic predicted cathepsin and genitourinary carcinoma in genomic regions

Trait1	Trait2	Colocalization window	nSNPs	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf
Cathepsin B	Testicular non-seminoma	± 100 kb	6190	0.000	0.528	0.000	0.442	0.030
Cathepsin D	Prostate cancer	± 100 kb	6297	0.327	0.392	0.123	0.147	0.011
Cathepsin D	Testicular seminal	± 100 kb	6290	0.289	0.346	0.155	0.186	0.024
Cathepsin S	chRCC	± 100 kb	4128	0.000	0.692	0.000	0.246	0.063
Cathepsin S	pRCC	± 100 kb	4128	0.000	0.576	0.000	0.221	0.203
Cathepsin L2	Overall kidney cancer	± 200 kb	4527	0.452	0.184	0.245	0.100	0.019
Cathepsin L2	pRCC	± 200 kb	4527	0.451	0.183	0.243	0.099	0.024

abf absolute value, *MR* mendelian randomization, *CTSB* cathepsin B, *CTSH* cathepsin H, *CTSS* cathepsin S, *CTSL2* cathepsin L2, *pRCC* papillary renal cell carcinoma, *chRCC* chromophobe renal cell carcinoma

Table 4 mutivariate Cox regression analysis showing correlations between cathepsin levels and risk of prostate cancer and kidney cancer

Cancer	Variable	Sample_size	Beta	HR(95%CI)	p.value	Cases
Prostate cancer	Cathepsin B	35,235	− 0.086	0.917 (0.821, 1.03)	0.129	960
	Cathepsin E	29,955	− 0.113	0.893 (0.783, 1.02)	0.092	793
	Cathepsin F	35,332	− 0.087	0.916 (0.793, 1.06)	0.238	965
	Cathepsin H	34,492	− 0.010	0.99 (0.916, 1.07)	0.808	936
	Cathepsin L	35,214	− 0.041	0.959 (0.754, 1.22)	0.736	950
	Cathepsin O	35,602	− 0.094	0.911 (0.754, 1.1)	0.334	967
	Cathepsin S	27,658	− 0.308	0.735 (0.541, 0.999)	0.050	735
	Cathepsin L2	35,420	0.048	1.05 (0.912, 1.21)	0.503	966
	Cathepsin Z	35,233	− 0.170	0.844 (0.702, 1.01)	0.069	954
Kidney cancer	Cathepsin B	35,235	− 0.063	0.939 (0.678, 1.3)	0.704	109
	Cathepsin E	29,955	0.095	1.1 (0.758, 1.6)	0.616	97
	Cathepsin F	35,332	− 0.130	0.878 (0.578, 1.34)	0.544	111
	Cathepsin H	34,492	0.067	1.07 (0.848, 1.35)	0.569	106
	Cathepsin L	35,214	0.813	2.25 (1.2, 4.24)	0.012	110
	Cathepsin O	35,602	0.550	1.73 (1.05, 2.86)	0.031	111
	Cathepsin S	27,658	0.805	2.24 (0.952, 5.26)	0.065	81
	Cathepsin L2	35,420	− 0.568	0.567 (0.368, 0.873)	0.010	110
	Cathepsin Z	35,233	0.594	1.81 (1.09, 3.02)	0.023	111

4 Discussion

In this study, we performed a two-sample MR analysis investigating the causal relationship between cathepsins and genitourinary cancers. After univariate and multivariate MR analysis, we found that CTSL2 was negatively associated with overall kidney cancer and pRCC. We also identified an inverse association between genetically predicted concentration of CTSH and prostate cancer, a positive association between CTSB and testicular non-seminoma, and a positive association between CTSS and pRCC. CTSS was revealed to be genetically correlated with pRCC in LDSC. The transcriptomic and proteomics analysis showed consistent results, indicating significantly differential expression of cathepsins in above-mentioned causal associations in corresponding genitourinary cancers. The AUC values of ROC curves further suggested a potential diagnostic value of these cathepsins in genitourinary cancers. External validation from UK biobank databased indicated higher CTSL2 to be an independent predictor for reduced risk of kidney cancer. Findings from our study could make contributions to the diagnosis and prevention of carcinomas deprived from genitourinary organs. For instance, patients with lower CTSL2 levels might have higher risk of developing kidney cancer compared to those with elevated CTSL2 levels.

High CTSL2 expression was revealed to be associated with tumor cell proliferation, drug resistance and poor prognosis in breast cancer, liver cancer and lung adenocarcinoma [30–32]. However, the relationship between CTSL2 and renal cell carcinoma was poorly investigated. Our results indicated that lower level of CTSL2 could potentially lead to higher risk of occurrence of overall kidney cancer and pRCC, and the result was further validated by differential analysis using RNA-seq data in TCGA databases and UK biobank. Due to the likeness to CTSL, it was hard to individually examine the functional role of CTSL2 in various types of cancers, including renal cell carcinoma [33]. Previous studies have indicated that CTSL was highly expressed in renal carcinoma tissue rather than adjacent normal tissue or renal embryonic cell line [34, 35]. However, Kirschke et.al reported decreased level of cathepsins in renal cell carcinoma compared with normal kidney tissue, which seemed to conflict with previous findings [4]. Based on individual data from UK biobank, we have found that CTSL was associated with increased risk of kidney cancer [HR (95%CI) 2.31 (1.23, 4.36), P = 0.009], while for CTSL2, this trend seemed to be reverse [HR (95%CI) 0.562 (0.364, 0.867), p = 0.009]. These findings might partially explain the inconsistency observed in the previous studies. Further large-scale observational studies are still needed to elucidate the expression patterns and functional implications of CTSL2 in RCC and other cancers.

One previous study revealed an important role of CTSS in resisting factor-related apoptosis-inducing ligand in human renal carcinoma Caki cell [36]. Although the direct impact of serum CTSS levels on the occurrence and progression of pRCC has not been investigated, insights from previous research suggest a potential link between these two traits through similar mechanisms. Liang et.al. found CTSS to be significantly higher expressed in pRCC compared with normal tissue in TCGA databases, which was also in accordance with our findings [37]. The multivariable MR analysis results, together with the AUC values generated from TCGA databases, indicated that CTSS could serve as a potential diagnostic tool to be integrated into other non-invasive screening models for the detection of pRCC.

We observed a significant positive effect of CTSB on testicular non-seminoma. Although previous studies have not specifically investigated the role of serum CTSB in the carcinogenesis of non-seminoma or seminoma, elevated levels of CTSB have been reported in multiple cancer types and were associated with carcinogenesis, invasion and metastasis of cancers [8]. It should be mentioned that due to little number of cases with testicular non-seminoma development or deaths, we did not perform relevant survival analysis (OS, DSS, PFS) or population-based analysis in UK biobank as it might be inconclusive. However, since diagnostic values of cathepsins to genitourinary carcinomas were the main emphasis of this study, this would not affect the conclusions drawn in the article.

There are several limitations in our study. First, the GWAS data of some cancer types (e.g., renal pelvis cancer) were generated based on small case proportion, potentially leading to underestimation or overestimation of the MR results. Some subdivisions of cancer were not available or of poor sample size in UK biobank, so we did not perform corresponding prospective analysis. Additionally, by using genetic variants as instruments to carry out causality investigation, pleiotropy served as an unavoidable issue as SNPs may affect the risk of genitourinary carcinogenosis by other factors rather than cathepsins. Therefore, we performed sensitivity analysis like MR-PRESSO and MR-Egger to mitigate the effect of pleiotropy. Population stratification of the GWAS data may also cause bias, which we performed multivariate MR analysis accordingly to minimize the influence from other cathepsins.

5 Conclusion

In summary, based on large-scale GWAS data, we performed univariate and multivariate MR as well as colocalization and LDSC analysis. We found positive causal associations between CTSS and pRCC, CTSB and testicular non-seminoma, and negative causal association between CTSL2 and pRCC and overall kidney cancer. Transcriptomic analysis and observational analysis further validated the result from MR analyses. These findings provide implications that diagnosis and prevention of urogenital carcinomas could be improved by testing the serum cathepsin levels of patients. Further in-vivo or in-vitro experiments and Large-scale observational studies are warranted to ensure the robustness of this study.

Acknowledgements We expressed our sincere gratitude towards the Dr. Zhongyu Jian who provided technical consultation and all the participants who participated in the original studies.

Patient and public involvement Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Author contributions QYZ, DBL, HYL and XHP provide equal contributions to this study. JRC and NWT are the correspond authors of this manuscript. QYZ: Formal analysis, methodology, writing—original draft; DBL: Formal analysis, writing—original draft; HYL, YFS: Writing—review and

editing; GXZ: Writing—original draft; JJG: Writing—review and editing; JRC: Writing—review and editing; XHP: Writing—review and editing; NWT: Writing—review and editing; HZ: Writing—review and editing.

Funding This work was supported by the National Natural Science Foundation of China (NSFC 82203110, 82172785, and 81974398), 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYJC21020, ZYGD22004), Science and Technology Support Program of Sichuan Province (2021YFS0119), Clinical and Translational Medicine Research Project, Chinese Academy of Medical Sciences (2022-I2M-C&T-B-098), Bethune Foundation, Oncology Basic Research Program (X-J-2020-016), and Bethune Foundation, Urological Oncology Special Research Fund (mnzl202002, mnzl202007).

Data availability The Finngen database was publicly available from https://finngen.gitbook.io/ documentation/. GWAS summary data for cathepsins from INTERVAL study was available from IEU OpenGWAS project (mrcieu.ac.uk). Data sources for transcriptomic analysis were publicly available from TCGA: https://portal.gdc.cancer.gov, CPTAC: https://pdc.cancer.gov/pdc/browse and HPA: https://www.proteinatlas.org. This research has been conducted using the UK Biobank Resource under Application Number 99427. The UK Biobank data are available on application to the UK Biobank (https:// www.ukbiobank.ac.uk/).

Declarations

Ethic approval and consent to participate The ethic approval was waived by West China Hospital ethics committee for studies involving humans as ethical approval was obtained from Ethical Committees in each original study. The written inform consent from each patients were also waived by ethics committee/institutional review board of West China Hospital because all participants gave inform consent to institution which conduct the original studies.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49.
2. Li J, et al. Mendelian randomization analyses explore the relationship between cathepsins and lung cancer. Commun Biol. 2023;6(1):1019.
3. Li R, et al. Gut microbiota-stimulated cathepsin K secretion mediates TLR4-dependent M2 macrophage polarization and promotes tumor metastasis in colorectal cancer. Cell Death Differ. 2019;26(11):2447–63.
4. Kirschke H, et al. Concentrations of lysosomal cysteine proteases are decreased in renal cell carcinoma compared with normal kidney. J Cancer Res Clin Oncol. 1997;123(7):402–6.
5. Weiss RE, et al. Mechanisms of human bladder tumor invasion: role of protease cathepsin B. J Urol. 1990;144(3):798–804.
6. Wu N, et al. Cathepsin K regulates the tumor growth and metastasis by IL-17/CTSK/EMT axis and mediates M2 macrophage polarization in castration-resistant prostate cancer. Cell Death Dis. 2022;13(9):813.
7. Batista AAS, et al. Decreased levels of cathepsin Z mRNA expressed by immune blood cells: diagnostic and prognostic implications in prostate cancer. Braz J Med Biol Res. 2021;54(10): e11439.
8. Mijanović O, et al. Cathepsin B: A sellsword of cancer progression. Cancer Lett. 2019;449:207–14.
9. Aits S, Jäättelä M. Lysosomal cell death at a glance. J Cell Sci. 2013;126(Pt 9):1905–12.
10. Zhitomirsky B, Assaraf YG. Lysosomes as mediators of drug resistance in cancer. Drug Resist Updat. 2016;24:23–33.
11. Pan H, et al. Autophagy-associated immune responses and cancer immunotherapy. Oncotarget. 2016;7(16):21235–46.
12. Rudzinska-Radecka M, et al. In silico in vitro, and clinical investigations of cathepsin b and stefin A mRNA expression and a correlation analysis in kidney cancer. Cells. 2022;11(9):10.
13. Steuber T, et al. Thrombospondin 1 and cathepsin D improve prostate cancer diagnosis by avoiding potentially unnecessary prostate biopsies. BJU Int. 2019;123(5):826–33.
14. Chauhan SS, Goldstein LJ, Gottesman MM. Expression of cathepsin L in human tumors. Cancer Res. 1991;51(5):1478–81.
15. Sekula P, et al. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol. 2016;27(11):3253–65.
16. Skrivankova VW, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. JAMA. 2021;326(16):1614–21.
17. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. JAMA. 2017;318(19):1925–6.
18. Sun BB, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73–9.

19. Di Angelantonio E, et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. Lancet. 2017;390(10110):2360–71.
20. Astle WJ, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167(5):1415-1429.e19.
21. Clarke L, et al. The 1000 Genomes Project: data management and community access. Nat Methods. 2012;9(5):459–62.
22. Palmer TM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21(3):223–42.
23. Kurki MI, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508–18.
24. Verbanck M, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693–8.
25. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–5.
26. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5): e1004383.
27. Uhlén M, et al. Proteomic tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
28. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010;5(9):1315–6.
29. Eldjarn GH, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. Nature. 2023;622(7982):348–58.
30. Song J, et al. High CTSL2 expression predicts poor prognosis in patients with lung adenocarcinoma. Aging. 2021;13(18):22315–31.
31. Liu J, et al. Cathepsin V is correlated with the prognosis and tumor microenvironment in liver cancer. Mol Carcinog. 2024;63(3):400–16.
32. Toss M, et al. Prognostic significance of cathepsin V (CTSV/CTSL2) in breast ductal carcinoma in situ. J Clin Pathol. 2020;73(2):76–82.
33. Lecaille F, et al. Cathepsin V: molecular characteristics and significance in health and disease. Mol Aspects Med. 2022;88: 101086.
34. Santamaría I, et al. Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. Cancer Res. 1998;58(8):1624–30.
35. Frolova AS, et al. Expression, intracellular localization, and maturation of cysteine cathepsins in renal embryonic and cancer cell lines. Biochemistry. 2023;88(7):1034–44.
36. Seo BR, et al. Inhibition of cathepsin S induces mitochondrial ROS that sensitizes TRAIL-mediated apoptosis through p53-mediated downregulation of Bcl-2 and c-FLIP. Antioxid Redox Signal. 2017;27(4):215–33.
37. Liang S, et al. Prognostic value and immunological role of cathepsin S gene in pan-cancer. Oncol Lett. 2024;27(1):41.