Journal of Hematology & Oncology

CORRESPONDENCE

Open Access

Integrated clinical and proteomic-based model for diagnostic and prognostic prediction in pRCC

Zeya Xu[1†], Linhui Zhang[1,2†], Jiacheng Lyu[1†], Maoping Cai[1,2†], Tao Ji[1†], Lin Bai[1,5†], Liqing Li[1,2†], Yao Zhu[1,2], Huashan Xu[1], Subei Tan[1], Hualei Gan[2,3], Shujuan Ni[2,3], Wenhao Xu[1,2], Xi Tian[1,2], Aihetaimujiang Anwaier[1,2], Beiyan Liu[2,3], Qinqin Hou[2,3], Guohai Shi[1,2], Hailiang Zhang[1,2], Jianyuan Zhao[4], Dingwei Ye[1,2*], Yuanyuan Qu[1,2*] and Chen Ding[1*]

Abstract

Papillary renal cell carcinoma (pRCC), a main pathological subtype of non-clear cell RCC (nccRCC), has strong heterogeneity. Comparing to other nccRCC subtypes, advanced pRCC has the poorest prognosis. Due to its lower incidence compared to ccRCC, clinical research and exploration of non-invasive biomarkers for pRCC are limited, and it is often misclassified. Herein, we leveraged the advantages of non-invasive plasma samples and the extensive coverage of mass spectrometry (MS)-based proteomics to develop a series of predictive models. First, we established the RCC subtype diagnostic model, which accurately differentiates pRCC, ccRCC, chromophobe RCC (chRCC), and healthy controls, achieving robust performance with an area under the receiver operating characteristic curve (AUROC) of 0.96 and averaged precision (AP) score of 0.91. Furthermore, recognizing the pivotal role of TNM staging in pRCC clinical management, we developed the the TNM staging diagnostic model with AUROC was 0.92 as the complementary noninvasive strategy. Finally, to facilitate real-time clinical monitoring of progression-free survival (PFS), we integrated routine blood indicators and proteomic features to develop the time-clock progression model, which demonstrated high predictive performance (AUROC > 0.95, AP > 0.95). In summary, this study provides a comprehensive plasma proteomic analysis and establishes diagnostic and prognostic predictive models for pRCC.

†Zeya Xu, Linhui Zhang, Jiacheng Lyu, Maoping Cai, Tao Ji, Lin Bai, and Liqing Li contributed equally to this work.

*Correspondence:
Dingwei Ye
dwyeli@163.com
Yuanyuan Qu
quyy1987@163.com
Chen Ding
chend@fudan.edu.cn
[1] Clinical Research Center for Cell-based Immunotherapy of Shanghai Pudong Hospital, Fudan University Pudong Medical Center, State Key Laboratory of Genetics and Development of Complex Phenotypes, School of Life Sciences, Human Phenome Institute, Department of Urology, Fudan University Shanghai Cancer Center, Fudan University, Shanghai 200433, China
[2] Department of Oncology, Shanghai Medical College, Fudan University, Shanghai Genitourinary Cancer Institute, Shanghai 200032, China
[3] Tissue Bank & Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China
[4] Institute for Development and Regenerative Cardiovascular Medicine, MOE-Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China
[5] Precision Research Center for Refractory Diseases, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 201620, China

To the editor

Renal cell carcinoma (RCC) was classified into clear cell RCC (ccRCC) and non-clear cell RCC (nccRCC) based on the morphological characteristics of tumor cells [1, 2]. Papillary RCC (pRCC) was the most common subtype of nccRCC, accounting for 10–20% of all RCCs [3]. Due to its lower incidence compared to ccRCC, clinical studies and non-invasive biomarker exploration of pRCC were limited, and it was often misclassified. Advanced mass spectrometry-based proteomic technologies can identify a large number of proteins altered under pathological conditions, providing opportunities to explore potential biomarkers [4].

In this study, we performed a comprehensive and robust plasma proteomic profiling on a total of 713 patients across three cohorts: the discovery cohort (n = 479), independent temporal validation cohort (n = 140), and independent external validation cohort (n = 94). The cohorts encompassed three RCC subtypes (pRCC, ccRCC, and chRCC) as well as healthy controls. Blood routine indictors were also collected. Specifically, the discovery cohort comprised 164 pRCC patients, 77 ccRCC patients, 76 chRCC patients, and 162 healthy controls. Meanwhile, to improve the credibility of the biomarkers, we further collected 15 paired tumor and adjacent normal tissue for pRCC and performed the proteome profiling. Independent validation cohort I included 48 pRCC patients, 23 ccRCC patients, 22 chRCC patients, and 47 healthy controls, while independent validation cohort II consisted of 21 pRCC patients, 32 ccRCC patients, 17 chRCC patients, and 24 healthy controls (Fig. 1A). The RCC subtypes diagnostic model, TNM staging diagnostic model and the real-time progression prediction model were constructed based on the plasma proteome data [5–7]. A detailed description of materials and methods can be found in Additional file.

Findings

To build a predictive model, we designed a machine learning pipeline consisting of eight parts: classification-target, feature-selection, feature-preprocessing, model-benchmarking, model-selection, hyperparameter-tuning, model-refitting, and model-evaluation [8] (Fig. 1A).

First, based on this pipeline, we constructed a RCC subtype diagnostic model, which classifies RCC subtypes and healthy controls. Specifically, 16 proteins with low collinearity were selected as the features for the model construction (Fig. S2A and Table S1). As for the model evaluation, we employed multiple metrics including the area under receiver operating characteristic curve (AUROC), balanced accuracy, recall, precision, and F1. As shown in Fig. 1B, the bootstrapping strategy results indicated the model had the good performance on the discovery cohort (Table S2). Moreover, to reduce sample imbalance and enhance model stability, we applied a bootstrapping approach to adjust the sample ratio, making it more representative of real-world epidemiological trends. The model's performance was outstanding with macro-average, micro-average, and weighted-average AUROC all exceeding 0.95, indicating the RCC subtype diagnostic model was robust under the different sample ratio (Fig. S2C). Furthermore, after refitting the model on discovery cohort, the model also showed good generalization ability on both independent temporal and external validation cohorts (AUROC > 0.9) (Fig. 1C, D and S2D). The average precision (AP) was 0.91 and 0.88, respectively. Notably, the APs for the three subtypes and healthy controls were all greater than 0.80 on both two validation cohorts (Fig. S2E). All evaluation results demonstrated that the model for diagnosing RCC subtypes had good predictive performance and generalization ability. Figure 1E showed the feature importance ranking of the RCC subtype diagnostic model (Table S3). To assess tissue-plasma consistency, we collected tumor and adjacent normal tissues from 15 pRCC patients and performed proteome profiling (Fig. S3A). We found the protein features with high ranking have the same tendency in both plasma and tissue proteome (Fig. S4). We further investigated clinical factors influencing pRCC prognosis by setting multiple potential risk factors as covariates including TNM stage, ISUP grade, sex, age, BMI, and WHO classification. The results revealed advanced TNM stage was significantly associated with survival (Fig. S5A), prompting the development of the TNM staging diagnostic model. The model showed the good performance on both discovery and validation cohort manifested as the AUROC > 0.90 for each of stages (Fig S5D and S5E). These results indicated the plasma proteome data could provide the complementary noninvasive strategy for the existed clinical diagnosis for TNM stage.

Additionally, considering the long-term clinical management owing to the advanced pRCC has the higher malignancy compared to other RCC subtypes, we developed a pRCC time-clock progression model to predict progression-free survival (PFS) status at different time points. Specifically, we selected four time points 2, 3, 4, and 5 years for the construction (Fig. 2A and S6A). The pRCC time-clock progression model has the good performance. Detailly, on the discovery cohort, the AUROC and AP at 2, 3, 4, and 5 years all exceeded 0.89 (Fig. S6B and S6C) and the bootstrapping results showed the robustness of the models (median value of AUROC > 0.80 and standard deviation < 0.05) (Fig. S7B). To assess the model's generalization performance, we calculated the metrics on the independent temporal validation cohort and the AUROC and AP at all four time points were

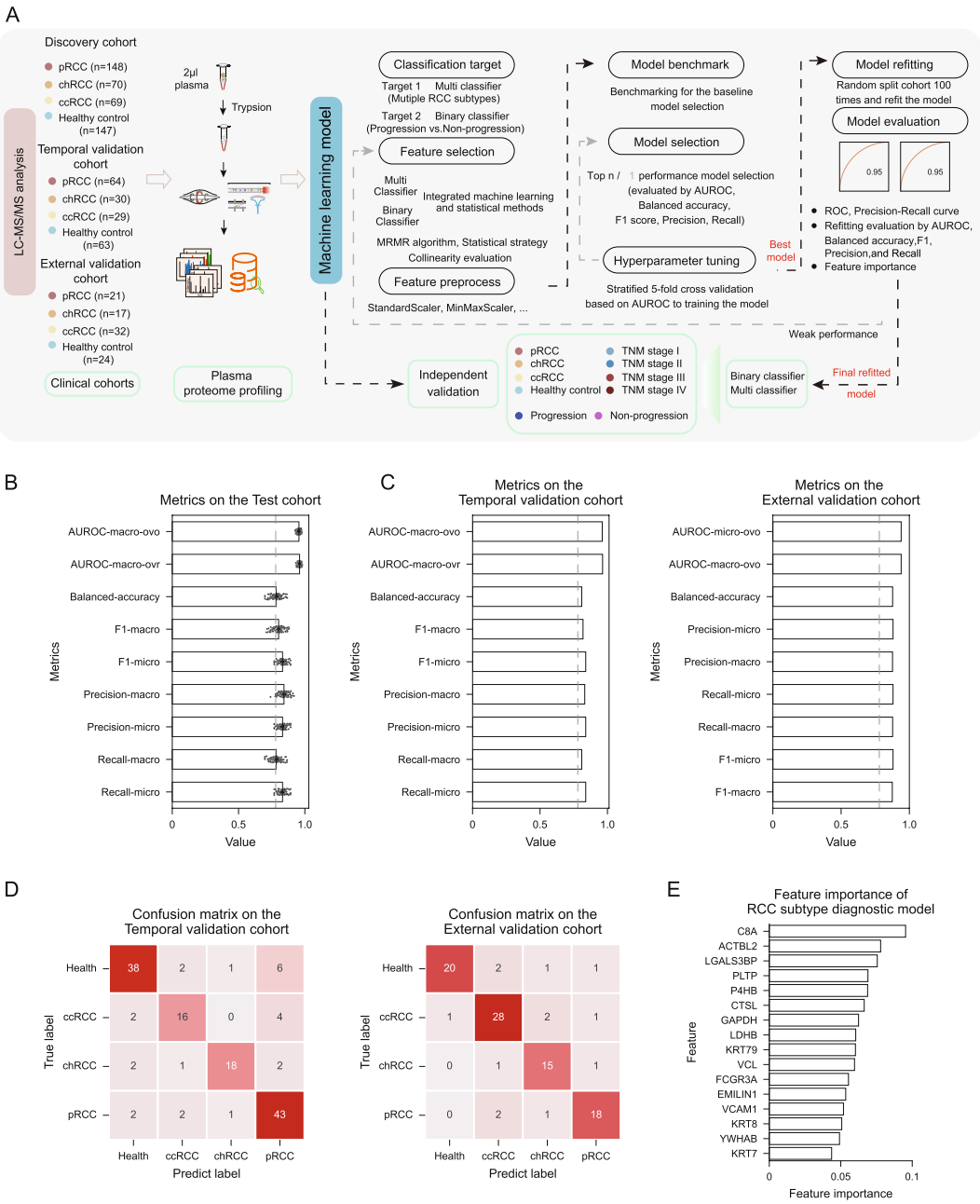

Fig. 1 Integrated model for discriminating RCC subtypes and healthy populations. **A** Schematic of the cohort design, proteome profiling, machine learning process for distinguishing models. The machine learning process consists of 8 parts: Classification target, Feature selection, Feature preprocess, Model benchmark, Model selection, Hyperparameter tuning, Model refitting and Model evaluation. **B** The evaluation for the RCC subtype diagnostic model on the test cohort by the bootstrap strategy. The area under the receiver operating characteristic curve (AUROC), balanced accuracy, precision, recall, and F1 score metrics were evaluated. **C** The model evaluation results of the RCC subtype diagnostic model on the independent temporal (left panel) and external (right panel) validation cohort. The AUROC, balanced accuracy, precision, recall, and F1 score metrics were evaluated. **D** The confusion matrix of the multi-classification model on the independent temporal (left panel) and external (right panel) validation cohort. **E** The bar plot depicting the feature importance of the 16 features in the decision process of the RCC subtype diagnostic model

greater than 0.9 (Fig. 2B and C). The confusion matrices of all four models demonstrated good generalization performance (Fig. 2D). Furthermore, we evaluated the feature importance of the four models (Fig. 2E and Table S4). These results indicate that the models exhibit robust generalization performance and may be beneficial

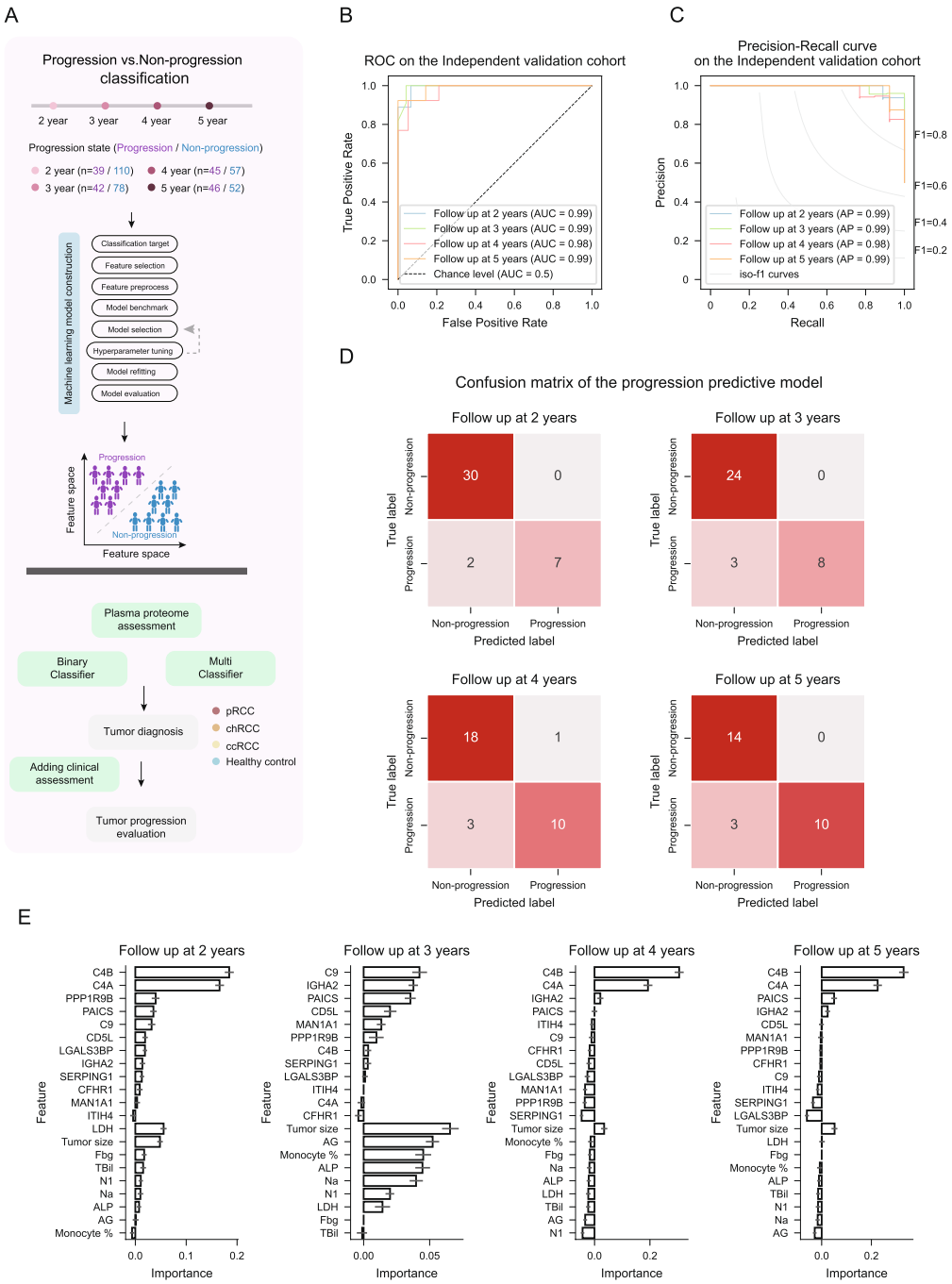

Fig. 2 Predicting the progression clock of pRCC. **A** Diagram describing the construction pipeline of the pRCC time-clock progression model used to predict progression and non-progression state in the pRCC population at 2, 3, 4, and 5 years, respectively. **B** The ROC curve of the four pRCC time-clock progression models on the independent temporal validation cohort. **C** The precision-recall curve of the four pRCC time-clock progression models on the independent temporal validation cohort. **D** Confusion matrix of the pRCC time-clock progression model at different time points in the pRCC population on the independent temporal validation cohort. **E** The bar chart depicting the important features of the different pRCC time-clock progression models

for prognostic management in the clinical practice of pRCC.

Overall, comprehensive proteomic analysis data combined with clinical indicators can facilitate the diagnosis of RCC subtypes (RCC subtype diagnostic model) and TNM staging (TNM staging diagnostic model), and the prediction of long-term PFS status (pRCC time-clock progression model), thereby enabling patient

stratification and contributing to personalized treatment of RCC, particularly for pRCC patients.

Abbreviations

pRCC	Papillary renal cell carcinoma
nccRCC	Non-clear cell RCC
ccRCC	Clear cell renal cell carcinoma
chRCC	Chromophobe renal cell carcinoma
AUROC	Area Under the Receiver Operating Characteristic Curve
PR	Precision-recall curve
AP	Average precision
PFS	Progression-free survival

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13045-025-01707-0.

Additional file 1. Supplemental file for detailed methods and results.

Acknowledgements
This work is supported by National Key Research and Development Program of China (2022YFA1303200 [C.D.], and 2022YFA1303201 [C.D.]), National Natural Science Foundation of China (32330062 [C.D.], and 31972933 [C.D.]), sponsored by Program of Shanghai Academic/Technology Research Leader (22XD1420100 [C.D.]), the Major Project of Special Development Funds of Zhangjiang National Independent Innovation Demonstration Zone (ZJ2019-ZD-004 [C.D.]), Shanghai Municipal Science and Technology Major Project (2023SHZDZX02 [C.D.]), the Fudan Original Research Personalized Support Project [C.D.], the National Ten Thousand Plan Young Top Talents [Y.Y.Q.], the Natural science foundation of China (No. 82172817 [Y.Y.Q.], 82172741 [W.D.Y.], 82473192 [W.D.Y.], 81972375 [Y.Z.]), Shanghai Rising-Star Program (No. 23QA1408900 [Y.Y.Q.]), Shanghai "Rising Stars of Medical Talent" Outstanding Youth Medical Talents [Y.Y.Q.], Shanghai "Science and Technology Innovation Action Plan" medical innovation research Project (22Y11905100 [Y.Y.Q.]), and Shanghai Municipal Health Bureau Project (No. 2020CXJQ03 [W.D.Y.]). This work is supported by Shanghai Municipal Science and Technology Major Project, the Human Phenome Data Center of Fudan university, and Shanghai Phenomic precision measurement professional technical service platform (23DZ2290800).

Authors' contributions
C.D., Y.Y.Q., and D.W.Y. conceived, designed and organized the study. Y.Y.Q., J.Y.Z., H.L.Z., L.H.Z., M.P.C., L.Q.L., H.L.G., Y.Z., S.J.N., W.H.X., X.T., A.A., B.Y.L., Q.Q.H., and G.H.S. were responsible for sample and clinical information collection. Z.Y.X., Y.C.J., T.J., and S.B.T. curated the data. Z.Y.X., H.S.X., and T.J. contributed to experiment. Z.Y.X., and Y.C.J. analysis the data. Z.Y.X., J.C.L., and L.B. interpreted the results and drafted the manuscript. C.D., and Y.Y.Q. supervised the study. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

Funding
This work is supported by National Key Research and Development Program of China (2022YFA1303200 [C.D.], and 2022YFA1303201 [C.D.]), National Natural Science Foundation of China (32330062 [C.D.], and 31972933 [C.D.]), sponsored by Program of Shanghai Academic/Technology Research Leader (22XD1420100 [C.D.]), the Major Project of Special Development Funds of Zhangjiang National Independent Innovation Demonstration Zone (ZJ2019-ZD-004 [C.D.]), Shanghai Municipal Science and Technology Major Project (2023SHZDZX02 [C.D.]), the Fudan Original Research Personalized Support Project [C.D.], the National Ten Thousand Plan Young Top Talents [Y.Y.Q.], the Natural science foundation of China (No. 82172817 [Y.Y.Q.], 82172741 [W.D.Y.], 82473192 [W.D.Y.], 81972375 [Y.Z.]), Shanghai Rising-Star Program (No. 23QA1408900 [Y.Y.Q.]), Shanghai "Rising Stars of Medical Talent" Outstanding Youth Medical Talents [Y.Y.Q.], Shanghai "Science and Technology Innovation Action Plan" medical innovation research Project (22Y11905100 [Y.Y.Q.]), and Shanghai Municipal Health Bureau Project (No. 2020CXJQ03 [W.D.Y.]). This work is supported by Shanghai Municipal Science and Technology Major Project, the Human Phenome Data Center of Fudan university, and Shanghai Phenomic precision measurement professional technical service platform (23DZ2290800).

Data availability
The proteome datasets have been deposited to the ProteomeXchange Consortium (https://www.iprox.cn/) under Project ID: IPX0009852000.
The link access to the raw data as follows:
https://www.iprox.cn/page/DSV021.html;?url=17280170749317vX9
The password: SUOg

Declarations

Ethical approval and consent to participate
The study was compliant with the ethical standards of Helsinki Declaration and was approved by the Institutional Review Board of Fudan University Shanghai Cancer Hospital (FUSCC) (050432-4-2307E). Written informed consent was obtained from each patient.

Competing interests
The authors declare no competing interests.

Received: 28 October 2024 Accepted: 26 April 2025
Published online: 28 May 2025

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70:7–30.
2. Padala SA, Barsouk A, Thandra KC, Saginala K, Mohammed A, Vakiti A, et al. Epidemiology of renal cell carcinoma. World J Oncol. 2020;11:79–87.
3. Courthod G, Tucci M, Di Maio M, Scagliotti GV. Papillary renal cell carcinoma: a review of the current therapeutic landscape. Crit Rev Oncol Hematol. 2015;96:100–12.
4. Ku X, Wang J, Li H, Meng C, Yu F, Yu W, et al. Proteomic portrait of human lymphoma reveals protein molecular fingerprint of disease specific subtypes and progression. Phenomics. 2023;3:148–66.
5. Li H-D, Liang C. Multigene panel predicting survival of patients with colon cancer. J Cancer. 2019;10:6792–800.
6. Wu M, Yuan H, Li X, Liao Q, Liu Z. Identification of a five-gene signature and establishment of a prognostic nomogram to predict progression-free interval of papillary thyroid carcinoma. Front Endocrinol. 2019;10:790.
7. Qu Y, Yao Z, Xu N, Shi G, Su J, Ye S, et al. Plasma proteomic profiling discovers molecular features associated with upper tract urothelial carcinoma. Cell Reports Med. 2023;4:101166.
8. Lyu J, Bai L, Li Y, Wang X, Xu Z, Ji T, et al. Plasma proteome profiling reveals dynamic of cholesterol marker after dual blocker therapy. Nat Commun. 2024;15:3860.

Publisher's Note