

# PRECISION MEDICINE AND TRANSLATIONAL RESEARCH IN UROLOGICAL ONCOLOGY

EDITED BY: Rong Na, Jonathan Olivier and Gong-Hong Wei  
PUBLISHED IN: Frontiers in Oncology





# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88976-672-7

DOI 10.3389/978-2-88976-672-7

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



# PRECISION MEDICINE AND TRANSLATIONAL RESEARCH IN UROLOGICAL ONCOLOGY

Topic Editors:

**Rong Na**, The University of Hong Kong, SAR China

**Jonathan Olivier**, Université de Lille, France

**Gong-Hong Wei**, Fudan University, China

**Citation:** Na, R., Olivier, J., Wei, G.-H., eds. (2022). Precision Medicine and Translational Research in Urological Oncology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-672-7

# Table of Contents

- 05 Editorial: Precision Medicine and Translational Research in Urological Oncology**  
Rong Na, Jonathan Olivier and Gong-Hong Wei
- 07 The Clinical Implications and Molecular Mechanism of CX3CL1 Expression in Urothelial Bladder Cancer**  
Guangliang Jiang, Hui Wang, Da Huang, Yishuo Wu, Weihong Ding, Qidong Zhou, Qiang Ding, Ning Zhang, Rong Na and Ke Xu
- 19 A Germline Variant at 8q24 Contributes to the Serum p2PSA Level in a Chinese Prostate Biopsy Cohort**  
Xiaoling Lin, Yishuo Wu, Fang Liu, Rong Na, Da Huang, Danfeng Xu, Jian Gong, Yao Zhu, Bo Dai, Dingwei Ye, Hongjie Yu, Haowen Jiang, Zujun Fang, Jie Zheng and Qiang Ding
- 27 Large Multicohort Study Reveals a Prostate Cancer Susceptibility Allele at 5p15 Regulating TERT via Androgen Signaling-Orchestrated Chromatin Binding of E2F1 and MYC**  
Xiaoming Dong, Qin Zhang, Jinglan Hao, Qianwen Xie, Binbing Xu, Peng Zhang, Haicheng Lu, Qilai Huang, Tielin Yang, Gong-Hong Wei, Rong Na and Ping Gao
- 44 Automated Classification of Papillary Renal Cell Carcinoma and Chromophobe Renal Cell Carcinoma Based on a Small Computed Tomography Imaging Dataset Using Deep Learning**  
Teng Zuo, Yanhua Zheng, Lingfeng He, Tao Chen, Bin Zheng, Song Zheng, Jinghang You, Xiaoyan Li, Rong Liu, Junjie Bai, Shuxin Si, Yingying Wang, Shuyi Zhang, Lili Wang and Jianhui Chen
- 54 Clinical Significance of Mesenchymal Circulating Tumor Cells in Patients With Oligometastatic Hormone-Sensitive Prostate Cancer Who Underwent Cytoreductive Radical Prostatectomy**  
Guanjie Yang, Jun Xie, Shun Zhang, Wenyu Gu, Jing Yuan, Ruiliang Wang, Changcheng Guo, Lin Ye, Bo Peng, Xudong Yao and Bin Yang
- 64 Deep Learning–Based Classification of Epithelial–Mesenchymal Transition for Predicting Response to Therapy in Clear Cell Renal Cell Carcinoma**  
Qiwei Chen, Yue Kuai, Shujing Wang, Xinqing Zhu, Hongyu Wang, Wenlong Liu, Liang Cheng and Deyong Yang
- 77 Identification of a Hypoxia-Related Gene Model for Predicting the Prognosis and Formulating the Treatment Strategies in Kidney Renal Clear Cell Carcinoma**  
Xiang-hui Ning, Ning-yang Li, Yuan-yuan Qi, Song-chao Li, Zhan-kui Jia and Jin-jian Yang
- 87 The Causal Relationships Between Extrinsic Exposures and Risk of Prostate Cancer: A Phenome-Wide Mendelian Randomization Study**  
Dongqing Gu, Mingshuang Tang, Yutong Wang, Huijie Cui, Min Zhang, Ye Bai, Ziqian Zeng, Yunhua Tan, Xin Wang and Ben Zhang

**97    *A Novel Nomogram for Prediction and Evaluation of Lymphatic Metastasis in Patients With Renal Cell Carcinoma***

Wenle Li, Bing Wang, Shengtao Dong, Chan Xu, Yang Song, Ximin Qiao, Xiaofeng Xu, Meijin Huang and Chengliang Yin

**108    *Fat Attenuation Index of Renal Cell Carcinoma Reveals Biological Characteristics and Survival Outcome***

Hongkai Wang, Yu Wei, Xiaoxin Hu, Jian Pan, Junlong Wu, Beihe Wang, Hailiang Zhang, Guohai Shi, Xiaohang Liu, Jinou Zhao, Yao Zhu and Dingwei Ye



# Editorial: Precision Medicine and Translational Research in Urological Oncology

Rong Na<sup>1\*</sup>, Jonathan Olivier<sup>2</sup> and Gong-Hong Wei<sup>3</sup>

<sup>1</sup> Division of Urology, Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, <sup>2</sup> Department of Urology, Centre Hospitalier Universitaire de Lille (CHU) Lille, Claude Huriez Hospital, University of Lille, Lille, France, <sup>3</sup> Fudan University Shanghai Cancer Center, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai, China

**Keywords:** precision medicine, translational research, prostate cancer, bladder cancer, kidney cancer, genetics, omics

## Editorial on the Research Topic

### Precision Medicine and Translational Research in Urological Oncology

Precision medicine has been widely acknowledged by scientists and physicians nowadays. The application of high-throughput omics technologies has provided a panorama of diseases at multiple levels. Thus, many diseases are reclassified based on their molecular spectrums for the purposes of personalized management. In recent years, plenty of research has been conducted to translate the omics/basic studies to clinical practice and precision treatment. Machine learning/artificial intelligence has been widely applied in the field of biomedical studies which helps overcome the large amount of manual computation. These approaches encourage the development of useful and practical clinical tools for personalized disease management.

The current Research Topic includes 10 multidisciplinary original research focusing on the “Precision Medicine and Translational Research in Urological Oncology”. Each study has addressed some of the critical issues in the field or provided useful tools for potential clinical applications. To be more specific, this special issue includes: 1) studies involving the discovery, mechanisms, and applications of biomarkers for genitourinary cancer diagnosis or prognosis (Dong et al., Gu et al., Lin et al., Yang et al. and Jiang et al.); 2) development of clinical tools using omics-based, big data, or machine learning approaches (Li et al, Ning et al, Chen et al., Wang et al., and Zuo et al.).

A series of original research in this issue investigated the discovery, mechanisms, and applications of biomarkers for genitourinary cancer diagnosis or prognosis. Dong et al. presented a comprehensive study bridging the GWAS findings and biological mechanisms in prostate cancer. They reported an interesting dual-directional regulation of androgen signaling pathway by a non-coding risk-associated germline common variant at TERT promoter region. The variant, therefore, could also serve as a potential biomarker for predicting prostate cancer risk and outcomes. A phenome-wide exposed-omics analysis of the risk factors for prostate cancer with subsequent causal inference by Mendelian randomization from Gu et al. discovered four potential novel biomarkers for prostate cancer. A germline common variant in 8q24.21 was identified to be associated with [-2]proPSA by Lin et al. via a GWAS in a Chinese cohort. This explains the phenomenon that baseline level of [-2]proPSA and its derivative prostate health index (phi) varies largely in individuals. It may also help us establish a genetic-adjusted biomarker with additional translational study in the future. As a significant biomarker for cancer prognosis, circulating tumor cells (CTCs) has been widely applied in clinical practice. Yang et al.

## OPEN ACCESS

### Edited and Reviewed by:

Ronald M Bukowski,  
Cleveland Clinic, United States

### \*Correspondence:

Rong Na  
yungna@hku.hk

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 12 June 2022

**Accepted:** 13 June 2022

**Published:** 05 July 2022

### Citation:

Na R, Olivier J and  
Wei G-H (2022) Editorial:  
Precision Medicine and  
Translational Research  
in Urological Oncology.  
Front. Oncol. 12:967278.  
doi: 10.3389/fonc.2022.967278

identified a subtype CTC called mesenchymal CTC that was significantly associated with poor survival of oligometastatic hormone-sensitive prostate cancer. Besides the studies in prostate cancer, Jiang et al. performed a systematic evaluation of a bladder cancer prognostic biomarker CX3CL1 including the functional study and the translational application.

Another series of important findings have been presented in this issue regarding the development of clinical tools using omics-based, big data, or machine learning approaches. A nomogram was established by Li et al. using clinical characteristics from big data for predicting the prognosis of renal cell carcinoma which was further confirmed in an independent cohort. At molecular level, a gene-based risk model based on large-scale omics studies was reported by Ning et al. which would have potential value for predicting the response to immune therapy in clear cell renal cell carcinoma. Meanwhile, Chen et al. provided an artificial intelligence approach to identify epithelial-mesenchymal transition subtype (poorer survival) of clear cell renal cell carcinoma *via* machine-based evaluation of H&E slides. Radiomics and imaging genomics is another popular topic during the recent years. In the current issue, Wang et al. reported an interesting imaging-based index named computed tomography (CT) fat attenuation index (FAI) for predicting renal cell carcinoma prognosis. This index was also found to be correlated with a molecular subtype of cancer. Incorporating with deep learning approach, Zuo et al. introduced a powerful tool distinguishing papillary renal cell carcinoma and chromophobe renal cell carcinoma based on CT imaging. These two subtypes of renal cell carcinoma are usually difficult to be distinguished manually *via* CT.

In summary, these 10 original studies collected in the current Research Topic provide invaluable evidence in terms of the precision medicine and translational research in different genitourinary cancers. Further translational studies are particularly promising based on the findings of these studies to apply the novel tools for personalized management to the diseases.

## AUTHOR CONTRIBUTIONS

RN wrote the editorial. JO and GW reviewed and finalized the editorial. All authors contributed to the article and approved the final version.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Na, Olivier and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# The Clinical Implications and Molecular Mechanism of CX3CL1 Expression in Urothelial Bladder Cancer

## OPEN ACCESS

### Edited by:

Francesca Sanguedolce,  
University of Foggia, Italy

### Reviewed by:

Sun Young Park,  
Pusan National University,  
South Korea  
Hongkai Wang,  
Shanghai Medical College of Fudan  
University, China

### \*Correspondence:

Ke Xu  
fudankexu@163.com  
Rong Na  
narong.hs@gmail.com  
Ning Zhang  
zhn\_hs@163.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 03 August 2021

**Accepted:** 14 September 2021

**Published:** 04 October 2021

### Citation:

Jiang G, Wang H, Huang D, Wu Y,  
Ding W, Zhou Q, Ding Q, Zhang N,  
Na R and Xu K (2021) The Clinical  
Implications and Molecular  
Mechanism of CX3CL1 Expression in  
Urothelial Bladder Cancer.  
Front. Oncol. 11:752860.  
doi: 10.3389/fonc.2021.752860

Guangliang Jiang<sup>1†</sup>, Hui Wang<sup>2,3,4†</sup>, Da Huang<sup>1†</sup>, Yishuo Wu<sup>2,3</sup>, Weihong Ding<sup>2,3</sup>,  
Qidong Zhou<sup>2,3</sup>, Qiang Ding<sup>2,3</sup>, Ning Zhang<sup>1\*</sup>, Rong Na<sup>1\*</sup> and Ke Xu<sup>2,3\*</sup>

<sup>1</sup> Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, <sup>2</sup> Department of Urology, Huashan Hospital, Fudan University, Shanghai, China, <sup>3</sup> Fudan Institute of Urology, Huashan Hospital, Fudan University, Shanghai, China, <sup>4</sup> Department of Nuclear Medicine, West China Hospital, Sichuan University, Chengdu, China

**Background:** CX3CL1 is a chemokine that may play important roles in cancer immune regulation. Its mechanism in bladder cancer (BCa) is poorly understood. The objective of the current study was to evaluate the association between CX3CL1 and BCa and the related biological mechanisms.

**Methods:** A total of 277 patients with BCa were enrolled in the present study. The association between CX3CL1 expression and disease outcome was evaluated. *In vitro* and *in vivo* experiments were performed using the TCCSUP cell line to investigate the function of CX3CL1 in BCa.

**Results:** Compared with low expression, high expression of CX3CL1 was significantly associated with poorer progression-free survival (hazard ratio [HR]=2.03, 95% confidence interval [95% CI]: 1.26-3.27, P=0.006), cancer-specific survival (HR=2.16, 95% CI: 1.59-2.93, P<0.001), and overall survival (HR=1.55, 95% CI: 1.08-2.24, P=0.039). Multivariable Cox regression analysis suggested that CX3CL1 was an independent prognostic factor for BCa outcomes. *In vitro* and *in vivo* experiments indicated that high expression of CX3CL1 was significantly associated with cell proliferation (P<0.001) and invasion (P<0.001). Gene expression profiling results showed that after CX3CL1 knockdown, CDH1 was significantly upregulated, while ETS1, RAF1, and EIF4E were significantly downregulated. Pathway enrichment analysis suggested that the ERK/MAPK signaling pathway was significantly inhibited (P<0.001).

**Conclusions:** CX3CL1 is an independent predictor of a poor prognosis in BCa and can promote the proliferation and invasion of BCa cells.

**Keywords:** CX3CL1, bladder cancer, survival analysis, ERK, MAPK

## INTRODUCTION

With an estimated 573,000 new cases and 212,000 deaths in 2020, urinary bladder cancer (BCa) has become one of the most common malignancies worldwide (1). In the United States, BCa accounted for 7% of all malignancy cases and ranked fourth among all types of malignancies in 2020 (2). Chemical components from cigarettes and industrial products are well-known risk factors; however, the etiology of BCa is poorly understood.

Previous evidence has shown that the immune system might play important roles in the development and progression of BCa, involving immune surveillance and immune tolerance. In immunoregulatory pathways, chemokines serve as key factors (3). Tumor immunotherapy represented by PD1/PD-L1 has attracted great attention in recent years. The effect of chemokines on tumors has been proven to produce a two-way regulation pattern: first, chemokines may enhance antitumor immunity by promoting immune cell recruitment to the tumor site (4); second, autonomous secretion of chemokines and receptors from tumor cells may promote tumor growth by inhibiting antitumor immunity (5).

A unique class of chemokines-the CX3C family has been reported playing a critical role in tumor immune regulation, in which CX3CL1 is the only member (6). It contains a unique 29nm long pedicle structure (mucus domain) in N prime/C prime. This structure allows chemokines to extend from epithelial cells to contact flowing white blood cells, and is independent of selectins and integrins (7). In addition, it can also induce intracellular calcium flow and chemotaxis without activating integrins (7). The receptor of CX3CL1, CX3CR1, is a transmembrane domain receptor, which mediates cell adhesion to CX3CL1 without G protein activation (8). Therefore, the unique structure and function of CX3CL1 has attracted the attention of many researchers.

As an important molecule in the chemokine family, C-X3-C Motif Chemokine Ligand 1 (CX3CL1) was found to be significantly associated with the development and progression of multiple cancers (9–13). However, the roles of CX3CL1 and its receptors are still controversial in terms of the biological regulation of tumors. Studies have suggested that CX3CL1 exerts its tumor-suppressive effect by activating immune cells, including natural killer T cells, and inhibiting tumor cell aggregation and adhesion (14, 15). For example, in a study of breast cancer, the high expression of CX3CL1 might recruit immune cells, such as CD8+ T cells and NK cells, to exert anti-tumor activity (16). Another study showed that patients with high expression of CX3CL1 and CX3CL1 receptors would have a significantly better prognosis in liver cancer (15). However, CX3CL1 may also produce a cancer-promotive effect as an adhesion molecule (9, 13). It was reported that high expression of CX3CL1 can induce tumor cell proliferation and promote the transformation of the cell cycle to S phase in gastric cancer (17). Shulby et al. indicated that CX3CL1 involved in the invasion process of prostate cancer metastasis to bone marrow which could be significantly reduced by anti-CX3CL1 antibody (13). This two-way regulation pattern of CX3CL1 is still unclear in BCa.

The objective of the current study was to evaluate the association between CX3CL1 and BCa and to investigate the underlying biological mechanisms involving CX3CL1 and its receptors in BCa.

## MATERIALS AND METHODS

### Fresh Tissue Sample

Seven fresh tumor tissue samples of bladder cancer and normal bladder mucosa tissue were collected by TURBT (transurethral resection of bladder tumor). In the end, the specimens due to severe burning and too small volume were eliminated. The qualified samples were divided into two groups (4 cases in the normal tissue group; 5 cases in the tumor tissue group) for Real-Time PCR experiment to determine the difference in CX3CL1 mRNA between the two groups.

### Clinical Samples

Patients with BCa undergoing surgical treatment were enrolled in the present study between December 2011 and December 2016 at Huashan Hospital (Shanghai, China). The inclusion criteria were as follows: (a) sufficient clinical information and follow-up information were available, (b) pathological diagnosis and grading were independently confirmed by three pathologists as transitional cell carcinoma of the bladder, and (c) sufficient formalin-fixed paraffin-embedded (FFPE) specimens could be obtained for further investigation. Informed consent was obtained from all patients for institutional biobank, and the study was approved by the Institutional Review Board at Huashan Hospital. Pathological grading and staging were performed according to the WHO/ISUP (2004) and UICC-TNM (2002) guidelines (18, 19). Ta-T1 was defined as non-muscle-invasive BCa (NMIBC), and T2-T4 were defined as muscle-invasive BCa (MIBC).

### Immunohistochemistry

Immunohistochemistry staining of FFPE samples was performed using an anti-CX3CL1 antibody (Abcam, Cambridge, UK; 1:100) and the avidin-biotin-complex method. The extent and intensity of tumor regions (without necrotic tissue) were further evaluated under a Nikon 80i microscope (Tokyo, Japan). Staining intensity was categorized as 0 (absent), 1 (weak), 2 (moderate), or 3 (strong). An H-score was then derived by multiplying the staining intensity and the percentage of cells stained. Given that there is still no established cutoff value for CX3CL1 expression, a cutoff value of 10% immuno-positive tumor cells was recommended by the 3 experienced pathologists (Figure S1). Disagreement as to whether >10% of cells were positively stained was resolved by consensus.

### RNA Isolation and Quantitative PCR (qPCR)

Total RNA was extracted from tissues with TRIzol (Pufei, Shanghai, China) and retrotranscribed using PrimeScript RT Master Mix (Perfect Real Time; TaKaRa, Shiga, Japan) following the manufacturer's recommendations. For qPCR, we used SYBR



Premix Ex Taq (Tli RNaseH Plus; TaKaRa) and the 7500 Real-Time PCR System with Dell Tower (Applied Biosystems, Foster City, USA) system. The primers used to amplify CX3CL1, GAPDH, etc. were purchased from Ji Kai (Shanghai, China). The sequences of the primers for CX3CL1 were 5'-GTAGCTTTGCTCATCCACTATCA-3' (upstream) and 5'-GACCACAGACTCGTCCATTCC-3' (downstream). For each sample, the average value of the threshold cycle was normalized to the GAPDH level with the  $2^{-\Delta\Delta Ct}$  method.

## Cell Lines and Viral Infection

T24 (RRID: CVCL\_0554), 5637 (RRID: CVCL\_0126) and TCCSUP (RRID: CVCL\_1738) bladder urothelial cancer cells were purchased from Cell Bank of Chinese Academy of Sciences (Shanghai, China) and cultured in DMEM (Gibco, Waltham, USA) supplemented with 10% fetal bovine serum (FBS; PAA).

Approximately  $1 \times 10^6$  cells were cultured for 12 h to allow attachment to plates. Cells were rinsed once with PBS, and 10 mL of serum-free medium was added to the plate. We used an adenovirus (Ji Kai, Shanghai, China) that overexpressed shRNA targeting CX3CL1 to infect the cells at a multiplicity of infection (MOI) of 10. Sixteen hours after infection, the medium was replaced with 10 mL of complete medium.

## Cell Viability Assays and Transwell Assays

Cell viability was assessed at the indicated time points using an MTT kit (Gen-View, Calimesa, USA). MTT reagent (20  $\mu$ L/well) was added to the medium, and the absorbance was read using a microplate reader at 490 nm for the test length and 570 nm for the reference. The percentage of viable cells was calculated based on the absorbance of the PBS control. Inserts in 24-well Transwell plates (Corning Costar, New York, USA) were coated with 500  $\mu$ L/well Matrigel (BD Bioscience, San Jose, USA). Rehydration was performed with 500  $\mu$ L/well serum-free medium for 2 h in a 37°C incubator (Sanyo, Japan). Samples were seeded into the inserts after removal of the rehydration medium. After 24 h of incubation, the cells that invaded through the membrane were stained with crystal violet (Beyotime, Shanghai, China) and observed at 200 $\times$  magnification using a light microscope (Olympus, Tokyo, Japan).

## RNA-Seq Expression Profiling and Data Analysis

CX3CL1 was knocked down in TCCSUP cells by lentiviral infection [virus strain: LV-CX3CL1-RNAi (52620-1)]. RNA was extracted using the TRIzol method used for qPCR. cDNA was obtained by reverse transcription (Promega M-MLV kit) *via* a real-time qPCR detection system (Type Roche-LightCycler 480). The knockdown efficiency of the CX3CL1 gene was also evaluated by qPCR.

GeneChip PrimeView detection: RNA samples were subjected to quality inspection analysis (Agilent 2100 Bioanalyzer), and the GeneChip 3'IVT Express Kit was used to prepare amplified RNA (aRNA). That is, cDNA was obtained through first-strand synthesis, a double-stranded DNA template was obtained through second-strand synthesis, and then biotinylated aRNA was obtained by *in vitro* inversion. The aRNA was purified, fragmented, hybridized, washed with a

chip probe (GeneChip Hybridization Wash and Stain Kit), and finally scanned to obtain images and raw data.

Statistical analysis and verification: Quality control was performed on the original data, statistical testing was used to identify differential genes between paired samples, and scatter plots, volcano plots, cluster plots and other chip analysis result plots were drawn. Ingenuity Pathway Analysis (IPA) database software was used to perform disease and function enrichment analysis and pathway enrichment analysis of the differentially expressed gene list; the results were verified by Western blot analysis.

## Xenografts and Animal Models

All experimental animals were obtained from Shanghai SLAC Laboratory Animal Co., Ltd. This work was carried out with 16 (8 male and 8 female) BALB/C nude mice aged 4-6 weeks and weighing 15 to 20 grams. All mice were kept under SPF living conditions. Mice were divided randomly and equally into 2 groups. The control group was named the negative control (NC) group and injected with normal TCCSUP cells, and the experimental group was named the KD group and injected with TCCSUP CX3CL1-knockdown cells. Each mouse was injected with 200  $\mu$ L of cell suspension in the shoulder, which contained  $1 \times 10^7$  cells. We measured tumor volume every 5 days after injection, and calculated tumor volume with the following formula: Volume =  $0.524 \times L \times W^2$  (L: long diameter; W: short diameter). The nude mice were sacrificed 35 days after injection, and tumor weight was measured. The subcutaneous tumors from each nude mouse were evaluated in pathological paraffin-embedded sections.

## Statistical Analysis

SPSS 21.0 (IBM, New York, USA) was used for statistical analyses. All data are presented as the mean  $\pm$  standard deviation (SD). Student's t-test and the Wilcoxon test were applied to compare gene expression and staining. Univariable and multivariable Cox/logistic regression analyses were used to evaluate the correlation between CX3CL1 expression and prognosis. Kaplan-Meier curve analyses and the log-rank test were also performed. A P value <0.05 was considered statistically significant.

## RESULTS

The expression level of CX3CL1 mRNA in bladder cancer tissues was significantly higher than that in normal tissues (**Figure S2**,  $P=0.0268$ ). A total of 277 patients were enrolled in the present study, with an average follow-up time of 49 months (1 to 72 months). Among them, 50 (18.1%) were negative for CX3CL1 expression, and 227 (81.9%) were considered positive for CX3CL1 expression. The demographic and clinical characteristics of the patients are summarized in **Table 1**. The patients were divided into 2 groups according to their expression of CX3CL1. By univariable analysis, the expression status of CX3CL1 was significantly associated with tumor stage [odds ratio (OR)=2.55, 95% confidence interval (95%CI): 1.36-4.76,  $P=0.003$ ], tumor size (OR=2.06, 95%CI: 1.10-3.86,  $P=0.025$ ), tumor grade (OR=1.91, 95%CI: 1.02-3.55,  $P=0.043$ ), disease recurrence (OR=2.05, 95%CI: 1.08-3.88,  $P=0.029$ ), metastatic disease (OR=3.98, 95%CI: 1.95-



**TABLE 1 |** Clinical characteristics of entire cohort and subgroups by CX3CL1 expression.

Characteristics, n (%)	Entire Cohort (N = 270)	CX3CL1		P value
		Negative (n = 50)	Positive (n = 227)	
Age				
≤65 years	129 (46.6)	11 (8.5)	118 (91.5)	<0.0001
>65 years	148 (53.4)	39 (26.4)	109 (73.6)	
Gender				
Male	234 (84.5)	45 (19.2)	189 (80.8)	0.233
Female	43 (15.5)	5 (11.6)	38 (88.4)	
BMI				
<24	158 (57.0)	27 (17.1)	131 (82.9)	0.631
≥24	119 (43.0)	23 (19.3)	96 (80.7)	
Smoking				
No	194 (70.0)	32 (16.5)	162 (83.5)	0.262
Yes	83 (30.0)	18 (21.7)	65 (78.3)	
Tumor size				
<3cm	123 (44.4)	27 (22.0)	96 (78.0)	0.131
≥3cm	154 (55.6)	23 (14.9)	131 (85.1)	
No. of tumor sites				
Single	125 (45.1)	21 (16.8)	104 (83.2)	0.624
Multiple	152 (54.9)	29 (19.1)	123 (80.9)	
Tumor stage				
Ta-T1	188 (67.9)	25 (13.3)	163 (86.7)	0.003
T2-T4	89 (32.1)	25 (28.1)	64 (71.9)	
Tumor grade				
Low	147 (53.1)	20 (13.6)	127 (86.4)	0.041
High	130 (46.9)	30 (23.1)	100 (76.9)	
Ki67 expression				
<30%	171 (61.7)	34 (19.9)	137 (80.1)	0.314
≥30%	106 (38.3)	16 (15.1)	90 (84.9)	
Carcinoma <i>in situ</i>				
No	253 (91.3)	47 (18.6)	206 (81.4)	0.459
Yes	24 (8.7)	3 (12.5)	21 (87.5)	
Recurrence				
No	192 (69.3)	40 (20.8)	152 (79.2)	0.07
Yes	85 (30.7)	10 (11.8)	75 (88.2)	
Metastasis				
No	236 (85.2)	45 (19.1)	191 (80.9)	0.291
Yes	41 (14.8)	5 (12.2)	36 (87.8)	
Death				
No	226 (81.6)	46 (20.4)	180 (79.6)	0.036
Yes	51 (18.4)	4 (7.8)	47 (92.2)	

BMI, body mass index.

8.13,  $P < 0.001$ ), Ki67 expression (OR=1.02, 95%CI: 1.01-1.04,  $P = 0.002$ ) and mortality (OR=4.90, 95%CI: 2.53-9.47,  $P < 0.001$ , **Table S1**). Multivariable analysis showed that CX3CL1 expression remained significantly associated with tumor size, tumor stage, Ki67 expression, disease recurrence, and mortality (all  $P < 0.05$ ).

CX3CL1 was an independent predictor of BCa recurrence in both a univariable analysis (OR=2.44, 95%CI: 1.24-2.83,  $P = 0.001$ ) and a multivariable analysis (OR=2.37, 95%CI: 1.18-2.75,  $P = 0.006$ ) after adjusting for multiple variables, as shown in **Table 2**. As shown in the cumulative hazard curve of disease recurrence (**Figure 1A**), the 6-year cumulative recurrence rate of the BCa CX3CL1-positive group was 62.9%, which was significantly higher than that of the BCa CX3CL1-negative group (31.8%,  $P = 0.006$ ). Similarly, the BCa-specific survival rate of CX3CL1-positive group was 23.2%, which was significantly lower than that of the negative expression group (41.1%,  $P < 0.0001$ , **Figure 1B**). Patients with high expression of

CX3CL1 had an ~2-fold increased risk of cancer-specific death based on univariable (HR=2.16, 95%CI: 1.59-2.93,  $P < 0.001$ ) and multivariable (HR=2.17, 95%CI: 1.08-2.39,  $P = 0.006$ ) analyses (**Table 3**). This indicated that CX3CL1 was also a significant and independent risk factor for disease-specific death. A total of 406 BCa samples were obtained from the TCGA database. Survival analysis suggested that the patients with high expression of CX3CL1 (FPKM $\geq 1.0$ ) were significantly associated with poor survival compared with the patients with low CX3CL1 expression (log-rank  $P$  value=0.039, **Figure 1C**). These results remained significant after adjusting for age at diagnosis and tumor stage *via* Cox regression (HR=1.76, 95%CI: 1.14-2.71,  $P = 0.011$ ).

To evaluate whether CX3CL1 expression is associated with disease aggressiveness with functional experiments, we evaluated the expression levels in 3 different types of cell lines, TCCSUP (grade IV cancer), 5637 (grade II cancer), and T24 cells (high-grade and invasive TCC). The mRNA expression of CX3CL1 in TCCSUP cells was significantly higher than that in 5637 and T24

**TABLE 2 |** Univariable and multivariable logistic regression analyses on tumor recurrence.

Predictors		Univariable analysis		Multivariable analysis	
		OR (95% CI)	P value	AOR (95% CI)	P value
Age	≤65 years	1.00 (Ref.)		/	/
	>65 years	1.11 (0.67-1.86)	0.679	/	/
Gender	Male	1.00 (Ref.)		/	/
	Female	1.35 (0.64-2.82)	0.431	/	/
BMI	<24	1.00 (Ref.)		1.00 (Ref.)	
	≥24	0.93 (0.86-1.00)	0.046	0.95 (0.88-1.03)	0.229
Smoking	No	1.00 (Ref.)		1.00 (Ref.)	
	Yes	1.80 (1.05-3.10)	0.033	1.98 (1.09-3.60)	0.025
Tumor size	<3cm	1.00 (Ref.)		1.00 (Ref.)	
	≥3cm	1.21 (1.03-1.42)	0.017	1.42 (1.05-1.91)	0.036
No. of tumor sites	Single	1.00 (Ref.)		/	/
	Multiple	1.45 (0.86-2.44)	0.162	/	/
Tumor stage	Ta-T1	1.00 (Ref.)		1.00 (Ref.)	
	T2-T4	2.41 (1.41-4.11)	0.001	1.83 (1.45-3.34)	0.027
Tumor grade	Low	1.00 (Ref.)		1.00 (Ref.)	
	High	2.30 (1.37-3.88)	0.002	1.82 (0.81-4.13)	0.011
Ki67 expression	<30%	1.00 (Ref.)		1.00 (Ref.)	
	≥30%	1.02 (1.01-1.04)	<0.001	1.02 (1.01-1.05)	0.002
Carcinoma <i>in situ</i>	No	1.00 (Ref.)		/	/
	Yes	0.92 (0.37-2.32)	0.866	/	/
CX3CL1	Negative	1.00 (Ref.)		1.00 (Ref.)	
	Positive	2.44 (1.24-2.83)	0.001	2.37 (1.18-2.75)	0.006

OR, odds ratio; 95% CI, 95% confidence interval; AOR, adjusted odds ratio; Ref, reference; BMI, body mass index.

cells ( $P < 0.001$ , **Figure 2A**). No significant difference in CX3CL1 expression was observed between the 5637 and T24 cell lines ( $P = 0.766$ ). Thus, TCCSUP cells were used for further evaluation. CX3CL1 was then knocked down by lentiviral infection. The knockdown efficiency was then verified by qPCR (efficiency reached 56.5%,  $P = 0.0032$ , **Figure 2B**). We then evaluated whether CX3CL1 expression is associated with cancer cell proliferation and invasiveness. MTT assay results showed that TCCSUP cell proliferation was significantly decreased in the KD group compared to the NC group (**Figure 2C**,  $P < 0.001$ ). Transwell assays suggested a significant decrease in cell invasion in the KD group compared with the NC group (**Figures 2D, E**,  $P < 0.001$ ).

To further evaluate whether CX3CL1 expression is associated with BCa aggressiveness *in vivo*, we established xenografts in the shoulder of immunocompromised nude mice. Subcutaneous tumor formation in the nude mice was verified by pathologists, and the success rate was 100% (**Figure 3**). From the 10th day to the 35th day, the tumor volume in the KD group was significantly smaller than that in the NC group (**Figure 3**, days 10-35,  $P < 0.05$ ). The tumor weight in the KD group was significantly lower than that in the NC group (**Figure 3**, day 35,  $P = 0.017$ ).

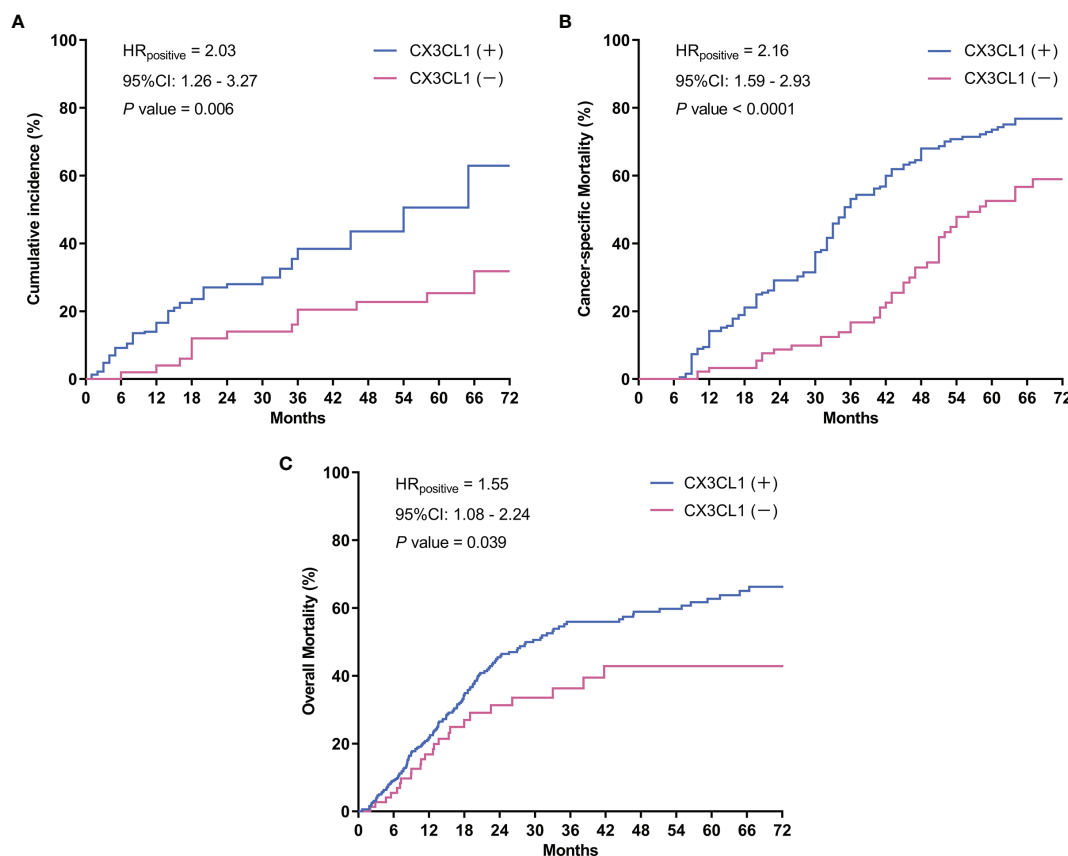
To investigate potential biological functions, we performed gene expression profiling using an mRNA microarray. Totals of 182 and 367 genes were found to be significantly upregulated and downregulated, respectively, in CX3CL1 KD TCCSUP cells. IPA-based disease and function analysis suggested that after CX3CL1 was inhibited, tumor cell proliferation ( $P = 1.62 \times 10^{-11}$ ,  $z\text{-score} = -2.32$ ), cell malignant transformation ( $P = 6.47 \times 10^{-13}$ ,  $z\text{-score} = -2.289$ ), tumor cell interphase ( $P = 1.57 \times 10^{-12}$ ,  $z\text{-score} = -2.078$ ) and differentiation of stem cells ( $P = 4.02 \times 10^{-5}$ ,  $z\text{-score} = -2.465$ ) were significantly inhibited, and tumor cell apoptosis ( $P = 2.35 \times 10^{-7}$ ,  $z\text{-score} = 2.236$ ) was significantly increased (**Figure 4A**). CDH1,

ETS1, RAF1, and EIF4E were genes with major effects in the enrichment analysis. In terms of classic pathway enrichment analysis based on IPA, the ERK/MAPK signaling pathway was significantly inhibited ( $P = 2.19 \times 10^{-4}$ ,  $z\text{-score} = -1.604$ ). ETS1, RAF1, and EIF4E were closely related to the ERK/MAPK pathway and significantly downregulated (**Figure 4A** and **Table S2**). A volcano plot and hierarchical clustering for significant difference analysis are shown in **Figure 4B**. Briefly, the tumor cell apoptosis ( $P = 2.35 \times 10^{-7}$ ,  $z\text{-score} = 2.236$ ) pathway was significantly activated, while tumor cell proliferation ( $P = 1.62 \times 10^{-11}$ ,  $z\text{-score} = -2.32$ ), tumor cell interphase ( $P = 1.57 \times 10^{-12}$ ,  $z\text{-score} = -2.078$ ), cell viability of tumor cell lines ( $P = 7.12 \times 10^{-14}$ ,  $z\text{-score} = -2.092$ ), differentiation of stem cells ( $P = 4.02 \times 10^{-5}$ ,  $z\text{-score} = -2.465$ ), cell malignant transformation ( $P = 6.47 \times 10^{-13}$ ,  $z\text{-score} = -2.289$ ), etc. were significantly inhibited.

To further confirm whether the genes in the ERK/MAPK pathway are influenced by the expression of CX3CL1, the protein expression of the enriched genes including RAF1, EIF4E, ETS1, and CDH1 was evaluated by Western blotting. The results suggested that RAF1, EIF4E, and ETS1 protein expression was significantly downregulated by 67.2%, 52.0%, and 43.6%, respectively (all  $P < 0.05$ ), while CDH1 protein expression was significantly upregulated by 74.6% ( $P < 0.05$ , **Figure 4C**).

## DISCUSSION

Tumor immunotherapy is considered as an important and hopeful curative therapy for cancers, such as PD-1/PD-L1. Immune surveillance and immune tolerance usually serve with a two-side effect throughout cancer development and progression. Scientists have found that chemokines family was the key factor in this process by affecting the immune system (3). Among different



**FIGURE 1** | Survival curves. **(A)** Cumulative tumor recurrence rate: the blue line represents the cumulative recurrence rate of 277 patients with bladder cancer with positive CX3CL1 expression; the red line represents the CX3CL1 expression-negative group. **(B)** Cancer-specific survival: the blue line represents the cancer-specific survival of the bladder cancer CX3CL1-positive group; the red line represents the CX3CL1 expression-negative group. **(C)** Overall mortality in the TCGA database: the blue line represents the overall mortality associated with 406 bladder cancer samples obtained from the TCGA database with positive CX3CL1 expression; the red line represents the CX3CL1 expression-negative group. HR, hazard ratio; 95%CI, 95% confidence interval.

chemokines, CX3CL1 is considered as one of the most important one, and may enhance local anti-tumor immunity by polarizing immunocompetent cells around the tumor microenvironment, and therefore inhibit tumor growth (4); on the other hand, tumors spontaneously secrete chemokines and receptors, which can promote the growth of cancer cells (5). We found that high expression of CX3CL1 was significantly associated with higher risk of disease recurrence and cancer-specific death; through *in vivo* and *in vitro* functional study, we further confirmed that high expression of CX3CL1 was associated with tumor cell proliferation and invasion. Enrichment analysis and functional study based on expression array data suggested that the regulation of tumor growth by CX3CL1 might due to the enhance of the ERK/MAPK signaling pathway.

CX3CL1 is a large cytokine (373 amino acids) and the only known member of the CX3C chemokine family. It is also known as fractalkine (in humans) and neurotactin (in mice) (6, 20). It has been proved correlating with cell migration and interaction in multiple diseases. Shulby *et al.* presented the first evidence of the expression of CX3CR1 (the specific receptor of fractalkine) in human prostate cancer cells and the expression of fractalkine in

human bone marrow endothelial cells. They also found that a neutralizing antibody against fractalkine might significantly reduce the adhesion of prostate cancer cells to human bone marrow endothelial cells, suggesting the important role of fractalkine in this process (13). By silencing CX3CR1, Jamieson *et al.* further defined the roles of this chemokine receptor in the adhesion of prostate cancer cells to bone marrow endothelial cells and prostate cancer cell migration towards media conditioned by different types of bone cells (21). Tardaguila *et al.* found that CX3CL1 expression was downregulated in HER2/neu tumors; however, paradoxically, adenovirus-mediated CX3CL1 expression in the tumor milieu enhanced mammary tumor numbers in a dose-dependent manner. Additionally, CX3CL1 triggered cell proliferation by inducing ErbB receptors through the proteolytic shedding of an ErbB ligand, and these findings support the conclusion that CX3CL1 acts as a positive modifier of breast cancer in concert with ErbB receptors (22). In another study on breast cancer, Tsang *et al.* reported that high CX3CL1 expression was detected in 33.3% of primary invasive cancers and that CX3CL1 expression correlated positively with increased tumor-infiltrating lymphocytes (TILs;  $P=0.005$ ). Furthermore, adverse features in breast cancers,

**TABLE 3 |** Univariable and multivariable-adjusted hazard ratios for cancer-specific mortality in Cox proportional hazards models.

Predictors		Univariable analysis		Multivariable analysis	
		HR (95% CI)	P value	AHR (95% CI)	P value
Age	≤65 years	1.00 (Ref.)	0.001	1.00 (Ref.)	0.031
	>65 years	2.99 (1.59-5.62)		2.22 (1.08-4.57)	
Gender	Male	1.00 (Ref.)	0.597	/	/
	Female	1.25 (0.55-2.86)		/	
BMI	<24	1.00 (Ref.)	0.223	/	/
	≥24	0.95 (0.87-1.03)		/	
Smoking	No	1.00 (Ref.)	0.207	/	/
	Yes	0.65 (0.34-1.27)		/	
Tumor size	<3cm	1.00 (Ref.)	0.012	1.00 (Ref.)	0.067
	≥3cm	2.19 (1.19-4.04)		2.26 (0.94-5.39)	
No. of tumor sites	Single	1.00 (Ref.)	0.040	1.00 (Ref.)	0.131
	Multiple	1.88 (1.03-3.41)		1.67 (1.29-3.51)	
Tumor stage	Ta-T1	1.00 (Ref.)	0.001	1.00 (Ref.)	0.178
	T2-T4	2.68 (1.49-4.82)		1.77 (1.30-3.96)	
Tumor grade	Low	1.00 (Ref.)	<0.001	1.00 (Ref.)	0.037
	High	3.42 (1.85-6.33)		2.59 (1.06-6.31)	
Ki67 expression	<30%	1.00 (Ref.)	0.001	1.00 (Ref.)	0.002
	≥30%	1.02 (1.01-1.03)		1.03 (1.01-1.04)	
Carcinoma <i>in situ</i>	No	1.00 (Ref.)	0.116	/	/
	Yes	0.81 (0.77-1.34)		/	
CX3CL1	Negative	1.00 (Ref.)	<0.001	1.00 (Ref.)	0.006
	Positive	2.16 (1.59-2.93)		2.17 (1.08-2.39)	

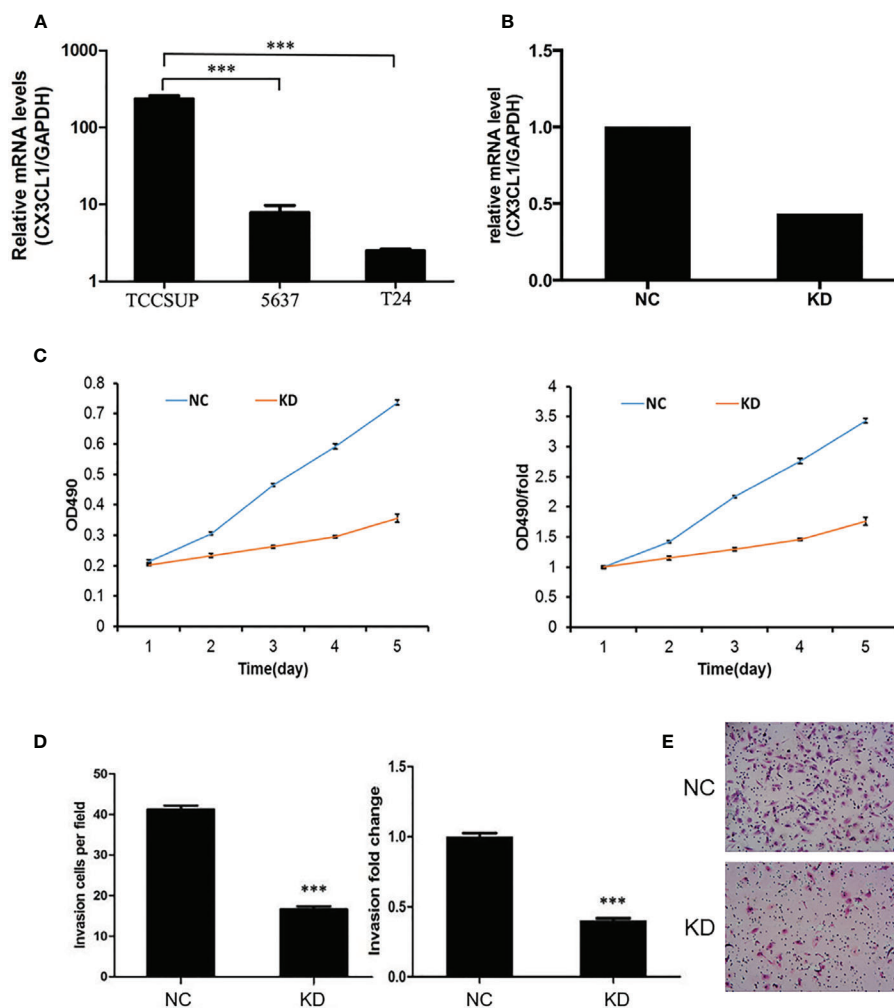
HR, hazard ratio; 95% CI, 95% confidence interval; AHR, adjusted hazard ratio; Ref, reference; BMI, body mass index.

including lymph node involvement,  $\alpha$ - $\beta$  crystalline expression and high Ki67, were positively correlated with the expression of CX3CL1. Notably, in the same study, the authors showed that patients with high levels of CX3CL1 had poorer overall survival. The findings indicated an oncogenic role for CX3CL1, despite its previously suggested role in enhancing antitumor immunity, and highlighted the complicated functions of CX3CL1 in breast carcinogenesis (23). However, a study on the prognosis of patients with liver cancer showed that patients with high expression of CX3CL1 and its receptor had a better prognosis than those with low expression, with a significantly lower local recurrence rate and distant metastasis rate and longer disease-free survival and overall survival (15). These results indicate that CX3CL1 has complex functions in tumors. At present, there are few in-depth studies on the role of CX3CL1 in BCa, and the mechanism of action of the chemokine CX3CL1 is still unclear. We conducted a series of experimental studies to investigate the role of CX3CL1 in BCa.

Previous studies have shown that CX3CL1 may promote or suppress tumors in different tissues. Our research confirms the value and function of CX3CL1 in BCa. Compared with previous studies, our study identified possible signaling pathways and downstream genes of CX3CL1 in BCa through GeneChip PrimeView Human technology and explored the possible mechanisms in more depth. Through bioinformatic data analysis and statistics, the significant differences were further verified. We found that after CX3CL1 was inhibited, tumor cell proliferation, cell malignant transformation, tumor cell interphase, and differentiation of stem cells were significantly inhibited, and tumor cell apoptosis was significantly increased. These results suggest that CX3CL1 may play a role in promoting the proliferation of BCa cells. In this experiment, the ERK/MAPK pathway was significantly inhibited, and the

downstream genes ETS1 ( $\downarrow$ ), RAF1 ( $\downarrow$ ), and EIF4E ( $\downarrow$ ), which are closely related to this pathway, were significantly downregulated. In addition, the downstream gene CDH1 ( $\uparrow$ ) was significantly upregulated. Further detection of the expression of the proteins encoded by the above genes was also confirmed. The RNA-Seq results, and protein expression results are consistent with a previous clinical prognostic analysis and *in vivo* and *in vitro* experimental results, which together indicate the tumor-promotive effect of CX3CL1 in BCa.

This study suggests that CX3CL1 might promote the cancer progression through the ERK/MAPK signaling pathway. Mitogen-activated protein kinase (MAPK) is a threonine/serine protein kinase that is widely present in cells. It is closely related to cell proliferation, apoptosis, tumor proliferation, invasion and metastasis. The MAPK pathway is one of the most studied signaling pathways in recent years (24, 25). Among MAPK family members, extracellular signal-regulated kinase (ERK) was the first discovered and is the most widely studied member. An abnormal ERK/MAPK pathway can lead to abnormal cell proliferation, which is related to tumorigenesis and can further affect the biological behavior of tumors (26). Current studies indicate that the ERK/MAPK signaling pathway plays an important role in tumor occurrence and development by promoting cancer cell proliferation, inhibiting tumor apoptosis, and inducing tumor neovascularization, invasion, and metastasis (27–30). Studies have also pointed out that the ERK/MAPK signaling pathway is also related to tumor resistance (31, 32). To date, many studies have found that the ERK/MAPK signaling pathway is related to many tumors, such as colon cancer, gastric cancer, liver cancer, melanoma, and BCa (33–38). According to previous research reports, among the four significantly regulated downstream genes found in this



**FIGURE 2 | (A)** The mRNA expression of CX3CL1 in TCCSUP, 5637 and T24 cells. The mRNA expression of CX3CL1 in TCCSUP cells was significantly higher than that in 5637 and T24 cells. **(B)** Difference in CX3CL1 mRNA levels between NC and KD TCCSUP cells. The CX3CL1 gene of TCCSUP cells in KD group was knocked down, and the knockdown rate was verified by qPCR. **(C)** Difference in TCCSUP cell proliferation between NC and KD groups determined by the MTT method. The results showed that cell proliferation in the KD group (Orange line) was significantly decreased. **(D, E)** Cell invasion of NC and KD cells determined by Transwell assays. The results showed that after CX3CL1 was knocked down, cell invasion in the KD group was significantly reduced. \*\*\* $P < 0.001$ .

experiment, CDH1 was considered to be an anti-oncogene, and ETS1, RAF1 and EIF4E play roles in promoting cancer (39–48). Our research results are consistent with those in previous research reports.

To the best of our knowledge, this was the first study investigating the association between CX3CL1 expression and BCa. The results showed that CX3CL1 overexpression was independently and significantly associated with poor prognosis (increased risk of recurrence and cancer-specific death rate). In addition, we were able to confirm our results in TCGA dataset. This finding may be applied to clinical practice as a valuable prognostic biomarker *via* additional translational research approaches.

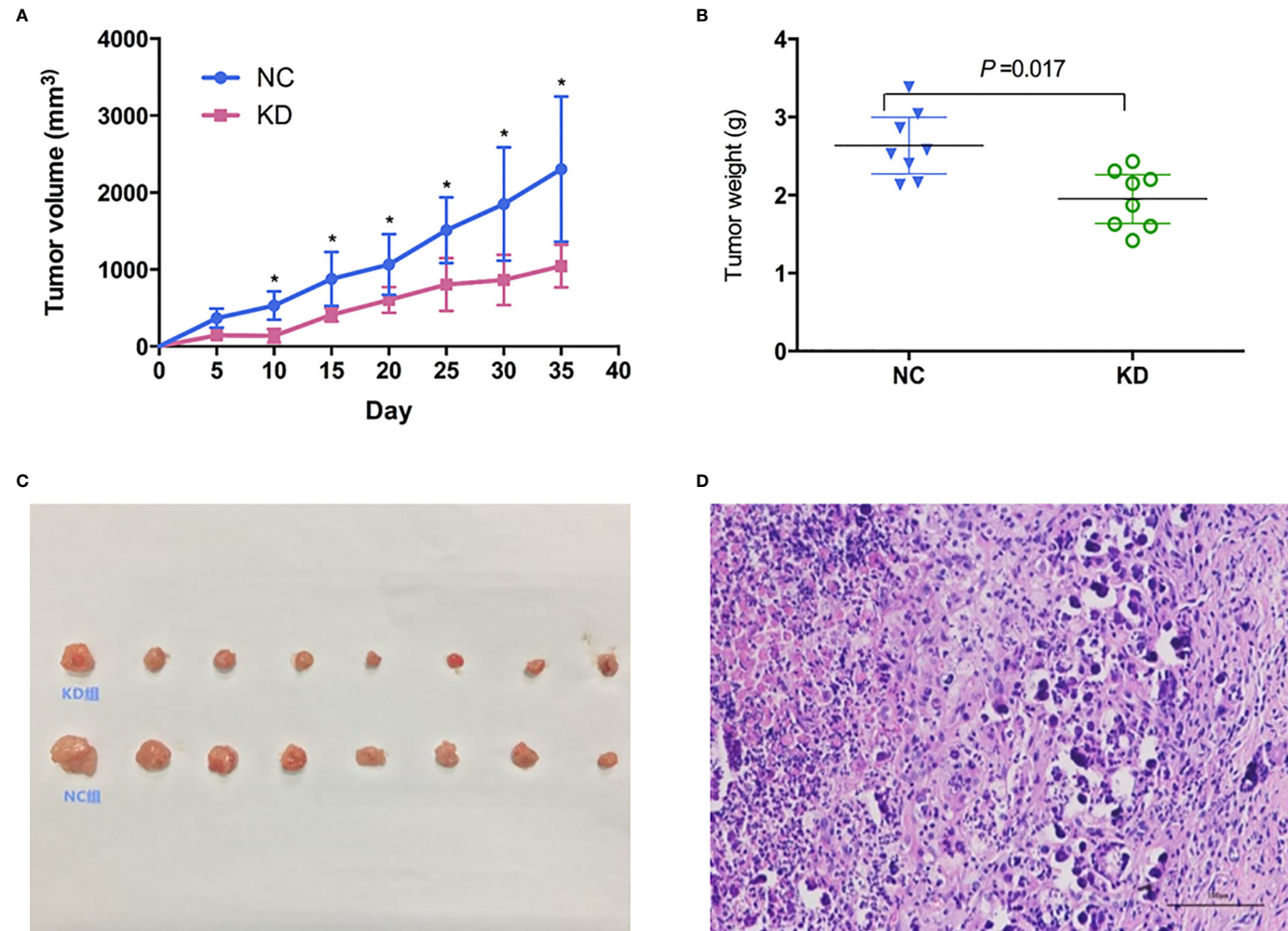
Our study has carried out a series of elaboration on the role of CX3CL1 in BCa. However, several limitations should be noted in the present study. First, the sample size was relatively small. However, we were able to validate our results in TCGA database.

The results should be investigated in a large-scale study in the future. Second, the subcutaneous xenograft animal model could not fully simulate the tumor microenvironment. Although orthotopic bladder cancer model would be a better option, the current method was the most common used model in BCa research. Last but not least, only one cell line was chosen in the present study. However, the main objective of this study was to evaluate the association of CX3CL1 expression and BCa prognosis.

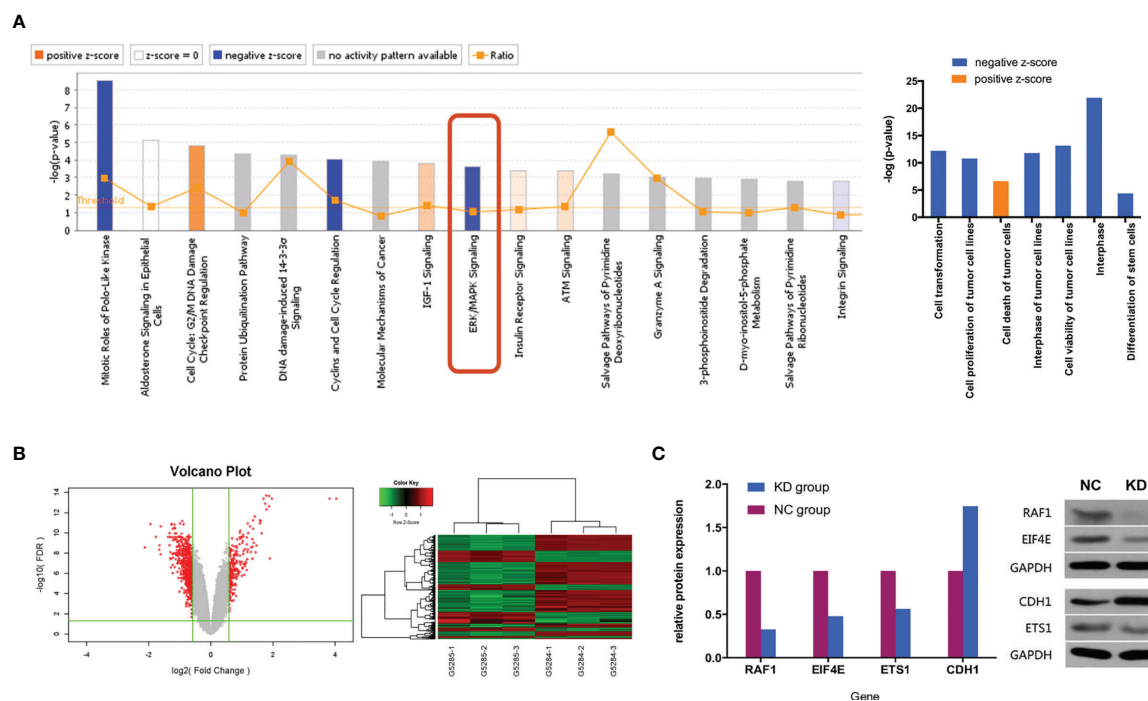
## CONCLUSION

CX3CL1 was a significant and independent predictor for bladder cancer prognosis. It could promote the proliferation and





**FIGURE 3** | Effects of CX3CL1 knockdown on (A) tumor volume and (B) weight changes in a nude mouse subcutaneous tumor model of bladder cancer. The results showed that compared with the control group, the volume and weight of subcutaneous tumors in the KD group decreased significantly. (C) Subcutaneous tumor in KD (up) and NC (down) group. (D) Hematoxylin & Eosin staining of the tumor cells. KD, knockdown; NC, negative control.



**FIGURE 4 | (A)** Classic pathway enrichment analysis and disease and function analysis based on IPA by GeneChip PrimeView. Our experiments suggested that when CX3CL1 is inhibited, the ERK/MAPK signaling pathway is inhibited, the function of promoting tumor growth is inhibited, and tumor death is enhanced. **(B)** Chip volcano plot (left) and hierarchical clustering (right): in the volcano plot, the abscissa represents the multiple of the gene expression difference (transformed by log<sub>2</sub>), the ordinate represents the significant FDR of the difference (transformed by log<sub>10</sub>), the red dots were selected with |Fold Change| > 1.5 and FDR < 0.05 as the screening criteria for genes that were significantly differentially expressed, and the gray dots are other genes that were not significantly differentially expressed; in the hierarchical clustering data, red means that gene expression was relatively upregulated, green means that gene expression was relatively downregulated, black means that the degree of gene expression did not change significantly, and gray means that the signal intensity of the gene was not detected. **(C)** Relative expression of downstream genes and proteins in the ERK/MAPK signaling pathway between NC and KD cells (Western blot). NC, normal control group; KD, CX3CL1 gene knockdown group.

invasion of bladder cancer cells *via* inhibiting the ERK/MAPK signaling pathway.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board at Huashan Hospital. The written informed consent was obtained from all patients for institutional biobank. The animal study was reviewed and approved by The Institutional Review Board at Huashan Hospital.

## AUTHOR CONTRIBUTIONS

KX, RN, and NZ conceived and designed the study. GJ, HW, YW, WD, and QZ acquired the data. GJ, HW, and DH analyzed and

interpreted the data. GJ, HW, and DH drafted the manuscript. KX, RN, NZ, and QD contributed to the critical revision of the manuscript. KX and RN supervised the study. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by the National Natural Science Foundation of China (Grant No. 81472379 and 81402339).

## ACKNOWLEDGMENTS

We thank all of the subjects included in this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.752860/full#supplementary-material>

## REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
- Pardoll D. Does the Immune System See Tumors as Foreign or Self? *Annu Rev Immunol* (2003) 21(1):807–39. doi: 10.1146/annurev.immunol.21.120601.141135
- Salcedo R, Ponce ML, Young HA, Wasserman K, Ward JM, Kleinman HK, et al. Human Endothelial Cells Express CCR2 and Respond to MCP-1: Direct Role of MCP-1 in Angiogenesis and Tumor Progression. *Blood* (2000) 96(1):34–40. doi: 10.1182/blood.V96.1.34
- Schneider GP, Salcedo R, Welniak LA, Howard OZ, Murphy WJ. The Diverse Role of Chemokines in Tumor Progression: Prospects for Intervention. *Int J Mol Med* (2001) 8(3):235–44. doi: 10.3892/ijmm.8.3.235
- Bazan JF, Bacon KB, Hardiman G, Wang W, Soo K, Rossi D, et al. A New Class of Membrane-Bound Chemokine With a CX3 C Motif. *Nature* (1997) 385(6617):640–4. doi: 10.1038/385640a0
- Fong AM, Erickson HP, Zachariah JP, Poon S, Schamberg NJ, Imai T, et al. Ultrastructure and Function of the Fractalkine Mucin Domain in CX3(C) Chemokine Domain Presentation. *J Biol Chem* (2000) 275(6):3781–6. doi: 10.1074/jbc.275.6.3781
- Haskell CA, Cleary MD, Charo IF. Molecular Uncoupling of Fractalkine-Mediated Cell Adhesion and Signal Transduction. Rapid Flow Arrest of CX3CR1-Expressing Cells is Independent of G-Protein Activation. *J Biol Chem* (1999) 274(15):10053–8. doi: 10.1074/jbc.274.15.10053
- Andre F, Cabioglu N, Assi H, Sabourin JC, Delaloge S, Sahin A, et al. Expression of Chemokine Receptors Predicts the Site of Metastatic Relapse in Patients With Axillary Node Positive Primary Breast Cancer. *Ann Oncol* (2006) 17(6):945–51. doi: 10.1093/annonc/mdl053
- Celesti G, Di Caro G, Bianchi P, Grizzi F, Marchesi F, Basso G, et al. Early Expression of the Fractalkine Receptor CX3CR1 in Pancreatic Carcinogenesis. *Br J Cancer* (2013) 109(9):2424–33. doi: 10.1038/bjc.2013.565
- Kim M, Rooper L, Xie J, Kajdacsy-Balla AA, Barbolina MV. Fractalkine Receptor CX3CR1 is Expressed in Epithelial Ovarian Carcinoma Cells and Required for Motility and Adhesion to Peritoneal Mesothelial Cells. *Mol Cancer Res* (2012) 10(1):11–24. doi: 10.1158/1541-7786.MCR-11-0256
- Yao X, Qi L, Chen X, Du J, Zhang Z, Liu S. Expression of CX3CR1 Associates With Cellular Migration, Metastasis, and Prognosis in Human Clear Cell Renal Cell Carcinoma. *Urol Oncol* (2014) 32(2):162–70. doi: 10.1016/j.urolonc.2012.12.006
- Shulby SA, Dolloff NG, Stearns ME, Meucci O, Fatatis A. CX3CR1-Fractalkine Expression Regulates Cellular Mechanisms Involved in Adhesion, Migration, and Survival of Human Prostate Cancer Cells. *Cancer Res* (2004) 64(14):4693–8. doi: 10.1158/0008-5472.CAN-03-3437
- Ohta M, Tanaka F, Yamaguchi H, Sadanaga N, Inoue H, Mori M. The High Expression of Fractalkine Results in a Better Prognosis for Colorectal Cancer Patients. *Int J Oncol* (2005) 26(1):41–7. doi: 10.3892/ijo.26.1.41
- Matsubara T, Ono T, Yamanoi A, Tachibana M, Nagasue N. Fractalkine-CX3CR1 Axis Regulates Tumor Cell Cycle and Deteriorates Prognosis After Radical Resection for Hepatocellular Carcinoma. *J Surg Oncol* (2007) 95(3):241–9. doi: 10.1002/jso.20642
- Park MH, Lee JS, Yoon JH. High Expression of CX3CL1 by Tumor Cells Correlates With a Good Prognosis and Increased Tumor-Infiltrating CD8+ T Cells, Natural Killer Cells, and Dendritic Cells in Breast Carcinoma. *J Surg Oncol* (2012) 106(4):386–92. doi: 10.1002/jso.23095
- Tang J, Chen Y, Cui R, Li D, Xiao L, Lin P, et al. Upregulation of Fractalkine Contributes to the Proliferative Response of Prostate Cancer Cells to Hypoxia via Promoting the G1/S Phase Transition. *Mol Med Rep* (2015) 12(6):7907–14. doi: 10.3892/mmr.2015.4438
- Eble JN, Sauter G, Epstein JI, Sesterhenn IA. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs: World Health Organization Classification of Tumours. *Int Agency Res Cancer* (2004) 7:89–154.
- Edge SB, Compton CC. The American Joint Committee on Cancer: The 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Ann Surg Oncol* (2010) 17(6):1471–4. doi: 10.1245/s10434-010-0985-4
- Pan Y, Lloyd C, Zhou H, Dolich S, Deeds J, Gonzalo J-A, et al. Neurotactin, a Membrane-Anchored Chemokine Upregulated in Brain Inflammation. *Nature* (1997) 387(6633):611–7. doi: 10.1038/42491
- Jamieson WL, Shimizu S, D'Ambrosio JA, Meucci O, Fatatis A. CX3CR1 is Expressed by Prostate Epithelial Cells and Androgens Regulate the Levels of CX3CL1/fractalkine in the Bone Marrow: Potential Role in Prostate Cancer Bone Tropism. *Cancer Res* (2008) 68(6):1715–22. doi: 10.1158/0008-5472.CAN-07-1315
- Tardaguila M, Mira E, Garcia-Cabezas MA, Feijoo AM, Quintela-Fandino M, Azcoitia I, et al. CX3CL1 Promotes Breast Cancer via Transactivation of the EGF Pathway. *Cancer Res* (2013) 73(14):4461–73. doi: 10.1158/0008-5472.CAN-12-3828
- Tsang JYS, Ni Y-B, Chan S-K, Shao M-M, Kwok Y-K, Chan K-W, et al. CX3CL1 Expression is Associated With Poor Outcome in Breast Cancer Patients. *Breast Cancer Res Treat* (2013) 140(3):495–504. doi: 10.1007/s10549-013-2653-4
- Shen H, Xu W, Luo W, Zhou L, Yong W, Chen F, et al. Upregulation of Mdr1 Gene is Related to Activation of the MAPK/ERK Signal Transduction Pathway and YB-1 Nuclear Translocation in B-Cell Lymphoma. *Exp Hematol* (2011) 39(5):558–69. doi: 10.1016/j.exphem.2011.01.013
- Parakh S, Murphy C, Lau D, Cebon JS, Andrews MC. Response to MAPK Pathway Inhibitors in BRAF V600M-Mutated Metastatic Melanoma. *J Clin Pharm Ther* (2014) 40(1):121–3. doi: 10.1111/jcpt.12229
- Aebersold DM, Burri P, Beer KT, Laissue J, Djonov V, Greiner RH, et al. Expression of Hypoxia-Inducible Factor-1 $\alpha$ : A Novel Predictive and Prognostic Parameter in the Radiotherapy of Oropharyngeal Cancer. *Cancer Res* (2001) 61(7):2911–6.
- Wang YF, Jiang CC, Kiejda KA, Gillespie S, Zhang XD, Hersey P. Apoptosis Induction in Human Melanoma Cells by Inhibition of MEK is Caspase-Independent and Mediated by the Bcl-2 Family Members PUMA, Bim, and Mcl-1. *Clin Cancer Res* (2007) 13(16):4934–42. doi: 10.1158/1078-0432.CCR-07-0665
- Chen Y-J, Tsai Y-M, Kuo C-D, Ku K-L, Shie H-S, Liao H-F. Norcantharidin is a Small-Molecule Synthetic Compound With Anti-Angiogenesis Effect. *Life Sci* (2009) 85(17-18):642–51. doi: 10.1016/j.lfs.2009.09.003
- Yue P, Gao Z-H, Xue X, Cui S-X, Zhao C-R, Yuan YI, et al. Des- $\gamma$ -Carboxyl Prothrombin Induces Matrix Metalloproteinase Activity in Hepatocellular Carcinoma Cells by Involving the ERK1/2 MAPK Signalling Pathway. *Eur J Cancer* (2011) 47(7):1115–24. doi: 10.1016/j.ejca.2011.01.017
- Corona G, Deiana M, Incani A, Vauzour D, Dessi MA, Spencer JPE. Hydroxytyrosol Inhibits the Proliferation of Human Colon Adenocarcinoma Cells Through Inhibition of ERK1/2 and Cyclin D1. *Mol Nutr Food Res* (2009) 53(7):897–903. doi: 10.1002/mnfr.200800269
- Ochi N, Takigawa N, Harada D, Yasugi M, Ichihara E, Hotta K, et al. Src Mediates ERK Reactivation in Gefitinib Resistance in non-Small Cell Lung Cancer. *Exp Cell Res* (2014) 322(1):168–77. doi: 10.1016/j.yexcr.2014.01.007
- Fan D-P, Zhang Y-M, Hu X-C, Li J-J, Zhang W. Activation of AKT/ERK Confers non-Small Cell Lung Cancer Cells Resistance to Vinorelbine. *Int J Clin Exp Pathol* (2014) 7(1):134.
- McCubrey JA, Steelman LS, Abrams SL, Chappell WH, Russo S, Ove R, et al. Emerging MEK Inhibitors. *Expert Opin emerging Drugs* (2010) 15(2):203–23. doi: 10.1517/14728210903282760
- Simeonova PP, Wang S, Hulderman T, Luster MI. C-Src-Dependent Activation of the Epidermal Growth Factor Receptor and Mitogen-Activated Protein Kinase Pathway by Arsenic Role in Carcinogenesis. *J Biol Chem* (2002) 277(4):2945–50. doi: 10.1074/jbc.M109136200
- Drobna Z, Jaspers I, Thomas DJ, Styblo M. Differential Activation of AP-1 in Human Bladder Epithelial Cells by Inorganic and Methylated Arsenicals. *FASEB J* (2003) 17(1):67–9. doi: 10.1096/fj.02-0287fj
- Fujimori Y, Inokuchi M, Takagi Y, Kato K, Kojima K, Sugihara K. Prognostic Value of RKIP and P-ERK in Gastric Cancer. *J Exp Clin Cancer Res* (2012) 31(1):30. doi: 10.1186/1756-9966-31-30
- Chen C-C, Chen C-Y, Wang S-H, Yeh C-T, Su S-C, Ueng S-H, et al. Melatonin Sensitizes Hepatocellular Carcinoma Cells to Chemotherapy Through Long non-Coding RNA RAD51-AS1-Mediated Suppression of DNA Repair. *Cancers (Basel)* (2018) 10(9):320. doi: 10.3390/cancers10090320
- Gupta A, Hossain MM, Miller N, Kerin M, Callagy G, Gupta S. NCOA3 Coactivator is a Transcriptional Target of XBP1 and Regulates PERK-Eif2 $\alpha$ –



- ATF4 Signalling in Breast Cancer. *Oncogene* (2016) 35(45):5860–71. doi: 10.1038/onc.2016.121
39. Girolodi LA, Schalken JA. Decreased Expression of the Intercellular Adhesion Molecule E-Cadherin in Prostate Cancer: Biological Significance and Clinical Implications. *Cancer Metastasis Rev* (1993) 12(1):29–37. doi: 10.1007/BF00689788
  40. Girolodi LA, Bringuier PP, Schalken JA. Defective E-Cadherin Function in Urological Cancers: Clinical Implications and Molecular Mechanisms. *Invasion Metastasis* (1994) 14(1-6):71–81.
  41. Nigam AK, Savage FJ, Boulos PB, Stamp GW, Liu D, Pignatelli M. Loss of Cell-Cell and Cell-Matrix Adhesion Molecules in Colorectal Cancer. *Br J Cancer* (1993) 68(3):507–14. doi: 10.1038/bjc.1993.377
  42. Chen M, Cooper JA. Ser-3 is Important for Regulating Mos Interaction With and Stimulation of Mitogen-Activated Protein Kinase Kinase. *Mol Cell Biol* (1995) 15(9):4727–34. doi: 10.1128/MCB.15.9.4727
  43. Dent P, Reardon DB, Morrison DK, Sturgill TW. Regulation of Raf-1 and Raf-1 Mutants by Ras-Dependent and Ras-Independent Mechanisms *In Vitro*. *Mol Cell Biol* (1995) 15(8):4125–35. doi: 10.1128/MCB.15.8.4125
  44. von der Haar T, Gross JD, Wagner G, McCarthy JEG. The mRNA Cap-Binding Protein Eif4e in Post-Transcriptional Gene Expression. *Nat Struct Mol Biol* (2004) 11(6):503–11. doi: 10.1038/nsmb779
  45. De Benedetti A, Graff JR. eIF-4E Expression and its Role in Malignancies and Metastases. *Oncogene* (2004) 23(18):3189–99. doi: 10.1038/sj.onc.1207545
  46. Oettgen P. The Role of Ets Factors in Tumor Angiogenesis. *J Oncol* (2010) 2010:767384. doi: 10.1155/2010/767384
  47. Ahn KS, Bae E, Han D, Choi HY. 211 Increasing Expression Level of ETS-1 via HIF-1 $\alpha$  Induces the Angiogenesis and Invasiveness Human Bladder Cancer: It may be One of Useful Molecular Markers to Determine the Progression of Human Bladder Cancer. *Eur Urol Suppl* (2004) 3(2):55. doi: 10.1016/S1569-9056(04)90212-6
  48. Oikawa M, Abe M, Kurosawa H, Hida W, Shirato K, Sato Y. Hypoxia Induces Transcription Factor ETS-1 via the Activity of Hypoxia-Inducible Factor-1. *Biochem Biophys Res Commun* (2001) 289(1):39–43. doi: 10.1006/bbrc.2001.5927

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer HW declared a shared affiliation, with no collaboration, with several of the authors HW, YW, WD, QZ, QD, KX to the handling editor at the time of review.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Jiang, Wang, Huang, Wu, Ding, Zhou, Ding, Zhang, Na and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Germline Variant at 8q24 Contributes to the Serum p2PSA Level in a Chinese Prostate Biopsy Cohort

Xiaoling Lin<sup>1,2†</sup>, Yishuo Wu<sup>1,2†</sup>, Fang Liu<sup>1,2</sup>, Rong Na<sup>3</sup>, Da Huang<sup>3</sup>, Danfeng Xu<sup>3</sup>, Jian Gong<sup>1,2</sup>, Yao Zhu<sup>4</sup>, Bo Dai<sup>4</sup>, Dingwei Ye<sup>4</sup>, Hongjie Yu<sup>5</sup>, Haowen Jiang<sup>1,2</sup>, Zujun Fang<sup>1,2</sup>, Jie Zheng<sup>1,2\*</sup> and Qiang Ding<sup>1,2\*</sup>

## OPEN ACCESS

### Edited by:

Benyi Li,  
University of Kansas Medical Center,  
United States

### Reviewed by:

Xiaobo Yang,  
Guangxi Medical University, China  
Wen-Hui Su,  
Chang Gung University, Taiwan

### \*Correspondence:

Qiang Ding  
qiangd\_urology@126.com  
Jie Zheng  
atlantica@163.com

<sup>†</sup>These authors have contributed  
equally to this study

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 05 August 2021

**Accepted:** 27 September 2021

**Published:** 19 October 2021

### Citation:

Lin X, Wu Y, Liu F, Na R,  
Huang D, Xu D, Gong J, Zhu Y,  
Dai B, Ye D, Yu H, Jiang H,  
Fang Z, Zheng J and Ding Q  
(2021) A Germline Variant at  
8q24 Contributes to the Serum  
p2PSA Level in a Chinese  
Prostate Biopsy Cohort.  
Front. Oncol. 11:753920.  
doi: 10.3389/fonc.2021.753920

<sup>1</sup> Department of Urology, Huashan Hospital, Fudan University, Shanghai, China, <sup>2</sup> Fudan Institute of Urology, Huashan Hospital, Fudan University, Shanghai, China, <sup>3</sup> Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, <sup>4</sup> Department of Urology, Shanghai Cancer Center, Fudan University, Shanghai, China, <sup>5</sup> State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

**Introduction:** The clinical performance of [-2]proPSA (p2PSA) and its derivatives in predicting the presence and aggressiveness of prostate cancer (PCa) has been well evaluated in prostate biopsy patients. However, no study has been performed to evaluate the common genetic determinants that affect serum level of p2PSA.

**Materials and Methods:** Here, we performed a two-stage genome-wide association study (GWAS) on the p2PSA level in Chinese men who underwent a transperineal ultrasound-guided prostate biopsy at Huashan Hospital, Shanghai Cancer Center, and Ruijin Hospital in Shanghai, China. Germline variants significantly associated with the p2PSA level in the first stage ( $n = 886$ ) were replicated in the second stage ( $n = 1,128$ ). Multivariate linear regression was used to assess the independent contribution of confirmed single nucleotide polymorphisms (SNPs) and known covariates, such as age, to the level of p2PSA.

**Results:** A novel non-synonymous SNP, rs72725879, in region 8q24.21 of the *PRNCR1* gene was significantly associated with the serum level of p2PSA in this two-stage GWAS ( $p = 2.28 \times 10^{-9}$ ). Participants with homozygous “T” alleles at rs72725879 had higher p2PSA levels compared to allele “C” carriers. This variant was also nominally associated with PCa risk ( $p$ -combined =  $3.44 \times 10^{-18}$ ). The association with serum level of p2PSA was still significant after adjusting for PCa risk and age ( $p = 0.017$ ).

**Conclusions:** Our study shows that the genetic variants in the 8q24.21 region are associated with the serum level of p2PSA in a large-scale Chinese population. By taking inherited variations between individuals into account, the findings of these genetic variants may help improve the performance of p2PSA in predicting prostate cancer.

**Keywords:** genome-wide association study, p2PSA, polymorphism, prostate cancer, Chinese

## INTRODUCTION

Prostate cancer (PCa) is one of the most common tumors in men and one of the leading causes of cancer-related death, conferring 1,111,700 new cases and 307,500 deaths annually worldwide (1). Although the incidence of PCa in Chinese is lower than that in Caucasian and African populations, it has been rising progressively in recent decades, along with increasing mortality (2).

Prostate-specific antigen (PSA) screening is the most widely used biomarker for the early detection and surveillance of PCa. However, PSA can also be affected by prostatitis, benign prostate hyperplasia, age, ethnicity, and genetic factors, which means it is organ-specific rather than cancer-specific. Therefore, its low specificity in clinical applications leads to quite a number of unnecessary biopsies and overdiagnosis of indolent cancers (3). In terms of genetic influence on the serum PSA level, it is estimated that 40% of the variations between individuals can be explained by inherited factors (4). In previous genome-wide association studies (GWAS), multiple inherited variants had been demonstrated to influence the serum levels of PSA in European and Asian populations (5–8).

Regarding the issues of unnecessary biopsies and overdiagnosis caused by PSA screening, new biomarkers with higher specificities and better abilities to discriminate PCa and aggressive PCa are needed. A relatively new biomarker  $[-2]_{\text{pro}}\text{PSA}$  (p2PSA), a predominant precursor form of PSA, was found elevated in almost all of the peripheral zone cancers, but was largely undetectable in the transition zone (9). In previous studies, p2PSA and its derivative prostate health index (phi)  $[(\text{p2PSA}/\text{free PSA}) \times \sqrt{\text{tPSA}}]$  have been proven to have a better discriminating ability in predicting PCa in Caucasians (10–14). Later on, the clinical utilities of p2PSA and phi are also implicated in the Chinese (15–17). However, whether the serum p2PSA level is affected by genetic variance between different individuals remains unknown.

Discovery of novel genetic variants that influence the serum level of p2PSA may improve our understanding of the molecular mechanisms and clinical utility of p2PSA test. Therefore, to identify the genetic variants that influence the serum p2PSA level, we carried out a two-stage GWAS among Chinese men who underwent prostate biopsy.

## MATERIAL AND METHODS

### Study Cohorts and Design

Our study included two prostate biopsy cohorts from three medical centers, which were genotyped with the same genotyping array platform and denoted as stages 1 and 2.

Stage 1 consisted of 886 subjects from Huashan Hospital, Fudan University, and Shanghai Cancer Center, Fudan University, between 2010 and 2014. Stage 2 included 1,128 subjects from Huashan Hospital, Fudan University, and Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, between 2015 and 2018. All the patients ( $n = 2,014$ ) underwent initial prostate biopsies at the above-mentioned medical centers from 2010 to 2018.

The indications for prostate biopsy were the same across the three centers: 1) PSA  $>10.0$  ng/ml; 2) PSA  $>4.0$  ng/ml with a confirmation within 3 months; 3) PSA level ranging from 4.0 to 10.0 ng/ml, with suspicious %fPSA (free PSA divided by PSA)  $<0.16$ ; and 4) abnormal findings from digital rectal examination (DRE), ultrasound, or magnetic resonance imaging (MRI) with any level of PSA.

Blood specimens were obtained before biopsies and serum samples were extracted. Serum total PSA (tPSA), free PSA (fPSA), and p2PSA were measured with a Beckman Coulter DxI 800 Immunoassay System (Beckman Coulter, Brea, CA, USA). All assays and quality control (QC) were performed according to the manufacturer's instructions and standard QC protocols. Specifically, Access Hybritech Calibrators S0, S1–S6, i.e., blank and low to high concentrations, were run as internal known standards before each batch of the measurements. The measurements of the internal known control materials, Access Hybritech QC 1, 2, and 3, were below two standard deviations for each batch.

All epidemiological and clinical pieces of information were collected before biopsy. The patients would be excluded if pathologically diagnosed biopsy outcome was missing or if the tPSA, fPSA, and p2PSA were unable to be tested for bad quality. This study was approved by the institutional review board of each medical center, and written informed consent to participate in the present study was obtained from all participants.

### SNP Genotyping, Quality Control, Imputation, and GWAS Analysis

All DNA samples from stages 1 and 2 were extracted from blood samples and genotyped with the same genotyping array platform, Illumina Asian Screening Array (ASA) Beadchip, which included 659,184 single nucleotide polymorphisms (SNPs).

Genotyping QC was conducted together with data from stages 1 and 2. We used the following standard QC procedure to select qualified samples and SNPs for imputation analysis. Samples were excluded if they: i) had a genotyping call rate of  $<95\%$ ; ii) were duplicates or showed familial relationships [identity by state (IBS)  $>0.99$ ]; and iii) had ambiguous gender. SNPs were removed if they had: i) a genotyping rate of  $<95\%$  ( $n = 9,650$ ); ii) a minor allele frequency of  $<0.01$  ( $n = 152,901$ ); and iii) a  $p$ -value  $<10^{-3}$  with the Hardy–Weinberg equilibrium test in patients with negative biopsy results ( $n = 2,721$ ). After QC analysis, a total of 2,014 (886 in stage 1 and 1,128 in stage 2) samples with 493,912 genotyped SNPs were retained for imputation analysis.

Imputation was performed with the IMPUTE computer program (18) using the 1000 Genomes Project Han Chinese in Beijing (CHB) population as the reference. A total of 17,098,949 SNPs with imputation information score  $>0.90$  were included in the analysis.

### Genome-Wide Association Analysis and Replication

GWAS analysis was conducted in stage 1, which included 886 samples with 17,098,949 SNPs. Using the same sample and SNP

QC criteria above, 886 samples with 4,552,207 SNPs were left in the association analysis. PCAs were estimated using EIGENSTRAT. Beta values and *p*-values were estimated using quantitative linear regression for each SNP, adjusting for age and the first two PCAs.

We then performed a replication analysis using data from stage 2. As an independent set, stage 2 included 1,128 samples with 17,098,949 SNPs (including genotyped and imputed). The combined analysis of two-stage data was performed using linear regression, adjusting for age.

## Statistical Analysis

The associations between the serum p2PSA level and SNP genotypes were evaluated using a quantitative linear regression model assuming additive effects of the alleles (0, 1, and 2). In the regression models, log-transformed p2PSA levels were used as the dependent variable, each SNP as an independent variable, and age as a covariate. This analysis was performed by the PLINK V.1.90 software package (19). *P*-values less than  $5 \times 10^{-8}$  and 0.05 were regarded as significant levels in the GWAS and other analyses, respectively.

A principal component approach was used to evaluate population stratification in the first stage with the EIGENSTRAT software (20). The top two eigenvectors were adjusted as covariates in the quantitative linear regression analysis. Quantile–quantile (Q–Q) plots were performed using the R package (<http://www.R-project.org>). Linkage disequilibrium (LD)-based result clumping analysis was applied to test the independence of the respective SNPs in the 8q24.21 locus using PLINK (19). LocusZoom (21) and haploview (22) were used to create plots of genetic data.

## RESULTS

A total of 886 subjects were recruited in the first stage and 1,128 in the second stage. The clinical characteristics of the cohorts in the two stages are described in **Table 1**. A total of 886 PCa cases were detected in two biopsy cohorts, with an overall positive biopsy rate of 42.9%. No significant difference in the clinical characteristics was observed between the two cohorts (**Table 1**).

In the first stage, 886 subjects were genotyped with the Illumina Asian Screening Array. After quality control, 4,552,207 SNPs and 877 individuals were eligible for GWAS

analysis. We did not observe population structure in our cohort (**Supplementary Figure S1**). In addition, the Q–Q plots revealed an unadjusted inflation factor of 0.9999 (**Supplementary Figure S2**), indicating no evidence of systematic bias for the association of logp2PSA phenotype observed in the current study. The Manhattan plot of GWAS for the first stage is shown in **Figure 1**. We selected all signals associated with the logp2PSA level at  $p < 1 \times 10^{-5}$  for replication analysis. A total of 148 SNPs were selected, and 86 of them reached a *p*-value of  $< 5 \times 10^{-8}$ , which were mainly located in two regions, 8q21.3 and 8q24.21 (**Supplementary Table S1**).

In the second stage, 143 out of 148 SNPs remained qualified after QC in 1,128 subjects. Among the 143 candidate SNPs, rs72725879 at the 8q24 locus was confirmed to be significantly associated with the logp2PSA level at a *p*-value cutoff of 0.0003 (Bonferroni correction of 143 tests). We then combined the data of the two stages and found two SNPs that reached genome-wide significance ( $p = 2.28 \times 10^{-9}$  for rs72725879 and  $p = 5.31 \times 10^{-9}$  for rs13254738) (**Table 2**). Therefore, the 8q24.21 region was revealed to be significantly associated with p2PSA levels.

In the current study, the strongest association effects were observed for two SNPs, rs72725879 and rs13254738, both of which were located in region 8q24.21 of the *PRNCR1* (PCa-associated non-coding RNA 1) gene. rs72725879-T and rs13254738-C showed a significant association with increasing serum levels of p2PSA. The two SNPs were correlated with each other ( $R^2 = 0.79$ ). After adjusting the association results for rs72725879 as a covariate, rs13254738 became insignificant ( $p = 0.054$ ). We then performed an additional univariate analysis for rs72725879 and found that it could explain 2.2% of the total genetic variance for the p2PSA levels.

Variants in the 8q24.21 region have previously been reported to be associated with a risk of PCa (23–30). Due to the potential confounding effects of the p2PSA level and PCa, we also evaluated whether the p2PSA-associated SNPs found in this study were associated with PCa. In the association analysis of the total study population combining the two cohorts, we confirmed the association of rs72725879 with PCa, with a combined odds ratio (OR) of 1.90 and a *p*-value of  $3.44 \times 10^{-18}$  for the T allele. After adjusting for PCa risk and age as covariates, rs72725879 was still associated with the p2PSA level ( $p = 0.017$ ) (**Supplementary Table S2**).

We then grouped the subjects in the two stages into three groups based on their genotypes. The proportions of PCa cases

**TABLE 1** | Characteristics of the participants in the two stages.

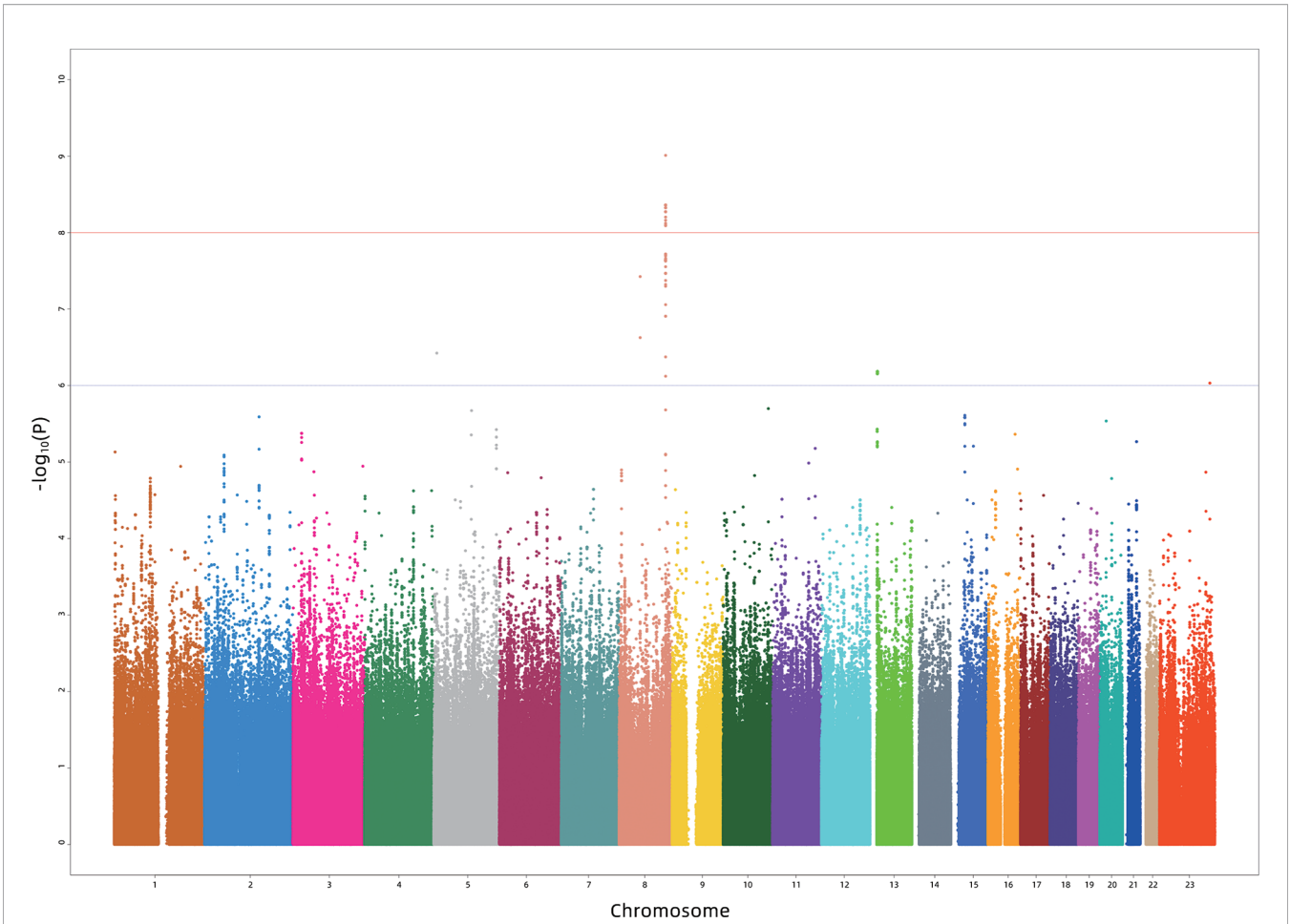
Characteristics	First stage (N = 886)				Second stage (N = 1,128)				<i>p</i> -value <sup>c</sup>
	All	Positive	Negative	<i>p</i> -value <sup>b</sup>	All	Positive	Negative	<i>p</i> -value <sup>b</sup>	
Biopsy result, N (%)	885	401 (45.31)	484 (54.69)		1124	462 (41.10)	662 (58.90)		0.68
p2PSA <sup>a</sup> , median (IQR), pg/ml	22.14 (49.56)	62.98 (312.52)	15.14 (13.97)	$2.02 \times 10^{-68}$	22.92 (40.02)	47.22 (181.97)	17.88 (16.73)	$1.65 \times 10^{-52}$	0.87
Age, mean $\pm$ SD, year	68.26 $\pm$ 9.64	70.87 $\pm$ 8.38	66.10 $\pm$ 10.05	$1.17 \times 10^{-13}$	67.73 $\pm$ 8.51	70.78 $\pm$ 7.92	65.60 $\pm$ 8.30	$8.81 \times 10^{-24}$	0.34

p2PSA [–2], proPSA; IQR, interquartile range; SD, standard deviation.

<sup>a</sup>p2PSA was measured in picograms per milliliter. Data were log-transformed for genome-wide association study (GWAS) analysis.

<sup>b</sup>*P*-values for the difference between positive and negative, Mann–Whitney *U* test for p2PSA, and *t*-test for age.

<sup>c</sup>*P*-values for the difference between two stages, chi-square test for the biopsy results, Mann–Whitney *U* test for p2PSA, and *t*-test for age.



**FIGURE 1 |** Manhattan plot for the genome-wide association study results for the levels of p2PSA in the Chinese population. The X-axis represents the chromosomal position and the Y-axis represents the  $-\log_{10} p$ -value from linear regression. The horizontal dashed line shows the preset threshold of  $p = 1 \times 10^{-6}$ . The horizontal solid line indicates the preset threshold of  $p = 1 \times 10^{-8}$ .

**TABLE 2 |** Results of genome-wide association study (GWAS) for logp2PSA in the cohort.

SNP	CHR	Position <sup>a</sup>	Loci	Gene	Stage	A/a <sup>b</sup>	RAF <sup>c</sup>	Counts (aa/Aa/AA) <sup>d</sup>	Mean levels (pg/ml) (aa/Aa/AA) <sup>d</sup>	Beta (SE)	p-value <sup>e</sup>
rs72725879	8	128103969	8q24.21	PRNCR1	1	T/C	0.76	51/321/495	17.02/29.11/41.78	0.18 (0.037)	$2.08 \times 10^{-6}$
					2	T/C	0.76	70/393/636	22.23/27.10/38.90	0.12 (0.030)	$6.24 \times 10^{-5}$
					Combined	T/C	0.76	121/714/1131	19.86/27.99/40.18	0.14 (0.023)	$2.28 \times 10^{-9}$
rs13254738	8	128104343	8q24.21	PRNCR1	1	C/A	0.75	50/341/485	15.74/29.38/43.05	0.19 (0.037)	$4.22 \times 10^{-7}$
					2	C/A	0.76	73/403/648	27.16/25.94/38.82	0.10 (0.029)	$3.90 \times 10^{-4}$
					Combined	C/A	0.76	123/744/1133	21.78/27.48/40.64	0.13 (0.023)	$5.31 \times 10^{-9}$

<sup>a</sup>Chromosome position based on human genome build 37.  
<sup>b</sup>A/a, Risk allele/reference allele.  
<sup>c</sup>RAF indicates the frequency for risk allele A.  
<sup>d</sup>aa, indicating homozygous with two reference alleles; Aa, heterozygous; AA, indicating homozygous with two risk alleles.  
<sup>e</sup>P-value was based on multivariate linear regression analysis, adjusted for age and eigen.

detected in the CC, CT, and TT groups were 24.18%, 33.77%, and 50.76%, respectively. The T allele of rs72725879 showed a significant positive association with the increasing detection rate of PCa in the prostate biopsy cohort ( $p_{\text{trend}} < 3.92 \times 10^{-21}$ ) (Supplementary Figure S3).

Then, we specifically looked into 14 SNPs associated with a risk of PCa within the 8q24.21 region reported by previous GWAS and assessed their effects on the levels of p2PSA (Table 3). The Manhattan plot of 1,236 SNPs within this region is also shown in Figure 2. These 14 SNPs belonged to



**TABLE 3** | Association analysis between logp2PSA and prostate cancer (PCa) risk-associated variants reported by previous genome-wide association study (GWAS) within the region 8q24.21.

SNP	LD cluster <sup>a</sup>	CHR	BP <sup>b</sup>	Status	MA	N	Beta (SE) <sup>c</sup>	p-value <sup>c</sup>	PCa risk-associated SNPs identified by previous GWAS <sup>d</sup>			
									RA	OR	p-GWAS	PMID
rs12543663	Block 1	8	127,924,659	Genotyped	C	1,919	-0.056 (0.043)	0.20	C	1.12	$2.60 \times 10^{-38}$	29892016
rs10086908	Block 1	8	128,011,937	Genotyped	C	1,919	-0.083 (0.027)	$1.82 \times 10^{-3}$	T	1.13	$1.80 \times 10^{-47}$	29892016
rs1016343	Block 2	8	128,093,297	Genotyped	T	1,919	0.096 (0.02)	$3.04 \times 10^{-6}$	T	1.25	$5.40 \times 10^{-21}$	26034056
rs13252298	Block 2	8	128,095,156	Genotyped	G	1,919	-0.11 (0.022)	$6.50 \times 10^{-7}$	A	1.11	$4.10 \times 10^{-10}$	19767752
rs72725879	Block 2	8	128,103,969	Imputed	C	1,886	-0.14 (0.023)	$2.28 \times 10^{-9}$	NA	NA	NA	NA
rs6983561	Block 2	8	128,106,880	Genotyped	C	1,919	0.11 (0.022)	$3.65 \times 10^{-7}$	C	1.13	$4.60 \times 10^{-19}$	26034056
rs16901979	Block 2	8	128,124,916	Imputed	A	1,919	0.11 (0.022)	$1.58 \times 10^{-6}$	A	1.56	$1.40 \times 10^{-105}$	29892016
rs16902094	Block 2	8	128,320,346	Genotyped	G	1,919	0.0054 (0.023)	0.81	G	1.20	$6.20 \times 10^{-15}$	19767754
rs445114	Block 2	8	12,832,3181	Genotyped	C	1,919	0.0015 (0.020)	0.94	T	1.19	$4.70 \times 10^{-10}$	19767754
rs620861	Block 3	8	128,335,673	Genotyped	A	1,919	0.0093 (0.020)	0.65	G	1.15	$4.90 \times 10^{-63}$	29892016
rs16902104	Block 3	8	128,340,908	Imputed	T	1,827	0.0041 (0.023)	0.86	T	1.21	$5.30 \times 10^{-10}$	26034056
rs6983267	Block 4	8	128,413,305	Genotyped	G	1,919	0.045 (0.020)	0.026	G	1.22	$2.80 \times 10^{-141}$	29892016
rs1447295	Block 5	8	128,485,038	Genotyped	A	1,919	0.094 (0.027)	$5.38 \times 10^{-4}$	A	1.41	$1.20 \times 10^{-179}$	29892016
rs11986220	Block 5	8	128,531,689	Genotyped	A	1,896	0.12 (0.027)	$1.13 \times 10^{-5}$	A	1.56	$2.30 \times 10^{-40}$	26034056
rs7837688	Block 5	8	128,539,360	Genotyped	T	1,919	0.13 (0.027)	$4.25 \times 10^{-6}$	T	1.43	$1.85 \times 10^{-14}$	17401363

LD, linkage disequilibrium; CHR, chromosome; BP, base pair; MA, minor allele; RA, risk allele; OR, odds ratio; NA, not applicable.

<sup>a</sup>LD block information referring to the report by Olama et al. (28).

<sup>b</sup>Chromosome position based on human genome build 37.

<sup>c</sup>P-value, beta and standard error (SE) were based on multivariate linear regression analysis, adjusted for age.

<sup>d</sup>The RA, OR, and p-GWAS were from previously reported GWAS.

five LD blocks in this locus according to previously reported results (28). In the association analysis adjusted for age, we revealed that eight SNPs associated with PCa risk also contributed to the serum p2PSA level at a *p*-value cutoff of 0.0036 (after a Bonferroni correction of 14 tests), although it did not reach a GWAS significance. Among the eight SNPs, rs72725879 in block 2 was found to be in weak LD with rs13252298 ( $r^2 = 0.26$ ) in the Chinese population (Figure 2).

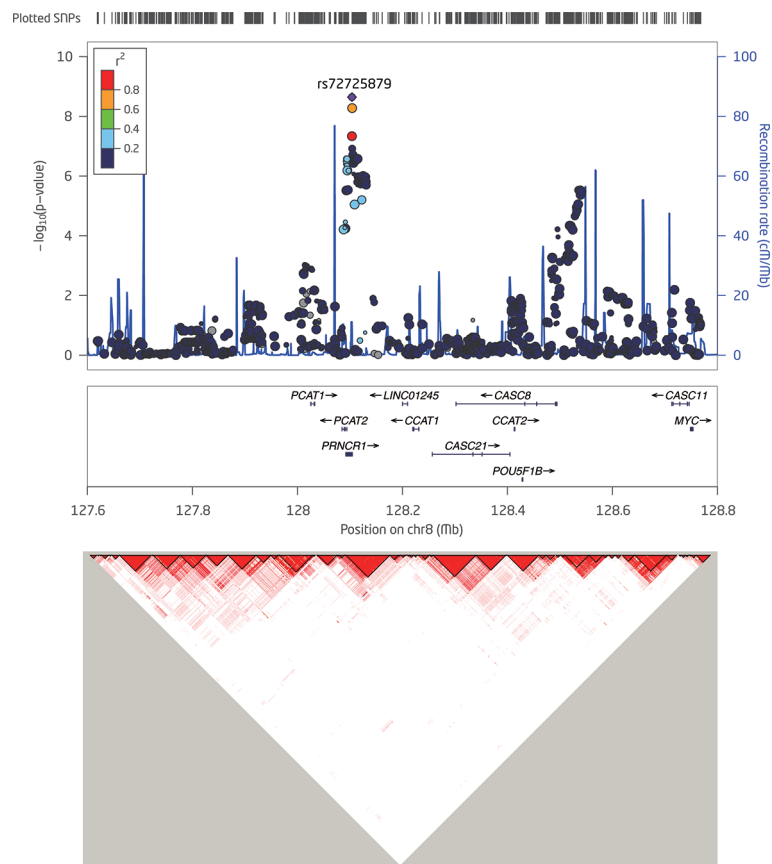
## DISCUSSION

Previous studies have found more than 40 SNPs associated with the serum PSA level. These findings provided important information on the genetic variations in PSA among different individuals and could help improve personalized PSA screening, thereby reducing unnecessary biopsies. Besides PSA, a relatively new biomarker, p2PSA and its derivative phi, have become important biomarkers for PCa diagnosis, especially for men with a PSA level in the range 2.0–10.0 ng/ml. Yet, there is no study on the genetic variants influencing the serum p2PSA level. To our knowledge, this is the first GWAS on the serum p2PSA level in the Chinese population. In this two-stage GWAS in Chinese men who underwent prostate biopsy, we identified one single locus, 8q24.21, associated with the serum p2PSA level at genome-wide significance.

Multiple GWAS and fine-mapping studies had identified that common genetic variations in 8q24 influenced the inherited risk of PCa independently (24–30), while only one SNP (rs17464492) in 8q24 that influences the serum PSA level had been identified in non-Hispanic whites previously (7). In our study, rs72725879 (in region 2) appeared to be the leading SNP that affected the serum p2PSA level in the identified locus, 8q24.21, and it was

located in the exon region of a non-coding RNA gene known as PCa-associated non-coding RNA 1 (*PRNCRI*). This SNP had previously been reported to be associated with PCa in men of African ancestry (30). According to data from The 1000 Genomes Project (1KGP), the risk allele frequency (RAF) of rs72725879 (T) was 0.76 in this study cohort, being higher than that in normal Asian (RAF = 0.66, ASN 1KGP), African (RAF = 0.33, AFR 1KGP), and European (RAF = 0.19, EUR 1KGP) populations. Previous GWAS had also identified that SNPs (rs1016343, rs13252298, rs13254738, and rs6983561) in *PRNCRI* were associated with PCa risk (23, 28, 29). In our two-stage combined analysis, SNP rs72725879 and the above three SNPs (rs1016343, rs13252298, and rs6983561) in *PRNCRI* were also found associated with PCa risk after adjustment for age (all  $p < 5.0 \times 10^{-8}$ ). *PRNCRI*, also known as *PCAT8*, is a long non-coding RNA (lncRNA) that is upregulated in aggressive PCa. This lncRNA could bind to the androgen receptor and enhance the androgen receptor-mediated gene activation programs and proliferation in PCa cells (31). In the current study cohort, the T allele of rs72725879 was also observed to be associated with an increased risk of PCa detected in biopsy. Since a higher PCa risk can have an impact on the serum p2PSA level, it is also possible that the association between rs72725879 and the serum p2PSA level observed in the current study may also reflect some latent or undiagnosed disease.

Distinguishing whether SNPs are associated with p2PSA, PCa, or both is relatively complicated. The levels of p2PSA can be influenced by a number of factors (e.g., age, prostate infection, prostate inflammation, cancerous status, and urological manipulations). Nevertheless, we performed both the association analysis (with p2PSA) after adjusting for age and the association analysis with PCa to address part of these issues. Our results showed that the association between the genetic variant (rs72725879) and the p2PSA level was still significant



**FIGURE 2** | Detailed regional plots of  $-\log_{10} p$ -values in the 8q24.21 region shown for logp2PSA. Colors indicate the linkage disequilibrium (LD) strength between rs7272589 and the other single nucleotide polymorphisms (SNPs) assessed. SNPs with red circle are reported to be associated with risk of prostate cancer (PCa) within the 8q24.21 region for the corresponding block. The right Y-axis shows the recombination rate from the 1000 Genomes Project data as reference. LD maps were based on  $D'$  values using data from the two-stage samples.

after adjusting for PCa and age. This indicated that, although the association between PCa and serum p2PSA level was stronger, which is plausible because p2PSA is a diagnostic predictor for PCa, genetic variance also contributed to the baseline p2PSA level among different individuals. It has been reported that PSA-associated SNPs discovered in GWAS could be used to help normalize an individual's PSA level, and incorporating these genetic factors into the application of PSA screening may increase the ability to classify individuals who should be biopsied (7). Here, we have proven that genetic variants also had an impact on p2PSA, so that personalizing the cutoff value for p2PSA by adjusting for genetic variants that were associated with the p2PSA levels in each individual might enhance the sensitivity and specificity of p2PSA in guiding biopsies.

There were several limitations in our study. Firstly, the study population was relatively small so that some signals might have been missed. Secondly, the overall p2PSA level in the current prostate biopsy cohort would be higher than that in a general population; thus, the findings from our study still need to be further validated in a larger general population. However, approximately 40% of the participants in our study were

PCa cases, which enabled us to evaluate the associations between all p2PSA-associated SNPs and PCa.

## CONCLUSIONS

In the current study, we described the first GWAS in a Chinese prostate biopsy population and identified one single locus at 8q24.21 that was associated with the serum p2PSA level at genome-wide significance. By taking inherited variations between individuals into account, the findings of these genetic variants may help calculate personalized cutoff values for serum p2PSA for patients, thus improving the performance of p2PSA to predict PCa risk.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Huashan Hospital, Fudan University; the Institutional Review Board of Shanghai Cancer Center, Fudan University; and the Institutional Review Board of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine. The patients/participants provided written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JZ, QD, and XL directed and designed the study. YW, FL, RN, DH, DX, JG, YZ, BD, DY, and HJ recruited the study subjects and managed the respective projects. XL and HY performed the bioinformatics and statistical analyses. XL and FL performed the genotyping and p2PSA testing. ZF and QD coordinated the project. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65:87–108. doi: 10.3322/caac.21262
- Liu X, Yu C, Bi Y, Zhang ZJ. Trends and Age-Period-Cohort Effect on Incidence and Mortality of Prostate Cancer From 1990 to 2017 in China. *Public Health* (2019) 172:70–80. doi: 10.1016/j.puhe.2019.04.016
- Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and Overtreatment of Prostate Cancer. *Eur Urol* (2014) 65:1046–55. doi: 10.1016/j.eururo.2013.12.062
- Pilia G, Chen WM, Scuteri A, Orru M, Albai G, Dei M, et al. Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. *PloS Genet* (2006) 2:e132. doi: 10.1371/journal.pgen.0020132
- Sun J, Tao S, Gao Y, Peng T, Tan A, Zhang H, et al. Genome-Wide Association Study Identified Novel Genetic Variant on SLC45A3 Gene Associated With Serum Levels Prostate-Specific Antigen (PSA) in a Chinese Population. *Hum Genet* (2013) 132:423–9. doi: 10.1007/s00439-012-1254-3
- Terao C, Terada N, Matsuo K, Kawaguchi T, Yoshimura K, Hayashi N, et al. A Genome-Wide Association Study of Serum Levels of Prostate-Specific Antigen in the Japanese Population. *J Med Genet* (2014) 51:530–6. doi: 10.1136/jmedgenet-2014-102423
- Hoffmann TJ, Passarelli MN, Graff RE, Emami NC, Sakoda LC, Jorgenson E, et al. Genome-Wide Association Study of Prostate-Specific Antigen Levels Identifies Novel Loci Independent of Prostate Cancer. *Nat Commun* (2017) 8:14248. doi: 10.1038/ncomms14248
- Li W, Bica M, Sjöberg DD, Vertosick E, Dahlin A, Melander O, et al. Genome-Wide Association Study Identifies Novel Single Nucleotide Polymorphisms Having Age-Specific Effect on Prostate-Specific Antigen Levels. *Prostate* (2020) 80:1405–12. doi: 10.1002/pros.24070
- Mikolajczyk SD ML, Wang TJ, Rittenhouse HG, Marks LS, Song W, Wheeler TM, et al. A Precursor Form of Prostate-Specific Antigen Is More Highly Elevated in Prostate Cancer Compared With Benign Transition Zone Prostate Tissue. *Cancer Res* (2000) 60:756–9.
- Guazzoni G, Nava L, Lazzeri M, Scattoni V, Lughezzani G, Maccagnano C, et al. Prostate-Specific Antigen (PSA) Isoform P2psa Significantly Improves the Prediction of Prostate Cancer at Initial Extended Prostate Biopsies in Patients With Total PSA Between 2.0 and 10 Ng/Ml: Results of a Prospective Study in a Clinical Setting. *Eur Urol* (2011) 60:214–22. doi: 10.1016/j.eururo.2011.03.052

## FUNDING

This work was funded by grants from the Clinical Science and Technology Innovation Project of Shanghai Shen Kang Hospital Development Center to Qiang Ding (no. SHDC12015105), and Clinical Research Project of Shanghai Health Commission to YW (No.20214Y0511).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.753920/full#supplementary-material>

**Supplementary Figure 1 |** The principal component analysis was conducted in 886 samples in the first stage and 1,978 from 1000 Genomes data by the first two principal components.

**Supplementary Figure 2 |** The quantile-quantile (Q-Q) plot of the expected and observed *P* values using imputed SNPs data in stage 1.

**Supplementary Figure 3 |** Plot for PCa detection rate by the genotypes of rs72725879.

- Catalona WJ, Partin AW, Sanda MG, Wei JT, Klee GG, Bangma CH, et al. A Multicenter Study of [-2]Pro-Prostate Specific Antigen Combined With Prostate Specific Antigen and Free Prostate Specific Antigen for Prostate Cancer Detection in the 2.0 to 10.0 Ng/Ml Prostate Specific Antigen Range. *J Urol* (2011) 185:1650–5. doi: 10.1016/j.juro.2010.12.032
- Guazzoni G, Lazzeri M, Nava L, Lughezzani G, Larcher A, Scattoni V, et al. Preoperative Prostate-Specific Antigen Isoform P2psa and its Derivatives, % P2psa and Prostate Health Index, Predict Pathologic Outcomes in Patients Undergoing Radical Prostatectomy for Prostate Cancer. *Eur Urol* (2012) 61:455–66. doi: 10.1016/j.eururo.2011.10.038
- Lazzeri M, Briganti A, Scattoni V, Lughezzani G, Larcher A, Gadda GM, et al. Serum Index Test %[-2]proPSA and Prostate Health Index Are More Accurate Than Prostate Specific Antigen and %fPSA in Predicting a Positive Repeat Prostate Biopsy. *J Urol* (2012) 188:1137–43. doi: 10.1016/j.juro.2012.06.017
- Fossati N, Buffi NM, Haese A, Stephan C, Larcher A, McNicholas T, et al. Preoperative Prostate-Specific Antigen Isoform P2psa and Its Derivatives, % P2psa and Prostate Health Index, Predict Pathologic Outcomes in Patients Undergoing Radical Prostatectomy for Prostate Cancer: Results From a Multicenter European Prospective Study. *Eur Urol* (2015) 68:132–8. doi: 10.1016/j.eururo.2014.07.034
- Na R, Ye D, Liu F, Chen H, Qi J, Wu Y, et al. Performance of Serum Prostate-Specific Antigen Isoform [-2]proPSA (P2psa) and the Prostate Health Index (PHI) in a Chinese Hospital-Based Biopsy Population. *Prostate* (2014) 74:1569–75. doi: 10.1002/pros.22876
- Chiu PK, Roobol MJ, Teoh JY, Lee WM, Yip SY, Hou SM, et al. Prostate Health Index (PHI) and Prostate-Specific Antigen (PSA) Predictive Models for Prostate Cancer in the Chinese Population and the Role of Digital Rectal Examination-Estimated Prostate Volume. *Int Urol Nephrol* (2016) 48:1631–7. doi: 10.1007/s11255-016-1350-8
- Na R, Ye D, Qi J, Liu F, Helfand BT, Brendler CB, et al. Prostate Health Index Significantly Reduced Unnecessary Prostate Biopsies in Patients With PSA 2–10 Ng/mL and PSA >10 Ng/Ml: Results From a Multicenter Study in China. *Prostate* (2017) 77:1221–9. doi: 10.1002/pros.23382
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A New Multipoint Method for Genome-Wide Association Studies by Imputation of Genotypes. *Nat Genet* (2007) 39:906–13. doi: 10.1038/ng2088
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* (2007) 81:559–75. doi: 10.1086/519795



20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nat Genet* (2006) 38:904–9. doi: 10.1038/ng1847
21. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: Regional Visualization of Genome-Wide Association Scan Results. *Bioinformatics* (2010) 26:2336–7. doi: 10.1093/bioinformatics/btq419
22. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and Visualization of LD and Haplotype Maps. *Bioinformatics* (2005) 21:263–5. doi: 10.1093/bioinformatics/bth457
23. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-Wide Association Study Identifies a Second Prostate Cancer Susceptibility Variant at 8q24. *Nat Genet* (2007) 39:631–7. doi: 10.1038/ng1999
24. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-Wide Association Study of Prostate Cancer Identifies a Second Risk Locus at 8q24. *Nat Genet* (2007) 39:645–9. doi: 10.1038/ng2022
25. Zheng SL, Sun J, Cheng Y, Li G, Hsu FC, Zhu Y, et al. Association Between Two Unlinked Loci at 8q24 and Prostate Cancer Risk Among European Americans. *J Natl Cancer Inst* (2007) 99:1525–33. doi: 10.1093/jnci/djm169
26. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, et al. Multiple Regions Within 8q24 Independently Affect Risk for Prostate Cancer. *Nat Genet* (2007) 39:638–44. doi: 10.1038/ng2015
27. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, et al. Multiple Newly Identified Loci Associated With Prostate Cancer Susceptibility. *Nat Genet* (2008) 40:316–21. doi: 10.1038/ng90
28. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, et al. Multiple Loci on 8q24 Associated With Prostate Cancer Susceptibility. *Nat Genet* (2009) 41:1058–60. doi: 10.1038/ng452
29. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E, Habel LA, Graff RE, et al. A Large Multiethnic Genome-Wide Association Study of Prostate Cancer Identifies Novel Risk Variants and Substantial Ethnic Differences. *Cancer Discovery* (2015) 5:878–91. doi: 10.1158/2159-8290.CD-15-0315
30. Han Y, Rand KA, Hazelett DJ, Ingles SA, Kittles RA, Strom SS, et al. Prostate Cancer Susceptibility in Men of African Ancestry at 8q24. *J Natl Cancer Inst* (2016) 108. doi: 10.1093/jnci/djv431
31. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-Dependent Mechanisms of Androgen-Receptor-Regulated Gene Activation Programs. *Nature* (2013) 500:598–602. doi: 10.1038/nature12451

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lin, Wu, Liu, Na, Huang, Xu, Gong, Zhu, Dai, Ye, Yu, Jiang, Fang, Zheng and Ding. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Large Multicohort Study Reveals a Prostate Cancer Susceptibility Allele at 5p15 Regulating *TERT* via Androgen Signaling-Orchestrated Chromatin Binding of E2F1 and MYC

## OPEN ACCESS

### Edited by:

Benyi Li,  
University of Kansas Medical Center,  
United States

### Reviewed by:

Jiajia Zhao,  
Huazhong University of Science and  
Technology, China  
Qi Ma,  
Ningbo First Hospital, China

### \*Correspondence:

Ping Gao  
ping.gao@snnu.edu.cn  
Rong Na  
narong.hs@gmail.com  
Gong-Hong Wei  
gonghong\_wei@fudan.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 06 August 2021

**Accepted:** 19 October 2021

**Published:** 10 November 2021

### Citation:

Dong X, Zhang Q, Hao J, Xie Q, Xu B,  
Zhang P, Lu H, Huang Q, Yang T,  
Wei G-H, Na R and Gao P (2021)  
Large Multicohort Study Reveals a  
Prostate Cancer Susceptibility Allele at  
5p15 Regulating *TERT* via Androgen  
Signaling-Orchestrated Chromatin  
Binding of E2F1 and MYC.  
Front. Oncol. 11:754206.  
doi: 10.3389/fonc.2021.754206

Xiaoming Dong<sup>1†</sup>, Qin Zhang<sup>2†</sup>, Jinglan Hao<sup>1†</sup>, Qianwen Xie<sup>1</sup>, Binbing Xu<sup>1</sup>, Peng Zhang<sup>3</sup>,  
Haicheng Lu<sup>1</sup>, Qilai Huang<sup>4</sup>, Tielin Yang<sup>5</sup>, Gong-Hong Wei<sup>2,3\*</sup>, Rong Na<sup>6\*</sup> and Ping Gao<sup>1\*</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biology, College of Life Sciences, Shaanxi Normal University, Xi'an, China,

<sup>2</sup> Biocenter Oulu, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland, <sup>3</sup> Fudan University Shanghai Cancer Center, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai, China, <sup>4</sup> Shandong Provincial Key Laboratory of Animal Cell and Developmental Biology, Department of Animal Science, School of Life Sciences, Shandong University, Qingdao, China, <sup>5</sup> Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China, <sup>6</sup> Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Aberrant telomerase reverse transcriptase (*TERT*) expression is crucial for tumor survival and cancer cells escaping apoptosis. Multiple *TERT*-locus variants at 5p15 have been discovered in association with cancer risk, yet the underlying mechanisms and clinical impacts remain unclear. Here, our association studies showed that the *TERT* promoter variant rs2853669 confers a risk of prostate cancer (PCa) in different ethnic groups. Further functional investigation revealed that the allele-specific binding of MYC and E2F1 at *TERT* promoter variant rs2853669 associates with elevated level of *TERT* in PCa. Mechanistically, androgen stimulations promoted the binding of MYC to allele T of rs2853669, thereby activating *TERT*, whereas hormone deprivations enhanced E2F1 binding at allele C of rs2853669, thus upregulating *TERT* expression. Notably, E2F1 could cooperate with AR signaling to regulate *MYC* expression. Clinical data demonstrated synergistic effects of *MYC*/*E2F1*/*TERT* expression or with the TT and CC genotype of rs2853669 on PCa prognosis and severity. Strikingly, single-nucleotide editing assays showed that the CC genotype of rs2853669 obviously promotes epithelial-mesenchymal transition (EMT) and the development of castration-resistant PCa (CRPC), confirmed by unbiased global transcriptome profiling. Our findings thus provided compelling evidence for understanding the roles of noncoding variations coordinated with androgen signaling and oncogenic transcription factors in mis-regulating *TERT* expression and driving PCa.

**Keywords:** *TERT*, rs2853669, prostate cancer, E2F1, MYC, AR signaling, EMT, crpc

## INTRODUCTION

Prostate cancer (PCa) is one of the leading causes of cancer deaths and a highly heritable cancer in men (1). Genetic heritability was estimated to account for 57% of familial risk in PCa (1). Thus, identification of genetic loci in association with PCa risk and pathogenesis, and illustration of the underlying mechanisms are expected to have substantial influence on our understanding of PCa and on the prevention and treatment of the disease. Thanks mainly to genome-wide association studies (GWAS), a substantial amount of single-nucleotide polymorphism (SNP) loci have thus far been identified to be significantly associated with PCa (2). To date, even though the detailed functions of several PCa risk SNPs have been unveiled, the regulatory mechanisms underlying many SNPs remain unidentified. Focusing our research scope on PCa, important oncogene-associated risk SNPs might be a way to speed up our understanding of the functions and the risk of SNPs.

The human telomeres are a safeguard of chromosome ends, whose main function is the maintenance of telomeric DNA length and chromosomal stability (3, 4). Most malignant cells in cancer, including PCa, achieve unlimited replicative capacity, a hallmark of cancer (5), through activating telomerase for telomere maintenance (6, 7). The key rate-limiting element for telomerase activity was found to be *TERT*, encoding an essential catalytic subunit of telomerase (8–10) that is aberrantly expressed in many types of cancer. Understanding the mechanisms of aberrant *TERT* expression is a fundamental question for human cancer, and *TERT* is also a potential clinical target for improving cancer diagnosis and prognosis (11–13). At present, a variety of transcription factors have been found to regulate the transcription of *TERT* by directly acting on the promoter region of *TERT* or indirectly in the form of complexes. There are two canonical E-box consensus sites at the hTERT core promoter of MYC family transcription factors, which are located at upstream 165 bp and downstream 45 bp relative to the hTERT transcriptional start site (TSS). To date, the mechanisms underlying high *TERT* expression in PCa are still not fully investigated.

In recent years, large-scale genome-wide association studies (GWAS) have made important contributions to the identification of common mutations at *TERT* sites. Many studies have confirmed that somatic mutations and functional SNPs of *TERT* genes are associated with multiple cancer risks. For example, rs2736100 was found to be associated with high risk of glioma, lung adenocarcinoma, and testicular germ cell cancer; rs402710 was found to be associated with lung cancer; and T allele of rs10069690 increased the risk of estrogen-receptor-negative breast cancer. rs2242652, rs7725218, rs2853677, and rs401681 have been reported to be associated with PCa risk. These four SNPs are located at the intron or 3'UTR region of *TERT*. To further validate whether there are SNPs located at the promoter region of *TERT* associated with PCa risk, we focused our research on the SNPs located around *TERT*.

## MATERIALS AND METHODS

### Study Population, SNP Genotyping, and Quality Control

A total of 209,741 male participants with Caucasian ancestry from UK Biobank (release V3) with GWAS genotyping array data and imputation information were included in the present study. A detailed description of the population of UK Biobank was reported elsewhere (14). Briefly, this is a prospective cohort with genetic and phenotypic data from 500,000 individuals across UK between 40 and 69 years at the age of recruitment. A PCa diagnosis was identified as having a record of PCa from national cancer registries and self-report based on ICD-10 code (C61) before or after recruitment in patients with European Ancestry. The current study was based on the update of the database in December 2020. Genome-wide genotype data with imputation and quality control (QC) were provided by UK Biobank. Genotype information of rs2853669 and its nearby region ( $\pm 50$  kb) was then obtained from the dataset for further analysis.

An additional independent population with Chinese Han ancestries was also included in the present study for further validation. It included 2,510 male patients from a biopsy cohort based on Chinese Han population (1,100 PCa patients vs. 1,410 healthy men as controls). Clinical information was collected including diagnosis of PCa, Gleason Score, and prostate-specific antigen. SNPs were genotyped using the Illumina Asian Screening Array (ASA) for Chinese patients. Subjects were removed from subsequent analyses if they met any of the following criteria: (1) overall genotype call rate  $<90\%$ ; (2) duplicates or familial relationship ( $PI\_HAT > 0.025$ ). Genotypic QC of rs2853669 was also performed (1) if the genotype call rate  $>95\%$  in the three populations; and (2) if it passes a Hardy–Weinberg Equilibrium test ( $p > 1 \times 10^{-3}$ ). Finally, this SNP could be further evaluated in these populations.

### Association Study and Fine Mapping

Logistic regression models including additive model, dominant model, and recessive model were used to evaluate the association between the SNPs and different phenotypes of PCa performing by using Plink 1.09. Cox regression survival analyses were performed using R (3.6.3). A two-tailed  $p < 0.05$  was considered statistically significant. Fine mapping for the  $\pm 50$ -kb region of rs2853669 was performed based on the association results from UK Biobank using LocusZoom (<http://locuszoom.sph.umich.edu/>). The linkage disequilibrium (LD) in this region was evaluated by LDlink (15).

### Cell Culture

All cell lines were originally purchased from the American Type Culture Collection (ATCC), and none of these cell lines were found to be contaminated with mycoplasma during our study. HEK293T cells were grown in DMEM (Invitrogen, catalog no. 10569010); LNCaP and 22Rv1 cells were grown in RPMI1640 (Merck, catalog no. R8758). DMEM and RPMI1640 were supplemented with 10% fetal bovine serum (Gibco, catalog no. 10099-141C) and antibiotics (Thermo Fisher Scientific, catalog

no. 15140122). RWPE1 cells were grown in Keratinocyte-Serum Free Medium (10724-011, Gibco) with 0.05 µg/ml epidermal growth factor (PHG0311, Gibco), 0.05 mg/ml bovine pituitary extract (02-104, Merck), and antibiotics. In order to study AR activity, we cultured the LNCaP and 22Rv1 cells in charcoal (Merck, catalog no. C6241-5G) stripping media up to at least 48 h. AR activity was induced by treating cells with 100 nM dihydrotestosterone (DHT; dissolved in methanol, Merck, catalog no. D-077-1ML).

## Transfection and Luciferase Promoter Reporter Assays

DNA fragments surrounding rs2853669 (T allele or C allele) were inserted into the pGL3 Basic vector (Promega, catalog no. E1751) (**Supplementary Table 1**). For plasmid transfection in the white 96-well tissue culture plates,  $0.4 \times 10^5$  22Rv1 and  $0.4 \times 10^5$  LNCaP cells per well were applied to reverse transfection with luciferase reporter plasmids together with pGL4.75 (Promega, catalog no. E6931) using X-tremeGENE HP DNA Transfection Reagent (Roche, catalog no. 11668019) following the manufacturer's instructions. After 48 h, cells were analyzed for luciferase activity using the Dual-Glo<sup>®</sup> Luciferase Assay System (Promega, catalog no. 06366236001) following the manufacturer's instructions. All data came from at least three replicate wells, and statistical analyses were performed by a two-tailed *t* test.

## Immunofluorescence

For immunofluorescence analysis, cells were fixed in 4% paraformaldehyde (Merck, catalog no. P6148) for 15 min at room temperature, permeabilized with 0.2% Triton X-100 (Merck, catalog no. T8787-250ML) for 30 min, washed three times with PBS, and incubated in 3% BSA for 1 h. Samples were then sequentially incubated with primary antibodies (E-Cadherin, Cell Signaling Technology, catalog no. 14472; N-Cadherin, Cell Signaling Technology, catalog no. 14215) overnight at 4°C, washed three times with PBS, and then fluorescent-conjugated with secondary antibodies for 50 min at room temperature. Nuclei were stained with DAPI (Merck, catalog no. 10236276001) for 15 min at room temperature. Slides were examined with an Axio Imager Upright Microscope (Carl Zeiss, Axio Imager M2).

## Cell Viability and Proliferation Assays

Mutated 22Rv1 cells treated with DHT or ethanol ( $2 \times 10^3$  per well) were seeded in 96-well plates. Cell viability and proliferation were determined with XTT assays (Roche, 11465015001) at designed time points by measuring the absorbance at 450 nm, following the manufacturer's instructions. Values were obtained from three replicate wells for each treatment and time point. The results were representative of three independent experiments. Significance was calculated by a two-tailed *t* test.

## Western Blots

Cell pellets were harvested and resuspended in lysis buffer (600 mM NaCl, 1% Triton X-100 in PBS,  $1 \times$  protease inhibitor).

The protein extracts were separated by electrophoresis in a 12% polyacrylamide gel before transferring to PVDF membrane. Membranes were blocked with 5% nonfat milk in TBST for 1 h at room temperature. Incubation with primary antibodies was performed at 4°C overnight (E2F1, ABclonal, catalog no. A19579; MYC, ABclonal, catalog no. A17332; TERT, ABclonal, catalog no. A2979; V5, Invitrogen, catalog no. R960-25; E-Cadherin, Cell Signaling Technology, catalog no. 14472; N-Cadherin, Cell Signaling Technology, catalog no. 14215; GAPDH, ABclonal, catalog no. AC002). Membranes were washed three times with TBST every 10 min before secondary incubation with antibodies fused to horseradish peroxidase (HRP) (goat anti-mouse IgG, ABclonal, catalog no. AS003; goat anti-rabbit IgG, Thermo Fisher Scientific, catalog no. 31460) for 1 h at room temperature. After three final washes, Chemiluminescence signal was developed with SuperSignal<sup>™</sup> West Pico Plus Sensitivity Substrate (Thermo Fisher Scientific, catalog no. 34094).

## Invasion Assay

Cells were detached by trypsinization and resuspended in serum free charcoal stripping media growth medium at  $2.5 \times 10^5$  cells/ml. Two hundred microliters of cell suspension with or without androgen treatment was transferred into 8-µm Transwell inserts (Corning Costar, catalog no. 3422) with 100 µl of Matrigel (diluted with serum free medium to 250µg/ml) coating (BD Biosciences, catalog no. 356230). Charcoal stripping media growth medium (750 µl) with or without androgen was added in the lower chambers. After 48 h, 3.7% formaldehyde was used to fix the cells; after being permeabilized with methanol, cells were stained with Wright-Giemsa (WG16-500ml, Sigma). Cells on the upper surface of the membranes were removed and the cells at the bottom surface of the filters were quantified by counting the number of cells that penetrated the membrane in five microscopic fields (acquired at 20× magnification) per membrane. A two-tailed *t*-test was employed to perform statistical analysis from three replicate inserts.

## Chromatin Immunoprecipitation

The ChIP experiment was performed as previously described (16). Briefly, cells were crosslinked with 1% formaldehyde (Merck, catalog no. F8775) and the reaction was stopped with 125 mM glycine (Merck, catalog no. G8898-1KG). The nuclei were isolated from cells and then suspended in SDS lysis buffer. Nuclear extracts were sonicated to generate an average size of 400-bp chromatin fragment. In each reaction, 6 µg of indicated antibodies (E2F1, Invitrogen, catalog no. 32-1400; MYC, Invitrogen, catalog no. MA1-980) or control IgG (ABclonal, catalog no. AC011) was incubated with 70 µl of Dynabead protein G (10004D, Invitrogen) slurry in IP buffer for 10 h. Then, the supernatant was removed, and fragmented chromatin was diluted in 1.3 ml of IP buffer, which was added onto bead/antibody complexes. After 12 h of incubation, the complex was washed once with wash buffer I (20 mM Tris-HCl, pH 8.0, with 2 mM EDTA, 0.1% SDS, 1% Triton X-100, and 150 mM NaCl) and one time with buffer II (20 mM Tris-HCl, pH 8.0, with 2 mM EDTA, 0.1% SDS, 1% Triton X-100, and 500 mM NaCl),



followed by two washes with buffer III (10 mM Tris-HCl, pH 8.0, with 1 mM EDTA, 250 mM LiCl, 1% Deoxycholate, and 1% NP-40) and buffer IV (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA). DNA was purified with MinElute PCR Purification Kit (Qiagen, catalog no. 28006) and the target DNA fragments were analyzed by qPCR or direct Sanger sequencing of PCR fragments harboring these SNPs.

## Quantitative PCR

Each target fragment was amplified using SYBR Select Master Mix (Applied Biosystems, catalog no. 4472908). All target primers had three technical replicates, and the data were normalized to the control regions; then, the relative enrichment of the target antibodies at target DNA fragment was determined by comparison with the IgG control. Primer sequences used in this experiment can be found in **Supplementary Table 2**.

## Wound Healing Assays

Cells were seeded into 24-well plates and allowed to grow to near confluency. P200 pipette tip was used to scrape the cells and the same field was imaged after 72 h. To induce AR activity, the cells were treated with 100 nM DHT. The area of the wound in each well was analyzed using ImageJ. Results are representative of three independent experiments. Statistical significance was calculated by the two-tailed *t* test.

## Single-Nucleotide Mutation Using CRISPR/Cas9

The experiment was performed according to the previous protocol (17). Briefly, one pair of sgRNA (sgRNA1-top: CACCGCCAGGACCGCGCTTCCCACG, sgRNA1-bottom: AAACCGTGGGAAGCGCGTCTGGC) was designed (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgRNA-design>) and inserted into pSpCas9 (BB)-2A-Puro (PX459) (Addgene plasmid ID: 48139). rs2853669 (T or C) centered DNA fragments were cloned into pGL3 basic vector to generate repair templates (**Supplementary Table 1**). Three hundred nanograms of indicated Cas9 plasmid [pSpCas9 (sgRNA) with sgRNA] and 300 ng of pGL3 vector with repair templates were co-transfected into 22Rv1 cells using Lipofectamine 2000. Medium was changed 24 h later. Puromycin (0.8 µg/ml) (Merck, catalog no. P9620) was added onto transfected cells after 48 h. After non-transfected cells were killed by puromycin, the remaining cells were sorted using flow cytometry to establish single-cell clones. The single cells were seeded in 96-well plates and checked for 9–14 days to rule out the non-single clone. Finally, the single clones were picked up for subculture and genotyping.

## Electrophoretic Mobility Shift Assays

The experiment was performed with LightShift® Chemiluminescent EMSA kit (Thermo Fisher Scientific, catalog no. 20148), following the manufacturer's instructions. Full length of *E2F1* or *MYC* was cloned into vector pcDNA3.1/V5-HisA (Invitrogen, catalog no. V81020). Briefly, double-stranded biotin-labeled consensus DNA was incubated with HEK293T cell nuclear extract with ectopically expressed *E2F1* or *MYC* in a

1× binding buffer, 50 ng/µl poly(dI:dC)-poly(dI:dC), 2.5% glycerol, 0.05% NP-40, 5 mM MgCl<sub>2</sub>, and 10 mM EDTA, and 200-fold excess of unlabeled probes was used for competition assays. The protein complexes were resolved on 6% DNA retardation gels for 1 h at 100 V, transferred to Biodyne B Nylon Membranes (PALL, catalog no. 60208), cross-linked, and detected using streptavidin-HRP conjugate and a chemiluminescent substrate. The oligos used are listed in **Supplementary Table 3**.

## Lentiviral Constructs, Lentivirus Production, and Infection

The shRNA constructs targeting *E2F1* and *MYC* were ordered from Merck. The shRNA sequences used are listed in **Supplementary Table 4**. Second-generation lentiviral vectors were packaged using HEK293T cells. In detail, HEK293T cells were trypsinized and seeded into 3.5-cm plates; 24 h later, the normal medium was replaced with 2 ml of low-glucose DMEM (Invitrogen, catalog no. 21885025) containing 10% FBS and 0.1% penicillin and streptomycin. Cells were co-transfected with indicated shRNA construct (1.5 µg each), pMD2.G (envelope plasmid, 0.375 µg) (Addgene, catalog no. 12259), and psPAX2 (packaging plasmid, 1.125 µg) (Addgene, catalog no. 12260) plasmids using 8 µl of Lipofectamine 2000 (Invitrogen, catalog no. 11668019). The medium was changed to fresh medium after 24 h, and the virus-containing medium was harvested every 24 h up to three times. Lentivirus was passed through a 0.45-µm filter unit and stored at –80°C. For viral transduction, LNCaP cells were seeded in six-well plates at a density of 60%–70%. Sixteen to 20 h later, cell culture medium was replaced with lentivirus-containing medium with final 8 µg/ml polybrene (Merck, catalog no. H9268). For lentivirus-mediated knockdown experiment, virus was removed and replaced by normal medium containing final 1 µg/ml puromycin (Merck, catalog no. P9620) after 24 h. When uninfected control cells were completely killed, the target cells were cultured in normal growth medium with 0.5 µg/ml puromycin.

## Motif Analysis

The effect of rs2853669 on transcription factor binding motifs was analyzed using R package at SNP v1.2.0 (affinity test for regulatory SNP detection) (18) in R (v.3.6.3). Binding affinity tests were performed for the motif matches between *MYC* and *E2F1* with alleles of rs2853669 using the derived motif library ENCODE. R packages “seq Logo” (v. 1.52.0) (19) and “universal motif” (v. 1.4.10) (20) were applied to create and plot motif logos.

## Genotype Imputation

IMPUTE2 (v. 2.3.2)<sup>64</sup> was used to perform the genotype imputation of rs2853669 from three cohorts, TCGA, Stockholm, and Cambridge, composed of 389, 94, and 119 prostate samples, respectively. QCTool (v.2.0.7) (20) was used to assess and perform quality control by setting parameters “-threshold 0.9, -snp-stats”. SNPs that failed to pass the quality checking were excluded from imputation. 1000 Genomes Phase 3 data were selected as reference panel (18). Parameters were set as default “-Ne = 2000 and -k hap = 500”. A SNP-centered 2-MB

region was set for the imputation on chromosome 5. Genotypes in the Gen format were converted to dosage format for downstream analysis. The transcriptional profiling was assessed by Illumina Expression Bead Chip in Swedish and Cambridge human prostate tissue samples, while RNA-seq was used in TCGA samples. The Stockholm and TCGA cohorts were genotyped on Illumina Omni 2.5 and Affymetrix SNP array 6, respectively.

## RNA-Seq and Differential Expression Analysis

Mutated T/T cells and C/C cells were harvested, total RNA was extracted with Trizol Reagent (Roche, catalog no. 11667165001) and ethanol precipitation, and then samples were sent to BGI Group for sequencing. For the RNA sequencing of the T/T and C/C genotyped cells, single-end raw sequence reads were first pre-processed with FastQC (21) to assess the read quality. SOAPnuke (22) was employed to process reads for quality trimming and adapter removal with the following criteria: Reads with adaptors were removed, reads with more than 5% of Ns were filtered out, and low-quality reads with more than 20% of bases with quality score smaller than 10 were ruled out. Cleaned reads were further trimmed to 50 bp. A final FastQC run was performed to ensure the success of previous quality control steps. The processed reads were aligned against the human genome assembly hg38 using STAR version 2.7.2a (23) with default settings. HTSeq (htseq-count) was employed to quantitate aligned sequencing reads against gene annotation from Encode and with parameters “-s no, -i gene\_name”. Differential expression analysis was performed from read count matrix using Bioconductor package DESeq2 (1.26.0) (24). Genes with low expressions (<2 cumulative read count across samples) were filtered out prior to differential expression analysis. A threshold of FDR < 0.05 was applied to generate the differentially expressed gene list. Data were normalized using variance Stabilizing Transformation (VST) method from DESeq2. A sample-to-sample distance matrix using hierarchical clustering with Euclidian distance metric from normalized total transcriptome of each sample was applied to examine expression correlations among biological replicates. Heatmap displaying differentially expressed genes between C/C and T/T was generated using the R package “pheatmap” (1.0.12). We have deposited the RNA-seq data to European Nucleotide Archive (ENA) with study accession number PRJEB47829.

## Gene Set Enrichment Analysis

We applied Gene Set Enrichment Analysis (GSEA) to interpret the biological mechanisms underlying the RNA-Sequencing of T/T and C/C genotyped cells from differential expression results obtained from DESeq2. The pre-ranked gene list was obtained by sorting the “stat” statistics from the differential gene list in a descending order. GSEAPreranked test (25) was applied to examine the enrichment of upregulated genes in C/C allele compared to T/T allele from the MSigDB database including Hallmark gene sets, WikiPathways gene sets, chemical and genetic perturbations gene sets, and Biocarta gene sets.

Parameters were set as follows: Enrichment statistic = “weighted”, Max size (exclude larger sets) = 5,000, number of permutations = 1,000. All other parameters were kept as default. The GSEA enrichment plots were generated using R packages “clusterProfiler” (3.14.3) (26) and “enrichplot” (1.6.1) (27).

## Generation of the CC Genotype Signature

The CC genotype signature was developed from the differentially expressed genes upon C/C vs. T/T alleles from RNA-seq. A list of 311 genes was initially generated by applying cutoff with FDR ≤ 0.01 and  $\log_e \geq 2$  to the 5016-DE gene list. We defined this 311-gene list as CC genotype signature and acquired available gene expression data from cBioportal, which further resulted in a 205-gene signature. We then calculated the Z score value of this signature by summing up the normalized gene expression data and further examined its correlation with AR signaling, CCP, hypoxia, and EMT scores.

## Statistical Analysis

All statistical analyses were performed using RStudio (v. 1.2.5033) (28) with R (v. 3.6.3) (29). Data were obtained from the cBioPortal for Cancer Genomics (30, 31), Oncomine database (32), and GEO database (33, 34). Differential gene expression analyses were applied among normal prostate, tumor, and metastatic tissues from various independent cohorts. Statistical analyses were performed to study the correlation between gene expression levels and clinical features including lymph nodes. Statistical tests for patients with gene expression defined to two groups were calculated using Mann–Whitney *U* test, while Kruskal–Wallis *H* test was applied for cohorts with more than two groups. Kaplan–Meier survival analysis was used to assess the impact of gene expression levels on PCa prognosis and survival. Patients were stratified based on median expression of genes or genotype of rs2853669. For the association between rs2853669 genotype and the prognosis survival, we tested several scenarios considering the synergistic effects of gene expression levels and rs2853669 genotype. The survival analyses were performed and visualized as Kaplan–Meier plots by using the R package “Survival” (v. 3.2.3) (35, 36) and “Survminer” (v. 0.4.7) (37). Function “Surv” was first used to generate the survival models with “time-to-event” and “event status” as input from clinical cohorts. Median expression of genes was further followed to fit the models by function “survfit”. Statistical analyses for all Kaplan–Meier curves were calculated using log-rank test. Cox proportional hazards model was employed to assess the hazard ratio (HR) (38). To examine the association of expression of *TERT*, *E2F1*, and *MYC* with androgen signaling, we devised a representative AR signaling signature with a panel of 10 genes, including *SOX9*, *RAN*, *TNK2*, *EP300*, *PXN*, *NCOA2*, *AR*, *NR1P1*, *NCOR1*, and *NCOR2*. The Z score sum of these 10 genes was calculated, and patients were grouped by the median expression of the AR signaling signature. For correlation analysis of gene expressions, we tested the linear correlation among the expression levels of *TERT*, *MYC*, and *E2F1* in benign prostate and tumor tissues in several independent cohorts from cBioPortal and Oncomine databases. Both Pearson

and Spearman methods were applied to assess the co-expression correlations between gene expression levels. For microarray-based expression profiling, we selected gene probes with lowest *p*-values. Samples with missing genotype, expression, or patient survival data were excluded from analyses. *p*-value < 0.05 was considered to be statistically significant. Statistical tests and figures were generated in R (3.6.3)

## RESULTS

### Discovering an Association of the *TERT* Promoter Variant rs2853669 With PCa Susceptibility

To investigate whether *TERT* expression was affected by SNPs, we performed an association study and fine mapping analysis in several independent populations (see Methods) using the SNPs resided in *TERT* or far from *TERT* 5' promoter region. We found several SNPs associated with PCa risk; however, among these SNPs, only one SNP named rs2853669 (located at 245 bp to ATG site of *TERT*) was residing in the functional region. Fine mapping showed strong linkage disequilibrium (LD) between rs2853669 and SNPs at 5'-UTR, but not the coding region of *TERT* (Figures 1A, B). Interestingly, our results indicated an opposite association of C or T allele of rs2853669 with PCa in different populations. The C allele of rs2853669 was significantly associated with PCa susceptibility (odds ratio, OR = 1.10 (95% CI 1.07–1.13), *p* =  $4.76 \times 10^{-10}$ , Table 1) in Caucasians from UK Biobank (10,207 cases and 199,534 controls), and the T allele of rs2853669 was considered as a risk allele [OR = 1.29 (95% CI 1.01–1.64), *p* = 0.038, Table 1] in the Chinese population (1,100 cases and 1,410 controls). In addition, we observed the similar OR value of T allele at rs2853669 in another Chinese population cohort (39) even though the *p*-value is not significant [OR = 1.28 (95% CI 0.89–1.83), *p* = 0.18, Table 1]. Thus, the regulatory mechanism underlying rs2853669 needs to be deeply investigated to explain these contradictory results.

### rs2853669 Altered Allele-Specific Chromatin Binding of E2F1 and MYC

To shed light on the mechanisms underlying rs2853669 in PCa, we first sought to find whether there might be oncogenic transcription factors binding at this SNP-containing region. We thus observed ChIP-seq enrichment signals of MYC and E2F1 at the rs2853669-containing region, and this observation was further verified in LNCaP cells (Figures 1C, D). We next performed SNPs and transcription factor DNA-binding motif matching analysis (40) and examined whether variation at rs2853669 directly modulates transcription factor binding. This analysis revealed that rs2853669 maps within the binding motifs of both E2F1 and MYC (Figure 1E). Interestingly, E2F1 had a higher preference for the C allele whereas MYC favored the T allele of rs2853669. To confirm this allele-specific DNA binding, we performed electrophoretic mobility shift assays (EMSAs) with HEK293T cell nuclear extract containing ectopically expressed E2F1 or MYC, respectively. Consistent with the motif analysis

results, we observed the DNA binding preference of E2F1 to the C or A allele and MYC to the rs2853669 G or T allele (Figures 1F, G). In line with *in silico* motif analysis and *in vitro* EMSA results, our ChIP-qPCR followed by Sanger sequencing results further proved that MYC or E2F1 preferentially bound to different alleles of rs2853669 *in vivo* in LNCaP cells (Figure 1H). Altogether, we demonstrated that both MYC and E2F1 can bind at the *TERT* promoter region, and MYC prefers binding to T allele while E2F1 favors the C allele of rs2853669.

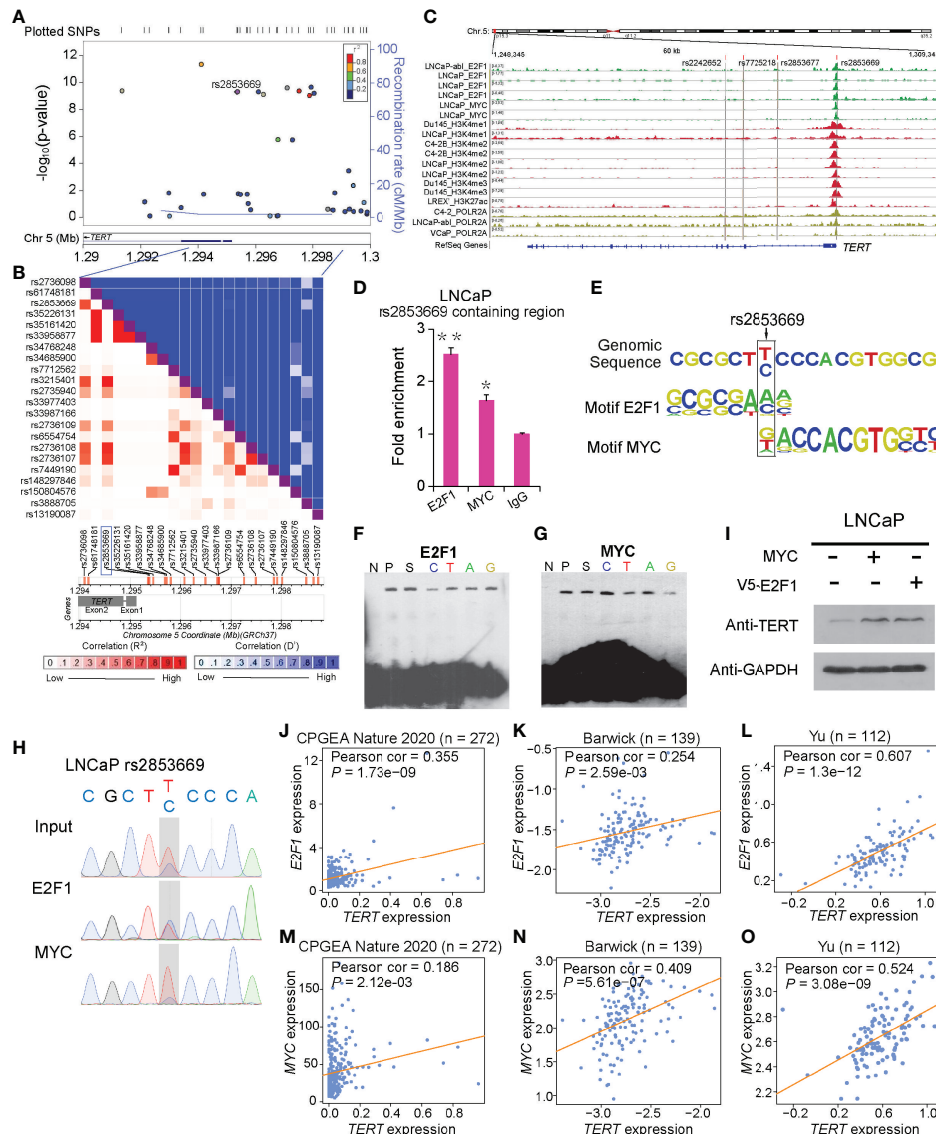
### *TERT* Was a Potential Target Gene of E2F1 and MYC

To assess whether MYC or E2F1 affects the expression of *TERT*, we performed both ectopic overexpression and short hairpin RNA mediated knockdown of MYC or E2F1 assays in PCa cells. The results showed that the protein level of *TERT* was upregulated upon overexpression of MYC or E2F1 and downregulated upon knockdown of MYC or E2F1 (Figure 1I and Supplementary Figures S1A–D). Furthermore, we observed significant positive correlations between the mRNA expression levels of *TERT* and MYC or E2F1 respectively in multiple large cohorts of clinical prostate tissue samples (Figures 1J–O and Supplementary Figures S1E–J), suggesting that MYC and E2F1 regulate the expression of *TERT* in the clinical settings. Interestingly, our genome-wide co-expression analysis revealed that *TERT* is the target gene in both TGCA and CPGA cohorts (Supplementary Figures S1K, L), whereas *TERT* is a top-ranking target gene positively correlated with E2F1 in one cohort (41) but not with MYC, while in the other cohort (42), *TERT* and MYC were highly positively correlated, and no significant expression correlation was observed between *TERT* and E2F1 (Supplementary Figures S1M, N). These results further indicated that MYC or E2F1 might regulate *TERT* expression through different regulatory mechanisms.

### The Role of E2F1 in Regulating MYC Disturbed by AR Signaling Pathway

Given that there are close regulatory associations between AR signaling and MYC or E2F1 in prostate tumors (43–45), we investigated the correlation between AR signaling intensity and the expression of MYC, E2F1, or *TERT*. We observed higher MYC and lower E2F1 expression levels in the PCa patient group with higher AR signaling activity compared to that of lower AR signaling group (Figures 2A, B, D, E and Supplementary Figures S2A–D). No clear expression correlation between *TERT* and AR signaling was observed despite the fact that *TERT* was regulated by MYC and E2F1 (Figures 2C, F and Supplementary Figures S2E, F). Moreover, E2F1 expressions were elevated in metastasis prostate tumors in several independent cohorts of PCa (Figure 2G and Supplementary Figures S2G–I). *TERT* expressions showed an elevated trend in metastasis prostate tumors in several independent cohorts of PCa even though the significance was not strong in Taylor and Yu datasets (Figure 2H and Supplementary Figures S2J–L). In contrast, the expressions of MYC in metastasis prostate tumors





**FIGURE 1** | rs2853669 modulated E2F1 and MYC binding to the *TERT* promoter region. **(A)** Regional plot of the associated SNPs at chr5:1.29mbp–1.30mbp. The plot showed the  $-\log_{10} p$ -values of association between the SNPs in this region and PCA. The intensity of red shading indicated the strength of LD ( $r^2$ ) with the index SNP (rs2853669). **(B)** LD matrix plot indicated the LD pattern in the 50-kb region around the 5'-UTR region of *TERT* based on 1000 Genome CEU population. **(C)** ChIP-seq enrichment signals showing the bindings of E2F1 and MYC at the *TERT* promoter region. **(D)** ChIP-qPCR confirmation of the chromatin bindings of E2F1 and MYC at the rs2853669-containing region in LNCaP cells over background levels using a nonspecific IgG antibody. **(E)** rs2853669 resided within the DNA-binding motifs of E2F1 and MYC. E2F1 was predicted to preferentially bind to C allele of rs2853669, while MYC favored T allele of rs2853669. **(F, G)** The C allele of rs2853669 showed stronger binding affinity for E2F1 than the T allele in an electrophoresis mobility shift assay (EMSA) **(F)**. The T allele of rs2853669 showed stronger binding affinity for MYC than the C allele in an EMSA **(G)**. Lane N represented no protein extract for DNA to bind. The binding of the consensus sequence to E2F1 **(F)** or MYC **(G)** (lane P) was competed by a scrambled sequence (lane S) and by sequences containing the C allele (lane C) and T allele (lane T) of rs2853669 or permutations of an A base (lane A) or G base (lane G) at the same chromosomal location as rs2853669. **(H)** E2F1 or MYC preferred the binding to C allele or T allele at rs2853669 respectively confirmed by ChIP followed by Sanger sequencing in LNCaP cells. **(I)** Overexpression of MYC or E2F1 in LNCaP cells elevated the protein level of TERT by Western blot. **(J–O)** Scatterplots showing a positive expression correlation between *TERT* and *E2F1* **(J)** or *MYC* **(K)** in prostate tumor tissues. Error bars, s.e.m.  $n = 3$  technical replicates.  $p < 0.05$ ,  $**p < 0.01$ . The  $p$ -values were assessed using two-tailed Student's  $t$  tests.

in different cohorts were inconsistent (**Figure 2I** and **Supplementary Figures S2M–O**). These inconsistent results might be attributable to the differences in AR signaling pathway in the metastasis prostate tumors. Therefore, the association of *MYC*, *E2F1*, and *TERT* with androgen-

dependent PCA progression to androgen-independent stage needs to be deeply investigated.

Thus, we next cultured LNCaP cells with hormone-deprived medium for 18 days and analyzed the expression of *MYC*, *E2F1*, and *TERT* at five time points (**Figure 2J**). Notably, the *MYC* level



**TABLE 1 |** The association between rs2853669 variants and PCa or disease aggressiveness.

Phenotypes of PCa	rs2853669 genotypes	UK Biobank-CEU ( <i>n</i> = 209,741)		China ( <i>n</i> = 2,510)		China ( <i>n</i> = 2,425) (39)	
		OR (95% CI)	<i>p</i> -value	OR (95% CI)	<i>p</i> -value	OR (95% CI)	<i>p</i> -value
PCa	Additive	1.10 (1.07–1.13)	4.76E–10	1.10 (0.97–1.25)	0.12	1.14 (0.97–1.34)	0.10
	Dominant	1.12 (1.07–1.16)	1.02E–07	1.07 (0.89–1.27)	0.49	1.14 (0.97–1.34)	0.10
	Resessive	–	–	1.29 (1.01–1.64)	0.038	1.28 (0.89–1.83)	0.18

Reference allele: T. Alternative allele: C.

UKB-case/control: 10,207 cases vs. 199,534 controls.

Chinese: 1,100 cases vs. 1,410 controls.

Chinese: 1,417 cases vs. 1,008 controls (39).

gradually decreased during androgen deprivation, whereas *TERT* and *E2F1* progressively increased during androgen deficiency for 18 days. Next, we aimed to examine the regulatory correlation between *MYC* and *E2F1*. Notably, a significant transcriptional expression correlation between *MYC* and *E2F1* was observed (**Figures 2K–P**). Our validation experiments showed that ectopic expression of *E2F1* markedly elevated *MYC* protein expression; in contrast, ectopic expression of *MYC* showed no effect on *E2F1* protein level (**Figures 2Q, R**). This was consistent with our observations that *MYC* protein level was significantly downregulated upon knockdown of *E2F1*, while knockdown of *MYC* had no impact on *E2F1* protein level (**Supplementary Figures S1C, D**). In addition, it was reported that *E2F1* physically interacts with AR (46), suggesting that the regulatory association between *E2F1* and *MYC* might correlate with AR signaling pathway. We thus treated PCa cells with or without dihydrotestosterone (DHT) and performed ChIP-qPCR experiment. The results demonstrated that *E2F1* has an apparent stronger binding at *MYC* promoter region in cells with DHT treatment than in control cells (**Figure 2S**), indicating an androgen signaling-dependent manner. Taken together, in androgen-responsive cells, *E2F1* might cooperate with AR signaling pathway to promote *TERT* expression through upregulating the expression level of *MYC* when cells were under higher androgen stimulation. However, in an androgen deprivation environment, *E2F1* could directly upregulate *TERT* expression without the involvement of *MYC* to assist androgen-dependent cells' survival in adverse environments.

### AR Signaling Pathway Coordinated rs2853669-Mediated Regulation of *TERT* Expression Through *E2F1* and *MYC*

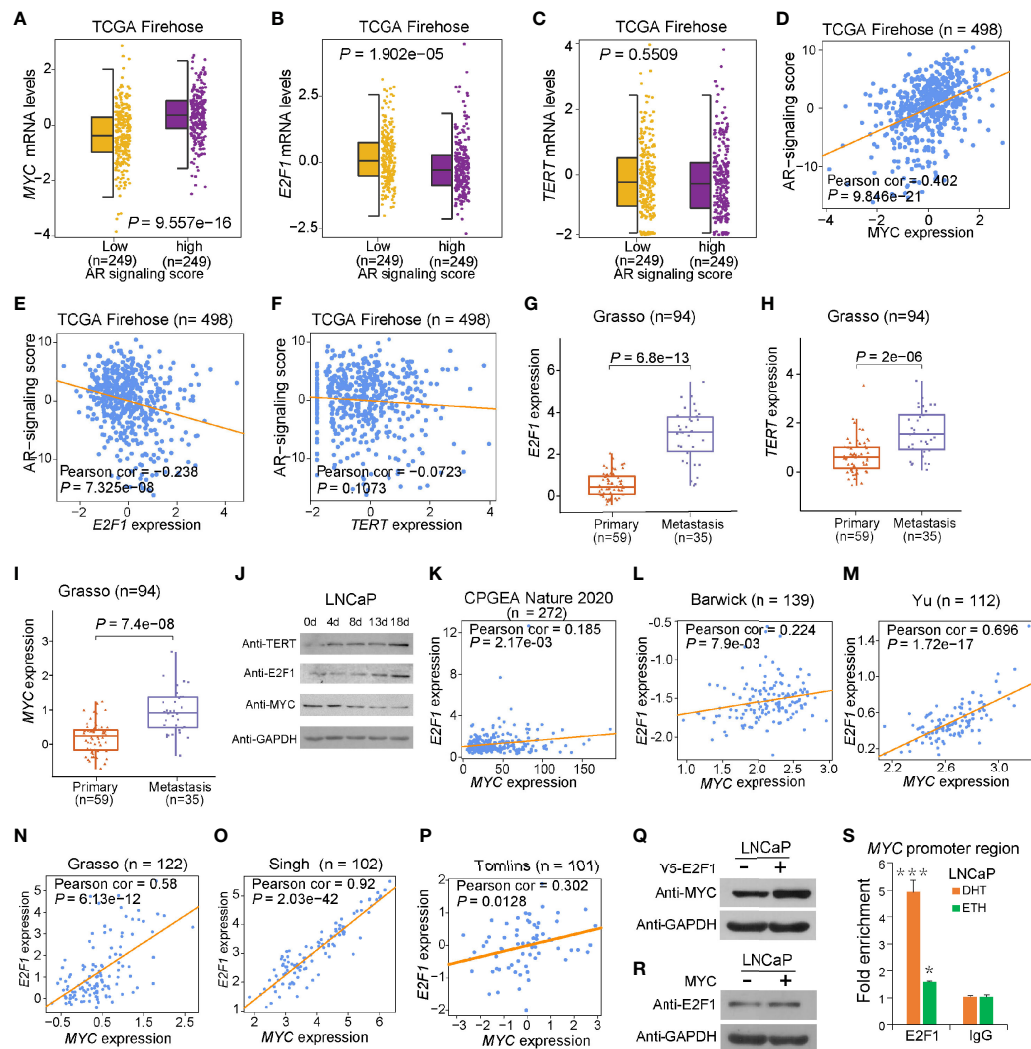
These findings together lead us to an important assumption whether the regulatory mechanisms underlying *MYC* and *E2F1* at the rs2853669 region were coordinated with the status of androgen signaling. To test this hypothesis, we performed luciferase-based promoter assays and observed that androgen stimulus obviously increased the promoter activity of the region harboring T allele but decreased the promoter activity of the region with the C allele in LNCaP and 22Rv1 cells, respectively. Conversely, the region harboring the C allele was observed with an increased promoter activity compared to the region with T allele under androgen deprivation conditions (**Figure 3A** and **Supplementary Figure S3A**). We then tested the promoter

activity in RWPE1 cells; T allele increased the promoter activity when undergoing DHT treatment, but both the plasmid with T and C allele had no obvious luciferase activity in normal prostate epithelial cells (**Supplementary Figure S4B**). We further performed ChIP followed by quantitative PCR and Sanger sequencing in androgen-sensitive PCa LNCaP cells. Androgen stimulation significantly promoted *MYC* binding at the fragment harboring rs2853669, whereas androgen deprivation obviously increased the enrichment of *E2F1* at this region (**Figures 3B, C**). Different from cancer cells, we did not observe the enrichment of either *E2F1* or *MYC* at this region in normal RWPE1 cells, whereas *MYC* showed obvious enrichment at this SNP containing region when the cells were treated with DHT (**Supplementary Figures S4C, D**). Moreover, *MYC* was preferentially recruited to T allele at rs2853669 under DHT treatment, whereas *E2F1* showed stronger binding affinity with C allele than T allele after androgen removal (**Figure 3D**).

Consistently, we observed that C allele at rs2853669 is associated with higher *TERT* expression in PCa patients with higher expression levels of *E2F1*, and *TERT* indicated a trend to be upregulated in patient group carrying T allele at rs2853669 with higher *MYC* expression levels (**Figures 3E, F**). There was no correlation between *TERT* expression and those two alleles at rs2853669 in patient group with lower *E2F1* or *MYC* expression levels (**Supplementary Figures S3E, F**). To further provide evidence to support our results, we overexpressed *E2F1* or *MYC* in T/T and C/C mutated cells; ectopic overexpression of *E2F1* in C/C cells could significantly enhance *TERT* expression level compared with that in T/T cells, while ectopic overexpression of *MYC* in T/T cells could significantly enhance *TERT* expression level compared with that in C/C cells (**Supplementary Figures S4F, G**). Taken together, these results suggested that T allele could promote *TERT* expression through increasing the binding of *MYC* to *TERT* promoter when PCa cells were androgen-stimulated. In contrast, *TERT* expression level was maintained through cooperating the C allele of rs2853669 with *E2F1* under androgen-deficient environment in PCa cells.

### Direct Effects of rs2853669 on EMT and CRPC

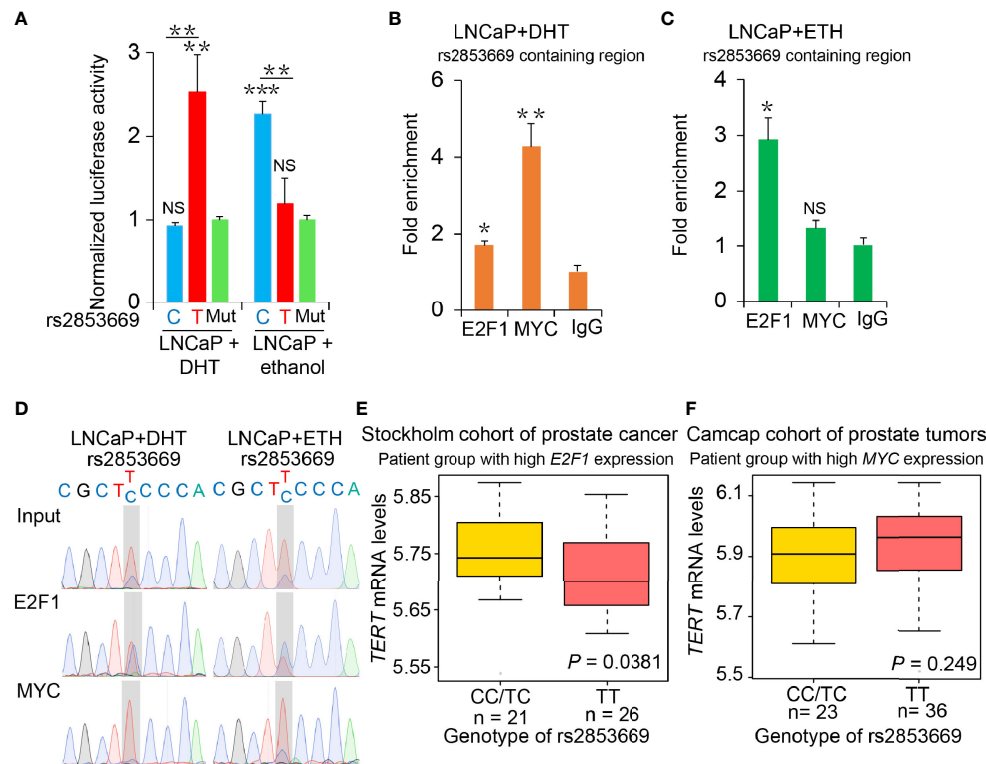
To further assess the phenotypic impacts of rs2853669 alteration, we first performed single-nucleotide mutation using CRISPR/Cas9-mediated genome editing approach and successfully



**FIGURE 2 |** Effect of AR signaling on *E2F1*, *MYC*, and *TERT* expression. (A, D) *MYC* was upregulated in the group with high AR signaling signature score (SOX9, *RAN*, *TNK2*, *EP300*, *PXN*, *NCOA2*, *AR*, *NR1P1*, *NCOR1*, and *NCOR2*). *MYC* was significantly positively correlated with AR signaling score. (B, E) *E2F1* was downregulated in AR signaling high group compared to AR signaling low group. *E2F1* was negatively correlated with AR signaling score. (C, F) No significant correlation was observed for *TERT* upon the AR signaling signature in the TCGA cohort. The Z score sum of the ten-gene AR signaling signature was stratified based on the median score. *p*-values were calculated using Mann–Whitney *U* test. (G–I) mRNA levels of *E2F1* (G), *TERT* (H), and *MYC* (I) were elevated in human benign, primary, and metastasis PCA in the Grasso cohort. *p*-values were calculated using Kruskal–Wallis test. (J) Western blot showing protein levels of *TERT*, *E2F1*, and *MYC* in LNCaP cells with hormone-deprived medium for the indicated days. GAPDH was used as a loading control. (K–O) Expression correlation of *MYC* with *E2F1* in human prostate tissues. Scatter plot showing the direct correlation between *MYC* with *E2F1* expression in the CPGA cohort ( $n = 272$ ) (K), Barwick cohort ( $n = 139$ ) (L), Yu cohort ( $n = 112$ ) (M), Grasso cohort ( $n = 122$ ) (N), Singh cohort ( $n = 102$ ) (O), and Tomlins ( $n = 101$ ) (P). (Q, R) Western blot showing the protein level of *MYC* in LNCaP cells overexpressing V5 tagged *E2F1* (P) and *E2F1* expression level in LNCaP cells overexpressing *MYC* plasmid (Q). GAPDH was used as a loading control. (S) ChIP–qPCR results showed *E2F1* chromatin binding at *MYC* promoter region in LNCaP cells. Error bars, s.e.m.  $n = 3$  technical replicates. \* $P < 0.05$ , \*\*\* $P < 0.001$ . The *P* values were assessed using two-tailed Student's *t* tests.

converted the genotype of rs2853669 from T/C to T/T or C/C in the PCa cell line 22Rv1 (Figure 4A). We applied ChIP–qPCR to examine the occupancy status of *MYC* or *E2F1* at the rs2853669-containing region in our mutated cells. We found that *MYC* chromatin occupancy at the rs2853669 locus was higher in the TT clones than in the TC and CC clones, while *E2F1* chromatin occupancy at the rs2853669 locus was higher in the CC clone than the TC and TT clones (Figure 4B).

Although there were no obvious morphology differences among rs2853669 T/T, parental T/C, and C/C 22Rv1 cell lines (Supplementary Figure S4A), cadherin switching, a major hallmark of EMT (36), was observed among those clones. Downregulation of E-cadherin and upregulation of N-cadherin were confirmed by immunofluorescence and Western blot studies, indicating that C allele of rs2853669 played a pivotal role in EMT (Figures 4C, D). To explore whether androgen had

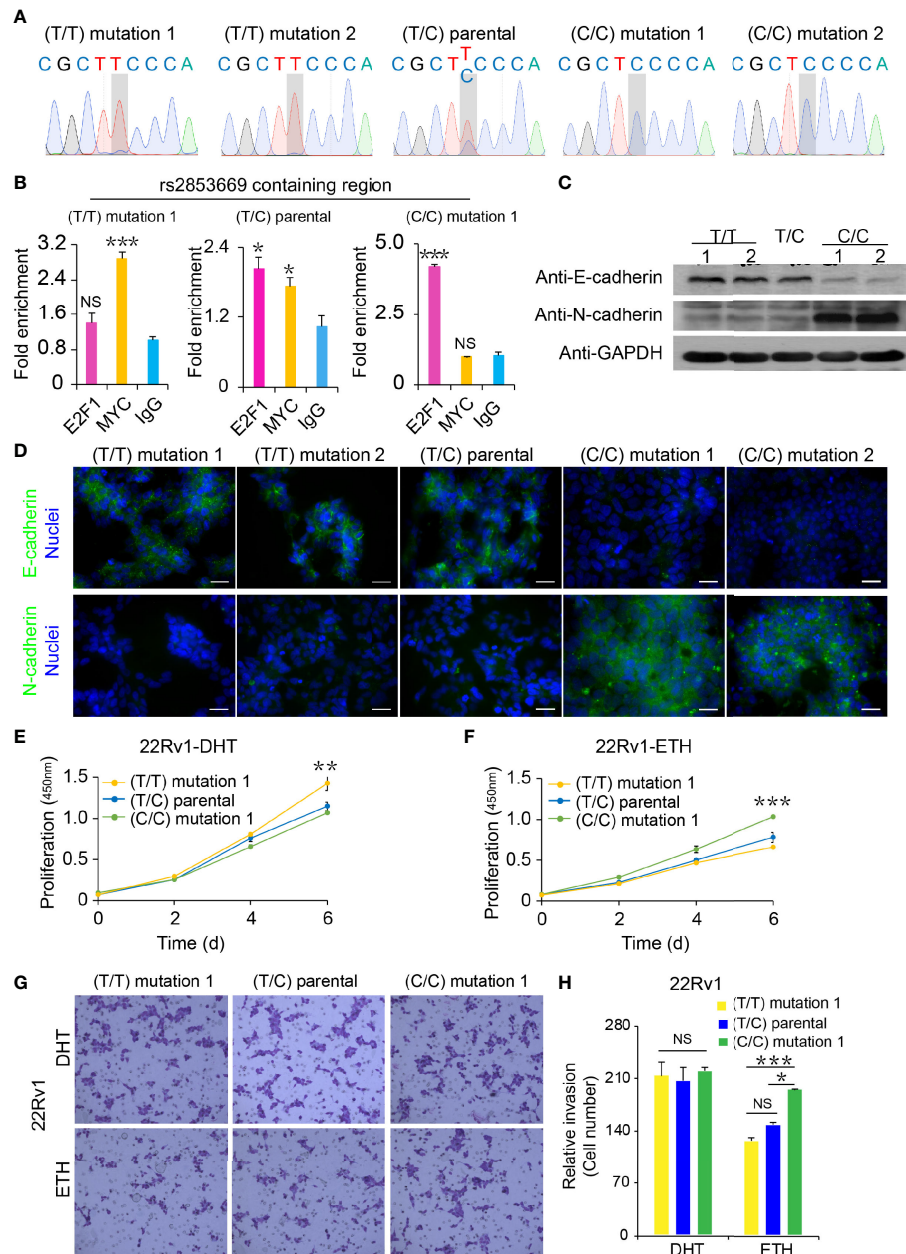


**FIGURE 3 |** The binding preference of E2F1 and MYC at rs2853669 affected *TERT* expression. **(A)** Luciferase reporter assays showing increased promoter activity of the T allele at rs2853669 relative to the C allele in LNCaP cells after androgen treatment. The promoter activity of the T allele at rs2853669 relative to the C allele was diminished in LNCaP cells after removing androgen. Mut, deletion of MYC or E2F1-binding site with rs2853669. **(B, C)** ChIP-qPCR results showed MYC and E2F1 chromatin binding at rs2853669-containing region in LNCaP cells under androgen stimulation **(B)** or withdrawal **(C)**. **(D)** MYC or E2F1 favored binding to the T or C allele at rs2853669 with or without androgen treatment determined by ChIP followed by Sanger sequencing. **(E, F)** The association between rs2853669 genotype and *TERT* expression in prostate tumor samples. C allele of rs2853669 was significantly associated with elevated mRNA expression of *TERT* in *E2F1* high expression group **(E)**; Homozygous TT genotype of rs2853669 was correlated with higher expression of *TERT* in MYC high expression group **(F)**. Patients were pre-stratified based on median expression of *E2F1* or MYC. *p*-values were examined by a log-rank test. Error bars, s.e.m.  $n = 3$  technical replicates. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , Student's *t* tests. NS, Non significance.

influence on cellular proliferation and migration of the 22Rv1 cells with different genotypes of rs2853669, we treated cells with or without DHT and performed cell proliferation, wound healing, and invasion assays. T/T cells grow faster than T/C and C/C cells under DHT treatment (Figure 4E), whereas C/C cells grow faster than the other two genotyped cell clones after removing androgen (Figure 4F). Consistently, we found an obvious inhibition of wound closure and invasion ability in both T/T and T/C cells after removing androgen, while wound closure and invasion ability in C/C cells was not affected by androgen deficiency (Figures 4G, H and Supplementary Figures S4B–E). However, we did not observe obvious difference between the wound closure and invasion ability among T/T, T/C, and C/C cells when treated cells with DHT.

To further explore the underlying biological mechanisms that were affected by the different alleles at rs2853669, we next performed RNA sequencing (RNA-seq) analysis of T/T and C/C genotyped cells. Two biological replicates were performed in each group and high correlations between replicates were observed (Supplementary Figure S5A). The differential gene expression

analysis identified 2,644 and 2,462 significantly upregulated and downregulated genes, respectively (FDR < 0.05; Figure 5A). Gene Set Enrichment Analysis (GSEA), performed in various gene sets from MSigDB, identified multiple pathways relevant to cell cycle and cancer development significantly enriched in upregulated genes of C/C cells compared to control cells with T/T alleles. Several cell growth-related genes in androgen response, hypoxia, Met, and EGF pathways were found upregulated in C/C cells (Supplementary Table S5, Figures 5B, C, and Supplementary Figures S5B, C). To explore the role of targeted genes affected by C/C alleles at rs2853669, we developed a CC genotype signature score and examined its relevance in various independent PCa cohorts. We found significant positive linear correlation between CC genotype signature and AR signaling (Figures 5D–F and Supplementary Figures S5D–F), cell cycle progression (CCP) (Figures 5G–I), and hypoxia scores (Figures 5J, K and Supplementary Figures S5G, H), which was in line with the results showing that C/C cells possessed more proliferative capability (Figure 4F). Moreover, in concordance with the above experimental results (Figures 4G, H), we found that

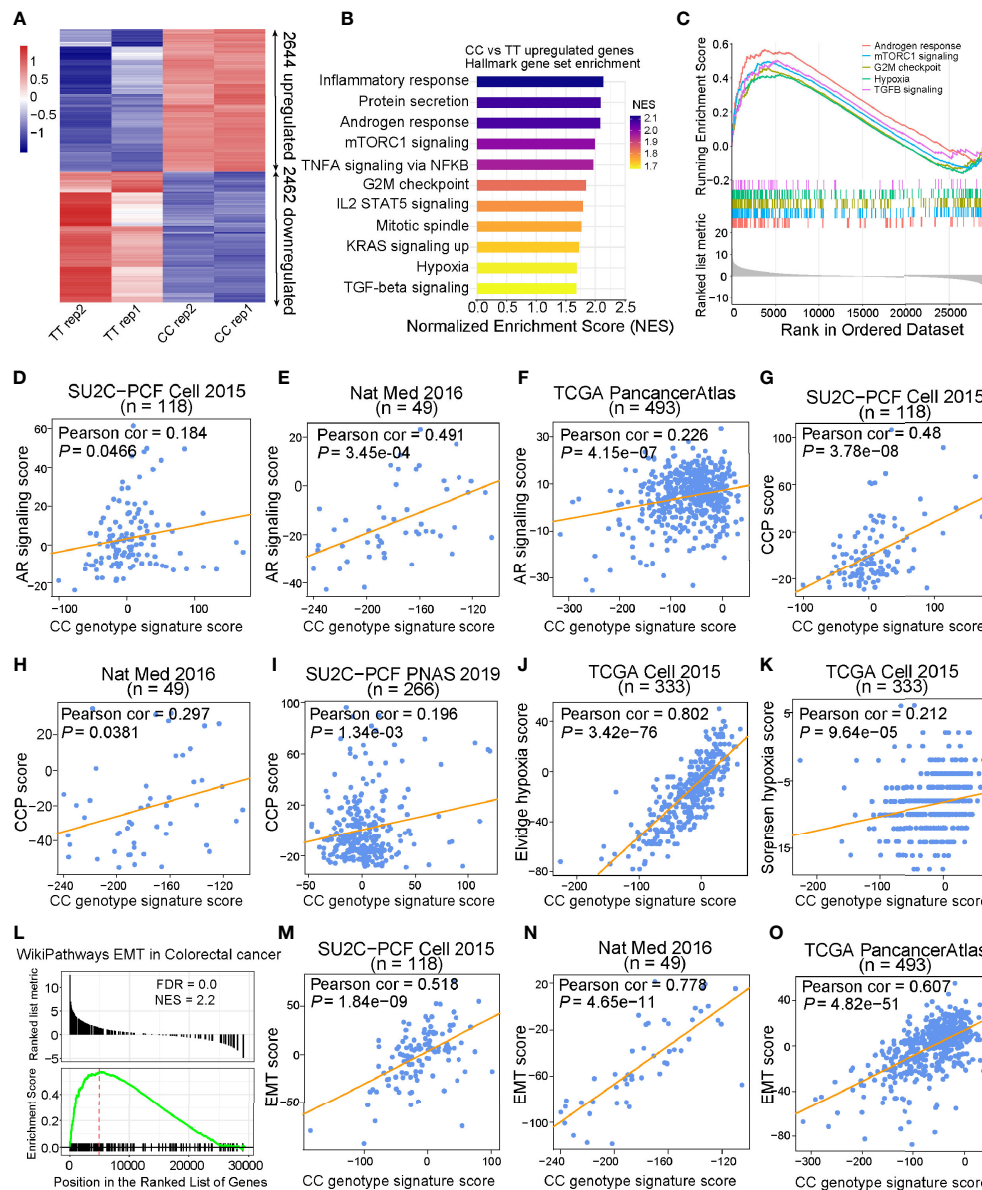


**FIGURE 4 |** Functional analysis of CRISPR/Cas9-modified PCa cells with different rs2853669 genotype. **(A)** Sanger sequencing of CRISPR/Cas9-modified and parental 22Rv1 cells. **(B)** Chromatin enrichment of E2F1 and MYC at the rs2853669 site measured by ChIP-qPCR. **(C)** Western blot showing the protein levels of E-cadherin and N-cadherin in CRISPR/Cas9-modified and parental 22Rv1 cell lines. GAPDH was used as a loading control. **(D)** Representative microscopy analysis of E-cadherin (green; upper panels) and N-cadherin (green; lower panels) expression in CRISPR/Cas9-modified and parental 22Rv1 cell lines. Nuclei were counterstained with DAPI (blue). Scale bars, 20  $\mu$ m. **(E, F)** Cell proliferation analysis of the TT genotype, TC genotype, and CC genotype at rs2853669 in 22Rv1 cells under androgen stimulation **(E)** or withdrawal **(F)**, mean  $\pm$  SD of triplicate experiments. **(G, H)** Representative images of invasion **(G)** assays for cells under androgen stimulation or withdrawal. Scale bars, 100  $\mu$ m. The number of cells in invasion **(H)** assays.  $\pm$  s.e.m. from three biological replicates, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , Student's  $t$  test. NS, Non significance.

genes involved in EMT, metastasis, and TGF- $\beta$  pathways were also upregulated in C/C cells (Figure 5L and Supplementary Figures S5I, J). The correlation analysis also revealed remarkable positive association between CC genotype signature score and EMT score in multiple PCa cohorts (Figures 5M–O and

Supplementary Figures S5K–M), further indicating a higher invasiveness capacity of C/C cells compared to T/T cells. Collectively, our data validated that CC alleles maintain cell survival in hormone deficiency environment and play a crucial role in EMT progression.





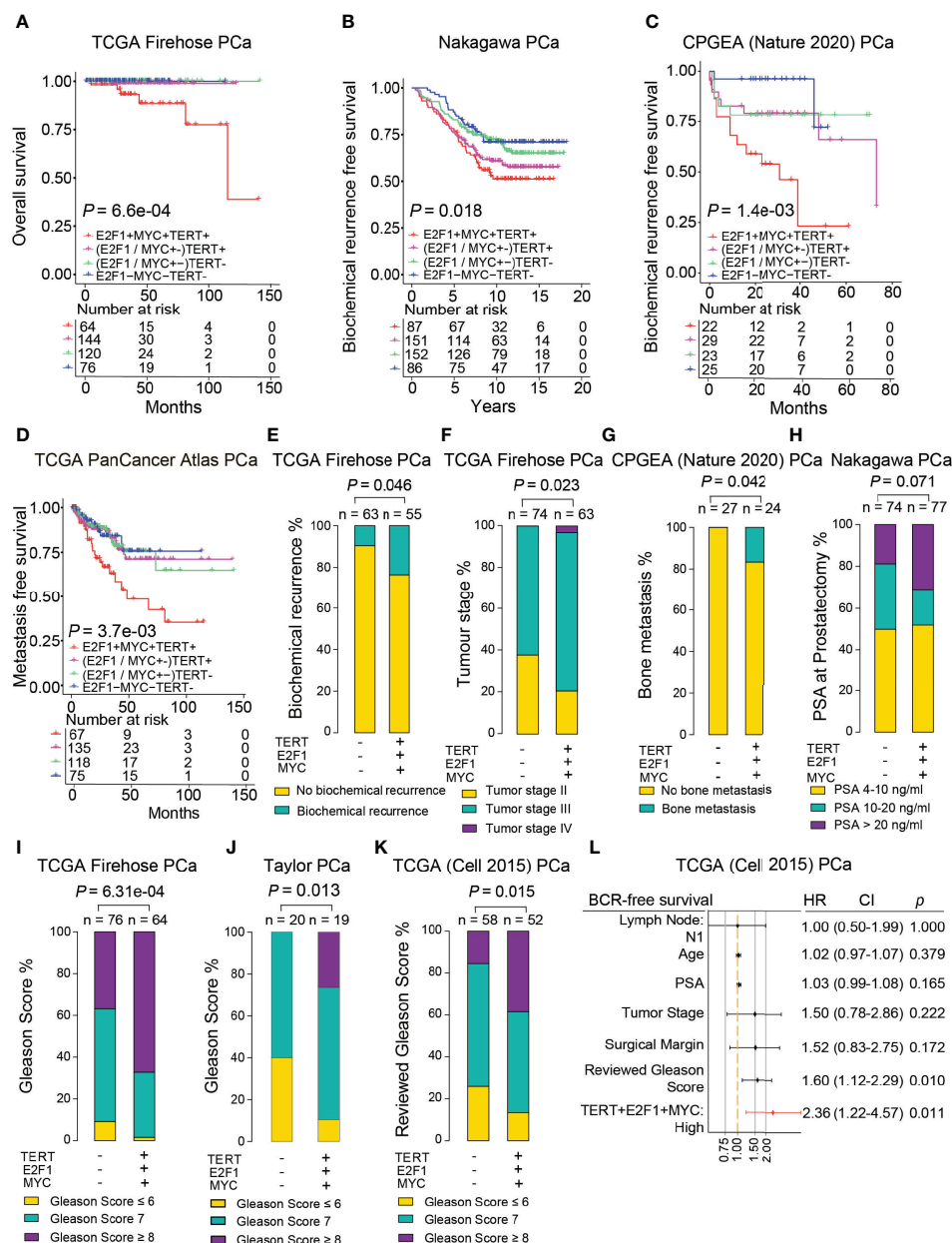
**FIGURE 5 |** CC alleles upregulated cell cycle and EMT genes that are associated with cancer. **(A)** Heatmap of CC allele target genes measured by RNA-Seq (FDR < 0.05). **(B)** Hallmark gene sets enriched in genes upregulated by CC allele. Gene sets were ranked by normalized enrichment score with FDR < 0.01. **(C)** Gene Set Enrichment Analysis (GSEA) of CC upregulated genes in Androgen response, mTORC1 signaling, G2M checkpoint, hypoxia, and TGFβ signaling from Hallmark gene sets. **(D–F)** CC genotype gene signature score based on z-score sum of the 205 differentially expressed genes targeted by CC alleles revealed strong linear positive correlation with AR signaling score in SU2C-PCF, Nat Med, and TCGA cohorts. **(G–I)** Scatter plots displaying significant positive correlation between CC genotype signature score and cell cycle progression (CCP) score in SU2C-PCF, Nat Med, and SU2C-PCF PNAS cohorts. **(J, K)** Pearson correlation examination demonstrated that there was positive linear correlation between CC genotype signature score and hypoxia score in the TCGA cohort. **(L)** CC upregulated genes enriched in epithelial-mesenchymal transition (EMT) in colorectal cancer from WikiPathways. **(M–O)** CC genotype signature score showed strong linear positive correlation with EMT score in SU2C-PCF, Nat Med, and TCGA cohorts, respectively. Genes were ranked based on the statistics “stat” of DESeq2 result between CC and TT samples.

## Synergistic Co-Overexpression of *TERT*, *E2F1*, and *MYC* Was Associated With Escalated Tumor Malignancy

Although *MYC* and *E2F1* are two well-studied oncogenic transcription factors that are frequently dysregulated in PCa cells (47–49), whether the two genes together with *TERT* have

synergistic effects on tumor severity are still unclear. To examine the synergistic effect of these three tumor oncogenic genes on patient survival time, we performed Kaplan–Meier analysis in several independent PCa cohorts. We found that patients with simultaneous triple high expression of *TERT*, *E2F1*, and *MYC* were significantly associated with shorter overall survival,

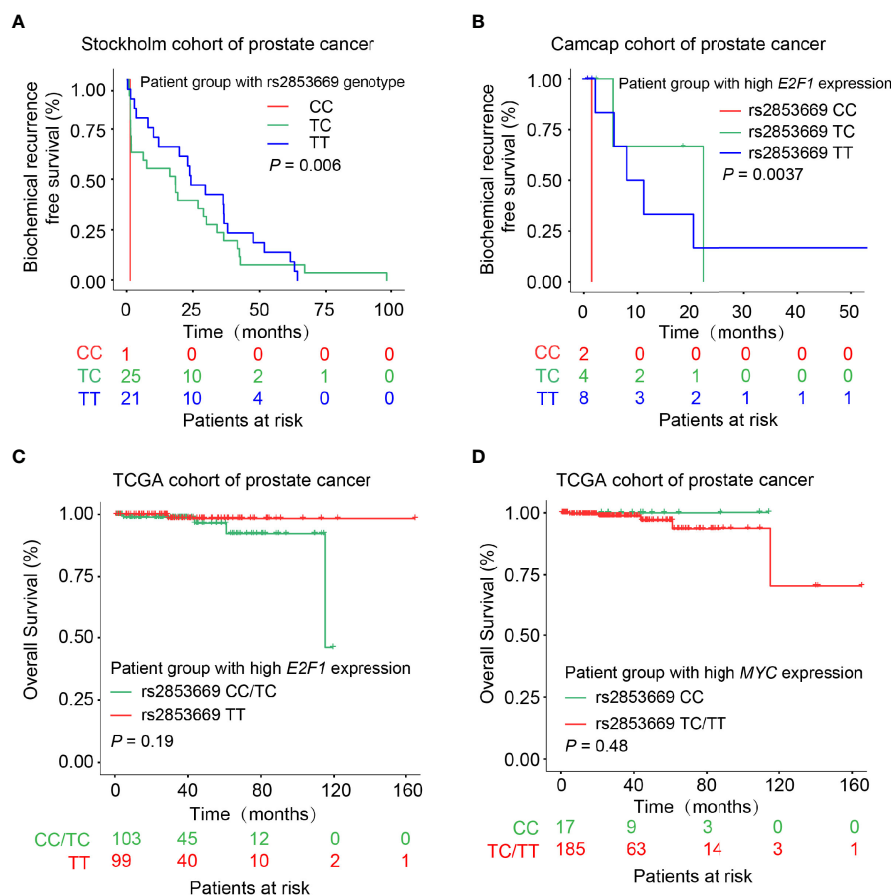




**FIGURE 6 |** Synergistic co-overexpression of *TERT*, *E2F1*, and *MYC* correlated with escalated tumor malignancy. (A–D) Synergistic triple high expression of *TERT*, *E2F1*, and *MYC* was associated with poorer overall survival (A), higher risks for biochemical relapse (B, C), and poorer metastatic-free survival (D) in PCa patients. (E–K) Proportion of PCa patients with biochemical relapse (E), advanced tumor stage (F), bone metastasis (G), higher PSA levels (H), and higher Gleason Score (I–K) was significantly higher in the group with triple high (+) expression of *TERT*, *E2F1*, and *MYC*. (L) Multivariate Cox regression analysis examined the synergistic effect of co-overexpression of *TERT*, *E2F1*, and *MYC* together with other clinicopathological features on the risk of BCR-free survival in PCa patients in TCGA cohort.

elevated risks of biochemical relapse, and metastasis. In contrast, the PCa patient group with simultaneous triple low expression appeared with better prognosis (Figures 6A–D and Supplementary Figures S6A, B). Notably, this observation was not found in normal prostates of the CPGEA cohort (Supplementary Figure S6C). To further investigate the synergistic effects of the three genes in clinical settings, we examined its correlation with several clinical features including

biochemical recurrence, tumor stage, metastasis, PSA, and patient neoplasm status in PCa cases. The results revealed that percentages of patients with biochemical relapse were significantly higher in groups with triple high expression of the three oncogenes (Figure 6E and Supplementary Figure S6D). The percentages of tumor stage III and IV were 62.2% and 79.4% in the low or high expression groups, respectively ( $p = 0.023$ ) (Figure 6F). Patients carrying tumors were also substantially



**FIGURE 7 |** Effect of rs2853669 genotype on PCa patient prognosis. **(A, B)** Kaplan–Meier plots demonstrating the biochemical recurrence-free survival of PCa patients grouped by the genotype of rs2853669. Patients carrying C allele of rs2853669 correlated with increased risk for biochemical recurrence in the Stockholm cohort **(A)**. Patient group with high *E2F1* expression and carrying rs2853669 CC genotype underwent a higher biochemical recurrence risk **(B)**. **(C, D)** PCa patient group carrying rs2853669 C allele with higher *E2F1* expression tumors or T allele with higher *MYC* expression levels indicated an increased risk of overall survival. *p*-values were examined by a log-rank test.

more in the triple high group (Supplementary Figures S6E, F). Proportions of patients with metastasis or PSA higher than 20 ng/ml were notably elevated in triple high group (Figures 6G, H and Supplementary Figure S6G). Moreover, the percentage of patients with advanced Gleason score was significantly higher in the triple high group in multiple cohorts (Figures 6I–K). These findings suggested that synergistic triple high expression of *E2F1*, *MYC*, and *TERT* were greatly associated with PCa severity and poorer prognosis. We next performed the multivariate Cox regression model to investigate the incorporated gene signature of *TERT*, *E2F1*, and *MYC* together with clinical prognostic factors influencing survival in PCa patients. The results revealed that the performance of the risk of the triple overexpression of the gene signature had the highest hazard ratio (Figure 6L), indicating a superior prognostic value in PCa compared to other clinicopathological features. Taken together, these findings indicated that the synergistic co-overexpression of *TERT*, *E2F1*, and *MYC* in PCa

is strongly associated with poor prognosis, tumor progression, metastasis, and patient survival.

### Allele-Specific Impact of rs2853669 on PCa Survival

We next asked whether the genotype of rs2853669 impacts PCa prognosis and thus examined the correlation of the CC genotype at rs2853669 with clinical features in PCa patients. This analysis showed that patients carrying genotype CC or TC at rs2853669 have a shorter time for biochemical relapse than patients with the TT genotype in the Stockholm cohort (Figure 7A). We also observed a clear trend in the CPGEA Nature 2020 dataset showing C allele associating with poor prognosis (Supplementary Figure S7A). Considering that *E2F1* and *MYC* are essential factors driving the severity of various tumors (50, 51), we thus explored whether the correlation of rs2853669 genotype and clinical variables was affected by *MYC* or *E2F1* expression status. We observed that the CC genotype at

rs2853669 was correlated with increased risk of biochemical relapse and overall survival rate in the PCa patient group with higher *E2F1* expression (**Figures 7B, C**). We additionally tested the association between the SNP and patient survival with the consideration of *E2F1* expression levels in Stockholm and CPGA Nature 2020 datasets shown in **Supplementary Figures S7B, C**. The results from both datasets showed that C allele is associated with poor prognosis. We observed a similar trend between the TT genotype at rs2853669, and the overall rate was observed in the patients with high expression of *MYC* (**Figure 7D**). These results further proved that *E2F1* and *MYC* drive the progression of PCa through different alleles at rs2853669.

## DISCUSSION

Over the past 15 years, GWASs have successfully identified many pleiotropic loci associated with PCa severity, despite the fact that a certain number of association results are inconsistent across different study cohorts. In this work, we also observed the opposite associations between rs2853669 and PCa in different populations. The C allele of rs2853669 was significantly associated with PCa severity observed in Caucasians from UK Biobank whereas T allele was found as a PCa risk allele in Chinese population.

We mechanistically determined that *MYC* and *E2F1* regulate *TERT* expression in PCa via a regulatory element that is disrupted by the T allele or C allele at rs2853669, respectively. The switch with T or C allele playing roles in regulating *TERT* expression was determined by AR signaling though other unknown factors may not be ruled out. These findings suggested that either T or C allele at rs2853669 can be the risk allele when cells are under certain molecular conditions. Disease severity, hormone levels, and biological contexts among the clinical samples might cause different observations regarding the association between rs2853669 and PCa risk. Importantly, while upregulation of *MYC* by *E2F1* was reported in other cancer types (52, 53), our findings deeply uncovered a critical role for *E2F1* in regulating *MYC* expression in PCa providing a new perspective for exploring the crosstalk between *E2F1* and *MYC* that function as two regulators in G1/S transition and tumor cell growth. Moreover, our study uncovered the critical role of C allele at rs2853669 in CRPC progression and EMT, dispensing new clues to unveil the molecular mechanisms underlying PCa development and progression. Both T and C alleles were essential for cell growth at the primary stage in PCa. However, C allele at rs2853669 showed higher risk association than that of T allele, implying that it plays a vital role in maintaining cell growth and enhancing invasion ability in hormone-free environment by recruiting *E2F1*. Future studies are required to discover whether there might be other factors affecting the regulatory mechanism of *MYC* and *E2F1* at this SNP considering the biological context of individuals are unique and complicated. The androgen response genes were upregulated in C/C cells, thereby explaining why C/C cells had hormone castration

resistant capacity in view of the fact that AR plays an important role in CRPC. In addition to androgen response genes, many growth and metastasis-associated genes were also upregulated in C/C cells, implying that the fragment harboring rs2853669 might be an enhancer element for other genes; alternatively, this SNP could change the three-dimensional structure of genome and may warrant further investigation.

Here, we proposed a novel model for studying germline variants in PCa; this model might be popular in other sex hormone relevant cancer types, such as breast cancer. For example, the roles of rs2853669 in breast cancer are still inclusive; one study reported an increased breast cancer risk in patients carrying the CC genotype (54), whereas other studies found that the CC genotype is not related to breast cancer risk (55). Our findings might provide more clues for unveiling the regulatory mechanisms underlying rs2853669 in breast cancer, and these inclusive results may be due to altered estrogen receptor signaling pathway, thereby facilitating us to fully understand the impact of the genetic variants on human diseases.

We reported here for the first time that the hormone level is one of the important factors in regulating the function of the variants. These results also proposed that there could be complicated crosstalk between one SNP and certain important signaling pathways in PCa. Complex factors such as ethnic groups, biological characters, genetic factors, and living surroundings between individuals brought more difficulties to explain the real biological functions of these loci. In future studies, detailed classification of tumor samples is essential because there might be more risk SNPs functioning like rs2853669 undiscovered. Moreover, more efficient genome editing tools need to be explored to convert the genotypes of more SNPs at one time to speed up the research of the synergies between two or more SNPs at the cellular level. Furthermore, the biological differences between individuals should also be considered when we implicate association studies to improve the detection, prognosis, and risk evaluation in PCa.

In summary, we interpreted the functional mechanisms of a 5p15 locus SNP rs2853669 in regulating *TERT* expression and PCa development, which may provide potential clues for improving PCa risk prediction and prognosis.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

PG, RN, and G-HW supervised the project. XD, JH, QX, BX, PZ, and HL performed the experiments with the help from

QH, RN performed cohort studies. TY provided advice on bioinformatics. QZ performed bioinformatics analysis. PG, RN, G-HW, XD, and QZ designed the studies and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by the National Natural Science Foundation of China (81972417 and 82073082), Jane and Aatos Erkko Foundation, Sigrid Juselius Foundation, Natural Science Foundation of Shaanxi Province (2020JM-292 and 2020JQ-430), Fudan University recruit funds, the “1000 Young Scholars” Program of Shaanxi Province, Fundamental Research Funds for the Central Universities (GK201902002

and GK201903061), and College Students’ Innovative Entrepreneurial Training Plan Program (S202010718091).

## ACKNOWLEDGMENTS

We thank the College of Life Sciences at Shaanxi Normal University for the sharing platform of laboratory apparatus and the support from Provincial Key Laboratory of animal cell and developmental biology of Shandong Province.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.754206/full#supplementary-material>

## REFERENCES

- Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* (2016) 315:68–76. doi: 10.1001/jama.2015.17703
- Farashi S, Kryza T, Clements J, Batra J. Post-GWAS in Prostate Cancer: From Genetic Association to Biological Contribution. *Nat Rev Cancer* (2019) 19:46–59. doi: 10.1038/s41568-018-0087-3
- Tomás-Loba A, Flores I, Fernández-Marcos PJ, Cayuela ML, Maraver A, Tejera A, et al. Telomerase Reverse Transcriptase Delays Aging in Cancer-Resistant Mice. *Cell* (2008) 135:609–22. doi: 10.1016/j.cell.2008.09.034
- Bryan TM, Cech TR. Telomerase and the Maintenance of Chromosome Ends. *Curr Opin Cell Biol* (1999) 11:318–24. doi: 10.1016/S0955-0674(99)80043-X
- Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
- Shay JW, Reddel RR, Wright WE. Cancer and Telomeres—an ALternative to Telomerase. *Science* (2012) 336:1388–90. doi: 10.1126/science.1222394
- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PLC, et al. Specific Association of Human Telomerase Activity With Immortal Cells and Cancer. *Science* (1994) 266:2011–5. doi: 10.1126/science.7605428
- Bodnar AG. Extension of Life-Span by Introduction of Telomerase Into Normal Human Cells. *Science* (1998) 279:349–52. doi: 10.1126/science.279.5349.349
- Morales CP, Holt SE, Ouellette M, Kaur KJ, Yan Y, Wilson KS, et al. Absence of Cancer-Associated Changes in Human Fibroblasts Immortalized With Telomerase. *Nat Genet* (1999) 21:115–8. doi: 10.1038/5063
- Weinrich SL, Pruzan R, Ma L, Ouellette M, Tesmer VM, Holt SE, et al. Reconstitution of Human Telomerase With the Template RNA Component hTR and the Catalytic Protein Subunit hTERT. *Nat Genet* (1997) 17:498–502. doi: 10.1038/ng1297-498
- Williams SCP. No End in Sight for Telomerase-Targeted Cancer Drugs. *Nat Med* (2013) 19:6–6. doi: 10.1038/nm0113-6
- Harley CB. Telomerase and Cancer Therapeutics. *Nat Rev Cancer* (2008) 8:167–79. doi: 10.1038/nrc2275
- Graham MK, Meeker A. Telomeres and Telomerase in Prostate Cancer Development and Therapy. *Nat Rev Urol* (2017) 14:607–19. doi: 10.1038/nrurol.2017.104
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank Resource With Deep Phenotyping and Genomic Data. *Nature* (2018) 562:203–9. doi: 10.1038/s41586-018-0579-z
- Machiela MJ, Chanock SJ. LDlink: A Web-Based Application for Exploring Population-Specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants: Fig. 1. *Bioinformatics* (2015) 31:3555–7. doi: 10.1093/bioinformatics/btv402
- Whittington T, Gao P, Song W, Ross-Adams H, Lamb AD, Yang Y, et al. Gene Regulatory Mechanisms Underpinning Prostate Cancer Susceptibility. *Nat Genet* (2016) 48:387–97. doi: 10.1038/ng.3523
- Gao P, Xia J-H, Sipeky C, Dong X-M, Zhang Q, Yang Y, et al. Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* (2018) 174:576–589.e18. doi: 10.1016/j.cell.2018.06.003
- The 1000 Genomes Project Consortium. A Global Reference for Human Genetic Variation. *Nature* (2015) 526:68–74. doi: 10.1038/nature15393
- Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PloS Genet* (2009) 5:e1000529. doi: 10.1371/journal.pgen.1000529
- Wigginton JE, Cutler DJ, Abecasis GR. A Note on Exact Tests of Hardy-Weinberg Equilibrium. *Am J Hum Genet* (2005) 76:887–93. doi: 10.1086/429864
- Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* (2010). Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuc: A MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data. *Gigascience* (2018) 7:1–6. doi: 10.1093/gigascience/gix120
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
- Love M, Huber W, Anders S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With Deseq2. *Genome Biol* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *PNAS* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
- Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J Integr Biol* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Yu G. *Enrichplot: Visualization of Functional Enrichment Result. R Package Version 1.8.1* (2020). Available at: <https://github.com/GuangchuangYu/enrichplot>.
- RStudio Team. *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. (2019). Available at: <http://www.rstudio.com/>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (2017). Available at: <https://www.R-project.org/>.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The Cbio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095



31. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the Cbioportal. *Sci Signaling* (2013) 6:p11–1. doi: 10.1126/scisignal.2004088
32. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* (2004) 6:1–6. doi: 10.1016/S1476-5586(04)80047-2
33. Edgar R. Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res* (2002) 30:207–10. doi: 10.1093/nar/30.1.207
34. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res* (2012) 41:D991–5. doi: 10.1093/nar/gks1193
35. Therneau T. A Package for Survival Analysis in R. *R Package Version 3.2-3* (2020). Available at: <https://CRAN.R-project.org/package=survival>.
36. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer (2000), ISBN: .
37. Kassambara A, Kosinski M, Biecek P. *Survminer: Drawing Survival Curves Using 'Ggplot2'*. *R Package Version 0.4.7* (2020). Available at: <https://CRAN.R-project.org/package=survminer>.
38. Cox DR. Regression Models and Life Tables. *J R Stat Soc* (1972) 34:187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x
39. Xu J, Mo Z, Ye D, Wang M, Liu F, Jin G, et al. Genome-Wide Association Study in Chinese Men Identifies Two New Prostate Cancer Risk Loci at 9q31.2 and 19q13.4. *Nat Genet* (2012) 44:1231–5. doi: 10.1038/ng.2424
40. Zuo C, Shin S, Keleş S. atSNP: Transcription Factor Binding Affinity Testing for Regulatory SNP Detection. *Bioinformatics* (2015) 31:3353–5. doi: 10.1093/bioinformatics/btv328
41. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell* (2010) 18:11–22. doi: 10.1016/j.ccr.2010.05.026
42. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell* (2018) 34:996–1011.e8. doi: 10.1016/j.ccell.2018.10.016
43. Bernard D, Pourtier-Manzanedo A, Gil J, Beach DH. Myc Confers Androgen-Independent Prostate Cancer Cell Growth. *J Clin Invest* (2003) 112:1724. doi: 10.1172/JCI200319035
44. Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, et al. Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* (2015) 161:1215–28. doi: 10.1016/j.cell.2015.05.001
45. Ramos-Montoya A, Lamb AD, Russell R, Carroll T, Jurmeister S, Galeano-Dalmau N, et al. HES6 Drives a Critical AR Transcriptional Programme to Induce Castration-Resistant Prostate Cancer Through Activation of an E2F1-Mediated Cell Cycle Network. *EMBO Mol Med* (2014) 6:651–61. doi: 10.1002/emmm.201303581
46. Altintas DM, Shukla MS, Goutte-Gattat D, Angelov D, Rouault JP, Dimitrov S, et al. Direct Cooperation Between Androgen Receptor and E2F1 Reveals a Common Regulation Mechanism for Androgen-Responsive Genes in Prostate Cells. *Mol Endocrinol* (2012) 26:1531–41. doi: 10.1210/me.2012-1016
47. Rodriguez-Bravo V, Pippa R, Song W-M, Carceles-Cordon M, Dominguez-Andres A, Fujiwara N, et al. Nuclear Pores Promote Lethal Prostate Cancer by Increasing POM121-Driven E2F1, MYC, and AR Nuclear Import. *Cell* (2018) 174:1200–15. doi: 10.1016/j.cell.2018.07.015
48. Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R, et al. Myc-Driven Murine Prostate Cancer Shares Molecular Features With Human Prostate Tumors. *Cancer Cell* (2003) 4:223–38. doi: 10.1016/S1535-6108(03)00197-1
49. McNair C, Xu K, Mandigo AC, Benelli M, Leiby B, Rodrigues D, et al. Differential Impact of RB Status on E2F1 Reprogramming in Human Cancer. *J Clin Invest* (2018) 128:341–58. doi: 10.1172/JCI93566
50. Cho H, Herzka T, Zheng W, Qi J, Wilkinson JE, Bradner JE, et al. RapidCaP, A Novel GEM Model for Metastatic Prostate Cancer Analysis and Therapy, Reveals Myc as a Driver of Pten-Mutant Metastasis. *Cancer Discovery* (2014) 4:318–33. doi: 10.1158/2159-8290.CD-13-0346
51. Ma L, Young J, Prabhala H, Pan E, Mestdagh P, Muth D, et al. miR-9, a MYC/MYCN-Activated microRNA, Regulates E-Cadherin and Cancer Metastasis. *Nat Cell Biol* (2010) 12:247–56. doi: 10.1038/ncb2024
52. Santoni-Rugiu E, Duro D, Farkas T, Mathiasen IS, Jäättelä M, Bartek J, et al. E2F Activity is Essential for Survival of Myc-Overexpressing Human Cancer Cells. *Oncogene* (2002) 21:6498–509. doi: 10.1038/sj.onc.1205828
53. Malekinejad H, Fani M, Shafiee-Roodbari SK, Delkhosh-Kasmaie F, Rezaei-Golmishheh A. Crosstalk Between E2f1 and C-Myc Mediates Hepato-Protective Effect of Royal Jelly on Taxol-Induced Damages. *Hum Exp Toxicol* (2017) 36:626–37. doi: 10.1177/0960327116660752
54. NBCS Collaborators, ABCTB Investigators, ConFab/AOCS, Michailidou K, Lindström S, Dennis J. Association Analysis Identifies 65 New Breast Cancer Risk Loci. *Nature* (2017) 551:92–4. doi: 10.1038/nature24284
55. Shen J, Gammon MD, Wu H-C, Terry MB, Wang Q, Bradshaw PT, et al. Multiple Genetic Variants in Telomere Pathway Genes and Breast Cancer Risk. *Cancer Epidemiol Biomarkers Prev* (2010) 19:219–28. doi: 10.1158/1055-9965.EPI-09-0771

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Dong, Zhang, Hao, Xie, Xu, Zhang, Lu, Huang, Yang, Wei, Na and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





## OPEN ACCESS

## Edited by:

Jonathan Olivier,  
Université de Lille,  
France

## Reviewed by:

Francesco Del Giudice,  
Sapienza University of Rome, Italy  
Amir Baghdadi,  
University of Calgary, Canada

## \*Correspondence:

Jianhui Chen  
chenjianhui1983@qq.com  
Lili Wang  
751501231@qq.com

## †ORCID:

Teng Zuo  
orcid.org/0000-0001-7920-1259  
Yanhua Zheng  
orcid.org/0000-0003-1103-957  
Jianhui Chen  
orcid.org/0000-0001-7494-1258  
Lili Wang  
orcid.org/0000-0001-6936-9402

## Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

Received: 24 July 2021

Accepted: 29 October 2021

Published: 18 November 2021

## Citation:

Zuo T, Zheng Y, He L, Chen T,  
Zheng B, Zheng S, You J, Li X, Liu R,  
Bai J, Si S, Wang Y, Zhang S, Wang L  
and Chen J (2021) Automated  
Classification of Papillary Renal Cell  
Carcinoma and Chromophobe Renal  
Cell Carcinoma Based on a Small  
Computed Tomography Imaging  
Dataset Using Deep Learning.  
Front. Oncol. 11:746750.  
doi: 10.3389/fonc.2021.746750

# Automated Classification of Papillary Renal Cell Carcinoma and Chromophobe Renal Cell Carcinoma Based on a Small Computed Tomography Imaging Dataset Using Deep Learning

Teng Zuo<sup>1†</sup>, Yanhua Zheng<sup>1†</sup>, Lingfeng He<sup>2</sup>, Tao Chen<sup>3</sup>, Bin Zheng<sup>4</sup>, Song Zheng<sup>1</sup>, Jinghang You<sup>5</sup>, Xiaoyan Li<sup>6</sup>, Rong Liu<sup>1</sup>, Junjie Bai<sup>1</sup>, Shuxin Si<sup>1</sup>, Yingying Wang<sup>7</sup>, Shuyi Zhang<sup>8</sup>, Lili Wang<sup>4\*†</sup> and Jianhui Chen<sup>1\*†</sup>

<sup>1</sup> Department of Urology, Fujian Medical University Union Hospital, Fuzhou, China, <sup>2</sup> Institute for Empirical Social Science Research, Xi'an Jiaotong University, Xi'an, China, <sup>3</sup> School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China, <sup>4</sup> School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, United States, <sup>5</sup> Department of Radiology, Fujian Medical University Union Hospital, Fuzhou, China, <sup>6</sup> Department of Pathology, Fujian Medical University Union Hospital, Fuzhou, China, <sup>7</sup> School of Medicine, Fujian Medical University, Fuzhou, China, <sup>8</sup> School of Medicine, Xiamen University, Xiamen, China

**Objectives:** This study was conducted in order to design and develop a framework utilizing deep learning (DL) to differentiate papillary renal cell carcinoma (PRCC) from chromophobe renal cell carcinoma (ChRCC) using convolutional neural networks (CNNs) on a small set of computed tomography (CT) images and provide a feasible method that can be applied to light devices.

**Methods:** Training and validation datasets were established based on radiological, clinical, and pathological data exported from the radiology, urology, and pathology departments. As the gold standard, reports were reviewed to determine the pathological subtype. Six CNN-based models were trained and validated to differentiate the two subtypes. A special test dataset generated with six new cases and four cases from The Cancer Imaging Archive (TCIA) was applied to validate the efficiency of the best model and of the manual processing by abdominal radiologists. Objective evaluation indexes [accuracy, sensitivity, specificity, receiver operating characteristic (ROC) curve, and area under the curve (AUC)] were calculated to assess model performance.

**Results:** The CT image sequences of 70 patients were segmented and validated by two experienced abdominal radiologists. The best model achieved 96.8640% accuracy (99.3794% sensitivity and 94.0271% specificity) in the validation set and 100% (case accuracy) and 93.3333% (image accuracy) in the test set. The manual classification achieved 85% accuracy (100% sensitivity and 70% specificity) in the test set.

**Conclusions:** This framework demonstrates that DL models could help reliably predict the subtypes of PRCC and ChRCC.

**Keywords:** CNN—convolutional neural network, PRCC, papillary renal cell carcinoma, ChRCC, chromophobe-primary renal cell carcinoma, cancer image classification

## INTRODUCTION

With the continuous advancement of imaging technology and increasing awareness of the public for early cancer screening, the detection rate of renal masses is increasing (1). In China, most renal masses are kidney cancer. The incidence of kidney cancer in the Chinese population continues to increase (2). Existing methods can meet the need to distinguish clear cell carcinoma from non-clear cell carcinoma. However, the differentiation between subtypes of non-clear carcinoma may be difficult because of the lack of a quantitative evaluation of images, especially from the early-stage cancers, which usually present atypically (3). Papillary renal cell carcinoma (PRCC) and chromophobe renal cell carcinoma (ChRCC) are the most common types of non-clear cell carcinoma and are characterized by a unique molecular morphology (4). PRCC is associated with activating germline mutations in MET (type I) and activation of the NRF2–ARE pathway (type II) (5). Typical genetic changes in ChRCC are deletions of chromosomes Y, 1, 2, 6, 10, 13, 17, and 21 (6). The differences in originating factors and driver genes between the two subtypes lead to different treatment options and prognoses (7, 8). There is some differentiation between PRCC and ChRCC in imaging findings: PRCC presents as cysts, necrosis, and calcification, while ChRCC presents as central wheel-shape enhancement (9). In low stage or small size masses, however, these characteristics mentioned above are atypical, which usually cause a difficult diagnosis. In addition, according to previous reports (10), the accuracy and sensitivity of the manual classification of PRCC/ChRCC are 61.8% and 84.5%, respectively, which cannot meet this need. Therefore, in the clinic, it is difficult to provide a highly accurate manual subtype differentiation between PRCC and ChRCC, and this remains to be a challenge.

Recently, with the rapid development of computer hardware and deep learning (DL) theory, artificial intelligence (AI) has been widely applied in radiological image processing for classification and is rapidly developing (11). Notably, the efficacy of DL-based models for the radiological diagnosis of several tumors [e.g., breast cancer (12), liver cancer (13), and lung masses (14)] is superior to that of manual processing according to previous studies (15). Convolutional neural networks (CNNs) and improved models have been widely used for medical image processing (16). DL-based oncological radiological characterization has shown value in medical fields (11, 15, 16). CNNs and their improved models are currently one of the hot spots in the field of medical image processing. Image processing based on this type of model for assisting in renal tumor examinations has achieved promising results and suggests the possibility of solving the challenges associated with the radiological differentiation of PRCC and ChRCC.

In this study, DL was utilized to classify PRCC and ChRCC from computed tomography (CT) datasets. The current study aimed to exploit DL-based models for renal cell carcinoma subtype classification based on small datasets so that the classification can be implemented in some scenarios without high-performance hardware or shortage of rare subtypes cases, to better promote the accuracy of radiological diagnosis.

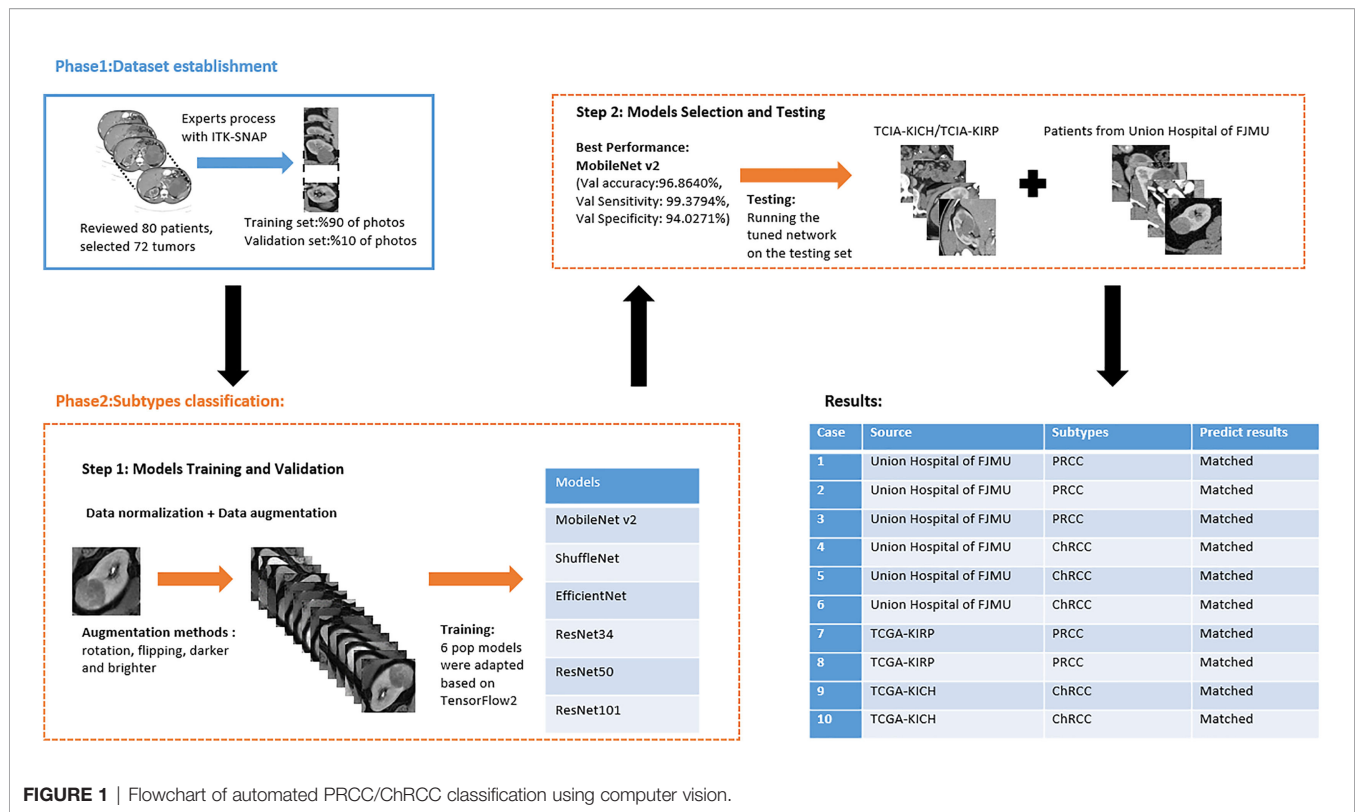
## METHODS

Institutional review board approval was obtained. The requirement for written informed patient consent was waived. A retrospective review of PRCC and ChRCC patients at Fujian Medical University Union Hospital was performed between 2012 and 2021. Ethical approval was obtained from the Institutional Ethics Committee of Fujian Medical University Union Hospital (No. 2021WSJK033). According to the Helsinki Declaration, all patients (or their legal clients) provided written informed consent before obtaining their clinical, radiological, and pathological data. The framework used to develop an automated method for the differentiation of these two subtypes was comprised of two phases (**Figure 1**): 1) CT scan data, clinical data, and pathological data were gathered and digitized, followed by tumor lesion segmentation and labeling by experts in the radiology department (dataset establishment); and 2) training neural networks; assessing the accuracy, sensitivity, and specificity of the models; and verifying model efficiency through comparison with the pathological diagnosis of new cases (subtype classification).

### Phase 1

#### Dataset Establishment

Patients with a pathological diagnosis obtained by biopsy or surgical resection were included in this study. In addition, 80 patients with available arterial/cortical/nephrogenic phase CT image sequences were reviewed (42 with PRCC and 38 with ChRCC). After randomly selecting 6 cases (3 PRCCs and 3 ChRCCs) for testing sets, the images of 74 tumors (39 PRCCs and 35 ChRCCs) were used to build the datasets. The CT images were obtained using various radiology scanners and non-standard protocols. Arterial phase sequences were preferred when multiple phases existed. Whole sequences were retrieved and exported utilizing the hospital radiological database. The window settings were 40 HU (width) and 400 HU (level). Based on the clinical and pathological data, ROIs of sequences were segmented, labeled, and exported with ITK-SNAP by two abdominal radiologists who have experience of more than 10 years in the diagnosis of urinary system tumor. After cross-validation, images that were exported in .jpg size included 857 images of ChRCCs and 997 images of PRCCs. Labeling was



applied in the non-graphical layer so that each slice filename contained the case number, gender, age, and histological subtypes. After resizing, images comprised matrices with 256 \* 256 pixels in the axial planes. The dataset was divided into the training set and validation set (90% for the training set and 10% for the validation set).

## Phase 2 Subtype Classification

### Model Training and Validation

Six pop models [MobileNetV2 (17), EfficientNet (18), ShuffleNet (19), ResNet-34, ResNet-50, and ResNet-101 (20)] were adapted for dichotomy based on TensorFlow 2.4.12. Preprocessing involved normalization and augmentation (including Gaussian blur, rotation, flipping, brighter, and darker) (**Figure 2**). In addition to data augmentation, ConvBNReLU (convolution + batch normalization + ReLU) was applied to avoid overfitting. The learning rate was initially set as 0.005 and was optimized by the Adaptive Moment Estimation (ADAM) optimization algorithm in every training phase. The batch size was set as 24. For model training, a desktop workstation with an Intel® Xeon® E5-2678 v3 CPU and an NVIDIA GeForce RTX 2080Ti (11 GB) GPU was used. A list of model parameters, training results, and validation/accuracy results is provided in **Table 1**.

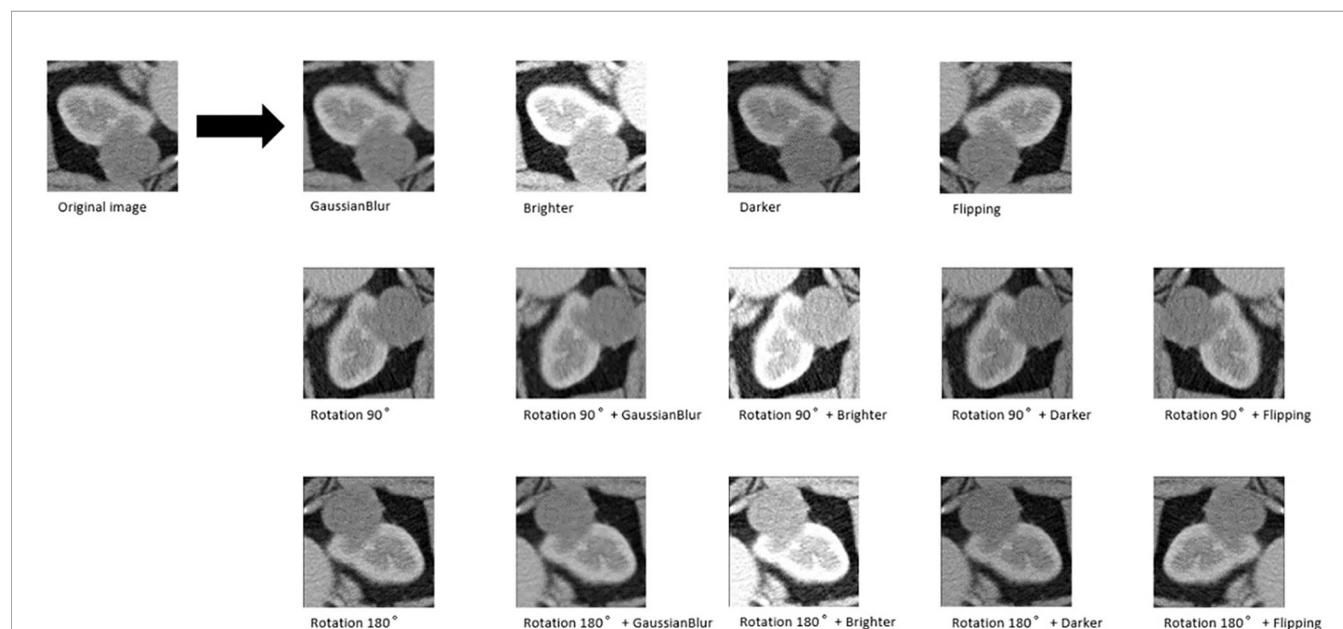
### Model Selection and Testing

Based on the results of the training step, MobileNetV2, ShuffleNet, and ResNet-34 were selected as the testing models. A special test set of PRCC/ChRCC samples was established in

two parts (**Table 2**): 1) reviewing the new cases in 2021, including six patients (three with PRCC, three with ChRCC); and 2) reviewing cases in The Cancer Imaging Archive (TCIA) datasets, including four patients (two with PRCC from the TCGA-KIRP dataset, two with ChRCC from the TCGA-KICH dataset). Slices were processed by abdominal radiologists, and for each case, three photographs were selected randomly. To assess efficiency from different views, two accuracy values were calculated. 1) Case accuracy: if correctly identified photographs were >2, this case was regarded as correctly identified. Case accuracy was used to reflect the percentage of correct cases. 2) Sample accuracy: this was used to show the proportion of correct images among all images. The accuracy, sensitivity, and specificity of these models were computed. In order to show the efficiency of manual processing, two radiologists were invited to distinguish these cases. Objective measure indexes of manual prediction were also calculated.

## RESULTS

The model based on MobileNetV2 (**Table 3** and **Figure 3**) performed best for tumor subtype diagnosis. The automated method achieved 96.8640% accuracy in the validation dataset (99.3794% sensitivity, 94.0271% specificity). Due to all correctly matching, case accuracy, case sensitivity, and case specificity were all achieved 100%. For every single photograph, image accuracy achieved 93.3333% in the testing dataset (88.2353% sensitivity and 86.6667% specificity). The AUC was 0.9489, and the *p*-value was



**FIGURE 2** | An example of data augmentation processing. Based on the geometric transformations (rotation and flipping), the Gaussian blur, brighter, and darker were applied, which finally achieved 15× amplification.

less than 0.001. Resource occupancy was less while training and predicting (less than 10 GB of accelerated graphics memory occupied), which means that this model can be applied to low-performance hardware. The manual method achieved 85% accuracy (100% sensitivity, 70% specificity) in the testing dataset. The results are provided in **Table 2** and **Figure 4**.

## DISCUSSION

Before a clinical treatment strategy is developed, the gold standard for the differentiation of subtypes is pathological

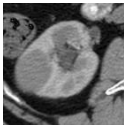
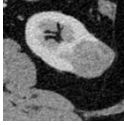

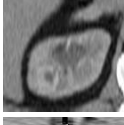
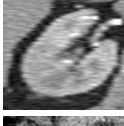
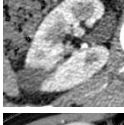
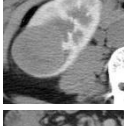
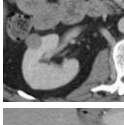
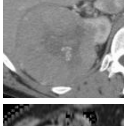
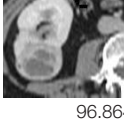
diagnosis by histological biopsy. Nevertheless, this invasive operation may increase the possibility of needle tract implantation and the metastasis of malignant tumors, as well as the risks of bleeding, infection, and damage to surrounding organs caused by puncture operations. Furthermore, the missed diagnosis rate is approximately 30% (21). An ideal renal tumor diagnosis method should avoid unnecessary damage to patients and potential risks as much as possible while ensuring high accuracy and a high detection rate, which points to the need to further improve auxiliary examination image processing technology to increase sensitivity and accuracy, as it has great prospects.

**TABLE 1** | The results of CNN-based networks for classification task training and validation and the testing results of the models.

Models	Parameters	Best validation accuracy	Testing results (case)
MobileNetV2	Total: 2,261,827 Trainable: 2,226,434	96.8640%	Accuracy: 100% Sensitivity: 100% Specificity: 100%
ShuffleNet	Total: 1,272,859 Trainable: 1,256,679	97.3074%	Accuracy: 83.3334% Sensitivity: 92.3077% Specificity: 72.7273%
EfficientNet	Total: 4,053,414 Trainable: 4,011,391	Cannot converge	NA
ResNet-34	Total: 21,829,058 Trainable: 21,812,034	93.6404%	Accuracy: 91.6667% Sensitivity: 84.6154% Specificity: 100%
ResNet-50	Total: 25,662,403 Trainable: 25,609,283	Cannot converge	NA
ResNet-101	Total: 44,706,755 Trainable: 44,601,411	Cannot converge	NA

NA, not available.

**TABLE 2 |** Information of test sets, comparison result of automated model prediction, and the result of model performance in the validation dataset.

Case	Source	Subtypes	Gender	Age	Sample	Automated prediction	Manual prediction
1	Union Hospital of FJMU	PRCC	Female	60		Matched	1. Matched 2. Matched
2	Union Hospital of FJMU	PRCC	Male	58		Matched	1. Matched 2. Matched
3	Union Hospital of FJMU	PRCC	Male	57		Matched	1. Matched 2. Matched
4	Union Hospital of FJMU	ChRCC	Male	62		Matched	1. Matched 2. Matched
5	Union Hospital of FJMU	ChRCC	Female	41		Matched	1. Matched 2. Mismatched
6	Union Hospital of FJMU	ChRCC	Female	62		Matched	1. Matched 2. Matched
7	TCGA-KIRP	PRCC	—	—		Matched	1. Matched 2. Matched
8	TCGA-KIRP	PRCC	—	—		Matched	1. Matched 2. Matched
9	TCGA-KICH	ChRCC	—	—		Matched	1. Matched 2. Mismatched
10	TCGA-KICH	ChRCC	—	—		Matched	1. Mismatched 2. Matched
Validation accuracy					96.8640%		
Validation sensitivity					99.3794%		
Validation specificity					94.0271%		
Test accuracy (case)					100%		
Test sensitivity (case)					100%		
Test specificity (case)					100%		
Test accuracy (image)					93.3333%		
Test sensitivity (image)					88.2353%		
Test specificity (image)					86.6667%		
Manual accuracy					85% (90% and 80%)		
Manual sensitivity					100%		
Manual specificity					70% (80% and 60%)		



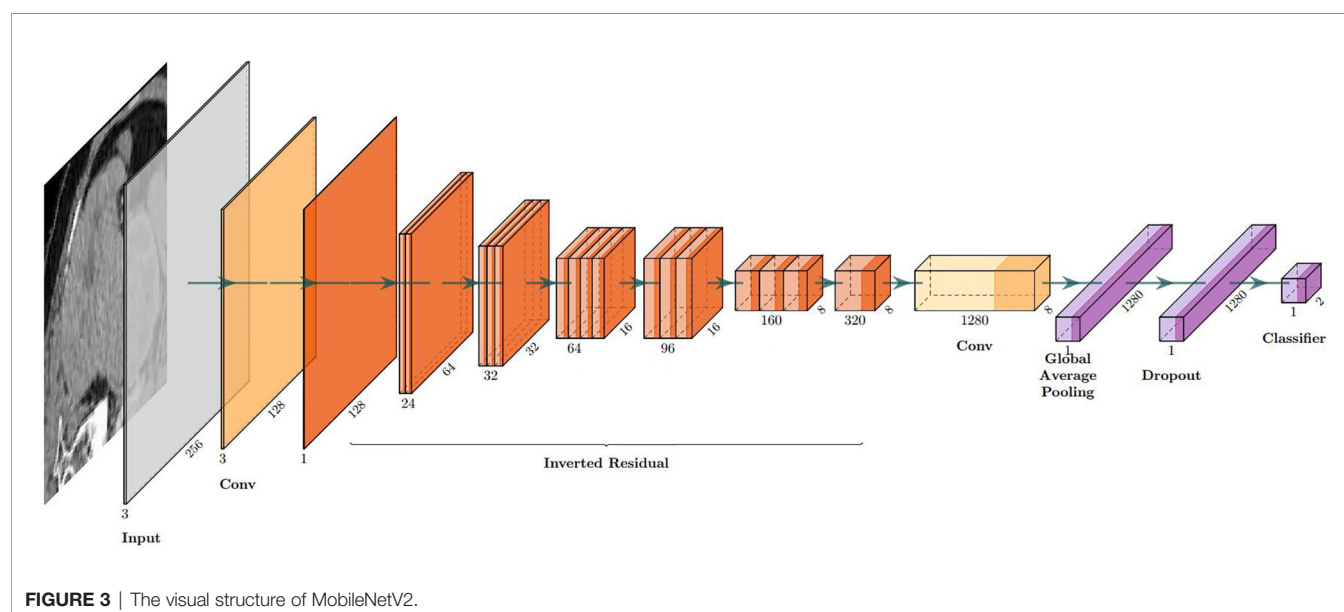
**TABLE 3 |** The structure of MobileNetV2.

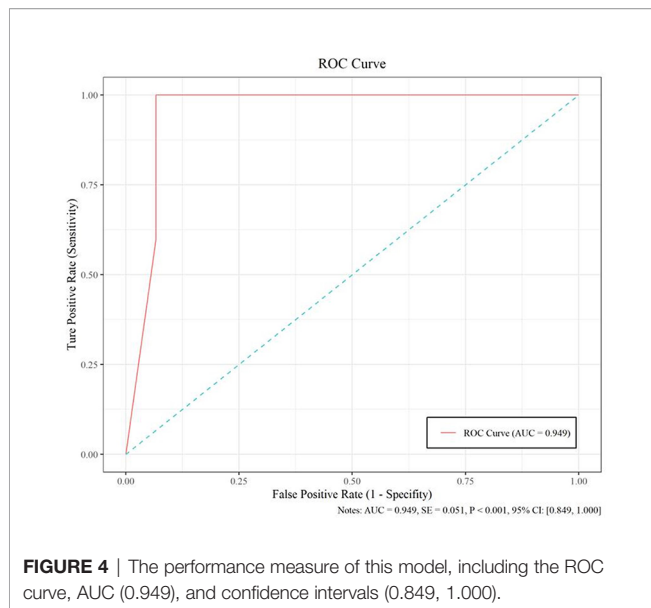
Layer (functions)	Output shape	Stride	Filter shape
Input layer	None, 256, 256, 3	/	/
Conv1 (Conv+BN+ReLU6)	None, 128, 128, 32	2	3 * 3 * 32
Inverted_residual (linear)	None, 128, 128, 16	1	1 * 1 * 32 * 16
Inverted_residual_1 (ReLU6)	None, 64, 64, 24	2	3 * 3 * 16 * 24
Inverted_residual_2 (linear)	None, 64, 64, 24	1	1 * 1 * 24
Inverted_residual_3 (ReLU6)	None, 32, 32, 32	2	3 * 3 * 24 * 32
Inverted_residual_4 (linear)	None, 32, 32, 32	1	1 * 1 * 32
Inverted_residual_5 (linear)	None, 32, 32, 32	1	1 * 1 * 32
Inverted_residual_6 (ReLU6)	None, 16, 16, 64	2	3 * 3 * 32 * 64
Inverted_residual_7 (linear)	None, 16, 16, 64	1	1 * 1 * 64
Inverted_residual_8 (linear)	None, 16, 16, 64	1	1 * 1 * 64
Inverted_residual_9 (linear)	None, 16, 16, 64	1	1 * 1 * 64
Inverted_residual_10 (linear)	None, 16, 16, 96	1	1 * 1 * 64 * 96
Inverted_residual_11 (linear)	None, 16, 16, 96	1	1 * 1 * 96
Inverted_residual_12 (linear)	None, 16, 16, 96	1	1 * 1 * 96
Inverted_residual_13 (ReLU6)	None, 8, 8, 160	2	3 * 3 * 96 * 160
Inverted_residual_14 (linear)	None, 8, 8, 160	1	1 * 1 * 160
Inverted_residual_15 (linear)	None, 8, 8, 160	1	1 * 1 * 160
Inverted_residual_16 (linear)	None, 8, 8, 320	1	1 * 1 * 160 * 320
Conv (ReLU6)	None, 8, 8, 1,280	1	1 * 1 * 320 * 1,280
Global average pooling	None, 1,280	1	Pool 8 * 8
Dropout	None, 1,280	1	Probability = 0.2
Classifier (ReLU)	None, 2	/	Classifier

The accuracy and sensitivity of a manual imaging diagnosis cannot meet current clinical diagnosis and treatment needs. In addition, there is still a lack of clinical and radiological features that can accurately predict histology. The current imaging diagnostic method has significant limitations. The accuracy and sensitivity of the manual classification of PRCC/ChRCC according to existing reports are 61.8% and 84.5% (10), which are significantly lower than those of our model. The average accuracy, sensitivity, and specificity of manual classification by our radiologists of these subtypes are 85% (90% and 80%), 100%, and 70% (80% and 60%), which are also lower than those of our model. Our MobileNet-based model also

showed better efficacy than manual processing. This result provides an automated approach to the dilemma of diagnosing subtypes with radiological data and may affect the selection of surgical methods and clinical decisions.

As a typical DL algorithm, due to their interlayer parameter sharing characteristics and sparse connection characteristics of the model architecture, CNNs can realize the automated extraction of pixel-level image features without the need to establish and engineer large-scale features in advance, and due to the real-time nature of the model itself, features such as flexibility, associative information storage, and backpropagation algorithm change weights can achieve





higher processing accuracy with manual data than traditional machine learning, prompting high-throughput automation based on the feasibility of CNN/DCNN models for imaging and omics analyses. The application of comprehensive digitized clinical data, radiological images, and pathological data has paved the way for automated processing methods based on AI for radiological data processing in the future. In recent years, various studies have started utilizing complete digital radiological and clinical data for segmentation and classification (22, 23), verifying the feasibility of this scheme. In nephrology oncology, study interests that incorporated AI started focusing on subtype classification. Tanaka et al. (24), based on the Inception-v3 CNN model and MR images, identified benign and malignant renal masses ( $\leq 4$  cm) on images with an accuracy rate of 88%. Based on a CNN model and CT images, Baghdadi et al. (25) identified benign renal oncocytoma and ChRCC on images with an accuracy rate of 95%. Zhou et al. applied transfer learning to classify benign and malignant kidney tumors with CT datasets, and the accuracy of the difference was reported to be 0.95 (26). Lee et al. developed a model that combined DL and manual feature machine learning to classify specific kidney tumor types, and the accuracy was 0.77 (27). These studies prove that the imaging differentiation of kidney tumors based on DL and dichotomy is feasible but lacks utility and requires high-performance hardware, limiting the research results to the clinic. The present models show the possibility of using a high-confidence DL-based diagnostic method for the radiological classification of PRCC/ChRCC and provide a feasible low-performance hardware program with high accuracy for different medical devices that can be applied even to a gaming laptop or a mobile workstation.

There were some new findings obtained during training and validation that have not been reported in research in the same field. First, according to the experimental results, we speculate that the valuable features of PRCC/ChRCC on CT images are commonly overlooked, which indicates that the fewer trainable parameters the model has, the better the accuracy it achieves. Although the feature

capacity of the models is correlated linearly with the number of parameters, the number of parameters is seemingly correlated with fitting situations in a parabola. In the lightest model, ShuffleNet, performance is the worst in these coverage models. The best performance is from the MobileNetV2, with a bigger capacity of parameters than ShuffleNet. However, as the number of parameters is continuously increasing, the accuracy is decreasing (ResNet-34). Importantly, too many trainable parameters in this task can cause model under-/overfitting (ResNet-50/ResNet-101). The relation between accuracy and parameters during the classification of small datasets needs further explanation and selection. However, this interesting finding shows the importance of feature capacity assessment of datasets and the right choice of models with a suitable size before promoting performances. Finally, we noticed that extreme data augmentation has little effect on training. We tried several ways to augment and amplify datasets to increase their size, which obviously did not affect the accuracy of the validation dataset.

This study had two main aims: 1) to make the automated classification methods easy to use with broad applicability to provide a highly accurate method that can be used in basic health units and deployed in medical centers with low-performance hardware and 2) to combine these methods with those used in federal studies, which can be used for multicenter studies and to increase model accuracy without the need to gather all the data. The deployment of such a model in devices at health centers will promote clinical treatment. Our future goal is to migrate this processing paradigm to other RCC subtypes. Although this study provides the first automated method for the radiological classification of PRCC/ChRCC subtypes, there are still some limitations. 1) The main limitation is the lack of multicenter validation, and the other limitations include the sizes of the training, validation, and testing datasets, which will be considered in future studies. Our methods used to avoid overfitting included data augmentation and ConvBNReLU but should have been more diverse. Also, due to the limited dataset capacity, the ROC curve was unsmoothed. 2) The underutilization of digital clinical data is another limitation. The conjunctive use of clinical and radiological data can further improve the prediction accuracy. 3) The underuse of multiphase sequences could be considered another limitation of our study. In our study, we exported images based on one patient-one phase. 4) Our dataset was mainly obtained from East Asian patients, and since a population-based analysis reported that racial disparities exist between black and white people in kidney cancer (28), the upshot of our study would have bias in the East Asian population. Multiple factors including race, gender, and age could be taken into consideration for further exploration. 5) Our validation dataset was based on a dataset from our hospital, but it was not the best choice and had a certain effect on the results of training. The ideal validation dataset should be based on three or more datasets from different institutes. The small size of the testing dataset could also have led to a controversy about the results of MobileNetV2, which need to be further tested in multiple centers. 6) Processing of datasets by experts may not be regarded as the best method. Ideally, an automated segmentation procedure contained in the pipeline may be a better choice. However, there were some barriers laid that could hardly be bypassed. We tried two proposals:

individual tumor output and area output. We applied U-Net-based methods to draw ROIs and found that the methods of existing reports (29–31) could not fit the need. In studies, tumor Dice scores were lower than 0.85 generally, which meant that some parts of tumor pixels could not be contained in the processed images and some radiological characters were lost unavoidably. The reason why we finally did not use this method was that this automated segmentation/classification-combined model had a performance lower than expected. Unless the method to improve the tumor Dice score obviously is developed, the automated segmentation–classification model efficiency has a rare possibility to reach the baseline of clinical application. We also tried YOLO-v3-based detection and area segmentation; however, it did not show better performances compared with existing ML-based methods, which finally led to its abandonment. Besides the technological challenges, the main reason why we did not introduce an automated segmentation into the pipeline was that in this study we focused more on the classification, and the key point was realizing the classification of subtypes with low capacity under a smaller feature engineering preprocessing and more automated processing compared with ML-based classification methods. As a challenge in DL-based radiomics, automated segmentation is our next study orientation. We are developing a possible method to realize our proposed DL-based radiological processing series, and we are also

trying to integrate several models into a DL-based radiomics workstation (Figure 5).

Overall, although there may be limitations in this study such as a small dataset and differences in races as well as in imaging single-center protocol, the research results may be biased to some extent. However, based on the result that a CNN-assisted diagnosis model with high diagnostic accuracy was developed in a single center of this study, it suggests that the AI research and development model adopted in this study has high clinical application potential in improving the accuracy of differential diagnosis of PRCC and ChRCC, at least in a single regional center. In the future, although there will be some difficult challenges in developing AI high diagnostic accuracy which was caused by some objective factors such as subtle potential differences in image feature led by the discrepancy between race and region and inability in high homogeneity in the imaging method, we still expect that the AI auxiliary renal tumor imaging diagnostic research can expand into different regions, different centers, and different races, together with bigger sample data to validate our conclusion, and can accurately classify as well as precisely and automatically segment multiple pathological types of renal tumor, with the aim of making it an auxiliary diagnostic imaging tool with wide clinical application prospects.



FIGURE 5 | A demo of a DL-based radiomics workstation (next study).

## CONCLUSIONS

To the best of our knowledge, this study provides the first automated framework for differentiating PRCC and ChRCC that could produce reliable results. This approach may be useful in improving the radiological diagnostic accuracy of RCC and, thus, benefit patients.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## REFERENCES

- Richard PO, Jewett MAS, Bhatt JR, Kachura JR, Evans AJ, Zlotta AR, et al. Renal Tumour Biopsy for Small Renal Masses: A Single-Center 13-Year Experience. *Eur Urol* (2015) 68(6):1007–13. doi: 10.1016/j.eururo.2015.04.004
- Yang Y, Xie L, Zheng JL, Tan YT, Zhang W, Xiang YB. Incidence Trends of Urinary Bladder and Kidney Cancers in Urban Shanghai, 1973-2005. *PLoS One* (2013) 8(12):e82430. doi: 10.1371/journal.pone.0082430
- Rossi SH, Prezzi D, Kelly-Morland C, Goh V. Imaging for the Diagnosis and Response Assessment of Renal Tumours. *World J Urol* (2018) 36:1927–42. doi: 10.1007/s00345-018-2342-3
- Li Y, Luo Q, Li Z, Wang Y, Zhu C, Li T, et al. Long Non-Coding RNA IRAIN Inhibits VEGFA Expression via Enhancing Its DNA Methylation Leading to Tumour Suppression in Renal Carcinoma. *Front Oncol* (2020) 10:1082. doi: 10.3389/fonc.2020.01082
- Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, Davis C, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* (2016) 374:135. doi: 10.1056/NEJMoa1505917
- Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur Urol* (2016) 70:93. doi: 10.1016/j.eururo.2016.02.029
- Turajlic S, Swanton C, Boshoff C. Kidney Cancer: The Next Decade. *J Exp Med* (2018) 215(10):2477–9. doi: 10.1084/jem.20181617
- Ljungberg B, Albiges L, Abu-Ghanem Y, Bensalah K, Dabestani S, Fernández-Pello S, et al. European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur Urol* (2019) 75(5):799–810. doi: 10.1016/j.eururo.2019.02.011
- Wang D, Huang X, Bai L, Zhang X, Wei J, Zhou J. Differential Diagnosis of Chromophobe Renal Cell Carcinoma and Papillary Renal Cell Carcinoma With Dual-Energy Spectral Computed Tomography. *Acta Radiol* (2020) 61(11):1562–9. doi: 10.1177/0284185120903447
- Sun XY, Feng QX, Xu X, Zhang J, Zhu F-P, Yang Y-H, et al. Radiologic-Radiomic Machine Learning Models for Differentiation of Benign and Malignant Solid Renal Masses: Comparison With Expert-Level Radiologists. *AJR Am J Roentgenol* (2020) 214(1):W44–54. doi: 10.2214/AJR.19.21617
- Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5
- Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. *Radiology* (2018) 290(1):52–8. doi: 10.1148/radiol.2018180694

## AUTHOR CONTRIBUTIONS

TZ and YZ: conceptualization, project administration, writing—original draft, and writing—review and editing. LH and TC: data curation, formal analysis, software, and visualization. BZ, JY, XL, RL, JB, SS, YW, SYZ, and SZ: investigation, methodology, and writing—review and editing. LW and JC: conceptualization, data curation, funding acquisition, investigation, project administration, supervision, validation, writing—original draft, and writing—review and editing. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by grants from the Natural Science Foundation of Fujian Province (2019J01153) and the Startup Fund for Scientific Research, Fujian Medical University (2019QH1053).

- Yasaka K, Akai H, Abe O, Kiryu S. Deep Learning With Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-Enhanced CT: A Preliminary Study. *Radiology* (2017) 286:887–96. doi: 10.1148/radiol.2017170706
- Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* (2017) 284:574–82. doi: 10.1148/radiol.2017162326
- Cruz-Roa A, Gilmore H, Basavanahally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and Reproducible Invasive Breast Cancer Detection in Whole-Slide Images: A Deep Learning Approach for Quantifying Tumour Extent. *Sci Rep* (2017) 7:46450. doi: 10.1038/srep46450
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* (2017).
- Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* (2019) 1905.11946.
- Ma N, Zhang X, Zheng H-T, Sun J. “ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design”. In: *European Conference on Computer Vision*. Cham: Springer (2018).
- He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *IEEE Conference on Computer Vision & Pattern Recognition IEEE Computer Society*.
- Soulen MC. Small Renal Masses: Challenges for the Radiologist. *Radiol* (2018) 288(1):91–2. doi: 10.1148/radiol.2018180602
- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. “3d U-Net: Learning Dense Volumetric Segmentation From Sparse Annotation”. In: *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*. In: *Lecture Notes in Computer Science*, vol. 9901. Cham: Springer (2016). p. 424–32. 1606.06650v1.
- Zhou X, Ito T, Takayama R, Wang S, Hara T, Fujita H. Threedimensional CT Image Segmentation by Combining 2D Fully Convolutional Network With 3D Majority Voting. In: *Proc Deep Learn Med Image Anal (DLMIA) In: Lecture Notes Comput Sci* (2016) 10 0 08:111–20. doi: 10.1007/978-3-319-46976-8\_12
- Tanaka T, Huang Y, Marukawa Y, Tsuboi Y, Masaoka Y, Kojima K, et al. Differentiation of Small ( $\leq 4$  Cm) Renal Masses on Multiphase Contrast-Enhanced CT by Deep Learning [Published Correction Appears in AJR Am J Roentgenol. *AJR Am J Roentgenol* (2020) 214(3):605–12. doi: 10.2214/AJR.19.22074

25. Baghdadi A, Aldhaam NA, Elsayed AS, Hussein AA, Cavuoto LA, Kauffman E, et al. Automated Differentiation of Benign Renal Oncocytoma and Chromophobe Renal Cell Carcinoma on Computed Tomography Using Deep Learning. *BJU Int* (2020) 125(4):553–60. doi: 10.1111/bju.14985
26. Zhou L, Zhang Z, Chen Y-C, Zhao Z-Y, Yin X-D, Jiang H-B. A Deep Learning-Based Radiomics Model for Differentiating Benign and Malignant Renal Tumours. *Translational Oncology* (2019) 12(2):292–300. doi: 10.1016/j.tranon.2018.10.012
27. Lee H, Hong H, Kim J, Jung DC. Deep Feature Classification of Angiomyolipoma Without Visible Fat and Renal Cell Carcinoma in Abdominal Contrast-Enhanced CT Images With Texture Image Patches and Hand-Crafted Feature Concatenation. *Med Phys* (2018) 45(4):1550–61. doi: 10.1002/mp.12828
28. Chung BI, Leow JJ, Gelpi-Hammerschmidt F, Wang Y, Del Giudice F, De S, et al. Racial Disparities in Postoperative Complications After Radical Nephrectomy: A Population-Based Analysis. *Urology* (2015) 85(6):1411–6. doi: 10.1016/j.urology.2015.03.001
29. Fabian I, Maier-Hein KH. An Attempt at Beating the 3D U-Net. *arXiv* 1908.02182v2 [eess.IV]. doi: 10.24926/548719.001
30. Ma J. Solution to the Kidney Tumour Segmentation Challenge 2019[C]// 2019 Kidney Tumour Segmentation Challenge: KiTS19 (2019). doi: 10.24926/548719.005
31. Hou X, Xie C, Li F, Nan Y. Cascaded Semantic Segmentation for Kidney and Tumour. doi: 10.24926/548719.002

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zuo, Zheng, He, Chen, Zheng, Zheng, You, Li, Liu, Bai, Si, Wang, Zhang, Wang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





## OPEN ACCESS

## Edited by:

Rong Na,  
Shanghai Jiao Tong University, China

## Reviewed by:

Charles Van Praet,  
Ghent University Hospital, Belgium  
Da Huang,  
Shanghai Jiao Tong University  
School of Medicine, China  
Yishuo Wu,  
Fudan University, China

## \*Correspondence:

Xudong Yao  
yaoxudong67@sina.com  
Bin Yang  
yangbnju@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

<sup>‡</sup>These authors have contributed  
equally to this work and share  
senior authorship

## Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

Received: 10 November 2021

Accepted: 15 December 2021

Published: 20 January 2022

## Citation:

Yang G, Xie J, Zhang S, Gu W,  
Yuan J, Wang R, Guo C, Ye L,  
Peng B, Yao X and Yang B (2022)  
Clinical Significance of Mesenchymal  
Circulating Tumor Cells in  
Patients With Oligometastatic  
Hormone-Sensitive Prostate Cancer  
Who Underwent Cytoreductive  
Radical Prostatectomy.  
Front. Oncol. 11:812549.  
doi: 10.3389/fonc.2021.812549

# Clinical Significance of Mesenchymal Circulating Tumor Cells in Patients With Oligometastatic Hormone-Sensitive Prostate Cancer Who Underwent Cytoreductive Radical Prostatectomy

Guanjie Yang<sup>1†</sup>, Jun Xie<sup>2†</sup>, Shun Zhang<sup>3†</sup>, Wenyu Gu<sup>1</sup>, Jing Yuan<sup>1</sup>, Ruiliang Wang<sup>1</sup>, Changcheng Guo<sup>1</sup>, Lin Ye<sup>1</sup>, Bo Peng<sup>1,2</sup>, Xudong Yao<sup>1,2\*‡</sup> and Bin Yang<sup>1\*‡</sup>

<sup>1</sup> Department of Urology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China,

<sup>2</sup> Shanghai Clinical College, Anhui Medical University, Shanghai, China, <sup>3</sup> Department of Urology, Affiliated Drum Tower Hospital, Medical School of Nanjing University, Nanjing, China

**Purpose:** Growing evidence shows that circulating tumor cells (CTCs) become more aggressive after the epithelial–mesenchymal transition (EMT), though the clinical significance of CTCs undergoing EMT in oligometastatic hormone-sensitive prostate cancer (omHSPC) patients has not yet been reported. Accordingly, the aim of this study was to detect the CTC level and investigate the clinical significance of mesenchymal CTCs in omHSPC patients who underwent cytoreductive radical prostatectomy (CRP).

**Materials and Methods:** Blood samples were drawn from 54 omHSPC patients who underwent CRP. The CanPatrol CTC enrichment technique was applied to isolate and identify different phenotypes of CTCs, which were classified as epithelial (E-CTCs), mesenchymal (M-CTCs), or biphenotypic epithelial/mesenchymal (Bi-CTCs). Univariable and multivariable Cox regression analyses were employed to investigate potential prognostic factors for metastatic castration-resistant prostate cancer (mCRPC)-free survival and cancer-specific survival (CSS). The prognostic value of CTCs for CSS and mCRPC-free survival was assessed using time-dependent receiver operating characteristic (ROC) curves and Kaplan–Meier analysis.

**Results:** CTCs were detected in 51 of 54 patients (94%). E-CTC, M-CTC, and Bi-CTC detection rates were 56%, 67%, and 85%, respectively. A positive correlation was found between the M-CTC count and number of bone metastases ( $p = 0.012$ ). Time-dependent ROC analysis showed that the M-CTC count had higher predictive power than E-CTC or Bi-CTC for mCRPC-free survival (3-year area under the curve [AUC] values: 0.64, 0.60, and 0.61) and CSS (3-year AUC: 0.86, 0.58, and 0.67). Additionally, time-dependent ROC analysis revealed total CTCs (T-CTCs)  $\geq 5$  and M-CTCs  $\geq 2$  to be the cutoff points with

optimal specificity and sensitivity. Based on multivariable Cox regression, T-CTC and M-CTC counts were both independently associated with CSS and mCRPC-free survival (all  $p < 0.05$ ), though E-CTCs and Bi-CTCs had no significant prognostic value (all  $p > 0.05$ ). Patients with T-CTC  $\geq 5$  or M-CTC  $\geq 2$  had significantly worse mCRPC-free survival and CSS than those with T-CTC  $< 5$  or M-CTC  $< 2$  (all  $p < 0.05$ ) after CRP.

**Conclusion:** CTC quantification and phenotype characterization provide prognostic information, and M-CTCs can be used as a novel biomarker for omHSPC patients who undergo CRP. The results need to be validated in prospective studies.

**Keywords:** oligometastatic prostate cancer, circulating tumor cell, epithelial–mesenchymal transition, androgen deprivation therapy, prostate-specific antigen, radical prostatectomy, liquid biopsy, biomarker

## INTRODUCTION

The oligometastatic state has been recognized as an intermediate state between localized disease and widespread metastases, suggesting the potential for preventing additional metastatic spread and improving survival with local treatment (such as surgery and radiotherapy) (1, 2). Retrospective studies have reported that cytoreductive radical prostatectomy (CRP) reduces the risk of clinical progression and improves cancer-specific survival (CSS) (3–5). According to a recent prospective registry, CRP is able to improve CSS and overall survival (OS) in newly diagnosed low-volume metastatic prostate cancer patients (6). Despite improvements in diagnosis and treatment, omHSPC patients comprise a heterogeneous population of men with different outcomes who will ultimately develop castration-resistant cancer (7, 8). Thus, it is necessary to develop a novel biomarker that can better predict the prognosis of these patients.

Circulating tumor cells (CTCs), which are shed from solid tumors, are presumed to constitute the mechanism for cancer metastasis (9). To date, CellSearch is the only method that has been analytically validated and cleared by the Food and Drug Administration (FDA) for use. The CTC level has been regarded as a surrogate biomarker for survival in patients with mCRPC (10, 11). It was also reported that CTCs might contribute to identifying high-risk prostate cancer patients with occult metastases at the time of diagnosis (12). In addition, CTCs might provide prognostic information for omHSPC and help in the selection of patients for CRP (13). However, the CTC isolation and capture techniques mentioned above depend on epithelial markers (such as epithelial cell adhesion molecule, EpCAM) for E-CTCs, which may fail to detect M-CTCs (14, 15).

Growing evidence shows that CTCs become more aggressive after adopting a mesenchymal phenotype during the epithelial–mesenchymal transition (EMT) (16, 17). Although mesenchymal CTCs (M-CTCs) are associated with tumor progression and poor prognosis in many carcinomas (18, 19), to the best of our knowledge, detection of M-CTCs in omHSPC has not been documented, and it remains unclear whether M-CTCs are involved in the progression of omHSPC after CRP. In this study, we used the CanPatrol CTC enrichment technique, which has been applied for a broad range of carcinomas based

on epithelial and mesenchymal markers (20–24), to detect the CTC level and to investigate the clinical significance of CTCs undergoing EMT in omHSPC patients treated with CRP.

## MATERIALS AND METHODS

### Patient Population

A total of 54 patients with omHSPC who underwent CRP at Shanghai Tenth People's Hospital of Tongji University from January 2015 to November 2017 were retrospectively enrolled in this study. All patients were examined by routine laboratory tests, serum prostate-specific antigen (PSA) and testosterone level measurement, thoracic, abdominal, and pelvic computed tomography (CT) scans, magnetic resonance imaging, and whole-body bone scan. The inclusion criteria were as follows: (a) resectable primary prostate cancer; (b) five or fewer bone metastases with or without suspicious pelvic nodal involvement confirmed by bone scan, CT scan, or magnetic resonance imaging; (c) no progression to mCRPC prior to CRP; (d) no visceral metastasis; (e) no local treatment for metastatic lesion prior to surgery; and (f) complete clinicopathological data and follow-up information. CRP was performed through open or laparoscopic surgery. Pathological stage was assessed using the American Joint Committee on Cancer 2010 TNM staging system and the Gleason grading system. All patients were treated with ADT until progression to CRPC after CRP. In the event of progression to mCRPC, systemic treatment was delivered by the treating physician using approved drugs for mCRPC. This study was carried out in accordance with the Declaration of Helsinki and approved by the ethical committee of our institution. All patients provided written informed consent.

### Isolation and Classification of CTCs

Isolation and classification of CTCs were performed as described in a previous study using the CanPatrol system (21, 24). Briefly, peripheral blood samples (5 ml, EDTA-anticoagulated) were collected at 12–14 days after surgery, and red blood cell lysis buffer (Sigma–Aldrich) was used to remove erythrocytes within 4 h of collection. The remaining cells were resuspended in PBS

with 4% formaldehyde for 5 min, and CTCs were isolated from the remaining cells using a filtration system consisting of a filtration tube containing a membrane (SurExam, Guangzhou, China), a manifold vacuum plate with valve settings (SurExam, Guangzhou, China), an E-Z 96 vacuum manifold (Omega, Norcross, USA), and a vacuum pump (Auto Science, Tianjin, China). An RNA-ISH assay was then performed to identify and classify CTCs based on the target sequences for leukocyte (CD45), epithelial (CK8/18/19 and EpCAM), and mesenchymal (Twist1 and Vimentin) markers; 4',6-diamidino-2-phenylindole (DAPI) was used to stain nuclei. The RNA-ISH assay was performed in a 24-well plate (Corning), and the cells were analyzed by fluorescence microscopy. CTCs were classified into three subgroups using the CanPatrol CTC enrichment technique: epithelial CTCs (epithelial biomarker positive, CD45 negative), mesenchymal CTCs (mesenchymal biomarker positive, CD45 negative), and biphenotypic epithelial/mesenchymal CTCs (epithelial/mesenchymal marker positive, CD45 negative).

## Statistical Analysis

mCRPC-free survival was defined as the time from initial diagnosis until CRPC. CRPC was defined as castration serum testosterone <50 ng/dl plus either radiological progression using RECIST (Response Evaluation Criteria in Solid Tumors) or three consecutive increases in PSA 1 week apart resulting in two 50% increases over the nadir and a PSA > 2 ng/ml (25). CSS was defined as the time from diagnosis to death from prostate cancer. Frequencies and proportions are used to describe categorical data, and medians and ranges are used to describe continuous data. Correlations of CTC count with continuous and categorical variables were evaluated using Pearson's test and Kruskal–Wallis *H* tests, respectively. Univariable Cox regression analyses were performed to assess prognostic factors for mCRPC-free survival and CSS; the significant individual (*p*-value <0.05) or clinically significant prognostic factors were assessed by multivariable Cox regression analysis. Time-dependent receiver operating characteristic (ROC) curve analysis was conducted to evaluate the optimal CTC cutoff point with the maximum Youden index value for predicting mCRPC-free survival and CSS after surgery. The Kaplan–Meier method with the log-rank test was applied to estimate mCRPC-free survival and CSS. SPSS ver. 22.0 (IBM Corporation, Armonk, NY, USA) and the R software environment for statistical computing were used for all statistical analyses. A value of *p* < 0.05 was considered statistically significant.

## RESULTS

### Patient Characteristics

Between January 2015 and November 2017, a total of 54 eligible patients with a median age of 68 years (IQR: 61–72 years) were enrolled in this study. The baseline and pathological characteristics of the patients are summarized in **Table 1**. Of the 54 patients, 41 (75.9%) had a pathologic Gleason score ≥8,

17 (31.5%) had pT4 disease, and 24 (44.4%) had lymph node metastasis. The positive surgical margin rate was 66.7%. The median follow-up period was 45 months (IQR: 43–49 months). In the overall cohort, 27 (50.0%) patients experienced progression to mCRPC during the follow-up period; 13 (24.1%) patients died. The 3-year mCRPC-free survival rate was 61.1% (**Figure 1A**), and the 3-year CSS rate was 79.6% (**Figure 1B**). The median time to mCRPC was 46.0 (95% CI: 37.7–54.6) months, though the median time to CSS was not reached. For patients with progression to mCRPC, 17 were treated with abiraterone, and 10 received chemotherapy.

### CTC Detection and Association With Clinicopathological Factors

CTCs were classified into three types through the CanPatrol technique (**Figure 2A**). CTCs were detected in 51 patients (94%), and the median CTC count was 4 (IQR: 3–9). As shown in **Figure 2B**, the detection rates of E-CTCs, M-CTCs, and Bi-CTCs were 56%, 67%, and 85%, respectively. The distribution of the three subtypes in each patient is depicted in **Figure 2C**. After using Pearson's test and Kruskal–Wallis *H* tests to evaluate the relationship between CTC count and clinical parameters, we found that both T-CTC count and Bi-CTC count correlated positively with lymph node invasion (both *p* < 0.05). In addition, T-CTC and M-CTC counts correlated positively with the number of bone metastases. No significant correlation was found between E-CTC count and clinicopathological factors, and there was no correlation between CTC count and PSA at diagnosis, pathologic Gleason score, or pathologic T stage (**Table 2**).

### Univariable and Multivariable Analyses

In univariable analysis, pathologic N stage, number of bone metastases, T-CTC count (continuous and categorical), and M-CTC count (continuous and categorical) were significantly associated with mCRPC-free survival and CSS (**Table 3**). Besides, postoperative adjuvant RT was an independent predictor of mCRPC-free survival (*p* = 0.013, **Table 3**) but not of CSS (*p* = 0.933, **Table 3**). After selecting the significant independent prognostic factors for multivariable Cox regression analysis, T-CTC count (continuous and categorical), M-CTC count (continuous and categorical), and postoperative adjuvant RT were significantly associated with mCRPC-free survival. Additionally, only T-CTC count (continuous and categorical) and M-CTC count (continuous and categorical) were significantly associated with CSS in multivariable Cox regression analysis.

### Prognostic Value of CTC Enumeration and Phenotype

Time-dependent ROC analysis was conducted to assess the role of CTCs in prognosis (**Figure 3** and **Table 4**). The results demonstrated that the M-CTC count had higher predictive power than E-CTC and Bi-CTC for mCRPC-free survival (3-year AUC: 0.64, 0.60, and 0.61; **Table 4**) and CSS (3-year AUC: 0.86, 0.58 and 0.67; **Table 4**). Additionally, M-CTC had higher

**TABLE 1 |** Patient characteristics.

Variables	Values
Number of patients	54
Age, years	
Median(range)	68 (61-72)
PSA at diagnosis (ng/ml)	
Median(range)	82.2 (38.5-100.4)
Pathologic Gleason score, n (%)	
≤7	13 (24.1)
8	14 (25.9)
≥9	27 (50.0)
Pathologic T stage, n (%)	
T2a-T3a	16 (29.6)
T3b	21 (38.9)
T4	17(31.5)
Pathologic N stage, n (%)	
N0	30 (55.6)
N1	24 (44.4)
Surgical margin, n (%)	
negative	18 (33.3)
positive	36 (66.7)
Number of bone metastases, n (%)	
1-3	36 (66.7)
4-5	18 (33.3)
preoperative ADT therapy, n (%)	
Yes	30 (55.6)
No	24 (44.4)
Postoperative adjuvant ADT, n (%)	54(100)
Postoperative adjuvant RT, n (%)	18(33.3)
First-line therapy for mCRPC (n=27), n (%)	
Abiraterone	17 (63.0)
Chemotherapy	10 (37.0)

PSA, Prostate-specific antigen; ADT, androgen deprivation therapy; mCRPC, metastatic castration-resistant prostate cancer; RT, radiotherapy.

predictive power than T-CTC for CSS (3-year AUC value: T-CTC, 0.74; M-CTC, 0.86; **Figure 3A**), whereas T-CTC had higher predictive power than M-CTC for mCRPC-free survival (3-year AUC value: T-CTC, 0.70; M-CTC, 0.64; **Figure 3A**). **Figure 3B** illustrates dynamic area under the curve (AUC) values over time, from 15 to 40 months, after surgery. According to time-dependent ROC analysis, the maximum Youden index value was applied to calculate the optimal CTC cutoff point, and the results showed that cutoffs of five T-CTCs ( $\geq 5$  vs.  $< 5$ ) and two M-CTCs ( $\geq 2$  vs.  $< 2$ ) had a statistically significant impact on mCRPC-free survival and CSS (all  $p < 0.05$ ).

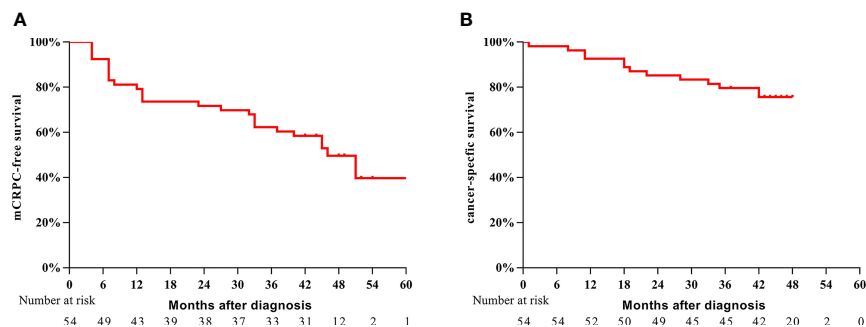
The CTC cutoff points were then used for survival analysis (**Figure 4**): patients with a T-CTC count  $\geq 5$  had significantly shorter 3-year mCRPC-free survival and CSS than those with a T-CTC count  $< 5$  (mCRPC-free survival: 39.1% vs. 77.4%,  $p < 0.001$ , **Figure 4A**; CSS: 60.9% vs. 93.5%,  $p = 0.007$ , **Figure 4B**). The median time to mCRPC was 23.0 (95% CI: 11.9–46.5) months for patients with a T-CTC count  $\geq 5$ , whereas the median time was not reached for those with a T-CTC count  $< 5$ .

Patients with an M-CTC count  $\geq 2$  had significantly shorter 3-year mCRPC-free survival and CSS than those with an M-CTC count  $< 2$  (mCRPC-free survival: 50.0% vs. 68.8%,  $p < 0.001$ , **Figure 4C**; CSS: 59.1% vs. 93.8%,  $p = 0.015$ , **Figure 4D**). The median mCRPC-free survival was 33.0 (95% CI: 16.4-65.2) months for patients with M-CTC count  $\geq 2$ ; for those with M-CTC count  $< 2$ , the median time was not reached.

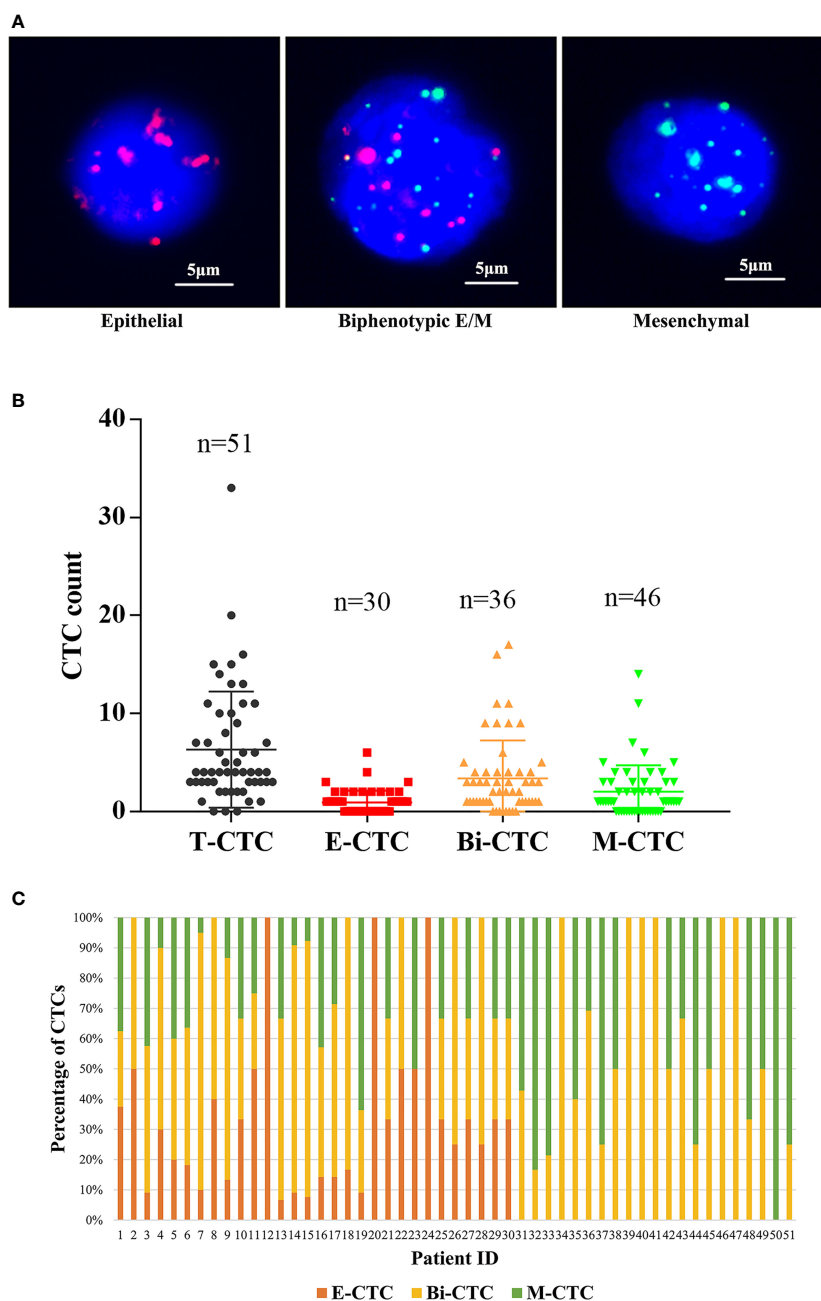
Additionally, we conducted subgroup analysis to further investigate the predictive value of CTCs. For the subgroup with M-CTC  $< 2$ , patients with T-CTC count  $\geq 5$  had shorter 3-year mCRPC-free survival (37.5% vs. 79.2%,  $p = 0.031$ , **Figure 5A**) and a trend of worse CSS (87.5% vs. 95.8%,  $p = 0.200$ , **Figure 5B**) than those with T-CTC count  $< 5$ . For the subgroup with T-CTC  $< 5$ , patients with M-CTC count  $\geq 2$  had shorter 3-year mCRPC-free survival (62.5% vs. 80.0%,  $p = 0.035$ , **Figure 5C**) and a trend of worse CSS (75.0% vs. 96.0%,  $p = 0.180$ , **Figure 5D**) than those with M-CTC count  $< 2$ .

## DISCUSSION

Despite increasing acknowledgment of the oligometastatic state in prostate cancer, there is no consensus on its definition. The majority of the published studies regards the prostate cancer with up to 3 to 5 metastatic lesions as the oligometastatic stage (2, 5, 13). In our study, an oligometastatic state was defined as five or fewer bone lesions, with or without suspicious pelvic nodal metastasis. This definition of oligometastatic prostate cancer is consistent with previous studies that investigated the impact of radical prostatectomy (3, 5). A phase 3 trial found that prostate-directed radiation can improve survival outcomes in patients with a low metastatic burden (26). Despite no hard evidence for the

**FIGURE 1 |** Kaplan–Meier survival analysis of mCRPC-free survival (A) and CSS (B) for the entire cohort of patients.





**FIGURE 2 | (A)** Representative fluorescence images of three types of CTCs isolated from the peripheral blood of omHSPC patients based on RNA-ISH staining for leukocytes (CD45, white), epithelial cells (EpCAM and CK8/18/19, red), and mesenchymal cells (vimentin and twist, green). 4',6-Diamidino-2-phenylindole was used to stain cell nuclei (blue). The scale bar indicates 5  $\mu$ m. **(B)** Levels of CTC subtypes. **(C)** The distribution of three subtypes of CTCs in each patient.

survival benefit of CRP in oligometastatic prostate cancer, several retrospective and prospective studies have suggested that CRP reduces the risk of clinical progression and improves long-term survival in patients with oligometastatic prostate cancer (3–6, 13). Currently, PSA kinetics are commonly used to follow disease progression, but they cannot be utilized to predict the prognosis of omHSPC patients well because of inherent limitations (24). Thus, it is extremely important to

identify an independent prognostic factor to help in therapeutic decision-making for omHSPC patients who undergo CRP.

It is reported that the number of metastatic lesions ( $\geq 10$  vs. 10) is not an independent predictor of mCRPC-free survival for metastatic hormone-sensitive prostate cancer (mHSPC) patients (24). One recent study also found that the number of metastases was not associated with overall survival in oligometastatic



**TABLE 2 |** Correlation of CTC count and phenotype with clinicopathological variables.

Variable	N	T-CTC P value	M-CTC P value	E-CTC P value	Bi-CTC P value
Age					
<70	30	0.30	0.46	0.45	0.43
≥70	24				
PSA at diagnosis	54	0.33	0.46	0.44	0.95
Pathologic Gleason score					
≤7	13	0.12	0.30	0.92	0.34
8	14				
≥9	27				
pT stage					
T2a-T3a	16	0.19	0.066	0.86	0.86
T3b	21				
T4	17				
pN stage					
N0	30	0.014	0.087	0.18	0.027
N1	24				
Surgical margin					
positive	36	0.42	0.38	0.37	0.17
negative	18				
Number of metastases					
1-3	36	0.012	0.009	0.50	0.12
4-5	18				

CTC, circulating tumor cell; T-CTC, total circulating tumor cell; M-CTC, mesenchymal circulating tumor cell; E-CTC, epithelial circulating tumor cell; Bi-CTC, biphenotypic circulating tumor cell; N, number; PSA, Prostate-specific antigen; pT stage, Pathologic T stage; pN stage, Pathologic N stage.

patients with prostate cancer who were treated with local and metastatic curative radiotherapy (27). In the present study, we found that the number of bone metastases was not a significant predictor of mCRPC-free survival or CSS in omHSPC patients. These findings suggest that the number of metastases might not be the primary prognostic factor for prostate cancer and support investigation of an independent biomarker for omHSPC when the metastatic burden varies from one to five. Since the discovery of CTCs, enormous attention has been given to investigating their potential as prognostic and treatment response biomarkers for mCRPC patients (11, 28). In this study, we detected the level of CTCs, characterized their phenotype, and further explored the clinical significance of CTCs in omHSPC patients who underwent CRP.

Although the CellSearch system has been widely used to detect CTCs depending on tumor cell epithelial markers such as EpCAM and CK, it fails to detect CTCs undergoing EMT. Thus, we used the CanPatrol enrichment technique, a filter-based method that uses a combination of epithelial and mesenchymal markers, to isolate and identify different phenotypes of CTCs; Bi-CTCs and M-CTCs can be simultaneously detected based on this novel system in addition to E-CTCs. The detection rate of CTCs was 94%, higher than that using the CellSearch system for mHSPC patients (38%–48.5%) (13, 29). Several reports have indicated an increase in CTC count with progression of the disease stage; for instance, the detection rate of CTCs was 0%–10% for healthy volunteers (30, 31), 5%–38.4% for nonmetastatic high-risk prostate cancer patients (30, 31), and 80% for mCRPC patients (15). In our cohort, all patients had metastatic prostate cancer, with 75.9% having disease with a Gleason score ≥8, and the aggressive clinicopathological characteristics might be another reason for the higher detection rate of CTCs.

CTCs have been proposed as prognostic biomarkers to predict treatment response and survival outcomes in mCRPC patients. A phase III trial (SWOG S0421) recruiting mCRPC patients treated with docetaxel indicated that baseline CTC count ≥5 was associated with shorter OS (28). Scher et al. (11) also found that a biomarker panel containing the CTC number and LDH level could be used as a surrogate for survival in mCRPC patients treated with abiraterone. The clinical significance of CTCs in mHSPC patients has recently been documented (13, 24). A prospective study of omHSPC patients receiving CRP demonstrated that CTC count ≥2 is associated with shorter mCRPC-free survival and OS (13). Similarly, a phase III prospective randomized trial (SWOG S1216) of ADT combined with orteronel or bicalutamide for mHSPC patients found that the baseline CTC count was highly prognostic of 7-month PSA and 2-year PFS (32). Consistent with the literature, our study found the T-CTC count to be an independent predictor of mCRPC-free survival and CSS in multivariable Cox regression analysis; in addition, an optimal cutoff of 5 had a statistically significant impact on survival outcomes. Overall, T-CTC count ≥5 was associated with early progression to mCRPC and shorter CSS, in accordance with the literature (11, 29).

It is well known that CTCs adopt a mesenchymal phenotype in the process of EMT, which endows cells with multiple malignant traits (16, 17). Recently, several studies have reported the significance of M-CTCs in a variety of malignancies, including breast (19), liver (21), colorectal (22), and prostate (24) cancer. These studies demonstrate that M-CTCs are significantly associated with early recurrence, progression, and metastasis (19, 22, 24). By using the CanPatrol system to detect M-CTCs, the current study found that a higher M-CTC count was associated with a higher number of metastases, though no significant correlation between E-CTCs

**TABLE 3 |** Univariable and multivariable Cox analysis for mCRPC-free survival and CSS.

Characteristics	mCRPC-free survival				CSS			
	Univariable analysis		Multivariable analysis		Univariable analysis		Multivariable analysis	
	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Age	1.030 (0.974-1.091)	0.294	1.087 (0.998-1.184)	0.154 <sup>a</sup>	0.995 (0.920-1.076)	0.901	0.983 (0.839-1.152)	0.833 <sup>c</sup>
PSA at diagnosis	0.997 (0.993-1.002)	0.256	0.994 (0.987-1.126)	0.256 <sup>a</sup>	0.995 (0.986-1.004)	0.270	0.941 (0.902-1.182)	0.125 <sup>c</sup>
Pathologic Gleason score								
≤7	Referent	0.045	Referent	0.198 <sup>a</sup>	Referent	0.759		
8	1.257 (0.856-2.175)	0.077	3.911 (0.722-7.176)	0.114	3.582 (0.749-7.936)	0.942		
≥9	4.236 (1.445-6.908)	0.014	4.006 (0.858-8.709)	0.078	2.535 (0.815-5.628)	0.940		
Pathologic T stage								
T2a-T3a	Referent	0.233			Referent	0.885		
T3b	2.330 (0.819-6.633)	0.113			2.559 (0.856-7.645)	0.932		
T4	2.341 (0.795-6.893)	0.123			1.745 (0.762-6.379)	0.931		
Pathologic N stage (N0 vs. N1)	3.786 (1.87-10.858)	0.010	0.370 (0.020-6.720)	0.502 <sup>a</sup>	0.363 (0.170-0.774)	0.009	3.245 (0.611-17.231)	0.167 <sup>c</sup>
Number of metastases (1-3 vs. 4-5)	2.552 (1.164-5.598)	0.019	1.285 (0.482-3.425)	0.617 <sup>a</sup>	5.551 (1.705-8.069)	0.004	2.174 (0.548-8.632)	0.270 <sup>c</sup>
Positive surgical margin (No vs. Yes)	1.358 (0.586-3.146)	0.475			3.383 (0.741-15.451)	0.116		
Postoperative adjuvant RT (No vs. Yes)	0.390 (0.147-0.835)	0.039	0.113 (0.130-0.422)	0.013 <sup>a</sup>	0.950 (0.286-3.156)	0.933		
T-CTC (continuous)	1.123 (1.063-1.187)	<0.001	1.182 (1.052-1.329)	0.035 <sup>a</sup>	1.179 (1.082-1.286)	<0.001	1.311 (1.110-1.549)	0.001 <sup>c</sup>
E-CTC (continuous)	1.234 (0.931-1.636)	0.144			1.341 (0.906-1.985)	0.143		
M-CTC (continuous)	1.303 (1.152-1.473)	<0.001	1.259 (1.081-1.466)	0.038 <sup>a</sup>	1.455 (1.215-1.743)	<0.001	1.386 (1.135-1.693)	0.001 <sup>c</sup>
Bi-CTC (continuous)	1.128 (1.036-1.228)	0.005	0.873 (0.730-1.044)	0.137 <sup>a</sup>	1.145 (1.026-1.278)	0.015	1.050 (0.913-1.208)	0.491 <sup>c</sup>
T-CTC (<5 vs. ≥5)	4.404 (1.946-9.969)	<0.001	4.150 (1.453-7.852)	0.020 <sup>b</sup>	5.005 (1.374-8.232)	0.015	3.362 (1.684-8.159)	0.024 <sup>d</sup>
M-CTC (<2 vs. ≥2)	3.277 (1.495-7.182)	0.003	3.341 (1.334-8.363)	0.011 <sup>b</sup>	3.911 (1.198-6.769)	0.024	3.912 (1.160-7.194)	0.028 <sup>d</sup>

mCRPC, metastatic castration-resistant prostate cancer; CSS, cancer specific survival; HR, hazard ratio; CI, confidence interval; CTC, circulating tumor cell; T-CTC, total circulating tumor cell; M-CTC, mesenchymal circulating tumor cell; E-CTC, epithelial circulating tumor cell; Bi-CTC, biphenotypic circulating tumor cell; RT, radiotherapy.

<sup>a</sup>Adjusted for: T-CTC (continuous), M-CTC (continuous), Bi-CTC (continuous), age, PSA at diagnosis, pathologic Gleason score, pathologic N stage, number of metastases and postoperative adjuvant RT.

<sup>b</sup>Adjusted for: T-CTC (<5 vs. ≥5), M-CTC (<2 vs. ≥2), Bi-CTC (continuous), age, PSA at diagnosis, pathologic Gleason score, pathologic N stage, number of metastases and postoperative adjuvant RT.

<sup>c</sup>Adjusted for: T-CTC (continuous), M-CTC (continuous), Bi-CTC (continuous), age, PSA at diagnosis, pathologic N stage and number of metastases.

<sup>d</sup>Adjusted for: T-CTC (<5 vs. ≥5), M-CTC (<2 vs. ≥2), Bi-CTC (continuous), age, PSA at diagnosis, pathologic N stage and number of metastases.

and the number of metastases was detected, which might be attributed to the higher invasion and migration potential of M-CTCs than E-CTCs (16, 17). According to time-dependent ROC analysis, we found that the M-CTC count had higher predictive power than E-CTC and Bi-CTC counts for mCRPC-free survival and CSS. Indeed, E-CTCs and Bi-CTCs did not show a significant relationship with prognosis for omHSPC patients, whereas M-CTCs were independently associated with mCRPC-free survival and CSS. Specifically, patients with M-CTC count ≥2 had worse mCRPC-free survival and CSS. We also explored the predictive value of M-CTC in the subgroup of patients with T-CTC count <5, and the results again demonstrated that those with M-CTC count ≥2 had a significantly shorter time to mCRPC and a trend of worse CSS. The reason that there was no statistical significance for CSS may be due to the small sample size. Thus, a prospective study with a large sample size is needed for further exploration. Our findings demonstrate the prognostic significance of M-CTCs for omHSPC, even in patients with T-CTC count <5, supporting the potential use of M-CTCs as a novel biomarker for omHSPC patients who undergo CRP.

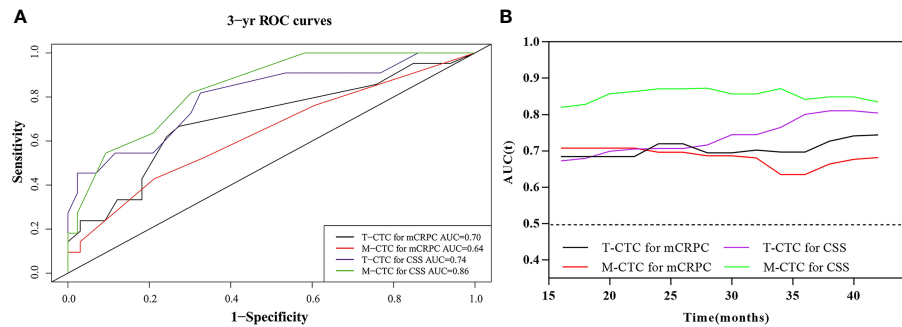
It is reported that the incidence rate of positive surgical margins (PSM+) ranges from 72.7% to 78.9% in oligometastatic prostate cancer patients treated with CRP (2, 3, 5, 13). In our study, the rate of PSM+ was 66.7%, which was in accordance with the data in these previous studies. In addition, 50% of the patients with PSM+ in our study had received postoperative adjuvant RT plus ADT, whereas the others received ADT alone. Overall, postoperative

adjuvant RT is an independent predictor of mCRPC-free survival, though there was no significant association between RT and CSS. The results need to be validated in larger prospective studies.

There were some limitations in our research. First, this was a retrospective study, and the results need to be validated in future prospective studies. Second, the small cohort size might have caused bias and influenced the results of multivariable analyses. Third, the impact of additional therapy (chemotherapy and abiraterone) after mCRPC could not be adjusted in multivariable analysis because the number of mCRPC patients was small and the patients were treated differently according to their physician's choice. Fourth, the optimal time interval for CTC quantification after surgery has not been conclusively established. Dynamic monitoring of CTC changes might be essential in the future (13).

## CONCLUSION

This study is the first to demonstrate that both T-CTC count ≥5 and M-CTC count ≥2 are independent predictors of early progression to mCRPC and shorter CSS after CRP for omHSPC patients. The findings support the use of CTC quantification and phenotype characterization as a prognostic biomarker to identify patients' cases progressing early and to select intensive treatment after surgery. The results need to be validated in prospective, randomized trials in the future.

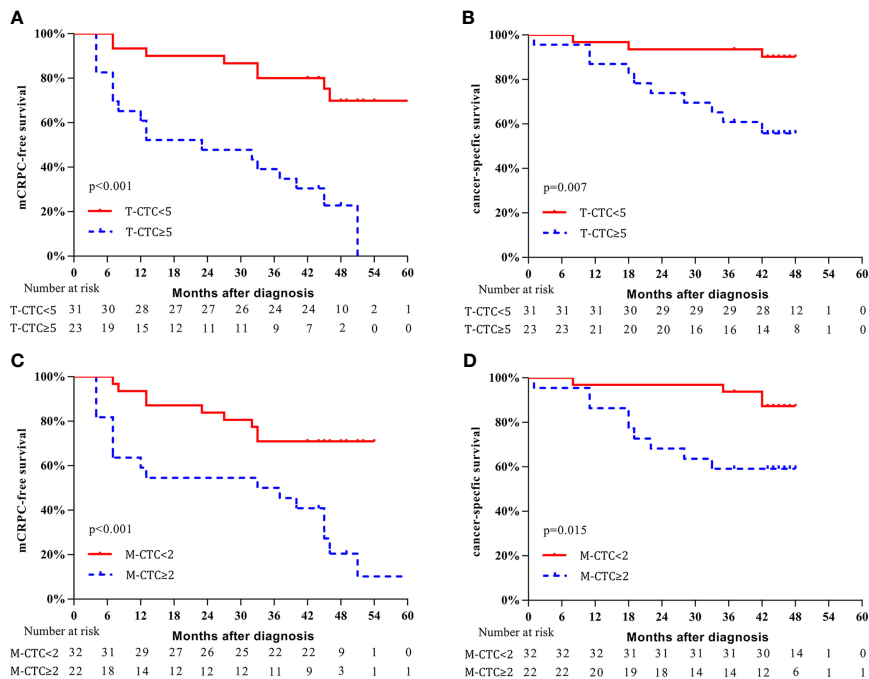


**FIGURE 3 | (A)** Comparison of 3-year predictive efficiency among CTC subtypes according to time-dependent receiver operating characteristic curve analysis. **(B)** The distribution of the dynamic AUC over time.

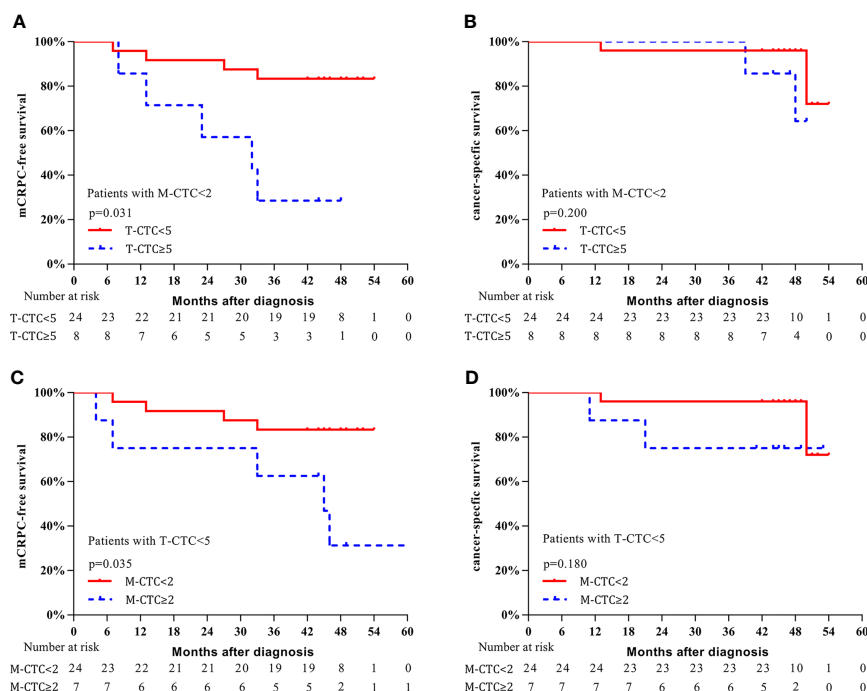
**TABLE 4 |** Comparison of predictive efficiency among different CTC subtypes according to time-dependent receiver operating characteristic curve analysis.

	AUC for 3 years		AUC for 4 years		AUC for 5 years	
	mCRPC-free survival	CSS	mCRPC-free survival	CSS	m-CRPC-free survival	CSS
E-CTC	0.60	0.58	0.53	0.55	0.61	0.55
Bi-CTC	0.61	0.67	0.60	0.67	0.65	0.67
M-CTC	0.64	0.86	0.76	0.76	0.78	0.76
T-CTC	0.70	0.74	0.74	0.77	0.82	0.77

CTC, circulating tumor cell; T-CTC, total circulating tumor cell; M-CTC, mesenchymal circulating tumor cell; E-CTC, epithelial circulating tumor cell; Bi-CTC, biphenotypic circulating tumor cell.



**FIGURE 4 |** Kaplan-Meier curves of survival outcomes according to CTC phenotype. **(A)** mCRPC-free survival stratified according to T-CTC count. **(B)** CSS stratified according to T-CTC count. **(C)** mCRPC-free survival stratified according to M-CTC count. **(D)** CSS stratified according to M-CTC count.



**FIGURE 5 |** Kaplan-Meier curves of survival outcomes. **(A)** mCRPC-free survival stratified according to T-CTC count in the subgroup of M-CTC < 2. **(B)** CSS stratified according to T-CTC count in the subgroup of M-CTC < 2. **(C)** mCRPC-free survival stratified according to M-CTC count in the subgroup of T-CTC < 5. **(D)** CSS survival stratified according to M-CTC count in the subgroup of T-CTC < 5.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee at the Tenth People's Hospital of Shanghai (SHSY-IEC-4.1/20-22/01). The patients/participants provided their written informed consent to participate in this study.

## REFERENCES

- Hellman S, Weichselbaum RR. Oligometastases. *J Clin Oncol: Off J Am Soc Clin Oncol* (1995) 13(1):8–10. doi: 10.1200/jco.1995.13.1.8
- Sundahl N, Tree A, Parker C. The Emerging Role of Local Therapy in Metastatic Prostate Cancer. *Current Onco Reports* (2020) 22(1):2. doi: 10.1007/s11912-020-0868-1
- Gandaglia G, Fossati N, Stabile A, Bandini M, Rigatti P, Montorsi F, et al. Radical Prostatectomy in Men With Oligometastatic Prostate Cancer: Results of a Single-Institution Series With Long-Term Follow-Up. *Eur Urol* (2017) 72(2):289–92. doi: 10.1016/j.eururo.2016.08.040

## AUTHOR CONTRIBUTIONS

Conception and design: XY and BY. Acquisition of data: GY, JX, and WG. Analysis and interpretation of data: GY, SZ, JY, RW, and CG. Writing, review, and/or revision of the manuscript: GY, JX, SZ, and BY. Administrative, technical, or material support: XY and BP. Study supervision: LY, BP, XY, and BY. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was partly supported by the National Natural Science Foundation of China (Grant No. 31570993) and Shanghai Pujiang Program (Grant No. 15PJ1407000).



- Treatment of Metastatic Prostate Cancer (LoMP) Registry. *Eur Urol Open Sci* (2021) 29:68–76. doi: 10.1016/j.euros.2021.05.006
7. Armstrong AJ, Szmulewitz RZ, Petrylak DP, Holzbeierlein J, Villers A, Azad A, et al. ARCHES: A Randomized, Phase III Study of Androgen Deprivation Therapy With Enzalutamide or Placebo in Men With Metastatic Hormone-Sensitive Prostate Cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2019) 37(32):2974–86. doi: 10.1200/jco.19.00799
  8. James ND, Spears MR, Clarke NW, Dearnaley DP, De Bono JS, Gale J, et al. Survival With Newly Diagnosed Metastatic Prostate Cancer in the “Docetaxel Era”: Data From 917 Patients in the Control Arm of the STAMPEDE Trial (MRC PR08, CRUK/06/019). *Eur Urol* (2015) 67(6):1028–38. doi: 10.1016/j.eururo.2014.09.032
  9. Alix-Panabières C, Pantel K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discovery* (2016) 6(5):479–91. doi: 10.1158/2159-8290.Cd-15-1483
  10. de Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, et al. Circulating Tumor Cells Predict Survival Benefit From Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clin Cancer Res: An Off J Am Assoc Cancer Res* (2008) 14(19):6302–9. doi: 10.1158/1078-0432.Ccr-08-0872
  11. Scher HI, Heller G, Molina A, Attard G, Danila DC, Jia X, et al. Circulating Tumor Cell Biomarker Panel as an Individual-Level Surrogate for Survival in Metastatic Castration-Resistant Prostate Cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2015) 33(12):1348–55. doi: 10.1200/jco.2014.55.3487
  12. Cieřlikowski WA, Budna-Tukan J, Świerczewska M, Ida A, Hrab M, Jankowiak A, et al. Circulating Tumor Cells as a Marker of Disseminated Disease in Patients With Newly Diagnosed High-Risk Prostate Cancer. *Cancers* (2020) 12(1):160. doi: 10.3390/cancers12010160
  13. Mandel PC, Huland H, Tiebel A, Haese A, Salomon G, Budaus L, et al. Enumeration and Changes in Circulating Tumor Cells and Their Prognostic Value in Patients Undergoing Cytoreductive Radical Prostatectomy for Oligometastatic Prostate Cancer-Translational Research Results From the Prospective ProMPT Trial. *Eur Urol Focus* (2021) 7(1):55–62. doi: 10.1016/j.euf.2019.05.008
  14. Alix-Panabières C, Pantel K. Challenges in Circulating Tumour Cell Research. *Nat Rev Cancer* (2014) 14(9):623–31. doi: 10.1038/nrc3820
  15. Liu W, Yin B, Wang X, Yu P, Duan X, Liu C, et al. Circulating Tumor Cells in Prostate Cancer: Precision Diagnosis and Therapy. *Oncol Lett* (2017) 14(2):1223–32. doi: 10.3892/ol.2017.6332
  16. Książkiewicz M, Markiewicz A, Zaczek AJ. Epithelial-Mesenchymal Transition: A Hallmark in Metastasis Formation Linking Circulating Tumor Cells and Cancer Stem Cells. *Pathobiol: J Immunopathol Mol Cell Biol* (2012) 79(4):195–208. doi: 10.1159/000337106
  17. Lozar T, Gersak K, Cemazar M, Kuhar CG, Jesenko T. The Biology and Clinical Potential of Circulating Tumor Cells. *Radiol Oncol* (2019) 53(2):131–47. doi: 10.2478/raon-2019-0024
  18. Danila DC, Fleisher M, Scher HI. Circulating Tumor Cells as Biomarkers in Prostate Cancer. *Clin Cancer Res: An Off J Am Assoc Cancer Res* (2011) 17(12):3903–12. doi: 10.1158/1078-0432.Ccr-10-2650
  19. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Sci (New York NY)* (2013) 339(6119):580–4. doi: 10.1126/science.1228522
  20. Ni C, Shen Y, Fang Q, Zhang M, Yuan H, Zhang J, et al. Prospective Study of the Relevance of Circulating Tumor Cell Status and Neoadjuvant Chemotherapy Effectiveness in Early Breast Cancer. *Cancer Med* (2020) 9(7):2290–8. doi: 10.1002/cam4.2876
  21. Qi LN, Xiang BD, Wu FX, Ye JZ, Zhong JH, Wang YY, et al. Circulating Tumor Cells Undergoing EMT Provide a Metric for Diagnosis and Prognosis of Patients With Hepatocellular Carcinoma. *Cancer Res* (2018) 78(16):4731–44. doi: 10.1158/0008-5472.Can-17-2459
  22. Wang W, Wan L, Wu S, Yang J, Zhou Y, Liu F, et al. Mesenchymal Marker and LGR5 Expression Levels in Circulating Tumor Cells Correlate With Colorectal Cancer Prognosis. *Cell Oncol (Dordrecht)* (2018) 41(5):495–504. doi: 10.1007/s13402-018-0386-4
  23. Wu S, Liu S, Liu Z, Huang J, Pu X, Li J, et al. Classification of Circulating Tumor Cells by Epithelial-Mesenchymal Transition Markers. *PLoS One* (2015) 10(4):e0123976. doi: 10.1371/journal.pone.0123976
  24. Yang YJ, Kong YY, Li GX, Wang Y, Ye DW, Dai B. Phenotypes of Circulating Tumour Cells Predict Time to Castration Resistance in Metastatic Castration-Sensitive Prostate Cancer. *BJU Int* (2019) 124(2):258–67. doi: 10.1111/bju.14642
  25. Cornford P, van den Bergh R, Briers E, Van den Broeck T, Cumberbatch M, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer. Part II-2020 Update: Treat Relapsing Metastatic Prostate Cancer (2021) 79(2):263–82. doi: 10.1016/j.eururo.2020.09.046
  26. Parker CC, James ND, Brawley CD, Clarke NW, Hoyle AP, Ali A, et al. Radiotherapy to the Primary Tumour for Newly Diagnosed, Metastatic Prostate Cancer (STAMPEDE): A Randomised Controlled Phase 3 Trial. *Lancet (London England)* (2018) 392(10162):2353–66. doi: 10.1016/s0140-6736(18)32486-3
  27. Reverberi C, Massaro M, Osti MF, Anzellini D, Marinelli L, Montalto A, et al. Local and Metastatic Curative Radiotherapy in Patients With De Novo Oligometastatic Prostate Cancer. *Sci Rep* (2020) 10(1):17471. doi: 10.1038/s41598-020-74562-3
  28. Vogelzang NJ, Fizazi K, Burke JM, De Wit R, Bellmunt J, Hutson TE, et al. Circulating Tumor Cells in a Phase 3 Study of Docetaxel and Prednisone With or Without Lenalidomide in Metastatic Castration-Resistant Prostate Cancer. *Eur Urol* (2017) 71(2):168–71. doi: 10.1016/j.eururo.2016.07.051
  29. Goldkorn A, Tangen C, Plets M, Morrison G, Cunha A, Xu T, et al. Baseline Circulating Tumor Cell (CTC) Count as a Prognostic Marker of PSA Response and Progression in Metastatic Castrate Sensitive Prostate Cancer (mCSPC): Results From SWOG S1216, a Phase III Randomized Trial of Androgen Deprivation Plus Orteronel (Cyp17 Inhibitor) or Bicalutamide. (2020) 38(15\_suppl):5506. doi: 10.1200/JCO.2020.38.15\_suppl.5506
  30. Resel Folkersma L, San José Manso L, Galante Romo I, Moreno Sierra J, Olivier Gómez C. Prognostic Significance of Circulating Tumor Cell Count in Patients With Metastatic Hormone-Sensitive Prostate Cancer. *Urol Dec* (2012) 80(6):1328–32. doi: 10.1016/j.urology.2012.09.001
  31. Thalgott M, Rack B, Maurer T, Souvatzoglou M, Eiber M, Kreß V, et al. Detection of Circulating Tumor Cells in Different Stages of Prostate Cancer. *J Cancer Res Clin Oncol May* (2013) 139(5):755–63. doi: 10.1007/s00432-013-1377-5
  32. Agarwal N, Tangen C, Hussain MHA, Gupta S, Plets M, Lara PLN, et al. SWOG S1216: A Phase III Randomized Trial Comparing Androgen Deprivation Therapy (ADT) Plus TAK-700 With ADT Plus Bicalutamide in Patients (Pts) With Newly Diagnosed Metastatic Hormone-Sensitive Prostate Cancer (mHSPC) (Nct01809691). *J Clin Onco* (2021) 39(15\_suppl):5001. doi: 10.1200/JCO.2021.39.15\_suppl.5001

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Xie, Zhang, Gu, Yuan, Wang, Guo, Ye, Peng, Yao and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Deep Learning–Based Classification of Epithelial–Mesenchymal Transition for Predicting Response to Therapy in Clear Cell Renal Cell Carcinoma

## OPEN ACCESS

### Edited by:

Gong-Hong Wei,  
Fudan University, China

### Reviewed by:

Andrea Benedetto Galosi,  
Marche Polytechnic University, Italy  
Wenying Yan,  
Soochow University, China  
Shangwei Ning,  
Harbin Medical University, China

### \*Correspondence:

Hongyu Wang  
whyu@dlut.edu.cn  
Wenlong Liu  
liuwl@dlut.edu.cn  
Deyong Yang  
yangdeyong@dmu.edu.cn  
Liang Cheng  
liang\_cheng@yahoo.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 24 September 2021

**Accepted:** 31 December 2021

**Published:** 24 January 2022

### Citation:

Chen Q, Kuai Y, Wang S, Zhu X,  
Wang H, Liu W, Cheng L and Yang D  
(2022) Deep Learning–Based  
Classification of Epithelial–  
Mesenchymal Transition for Predicting  
Response to Therapy in Clear Cell  
Renal Cell Carcinoma.  
*Front. Oncol.* 11:782515.  
doi: 10.3389/fonc.2021.782515

Qiwei Chen<sup>1,2†</sup>, Yue Kuai<sup>3†</sup>, Shujing Wang<sup>4</sup>, Xinqing Zhu<sup>1</sup>, Hongyu Wang<sup>3\*</sup>,  
Wenlong Liu<sup>3\*</sup>, Liang Cheng<sup>5\*</sup> and Deyong Yang<sup>1\*</sup>

<sup>1</sup> Department of Urology, First Affiliated Hospital of Dalian Medical University, Dalian, China, <sup>2</sup> School of Information Science and Technology of Dalian Maritime University, Dalian, China, <sup>3</sup> School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, <sup>4</sup> Department of Biochemistry and Molecular Biology, Dalian Medical University, Dalian, China, <sup>5</sup> Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Epithelial–mesenchymal transition (EMT) profoundly impacts prognosis and immunotherapy of clear cell renal cell carcinoma (ccRCC). However, not every patient is tested for EMT status because this requires additional genetic studies. In this study, we developed an EMT gene signature to classify the H&E-stained slides from The Cancer Genome Atlas (TCGA) into epithelial and mesenchymal subtypes, then we trained a deep convolutional neural network to classify ccRCC which according to our EMT subtypes accurately and automatically and to further predict genomic data and prognosis. The clinical significance and multiomics analysis of the EMT signature was investigated. Patient cohorts from TCGA (n = 252) and whole slide images were used for training, testing, and validation using an algorithm to predict the EMT subtype. Our approach can robustly distinguish features predictive of the EMT subtype in H&E slides. Visualization techniques also detected EMT-associated histopathological features. Moreover, EMT subtypes were characterized by distinctive genomes, metabolic states, and immune components. Deep learning convolutional neural networks could be an extremely useful tool for predicting the EMT molecular classification of ccRCC tissue. The underlying multiomics information can be crucial in applying the appropriate and tailored targeted therapy to the patient.

**Keywords:** clear cell renal cell carcinoma, epithelial–mesenchymal transition, deep learning, histopathology, immune checkpoint inhibitor

**Abbreviations:** TCGA, The Cancer Genome Atlas; EMT, Epithelial–mesenchymal transition; ccRCC, clear cell renal cell carcinoma; Mes, mesenchymal; Epi, epithelial; CAM, class activation mapping; AMPK, AMP-activated protein kinase; CTA, cancer-testis antigen; HRD, homologous recombination deficiency; ITH, intratumoral heterogeneity; TMB, tumor mutation burden; ICIs, immune checkpoint inhibitors.

## HIGHLIGHTS

In this study, we trained a deep convolutional neural network on hematoxylin and eosin (H&E) histology slides obtained from The Cancer Genome Atlas (TCGA) to classify their EMT status accurately and automatically and to further predict genomic data and prognosis. We revealed that Mes cluster demonstrated truncating mutations in PBRM1, and high expression of immune checkpoint molecules might lead to the immune escape of this cluster. Further we suggested that patients in the Mes subtype might respond better to ICIs combined with antiangiogenic therapy. Deep learning convolutional neural networks could be an extremely useful tool for predicting the EMT molecular classification of ccRCC tissue. The underlying multiomics information can be crucial in applying the appropriate and tailored targeted therapy to the patient.

## INTRODUCTION

Clear cell renal cell carcinomas (ccRCC) account for approximately 80% of all renal cancer cases, with approximately 3.8% of all cancers in United States (1). Metastatic ccRCC are pharmacologically managed, targeted therapy utilizing tyrosine kinase inhibitors (TKIs), anti-VEGF antibodies, mammalian target of rapamycin (mTOR), and/or immune checkpoint inhibitors are widely used in first and second line treatments, suggesting that treatment strategy is crucial for ccRCC treatment (2).

Epithelial–mesenchymal transition (EMT) is a multistep process in which epithelial cells gain a range of mesenchymal characteristics (3). EMT molecular stratification can predict whether patients respond to immunotherapy in several tumor types (4–6). Thus, we sought to develop an EMT gene signature that can predict genomic data and prognosis of patients with ccRCC.

All EMT stratification systems are based on complex molecular experiments (7, 8). Therefore, there is a tremendous medical need for simplifying procedures. One key could be the use of deep neural networks. Recently, Zhang et al. presented a comprehensive morphological analysis using computer vision methods including random decision forests and artificial neural networks to establish the correlation between cellular morphological features and EMT (9). Kather et al. predicted microsatellite instability (MSI) directly from histology in gastrointestinal cancer using convolutional neural networks (10). Nevertheless, unlike the typical MSI tumors, there are no standard histological criteria for EMT molecular subtypes in ccRCC patients. Herein, we investigated the deep learning neural network to precisely recognize the ccRCC EMT subtypes from whole-slide images of hematoxylin and eosin (H&E)–stained tissue from TCGA (The Cancer Genome Atlas). Additionally, we compared subtype comprehensive genomic, phenotypic, and clinical data.

## MATERIALS AND METHODS

### RNA Expression Data and Somatic Exome Mutation Analysis

The RNA-Seq upper quartile normalized RSEM data was available for 539 ccRCC; all data is accessible *via* the NCI genome data commons and the Gene Expression Omnibus (<https://gdc.cancer.gov/> and <https://www.ncbi.nlm.nih.gov/geo/>). GSE150404 has 60 samples with different stages. Clinicopathologic data for the corresponding patients, including gender, race, age, tumor location, histology classification, differentiation grade, tumor stage, and survival information, were also retrieved from the database. Only patients with both survival information and expression data available were included in this study. Analysis of the RNA data was split into miRNA analysis, mRNA signature analysis, and immune gene signature analysis. All subsequent analyses were performed in R open-source programming language. For differential expression analysis, RPKM values were calculated from RNaseq raw counts and upper quantile normalized. For hierarchical clustering and WGCNA, raw count data were processed and normalized using the variance stabilizing transformation algorithm implemented by the DESeq2 package.

Somatic exome sequencing data were downloaded for the 255 ccRCC. The tumors with sequencing data are accessible *via* the NCI genome data commons (<https://gdc.cancer.gov/>). The SMG that had been previously identified by the MutSigCV algorithm in the previous TCGA KIRC publications were used as the reference SMG.

### DNA Methylation Analysis

All data is available from TCGA data. To minimize the influence of tumor purity, we dichotomized the methylation data with a beta value cut off 0.3 and used the Ward method to cluster the distance matrix computed with the Jaccard Index. Heatmaps were generated based on row and column orders calculated as above and colored by dichotomized beta values. All methylation genes we selected with a beta value of 0.3 or more were considered evidence for epigenetic silencing and associated with poor survival.

### Development of EMT-Gene Signature in ccRCC

To find EMT-specific genes in ccRCC, gene expression data were analyzed from TCGA cohort. The EMT-related gene expression signature comprised 200 genes obtained from gene set “hallmark epithelial mesenchymal transition” in The Molecular Signatures Database (MSigDB, software. [broadinstitute.org/gsea/msigdb](http://broadinstitute.org/gsea/msigdb)). Genes were selected when the mRNA expression levels were either positively or negatively correlated with at least one of the well-known EMT markers: E-cadherin (CDH1), vimentin (VIM), N-cadherin (CDH2), and fibronectin 1 (FN1). Using a gene feature and its correlated genes, hierarchical clustering analysis was performed with the centered correlation coefficient as the measure of similarity (**Supplementary Table S1**). The patient clustering result divided the patients into two subtypes

(mesenchymal and epithelial). We also calculated EMT scores based on previous studies and obtained comparable results.

## Image Preprocessing

All the slides were tiled in non-overlapping 256×256-pixel windows at a magnification of 10× using the openslide library. The tiles with more than 50% background (for which all the values are below 220 in the RGB color space) were considered to have insufficient information and were dropped while tiling (11). For the remaining tiles, we yielded 1150952 tiles, with an average of 3326 tiles per slide (minimum 174 tiles, maximum 9500 tiles). We used the Macenko method for color normalization to convert all images to a reference color space for all these tiles (12).

## Tumor Detection and EMT Subtype Detection

For tumor detection in ccRCC, we selected 1000 tiles of necrosis, stroma and tumor from the tiles respectively, and randomly split them into training, validation and testing set at a ratio of 7:1.5:1.5. We trained a convolutional neural network with the Inception V3 model to classify necrosis, stroma, and tumor tiles by transfer learning. This network was then used to detect all the tumor tiles generated from the whole-slide images. This process resulted in 1031666 tiles, with an average of 2982 tiles per slide (minimum 170 tiles, maximum 7930 tiles).

For EMT subtype detection, we trained another Inception v3 model using just the tumor tiles detected before. As we already obtained the EMT subtype for patients' slides using our gene signature, we labeled the tiles with Epi or Mes according to the slides' EMT subtype. That is, the label of tiles was the EMT subtype of the slides the tiles generated from. Due to the small number of Epi slides, training and validation were performed slide-wise using six-fold cross-validation. That is, the tiles associated with the same slide were grouped into the same fold. Because there were different numbers of Epi and Mes slides (6 Epi and 52 Mes in each fold) and different numbers of tumor tiles of each slide, there was an imbalance between these two classes. To minimize such imbalance, while training, we extracted all the Epi slides and downsampled the Mes slides at a ratio of 1:2 (Epi : Mes), and then randomly selected 600 tiles per Epi slide and 300 tiles per Mes slide. However, for slide-wise validation, all the tumor tiles were used. After getting the classification of all tumor tiles of a given slide, the mean prediction value of all these tiles was regarded as the slide's final prediction result.

## Neural Network and Transfer Learning on Inception V3

Both the tumor detection and EMT subtype detection models were based on Inception V3 and trained using transfer learning. All the convolutional neural networks were pretrained on the ImageNet ([www.image-net.org](http://www.image-net.org)) database. Only the last softmax layer was changed and retrained by our dataset. We used Stochastic gradient descent (SGD) optimization with a learning rate of 0.01 for the classification model. All codes were implemented in Python 3.7 and ran on desktop workstations with Nvidia graphics-processing units

(GPUs; NVIDIA 2080). Performance was scored using accuracy, AUC, specificity, and sensitivity.

## Immune Gene Signature Analysis

Immune gene signatures were derived from previously published work (13). RSEM upper quartile normalized, log2 transformed, and mean centered RNA-seq data were matched to predefined immune gene signature clusters *via* Entrez IDs. Each gene signature was included in **Supplementary Table S2**. Differential expression for each gene signature was analyzed between subtypes *via* one-way ANOVA. These P values were adjusted for multiple tests using the Benjamini-Hochberg procedure. For hazard ratio forest plots, a univariate Cox proportional hazards (CoxPH) model was used with signature/clinical variable as a continuous variable compared to patient overall survival.

## CIBERSORT

CIBERSORT is an analytical tool that accurately quantifies the relative levels of distinct immune cell types within a complex gene expression mixture to characterize and quantify each immune cell subtype. Here, we applied the original CIBERSORT gene signature file, LM22, which defined 24 immune cell subtypes, and analyzed datasets from ccRCC. The data were normalized using the cubic spline algorithm. All samples were analyzed for immune cell profiles by CIBERSORT, the number of permutations being set to 100. Twenty-two immune cell types together with CIBERSORT metrics as Pearson correlation coefficient, CIBERSORT P value and root mean squared error (RMSE) were quantified for each sample. CIBERSORT P value reflects the statistical significance of the deconvolution results across all cell subsets and was useful for filtering out deconvolution with less significant fitting accuracy (<https://cibersort.stanford.edu>). Immune cell profile was calculated for each sample, and mean values were calculated.

## ESTIMATE

The ESTIMATE algorithm was applied to the normalized expression matrix for estimating the stromal and immune scores for each clear cell renal cell carcinoma sample. Access to the deidentified linked dataset was obtained from TCGA in accordance with the database policy. For analyses of deidentified data from TCGA database, Institutional Review Board approval and informed consent were not required.

## Quantification and Statistical Analysis

Student T test, Wilcoxon test, and Kruskal-Wallis test were utilized to compare continuous variables and ordered categorical variables, such as mutation load, neoantigen load, HRD score,

CTAs number, and ITH. Permutation test was conducted in the comparison of gene mutation frequencies among clusters. Correlation matrices were created with Pearson or Spearman correlation. Survival analysis was performed using the Kaplan-Meier method, and the survival of the clusters was compared using the log-rank test. For all analyses, significance was determined as a P value < 0.05 and corrected for multiple testing where specified. Univariate analysis was performed unless otherwise specified. Survival analyses were performed using



GraphPad Prism (GraphPad Software, Inc.) or by individually specified methodologies.

## Data Availability

Data are available from the authors upon request but may require data transfer agreements. No personalized health information will be shared.

## Code Availability

The code used during the current study are available from the corresponding author on reasonable request.

## RESULTS

### Discovery of EMT-Related Gene Signature in Patients With ccRCC

An EMT-related gene signature was developed and consisted of 63 genes (**Supplementary Table S1**) and the flowchart of strategy to identify EMT-related biomarkers in ccRCC was shown in **Figure 1A**. Two distinct subtypes of ccRCC, mesenchymal (Mes) and epithelial (Epi), were identified using hierarchical clustering of gene expression data. The Kaplan-Meier plots showed that the patients in the Mes subtype showed worse overall survival than those in the Epi and Intermediate subtypes ( $P = 0.009$ ; **Figure 1B**).

### Validation of the EMT-Related Molecular Subtype

The independent cohort (GSE150404) was used to validate the accuracy of the EMT gene signature ( $P < 0.0001$ ; **Supplementary Figure S1**). Furthermore, higher grade (grade 3/grade 4) and stage (stage III/stage IV) samples were in the Mes group (**Figures 1C, D**). Univariate and multivariate analyses were performed and EMT gene signature was significantly associated with outcome in the multivariate analysis (**Table 1**).

According to TCGA Research Network article, unsupervised clustering methods identified subsets in mRNA (m1-m4) and miRNA (mi1-mi4) expression datasets. In the mRNA cluster, m3 accounted for roughly 50% of samples in the Mes classification and was associated with the worst survival outcome (**Figure 1E**). Survival differences were also evident in miRNA-based subtypes (**Figure 1F**). DNA methylation (me1-me3) expression datasets were detected (**Figure 1G**), and almost one-half of the Mes group was me1 type and associated with significantly poorer survival.

Thorsson et al. identified six immune signature sets (C1-C6) associated with overall survival and progression-free interval (14). C3 had the best prognosis and was enriched in most ccRCC, the more mixed-signature subtypes, C4 and C6, had the least favorable outcome. In our model, compared with Epi, Mes had the most significant proportion of C4+C6 (**Figure 1H**).

Further, we quantified the EMT levels by calculating EMT scores described by Tan et al (15). Positive EMT scores corresponded to the mesenchymal phenotype, whereas negative scores reflected the epithelial phenotype. In general, All Mes subtype was characterized by positive EMT scores corresponding to their phenotype. In contrast, Epi phenotype had intermediate and low EMT scores

(**Figure 1I**), reflecting their partial EMT and complete epithelial states. These results demonstrated the robustness of the prognostic value of EMT signature.

### Deep Learning Can Predict EMT Status Directly From Histopathology Images

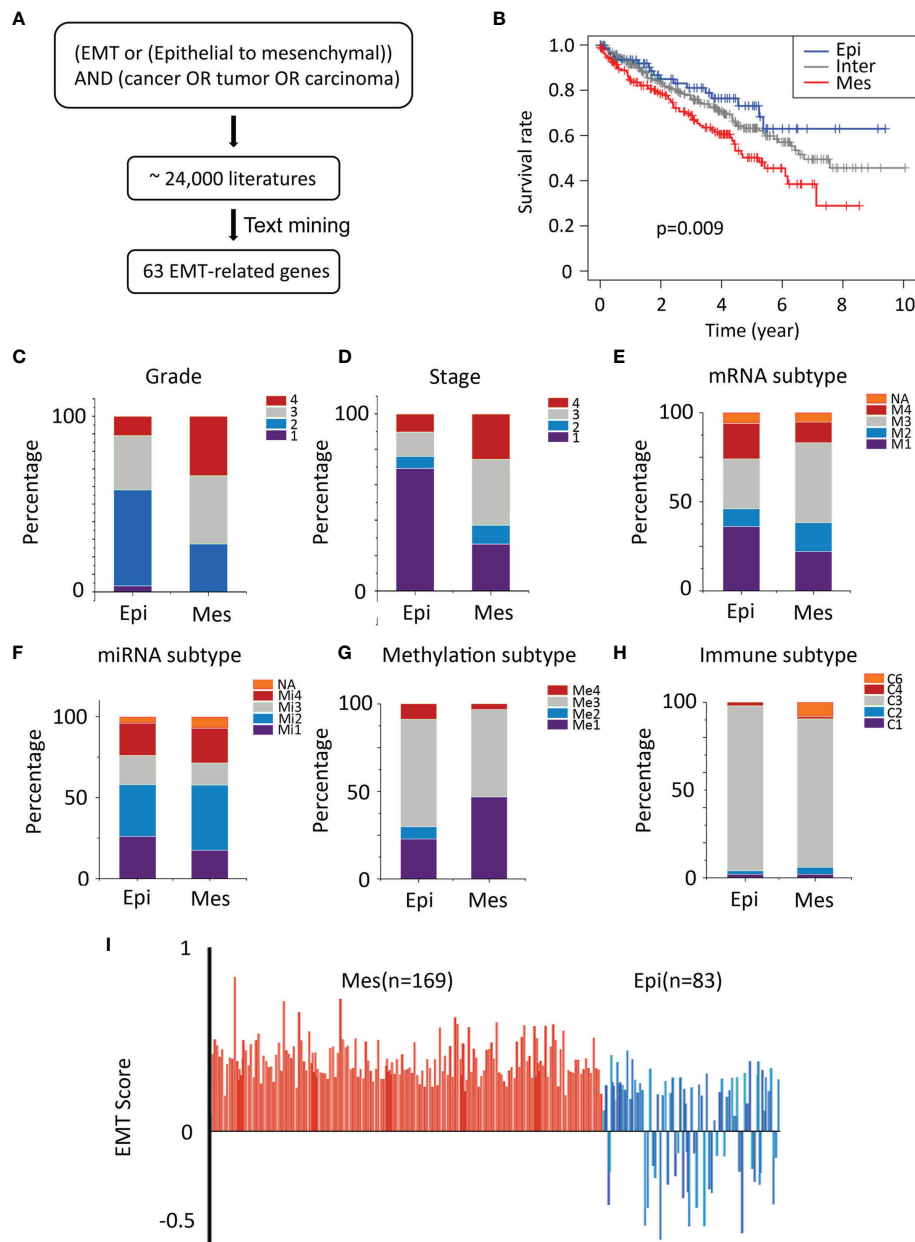
Recently, deep learning has overmatched humans in some medical data processing, especially the ability to predict some genetic information efficiently and economically using histopathology images (16, 17). To facilitate applying the EMT gene signature, we investigated whether deep learning could directly predict EMT status which according to our molecular subtype from H&E-stained slides.

In this study, we used 346 whole-slide images from TCGA and labeled them with the EMT types which according to our gene signature (**Figure 2**). Before classifying the EMT status, 3000 tiles of tumor, necrosis, and stroma were selected and labeled by a urologist and a pathologist. We trained an Inception V3 model to recognize tumor from the other two classes by transfer learning, which yielded an average AUC of 0.99 at the tile level in five independent experiments. Then we used the tumor tiles detected by the model with best performance among the five experiments for EMT subtype classification. For the slide-wise six-fold validation of the EMT subtype, the AUC value of the ROC curve for all classes was 0.84 ( $\pm 0.07$ ). The mean validation accuracy at slide level was 74.90%, mean specificity was 72.23%, and mean sensitivity was 75.32% (**Figures 3A–C**). The predicted EMT status and the classification probability, were aggregated to extract the heatmap for visualization by the tiles' raw position of the slide (**Figure 3D**).

### Histopathological Features Detection for Stratification

To detect which histopathological features were the most relevant in our algorithm to identify molecular subtypes, we used class activation mapping (CAM) (18). For the Epi subtype, these regions had a looser arrangement, big cell gap, nucleoli absent or inconspicuous which were mainly in pink, and granular eosinophilic cytoplasm areas (**Figure 4A**). Several studies suggested that granular eosinophilic cytoplasm reflects the presence of abundant mitochondria, which are required to supply energy for acid secretion (19, 20). For the Mes subtype, the regions were densely packed, surrounded by arborizing vasculature, the large multinucleate cells with empty cytoplasm (different in shape and size), and were often surrounded by abundant immune infiltration (**Figure 4B**).

We further investigated the molecular alterations that could be correlated with the histopathological features. Notably, the genes located in the Mes subtype were annotated to various immune associated pathways and biologic processes (**Supplementary Figure S2**). In addition, we noticed that Mes group exhibited higher artificial intelligence (AI) score ( $> 0.5$  Mes possibility;  $< 0.5$  Epi possibility) and was positively correlated with immune infiltration (Spearman correlation = 0.275,  $P < 0.001$ , **Figure 4B**). Moreover, gene set enrichment analysis (GSEA) validated that abnormality of the mitochondrion and varieties of



**FIGURE 1** | Characteristics of TCGA cohort. **(A)** Text-mining literature abstracts associated with EMT from the PubMed database identified 63 EMT-related genes. **(B)** Overall survival analysis for sample clusters. **(C, D)** General TNM and grade characteristics of patients of the two clusters (left and middle). **(E–G)** Significant differences in mRNA-based, miRNA-based, and methylation classification were identified for both the Epi and Mes clusters. **(H)** The proportion of samples belonging to each immune subtype in two clusters. **(I)** Plot of EMT scores (mean  $\pm$  SEM; y-axis) of in samples. EMT score nearer to +1.0 is more mesenchymal-like (Mes), whereas EMT score nearer to -1.0 is more epithelial-like (Epi), Epi, blue; Mes, red.

metabolic process was enriched in the Epi cluster, which was in accordance with the detection above (**Figure 4A**).

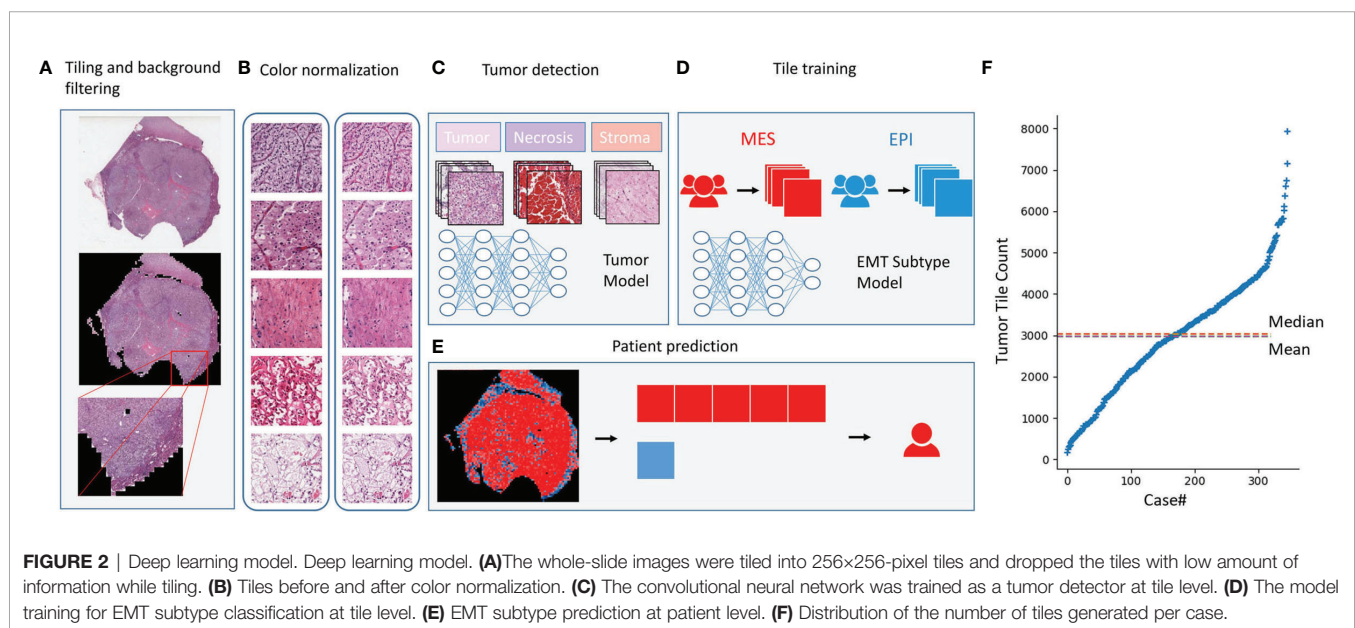
Our tests showed promising results on tumor EMT classification from sections that could be predicted from H&E images. Furthermore, we analyzed the connection between DNA and RNA-level alternations and histopathological features to explore the clinical relevance of the classification.

## Methylation and Metabolism

Heatmapping demonstrated clustering of 89 highly variable DNA methylation markers that were hypermethylated in the tumor. A  $\beta$ -value  $R$  greater than 0.3 was considered hypermethylated, and the markers were associated with significantly poorer survival (13). The methylation Mes subtype had a pronounced increase in hypermethylation across the selected genome (**Figure 5A**).

**TABLE 1 |** Univariate and multivariate Cox proportional hazard regression analysis of overall survival in the cohorts (n = 252).

Variables	Univariate		Multivariate	
	HR (95% CI)	P-value	HR (95% CI)	P-value
EMT signature	1.71 (1.27-2.31)	1.66e-09*	1.22 (1.02-1.54)	7.95e-06*
Age	1.04 (1.02-1.06)	0.00026*	1.03 (1.01-1.05)	0.00091*
Gender	1.14 (0.70-1.83)	0.603	0.904 (0.54-1.5)	0.696
Grade	2.57 (1.90-3.47)	8.13e-10*	1.60 (1.12-2.30)	0.01*
Stage	2.01 (1.65-2.46)	8.56e-12*	1.56 (1.12-2.17)	0.0089*
Metastasis	2.89 (1.96-4.25)	6.87e-08*	0.97 (0.5-1.93)	0.97
Lymph node	3.68 (1.69-8.04)	0.001*	1.86 (0.82-4.23)	0.14

\* $P < 0.05$ .

Evaluation of metabolic differences was performed by 15 major metabolic processes. Expression of the Krebs cycle and the electron transport chain genes (complex I – complex IV) provided a clear distinction between the subtypes (**Figure 5B**).

AMP-activated protein kinase (AMPK) acts as an intracellular energy sensor and was significantly lower in Mes subtype compared to Epi subtype ( $P < 0.01$ ; **Figure 5B**), which negatively regulates fatty acid synthesis and positively regulates mitochondria production (21). The metabolic shift identified in the Mes group is reported to contribute to the Warburg metabolic phenotype, further enhanced malignancy, immune protection of cancer cells (22). Furthermore, the GSEA and AMPK complex genes expression validated some of these results (**Figures 5C, D**).

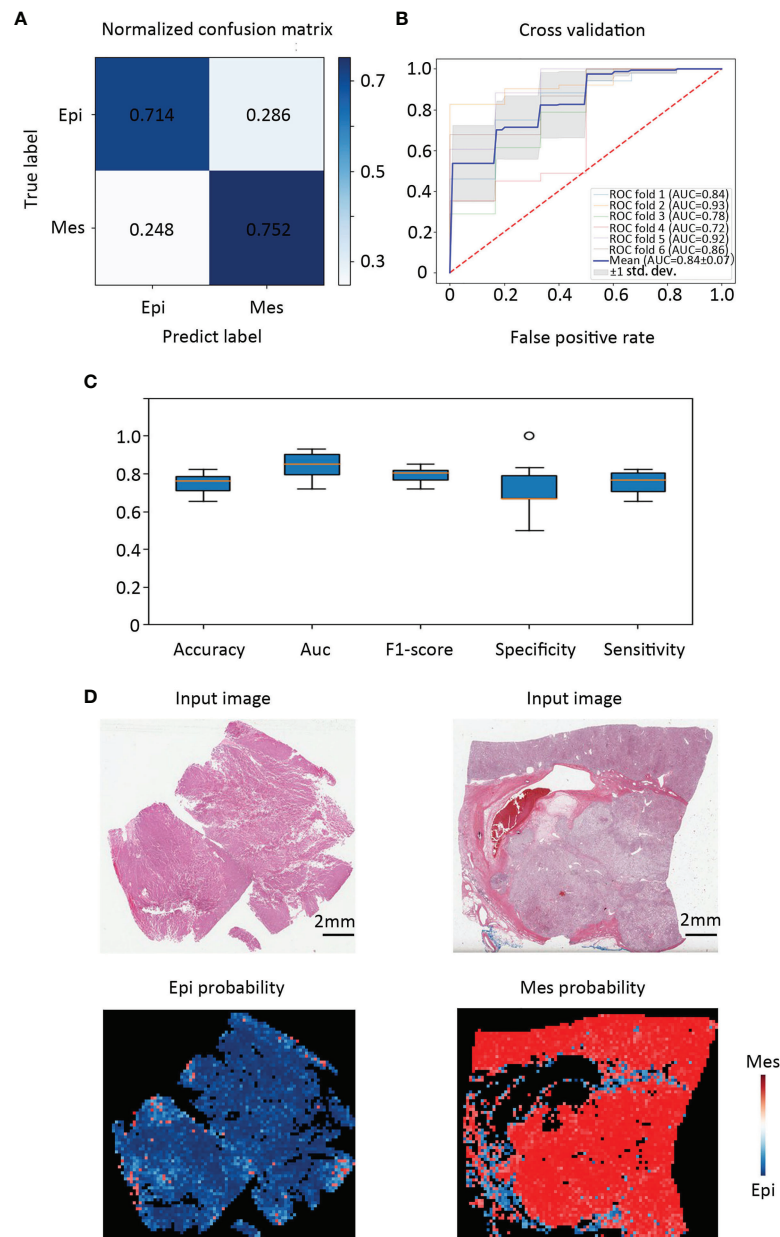
## Mutation and Immune Signature Analysis

The pattern of somatic alterations was determined from the analysis of 104 samples (24 Epi and 80 Mes). Based on previous study methodologies, we identified the eight most significantly mutated genes ( $P < 0.00001$ ) (23). As shown in **Figures 6A, B**, patients with PBRM1 mutation tended to be in the Mes subtype, while the Epi subtype had higher mTOR mutant rates ( $P < 0.05$ ).

PBRM1 encodes the bromodomain-containing protein BAF180, a subtype of the switch/sucrose non-fermentable (SWI/SNF) chromatin remodeling complex (24). Importantly, retrospective data have shown that patients with metastatic ccRCC harboring truncating mutations in PBRM1 experienced increased benefit from immune checkpoint therapy (25, 26). Moreover, the Epi subtype tended to have a better response to the mTOR inhibitor.

To further investigate the relationship between the EMT-associated subtype and the immune, we first established a microenvironment phenotype calculated by ssGSEA (**Supplementary Table S2**). Mes subtype was characterized by high innate and adaptive immune cells infiltration (**Figure 6C**). The ESTIMATE algorithm showed that the Mes subtype had the highest immune and stromal scores for the analyzed cohort ( $P < 0.01$ ; **Figures 6D–G**).

To identify whether different clusters of EMT had distinct tumor extrinsic immune escape mechanisms (27), we focused on the adaptive immune cells in ccRCC (**Figure 6H**). Analysis revealed that the fraction of naïve B cells was higher in Mes subtype tissue than in Epi tissue. The three main T cell subpopulations in tissue (CD4+ memory resting T cells, CD8+



**FIGURE 3 |** Classification performance in an external validation set and class activation maps for morphological features. **(A)** Confusion matrix of the classification results of the validation patients. **(B)** Training and validation on TCGA cohort were performed using six-fold repeated validation with a 90% (training) to 10% (validation) stratified, random split for each fold. **(C)** Classifier performance was assessed including accuracy, F1 score, AUC, specificity, and sensitivity separately for each EMT molecular subtype. **(D)** Tissue slides of patients with Epi and Mes tumors in TCGA test set show the spatial patterns of predicted MSI score. The top panel shows the H&E input image. Corresponding predicted Epi and Mes map for the image shown in the bottom panel. Class activation maps for morphological features.

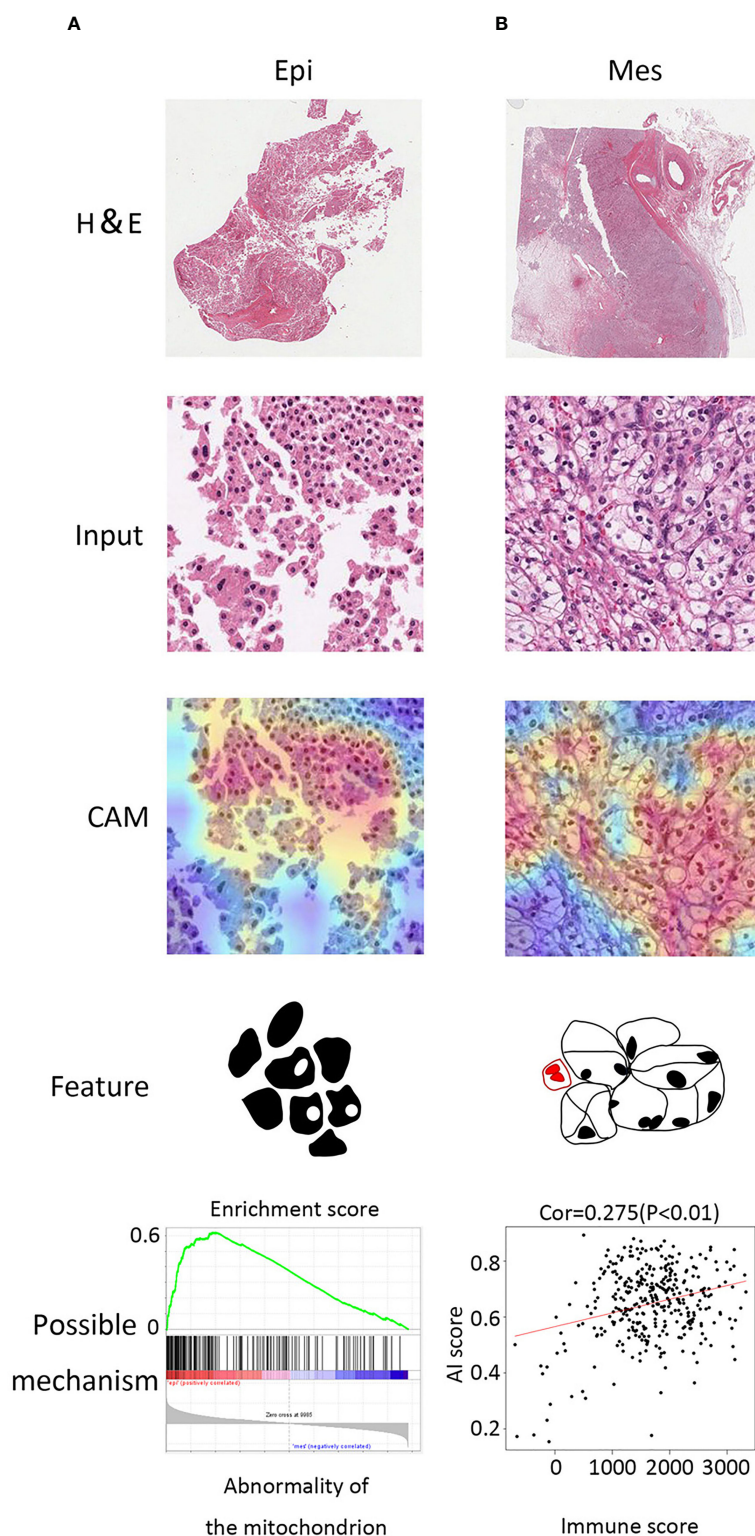
T cells, and follicular helper T cells) were increased in Mes compared to Epi. The results also showed that innate immune components, neutrophils, M2 macrophages and resting NK fractions increased. In general, Mes had abundant active innate and adaptive immune cells and immunosuppressive cells.

Consistent with these results (**Figure 6I**), Mes had a higher expression of chemokines, including CCL4, CXCL9, and

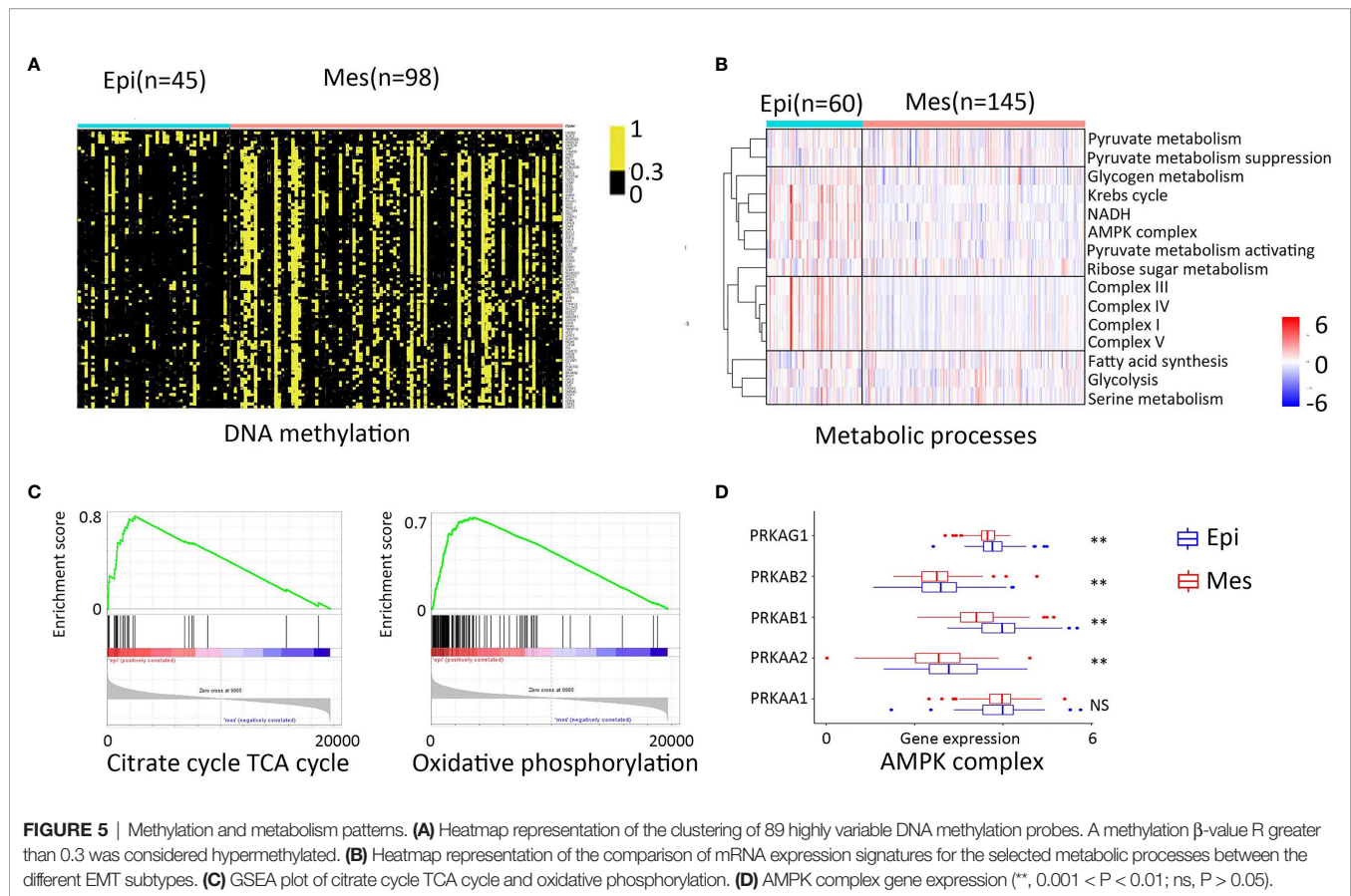
CXCL10, which have been proven to attract dendritic cells and CD8 T cells (28). Overall, the increase of chemokines might contribute to the extrinsic immune escape of the Mes clusters.

We further investigated the two aspects of intrinsic immune escape: tumor immunogenicity and immune checkpoint molecule expression (29). We first compared the factors that were the primary source of tumor antigens: mutation load, neoantigen





**FIGURE 4** | Class activation maps for morphological features. **(A)** Macrograph (top), input image (second from top), CAM (third from top), and corresponding sketch maps of predominant morphological feature (bottom) for EMT subtypes (from left to right: Epi and Mes). Epi region is a granular eosinophilic cytoplasm area with a looser arrangement, big cell gap, nucleoli absent. Mes region contained mostly tumor cells with empty cytoplasm and mixed with immune cells. **(B)** GSEA plot of abnormality of the mitochondrion in Epi cluster (left) and positive correlation between AI score and immune score (estimated by the ESTIMATE algorithm).



load, CTA (cancer-testis antigen) level, HRD (homologous recombination deficiency), ITH (intratumoral heterogeneity), and tumor antigen-presenting capability. In general, the difference in the tumor antigen burden among the clusters was significant between the two groups (**Figures 7A–E**). Considering that high tumor mutation burden (TMB) correlates with a greater probability of displaying tumor neoantigens, it is rational to hypothesize that the tumors with Mes subtype are more likely to respond to immune checkpoint inhibitors (ICIs) as this greater mutation load.

Therefore, we referred to a database of immune checkpoint molecules among clusters (30) (**Figure 7F**). In many cancers, the PD-1 pathway is involved in tumor evasion from immune activity (31). PD-1 (PDCD1) and CTLA-4 (CTLA4) levels were significantly elevated in the Mes subtype as compared with those in the Epi subtype ( $p < 0.01$ ); Furthermore, we demonstrated that PD-1, CTLA4, LAG3, TIGIT, CD80, and CD86 and the expression of most checkpoint molecules were positively correlated (**Figure 7G**). In addition, analysis of overall survival of CTLA4 was positively correlated with poor patient survival (**Supplementary Figure S3**).

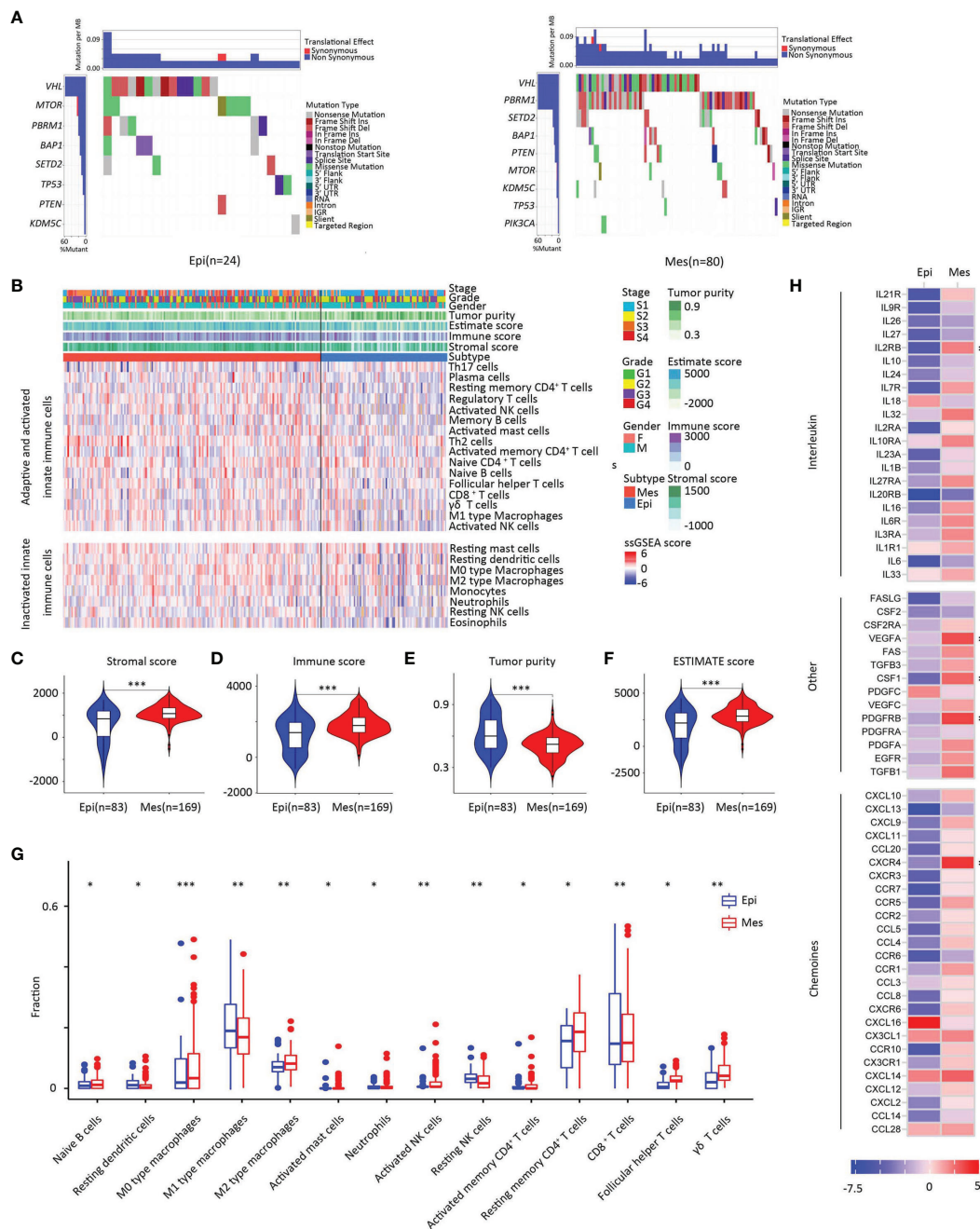
To further test the clinical relevance, a heatmap of genes previously defined and representing angiogenesis and immune biology was investigated (**Figure 7H**) (32). Our data suggested the association of Mes subtype with improved response to antiangiogenic drugs. We further conducted confirmatory studies to evaluate molecular subtypes by deep learning method in 34 stage

IV patients enrolled in the study from 2 January 2020 to 2 June 2021. Patients received more than one dose of pazopanib. In our cohort, only four patients in the pazopanib arm were identified as Epi term, the average score of the remaining patients was 0.72 (95% CI, 0.54–0.83) for Mes tiles and 0.28 (95% CI, 0.16–0.40) for Epi tiles (**Supplementary Table S3**). In Mes-type patients, the objective response rates were 47% (10% complete response; 37% partial response); no Epi-type patients benefitted from the pazopanib. Of note, we observed a trend for improved efficacy (objective response rates) with Mes subtype, indicating that our approach can be beneficial to targeted cancer therapy.

## DISCUSSION

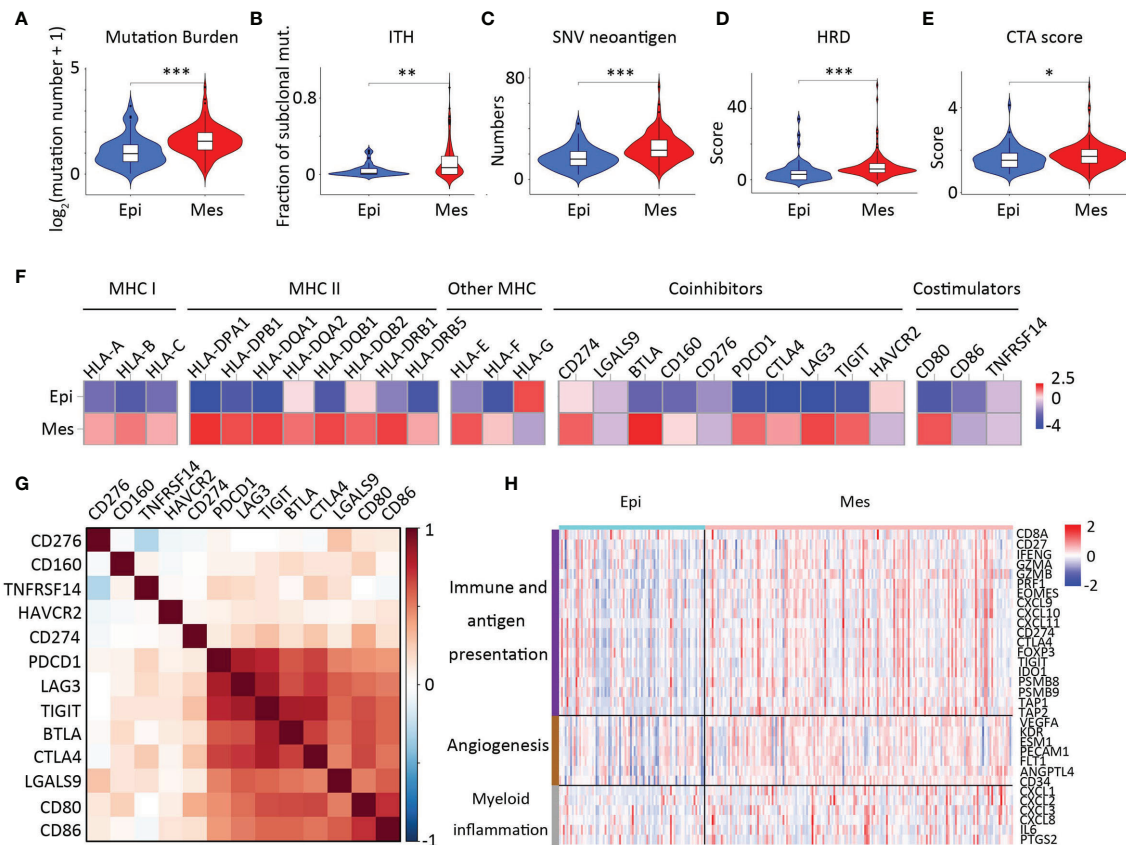
Although pathology divides tumors into distinct ccRCC grades, it is limited in its ability to provide an in-depth analysis (33). Using deep learning methods, our study revealed EMT subtypes and their clinical significance.

In the present study, a robust EMT gene signature clinically significant to patients with ccRCC was developed. Moreover, further analysis demonstrated that Mes and Epi ccRCC subtypes were characterized by distinctive mutations, chromosomal copy number alterations, and mRNA, miRNA, methylation expression, and metabolic patterns. The Mes subtype metabolic alterations were in accordance with the histopathological feature detection. These



**FIGURE 6 |** Somatic alterations in EMT subtypes. (A) Upper heat map, mutation events per sample; lower heat map, mutation types, and events (left, Epi group; right, Mes group). (B) Landscape of the microenvironment phenotypes in EMT subtypes, K-means clustering of microenvironment phenotypes based on the estimated numbers of 24 microenvironment cell subsets calculated by ssGSEA (Epi type, blue; Mes type red). Tumor stage (stage I, light blue; stage II yellow; stage III, orange; stage IV, red), grade (grade 1, green; grade 2, yellow; grade 3, purple; grade 4, red), gender (male, blue; female, pink), and scores estimated by the ESTIMATE algorithm. (C–F) Scores of stromal, immune, tumor purity, and ESTIMATE between clusters. In the violin plots, the mean values are plotted as dots, and the boxplot is drawn inside of the violin plot. (G) Boxplot plot for comparison of the immune cell fraction difference between Epi and Mes tissues. Fractions of each immune cell type were compared by means of a two-sided Mann–Whitney U test. Blue color indicates Epi type and red indicates Mes type. (H) Log2-fold change in mRNA expression in the tumor tissues of chemokines, interleukins, interferons, and other important cytokines and their receptors for each cluster. Molecules with significantly differential expression between the tumor site and the paired normal site ( $P < 0.01$ ) were illustrated. (\*\*\*,  $P < 0.001$ ; \*\*,  $0.001 < P < 0.01$ ; \*,  $0.01 < P < 0.05$ ; ns,  $P > 0.05$ ).





**FIGURE 7 |** Intrinsic immune escape mechanisms of EMT subtypes. Comparison of mutation burden (A), ITH (B), SNV neoantigen (C), HRD scores (D), and CTA numbers (E) among the two clusters. In the violin plots, the mean values are plotted as red dots, and the boxplot was drawn inside the violin plot. (F) Comparison of the log2-fold changes in mRNA expression at the tumor sites of the MHC molecules, costimulators, and coinhibitors for each cluster. For costimulators, only molecules having significantly differential expression between two clusters (except CD274 and CD276,  $P < 0.01$ ) were illustrated. (G) Correlations between expression of immune checkpoint molecules. (H) Heatmap showing expression of genes related to angiogenesis (brown), immune and antigen presentation (purple), and myeloid inflammation (gray) were z-score transformed before visualization.

findings suggested that treatment with demethylating agents and targeting the metabolic pathway could benefit patients with Mes phenotype.

The role of the features in determining the therapeutic responsiveness of ccRCC will be important in future therapeutic planning (34, 35). Studies have reported that EMT are related to immunosuppressive cytokines in several cancer types (36). It has also been reported that EMT can induce PD-L1 expression in non-small cell lung cancer (37). Our study has important implications for clinical translations. First, our results might facilitate the selection of suitable patients for ICIs. We revealed that the Mes cluster demonstrated truncating mutations in PBRM1, and high expression of immune checkpoint molecules might lead to the immune escape of this cluster. Further we suggested that patients in the Mes subtype might respond better to ICIs combined with antiangiogenic therapy. Notably, the CheckMate 9ER trial demonstrated improved progression-free survival and overall survival benefits with the combination of cabozantinib plus nivolumab (38). A Keynote-426 update demonstrated ongoing overall survival benefits of pembrolizumab

plus axitinib in the intention-to-treat population (38). These combination therapies are recommended as first-line treatment for advanced kidney cancer. Therefore, our data suggest that EMT signature-based biomarkers may be valuable for identifying patients who can benefit from immune checkpoint blockade and antiangiogenesis agents.

However, the EMT molecular subtype analysis in ccRCC specimens is far from being implemented in routine pathology due to the high costs. We demonstrated how the EMT gene signature could further benefit from deep learning by presenting a strategy based on convolutional neural networks. The development of inexpensive and more powerful technologies has made the training of larger and more complex neural networks possible.

## CONCLUSIONS

In conclusion, our analysis showed that the EMT molecular subtype of ccRCC specimens *via* deep-learning convolutional



neural networks could be an extremely useful tool for patient-tailored therapy strategies.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Chinese Clinical Trial Registry. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DY had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of

the data analysis. Study concept and design: QC. Acquisition of data: YK, SW, XZ, and HW. Analysis and interpretation of data: LC, QC, YK, SW, XZ, HW, and WL. Drafting of the manuscript: QC and LC. Critical revision of the manuscript for important intellectual content: DY and LC. Statistical analysis: QC and WL. Administrative, technical, or material support: YK. Supervision: HW. All authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.782515/full#supplementary-material>

**Supplementary Figure 1** | Kaplan-Meier plots of OS of the cohort.

**Supplementary Figure 2** | GSEA plot of cytokine receptor interaction, JAK-STAT pathway, chemokine signaling pathway, regulation of alpha beta T cell activation and differentiation in Mes cluster.

**Supplementary Figure 3** | Kaplan-Meier plots of OS of the different expression levels of immune checkpoint molecules.

## REFERENCES

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
- Ljungberg B, Albiges L, Abu-Ghanem Y, Bensalah K, Dabestani S, Fernández-Pello S, et al. European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur Urol* (2019) 75:799–810. doi: 10.1016/j.eururo.2019.02.011
- Dongre A, Weinberg RA. New Insights Into the Mechanisms of Epithelial-Mesenchymal Transition and Implications for Cancer. *Nat Rev Mol Cell Biol* (2019) 20:69–84. doi: 10.1038/s41580-018-0080-4
- Dongre A, Rashidian M, Eaton EN, Reinhardt F, Thiru P, Zagorulya M, et al. Direct and Indirect Regulators of Epithelial-Mesenchymal Transition-Mediated Immunosuppression in Breast Carcinomas. *Cancer Discov* (2021) 11:1286–305. doi: 10.1158/2159-8290.CD-20-0603
- Ravi R, Noonan KA, Pham V, Bedi R, Zhavoronkov A, Ozerov IV, et al. Bifunctional Immune Checkpoint-Targeted Antibody-Ligand Traps That Simultaneously Disable Tgfb Enhance the Efficacy of Cancer Immunotherapy. *Nat Commun* (2018) 9:741. doi: 10.1038/s41467-017-02696-6
- Ruiz de Galarreta M, Bresnahan E, Molina-Sánchez P, Lindblad KE, Maier B, Sia D, et al.  $\beta$ -Catenin Activation Promotes Immune Escape and Resistance to Anti-PD-1 Therapy in Hepatocellular Carcinoma. *Cancer Discov* (2019) 9:1124–41. doi: 10.1158/2159-8290.CD-19-0074
- Albaradei S, Thafar M, Alsaedi A, Van Neste C, Gojobori T, Essack M, et al. Machine Learning and Deep Learning Methods That Use Omics Data for Metastasis Prediction. *Comput Struct Biotechnol J* (2021) 19:5008–18. doi: 10.1016/j.csbj.2021.09.001
- Qiu Y, Jiang H, Ching WK. Unsupervised Learning Framework With Multidimensional Scaling in Predicting Epithelial-Mesenchymal Transitions. *IEEE/ACM Trans Comput Biol Bioinf* (2020) 18(6):2714–23. doi: 10.1109/TCBB.2020.2992605
- Zhang Z, Chen L, Humphries B, Brien R, Wicha MS, Luker KE, et al. Morphology-Based Prediction of Cancer Cell Migration Using an Artificial Neural Network and a Random Decision Forest. *Integr Biol: Quant Biosci Nano Macro* (2018) 10(12):758–67. doi: 10.1039/C8IB00106E
- Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep Learning can Predict Microsatellite Instability Directly From Histology in Gastrointestinal Cancer. *Nat Med* (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y
- Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and Mutation Prediction From non-Small Cell Lung Cancer Histopathology Images Using Deep Learning. *Nat Med* (2018) 24(10):1559–67. doi: 10.1038/s41591-018-0177-5
- Fernández-Carrobles MM, Bueno G, García-Rojo M, González-López L, López C, Déniz O. Automatic Quantification of IHC Stain in Breast TMA Using Colour Analysis. *Comput Med Imaging Graph: Off J Comput Med Imaging Soc* (2017) 61:14–27. doi: 10.1016/j.compmedimag.2017.06.002
- Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* (2018) 23:313–26.e5. doi: 10.1016/j.celrep.2018.03.075
- Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
- Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, et al. Epithelial-Mesenchymal Transition Spectrum Quantification and its Efficacy in Deciphering Survival and Drug Responses of Cancer Patients. *EMBO Mol Med* (2014) 6:1279–93. doi: 10.15252/emmm.201404208
- Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, et al. Multi-Institutional Validation of Deep Learning for Pretreatment Identification of Extranodal Extension in Head and Neck Squamous Cell Carcinoma. *J Clin Oncol* (2020) 38:1304–11. doi: 10.1200/JCO.19.02031
- Jiao W, Atwal G, Polak P, Karlic R, Cuppen E, Danyi A, et al. A Deep Learning System Accurately Classifies Primary and Metastatic Cancers Using Passenger Mutation Patterns. *Nat Commun* (2020) 11:728. doi: 10.1038/s41467-019-13825-8
- Shi J, Chen D, Wang MJS. Pre-Impact Fall Detection With CNN-Based Class Activation Mapping Method. *Sensors (Basel)* (2020) 20(17):4750. doi: 10.3390/s20174750
- Chan JK. The Wonderful Colors of the Hematoxylin-Eosin Stain in Diagnostic Surgical Pathology. *Int J Surg Pathol* (2014) 22:12–32. doi: 10.1177/1066896913517939
- Damiani S, Dina R, Eusebi V. Eosinophilic and Granular Cell Tumors of the Breast. *Semin Diagn Pathol* (1999) 16:117–25. doi: 10.1158/0008-5472.CAN-19-1023

21. Qu YY, Zhao R, Zhang HL, Zhou Q, Xu FJ, Zhang X, et al. Inactivation of the AMPK-GATA3-ECHS1 Pathway Induces Fatty Acid Synthesis That Promotes Clear Cell Renal Cell Carcinoma Growth. *Cancer Res* (2020) 80:319–33.
22. Curran CS, Kopp JB. PD-1 Immunobiology in Glomerulonephritis and Renal Cell Carcinoma. *BMC Nephrol* (2021) 22:80. doi: 10.1186/s12882-021-02257-6
23. Chad JC, Margaret M, Preethi HG, David AW, Richard AG, Gordon R, et al. Comprehensive Molecular Characterization of Clear Cell Renal Cell Carcinoma. *Nature* (2013) 499:43–9. doi: 10.1038/nature12222
24. Kim JY, Lee SH, Moon KC, Kwak C, Kim HH, Keam B, et al. The Impact of PBRM1 Expression as a Prognostic and Predictive Marker in Metastatic Renal Cell Carcinoma. *J Urol* (2015) 194:1112–9. doi: 10.1016/j.juro.2015.04.114
25. Braun DA, Ishii Y, Walsh AM, Van Allen EM, Wu CJ, Shukla SA, et al. Clinical Validation of PBRM1 Alterations as a Marker of Immune Checkpoint Inhibitor Response in Renal Cell Carcinoma. *JAMA Oncol* (2019) 5:1631–3. doi: 10.1001/jamaoncol.2019.3158
26. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, et al. Genomic Correlates of Response to Immune Checkpoint Therapies in Clear Cell Renal Cell Carcinoma. *Science* (2018) 359:801–6. doi: 10.1126/science.aan5951
27. Spranger S, Gajewski TF. Impact of Oncogenic Pathways on Evasion of Antitumor Immune Responses. *Nat Rev Cancer* (2018) 18:139–47. doi: 10.1038/nrc.2017.117
28. Fu C, Jiang A. Dendritic Cells and CD8 T Cell Immunity in Tumor Microenvironment. *Front Immunol* (2018) 9:3059. doi: 10.3389/fimmu.2018.03059
29. Pitt JM, Vétizou M, Daillère R, Roberti MP, Yamazaki T, Routy B, et al. Resistance Mechanisms to Immune-Checkpoint Blockade in Cancer: Tumor-Intrinsic and -Extrinsic Factors. *Immunity* (2016) 44:1255–69. doi: 10.1016/j.immuni.2016.06.001
30. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, et al. Multi-Omics Profiling Reveals Distinct Microenvironment Characterization and Suggests Immune Escape Mechanisms of Triple-Negative Breast Cancer. *Clin Cancer Res: Off J Am Assoc Cancer Res* (2019) 25(16):5002–14. doi: 10.1158/1078-0432.CCR-18-3524
31. Burrack AL, Spartz EJ, Raynor JF, Wang I, Olson M, Stromnes IM. Combination PD-1 and PD-L1 Blockade Promotes Durable Neoantigen-Specific T Cell-Mediated Immunity in Pancreatic Ductal Adenocarcinoma. *Cell Rep* (2019) 28:2140–55.e6. doi: 10.1016/j.celrep.2019.07.059
32. McDermott DF, Huseni MA, Atkins MB, Motzer RJ, Rini BI, Escudier B, et al. Clinical Activity and Molecular Correlates of Response to Atezolizumab Alone or in Combination With Bevacizumab Versus Sunitinib in Renal Cell Carcinoma. *Nat Med* (2018) 24:749–57. doi: 10.1038/s41591-018-0053-3
33. Serie DJ, Joseph RW, Cheville JC, Ho TH, Parasramka M, Hilton T, et al. Clear Cell Type A and B Molecular Subtypes in Metastatic Clear Cell Renal Cell Carcinoma: Tumor Heterogeneity and Aggressiveness. *Eur Urol* (2017) 71:979–85. doi: 10.1016/j.eururo.2016.11.018
34. Ross-Macdonald P, Walsh AM, Chasalow SD, Ammar R, Papillon-Cavanagh S, Szabo PM, et al. Molecular Correlates of Response to Nivolumab at Baseline and on Treatment in Patients With RCC. *J Immunother Cancer* (2021) 9(3):e001506. doi: 10.1136/jitc-2020-001506
35. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, et al. Interplay of Somatic Alterations and Immune Infiltration Modulates Response to PD-1 Blockade in Advanced Clear Cell Renal Cell Carcinoma. *Nat Med* (2020) 26:909–18. doi: 10.1038/s41591-020-0839-y
36. Martinelli P, Carrillo-de Santa Pau E, Cox T, Sainz B Jr, Dusetti N, Greenhalf W, et al. GATA6 Regulates EMT and Tumour Dissemination, and is a Marker of Response to Adjuvant Chemotherapy in Pancreatic Cancer. *Gut* (2017) 66:1665–76. doi: 10.1136/gutjnl-2015-311256
37. Chae YK, Chang S, Ko T, Anker J, Agte S, Iams W, et al. Epithelial-Mesenchymal Transition (EMT) Signature is Inversely Associated With T-Cell Infiltration in non-Small Cell Lung Cancer (NSCLC). *Sci Rep* (2018) 8:2918. doi: 10.1038/s41598-018-21061-1
38. Bedke J, Albiges L, Capitanio U, Giles RH, Hora M, Lam TB, et al. Updated European Association of Urology Guidelines on Renal Cell Carcinoma: Nivolumab Plus Cabozantinib Joins Immune Checkpoint Inhibition Combination Therapies for Treatment-Naïve Metastatic Clear-Cell Renal Cell Carcinoma. *Eur Urol* (2021) 79:339–42. doi: 10.1016/j.eururo.2020.12.005

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AG declared a past co-authorship with one of the authors LC to the handling editor.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Kuai, Wang, Zhu, Wang, Liu, Cheng and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Identification of a Hypoxia-Related Gene Model for Predicting the Prognosis and Formulating the Treatment Strategies in Kidney Renal Clear Cell Carcinoma

Xiang-hui Ning<sup>1\*</sup>, Ning-yang Li<sup>1</sup>, Yuan-yuan Qi<sup>2</sup>, Song-chao Li<sup>1</sup>, Zhan-kui Jia<sup>1</sup> and Jin-jian Yang<sup>1\*</sup>

<sup>1</sup> Department of Urology, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, <sup>2</sup> Department of Nephrology, the First Affiliated Hospital, Zhengzhou University, Zhengzhou, China

## OPEN ACCESS

### Edited by:

Jonathan Olivier,  
Université de Lille,  
France

### Reviewed by:

Di Gu,  
First Affiliated Hospital of Guangzhou  
Medical University, China  
Pasquale Ditunno,  
University of Bari, Italy

### \*Correspondence:

Xiang-hui Ning  
ningxianghui@126.com  
Jin-jian Yang  
yangjinjian2011@163.com

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 31 October 2021

**Accepted:** 23 December 2021

**Published:** 24 January 2022

### Citation:

Ning X-h, Li N-y, Qi Y-y, Li S-c, Jia Z-k  
and Yang J-j (2022) Identification of a  
Hypoxia-Related Gene Model for  
Predicting the Prognosis and  
Formulating the Treatment Strategies  
in Kidney Renal Clear Cell Carcinoma.  
Front. Oncol. 11:806264.  
doi: 10.3389/fonc.2021.806264

**Purpose:** The present study aimed to establish a hypoxia related genes model to predict the prognosis of kidney clear cell carcinoma (KIRC) patients using data accessed from The Cancer Genome Atlas (TCGA) database and International Cancer Genome Consortium (ICGC) database.

**Methods:** Patients' data were downloaded from the TCGA and ICGC databases, and hypoxia related genes were accessed from the Molecular Signatures Database. The differentially expressed genes were evaluated and then the differential expressions hypoxia genes were screened. The TCGA cohort was randomly divided into a discovery TCGA cohort and a validation TCGA cohort. The discovery TCGA cohort was used for constructing the hypoxia genes risk model through Lasso regression, univariate and multivariate Cox regression analysis. Receiver operating characteristic (ROC) curves were used to assess the reliability and sensitivity of our model. Then, we established a nomogram to predict the probable one-, three-, and five-year overall survival rates. Lastly, the Tumor Immune Dysfunction and Exclusion (TIDE) score of patients was calculated.

**Results:** We established a six hypoxia-related gene prognostic model of KIRC patients in the TCGA database and validated in the ICGC database. The patients with high riskscore present poorer prognosis than those with low riskscore in the three TCGA cohorts and ICGC cohort. ROC curves show our six-gene model with a robust predictive capability in these four cohorts. In addition, we constructed a nomogram for KIRC patients in the TCGA database. Finally, the high risk-group had a high TIDE score than the patients with low riskscore.

**Conclusions:** We established a six hypoxia-related gene risk model for independent prediction of the prognosis of KIRC patients was established and constructed a robust nomogram. The different riskscores might be a biomarker for immunotherapy strategy.

**Keywords:** kidney cancer, hypoxia gene, prognosis, immunotherapy, nomogram

## INTRODUCTION

Renal cell carcinoma is one of the most common malignant tumors in the urinary system (1). Clear cell renal cell carcinoma (ccRCC) is the most common type of RCC, and accounts for approximately 80% of cases, with a 10-year cancer specific-survival rate of 62% (2). Radical nephrectomy is the standard surgical treatment for localized and locally advanced ccRCC. However, there is a lack of sufficient effective treatment for metastatic ccRCC and its 5-year survival rate of 12% (1). Now the immune checkpoint inhibition agents have been recommended as the first line agents for metastatic RCC instead of targeted therapy (3). Constructing a prognostic model that can accurately predict the prognosis has become a key issue in the diagnosis and formulation of treatment strategies for renal cancer.

Nowadays, a few prognostic systems and nomograms have been developed based on the clinical characters (such as TNM stage and Fuhrman grade) and laboratory factors (such as hemoglobin, neutrophil count, and platelet count) (2). With the development of high-throughput sequencing technology, the nomogram based on the RNA-seq data or methylation data have been constructed (4–9). These nomograms have not only improved the capability of predicting the prognosis of ccRCC but have also indicated the pathogenic mechanism of ccRCC which may help to develop drug candidates.

VHL gene mutation plays a central role in the initiation and progression of ccRCC (10, 11). As VHL gene is a key regulator of hypoxia, and hypoxia related genes can facilitate tumorigenesis (12), whether the hypoxia pathway could predict the prognosis of ccRCC was interesting. Now, we used the TCGA and ICGC data to investigate the role of hypoxia pathway in predicting the prognosis of ccRCC, and our results constructed a hypoxia related genes model and a nomogram to improve the prognostic value of ccRCC based on bioinformatics approaches. We also investigate the role of our prognosis model in the immunotherapy strategy.

## MATERIALS AND METHODS

### Patients and Public Datasets

RNA-seq data of 533 KIRC and 78 adjacent normal kidney tissues were downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) in October 2018. These original data were processed according to the guidelines and the data were merged into a single expression file for further study. All these expression data were used for screening the differential expression genes between normal and KIRC tissues. The clinical data of KIRC patients were also obtained, and 512 patients who had identified information of age, gender, clinical stage, TNM stage, overall survival status, and survival time were finally enrolled in the prognosis study (**Supplementary File; Supplementary Table**). The International Cancer Genome Consortium (ICGC) cohort (RECA-EU project, 91 patients) clinical information and RNA

expression data were downloaded from the ICGC data portal in April 2020 (**Supplementary File, Supplementary Table**). The TCGA was used as a public open database, and the relevant information retrieved from it did not require further ethical approval.

### Screening of Differential Expression Hypoxia Genes (DEHGs)

RNA-seq data were standardized using the “limma” package of R software. Then differentially expressed genes (DEGs) were evaluated using the “DESeq” package. Genes with adjusted p-value <0.05 and absolute log2 fold change (FC) >2 in normal and tumors tissues were considered as DEGs. The hypoxia related genes were downloaded from the Molecular Signatures Database ([https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK\\_HYPOXIA.html](https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_HYPOXIA.html)). The hypoxia related genes in DEGs and the ICGC genes list were considered as differential expression hypoxia genes (DEHGs).

### Prognosis Related Genes Filtering and Gene Risk Model Construction

The whole 512 patients were named Total TCGA cohort. Total TCGA cohort was then divided into a discovery TCGA cohort (256 patients) and a validation TCGA cohort (256 patients) by a ratio of 1 to 1 randomly (**Supplementary File**). The discovery TCGA cohort was used for constructing the hypoxia genes risk model. Firstly, univariate Cox expression was used to assess the prognosis role of these DEHGs. After identifying the prognosis related DEHGs, LASSO (least absolute shrinkage and selection operator) regression was used to select a panel of genes, and the analysis was performed using “glmnet” package in R software. Lastly, the genes which were extracted by lasso regression, were enrolled in multivariate Cox regression to further screen the prognosis related DEHGs and then construct a risk model. Riskscore was calculated according to the coefficient and expression value of each DEHG which had the significant meaning in multivariate Cox regression.

### Genes Risk Model Validation, Nomogram Construction and TIDE Score Calculation

Risk-groups were deemed as two categorical variables (high risk-group and low risk-group) according to the median value of riskscore. Kaplan–Meier curve and log-rank test were used to assess the prognosis role of risk-group. Next, univariate and multivariate Cox regression were used to identify whether riskscore was an independent factor for the prognosis of KIRC patients. In the Cox regression analysis, gender, T stage, N stage, M stage, clinical stage, and risk-group were considered as categorical variables. Meanwhile, age and riskscore were considered as continuous variables. Validation TCGA cohort, total TCGA cohort, and ICGC cohort underwent the same analysis procedure to validate the genes risk model. Moreover, Receiver operating characteristic curve (ROC) analysis was performed to assess the predictive accuracy of the riskscore in all these four cohorts. In addition, a nomogram was established based on the total TCGA cohort for clinical application. Statistical



analyses were performed using R software. Tumor Immune Dysfunction and Exclusion (TIDE) score was calculated online ([HTTP://tide.dfci.harvard.edu/](http://tide.dfci.harvard.edu/)) (13). All tests were two-tailed, and a P-value of <0.05 was considered statistically significant.

## RESULTS

### Data Processing and Screening of Differentially Expressed Genes

A total of 5,825 DEGs were screened (3,370 upregulated and 2,455 downregulated) from a total of 34,827 genes using the adjust p-value <0.05 and log FC (fold change) >1 threshold (**Supplementary File**). Two hundred hypoxia-related genes were acquired from the Molecular Signatures Database. Finally, 72 genes DEGs were identified both in total TCGA and ICGC cohorts (**Supplementary File; Supplementary Figure**).

### Screening of Prognostic DEGs

The processed survival data of each tumor sample in the training set were subjected to univariate Cox proportional hazards regression analysis, in which the significant threshold was set at value <0.05. Therefore, 15 prognosis-related DEGs containing 10 risky genes and 5 protective genes were identified (**Supplementary File**).

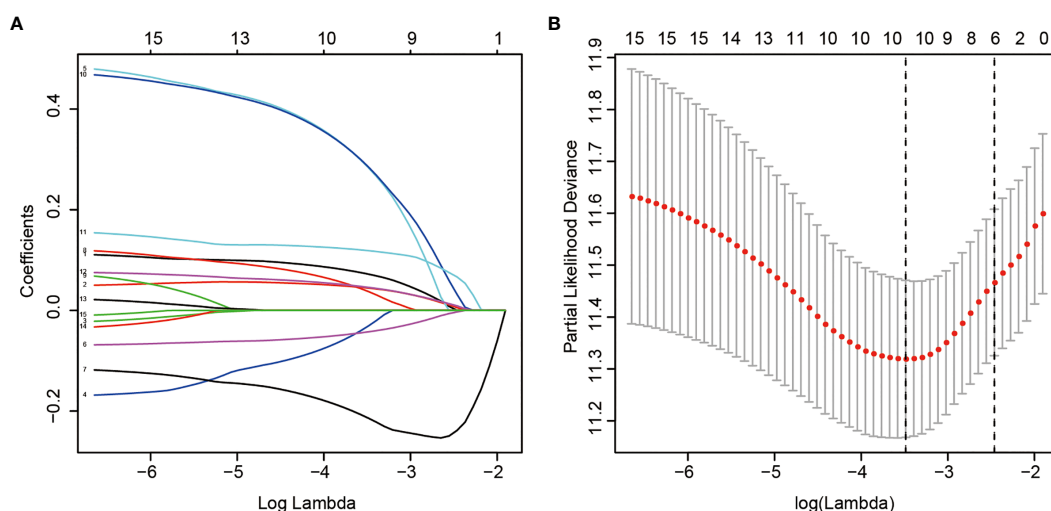
### Establishment and Validation of the 6-DEHGs Prognosis Model

LASSO regression with tenfold cross validation was performed to further screen the DEHGs that significantly correlated with the prognosis of ccRCC patients. The optimal lambda value was obtained from the minimum partial likelihood deviance (**Figures 1A, B**). Then, the optimal 6-DEHG model was

obtained which contained *GPC3*, *KIF5A*, *PLAUR*, *ANKZF1*, *ETS1*, and *SELENBP1*. The Cox coefficients of the DEHGs were obtained from the multivariate Cox proportional hazards regression analysis. Then, the riskscore was constructed based on the coefficients and categorical values of expression level as the following:  $\text{Riskscore} = (0.1611 * \text{GPC3}) + (0.0781 * \text{KIF5A}) + (0.2598 * \text{PLAUR}) + (0.4296 * \text{ANKZF1}) - (0.2765 * \text{ETS1}) + (0.4578 * \text{SELENBP1})$ . The riskscore of each tumor patient in the discovery TCGA cohort was calculated and the patients were divided into a high risk-group and a low risk-group based on the median riskscore (**Figure 2**). Then, the other three cohorts were also divided into a high risk-group and a low risk-group based on the median riskscore of each cohort individually. Patients with the high risk-group presented with poorer survival (**Figure 3A**) and it had been validated in the other three cohorts (**Figures 3B–D**).

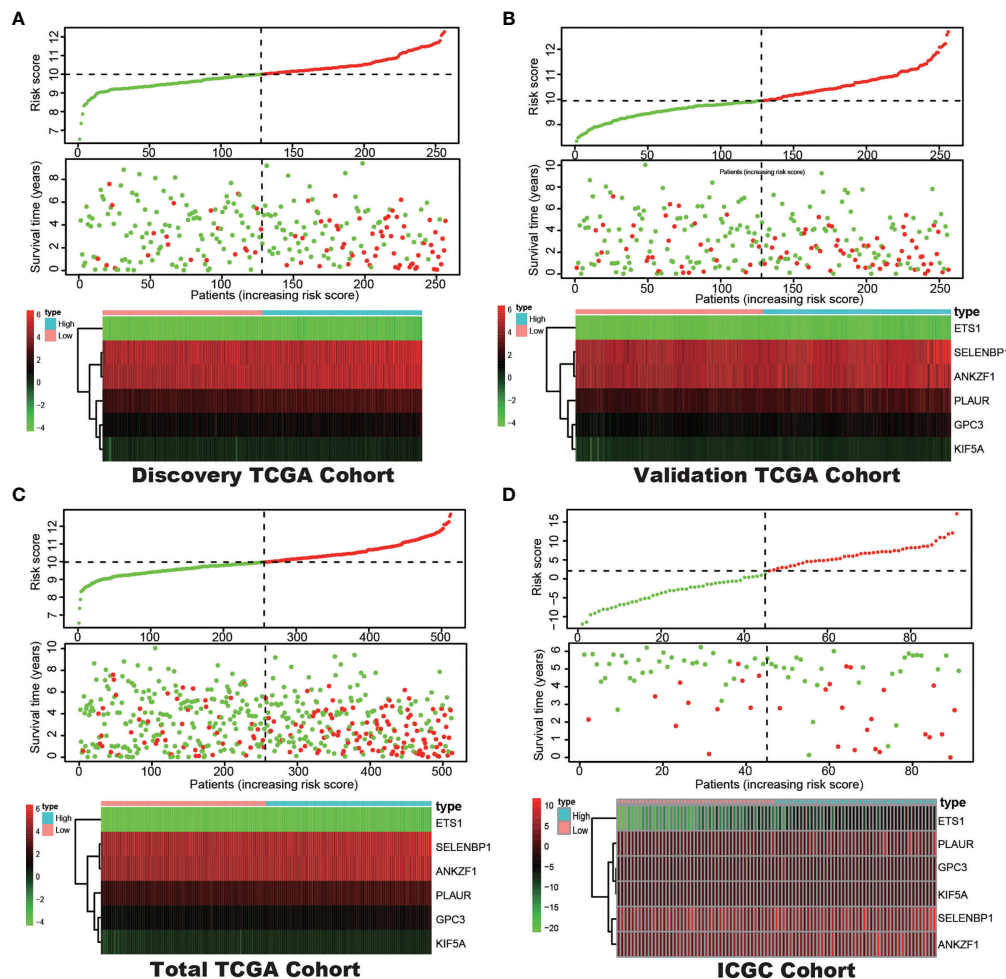
### Construction of a 6-DEHG Prognosis Model-Based Nomogram

In addition, Cox regression analysis remained that riskscore was an independent prognostic factor influencing patients with ccRCC in these four cohorts (**Figure 4, Table 1**). The AUC value for the DEHG Prognosis model was 0.711 in the 1-year ROC curve, 0.708 in the 3-year ROC curve, and 0.779 in the 5-year ROC curve of discovery TCGA cohort (**Figure 5A**). The AUC value of the other three validation cohorts is shown in **Figures 5B–D**. The riskscore, patients' age, and clinical stage could be an independent prognostic factor, respectively, for OS prediction of ccRCC patients in the training set after the univariate and multivariate Cox proportional hazards regression analyses. Then, these independent prognostic factors were integrated together into this nomogram to predict the 1-, 3-, and 5-year OS of ccRCC patients (**Figure 6**).



**FIGURE 1** | LASSO regression analysis of 15 DEHGs. Ten-fold cross-validation was applied to calculate the best lambda, which leads to a minimum mean cross-validated error (**A**). A total of 6 DEHGs were adopted for the LASSO model (**B**).





**FIGURE 2** | Correlation between the prognostic model and the overall survival (OS) of patients in the Discovery TCGA Cohort (A), the Validation TCGA Cohort (B), the Total TCGA Cohort (C) and the ICGC Cohort (D). The distribution of riskscore (upper), survival time (middle) and hypoxia genes expression levels (below). Patients were classified into low-risk and high-risk groups by using the median score as a cut-off value. The red dots and lines represent the patients in high-risk groups. The green dots and lines represent the patients in low-risk groups.

## The Probably Benefit of Patients for Immune Checkpoint Inhibitor (ICI) Therapy in Different Risk Groups

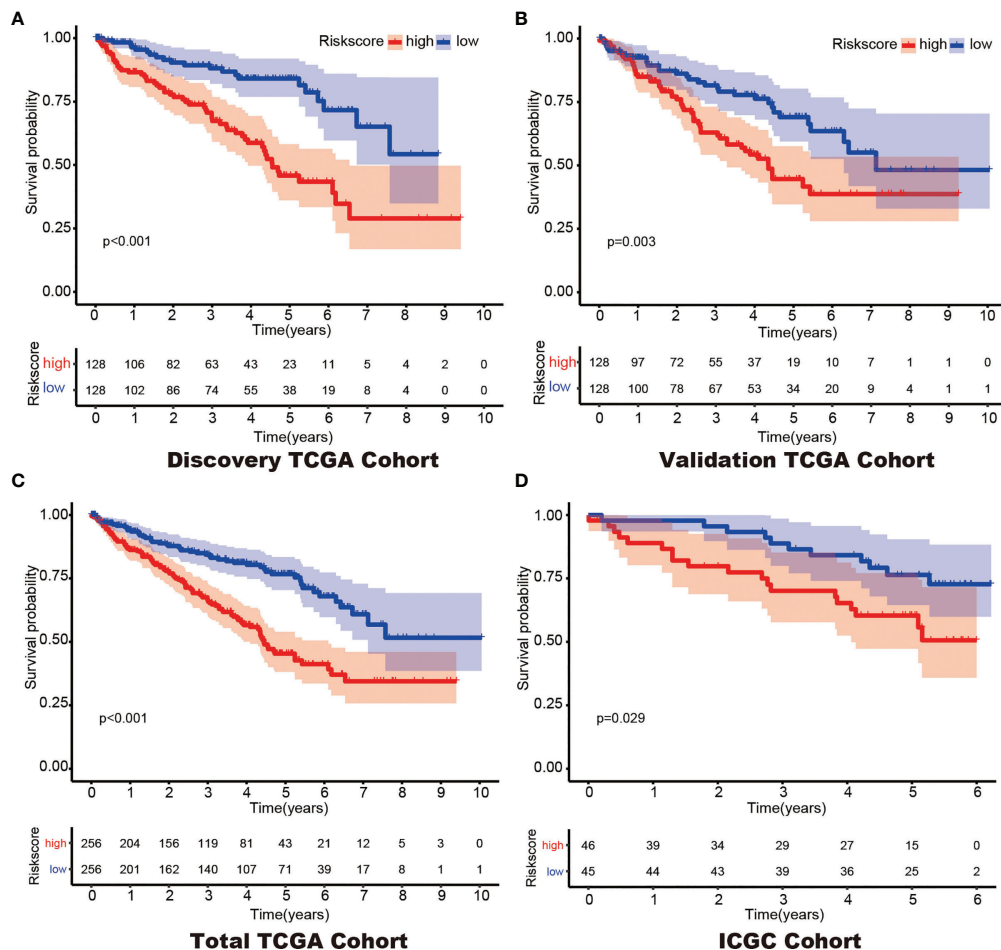
The TIDE-score was used to assess the potential clinical efficacy of immunotherapy in different risk groups. In our results, the high-risk group had a higher TIDE score than the low risk-group (Figure 7A). Also, we found that the high risk-group had a lower microsatellite instability (MSI) score, while the high risk-group had a higher T cell dysfunction score, but there was no difference in T cell exclusion between the two groups (Figures 7B–D).

## DISCUSSION

To achieve the precision medicine of kidney cancer, accurate diagnosis, individualized therapeutic strategies are needed.

Providing the detail prognosis information for patients make the precision medicine more meaningful. Hypoxia pathway plays the central function in ccRCC pathogenesis and development. In our study, we screened the hypoxia-related genes for predicting the prognosis of ccRCC and then constructed a six hypoxia-related gene prognosis model using the LASSO regression method. The model was validated by the validation TCGA cohort, the total TCGA cohort, and the ICGC cohort. A nomogram, based on the clinical features and our model, was built for predicting the overall survival probability of ccRCC patients. In addition, we found the different immune function statuses in the different risk groups.

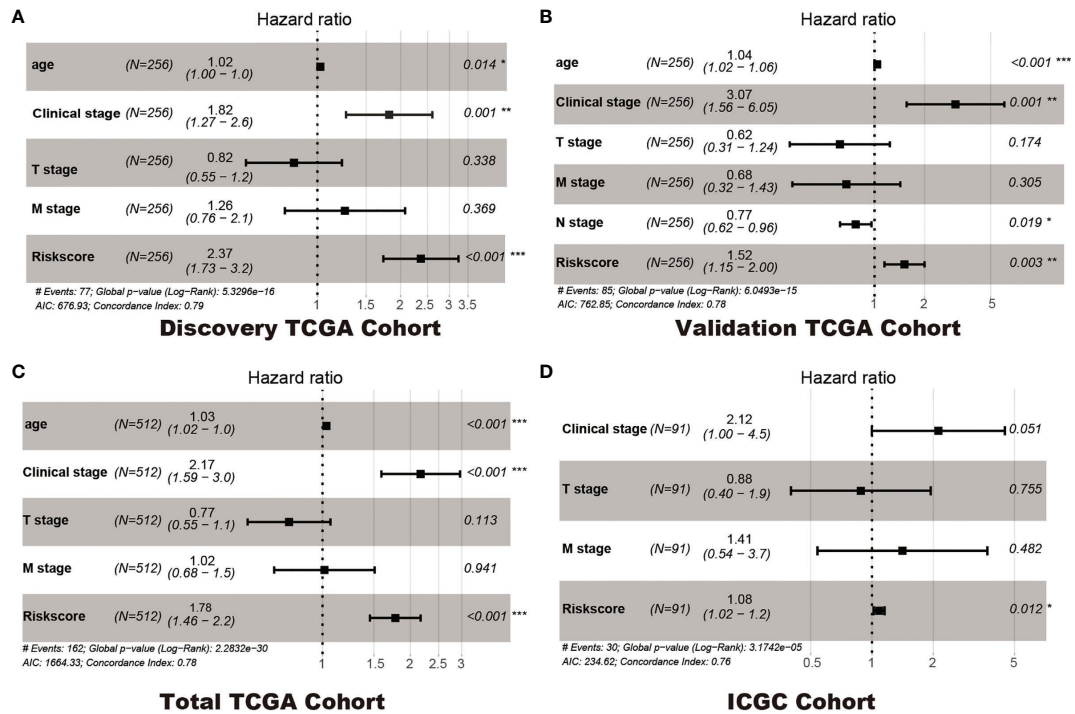
Recently, prognostic model integrating hypoxia-related gene expression and clinical features emerged (14–16). In our study, the hypoxia-related genes prognosis model was made up of six genes: *GPC3*, *KIF5A*, *PLAUR*, *ANKZF1*, *ETS1*, and *SELENBP1*. In the calculation formula of riskscore, the coefficient of *ETS1*



**FIGURE 3 |** Kaplan–Meier survival of the 6-gene prognostic model in the Discovery TCGA Cohort (A), the Validation TCGA Cohort (B), the Total TCGA Cohort (C) and the ICGC Cohort (D).

was a negative numbers, while the coefficient of the other five genes was a positive number. Therefore, there was a negative relationship between riskscore and ETS1, while there was a positive relationship between riskscore and GPC3, KIF5A, PLAUR, ANKZF1, and SELENBP1. Glypican 3 (GPC3) is an oncofetal proteoglycan anchored on the cell membrane (17). High GPC3 expression was also extensively associated with worse tumor differentiation, later tumor stage, presence of vascular invasion, and hepatitis B virus (HBV) infection (18). This protein is expressed in the liver and the kidney of healthy fetuses but is hardly expressed in adults, except in the placenta. Contrarily, GPC3 is specifically expressed in hepatocellular carcinoma (HCC), ovarian clear cell carcinoma, melanoma, squamous cell carcinoma of the lung, hepatoblastoma, nephroblastoma (Wilms tumor), yolk sac tumor, and some pediatric cancers (19). In addition, GPC3 has a low expression in ccRCC tissues than normal kidney tissues, and it can reduce the proliferation of RCC cell lines (20). Kinesin family member 5A (KIF5A) is a member of the kinesin superfamily which can

modulate many cell physiological behavior such as the proliferation of cell cycle (21). KIF5A mutation causes familial Amyotrophic lateral sclerosis (ALS; OMIM: 05400) (22). Tian et al. have proved that KIF5A can regulate the bladder cancer development and progression (21). Plasminogen activator, urokinase receptor (PLAUR) encodes the receptor for urokinase plasminogen activator and is a key molecule in regulating of cell-surface plasminogen activation (23). Previous studies have shown that PLAUR was involved in cancer cell migration, invasion and metastasis processing and could predict the prognosis of many cancers, such as glioma and oral squamous cell carcinoma (24–27). Ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1 (ANKZF1) is essential for the oxidative stress and the maintenance of mitochondrial integrity (28). Some bioinformatics analyses have proved that ANKZF1 based genes signature can predict the prognosis of patients with colon cancer, renal cell carcinoma, and prostate cancer (14, 29–33). ETS proto-oncogene 1, transcription factor (ETS1) is a transcription factor belong to the ETS domain family.

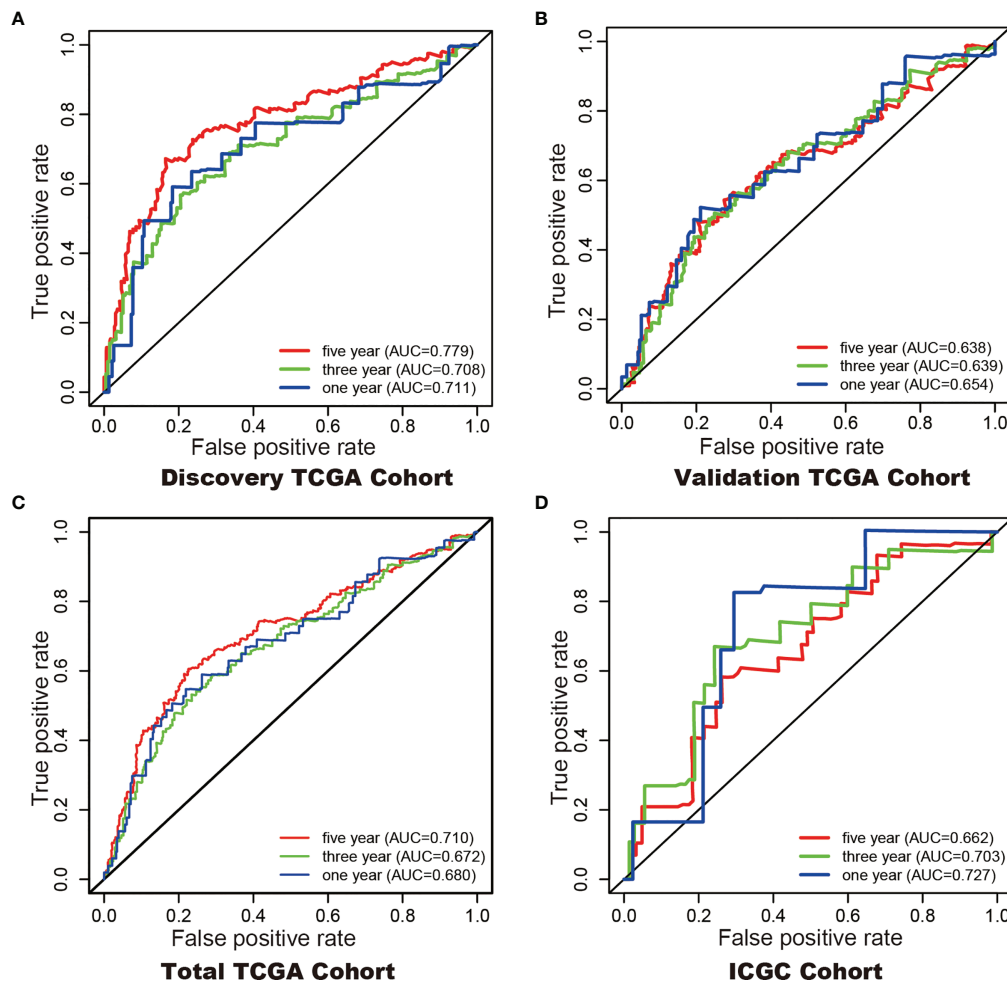


**FIGURE 4 |** Multivariate Cox regression analyses. Discovery TCGA Cohort (A), Validation TCGA Cohort (B), Total TCGA Cohort (C) and ICGC Cohort (D). \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001

**TABLE 1 |** Univariate and Multivariate cox regression in the Discovery and Validation cohorts.

Cohorts	Variables	Univariate Cox Regression		Variables	Multivariate Cox Regression	
		HR (95%CI)	p-value		HR (95%CI)	p-value
Discovery TCGA cohort	Age	1.028 (1.009–1.047)	<b>0.003</b>	Age	1.024 (1.005–1.044)	<b>0.014</b>
	Gender	1.025 (0.651–1.612)	0.916			
	Clinical stage	1.885 (1.556–2.284)	<b>&lt;0.001</b>	Clinical stage	1.817 (1.269–2.602)	<b>0.001</b>
	T stage	1.893 (1.487–2.409)	<b>&lt;0.001</b>	T stage	0.823 (0.552–1.227)	0.338
	M stage	1.957 (1.392–2.752)	<b>&lt;0.001</b>	M stage	1.258 (0.763–2.075)	0.369
	N stage	0.941 (0.75–1.18)	0.598			
	Riskscore	2.718 (2.056–3.594)	<b>&lt;0.001</b>	Riskscore	2.366 (1.73–3.235)	<b>&lt;0.001</b>
Validation TCGA cohort	Age	1.03 (1.01–1.051)	<b>0.003</b>	Age	1.038 (1.016–1.061)	<b>0.001</b>
	Gender	1.124 (0.709–1.784)	0.618			
	Clinical stage	1.984 (1.633–2.41)	<b>&lt;0.001</b>	Clinical stage	3.072 (1.559–6.053)	<b>0.001</b>
	T stage	2.12 (1.657–2.713)	<b>&lt;0.001</b>	T stage	0.618 (0.309–1.237)	0.174
	M stage	2.407 (1.724–3.362)	<b>&lt;0.001</b>	M stage	0.676 (0.32–1.428)	0.305
	N stage	0.772 (0.622–0.957)	<b>0.019</b>	N stage	0.77 (0.619–0.958)	<b>0.019</b>
	Riskscore	1.55 (1.21–1.986)	<b>0.001</b>	Riskscore	1.515 (1.147–2.001)	<b>0.003</b>
Total TCGA cohort	Age	1.029 (1.016–1.043)	<b>&lt;0.001</b>	Age	1.031 (1.017–1.046)	<b>&lt;0.001</b>
	Gender	1.045 (0.758–1.439)	0.789			
	Clinical stage	1.939 (1.693–2.221)	<b>&lt;0.001</b>	Clinical stage	2.17 (1.590–2.96)	<b>&lt;0.001</b>
	T stage	1.993 (1.683–2.361)	<b>&lt;0.001</b>	T stage	0.768 (0.554–1.064)	0.113
	M stage	2.145 (1.696–2.713)	<b>&lt;0.001</b>	M stage	1.015 (0.683–1.509)	0.941
	N stage	0.859 (0.735–1.004)	0.057			
	Riskscore	1.975 (1.645–2.37)	<b>&lt;0.001</b>	Riskscore	1.777 (1.455–2.17)	<b>&lt;0.001</b>
ICGC cohort	Age	1.031 (0.993–1.071)	0.109			
	Gender	0.939 (0.456–1.933)	0.863			
	Clinical stage	2.094 (1.515–2.896)	<b>&lt;0.001</b>	Clinical stage	2.116 (0.997–4.49)	0.051
	T stage	1.989 (1.402–2.821)	<b>&lt;0.001</b>	T stage	0.882 (0.401–1.941)	0.755
	M stage	2.522 (1.394–4.562)	<b>0.002</b>	M stage	1.411 (0.54–3.686)	0.482
	N stage	1.162 (0.696–1.938)	0.566			
	Riskscore	1.087 (1.023–1.154)	<b>0.007</b>	Riskscore	1.083 (1.018–1.153)	<b>0.012</b>

The bold values indicated the correspondence variable was significant difference in Univariate Cox regression or Multivariate Cox regression analysis.



**FIGURE 5** | ROC curves with calculated AUCs were established to evaluate the prognostic value of the 6-gene model in 5-, 3-, 1-years in the Discovery TCGA Cohort (A), the Validation TCGA Cohort (B), the Total TCGA Cohort (C), and the ICGC Cohort (D).

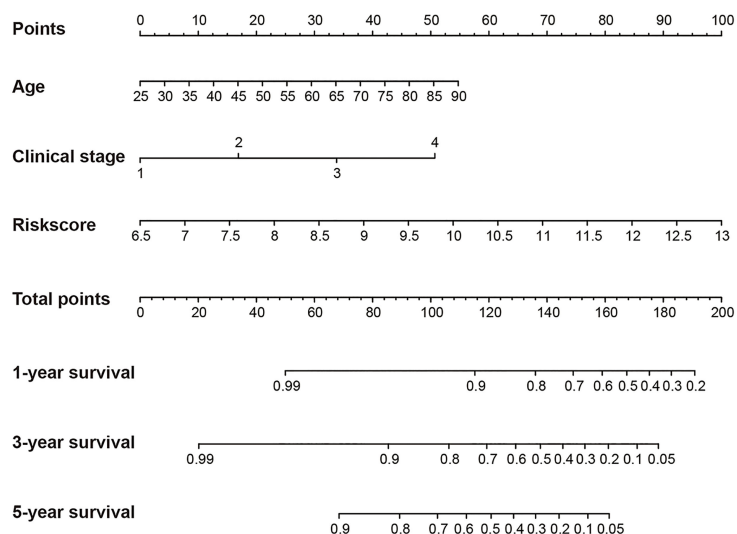
Ets1 expression is linked to poor survival of some cancers and contributes to the cancer cell invasiveness, EMT, and drug resistance. In addition, as a major MAPK downstream molecule, ETS1 might be a prospective therapeutic target for cancer (34, 35). Moreover, high expression levels of ETS1 were associated with poor survival in RCC tissues (36). Selenium binding protein 1 (SELENBP1) encodes protein and plays a selenium-dependent role in many physiological functions, such as protein degradation, cell differentiation and redox modulation. SELENBP1 is downregulated in many cancer types and its low expression levels are associated with poorer prognosis (37). In RCC, SELENBP1 has been proved as a tumor suppressor gene and low SELENBP1 mRNA expression predicts a worse cancer-specific survival (38).

Although tumor stage, lymph node metastasis and grade are significantly associated with the prognosis of KIRC (39), prognostic model integrating gene expression and clinical features are gained increasing attention along with the

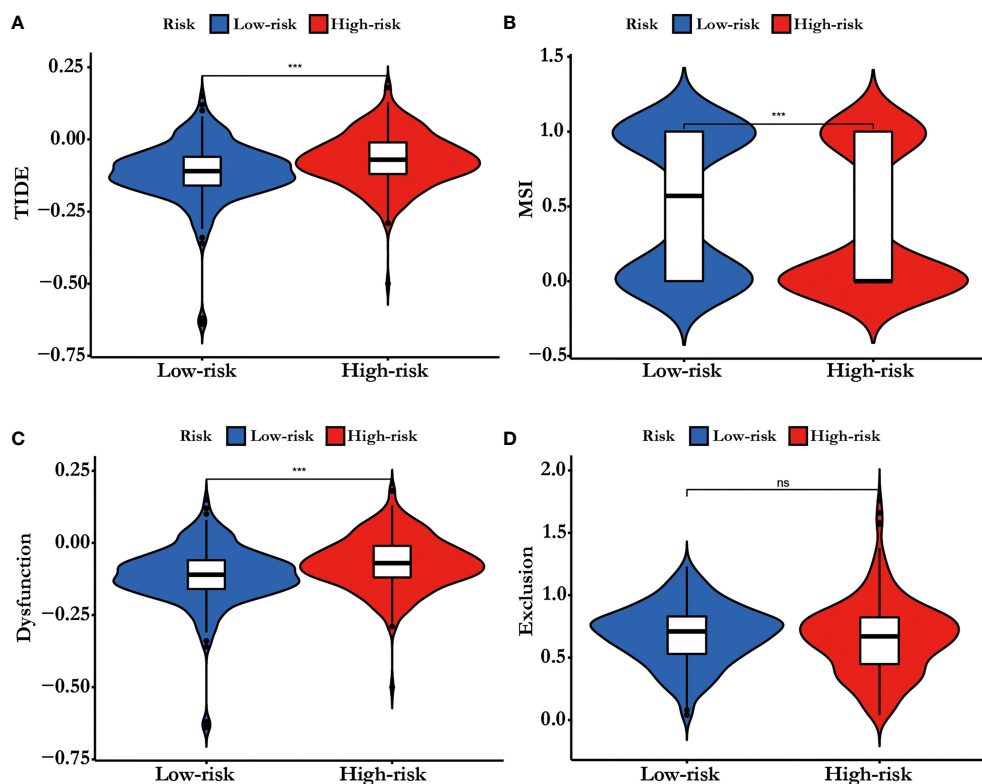
development of sequencing technology. Here, we developed a hypoxia-related genes prognostic model and a significant difference in overall survival was observed between high- and low-risk subgroups. The 1- (TCGA 0.68, ICGC 0.72), 3- (TCGA 0.67, ICGC 0.70), and 5- (TCGA 0.71, ICGC 0.66) year AUC showed a good accuracy of ROC curves in the TCGA cohort and the ICGC cohort, respectively. Hypoxia-related genes can be promising prognostic biomarkers for KIRC which gain evidence for hypoxia-related genes targeted therapy.

TIDE score could predict patient response to immunotherapy as it can reflect the potential capacity for the tumor's immune evasion (13). In our results, the high risk-group had a higher TIDE score than the low risk-group, indicating that patients with low riskscore could benefit more from ICI therapy than patients with high riskscore (Figure 7A). In addition, high TIDE score predicted a worse outcome, which was accordance with our results and it might interpret the high riskscore group presented a worse prognosis.





**FIGURE 6** | Construction of a nomogram for overall survival prediction in KIRC. The nomogram consists of age, clinical stage and the riskscore based on the six-hypoxia gene model.



**FIGURE 7** | (A) TIDE score, (B) MSI, (C) T cell Dysfunction and (D) Exclusion in different risk-groups. The score between the two risk-groups were compared through Wilcoxon test (ns, not significant, \*\*\*p < 0.001).

In conclusion, a six hypoxia-related gene risk model for independent prediction of the prognosis of KIRC patients was established. The different riskscores might be a biomarker for immunotherapy strategy.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the TCGA Research Network: <http://cancergenome.nih.gov/>. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

X-hN and J-yY conceived and designed the experiment. X-hN, Y-yQ, N-yL, S-cL, and Z-kJ performed the experiments and

analyzed the data. Interpretation of the findings were done by X-hN, Y-yQ, and Z-kJ. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Postdoctoral Research Grant in Henan Province (grant numbers 1901004, 1902005), the Henan Science and Technology Research Program (grant number 2018020142) and The Natural Science Foundation of Henan Province (212300410265). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## ACKNOWLEDGMENTS

The results shown here are in whole or part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.806264/full#supplementary-material>

## REFERENCES

- 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. doi: 10.14740/wjon1279
- Heng DY, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, et al. External Validation and Comparison With Other Models of the International Metastatic Renal-Cell Carcinoma Database Consortium Prognostic Model: A Population-Based Study. *Lancet Oncol* (2013) 14:141–8. doi: 10.1016/S1470-2045(12)70559-4
- Bedke J, Albiges L, Capitanio U, Giles RH, Hora M, Lam TB, et al. Updated European Association of Urology Guidelines on Renal Cell Carcinoma: Nivolumab Plus Cabozantinib Joins Immune Checkpoint Inhibition Combination Therapies for Treatment-Naive Metastatic Clear-Cell Renal Cell Carcinoma. *Eur Urol* (2021) 79:339–42. doi: 10.1016/j.eururo.2020.12.005
- Gao R, Qin H, Lin P, Ma C, Li C, Wen R, et al. Development and Validation of a Radiomic Nomogram for Predicting the Prognosis of Kidney Renal Clear Cell Carcinoma. *Front Oncol* (2021) 11:613668. doi: 10.3389/fonc.2021.613668
- Li H, Hu J, Yu A, Othmane B, Guo T, Liu J, et al. RNA Modification of N6-Methyladenosine Predicts Immune Phenotypes and Therapeutic Opportunities in Kidney Renal Clear Cell Carcinoma. *Front Oncol* (2021) 11:642159. doi: 10.3389/fonc.2021.642159
- Yoo SH, Yun J, Keam B, Hong SP, Ock CY, Koh J, et al. Discovery of Acquired Molecular Signature on Immune Checkpoint Inhibitors in Paired Tumor Tissues. *Cancer Immunol Immunother* (2021) 70:1755–69. doi: 10.1007/s00262-020-02799-y
- Yu J, Mao W, Sun S, Hu Q, Wang C, Xu Z, et al. Identification of an M6a-Related lncRNA Signature for Predicting the Prognosis in Patients With Kidney Renal Clear Cell Carcinoma. *Front Oncol* (2021) 11:663263. doi: 10.3389/fonc.2021.663263
- Zhang W, Li C, Wu F, Li N, Wang Y, Hu Y, et al. Analyzing and Validating the Prognostic Value of a TNF-Related Signature in Kidney Renal Clear Cell Carcinoma. *Front Mol Biosci* (2021) 8:689037. doi: 10.3389/fmolb.2021.689037
- Zheng W, Zhang S, Guo H, Chen X, Huang Z, Jiang S, et al. Multi-Omics Analysis of Tumor Angiogenesis Characteristics and Potential Epigenetic Regulation Mechanisms in Renal Clear Cell Carcinoma. *Cell Commun Signal* (2021) 19:39. doi: 10.1186/s12964-021-00728-9
- Clark PE. The Role of VHL in Clear-Cell Renal Cell Carcinoma and its Relation to Targeted Therapy. *Kidney Int* (2009) 76:939–45. doi: 10.1038/ki.2009.296
- Gossage L, Eisen T, Maher ER. VHL, the Story of a Tumour Suppressor Gene. *Nat Rev Cancer* (2014) 15:55–64. doi: 10.1038/nrc3844
- Schodel J, Grampp S, Maher ER, Moch H, Ratcliffe PJ, Russo P, et al. Hypoxia, Hypoxia-Inducible Transcription Factors, and Renal Cancer. *Eur Urol* (2016) 69:646–57. doi: 10.1016/j.eururo.2015.08.007
- Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat Med* (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1
- Li Z, Du G, Zhao R, Yang W, Li C, Huang J, et al. Identification and Validation of a Hypoxia-Related Prognostic Signature in Clear Cell Renal Cell Carcinoma Patients. *Med (Baltimore)* (2021) 100:e27374. doi: 10.1097/MD.00000000000027374
- Zhang Z, Li Q, Wang F, Ma B, Meng Y, Zhang Q. Identifying Hypoxia Characteristics to Stratify Prognosis and Assess the Tumor Immune Microenvironment in Renal Cell Carcinoma. *Front Genet* (2021) 12:606816. doi: 10.3389/fgene.2021.606816
- Zhong W, Zhong H, Zhang F, Huang C, Lin Y, Huang J. Characterization of Hypoxia-Related Molecular Subtypes in Clear Cell Renal Cell Carcinoma to Aid Immunotherapy and Targeted Therapy via Multi-Omics Analysis. *Front Mol Biosci* (2021) 8:684050. doi: 10.3389/fmolb.2021.684050
- Mu W, Jiang D, Mu S, Liang S, Liu Y, Zhang N. Promoting Early Diagnosis and Precise Therapy of Hepatocellular Carcinoma by Glypican-3-Targeted

- Synergistic Chemo-Photothermal Theranostics. *ACS Appl Mater Interf* (2019) 11:23591–604. doi: 10.1021/acsami.9b05526
18. Zhang J, Zhang M, Ma H, Song X, He L, Ye X, et al. Overexpression of Glypican-3 is a Predictor of Poor Prognosis in Hepatocellular Carcinoma: An Updated Meta-Analysis. *Med (Baltimore)* (2018) 97:e11130. doi: 10.1097/MD.00000000000011130
  19. Shimizu Y, Suzuki T, Yoshikawa T, Endo I, Nakatsura T. Next-Generation Cancer Immunotherapy Targeting Glypican-3. *Front Oncol* (2019) 9:248. doi: 10.3389/fonc.2019.00248
  20. Valsechi MC, Oliveira AB, Conceicao AL, Stuqui B, Candido NM, Provazzi PJ, et al. GPC3 Reduces Cell Proliferation in Renal Carcinoma Cell Lines. *BMC Cancer* (2014) 14:631. doi: 10.1186/1471-2407-14-631
  21. Tian DW, Wu ZL, Jiang LM, Gao J, Wu CL, Hu HL. KIF5A Promotes Bladder Cancer Proliferation *In Vitro* and *In Vivo*. *Dis Markers* (2019) 2019:4824902. doi: 10.1155/2019/4824902
  22. Brenner D, Yilmaz R, Muller K, Grehl T, Petri S, Meyer T, et al. Hot-Spot KIF5A Mutations Cause Familial ALS. *Brain* (2018) 141:688–97. doi: 10.1093/brain/awx370
  23. Borglum AD, Byskov A, Ragno P, Roldan AL, Tripputi P, Cassani G, et al. Assignment of the Urokinase-Type Plasminogen Activator Receptor Gene (PLAUR) to Chromosome 19q13.1-Q13.2. *Am J Hum Genet* (1992) 50:492–7. doi: 0002-9297/92/5003-0006
  24. Gonias SL, Hu J. Urokinase Receptor and Resistance to Targeted Anticancer Agents. *Front Pharmacol* (2015) 6:154. doi: 10.3389/fphar.2015.00154
  25. Gilder AS, Natali L, Van Dyk DM, Zalfa C, Banki MA, Pizzo DP, et al. The Urokinase Receptor Induces a Mesenchymal Gene Expression Signature in Glioblastoma Cells and Promotes Tumor Cell Survival in Neurospheres. *Sci Rep* (2018) 8:2982. doi: 10.1038/s41598-018-21358-1
  26. Tan YQ, Li YT, Yan TF, Xu Y, Liu BH, Yang JA, et al. Six Immune Associated Genes Construct Prognostic Model Evaluate Low-Grade Glioma. *Front Immunol* (2020) 11:606164. doi: 10.3389/fimmu.2020.606164
  27. Wang J, Wang Y, Kong F, Han R, Song W, Chen D, et al. Identification of a Six-Genes Prognostic Signature for Oral Squamous Cell Carcinoma. *J Cell Physiol* (2020) 235:3056–68. doi: 10.1002/jcp.29210
  28. van Haaften-Visser DY, Harakalova M, Mocholi E, van Montfrans JM, Elkadri A, Rieter E, et al. Ankyrin Repeat and Zinc-Finger Domain-Containing 1 Mutations are Associated With Infantile-Onset Inflammatory Bowel Disease. *J Biol Chem* (2017) 292:7904–20. doi: 10.1074/jbc.M116.772038
  29. Guo K, Lai C, Shi J, Tang Z, Liu C, Li K, et al. A Novel Risk Factor Model Based on Glycolysis-Associated Genes for Predicting the Prognosis of Patients With Prostate Cancer. *Front Oncol* (2021) 11:605810. doi: 10.3389/fonc.2021.605810
  30. Xing Q, Zeng T, Liu S, Cheng H, Ma L, Wang Y. A Novel 10 Glycolysis-Related Genes Signature Could Predict Overall Survival for Clear Cell Renal Cell Carcinoma. *BMC Cancer* (2021) 21:381. doi: 10.1186/s12885-021-08111-0
  31. Chen S, Cao G, Wu W, Lu Y, He X, Yang L, et al. Mining Novel Cell Glycolysis Related Gene Markers That Can Predict the Survival of Colon Adenocarcinoma Patients. *Biosci Rep* (2020) 40(8):BSR20201427. doi: 10.1042/BSR20201427
  32. Xu F, Guan Y, Xue L, Huang S, Gao K, Yang Z, et al. The Effect of a Novel Glycolysis-Related Gene Signature on Progression, Prognosis and Immune Microenvironment of Renal Cell Carcinoma. *BMC Cancer* (2020) 20:1207. doi: 10.1186/s12885-020-07702-7
  33. Zhou X, Shang YN, Lu R, Fan CW, Mo XM. High ANKZF1 Expression Is Associated With Poor Overall Survival and Recurrence-Free Survival in Colon Cancer. *Future Oncol* (2019) 15:2093–106. doi: 10.2217/fon-2018-0920
  34. Dittmer J. The Role of the Transcription Factor Ets1 in Carcinoma. *Semin Cancer Biol* (2015) 35:20–38. doi: 10.1016/j.semcancer.2015.09.010
  35. Tetsu O, McCormick F. ETS-Targeted Therapy: Can it Substitute for MEK Inhibitors? *Clin Transl Med* (2017) 6:16. doi: 10.1186/s40169-017-0147-4
  36. Zhai W, Ma J, Zhu R, Xu C, Zhang J, Chen Y, et al. MiR-532-5p Suppresses Renal Cancer Cell Proliferation by Disrupting the ETS1-Mediated Positive Feedback Loop With the KRAS-Nap11/P-ERK Axis. *Br J Cancer* (2018) 119:591–604. doi: 10.1038/s41416-018-0196-5
  37. Elhodaky M, Diamond AM. Selenium-Binding Protein 1 in Human Health and Disease. *Int J Mol Sci* (2018) 19(11):3437. doi: 10.3390/ijms19113437
  38. Ha YS, Lee GT, Kim YH, Kwon SY, Choi SH, Kim TH, et al. Decreased Selenium-Binding Protein 1 mRNA Expression Is Associated With Poor Prognosis in Renal Cell Carcinoma. *World J Surg Oncol* (2014) 12:288. doi: 10.1186/1477-7819-12-288
  39. Ficarra V, Galfano A, Mancini M, Martignoni G, Artibani W. TNM Staging System for Renal-Cell Carcinoma: Current Status and Future Perspectives. *Lancet Oncol* (2007) 8:554–8. doi: 10.1016/S1470-2045(07)70173-0

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ning, Li, Qi, Li, Jia and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Causal Relationships Between Extrinsic Exposures and Risk of Prostate Cancer: A Phenome-Wide Mendelian Randomization Study

Dongqing Gu<sup>1</sup>, Mingshuang Tang<sup>1</sup>, Yutong Wang<sup>1</sup>, Huijie Cui<sup>2</sup>, Min Zhang<sup>3</sup>, Ye Bai<sup>3</sup>, Ziqian Zeng<sup>1</sup>, Yunhua Tan<sup>1</sup>, Xin Wang<sup>2</sup> and Ben Zhang<sup>1\*</sup>

<sup>1</sup> Department of Epidemiology and Biostatistics, First Affiliated Hospital, Army Medical University, Chongqing, China, <sup>2</sup> Department of Epidemiology and Biostatistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China, <sup>3</sup> School of Public Health and Management, Chongqing Medical University, Chongqing, China

## OPEN ACCESS

### Edited by:

Rong Na,  
Shanghai Jiao Tong University, China

### Reviewed by:

Xian-Tao Zeng,  
Wuhan University, China  
Peng Wang,  
Anhui Medical University, China

### \*Correspondence:

Ben Zhang  
benzhang@vip.163.com

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 05 December 2021

**Accepted:** 19 January 2022

**Published:** 14 February 2022

### Citation:

Gu D, Tang M, Wang Y, Cui H, Zhang M, Bai Y, Zeng Z, Tan Y, Wang X and Zhang B (2022) The Causal Relationships Between Extrinsic Exposures and Risk of Prostate Cancer: A Phenome-Wide Mendelian Randomization Study. *Front. Oncol.* 12:829248. doi: 10.3389/fonc.2022.829248

**Background:** Prostate cancer is the second most common cancer in males worldwide, and multitudes of factors have been reported to be associated with prostate cancer risk.

**Objectives:** We aim to conduct the phenome-wide exposed-omics analysis of the risk factors for prostate cancer and verify the causal associations between them.

**Methods:** We comprehensively searched published systematic reviews and meta-analyses of cohort studies and conducted another systematic review and meta-analysis of the Mendelian randomization studies investigating the associations between extrinsic exposures and prostate cancer, thus to find all of the potential risk factors for prostate cancer. Then, we launched a phenome-wide two-sample Mendelian randomization analysis to validate the potentially causal relationships using the PRACTICAL consortium and UK Biobank.

**Results:** We found a total of 55 extrinsic exposures for prostate cancer risk. The causal effect of 30 potential extrinsic exposures on prostate cancer were assessed, and the results showed docosahexaenoic acid (DHA) [odds ratio (OR)=0.806, 95% confidence interval (CI): 0.661-0.984,  $p=0.034$ ], insulin-like growth factor binding protein 3 (IGFBP-3) (OR=1.0002, 95%CI: 1.00004-1.0004,  $p=0.016$ ), systemic lupus erythematosus (SLE) (OR=0.9993, 95%CI: 0.9986-0.99997,  $p=0.039$ ), and body mass index (BMI) (OR=0.995, 95%CI: 0.990-0.9999,  $p=0.046$ ) were associated with prostate cancer risk. However, no association was found between the other 26 factors and prostate cancer risk.

**Conclusions:** Our study discovered the phenome-wide exposed-omics risk factors profile of prostate cancer, and verified that the IGFBP-3, DHA, BMI, and SLE were causally related to prostate cancer risk. The results may provide new insight into the study of the pathogenesis of prostate cancer.

**Keywords:** prostate cancer, Mendelian randomization, risk factor, causal relationship, systematic review



## INTRODUCTION

Prostate cancer is the second most frequently diagnosed malignancy and the fifth leading cause of cancer-related death among males worldwide (1). In the United States, the estimated new prostate cancer cases reached 191 930, and prostate cancer-related death achieved 33 330 in 2020, making it a malignancy with the highest incidence and second leading cause of mortality in males (2). In 2019, the regions with the most incident cases of prostate cancer were High-income North America, Western Europe, and East Asia. It was reported that the global incident cases were 169.11% higher for prostate cancer during the past 30 year, making it a major global public health challenges (3).

Investigating the risk factors and potential etiological factors for prostate cancer may provide the basis for identifying high-risk populations, developing disease control strategies, and even cognizing the pathogenesis. The burden of prostate cancer was mainly distributed among older men. In addition, epidemiological evidence have established some attributable risk factors for prostate cancer, such as smoking, high body mass index (BMI) and high fasting glucose (3, 4). However, due to the inherent defect of the temporal problem and inadequately controlled confounders in conventional observational studies, the causality between these factors and prostate cancer remains debated.

Mendelian randomization (MR) analysis is a widely used method that uses genetic variants as instrumental variables to inference the causal relationships between potential risk factors and outcomes in observational data in recent years (5). Since the genotypes are presumed to be randomly allocated in gamete formation, Mendelian randomization analysis is not affected by reverse causation. In addition, the inheritance of one exposure predicted by the SNPs is usually independent from the inheritance of another exposure, it is less susceptible to confounding factors (6). Two-sample Mendelian randomization analysis has the additional advantage that access to individual-level data or trait measurements in all samples is not required. Therefore, it can be implemented using summary information for the required genotype-exposure and genotype-outcome associations from separate samples, which significantly increases the scope and efficiency of the approach (7, 8).

Several previous Mendelian randomization studies have identified the etiological factors for prostate cancer, such as serum 25-Hydroxyvitamin D, body mass index (BMI), alcohol consumption, and vitamin B12. However, due to the relatively smaller sample size and lower proportion of variance explained

by the instrumental variables, the results are usually inconclusive, and the evidence was insufficient. In this scenario, we aim to first review the published systematic review and meta-analyses of cohort studies and conduct another systematic review and meta-analyses of Mendelian randomization studies to summarize the phenome-wide exposed-omics risk factors for prostate cancer. Next, we conducted two-sample Mendelian randomization analyses to verify the causal relationships using Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium covering 44 825 prostate cancer cases and 27 904 controls, as well as UK Biobank including 6879 prostate cancer cases and 199 891 controls.

## METHODS

We obtained summary GWAS statistics from PRACTICAL consortium and UK Biobank (application ID 45973), and all participants included in the consortia were of European ancestry, relevant ethics approval can be found in the original publications (9, 10). Any additional ethical approval was adjudged unnecessary for the present study.

### Potential Risk Factors Identified by the Published Meta-Analysis of Cohort Studies

We searched PubMed, Embase, and Web of Science databases to identify all potential risk factors for prostate cancer reported by the published meta-analysis of cohort studies published in print or online before October 31, 2019. The key terms were as follows: “meta- OR review OR pooled OR consortium OR consortia OR collaboration” AND “Prostate cancer OR prostate adenocarcinoma OR prostate carcinoma OR prostate tumor OR prostate malignancy OR prostate neoplasm”. Inclusion criteria are as follows: (1) meta-analysis of cohort studies; (2) the outcome of interest was prostate cancer; (3) written in the English language. For multiple publications investigating the same factor, the latest publication or publication with the largest sample size was included.

### Systematic Review and Meta-Analysis of Mendelian Randomization Studies

We conducted a systematic review and meta-analysis of published Mendelian randomization studies, and this review was registered in PROSPERO (CRD42021287713). We searched PubMed, Embase, and Web of Science databases to identify all potential risk factors for prostate cancer reported by the Mendelian randomization studies (published in print or online before October 31, 2019) with the following key terms “Prostate cancer OR prostate adenocarcinoma OR prostate carcinoma OR prostate tumor OR prostate malignancy OR prostate neoplasm” AND “Mendelian randomization OR instrumental variable OR causal”. Inclusion criteria are as follows: (1) Mendelian randomization studies to assess the association between exposures and risk of prostate cancer; (2) reported results included odds ratios (ORs) with 95% CIs, which were estimated using an instrumental variable method. When one more study reported data from the

**Abbreviations:** ALA, alpha-linolenic acid; BMI: body mass index; BPH, Benign prostatic hyperplasia; CI, confidence interval; DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; FG, fasting glucose; GWAS, genome-wide association study; HDL, high-density lipoprotein; IGF-I, Insulin-like growth factor-I; IGFBP-3, insulin-like growth factor binding protein 3; IVW, inverse variance weighted; LDL, low-density lipoprotein; MR, Mendelian randomization; OR, odds ratio; PRACTICAL, Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; ω-3 PUFAs, omega-3 polyunsaturated fatty acids; SHBG, sex hormone-binding globulin; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphisms; T2DM, Type 2 diabetes mellitus; WC, waist circumference.

same source or databank, only the study with the most participants was included in the analysis. When more than two datasets reported the same factor, the odds ratio (OR) from individual studies were pooled using a random-effects model. Statistical analyses were done using Stata version 15 (Stata, College Station, TX, USA).

## Selection of Factors

Inclusion criteria of factors are as follows: (1) for the same factor, we only included the factor reported to be positive by the largest meta-analysis; (2) dietary factors or internal exposures were excluded. We selected all potential risk factors for prostate cancer identified by the published meta-analysis of cohort studies and Mendelian randomization studies. Then, we searched for each of the risk factors in the GWAS catalog ([www.ebi.ac.uk/gwas](http://www.ebi.ac.uk/gwas)) to identify the associations between SNPs and the specific risk factor of interest, and any factor without related GWAS or the GWAS with incomplete information was excluded.

## Defining Genetic Instruments

The SNPs for each exposure identified by the largest GWAS in populations of European ancestry were used to conduct instrumental variables. Further details of the exposures and how we defined the genetic instruments are provided in the **Supplementary Methods** and **Supplementary Table S1**. Inclusion criteria of the SNPs as follows: (1) independent loci: defined as  $r^2 < 0.1$  based on European ancestry reference data from the 1000 Genomes Project (11, 12), and for a locus in which multiple SNPs in linkage disequilibrium, we selected the SNP with the strongest effect; (2) GWAS  $p$ -value threshold of  $< 5 \times 10^{-8}$ , and for the SNPs of risk factor less than ten, we set GWAS-significant  $p$ -value threshold of  $< 5 \times 10^{-6}$ ; (3) having the rs numbers (or position information); (4) providing beta-coefficient ( $\beta$ ), and standard error (SE) (or sufficient data to calculate them). After selecting the set of SNPs for each risk factor, we extracted the following information for each SNP-risk factor association: rs numbers, effect allele, other alleles, effect allele frequency,  $\beta$ , SE, and  $p$ -value. Any SNP missing the information was removed.

For the SNP(s) extracted for use in the MR-analysis, we calculated the proportion of variance explained ( $R^2$ ) in the risk factor by the SNP(s) and the strength of the instrument (F-statistic) (13). The formulas to calculate  $R^2$  and F-statistic were:

$$R^2 = [2 \times \beta^2 \times \text{MAF} \times (1 - \text{MAF})] / (2 \times \beta^2 \times \text{MAF} \times [1 - \text{MAF}] + (\text{SE}(\beta))^2 \times (2 \times N) \times \text{MAF} \times (1 - \text{MAF})),$$

where  $\beta$  is the effect size (beta coefficient) for a given SNP, MAF is the minor allele frequency,  $\text{SE}(\beta)$  is the standard error of the effect size, and  $N$  is the sample size of the GWAS for the SNP-risk factor association.

$F = R^2 \times (N - 1 - k) / ((1 - R^2) \times k)$ , where  $R^2$  is the proportion of variance explained in the risk factor by the genetic instrument,  $N$  is the sample size of the GWAS,  $k$  is the number of SNPs included in the instrument.

## Outcome Trait

GWAS results for prostate cancer were obtained from fixed-effects meta-analyses based on individuals of European ancestry in the PRACTICAL consortium (44 825 prostate cancer cases and 27 904 controls) (10), and UK Biobank (6879 prostate cancer cases and 199 891 controls) (9). We extracted the following information for each SNP of risk factor: rs numbers, effect allele, other alleles, effect allele frequency,  $\beta$ , SE, and  $p$ -value. We removed any SNP missing this information, and the one reached a  $p$ -value threshold of  $< 5 \times 10^{-8}$ .

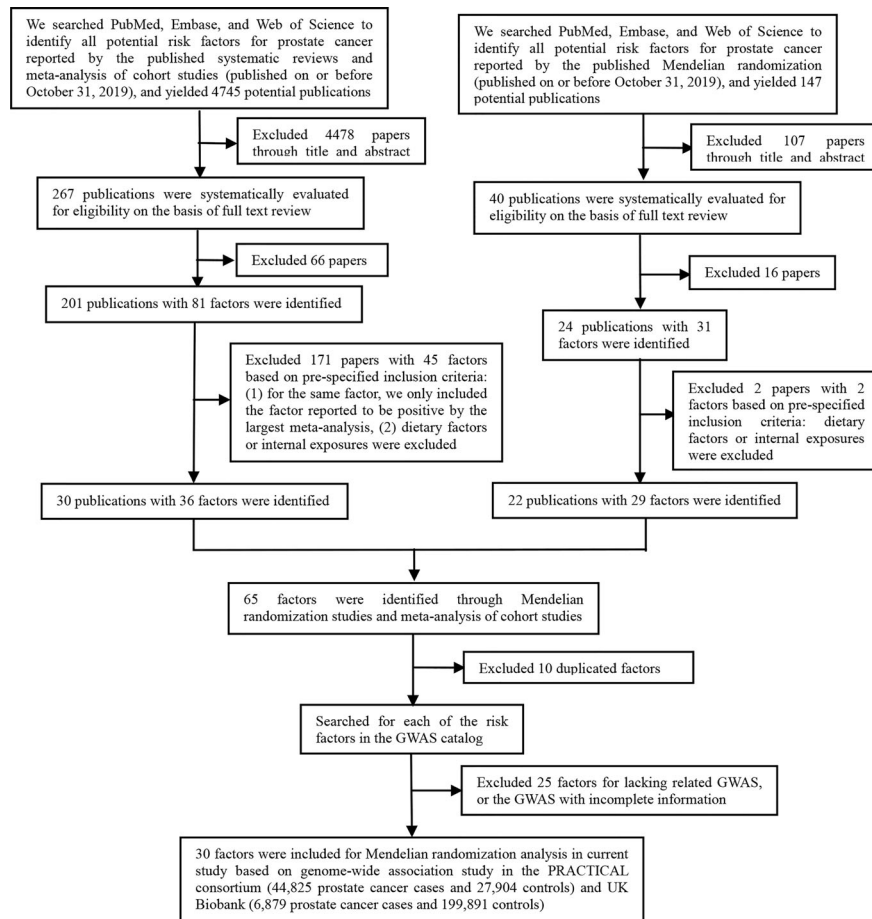
## Two-Sample Mendelian Randomized Analysis

The inverse variance weighted (IVW) fixed-effect method was used as the main method to estimate the effect of genetically predicted exposure on prostate cancer in our Mendelian randomization analysis. The IVW method estimates the effect of the exposure on the outcome from the slope of the relationship between bXG (SNP-exposure association) and bYG (SNP-outcome association). Casual estimates were presented as an OR and its 95% CI. OR estimates were reported per standard deviation (SD) increment for continuous variable and per log-odds increment for categorical variable in genetically determined risk of the exposures. In addition, other Mendelian randomization methods including MR-Egger, weighted median, and weighted mode method were used to check the consistency of the direction of effect estimates. We assessed horizontal pleiotropy, heterogeneity tests, funnel plots, scatter plots, and leave-one-out plots in sensitivity analyses. In addition, scatter plots of effect estimates of individual SNPs with outcome versus effect estimates of individual SNPs with exposure are provided as a comparative visual assessment of the effect estimates generated from different Mendelian randomization methods. All analyses were conducted using the package TwoSampleMR (version 0.5.6) in R (version 4.1.2).

## RESULTS

### Exposed-Omics Analysis of the Extrinsic Exposures for Prostate Cancer

As shown in **Figure 1**, the present study conducted two parts of investigation: (1) a total of 4745 published meta-analyses of cohort studies were acquired from the PubMed, Embase, and Web of Science databases. After excluding the 4715 publications through title, abstract, and full-text reading, 30 articles including 36 factors were identified (**Figure 2A**). (2) Another systematic review and meta-analysis of the Mendelian randomization study incorporated 24 publications with 31 factors. The characteristics of these studies are shown in **Supplementary Table S2**. Of these studies, 18 studies outcome data source was generated from PRACTICAL consortium, two from UK-based ProtecT study, one from UK Biobank, and six from other sources (among them three studies from two sources). For these studies, eight studies involving eight factors with 140 036 cases and 279 025 controls were eligible for the meta-analysis. Results showed coffee consumption (OR=0.91,



**FIGURE 1** | The flowchart of factors selection.

95%CI: 0.83-0.99), microseminoprotein-beta (OR=0.96, 95%CI: 0.95-0.98), and pubertal development (OR=0.97, 95%CI: 0.94-1.00) may be causal protective factors of prostate cancer. However, we found no association of triglycerides (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL), and height with risk of prostate cancer (**Figure 2B**).

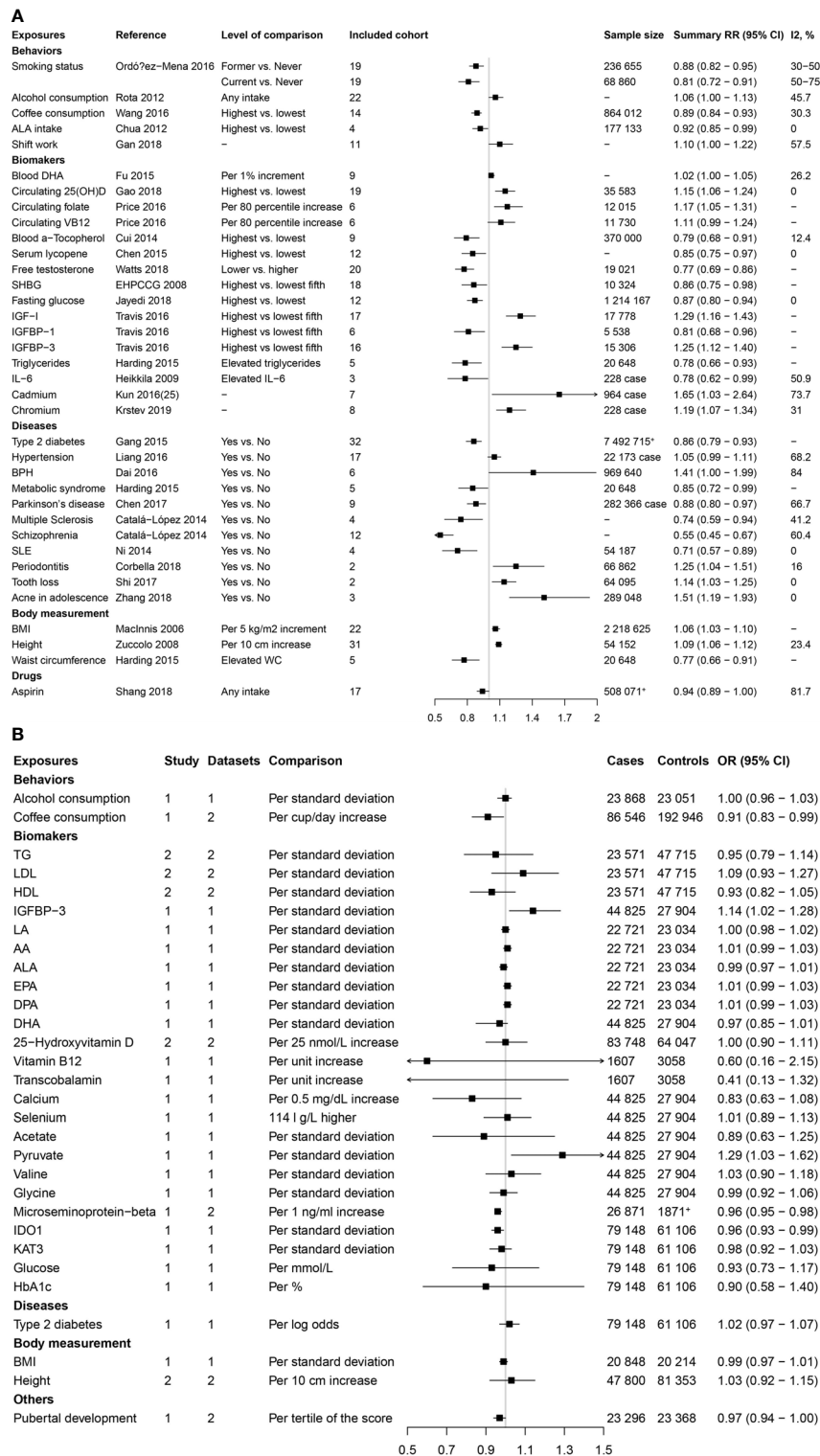
These two parts get 65 extrinsic exposures for prostate cancer. After excluding 10 duplicated factors and 25 factors without GWAS, a total of 30 exposures were included for Mendelian randomization analysis in the current study. Of the included 30 factors, 13 were risk factors [including alcohol consumption, blood docosahexaenoic acid (DHA), 25-Hydroxyvitamin D, circulating folate, circulating vitamin B12, Insulin-like growth factor-I (IGF-1), insulin-like growth factor binding protein 3 (IGFBP-3), hypertension, benign prostatic hyperplasia (BPH), periodontitis, BMI, height, and LDL] for prostate cancer, while 17 were protective factors [including smoking status, coffee consumption, alpha-linolenic acid (ALA), free testosterone, sex hormone-binding globulin (SHBG), fasting glucose (FG), TG, HDL, Interleukin-6, Type 2 diabetes mellitus (T2DM), metabolic syndrome, Parkinson's disease, multiple sclerosis, schizophrenia,

systemic lupus erythematosus (SLE), waist circumference (WC), and aspirin intake].

## Mendelian Randomization Analysis

The genetic instruments of the selected exposures used in the Mendelian randomization analysis could explain 0.42%-54.81% of variability, and the F-statistic ranged from 8.12 to 286.33 (**Table 1**). The detailed information of variants used to conduct instrumental variables for each exposure was shown in **Supplementary Table S3**. The results of Mendelian randomization analyses are shown in **Table 2** and **Supplementary Figures S1-S4**, and the effect estimates using different MR methods are provided in **Supplementary Table S4**.

In the PRACTICAL consortium dataset, of the 30 potential extrinsic exposures examined in our study, DHA was causally associated with a decreased risk of prostate cancer (OR=0.806, 95%CI: 0.661-0.984,  $p=0.034$ ) with the wald ratio method. Consistent with the findings in the previous meta-analysis, the conventional IVW method indicated a causal association between genetically predisposed IGFBP-3 and prostate cancer (OR=1.0002, 95%CI: 1.00004-1.0004,  $p=0.016$ ), and weighted



**FIGURE 2 |** Factors of prostate cancer that were identified by the published systematic reviews and meta-analyses of cohort studies and mendelian randomization studies. **(A)** Factors identified by the published systematic reviews and meta-analyses of cohort studies. **(B)** Factors identified by the published mendelian randomization studies.



**TABLE 1 |** Summary of genetic instruments used in the present Mendelian randomization analysis.

Extrinsic exposures	PMID	Maximum sample size	SNPs in genetic instrument	% explained variability	F statistic
Coffee consumption	31046077	375 833	13	0.55	160.69
Alcohol consumption	23743675	4915	12	4.62	19.80
Smoking status	30643258	518 633	223	1.73	40.99
25-Hydroxy vitamin D	32059762	443 734	138	8.18	286.33
Vitamin B12	19303062	3613	3	23.62	372.11
Folate	19303062	3617	1	0.42	15.26
Fasting blood glucose	20081858	128 613	33	3.3	132.9
IGF-I	21216879	10 280	4	5.93	161.99
IGFBP-3	21216879	10 280	4	1.07	27.86
TG	30275531	617 303	151	2.88	121.33
HDL	30275531	617 303	156	3.71	152.61
LDL	30275531	617 303	119	2.60	138.49
Testosterone	31169883	4291	9	6.29	35.91
SHBG	22829776	29 966	12	5.91	156.79
ALA	21829377	8866	4	3.76	86.46
DHA	21829377	8866	1	0.72	64.42
Interleukin-6	27989323	8189	23	5.99	22.64
Hypertension	31879980	185 565	35	1.14	61.20
Type 2 diabetes	30054458	659 316	149	1.54	68.89
Periodontitis	30218097	15 003	9	1.47	24.81
BPH	30988330	10 419	7	1.59	24.03
SLE	26502338	14 267	69	54.81	249.52
Schizophrenia	29483656	265 304	143	0.44	8.12
Parkinson's disease	31701892	1 474 097	106	0.6	84.38
Multiple sclerosis	31604244	115 804	192	11.48	76.13
Metabolic syndrome	31589552	291 107	93	2.53	81.11
BMI	30108127	793 208	191	1.48	62.20
Height	25282103	333 355	169	5.89	123.48
Waist circumference	28443625	362 932	51	0.69	49.36
Aspirin use	31015401	112 010	10	0.43	48.18

ALA, alpha-linolenic acid; BMI, body mass index; BPH, benign prostatic hyperplasia; SLE, systemic lupus erythematosus; DHA, docosahexaenoic acid; IGF-I, Insulin-like growth factor-I; IGFBP-3, insulin-like growth factor binding protein 3; PCa, Prostate cancer; SHBG, sex hormone binding globulin; SNP, single nucleotide polymorphism.

median methods also generated similar effect estimation (OR=1.0002, 95%CI: 1.0001-1.0004,  $p=0.0002$ ). In addition, in the UK Biobank dataset, we found inverse associations for systemic lupus erythematosus (OR=0.9993, 95%CI: 0.9986-0.9999,  $p=0.039$ ) and BMI (OR=0.995, 95%CI: 0.990-0.9999,  $p=0.046$ ) with prostate cancer risk using IVW method, and weighted median methods also supported these associations. Likewise, the MR Egger method indicated a causal association between genetically predisposed SLE and prostate cancer in both PRACTICAL consortium (OR=0.96, 95%CI: 0.93-0.99,  $p=0.003$ ) and UK Biobank (OR=0.999, 95%CI: 0.997-0.999,  $p=0.002$ ). Besides, no causal relationship was found between other exposures and prostate cancer.

## DISCUSSION

In the present study, we summarized the previous meta-analysis of cohort studies and performed a systematic review and meta-analysis of published Mendelian randomization studies, thus finding a total of 55 risk factors for prostate cancer. Besides, we conducted a comprehensive two-sample MR analysis to evaluate the potential causal effect of 30 extrinsic exposures on the risk of prostate cancer based on European-descent individuals in the PRACTICAL consortium and UK Biobank.

The IGF pathway plays a critical role in somatic growth and activates carcinogenic intracellular signaling networks. Published results have shown an association between circulating insulin-like growth factors (IGFs) and their binding proteins (IGFBPs) and the subsequent prostate cancer risk (14–16). Our Mendelian randomization results showed a positive association between IGFBP-3 levels and prostate cancer, as previously reported from observational and Mendelian randomization studies (15–18). IGFBP-3 is the most abundant circulating IGFBP and modulates the bioactivity of IGFs. Independent of IGFs, IGFBP-3 could regulate cell proliferation and apoptosis, leading to the carcinogenesis of certain common cancers (19, 20). Furthermore, experimental pieces of evidence suggested that IGFBP-3 might contribute to the growth and progression of prostate cancer cells (21, 22). Although our Mendelian randomization results were unable to support previous evidence of an association between genetically predicted serum IGF-1 levels and prostate cancer risk, recent published Mendelian randomization studies reported a causal association of IGF-1 levels with prostate cancer (23, 24). The inconsistency might be attributed to the proportions of advanced-stage prostate cancer cases across datasets (23).

Dietary fatty acids, especially omega-3 polyunsaturated fatty acids ( $\omega$ -3 PUFAs), are one of the most intensively studied dietary factors closely related to prostate cancer risk.  $\omega$ -3

**TABLE 2 |** Mendelian randomization analyses of the association between extrinsic exposures and prostate cancer risk.

Extrinsic exposures	PRACTICAL consortium			UK Biobank			Combined	
	N SNPs	OR (95%CI)	p	N SNPs	OR (95%CI)	p	OR (95%CI)	p
<b>Behaviors</b>								
Coffee consumption	15	0.855 (0.503-1.452)	0.561	15	0.987 (0.972-1.002)	0.081	1.005 (0.985-1.025)	0.078
Alcohol consumption (drinkers vs non-drinkers)	16	0.998 (0.994-1.003)	0.440	15	1.000 (1.000-1.000)	0.726	1.000 (1.000-1.000)	0.700
Smoking status (ever vs never smokers)	214	1.002 (0.872-1.152)	0.973	213	1.003 (0.997-1.009)	0.372	0.999 (0.999-1.000)	0.372
ALA	3	0.655 (0.176-2.442)	0.528	3	0.655 (0.176-2.442)	0.528	0.665 (0.258-1.661)	0.373
DHA	1	<b>0.806 (0.661-0.984)</b>	<b>0.034</b>	1	1.008 (0.998-1.017)	0.117	0.987 (0.972-1.002)	0.460
<b>Biomarkers</b>								
Serum 25-Hydroxyvitamin D levels	106	1.012 (0.949-1.080)	0.708	85	0.999 (0.995-1.003)	0.702	0.999 (0.995-1.003)	0.719
Vitamin B12	3	1.000 (1.000-1.001)	0.081	3	1.000 (1.000-1.000)	0.272	1.000 (1.000-1.000)	0.506
Folate	1	0.991 (0.971-1.013)	0.431	1	0.999 (0.998-1.000)	0.223	0.997 (0.989-1.006)	0.210
Fasting blood glucose	32	0.927 (0.730-1.178)	0.536	32	0.997 (0.989-1.006)	0.560	0.923 (0.745-1.142)	0.546
IGF-I	4	1.000 (0.997-1.003)	0.864	4	1.000 (1.000-1.000)	0.423	1.000 (0.997-1.004)	0.416
IGFBP-3	4	<b>1.000 (1.000-1.000)</b>	<b>0.016</b>	4	1.000 (1.000-1.000)	0.743	1.000 (1.000-1.000)	0.404
Triglycerides	124	0.973 (0.886-1.068)	0.562	121	1.001 (0.997-1.005)	0.634	1.001 (0.997-1.005)	0.653
HDL	131	0.970 (0.891-1.056)	0.488	129	1.001 (0.997-1.006)	0.535	0.999 (0.998-1.000)	0.559
LDL	101	0.981 (0.888-1.084)	0.707	102	1.003 (1.000-1.007)	0.054	0.999 (0.997-1.002)	0.055
Testosterone levels	40	1.000 (0.996-1.004)	0.854	17	1.000 (0.997-1.003)	0.836	1.000 (0.998-1.003)	0.783
Sex hormone-binding globulin levels	23	0.912 (0.791-1.051)	0.202	21	1.001 (0.995-1.008)	0.659	0.970 (0.892-1.054)	0.641
Interleukin-6 levels	21	1.025 (0.951-1.104)	0.520	19	0.999 (0.997-1.002)	0.694	1.000 (1.000-1.000)	0.712
<b>Diseases</b>								
Hypertension	41	1.058 (0.994-1.187)	0.332	40	1.000 (0.996-1.004)	0.854	1.000 (0.997-1.003)	0.828
Type 2 diabetes	141	1.029 (0.967-1.095)	0.363	147	1.000 (0.998-1.002)	0.863	1.003 (0.997-1.009)	0.889
Periodontitis	110	1.006 (0.991-1.021)	0.451	105	1.001 (1.000-1.001)	0.078	1.000 (0.999-1.002)	0.073
Benign prostatic hyperplasia	7	1.027 (0.985-1.070)	0.212	5	1.000 (0.997-1.003)	0.898	0.995 (0.990-1.000)	0.647
Systemic lupus erythematosus	66	0.999 (0.984-1.013)	0.860	63	<b>0.999 (0.999-1.000)</b>	<b>0.039</b>	<b>0.999 (0.999-1.000)</b>	<b>0.039</b>
Schizophrenia	134	0.916 (0.836-1.004)	0.062	111	1.001 (0.997-1.005)	0.593	1.001 (1.000-1.001)	0.469
Parkinson's disease	101	0.992 (0.960-1.025)	0.637	101	1.000 (0.999-1.002)	0.733	0.999 (0.997-1.001)	0.750
Multiple sclerosis	310	1.011 (0.967-1.057)	0.623	290	0.999 (0.997-1.001)	0.280	1.000 (0.997-1.003)	0.290
Metabolic syndrome	76	0.999 (0.939-1.062)	0.975	68	1.000 (0.997-1.003)	0.818	1.003 (1.000-1.007)	0.819
<b>Body measurement</b>								
Body mass index	195	1.023 (0.906-1.156)	0.711	192	<b>0.995 (0.990-1.000)</b>	<b>0.047</b>	<b>0.995 (0.990-1.000)</b>	<b>0.048</b>
Height	164	1.015 (0.950-1.086)	0.655	164	1.000 (0.997-1.003)	0.916	1.001 (0.997-1.006)	0.993
Waist circumference	49	0.913 (0.786-1.060)	0.230	48	1.000 (0.992-1.008)	0.922	0.986 (0.925-1.051)	0.665
<b>Drugs</b>								
Aspirin use measurement	9	1.108 (0.970-1.265)	0.130	10	1.006 (0.996-1.016)	0.230	1.031 (0.950-1.119)	0.466

ALA, alpha-linolenic acid; CI, confidence interval; DHA, docosahexaenoic acid; HDL, high density lipoprotein; IGF-I, Insulin-like growth factor-I; IGFBP-3, insulin-like growth factor binding protein 3; LDL, low density lipoprotein; OR, odds ratio; PRACTICAL, Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; SNP, single nucleotide polymorphisms.

The bold means statistical significant.

PUFAs mainly include ALA, EPA, docosapentaenoic acid (DPA), and DHA. Interestingly, our study suggested an inverse association of blood DHA concentration and prostate cancer, whereas no association was observed between the genetically predicted ALA levels and prostate cancer risk. Nevertheless,  $\omega$ -3 PUFAs were demonstrated to have anti-inflammatory and anti-tumor effects (25). A considerable number of studies, including both animal and *in vitro* cell studies, have indicated that  $\omega$ -3 PUFAs are the most promising type of nutrients to suppress carcinogenesis and can reduce prostate cancer risk (26–28). Results from observational studies, however, have been inconsistent. Therefore, studies with larger sample sizes and longer follow-up times are warranted to confirm the results.

In the present study, we found that higher BMI was associated with a reduced prostate cancer risk, and the results were consistent with previous Mendelian randomization studies (29, 30). However, no strong evidence was found in a recent Mendelian randomization study of a causal effect of either early

or later life BMI on prostate cancer (31). Besides, observational studies also reported inconsistent results since the association between BMI and prostate cancer is complex. This complex relation might be owing to the different effects of obesity on various hormones in men, such as a positive association with estrogen concentrations (32) but an inverse association with prostate-specific antigen (33). Another explanation may be the dual effect of BMI on prostate cancer. A meta-analysis of prospective studies suggested that high BMI may protect against localized prostate cancer, whereas it was a risk factor for advanced prostate cancer (34).

The relationship between SLE and cancer is also intriguing. Epidemiological evidence has suggested an increased risk of some malignancies, such as lung cancer, liver cancer, cervical cancer, and especially some hematologic cancers among patients with SLE. However, several studies found a decreased risk of some hormone-sensitive cancers, such as breast, ovarian, and endometrial cancer, in patients with SLE. Interestingly, as reported by the largest

meta-analysis of the cohort study, our Mendelian randomization analysis further supported a protective effect of genetically predicted SLE on prostate cancer risk (35). However, the underlying mechanism remains unclear. Sex hormones might play a putative role in the pathogenesis of prostate cancer in males with SLE (36). As we know, androgens mediate cell proliferation in prostate tissue and are thus important in the development and progression of prostate cancer (37–40). In particular, there is some evidence that males with SLE tend to have low testosterone levels, as compared to males without SLE (41, 42), and men with low circulating free testosterone may carry a lower risk of prostate cancer (43). Further investigations are warranted.

Although our study identified several causal factors for prostate cancer, several limitations should be concerned. A total of 30 factors were included in current study, and a Bonferroni-corrected *p*-value was considered significant to address multiple testing, with a *p*-value <0.0016 being considered suggestive of an association ( $0.05/30 = 0.0016$ ).

Nevertheless, we found no evidence in support of a relationship between other factors and prostate cancer risk. On the one hand, although we identify all potential risk factors for prostate cancer reported by the most recent and largest published meta-analysis of cohort studies, some meta-analyses were still limited by the small amount of literature or studies with small sample size or large heterogeneity among studies. On the other hand, the results of several exposures, such as DHA and folate, were based on one single genetic variant, which might lead to lower precision. Besides, the *F*-statistics for all the genetic instruments were large (>10) in our study, except for schizophrenia, indicating strong genetic instruments that are associated with the exposure. However, the percentage of variation explained was low (<3%) for most of the exposures-specific instruments, and future investigations are needed to identify additional variants to further improve the instrument strength. Considering the inconsistent results reported by previous meta-analyses of cohort studies and our research, well-designed cohort studies with larger sample sizes and Mendelian randomization analysis using more genetic variants are needed to verify these associations further.

Our study also has other limitations. First, all GWAS summary statistics used in our study were from European ancestry participants, limiting the inference of findings in other populations. Second, though our study included as many as 30 extrinsic exposures, several other important exposures, such as dietary calcium, physical activity, Cadmium, Chromium, and plasma/serum lycopene, were not included due to unavailable genetic instruments for analyses. Third, due to the lack of individual data, we were unable to test the association of genetic instruments with other confounders such as BMI, smoking, alcohol consumption, and other lifestyle-related factors. Fourth, since the data on advanced-stage prostate cancer were not available, we only investigated the associations between extrinsic exposures and the overall prostate cancer risk. Finally, although our meta-analysis suggested a positive association between IGF-1 and prostate cancer, however, the

data generated from one study with three datasets and our Mendelian randomization results were unable to support this association, and therefore, more studies are required to confirm this finding.

In conclusion, we conducted a phenome-wide exposed-omics analysis and found a total of 55 factors for prostate cancer risk. The Mendelian randomization analysis verified the IGFBP-3, DHA, BMI, and SLE were causally related to prostate cancer risk. The results could help the clinicians to tailor individualized prophylactic strategies and may provide new insight into the study of the pathogenesis of prostate cancer. More Mendelian randomization studies with larger sample size and stronger power to explain the variance were needed to confirm the results further.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

Our study is a secondary analysis of existing, de-identified article data, or summary-level GWAS data. Specific ethics in this study can be found in the original publications.

## AUTHOR CONTRIBUTIONS

All authors contributed significantly to this work. BZ and DG designed the research study. DG, MT, and YW collected the data. HC, MZ, YB, ZZ, YT, and XW analyzed the data. DG wrote the first draft of the manuscript. All authors reviewed, edited, and approved the manuscript.

## FUNDING

This study was supported by the National Natural Science Foundation of China (81903393, and 81903398), Chongqing Natural Science Foundation Program (cstc2020jcyj-msxmX0021), and Chongqing Special Postdoctoral Science Foundation (XmT2018068). The sponsors of this study had no role in study design, data collection, analysis, interpretation, writing of the report, or the decision for submission.

## ACKNOWLEDGMENTS

We would thank all participants in the PRACTICAL consortium, CRUK, BPC3, CAPS, PEGASUS, and UK Biobank. The Prostate cancer genome-wide association analyses are supported by the Canadian Institutes of Health Research, European Commission's

Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative). We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now PCUK), The Orchid Cancer Appeal, Rosetrees Trust, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK acknowledge the Intramural Program of the National Human Genome Research Institute for their support. Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for ELucidating Loci Involved in Prostate cancer Susceptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN2682012000081] and by Cancer Research UK grant A8197/A16565. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). Funding for the iCOGS infrastructure came from: the European Community's

Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 – the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The BPC3 was supported by the U.S. National Institutes of Health, National Cancer Institute (cooperative agreements U01-CA98233 to D.J.H., U01-CA98710 to S.M.G., U01-CA98216 to E.R., and U01-CA98758 to B.E.H., and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics). CAPS GWAS study was supported by the Swedish Cancer Foundation (grant no 09-0677, 11-484, 12-823), the Cancer Risk Prediction Center (CRisP; [www.crispcenter.org](http://www.crispcenter.org)), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, Swedish Research Council (grant no K2010-70X-20430-04-3, 2014-2269). PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.829248/full#supplementary-material>

## REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistic. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
- Zi H, He SH, Leng XY, Xu XF, Huang Q, Weng H, et al. Global, Regional, and National Burden of Kidney, Bladder, and Prostate Cancers and Their Attributable Risk Factors 1990–2019. *Mil Med Res* (2021) 8(1):60. doi: 10.1186/s40779-021-00354-z
- GBD 2019 Risk Factors Collaborators. Global Burden of 87 Risk Factors in 204 Countries and Territories 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *Lancet* (2020) 396(10258):1223–49. doi: 10.1016/S0140-6736(20)30752-2
- Smith GD, Ebrahim S. 'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease? *Int J Epidemiol* (2003) 32(1):1–22. doi: 10.1093/ije/dyg070
- Davies NM, Holmes MV, Davey Smith G. Reading Mendelian Randomisation Studies: A Guide, Glossary, and Checklist for Clinicians. *BMJ* (2018) 362: k601. doi: 10.1136/bmj.k601
- Pierce BL, Burgess S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. *Am J Epidemiol* (2013) 178(7):1177–84. doi: 10.1093/aje/kwt084
- Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG Consortium, E.-I. Using Published Data in Mendelian Randomization: A Blueprint for Efficient Identification of Causal Risk Factors. *Eur J Epidemiol* (2015) 30(7):543–52. doi: 10.1007/s10654-015-0011-z
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PloS Med* (2015) 12(3): e1001779. doi: 10.1371/journal.pmed.1001779
- Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association Analyses of More Than 140,000 Men Identify 63 New Prostate Cancer Susceptibility Loci. *Nat Genet* (2018) 50(7):928–36. doi: 10.1038/s41588-018-0142-8
- Karlsson Linner R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-Wide Association Analyses of Risk Tolerance and Risky Behaviors in Over 1 Million Individuals Identify Hundreds of Loci and Shared Genetic Influences. *Nat Genet* (2019) 51(2):245–57. doi: 10.1038/s41588-018-0309-3
- Larsson SC, Carter P, Kar S, Vithayathil M, Mason AM, Michaelsson K, et al. Smoking, Alcohol Consumption, and Cancer: A Mendelian Randomisation Study in UK Biobank and International Genetic Consortia Participants. *PloS Med* (2020) 17(7):e1003178. doi: 10.1371/journal.pmed.1003178
- Yarmolinsky J, Bonilla C, Haycock PC, Langdon RJQ, Lotta LA, Langenberg C, et al. Circulating Selenium and Prostate Cancer Risk: A Mendelian Randomization Analysis. *J Natl Cancer Inst* (2018) 110(9):1035–8. doi: 10.1093/jnci/djy081
- Roddam AW, Allen NE, Appleby P, Key TJ, Ferrucci L, Carter HB, et al. Insulin-Like Growth Factors, Their Binding Proteins, and Prostate Cancer Risk: Analysis of Individual Patient Data From 12 Prospective Studies. *Ann Intern Med* (2008) 149(7):461–71, W483–68. doi: 10.7326/0003-4819-149-7-200810070-00006



15. Rowlands MA, Holly JM, Gunnell D, Donovan J, Lane JA, Hamdy F, et al. Circulating Insulin-Like Growth Factors and IGF-Binding Proteins in PSA-Detected Prostate Cancer: The Large Case-Control Study ProtecT. *Cancer Res* (2012) 72(2):503–15. doi: 10.1158/0008-5472.CAN-11-1601
16. Travis RC, Appleby PN, Martin RM, Holly JMP, Albanes D, Black A, et al. A Meta-Analysis of Individual Participant Data Reveals an Association Between Circulating Levels of IGF-I and Prostate Cancer Risk. *Cancer Res* (2016) 76(8):2288–300. doi: 10.1158/0008-5472.CAN-15-1551
17. Bonilla C, Lewis SJ, Rowlands MA, Gaunt TR, Davey Smith G, Gunnell D, et al. Assessing the Role of Insulin-Like Growth Factors and Binding Proteins in Prostate Cancer Using Mendelian Randomization: Genetic Variants as Instruments for Circulating Levels. *Int J Cancer* (2016) 139(7):1520–33. doi: 10.1002/ijc.30206
18. Tan VY, Biernacka KM, Dudding T, Bonilla C, Gilbert R, Kaplan RC, et al. Reassessing the Association Between Circulating Vitamin D and IGFBP-3: Observational and Mendelian Randomization Estimates From Independent Sources. *Cancer Epidemiol Biomarkers Prev* (2018) 27(12):1462–71. doi: 10.1158/1055-9965.EPI-18-0113
19. Firth SM, Baxter RC. Cellular Actions of the Insulin-Like Growth Factor Binding Proteins. *Endocr Rev* (2002) 23(6):824–54. doi: 10.1210/er.2001-0033
20. Ali O, Cohen P, Lee KW. Epidemiology and Biology of Insulin-Like Growth Factor Binding Protein-3 (IGFBP-3) as an Anti-Cancer Molecule. *Horm Metab Res* (2003) 35(11-12):726–33. doi: 10.1055/s-2004-814146
21. Pollitt NS, Inouye M. Synthesis of an Escherichia Coli Protein Carrying a Signal Peptide Mutation Causes Depolarization of the Cytoplasmic Membrane Potential. *J Bacteriol* (1988) 170(5):2051–5. doi: 10.1128/jb.170.5.2051-2055.1988
22. Fang P, Hwa V, Little BM, Rosenfeld RG. IGFBP-3 Sensitizes Prostate Cancer Cells to Interferon-Gamma-Induced Apoptosis. *Growth Horm IGF Res* (2008) 18(1):38–46. doi: 10.1016/j.ghir.2007.07.002
23. Larsson SC, Carter P, Vithayathil M, Kar S, Mason AM, Burgess S. Insulin-Like Growth Factor-1 and Site-Specific Cancers: A Mendelian Randomization Study. *Cancer Med* (2020) 9(18):6836–42. doi: 10.1002/cam4.3345
24. Watts EL, Fensom GK, Smith Byrne K, Perez-Cornago A, Allen NE, Knuppel A, et al. Circulating Insulin-Like Growth Factor-I, Total and Free Testosterone Concentrations and Prostate Cancer Risk in 200 000 Men in UK Biobank. *Int J Cancer* (2021) 148(9):2274–88. doi: 10.1002/ijc.33416
25. Berquin IM, Edwards IJ, Kridel SJ, Chen YQ. Polyunsaturated Fatty Acid Metabolism in Prostate Cancer. *Cancer Metastasis Rev* (2011) 30(3-4):295–309. doi: 10.1007/s10555-011-9299-7
26. Akinsete JA, Ion G, Witte TR, Hardman WE. Consumption of High Omega-3 Fatty Acid Diet Suppressed Prostate Tumorigenesis in C3(1) Tag Mice. *Carcinogenesis* (2012) 33(1):140–8. doi: 10.1093/carcin/bgr238
27. Bianchini F, Giannoni E, Serni S, Chiarugi P, Calorini L. 22 : 6n-3 DHA Inhibits Differentiation of Prostate Fibroblasts Into Myofibroblasts and Tumorigenesis. *Br J Nutr* (2012) 108(12):2129–37. doi: 10.1017/S0007114512000359
28. Wang S, Wu J, Suburu J, Gu Z, Cai J, Axanova LS, et al. Effect of Dietary Polyunsaturated Fatty Acids on Castration-Resistant Pten-Null Prostate Cancer. *Carcinogenesis* (2012) 33(2):404–12. doi: 10.1093/carcin/bgr290
29. Davies NM, Gaunt TR, Lewis SJ, Holly J, Donovan J, Hamdy FC, et al. The Effects of Height and BMI on Prostate Cancer Incidence and Mortality: A Mendelian Randomization Study in 20,848 Cases and 20,214 Controls From the PRACTICAL Consortium. *Cancer Causes Control* (2015) 26(11):1603–16. doi: 10.1007/s10552-015-0654-9
30. Kazmi N, Haycock P, Tsilidis K, Lynch BM, Truong T, Practical Consortium, C.B.C.P, et al. Appraising Causal Relationships of Dietary, Nutritional and Physical-Activity Exposures With Overall and Aggressive Prostate Cancer: Two-Sample Mendelian-Randomization Study Based on 79 148 Prostate-Cancer Cases and 61 106 Controls. *Int J Epidemiol* (2020) 49(2):587–96. doi: 10.1093/ije/dyz235
31. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of Genetic Variation to Separate the Effects of Early and Later Life Adiposity on Disease Risk: Mendelian Randomisation Study. *BMJ* (2020) 369:m1203. doi: 10.1136/bmj.m1203
32. Giovannucci E, Rimm EB, Liu Y, Leitzmann M, Wu K, Stampfer MJ, et al. Investigating the Prostate Specific Antigen, Body Mass Index and Age Relationship: Is an Age-BMI-Adjusted PSA Model Clinically Useful? *Cancer Causes Control* (2016) 27(12):1465–74. doi: 10.1007/s10552-016-0827-1
33. Harrison S, Tilling K, Turner EL, Lane JA, Simpkin A, Davis M, et al. Investigating the Prostate Specific Antigen, Body Mass Index and Age Relationship: Is an Age-BMI-Adjusted PSA Model Clinically Useful? *Cancer Causes Control* (2016) 27(12):1465–74. doi: 10.1007/s10552-016-0827-1
34. Discacciati A, Orsini N, Wolk A. Body Mass Index and Incidence of Localized and Advanced Prostate Cancer—A Dose-Response Meta-Analysis of Prospective Studies. *Ann Oncol* (2012) 23(7):1665–71. doi: 10.1093/annonc/mdr603
35. Ni J, Qiu LJ, Hu LF, Cen H, Zhang M, Wen PF, et al. Lung, Liver, Prostate, Bladder Malignancies Risk in Systemic Lupus Erythematosus: Evidence From a Meta-Analysis. *Lupus* (2014) 23(3):284–92. doi: 10.1177/0961203313520060
36. Bernatsky S, Ramsey-Goldman R, Gordon C, Clarke AE. Prostate Cancer in Systemic Lupus Erythematosus. *Int J Cancer* (2011) 129(12):2966–9. doi: 10.1002/ijc.25956
37. Pollard M, Luckert PH, Schmidt MA. Induction of Prostate Adenocarcinomas in Lobund Wistar Rats by Testosterone. *Prostate* (1982) 3(6):563–8. doi: 10.1002/pros.2990030605
38. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple Loci Identified in a Genome-Wide Association Study of Prostate Cancer. *Nat Genet* (2008) 40(3):310–5. doi: 10.1038/ng.91
39. Bu H, Narisu N, Schlick B, Rainer J, Manke T, Schafer G, et al. Putative Prostate Cancer Risk SNP in an Androgen Receptor-Binding Site of the Melanophilin Gene Illustrates Enrichment of Risk SNPs in Androgen Receptor Target Sites. *Hum Mutat* (2016) 37(1):52–64. doi: 10.1002/humu.22909
40. Dobbs RW, Malhotra NR, Greenwald DT, Wang AY, Prins GS, Abern MR. Estrogens and Prostate Cancer. *Prostate Cancer Prostatic Dis* (2019) 22(2):185–94. doi: 10.1038/s41391-018-0081-6
41. Mok CC, Lau CS. Profile of Sex Hormones in Male Patients With Systemic Lupus Erythematosus. *Lupus* (2000) 9(4):252–7. doi: 10.1191/096120300680198926
42. Mononen N, Schleutker J. Polymorphisms in Genes Involved in Androgen Pathways as Risk Factors for Prostate Cancer. *J Urol* (2009) 181(4):1541–9. doi: 10.1016/j.juro.2008.11.076
43. Watts EL, Appleby PN, Perez-Cornago A, Bueno-de-Mesquita HB, Chan JM, Chen C, et al. Low Free Testosterone and Prostate Cancer Risk: A Collaborative Analysis of 20 Prospective Studies. *Eur Urol* (2018) 74(5):585–94. doi: 10.1016/j.eururo.2018.07.024

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gu, Tang, Wang, Cui, Zhang, Bai, Zeng, Tan, Wang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Novel Nomogram for Prediction and Evaluation of Lymphatic Metastasis in Patients With Renal Cell Carcinoma

## OPEN ACCESS

### Edited by:

Gong-Hong Wei,  
Fudan University, China

### Reviewed by:

Felice Crocetto,  
Federico II University Hospital, Italy  
Ping Gao,  
Shaanxi Normal University, China

### \*Correspondence:

Chengliang Yin  
chengliangyin@163.com  
Meijin Huang  
hmj986@163.com  
Xiaofeng Xu  
XFxu2918@163.com  
Ximin Qiao  
qxmxxyy@163.com

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 10 January 2022

**Accepted:** 16 March 2022

**Published:** 11 April 2022

### Citation:

Li W, Wang B, Dong S, Xu C,  
Song Y, Qiao X, Xu X, Huang M  
and Yin C (2022) A Novel  
Nomogram for Prediction and  
Evaluation of Lymphatic  
Metastasis in Patients With  
Renal Cell Carcinoma.  
Front. Oncol. 12:851552.  
doi: 10.3389/fonc.2022.851552

Wenle Li<sup>1†</sup>, Bing Wang<sup>1†</sup>, Shengtao Dong<sup>2†</sup>, Chan Xu<sup>1</sup>, Yang Song<sup>3</sup>, Ximin Qiao<sup>1,4\*</sup>, Xiaofeng Xu<sup>1,4\*</sup>, Meijin Huang<sup>5\*</sup> and Chengliang Yin<sup>6\*</sup>

<sup>1</sup> Clinical Medical Research Center, Xianyang Central Hospital, Xianyang, China, <sup>2</sup> Department of Spine Surgery, Second Affiliated Hospital of Dalian Medical University, Dalian, China, <sup>3</sup> Department of Gastroenterology and Hepatology, Chinese People's Liberation Army (PLA) General Hospital, Beijing, China, <sup>4</sup> Department of Urology, Xianyang Central Hospital, Xianyang, China, <sup>5</sup> Department of Oncology, 920th Hospital of People's Liberation Army (PLA) Joint Logistics Support Force, Yunnan, China, <sup>6</sup> Faculty of Medicine, Macau University of Science and Technology, Macau, Macau SAR, China

**Background:** Lymphatic metastasis is an important mechanism of renal cell carcinoma (RCC) dissemination and is an indicator of poor prognosis. Therefore, we aimed to identify predictors of lymphatic metastases (LMs) in RCC patients and to develop a new nomogram to assess the risk of LMs.

**Methods:** This study included patients with RCC from 2010 to 2018 in the Surveillance, Epidemiology, and Final Results (SEER) database into the training cohort and included the RCC patients diagnosed during the same period in the Second Affiliated Hospital of Dalian Medical University into the validation cohort. Univariate and multivariate logistic regression analysis were performed to identify risk factors for LM, constructing a nomogram. The receiver operating characteristic (ROC) curves were generated to assess the nomogram's performance, and the concordance index (C-index), area under curve value (AUC), and calibration plots were used to evaluate the discrimination and calibration of the nomogram. The nomogram's clinical performance was evaluated by decision curve analysis (DCA), probability density function (PDF) and clinical utility curve (CUC). Furthermore, Kaplan-Meier curves were performed in the training and the validation cohort to evaluate the survival risk of the patients with lymphatic metastasis or not. Additionally, on the basis of the constructed nomogram, we obtained a convenient and intuitive network calculator.

**Results:** A total of 41837 patients were included for analysis, including 41,018 in the training group and 819 in the validation group. Eleven risk factors were considered as predictor variables in the nomogram. The nomogram displayed excellent discrimination power, with AUC both reached 0.916 in the training group (95% confidence interval (CI) 0.913 to 0.918) and the validation group (95% CI 0.895 to 0.934). The calibration curves presented that the nomogram-based prediction had good consistency with

practical application. Moreover, Kaplan-Meier curves analysis showed that RCC patients with LMs had worse survival outcomes compared with patients without LMs.

**Conclusions:** The nomogram and web calculator (<https://liwenle0910.shinyapps.io/DynNomapp/>) may be a useful tool to quantify the risk of LMs in patients with RCC, which may provide guidance for clinicians, such as identifying high-risk patients, performing surgery, and establishing personalized treatment as soon as possible.

**Keywords:** nomogram, renal cell carcinoma (RCC), lymphatic metastasis, multicenter, web calculator

## INTRODUCTION

Renal cell carcinoma (RCC) is the most common malignant tumor of the kidney, ranking sixth in men and tenth in women, accounting for 5% to 3% of all tumors (1). The most common subtype of RCC is clear cell RCC, which accounts for approximately 70-80%. Other subtypes include papillary RCC (pRCC, 10-15%) (2), chromophobic RCC (chRCC, 5-10%) (3), the rare collecting duct RCC (cdRCC) and sarcomatoid RCC (srRCC) (4). According to the latest report, more than 140,000 people die from RCC yearly, ranking the 13th most common cause of cancer death worldwide (5). With the improvement of examination methods, despite the fact that most of the lesions are found to be small, there are still a considerable number of patients diagnosed as locally advanced stage, and up to 17% of patients had distant metastases (6). The most common metastatic sites of RCC involve lung, lymph nodes, liver, bone and adrenal glands. Among them, local lymph node metastasis is a most important adverse prognostic factors for adult RCC, resulting in an 8-fold risk of death (7, 8). Therefore, it is critical for clinicians to accurately evaluate the risk of lymph node metastasis and formulate the optimal treatment plan. Anatomically, the lymphatic drainage structure of the kidney is complex, variable and inconsistent, making the discovery, diagnosis, and evaluation of LMs in RCC particularly difficult. As we all know, computed tomography (CT) and magnetic resonance (MRI) are currently the ideal tools for the diagnosis and staging of RCC, but they cannot accurately predict LMs, and their ability to distinguish normal size or micrometastasis is limited (9). In addition to the unclear imaging findings, the low positive rate of intraoperative biopsy can also lead to the failure of detection and diagnosis of lymphatic metastasis in RCC early, thereby limiting the therapeutic effect. Thus, improving the awareness and monitoring of LMs will contribute to improve the prognosis of RCC patients. However, there is currently no relevant research focused on developing an ideal predictive model to predict the risk of LMs in RCC, which means that the probability of occurrence of LMs cannot be quantified. Recently, nomogram is a novel type of prognostic tool, which is widely used in oncology and medicine to help clinicians predict prognosis and make medical decisions (10-14). Therefore, we utilized the Surveillance, Epidemiology, and End Results (SEER) database, which is often used to study rare tumors. The database provides data from 18 cancer registries, including approximately 30% of the U.S. population. To address

this, by integrating different clinical variables, our study first developed a nomogram to predict LMs in RCC patients to provide an individual risk assessment and medical decision-making for patients.

## METHODS

### Study Design and Participants

Patients diagnosed with RCC between 2010 and 2018 from the SEER database through the SEER&STAT software (version 8.3.9.2) were collected and the patients who met the following inclusion criteria were grouped into a training group. Patients diagnosed in the Second Affiliated Hospital of Dalian Medical University at the same time as the SEER database were included in the external verification group.

The inclusion criteria were as follows: (1) the patient was older than 18 years, (2) patients with primary kidney cancer (International Classification of Diseases for Oncology ICD-O. 8120/3 represents transitional cell carcinoma, 8130/3 represents papillary transitional cell carcinoma, 8260/3 represents papillary adenocarcinoma, 8310/3 represents clear cell adenocarcinoma, 8312/3 represents renal cell carcinoma, 8317/3 represents chromophobe renal cell carcinoma) diagnosed between January 1, 2010 and December 31, 2017, (3) no previous or concurrent history of other malignant tumors, (4) according to the 8th American Joint Committee on Cancer (AJCC) TNM Staging Manual, re-staging the enrolled patients, and (5) there were sufficient imaging and pathological results during the follow-up period to assess whether the metastasis was in progress happen.

The exclusion criteria were as follows: (1) patients younger than 18, (2) multiple malignant tumor history or the same period, (3) unable to obtain complete demographic characteristics, including age, gender, race, etc., (4) unable to obtain tumor information, including size, stage, histological type, TNM stage, etc., (5) diagnosis was from cadavers, (6) with unknown LMss and survival time, and (7) cause of death unrelated to RCC or unknown.

This study was approved by the institutional ethics committee.

### Data Collection

A total of 41837 RCC patients fulfilled the inclusion criteria were incorporated into the final analyses. All data of the training group were obtained from the SEER database, and the data of the verification group were obtained from the Second Affiliated Hospital of Dalian Medical University. Fifteen variables that

might be related to the development of LMs in RCC patients were included in the study. Demographic characteristics and clinical variables included age, race, gender, marriage, One primary only or more, primary site, the degree of tumor differentiation, tumor size, histological type, T stage, M stage, with or without LMs, alive or not. The data of the verification group were collected by two researchers and one researcher was responsible for verification. Tumor-related information is provided by clinicians, and pathological information is diagnosed by two pathologists using a double-blind method and reviewed by a senior pathologist.

## Construction of Nomogram and Statistical Analysis

R language (version 4.0.5) and SPSS 25.0 were used for all statistical analyses in this study. The prediction nomogram was constructed based on the patients in the training group and tested by the patients in the validation group. The independent sample *t* test was utilized to analyze continuous variables, and the chi-square test was used to analyze categorical variables. Univariate logistic analysis was carried out to identify LMs-associated risk factors. Variables with a *P* value < 0.05 in univariate analysis were further incorporated into multivariate logistic regression analysis to identify the independent risk factors for LMs in RCC patients. Significant independent risk factors verified by multivariate logistic regression analysis were used to construct nomogram with the “rms” package in R software. Receiver operating characteristic (ROC) curve was drawing by medcalc to test the performance of the predictive model, and area under the curve (AUC) was used to express the recognition ability of the predictive model. The area was larger, the recognition ability was better. Probability density function (PDF) was plotted to identify the key points suitable for clinical application, and clinical utility curve (CUC) was used to compare the clinical benefits under different thresholds. Additionally, the consistency of the model was verified by drawing a calibration curve. The decision curve analysis (DCA) was used to verify and evaluate its clinical applicability. Meanwhile, we also performed Kaplan-Meier survival analysis on the overall survival rate of the included patients and used the log-rank test to determine the significance of the difference between the internal and external cohort survival curves. Furthermore, based on the constructed nomogram, we also provide a convenient and intuitive web calculator. Statistical analysis was performed using SPSS (version 20.0, Chicago, IL, USA). A *p*-value < 0.05 (two-tailed) was considered statically significant. The R language software packages applied for developing predictive model included plyr, rms, foreign, DynNom, regplot, caret, ggDCA, ggpubr, pROC, patchwork, eoffice, gLMsnet, survival.

## RESULTS

### Basic Characteristics of Patients

A total of 41837 RCC patients diagnosed from 2010 to 2017 were enrolled in this study, of which 41018 patients from the SEER

database were included in the training group, and 819 patients from the Second Affiliated Hospital of Dalian Medical University were included in the verification group. The demographic and clinical characteristics of the two groups were collected in **Table 1**. In the training and validation groups, the age of the patients ranged from 55 to 75 years, with a mean age of 64, 65 years, respectively. Most of the patients were male, and the ratio of male to female was roughly similar in the two groups. In the training group, most of them were white (78.11%), and only 1.2% were Chinese. The majority of the patients (58.86% and 65.57%, respectively) were married. The most common histologic subtype was clear cell adenocarcinoma (8310/3) (53.36% and 55.68%, respectively). Primary tumor location was mainly in the kidney, and the degree of differentiation was mostly moderately differentiated, accounting for 34.70% and 37.36% respectively. According to the guidelines of the American Joint Committee on Cancer (AJCC), the most common T stage was T1 (66.11% and 60.68%, respectively). Moreover, the whole population had a relatively low rate of lymph node metastasis, occurred in 2630 (6.41%) patients in the training set and 66 (8.06%) patients in the validation set. In the two groups, only 11.04% and 14.53% of the patients presented with metastatic tumors at diagnosis, respectively. Most of the patients were alive during the follow-up period (74.33% and 73.50%, respectively). There were no statistically significant differences in the lymph node metastasis rate, age, one primary only or more, time, alive or dead, sex and tumor size between the two groups (*P*>0.05). However, there were statistical differences in M-stage, marital status, race, primary site, grade, laterality, pathological and T-stage of the the training and validation groups (*P*<0.05). Additionally, according to the presence or absence of LMs in RCC, all patients were divided into two subgroups: lymph node metastasis negative (LNN=39141) and lymph node metastasis positive (LNP=2696). The difference between the two subgroups was shown in **Table 2**. With the exception of race, variables differed significantly between the two subgroups.

### Independent Risk Factors for Lymphatic Metastasis

In order to determine the LMs-related variables of RCC patients, 16 variables were analyzed. We conducted univariate and multivariate logistic regression analysis to explore independent risk factors for lymphatic metastasis. First, through univariate regression analysis, 15 variables were found to be significantly associated with lymphatic metastasis. Subsequently, after conducting multivariate regression analysis, 11 variables: age, marriage, one primary only or more, liver metastasis, lung metastasis, M staging, T staging, tumor differentiation grade, pathological classification, and tumor size, were identified as independent prognostic factors for lymphatic metastasis in RCC patients (all *P* < 0.05, **Table 3**).

### Construction and Validation of Nomogram

Meaningful clinical indicators after multivariate analysis were included in the constructing a nomogram (**Figure 1**), including: pathological subtype, single/multiple tumors, tumor T, M



**TABLE 1 |** Baseline of patients in the training and validation groups.

Characteristics	Level	Training group (N=41018)	Validation group (N=819)	p
Lymph.node.metastasis (%)	No	38388 (93.59)	753 (91.94)	0.0675
	Yes	2630 (6.41)	66 (8.06)	
M (%)	M0	36490 (88.96)	700 (85.47)	0.002
	M1	4528 (11.04)	119 (14.53)	
Marital (%)	Married	24143 (58.86)	537 (65.57)	<0.0001
	unknown	2002 (4.88)	0 (0.00)	
	unmarried	14873 (36.26)	282 (34.43)	
Age (median [IQR])	not available	64.000 [55.000, 73.000]	65.000 [55.000, 73.000]	0.383
Race.ethnicity (%)	black	5225 (12.74)	0 (0.00)	<0.0001
	Chinese	492 (1.20)	819 (100.00)	
	other	3263 (7.96)	0 (0.00)	
	white	32038 (78.11)	0 (0.00)	
Sequence.number (%)	more	13557 (33.05)	252 (30.77)	0.181
	One primary only	27461 (66.95)	567 (69.23)	
Time (mean (SD))	not available	39.842 (30.760)	37.827 (30.885)	0.0634
status (%)	alive	30487 (74.33)	602 (73.50)	0.6224
	dead	10531 (25.67)	217 (26.50)	
Sex (%)	female	14530 (35.42)	299 (36.51)	0.5448
	male	26488 (64.58)	520 (63.49)	
Primary.Site (%)	C64.9-Kidney	39018 (95.12)	731 (89.26)	<0.0001
	C65.9-Renal pelvis	2000 (4.88)	88 (10.74)	
Grade (%)	Moderately differentiated	14234 (34.70)	306 (37.36)	<0.0001
	Poorly differentiated	8662 (21.12)	242 (29.55)	
	Undifferentiated; anaplastic	3245 (7.91)	68 (8.30)	
	unknown	11602 (28.29)	126 (15.38)	
	Well differentiated	3275 (7.98)	77 (9.40)	
Pathological (%)	8120/3	1082 (2.64)	33 (4.03)	0.0014
	8130/3	998 (2.43)	29 (3.54)	
	8260/3	5130 (12.51)	75 (9.16)	
	8310/3	21888 (53.36)	456 (55.68)	
	8312/3	7398 (18.04)	139 (16.97)	
	8317/3	2160 (5.27)	50 (6.11)	
	other (n<1000)	2362 (5.76)	37 (4.52)	
T (%)	T1	27118 (66.11)	497 (60.68)	0.0021
	T2	4108 (10.02)	98 (11.97)	
	T3	8098 (19.74)	180 (21.98)	
	T4	1061 (2.59)	21 (2.56)	
	TX	633 (1.54)	23 (2.81)	
Tumor.Size (mean (SD))	not available	51.355 (41.109)	51.877 (37.304)	0.7186

IQR, interquartilerange; Other, less than 1,000 cases.

staging, differentiation grade, and tumor size. In the nomogram, the values of specific patients were positioned along each variable axis, and a vertical line was drawn up to the dot axis to obtain the score for each variable. The score of each variable was added to get the total score, which was displayed on the total score line at the bottom of the nomogram. Then we would get the probability by drawing a vertical line from the total score to the LMs axis. In order to evaluate and verify the nomogram, the ROC curve of each independent LMs-associated risk factor was drawn in **Figure 2**. The AUC of the training group and the validation group reached 0.916, with 95% CI (0.913 to 0.918) and (0.895 to 0.934) respectively, indicating that the risk model possessed excellent discriminative ability (**Table 4**). What's more, it showed the univariant association and the discrimination power measured by the AUC for each predictor variable in the training and verification groups (**Table 4**). As shown in **Figure 3**, the calibration chart verified that the predictive ability of the nomogram in the training group was highly consistent with the actual results. The results of DCA indicated that the nomogram

had a significant positive net benefit in the process of predicting risk, confirming its good clinical application value (**Figure 4**). The probability density function (PDF) showed that the distribution of the nomogram probability in non-metastatic patients was sharply clustered, while the distribution in metastatic patients was relatively flat (**Figure 5**). Clinical utility curve (CUC), as a means to assist the translation of model information to the clinician, was used for determining the optimal prediction score threshold for each subgroup. For example, it showed that under the same threshold, the percentage of non-metastatic patients and metastatic patients could be detected (**Figure 5**). Furthermore, in order to assess the effect of lymphatic metastasis on the OS of RCC patients, we performed Kaplan-Meier survival analysis in the two groups of patients. As shown in **Figure 6**, whether in the training group or in the validation group, the OS of different lymph node metastasis status was significantly different ( $P < 0.0001$ ), and the survival rate of patients without lymph node metastasis was significantly higher than that of patients with lymph node

**TABLE 2 |** Baseline renal cancer patients with and without lymph node metastasis.

Characteristics	Level	NLMs (N=39141)	LMs (N=2696)	p
category (%)	Training group	38388 (98.08)	2630 (97.55)	0.0675
	Validation group	753 (1.92)	66 (2.45)	
Marital (%)	Married	23163 (59.18)	1517 (56.27)	<0.0001
	unknown	1916 (4.90)	86 (3.19)	
	unmarried	14062 (35.93)	1093 (40.54)	
Age (median [IQR])	not available	64.000 [55.000, 72.000]	66.000 [57.000, 76.000]	<0.0001
Race.ethnicity (%)	black	4918 (12.56)	307 (11.39)	0.1844
	Chinese	1215 (3.10)	96 (3.56)	
	other	3057 (7.81)	206 (7.64)	
	white	29951 (76.52)	2087 (77.41)	
Sequence.number (%)	more	13160 (33.62)	649 (24.07)	<0.0001
	One primary only	25981 (66.38)	2047 (75.93)	
times (mean (SD))	not available	41.480 (30.663)	15.446 (20.027)	<0.0001
status (%)	alive	30431 (77.75)	658 (24.41)	<0.0001
	dead	8710 (22.25)	2038 (75.59)	
Sex (%)	female	13956 (35.66)	873 (32.38)	0.0006
	male	25185 (64.34)	1823 (67.62)	
Primary.Site (%)	C64.9-Kidney	37455 (95.69)	2294 (85.09)	<0.0001
	C65.9-Renal pelvis	1686 (4.31)	402 (14.91)	
Grade (%)	Moderately differentiated	14373 (36.72)	167 (6.19)	<0.0001
	Poorly differentiated	8286 (21.17)	618 (22.92)	
	Undifferentiated; anaplastic	2679 (6.84)	634 (23.52)	
	unknown	10472 (26.75)	1256 (46.59)	
	Well differentiated	3331 (8.51)	21 (0.78)	
Pathological (%)	8120/3	783 (2.00)	332 (12.31)	<0.0001
	8130/3	936 (2.39)	91 (3.38)	
	8260/3	4972 (12.70)	233 (8.64)	
	8310/3	21526 (55.00)	818 (30.34)	
	8312/3	6774 (17.31)	763 (28.30)	
	8317/3	2168 (5.54)	42 (1.56)	
	other(n<1000)	1982 (5.06)	417 (15.47)	
T (%)	T1	27177 (69.43)	438 (16.25)	<0.0001
	T2	3808 (9.73)	398 (14.76)	
	T3	7075 (18.08)	1203 (44.62)	
	T4	612 (1.56)	470 (17.43)	
	TX	469 (1.20)	187 (6.94)	
Tumor.Size (mean (SD))	not available	48.881 (39.220)	87.434 (49.112)	<0.0001

NLMs, no lymph node metastasis; LMs, lymph node metastasis; Other, less than 1,000 cases.

metastasis. In addition, we created a network calculator (<https://liwenle0910.shinyapps.io/DynNomapp/>) using independent risk factors obtained from the previous analysis, which could quickly and easily obtain the probability of lymph node metastasis in RCC patients.

## DISCUSSION

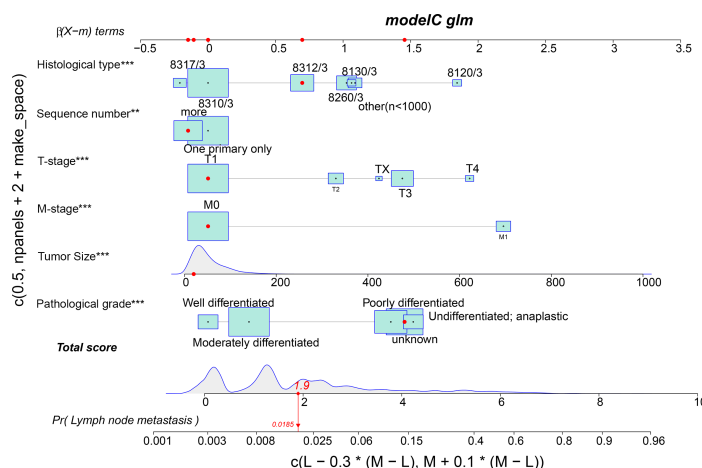
RCC is a deadliest malignant urinary system tumor with high tumor heterogeneity and high recurrence rate (15), especially clear cell renal cell carcinoma (ccRCC) (16). Studies have shown that about 25% of patients with RCC have metastasised at the time of diagnosis, and 35% of them will develop distant metastases (DMs) during the process of tumor progression, resulting in a 5-year survival rate dropping about 10% (17). Due to resistance to chemotherapy and hormone therapy, surgical resection is still the main treatment for RCC at present. Considering the patient's pathology and lymphatic metastasis, clinicians are often confronted with the difficulty of selecting surgical methods and

scope (18). However, there are still quite a few patients undergoing recurrence and metastasis after surgery, which makes it difficult to accurately predict the survival rate of RCC patients. In recent years, with the advent of targeted therapy, median survival for metastatic kidney cancer has roughly doubled. Furthermore, immunotherapy based on anti-PD-1/PD-L1 inhibitors has been shown to be more effective than sunitinib in the first-line treatment of advanced renal cell carcinoma (RCC). To our knowledge, sarcomatoid RCC (srRCC) is prone to metastases with poor prognosis and limited treatment options. A systematic review and meta-analysis found out that sarcomatoid histology might be associated with improved response to PD-1/PDL-1 compared with sunitinib (19). TNM staging is an excellent cancer staging and prognostic system and is determined by the primary tumor stage (the size and extent of tumor expansion), lymph node metastasis and distant metastasis (20, 21). Lymph node metastasis doubles the risk of distant metastasis in patients and has a significant negative impact on the progression-free survival and overall survival of patients with metastatic RCC (15). With the rapid development of multiple imaging methods, the

**TABLE 3 |** Univariate and multivariate Logistic regression for lymphatic metastasis of renal carcinoma.

Characteristics	Univariate logistics			Multivariable logistics		
	OR	CI	P	OR	CI	P
Age	1.01	1.01-1.02	<0.001	1	0.99-1	0.022
Bone.metastases						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	9.67	8.68-10.78	<0.001	1.07	0.93-1.23	0.34
Brain.metastases						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	7.51	6.17-9.14	<0.001	0.91	0.73-1.14	0.41
Unknown	8.37	4.26-16.46	<0.001	0.93	0.43-2	0.845
Grade						
Well differentiated	Ref	Ref	Ref	Ref	Ref	Ref
Moderately differentiated	1.79	1.14-2.83	0.012	1.34	0.84-2.14	0.219
Poorly differentiated	11.31	7.3-17.5	<0.001	3.85	2.45-6.04	<0.001
Undifferentiated; anaplastic	36.45	23.53-56.48	<0.001	4.55	2.88-7.18	<0.001
unknown	18.53	12.01-28.57	<0.001	4.24	2.7-6.64	<0.001
Liver.metastasis						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	16.35	14.26-18.75	<0.001	1.42	1.2-1.67	<0.001
Unknown	10.7	6.04-18.97	<0.001	1.12	0.57-2.2	0.744
M						
M0	Ref	Ref	Ref	Ref	Ref	Ref
M1	21.85	20-23.86	<0.001	7.37	6.31-8.61	<0.001
Marital						
Married	Ref	Ref	Ref	Ref	Ref	Ref
Unmarried	1.2	1.11-1.3	<0.001	1.11	1-1.22	0.047
Unknown	0.69	0.55-0.86	0.001	0.83	0.64-1.07	0.15
Pathological						
8310/3	Ref	Ref	Ref	Ref	Ref	Ref
8312/3	2.95	2.66-3.27	<0.001	2	1.75-2.28	<0.001
8260/3	1.23	1.05-1.43	0.008	2.76	2.32-3.29	<0.001
8317/3	0.5	0.36-0.69	<0.001	0.81	0.58-1.14	0.232
8120/3	11.02	9.5-12.79	<0.001	4.58	3.26-6.45	<0.001
8130/3	2.58	2.05-3.24	<0.001	1.99	1.32-2.99	0.001
other(n<1000)	5.52	4.86-6.27	<0.001	2.87	2.45-3.38	<0.001
Primary.Site						
C64.9-Kidney	Ref	Ref	Ref	Ref	Ref	Ref
C65.9-Renal pelvis	3.9	3.47-4.4	<0.001	1.52	1.08-2.13	0.015
Pulmonary.metastasis						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	15.22	13.89-16.67	<0.001	1.21	1.05-1.4	0.007
Race.ethnicity						
White	Ref	Ref	Ref	Ref	Ref	Ref
Black	0.9	0.79-1.01	0.081	NA	NA	NA
Chinese	0.93	0.64-1.35	0.71	NA	NA	NA
Other	0.97	0.83-1.12	0.657	NA	NA	NA
Sequence number						
One primary only	Ref	Ref	Ref	Ref	Ref	Ref
more	0.62	0.57-0.68	<0.001	0.89	0.8-0.99	0.039
Sex						
Male	Ref	Ref	Ref	Ref	Ref	Ref
Female	0.86	0.79-0.94	0.001	0.96	0.87-1.07	0.475
T						
T1	Ref	Ref	Ref	Ref	Ref	Ref
T2	6.35	5.51-7.32	<0.001	2.53	2.14-2.99	<0.001
T3	10.42	9.3-11.67	<0.001	4.18	3.64-4.8	<0.001
T4	47.28	40.52-55.16	<0.001	6.7	5.53-8.11	<0.001
TX	25.7	21.14-31.26	<0.001	3.49	2.77-4.39	<0.001
Tumor.Size	1.02	1.02-1.02	<0.001	1.00	1.00-1.00	<0.001

OR, odds ratio; 95% CI, 95% confidence interval.



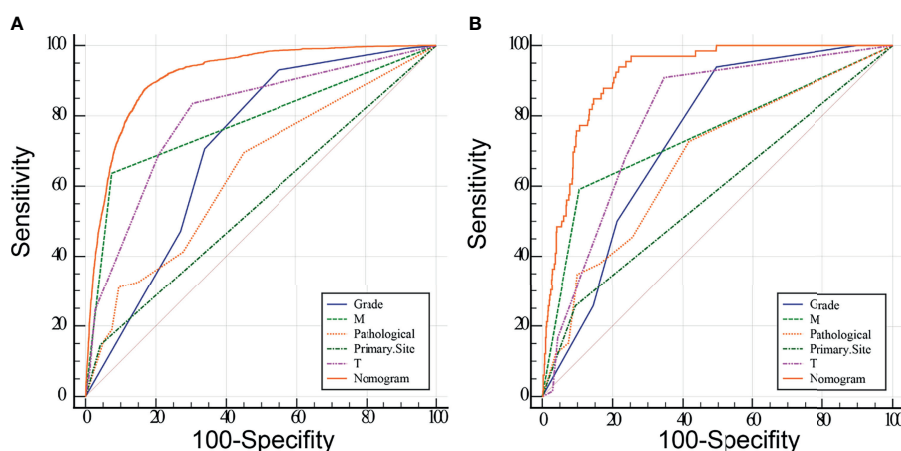
**FIGURE 1** | A nomogram for predicting the risk of lymphatic metastasis in patients with kidney cancer. 8317/3 represents chromophobe renal cell carcinoma, 8310/3 represents clear cell adenocarcinoma, 8312/3 represents renal cell carcinoma, 8260/3 represents papillary adenocarcinoma, 8130/3 represents papillary transitional cell carcinoma, 8120/3 represents transitional cell carcinoma, and other represents the number of patients is less than 1,000. Independent factors, \*\*,  $<0.01$ ; \*\*\*,  $<0.001$ .

identification and detection of lymph node metastasis have improved, but micrometastasis is often overlooked. Therefore, exploring LMs-related predictors and identifying RCC patients with high risk of LMs seem to be of great significance for clinical decision-making and personalized management.

RCC is a type of tumor with gender-biased characteristics (22). According to statistics, the number of cases in men is almost twice that of women. Compared with women, male RCC patients show poorer initial tumor characteristics and higher cancer-specific mortality and worse disease outcomes after surgical treatment (23). Miki et al. analyzed the differences in age and gender of RCC, and the results showed that women had an older age of RCC than men, but the tumor stage and size were

smaller than men (24). Smoking and drinking have been widely recognized as independent risk factors for RCC (22), and these behaviors are mostly found in males, so the incidence of male patients is higher. A study on androgen receptor (AR) overexpression increased blood metastasis but reduced LMs showed that there was also a gender difference between lung metastasis and lymph node metastasis in RCC patients. The results suggested that if the AR was overexpressed, RCC was more likely to metastasize to the lung, and conversely, it was more prone to LMs (25).

The occurrence of RCC is related to a variety of gene mutations and exposure to environmental risk factors (22). Our nomogram showed that the most common pathological



**FIGURE 2** | Receiver operating characteristic (ROC) analyses of the nomogram and each independent predictor based on the training (A) and validation (B) group. The results show that the nomogram has better predictive performance than any single variable.



**TABLE 4 |** Area under the ROC Curve (AUC) of the training and verification groups.

Variable	AUC	SE	95% CI	AUC	SE	95% CI
Grade	0.713	0.00384	0.708 to 0.717	0.74	0.0229	0.708 to 0.769
M	0.781	0.00474	0.777 to 0.785	0.742	0.031	0.711 to 0.772
Pathological	0.64	0.00542	0.635 to 0.644	0.673	0.0334	0.640 to 0.705
Primary.Site	0.552	0.00348	0.547 to 0.557	0.582	0.0276	0.547 to 0.616
T	0.799	0.00429	0.796 to 0.803	0.794	0.0231	0.765 to 0.821
Nomogram	0.916	0.00251	0.913 to 0.918	0.916	0.0133	0.895 to 0.934

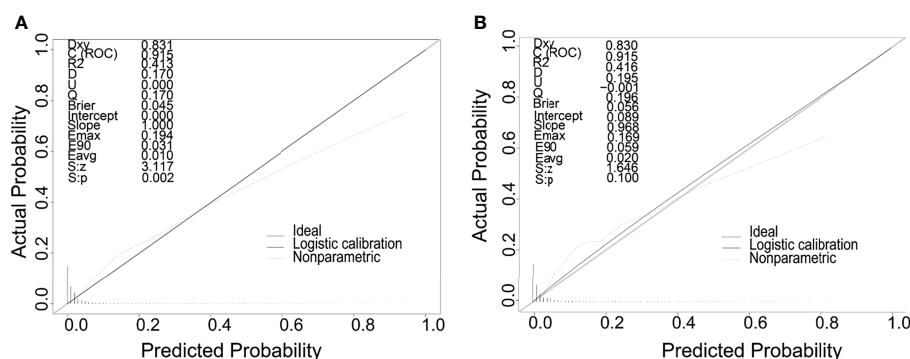
SE, standard error; 95% CI, 95%, confidence interval.

subtype of RCC was ccRCC. The protein-coding mutations for ccRCC have been widely characterized, involving the inactivation of Von Hippel Lindau (VHL) tumor suppressor (16), and the induction of HIF and VEGF. The metastasis of RCC mainly occurs through hematogenous and lymphatic pathways, and the occurrence of these two types of metastases is related to different microvessel density and angiogenesis-specific factors. Among them, the most common site of blood-borne metastasis is the lung. Studies shown that when RCC patients developed LMs, VEGF-C increased, and VEGF-A decreased, When PM occurred, VEGF-A increased, and VEGF-C decreased (25). According to reports, the appearance of PM was significantly related to the difference in progression-free survival (26). Zhang et al. discovered and characterized 17 ccRCC key metastasis-associated genes (MAGs) through single-cell sequencing and found that the increase in MAGs scores was associated with higher T staging, higher lymph node positive rate, late metastasis, poor pathological staging, and tumor grade. Finally, four independent risk factors related to RCC metastasis were determined, including age, tumor grade, pathological stage, and MAG score (27). Our findings were consistent with increasing evidence that the presence of metastasis predicted a worse clinical outcome.

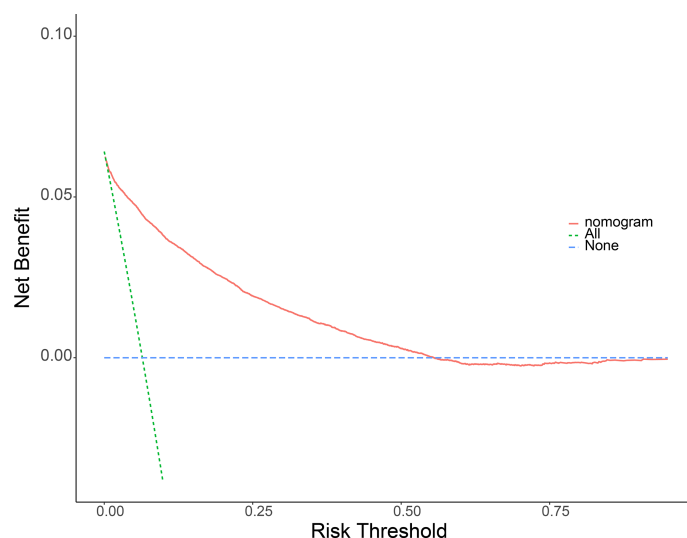
The distribution and drainage of lymph nodes around the kidney are complex and cumbersome, and the existing imaging techniques are still very limited in the ability to identify LMs early. Lymph vessels and lymph nodes are mainly distributed around the veins. Given that lymphatic distribution is closely

related to the course of the intrarenal veins, venous infiltration and lymphatic infiltration are inseparable from the LMs (28, 29). Venous infiltration is common in advanced RCC. It is not only an independent prognostic indicator of patient survival, but also a predictor of recurrence after radical surgery. By using immunohistochemistry to study the relationship between LMs and lymphatic invasion and lymphatic proliferation, it was found that tumor size, tumor cell type, tumor growth pattern, venous invasion, lymphatic invasion, and primary tumor stage were all related to LMs. Ultimately, lymphatic invasion was found to be an independent predictor of LMs in RCC. Moreover, it was considered that the expansion of the tumor and proliferation of lymph nodes around the tumor may increase the chance of tumor cells leaving the primary site (30). It was in line with the risk factors of LMs in RCC patients found in our research.

Radical nephrectomy is the main treatment for RCC. Ideally, kidney disease and lymph nodes in the lymphatic drainage area must be removed, which is one of the important conditions for curing (31). Although LMs is a major factor in determining the clinical stage and predicting the prognosis of patients, there are still controversies about the role of extensive lymphadenectomy in the surgical treatment of RCC and whether it affects the survival of patients. A prospective randomized controlled trial evaluated whether complete lymph node dissection combined with radical nephrectomy was more effective than radical nephrectomy alone. The results could not prove the survival advantage of complete lymph node dissection combined with radical nephrectomy, which might be due to the low incidence of



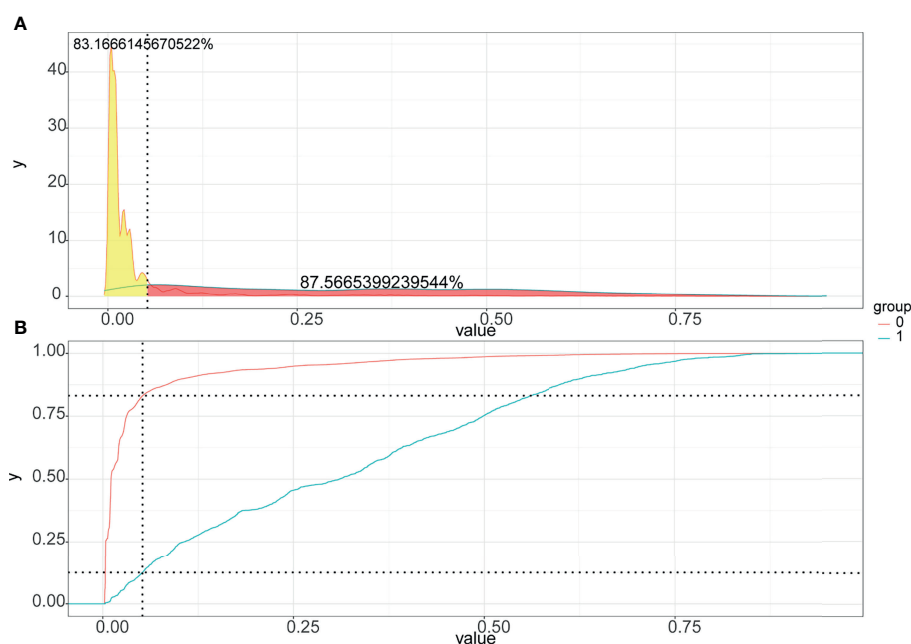
**FIGURE 3 |** Calibration curves of the nomogram for predicting LMs in patients with RCC in the training cohort (A) and the validation cohort (B). The x-axis represents the nomogram-predicted probability of LMs; the y-axis represents the actual probability of LMs.



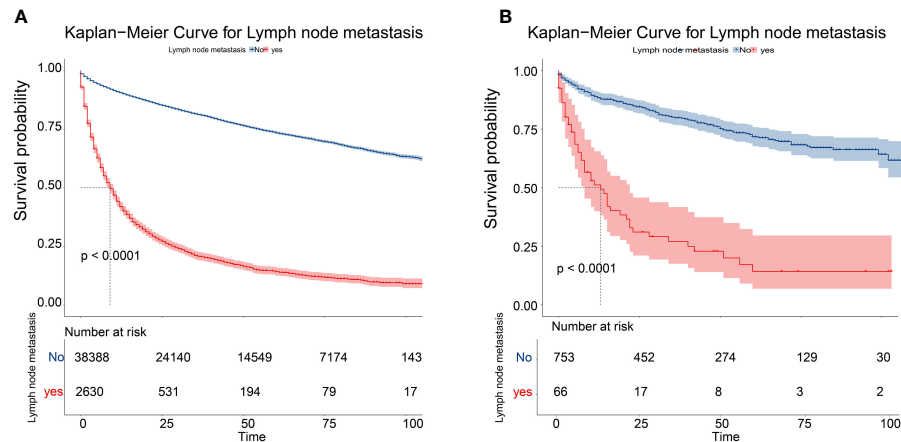
**FIGURE 4** | Decision curve analysis (DCA) of the nomogram for predicting LMs in patients with RCC in the training cohort **(A)** and the validation cohort **(B)**. The blue line represents the hypothesis that all RCC patients do not have lymphatic metastasis. The green line represents the hypothesis that all patients with RCC present lymphatic metastasis. The red line represents the nomogram. The y-axis represents net benefit, and the x-axis represents threshold probability. This diagnostic nomogram shows a notable positive net benefit, indicating that it has a good clinical utility in predicting estimating the risk of LMs in patients with RCC.

unexpected LMs after proper preoperative staging (4.0%) (32). Regardless of the fact that the incidence of LMs is low, in some studies, lymph node involvement has been determined by some studies as an independent risk factor for poor tumor prognosis, and it still needs our attention.

Predicting risks of LMs in RCC patients is crucial for the patient's prognostic consultation. It is also of valuable significance in designing clinical trials, evaluating the clinical results, patient psychological counseling, and programmed management and treatment. Medical nomogram is a model that uses biological and



**FIGURE 5** | Probability density function graph **(A)** and Clinical utility curve **(B)** of the nomogram.



**FIGURE 6 |** Kaplan-Meier overall survival curves of patients with LMs or not in the training cohort (A) and the validation cohort (B).

clinical variables (such as tumor grade and patient age) to graphically describe statistical prognosis and generates the probability of occurrence of individual patients' clinical events (such as cancer recurrence or death), which is widely used in various malignant tumors. Additionally, radiomics and genomics have shown great promise in cancer research, such as improving risk stratification and disease management in prostate cancer (PCa) patients. In the near future, it is also hoped that it can be applied to kidney cancer (33).

Based on the information of patients diagnosed as RCC in the SEER database, this study constructed a nomogram to quantify the risk of LMs in RCC patients and verified it in the patient population from the Second Affiliated Hospital of Dalian Medical University. The total score obtained by combining different risk factors predicted the probability of developing LMs in RCC patients. The higher the score, the higher the risk of LMs. The ROC and AUC analysis showed that the nomogram had excellent predictive ability. The calibration chart indicated that the nomogram had a high degree of consistency in prediction and practical applications. DCA showed that the predictive model had a significant positive net benefit in its application. All these results showed that these independent risk factors were well in predicting LMs in patients with RCC, not only in the training group, but also in the validation group. Although the verification group is small, it can be well verified on the results of the training group. In addition, we will continue to collect more clinical data for Prospective research. Additionally, in conjunction with the risk factors that played an important role, we also created a web calculator (<https://liwenle0910.shinyapps.io/DynNomapp/>) to help clinicians easily and quickly obtain the probability of LMs in RCC patients. Overall, our nomogram may be the first useful method for accurately predicting LMs in patients with RCC to date. However, as a retrospective analysis, there were several limitations in our study, including selection bias, information bias, lack of standardization of diagnosis, treatment and follow-up, missing or unavailability of some information (such as smoking, drinking history), tumors markers, etc. Despite these limitations, our

nomogram was based on a large number of samples, and internal and external verification to ensure the credibility. In the future, a more complete experimental design will be needed to facilitate clinical application.

## CONCLUSION

Through retrospective analysis of RCC patient information from the SEER database and the Second Affiliated Hospital of Dalian Medical University, we obtained the independent factors for lymph node metastasis in patients with RCC. By integrating these factors, we constructed a nomogram to predict the risk of lymphatic metastasis in RCC patients. After drawing a series of verification curves, it was confirmed that the nomogram had good calibration and discrimination. The poor prognosis of LMs patients was confirmed by Kaplan-Meier curve. Moreover, a web version of the nomogram, a simple network calculator, had likewise been established to facilitate clinical application. The nomogram we made can uniquely, conveniently, and intuitively quantify the risk of LMs in RCC patients, and then guide clinicians to predict prognosis and make individualized treatment decisions for patients.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

CY, MH carried out the study design. WL, BW and SD conducted the research and collected and analyzed the data. WL performed the statistical analysis and drafted the manuscript. XQ and XX provided the expert consultations and

suggestions. All conceived the study, participated in its design and coordination, and helped shape the language. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2018. *CA Cancer J Clin* (2018) 68:7–30. doi: 10.3322/caac.21442
- Cancer Genome Atlas Research N, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* (2016) 374(2):135–45. doi: 10.1056/NEJMoa1505917
- Larkin JM, Chowdhury S, Gore ME. Drug Insight: Advances in Renal Cell Carcinoma and the Role of Targeted Therapies. *Nat Clin Pract Oncol* (2007) 4(8):470–9. doi: 10.1038/nponc0901
- Hsieh JJ, Le V, Cao D, Cheng EH, Creighton CJ. Genomic Classifications of Renal Cell Carcinoma: A Critical Step Towards the Future Application of Personalized Kidney Cancer Care With Pan-Omics Precision. *J Pathol* (2018) 244(5):525–37. doi: 10.1002/path.5022
- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S. *GLOBOCAN 2012 V1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*. Lyon, France: International Agency for Research on Cancer (2013).
- Capitanio U, Montorsi F. Renal Cancer. *Lancet* (2016) 387:894–906. doi: 10.1016/S0140-6736(15)00046-X
- Dudani S, de Velasco G, Wells JC, Gan CL, Donskov F, Porta C, et al. Evaluation of Clear Cell, Papillary, and Chromophobe Renal Cell Carcinoma Metastasis Sites and Association With Survival. *JAMA Netw Open* (2021) 4(1):e2021869. doi: 10.1001/jamanetworkopen.2020.21869
- Türkvtan A, Akdur PO, Altinel M, Olçer T, Turhan N, Cumhur T, et al. Preoperative Staging of Renal Cell Carcinoma With Multidetector CT. *Diagn Interv Radiol* (2009) 15(1):22–30.
- Coll DM, Smith RC. Update on Radiological Imaging of Renal Cell Carcinoma. *BJU Int* (2007) 99(5 Pt B):1217–22. doi: 10.1111/j.1464-410X.2007.06824.x
- Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in Oncology: More Than Meets the Eye [J]. *Lancet Oncol* (2015) 16(4):e173–180. doi: 10.1016/S1470-2045(14)71116-7
- Yang J, Pan Z, Zhou Q, Zhao F, Feng X, Lyu J, et al. Nomogram for Predicting the Survival of Patients With Malignant Melanoma: A Population Analysis [J]. *Oncol Lett* (2019) 18(4):3591–98. doi: 10.3892/ol.2019.10720
- Pan YX, Chen JC, Fang AP, Wang X-H, Chen J-B, Wang J-C, et al. A Nomogram Predicting the Recurrence of Hepatocellular Carcinoma in Patients After Laparoscopic Hepatectomy [J]. *Cancer Commun (Lond)* (2019) 39(1):55. doi: 10.1186/s40880-019-0404-6
- Kong J, Zheng J, Cai J, Wu S, Diao X, Xie W, et al. A Nomogram for Individualized Estimation of Survival Among Adult Patients With Adrenocortical Carcinoma After Surgery: A Retrospective Analysis and Multicenter Validation Study [J]. *Cancer Commun (Lond)* (2019) 39(1):80. doi: 10.1186/s40880-019-0426-0
- Narita Y, Kadowaki S, Oze I, Kito Y, Kawakami T, Machida N, et al. Establishment and Validation of Prognostic Nomograms in First-Line Metastatic Gastric Cancer Patients [J]. *J Gastrointest Oncol* (2018) 9(1):52–63. doi: 10.21037/jgo.2017.11.08
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A, et al. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262
- Yao X, Tan J, Lim KJ, Koh J, Ooi WF, Li Z, et al. VHL Deficiency Drives Enhancer Activation of Oncogenes in Clear Cell Renal Cell Carcinoma. *Cancer Discov* (2017) 7(11):1284–305. doi: 10.1158/2159-8290.CD-17-0375
- Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* (2018) 173(3):611–23.e17. doi: 10.1016/j.cell.2018.02.020
- Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and Prognostic Stratification of 670 Patients With Advanced Renal Cell Carcinoma. *J Clin Oncol* (1999) 17(8):2530–40. doi: 10.1200/JCO.1999.17.8.2530
- Buonerba C, Dolce P, Iaccarino S, Scafuri L, Verde A, Costabile F, et al. Outcomes Associated With First-Line Anti-PD-1/ PD-L1 Agents vs. Sunitinib in Patients With Sarcomatoid Renal Cell Carcinoma: A Systematic Review and Meta-Analysis. *Cancers (Basel)* (2020) 12(2):408. doi: 10.3390/cancers12020408
- Giuliani L, Giberti C, Martorana G, Rovida S. Radical Extensive Surgery for Renal Cell Carcinoma: Long-Term Results and Prognostic Factors. *J Urol* (1990) 143(3):468–73; discussion 473–4. doi: 10.1016/S0022-5347(17)39992-5
- Ficarra V, Righetti R, Pilloni S, D'amico A, Maffei N, Novella G, et al. Prognostic Factors in Patients With Renal Cell Carcinoma: Retrospective Analysis of 675 Cases. *Eur Urol* (2002) 41(2):190–8. doi: 10.1016/S0302-2838(01)00027-6
- Scelo G, Larose TLL. Epidemiology and Risk Factors for Kidney Cancer. *J Clin Oncol* (2018) 36(36):Jco2018791905. doi: 10.1200/JCO.2018.79.1905
- Marchioni M, Martel T, Bandini M, Pompe RS, Tian Z, Kapoor A, et al. Marital Status and Gender Affect Stage, Tumor Grade, Treatment Type and Cancer Specific Mortality in T(1–2) N(0) M(0) Renal Cell Carcinoma. *World J Urol* (2017) 35(12):1899–905. doi: 10.1007/s00345-017-2082-9
- Hew MN, Zonneveld R, Kümmerlin IP, Opondo D, de la Rosette JJ, Laguna MP. Age and Gender Related Differences in Renal Cell Carcinoma in a European Cohort. *J Urol* (2012) 188(1):33–8. doi: 10.1016/j.juro.2012.02.2573
- Huang Q, Sun Y, Ma X, Gao Y, Li X, Niu Y, et al. Androgen Receptor Increases Hematogenous Metastasis Yet Decreases Lymphatic Metastasis of Renal Cell Carcinoma. *Nat Commun* (2017) 8(1):918. doi: 10.1038/s41467-017-00701-6
- Motzer RJ, Bukowski RM, Figlin RA, Hutson TE, Michaelson MD, Kim ST, et al. Prognostic Nomogram for Sunitinib in Patients With Metastatic Renal Cell Carcinoma. *Cancer* (2008) 113(7):1552–8. doi: 10.1002/cncr.23776
- Zhang C, He H, Hu X, Liu A, Huang D, Xu Y, et al. Development and Validation of a Metastasis-Associated Prognostic Signature Based on Single-Cell RNA-Seq in Clear Cell Renal Cell Carcinoma. *Aging (Albany NY)* (2019) 11(22):10183–202. doi: 10.18632/aging.102434
- Cuttino JT Jr, Clark RL, Jennette JC. Microangiographic Demonstration of Human Intrarenal Microlymphatic Pathways. *Urol Radiol* (1989) 11(2):83–7. doi: 10.1007/BF02926482
- Ishikawa Y, Akasaka Y, Kiguchi H, Akishima-Fukasawa Y, Hasegawa T, Ito K, et al. The Human Renal Lymphatics Under Normal and Pathological Conditions. *Histopathology* (2006) 49(3):265–73. doi: 10.1111/j.1365-2559.2006.02478.x
- Ishikawa Y, Aida S, Tamai S, Akasaka Y, Kiguchi H, Akishima-Fukasawa Y, et al. Significance of Lymphatic Invasion and Proliferation on Regional Lymph Node Metastasis in Renal Cell Carcinoma. *Am J Clin Pathol* (2007) 128(2):198–207. doi: 10.1309/0FT8WTDKREFHHP4P
- Robson CJ. Radical Nephrectomy for Renal Cell Carcinoma. *J Urol* (1963) 89:37–42. doi: 10.1016/S0022-5347(17)64494-X
- Blom JH, van Poppel H, Maréchal JM, Jacqmin D, Schröder FH, de Prieck L, et al. Radical Nephrectomy With and Without Lymph-Node Dissection: Final Results of European Organization for Research and Treatment of Cancer (EORTC) Randomized Phase 3 Trial 30881. *Eur Urol* (2009) 55(1):28–3. doi: 10.1016/j.eururo.2008.09.052
- Ferro M, de Cobelli O, Dorin Vartolomei M, Lucarelli G, Crocetto F, Barone B, et al. Prostate Cancer Radiogenomics-From Imaging to Molecular Characterization. *Int J Mol Sci* (2021) 22(18):9971. doi: 10.3390/ijms22189971

## ACKNOWLEDGMENTS

We thank all individuals who took part in this research.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Wang, Dong, Xu, Song, Qiao, Xu, Huang and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Fat Attenuation Index of Renal Cell Carcinoma Reveals Biological Characteristics and Survival Outcome

## OPEN ACCESS

### Edited by:

Jonathan Olivier,  
Université de Lille,  
France

### Reviewed by:

Sonia Kiran,  
University of Tennessee Health  
Science Center (UTHSC),  
United States  
Rong Na,  
The University of Hong Kong,  
Hong Kong SAR, China

### \*Correspondence:

Dingwei Ye  
dwyeli@163.com  
Yao Zhu  
yaozhu09@fudan.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Genitourinary Oncology,  
a section of the journal  
Frontiers in Oncology

Received: 30 September 2021

Accepted: 09 May 2022

Published: 09 June 2022

### Citation:

Wang H, Wei Y, Hu X, Pan J, Wu J,  
Wang B, Zhang H, Shi G, Liu X,  
Zhao J, Zhu Y and Ye D (2022) Fat  
Attenuation Index of Renal Cell  
Carcinoma Reveals Biological  
Characteristics and Survival Outcome.  
Front. Oncol. 12:786981.  
doi: 10.3389/fonc.2022.786981

Hongkai Wang<sup>1,2†</sup>, Yu Wei<sup>1,2†</sup>, Xiaoxin Hu<sup>2,3†</sup>, Jian Pan<sup>1,2,4†</sup>, Junlong Wu<sup>1,2</sup>,  
Beihe Wang<sup>1,2</sup>, Hailiang Zhang<sup>1,2</sup>, Guohai Shi<sup>1,2</sup>, Xiaohang Liu<sup>1,2</sup>, Jinou Zhao<sup>1,2</sup>,  
Yao Zhu<sup>1,2\*</sup> and Dingwei Ye<sup>1,2\*</sup>

<sup>1</sup> Department of Urology, Fudan University Shanghai Cancer Center, Shanghai, China, <sup>2</sup> Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, <sup>3</sup> Department of Pathology, Fudan University, Shanghai, China, <sup>4</sup> Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China

**Purpose:** The computed tomography fat attenuation index (FAI) is an ideal quantifiable imaging factor to identify the inflammation degree of peri-tumor adipose tissue. We aimed to verify whether FAI could reflect peri-tumor adipose inflammation, predict the survival outcome of renal cell carcinoma (RCC), and discover transcriptomic features of tumor tissues and adjacent adipocytes.

**Materials and Methods:** Two clinical cohorts (Fudan University Shanghai Cancer Center [FUSCC] cohort [n=129] and TCGA cohort [n=218]) were used to explore the association between FAI and clinical outcome. A prospective cohort (n = 19) was used to discover the molecular phenotyping of peri-tumor adipose tissue and tumor tissue according to their FAI value. A clinical cohort (n = 32) in which patients received cyto-reductive surgery was used to reveal the dynamic change of FAI.

**Results:** A high peri-tumor FAI was significantly associated with a worse outcome in both the FUSCC (HR = 2.28, p = 0.01) and the TCGA cohort (HR = 2.24, p < 0.001). The analysis of the RNA expression of paired RCC tissue and peri-tumor fat tissue showed synchronized alterations in pathways such as cytokine–cytokine receptor interaction and complement and coagulation cascades. RCC tissues showed significant alterations in the neuroactive ligand–receptor interaction pathway. Immune deconvolution analysis showed enhanced infiltration of macrophages in high FAI tumor tissues with a lower angiogenesis level. We also observed synchronous dynamic changes in FAI and tumor size after targeted therapy.

**Conclusion:** In summary, FAI could be used in RCC to reflect the biological characteristics and tumor immune micro-environment of both the tumor and the peri-

tumor adipose. High peri-tumor FAI had the potential to predict a worse survival outcome in various cohorts. This study demonstrates that the crosstalk exists between a tumor and its micro-environment and could be reflected easily by imaging procedures, which could facilitate clinical decision making.

**Keywords:** CT fat attenuation index, obesity paradox, renal cell carcinoma, immune micro-environment, adipocyte

## INTRODUCTION

Obesity is associated with an increased incidence of renal cell carcinoma (RCC) (1). However, a high body mass index (BMI) is believed to be a protective factor for RCC prognosis (2). The association between obesity and RCC is quite complex. Adipose tissue may have played an important role in the “obesity paradox” since it has multiple physiologic and pathophysiologic functions. Albiges and colleagues tried to explain the paradox in cohorts of metastatic RCC, and found that the fatty acid synthase (FASN) pathway activation is associated with BMI and survival, which is linked to lipogenesis of the tumor (3). On the other hand, adipose tissues have paracrine functions by secreting adipokines such as adiponectin and leptin or by secreting inflammatory cytokines such as tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), interleukin-6 (IL-6), interleukin-8 (IL-8), plasminogen activator inhibitor 1 (PAI1), etc., which could facilitate cancer growth (4). Recently, Sanchez and colleagues found that tumors of obese patients showed higher angiogenic scores and that inflammation in the peri-tumoral adipose tissue was increased in obese patients (5). These results showed that interactions exist between the tumor and peri-tumoral adipose tissues, and those interactions could be a good explanation for the obesity paradox.

It is crucial to understand how these interactions can be used in making clinical decisions and predicting outcomes. Hakimi et al. showed that different molecular subgroups of clear cell RCC, especially those with angiogenesis and macrophage infiltration, may be powerful predictors of outcome with tyrosine kinase inhibitor (TKI) efficacy (6). Clark and colleagues characterized the immune infiltration of clear cell RCC into four sub-types, discriminated by the presence or absence of cell types related to immune (CD8+ T cells, macrophages, dendritic cells) and stromal (fibroblast, endothelial) signatures. They announced that these sub-types could be leveraged to predict the therapeutic response, such as immunotherapy (7). Either immune infiltration or angiogenesis could release inflammatory mediators or oxidation products, which could directly modify the phenotype of peri-tumor adipocytes. Since multiple image-based scoring systems have been developed for RCC outcome prediction, namely, RENAL, PADUA, C-index, and Mayo Adhesive Probability Score, is it possible to simplify immune/angiogenesis variables to an easily obtained factor to predict the outcome? We noticed that Antonopoulos and colleagues had developed the computed tomography (CT) fat attenuation index (FAI), which has excellent sensitivity and specificity for detecting tissue inflammation in peri-vascular adipose tissue (8). The development of FAI is an ideal quantifiable factor to help us

identify the exact inflammation degree of peri-tumor adipose tissue.

We verified whether FAI could be used in RCC to reflect peri-tumor adipose inflammation, discovered transcriptomic features of tumor tissues and adjacent adipocytes, and evaluated whether FAI could be a predictor of tumor biological characteristics and survival outcome in various cohorts.

## METHODS

### Study Design, Inclusion Criteria, and Participants

The study design, inclusion criteria, and participants are shown in **Figure 1**. In this study, we analyzed the data from four independent clinical cohorts. All cohorts included patients with clear cell RCC aged 18 years and older. There were no duplicated cases among those cohorts.

#### FAI Cohort A—FUSCC

For the FAI cohort A, all patients were from the Fudan University Shanghai Cancer Center (FUSCC, Shanghai, China) and had been histologically confirmed by surgery from November 2013 to November 2015. All patients had undergone contrast-enhanced CT of the abdomen before surgery. The clinical and pathological characteristics of the patients were collected from the database. All patients were confirmed as having clear cell RCC. The cohort used in this retrospective study was approved by the Ethics Committee of FUSCC.

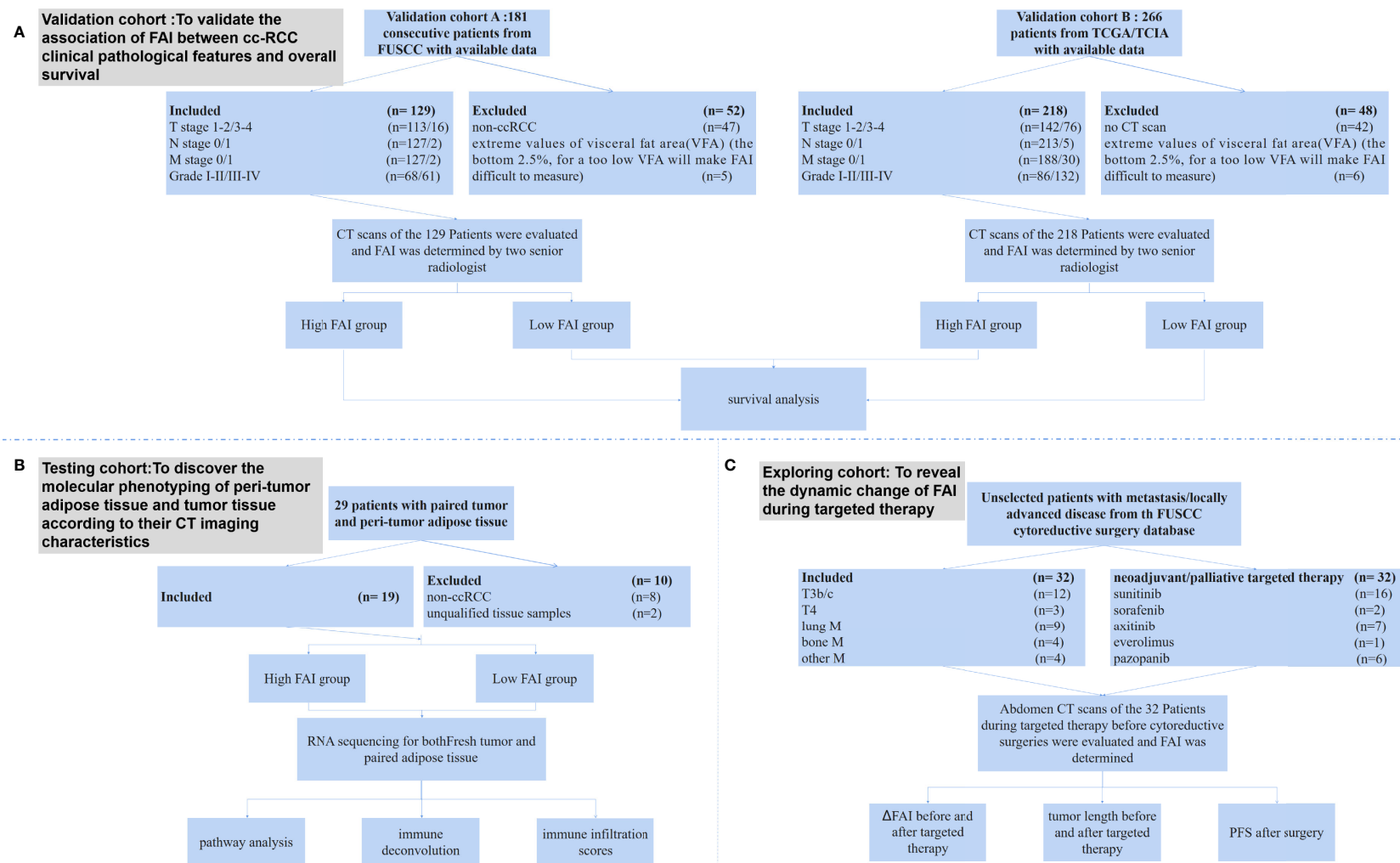
#### FAI Cohort B—TCGA

For the FAI cohort B, all patients were from the TCGA and TCIA databases. Only patients with RNA sequencing data, CT scans before surgery, and pathology confirmed clear cell RCC were included. The survival data of those patients were achieved.

The two validation cohorts were designed to analyze and validate the association between FAI and clinical pathological features and the overall survival of those patients.

#### Imaging Genomics Cohort

For the imaging genomics cohort, we prospectively included consecutive unselected patients from FUSCC who underwent nephrectomy or nephron sparing surgery from July 2019 to September 2019. All patients had CT scans before surgery. Fresh tumor and adipose tissue collected from those patients were used for RNA sequencing. Those who were confirmed to be clear cell RCC were finally included in the study.



**FIGURE 1 |** The study design, inclusion criteria, and participants. **(A)** The validation cohorts were composed by patients from the FUSCC database (FAI cohort A) and TCGA/TCIA database (FAI cohort B). We excluded patients who did not have CT scans and those who had extreme values of VFA. The final included patients were 129 from the FUSCC cohort and 218 from the TCGA/TCIA cohort. **(B)** The imaging genomics cohort prospectively included consecutive unselected patients from FUSCC who had nephrectomy or nephron sparing surgery. After excluding 8 patients who were not clear cell RCC and 2 patients with unqualified tissue samples, the final cohort included 19 patients. **(C)** The treatment response cohort included patients from a FUSCC cytoreductive surgery database. All patients had CT scans of the kidney tumor before and after two months of neoadjuvant/palliative targeted therapy, and then they undergone nephrectomies.

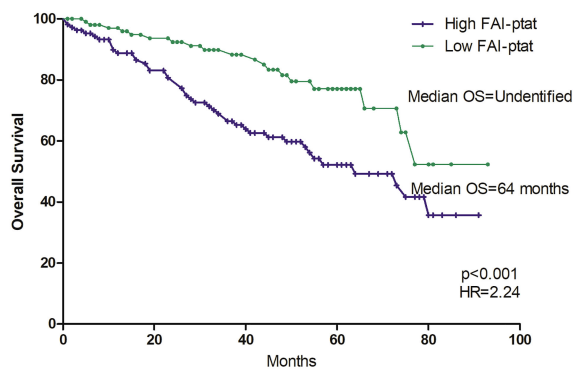


Figure 2A. Kaplan-Meier curves of patients with different FAI-ptat from TCGA cohort.

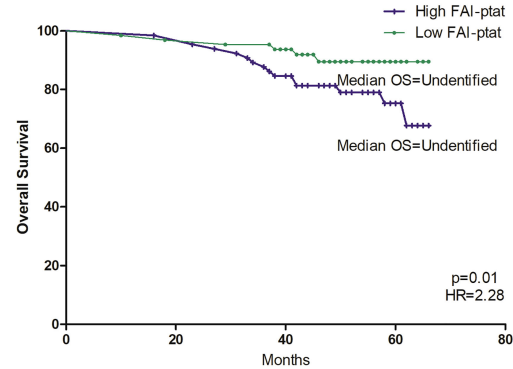


Figure 2B. Kaplan-Meier curves of patients with different FAI-ptat from FUSCC cohort.

**FIGURE 2 |** Kaplan-Meier curves of patients with different FAI<sub>PTAT</sub>. Fat attenuation index (FAI) was calculated and grouped according to the median FAT value of the FUSCC and TCGA cohorts. **(A)** Overall survival of the TCGA cohort, the 5-year median overall survival was 64 months for high FAI<sub>PTAT</sub> group and not reached for low FAI<sub>PTAT</sub> group (HR = 2.24,  $p < 0.001$ ). **(B)** Overall survival of the FUSCC cohort, the 5-year median overall survival was not reached for both groups in FUSCC cohort (HR = 2.28,  $p = 0.01$ ). The 3-year survival rate was 88%/93% for high FAI<sub>PTAT</sub> and low FAI<sub>PTAT</sub> groups in FUSCC cohort (HR = 1.81) and 72%/93% in TCGA cohort (HR = 5.65).

The imaging genomics cohort was to: 1) link molecular phenotyping of peri-tumor adipose tissue and tumor tissue with their CT imaging characteristics; 2) discover whether tumor-adipose interaction exists; and 3) observe how tumor immune micro-environment changes.

### Treatment Response Cohort

For the treatment response cohort, we included patients from a FUSCC cyto-reductive surgery database. All patients had CT scans of the kidney tumor before and after two months of neo-adjuvant/palliative targeted therapy, and then they underwent nephrectomies. After that, targeted therapy was performed to further treat metastasis sites. The clinical and pathological characteristics of the patients were collected.

The treatment response cohort was to discover how FAI changes after targeted therapies.

### Imaging Studies Using CT

Fat attenuation index (FAI): Adipose tissue was defined as voxels with attenuation between  $-190$  and  $-30$  HU. Voxel attenuation histograms were plotted, and FAI was defined as the average attenuation of the adipose tissue volume of interest as previously described (8). FAI<sub>PTAT</sub> was defined as the FAI of the first 5-mm-thick layer of peri-tumor adipose tissue (PTAT) (**Supplementary Figure 1**). All FAIs were determined by two senior radiologists who were blinded to patient characteristics and the mean count was adopted. The intraclass correlation coefficient between the two radiologists from the same CT scan was 0.927 (95% CI 0.909 to 0.937,  $P < 0.001$ ).

### Procedures

The RNA sequencing and data analysis of tumor tissue and peri-tumor fat from the imaging genomics cohort was described in **Supplementary Data 1**. GO (<http://www.geneontology.org/>)

and KEGG (<https://www.kegg.jp/>) enrichment analyses of annotated different expressed genes were performed by Phyper ([http://en.wikipedia.org/wiki/Hypergeometric\\_distribution](http://en.wikipedia.org/wiki/Hypergeometric_distribution)) based on the Hypergeometric test. The protein-protein interaction (PPI) network was predicted using the STRING (<http://string-db.org>) (version 10.0) online database (9).

RNAseq data of the TCGA cohort, mainly of the tumor tissue, were downloaded from the National Institutes of Health Genomic Data Commons. The methods for RNA extraction and processing for the TCGA cohort have previously been published (10). For both the FUSCC and TCGA cohorts, ssGSEA for immune deconvolution analyses was used as previously described (11). The immune infiltration score, the fraction of immune cells (ImmuneScore), each individual immune cell type, angiogenesis score, and hypoxia score were calculated according to previous studies (5).

Multiplex immunofluorescence (mIF) was done to confirm the status of immune cell infiltration. The procedures are shown in **Supplementary Data 2**.

### Statistical Analysis

Continuous data are presented as the median (range), and binary data are presented as proportions. The association between FAI, clinical features, transcriptomic, and genomic differences was tested using Fisher's exact tests, Pearson's tests, and  $\chi^2$  tests. The Kaplan-Meier method was used to determine the overall survival rate. Overall survival rates were compared using the log-rank test. Predictive parameters were assessed in the Cox proportional hazards model, and odds ratios with 95% confidence intervals were calculated. All other analyses were conducted using SPSS 20.0 software (IBM, Chicago, IL, USA). Two-tailed P-values were used, and a  $P < 0.05$  was considered to indicate statistical significance.



## RESULTS

The FAI cohort A from FUSCC included 181 consecutive patients who had CT scans before surgery, adequate clinical-pathological data, and overall survival data (**Figure 1**). After excluding 47 non-ccRCC patients and 5 patients who had extreme values of visceral fat area (VFA) (the bottom 2.5%, for a VFA too low will make FAI difficult to measure), our final cohort comprised 129 patients. The FAI cohort B from TCGA/TCIA included 266 patients with CT/magnetic resonance imaging (MRI) images. We excluded 42 patients who only had MRI scans and 6 patients who had extreme values of VFA (the bottom 2.5%). Our final cohort of cohort B comprised 218 patients. In the imaging genomics cohort for RNA sequencing, we prospectively collected 29 pairs of tumor and peri-tumor adipose tissue from 29 consecutive patients. After excluding 8 patients who were not clear cell RCC and 2 patients with unqualified tissue samples, the final cohort included 19 patients (**Supplementary Table S1**). The treatment response cohort included 32 patients, of which 17 patients had metastasis and 15 patients had tumors that extended into the vena cava or adjacent organs (**Supplementary Table S2**).

### Higher FAI<sub>PTAT</sub> Indicates Worse T Stage, M Stage, Tumor Grade, and Worse Overall Survival Both in the FAI Cohorts A and B

Patients from the FAI cohort A (FUSCC) had fewer T3–4 stage patients, fewer M stage patients, a lower FAI<sub>PTAT</sub> compared with the FAI cohort B (TCGA) cohort, as well as a lower rate of overweight and obese patients. FAI<sub>PTAT</sub> tends to be lower in obese patients. However, no association with BMI was observed in either cohort ( $p = 0.29/p = 0.056$ ). Data showed that FAI<sub>PTAT</sub> were significantly associated with T stage ( $p = 0.028/p = 0.00$ ), M stage ( $p = 0.01/p = 0.00$ ), and tumor grade ( $p = 0.028/p = 0.00$ ) in both the FUSCC and TCGA cohorts (**Table 1**). Necrosis status was available for the FUSCC cohort. Both necrosis ( $p = 0.01$ ) and SSIGN score ( $p = 0.001$ ) were significantly associated with FAI<sub>PTAT</sub>. We set the cutoff of FAI<sub>PTAT</sub> to be  $-93$  hu (median value, ranging from  $-33$  to  $-113$ ) for the FAI cohort A (FUSCC), and  $-79$  hu (median value, ranging from  $-31$  to  $-108$ ) for the FAI cohort B (TCGA), and divided the cohorts into high FAI<sub>PTAT</sub> and low FAI<sub>PTAT</sub> groups. Multivariate analysis showed that FAI<sub>PTAT</sub> ( $p = 0.007$ ;  $p = 0.027$ ) and M stage ( $p = 0.001$ ;  $p = 0.006$ ) were significantly associated with overall survival (**Table 2**) both in the TCGA and FUSCC cohorts. When Kaplan–Meier curves were made for the high FAI<sub>PTAT</sub> and low FAI<sub>PTAT</sub> groups, the 5-year median overall survival was not reached for both groups in the FUSCC cohort ( $HR = 2.28$ ,  $p = 0.01$ ), while the median overall survival was 64 months for high FAI<sub>PTAT</sub> group and not reached for the low FAI<sub>PTAT</sub> group in the TCGA cohort ( $HR = 2.24$ ,  $p < 0.001$ ) (**Figure 2**). We then calculated the 3-year survival rate. It was 88%/93% for high FAI<sub>PTAT</sub> and low FAI<sub>PTAT</sub> groups in the FUSCC cohort ( $HR = 1.81$ ) and 72%/93% for high FAI<sub>PTAT</sub> and low FAI<sub>PTAT</sub> groups in the TCGA cohort ( $HR = 5.65$ ). Those results showed that FAI<sub>PTAT</sub> may be an independent factor for overall survival.

**TABLE 1 |** Clinical Pathological Characteristics of FAI.

Characteristics	TCGA (n = 218)			FUSCC (n = 129)		
	No. (N)	FAI <sub>PTAT</sub>	p value	No. (N)	FAI <sub>PTAT</sub>	p-value
Male	145	-76.9	0.78	87	-87.7	0.033
Female	73	-77.6		42	-94.3	
Age						
<64	142	-77.6	0.62	104	-90.4	0.44
>64	76	-76.3		25	-87.5	
T stage						
1–2	142	-82.5	<0.001	113	-91.1	0.028
3–4	76	-67.1		16	-81.4	
N stage						
0	213	-77.3	0.27	127	-90.1	0.27
1	5	-68.6		2	-77.1	
M stage						
0	188	-78.9	<0.001	127	-90.3	0.01
1	30	-66.3		2	-60.6	
Necrosis						
No		NA		94	-92.1	0.01
Yes				35	-83.7	
Grade						
I–II	86	-83.6	<0.001	68	-91.1	0.028
III–IV	132	-72.9		61	-81.4	
BMI						
normal	36	-74.4	0.056	60	-88.3	0.29
overweight	61	-74.1		64	-90.5	
obese	84	-80.3		5	-99.8	
Hypertension						
No	80	-77.9	0.56	88	-91.5	0.09
Yes	101	-76.4		41	-86.3	
Diabetes						
No		NA		112	-90.2	0.61
Yes				17	-87.9	
Smoking						
No	92	-78.1	0.41	77	-90.8	0.44
Yes	89	-75.9		52	-88.5	
SSIGN						
0–2		NA		81	-94.6	0.001
3–4				20	-83.9	
5–6				16	-82.9	
7–9				10	-78.1	
≥10				2	-70.1	

NA, not available.

### Higher FAI<sub>PTAT</sub> Indicates Pathway Alterations in the Tumor Tissue and the Peri-Tumor Adipose Tissue

#### A) Neuroactive Ligand–Receptor Interaction Was Altered in the Tumor Tissue Both in the Imaging Genomics Cohort and FAI Cohort B (TCGA)

We then further explored the potential mechanism by assessing transcriptomic differences in the tumor tissue using the imaging genomics cohort and FAI cohort B (TCGA). In the imaging genomics cohort, KEGG pathway analysis showed that tumors of the high FAI<sub>PTAT</sub> group had significant alterations in pathways such as neuroactive ligand–receptor interaction, cytokine–cytokine receptor interaction, complement and coagulation cascades, and pathways of cancer (**Figure 3A**). The PPI networks were drawn using Cytoscape and key genes from those pathways such as

**TABLE 2 |** Univariate and Multivariate analysis for overall survival in patients with ccRCC.

Characteristics	Univariate analysis			Multivariate analysis		
	HR	95% CI	p-value	HR	95% CI	p-value
<b>TCGA cohort</b>						
Gender	1.46	0.88–2.42	0.14	0.93	0.52–1.69	0.82
Age	1.04	1.02–1.06	0.000	1.04	1.018–1.068	0.001
BMI	0.49	0.35–0.70	0.000	0.55	0.38–0.79	0.001
T stage	1.97	1.5–2.57	0.000	1.31	0.92–1.88	0.13
N stage	4.00	1.25–12.84	0.02	0.40	0.09–1.84	0.24
M stage	4.53	2.71–7.56	0.000	2.44	1.35–4.43	0.003
Pathological Grade	1.63	0.93–2.84	0.09	0.61	0.32–1.19	0.15
$\Delta$ FAI <sub>PTAT</sub>	2.33	1.36–4.01	0.002	1.28	0.61–2.72	0.51
FAI <sub>PTAT</sub>	2.81	1.59–4.96	0.000	2.14	1.15–3.99	0.017
<b>FUSCC cohort</b>						
Gender	1.57	0.66–3.73	0.31	2.49	0.97–6.16	0.056
Age, years	1.003	0.96–1.04	0.86	0.99	0.95–1.05	0.96
BMI	1.03	0.44–1.97	0.85	1.06	0.44–2.52	0.89
T stage	1.93	0.64–5.75	0.23	1.03	0.27–3.98	0.96
N stage	3.62	0.48–27.2	0.21	1.73	0.22–13.60	0.60
M stage	15.21	3.37–68.6	0.000	9.26	1.88–45.5	0.006
Pathological Grade	4.00	1.35–11.92	0.013	3.53	1.16–10.68	0.026
$\Delta$ FAI <sub>PTAT</sub>	2.42	0.94–6.23	0.068	1.18	0.34–4.11	0.79
FAI <sub>PTAT</sub>	3.31	1.21–9.05	0.019	3.14	1.11–8.88	0.031

HR, Hazard Ratio; CI, Confidence Interval; FAI, Fat Attenuation Index; BMI, Body Mass Index.

glucagon receptor (GCGR), Lysophosphatidic Acid Receptor 3 (LPAR3), Neuromedin U Receptor 2 (NMUR2), Epidermal Growth Factor (EGF), Proto-oncogene c-KIT (KIT), Kirsten rat sarcoma virus (KRAS), C-X-C Motif Chemokine Ligand 8 (CXCL8), IL-6, GCGR, Fibrinogen Alpha Chain (FGA), and Fibrinogen Beta Chain (FGB) were observed (**Figures 3B–E**). The pathway results were confirmed by the FAI cohort B from the TCGA database, which also presented neuroactive ligand–receptor interaction as the most significantly altered pathway (**Supplementary Figure 2**).

#### B) Complement and Coagulation Cascades, Cytokine–Cytokine Receptor Interaction and Multiple Metabolic Pathways Were Altered in the Peri-Tumor Adipose Tissue in the Imaging Genomics Cohort

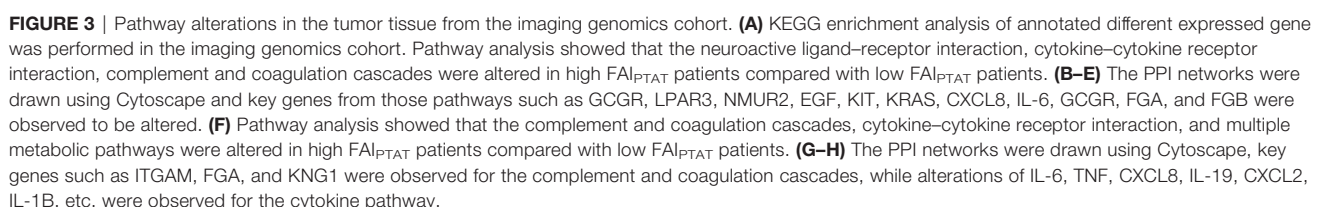
Transcriptomic differences in the peri-tumor adipose tissues were available using the imaging genomics cohort. Peri-tumor adipose tissue with higher FAI values showed significant alteration in pathways such as complement and coagulation cascades, cytokine–cytokine receptor interaction and multiple metabolic pathways (cholesterol metabolism, retinol metabolism, glycine, serine and threonine metabolism, arginine and proline metabolism, linoleic acid metabolism, etc.) (**Figure 3F**). Strong alterations of Integrin Subunit Alpha M (ITGAM), FGA, and kininogen 1 (KNG1) were observed for the complement and coagulation cascades (**Figure 3G**), while alterations of IL-6, Tumor Necrosis Factor (TNF), C-X-C Motif Chemokine Ligand 8 (CXCL8), interleukin 19 (IL-19), C-X-C Motif Chemokine Ligand 2 (CXCL2), interleukin 1 beta (IL-1B), etc. were observed for the cytokine pathway (**Figure 3H**).

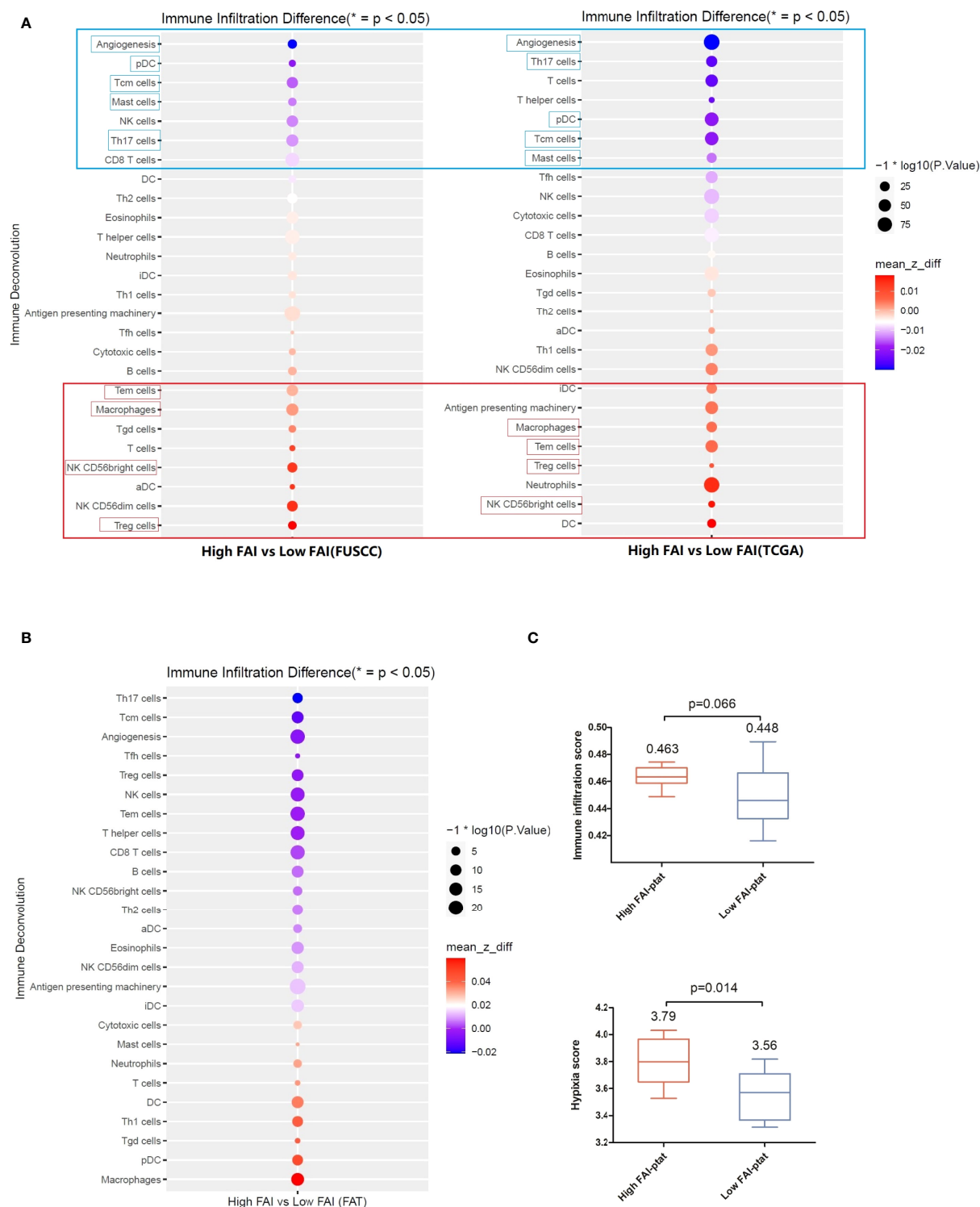
Thus, FAI<sub>PTAT</sub> could not only represent the characteristics of the tumor but also reveal the alterations of the peri-tumor adipose tissue. In the imaging genomics cohort, it was surprising that the cytokine–cytokine receptor interaction and the complement and coagulation cascades were both altered in the tumor and adipose tissues.

#### Higher FAI<sub>PTAT</sub> Indicates Diverse Immune Micro-Environment in the Tumor Tissue and the Peri-Tumor Adipose Tissue

##### A) Antigen Presenting and Angiogenesis may be Enhanced in Tumors of High FAI Patients Both in the Imaging Genomics Cohort and the TCGA Validation Cohort

We then analyzed the immune micro-environment of tumor tissue from the imaging genomics cohort (FUSCC) and the FAI cohort B (TCGA) by using immune deconvolution. We found good consistency between both the cohorts, tumors of high FAI<sub>PTAT</sub> patients showed enhanced infiltration of regulatory T cells (Treg), natural killer (NK), CD56 bright cells, effector memory T cells (Tem), and macrophages. Additionally, dendritic cells (DCs) were seen enhanced in the TCGA group, and B cells were seen enhanced in the FUSCC group. Reduced infiltration of central memory T cells (Tcms), T helper 17 cell (Th17) cells, and mast cells (**Figure 4A**) and a lower angiogenesis level were observed. We believe that antigen presentation was enhanced as macrophage, DC cells, and B cells were the main functional cells of antigen presentation. These results were then confirmed by multiplex immunofluorescence (mIF) in ten patients, that high FAI<sub>PTAT</sub> tumors may have more infiltrated macrophages and T cells (**Supplementary Figure 3**).





**FIGURE 4** | The immune microenvironment of tumor tissue of the imaging genomics cohort (FUSCC) and the FAI cohort B (TCGA). Immune deconvolution showed good consistency between the both cohort. **(A)** Tumors of high FAI<sub>PTAT</sub> patients showed enhanced infiltration of Treg, NK, CD56bright cells, Tem cells, and macrophages, and a reduced infiltration of pDC, Tcm cells, Th17 cells, and mast cells, as well as a lower angiogenesis level compared with low FAI<sub>PTAT</sub> patients. **(B)** Immune deconvolution showed an enhanced infiltration of macrophages, pDC cells, Th1 cells, and DC cells, and a reduced infiltration of Th17 cells, Tcm cells, Treg cells, CD8+ T cells, and NK cells, as well as a lower angiogenesis level in peri-tumor fat. **(C)** Hypoxia score and immune infiltration score were enhanced in fat of high FAI<sub>PTAT</sub> patients.



## B) High Hypoxia Score, Immune Infiltration Score and Macrophage Infiltration Scores Were Observed in Fat Tissues of High FAI Patients in the Imaging Genomics Cohort

Fat tissues from the imaging genomics cohort were available. Immune deconvolution showed enhanced infiltration of macrophages, pDC, Th1, DC, and reduced infiltration of Th17, Tcm, Treg, CD8+ T, and NK (**Figure 4B**), as well as a lower angiogenesis level. Hypoxia score and immune infiltration score were enhanced in the fat of high FAI<sub>PTAT</sub> patients compared with low FAI<sub>PTAT</sub> patients (**Figure 4C**).

## FAI<sub>PTAT</sub> Decreases After Effective Targeted Therapy

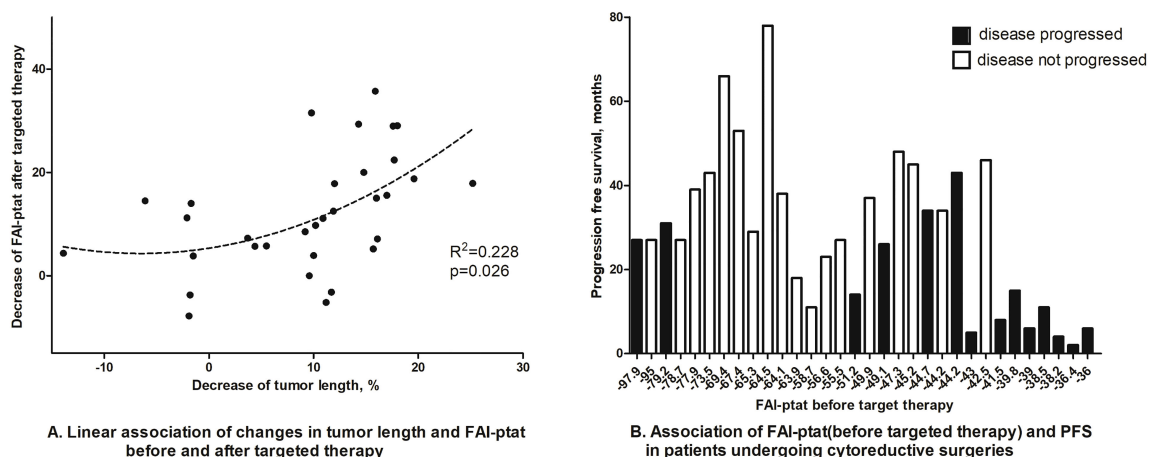
Finally, using the treatment response cohort, we tried to observe the dynamic changes of FAI<sub>PTAT</sub> after neo-adjuvant/palliative targeted therapy in patients who subsequently underwent nephrectomies and whether FAI<sub>PTAT</sub> could be altered by RCC treatment. Scattergram showed that there was an association between a decrease in FAI<sub>PTAT</sub> after targeted therapy and a decrease in tumor size (**Figure 5A**). Waterfall plot showed that patients with high FAI<sub>PTAT</sub> before targeted therapy will progress sooner; however, due to the limited number of participants, no significant statistics could be made (**Figure 5B**).

## DISCUSSION

In this study, we introduced a fat attenuation index imaging metric, which could quantify the degree of peri-tumor inflammation in renal cell carcinoma. We used four distinct cohorts to verify the outcome prediction ability of FAI in RCC patients and tried to discover the transcriptomic features of tumor tissues and adjacent adipocytes in high/low FAI groups. Our observations suggest that a high FAI<sub>PTAT</sub> (which indicates high inflammation status of peri-tumor fat) was significantly

associated with a worse outcome in both the FUSCC and TCGA cohorts, and it also indicated short PFS in patients who undertook cyto-reductive surgeries after targeted therapies. Molecular analysis of RNA expression of paired RCC tissue and peri-tumor fat tissue showed synchronized alterations in pathways such as cytokine–cytokine receptor interaction and complement and coagulation cascades. RCC tissues from the TCGA and FUSCC cohorts both showed significant alterations in the neuroactive ligand–receptor interaction pathway. Additionally, immunodeconvolution analysis showed enhanced infiltration of macrophages in high FAI<sub>PTAT</sub> tumor tissues with lower angiogenesis levels, consistent with previous studies indicating that Angio<sup>low</sup>Macrophage<sup>high</sup> patients had worse outcomes (6). We also observed synchronous dynamic changes in FAI<sub>PTAT</sub> and tumor size after targeted therapy, indicating that the tumor micro-environment would change after effective anti-cancer treatments. Whether the dynamic changes of FAI<sub>PTAT</sub> are the cause or results of primary tumor response should be further explored, but the strength of the link between FAI<sub>PTAT</sub> and RCC progression has improved.

CT imaging measurement of peri-tumor fat has been widely used in the assessment of adherent peri-nephric fat (APF), or so-called “sticky fat”, which would make partial nephrectomy difficult. The mayo adhesive probability (MAP) included posterior peri-nephric fat thickness and stranding and showed accuracy in predicting APF (12). Soon after, Thiel et al. reported that MAP may represent visceral obesity/inflammation and was associated with RCC prognosis (13). However, the introduction of stranding was quite subjective, and on the other hand, few had studied the relationship between stranding and inflammation as well as its molecular biological relationship with kidney tumors. FAI was initially introduced by Antonopoulos and colleagues to quantify vascular inflammation in peri-vascular adipose tissue by using CT imaging (8). They believe that high FAI indicates



**FIGURE 5 |** Dynamic changes of FAI<sub>PTAT</sub> after targeted therapy. **(A)** Scattergram showed that FAI<sub>PTAT</sub> changes synchronously with tumor size after targeted therapy. **(B)** Waterfall plot showed that patients with high FAI<sub>PTAT</sub> had short PFS during targeted therapy.

immature adipocytes with smaller sizes and less lipid accumulation, which is caused by inflammatory signals released by the coronary artery (14). Oikonomou et al. further indicated that high peri-vascular FAI values (cutoff  $\geq -70.1$  HU) were an indicator of increased cardiac mortality (15). Since FAI has been proven to be an objective, reproducible, and quantifiable factor, we believe it could demonstrate the peri-tumor fat stranding and inflammation degree. We tried to investigate how FAI could reflect the phenotypic character of peri-tumor adipose tissue and whether it could somehow reflect the tumor characteristics.

Kidney tumors with high FAI<sub>PTAT</sub> showed alterations in multiple pathways. The neuroactive ligand–receptor interaction pathway was the most significantly altered pathway of tumors in both the FUSCC and TCGA databases. After constructing the PPI network using Cytoscape, we observed that GCGR, LPAR3, and NMUR2 were the most significant modules in the pathway. GCGR is a receptor for glucagon and plays a central role in the regulation of blood glucose levels, glucose homeostasis, amino acid metabolism, and lipid metabolism (16). Further investigation is needed to determine whether glucagon metabolism plays a certain role in RCC development. Other key genes such as EGF, KIT, KRAS, CXCL8, and IL-6 were also shown to be altered, all of which indicated worse outcomes and could affect the effectiveness of TKIs (17–19). Research has indicated that up-regulation of the pro-inflammatory cytokines IL-6, TNF- $\alpha$ , and IFN- $\gamma$  secreted by cancer tissues could prevent the differentiation of pre-adipocytes and cause-altered phenotypes. Those so-called cancer-associated adipocytes could then contribute to promoting tumor aggressiveness by over-expression of pro-inflammatory cytokines [interleukin (IL)-6, IL-1 $\beta$ ] and vice versa (20). According to our results, we observed an activated cytokine pathway both in the tumor tissue and the paired fat tissue, promoting strong interaction between the tumor and peri-tumor fat tissue. Thus, peri-tumor fat morphology represented by FAI<sub>PTAT</sub> could present the tumor characteristics such that inflammation enhancement both in the tumor itself and the peri-tumor environment could be observed in highly aggressive kidney tumors.

We then tried to evaluate the immune infiltration of kidney tumors and peri-tumor fat in high FAI<sub>PTAT</sub> patients. We observed high consistency of immune infiltration of tumor tissue in the FUSCC and TCGA cohorts. For high FAI<sub>PTAT</sub> patients in both cohorts, higher scores of Treg, NK, CD56bright cells, Tem cells, macrophages, and antigen-presenting machinery were observed in tumor tissue; and a lower number of pDC, Tcm cells, Th17 cells, mast cells, and a lower angiogenesis status were observed. Considering that high FAI<sub>PTAT</sub> patients had short PFS after targeted therapy in the treatment response cohort, these results were supported by the research done by Hakimi et al., which showed that Angio<sup>low</sup>Macrophage<sup>high</sup> represented worse survival and poor TKI sufficiency (6). Immune deconvolution of peri-tumor fat tissue also showed higher scores of macrophages, pDC cells, Th1 cells, and DC cells in high FAI<sub>PTAT</sub> patients, as well as lower scores of Th17 cells, Tcm cells, Treg cells, CD8+ T cells, NK cells, and angiogenesis. Macrophages are often the most

abundant cell type in the tumor micro-environment, among which tumor associated macrophages (TAM) support angiogenesis, tissue remodeling, and immune suppression. It was reported that RCC tumors with high-infiltration TAMs were significantly associated with poor prognosis (18). One of the possible theories was that TAMs may accumulate in regions of hypoxia, and the initial hypoxic response may cause M2 polarization, which has pro-tumorigenic functions (21). Another significantly altered immune cell were Treg cells, which also suppress anti-tumor immune responses. There is solid evidence that infiltration of Treg cells is often associated with poor prognosis and poor immunotherapy effectiveness (22). However, due to its high expression of CTLA-4, it is presumed an anti-CTLA-4 antibody may reduce Treg cells and make immunotherapy more effective. Thus, understanding the relationship between FAI<sub>PTAT</sub> and the tumor micro-environment may help decision-making. However, further evaluations are needed.

This study had certain limitations. 1) The cohort size was quite small for both clinical analysis and RNA sequencing. However, the biological characteristics were validated and had good consistency between different cohorts. 2) We only included clear-cell RCC; further exploration should be made for non-clear RCC. 3) The immune infiltration status of immune cells was mainly calculated by immune deconvolution and immune infiltration scores, IHC was done for 10 patients. However, they were not validated by flow cytometry. 4) We also had concerns about whether the outcome distinguished by FAI<sub>PTAT</sub> was due to peri-nephric fat invasion. Landman and colleagues indicated that peri-nephric soft-tissue stranding was a significant factor for predicting peri-nephric fat invasion, especially in tumors 4 cm or less (23). However, the data was so controversial that Bradley et al. reported that the presence of peri-nephric stranding and tumor necrosis were not reliable signs for pT stage >T3a (24). On the other hand, the survival differences between T3a and T2 tumors were not so significant (25). We would rather believe that the survival differences between high FAI<sub>PTAT</sub> and low FAI<sub>PTAT</sub> patients were due to tumor transcriptomic characteristics or the micro-environments represented by peri-tumor fat.

In summary, FAI could be used in RCC to reflect the inflammation status of both the tumor and the peri-tumor adipose. Additionally, FAI has the potential to predict tumor biological characteristics and survival outcomes in various cohorts. This study demonstrates that the crosstalk exists between a tumor and its micro-environment and could be reflected easily by imaging procedures and then facilitates clinical decision making.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Ethics Committee of Fudan University Shanghai Cancer Center. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

YZ and DWY contributed to conception and design of the study. HW and YZ organized the database. YW, XH, HZ, GS, HL, and JOZ performed the statistical analysis. JP, BHW and JLW participated in laboratory experiments. HW wrote the first

draft of the manuscript. YW and YZ wrote sections of the manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

The manuscript is supported by the National Natural Science Foundation of China Projects 81972375 and 81802528.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.786981/full#supplementary-material>

## REFERENCES

1. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-Mass Index and Incidence of Cancer: A Systematic Review and Meta-Analysis of Prospective Observational Studies. *Lancet* (2008) 371(9612):569–78. doi: 10.1016/S0140-6736(08)60269-X
2. Choi Y, Park B, Jeong BC, Seo SI, Jeon SS, Choi HY, et al. Body Mass Index and Survival in Patients With Renal Cell Carcinoma: A Clinical-Based Cohort and Meta-Analysis. *Int J Cancer* (2013) 132(3):625–34. doi: 10.1002/ijc.27639
3. Albiges L, Hakimi AA, Xie W, McKay RR, Simantov R, Lin X, et al. Body Mass Index and Metastatic Renal Cell Carcinoma: Clinical and Biological Correlations. *J Clin Oncol* (2016) 34(30):3655–63. doi: 10.1200/JCO.2016.66.7311
4. Khandekar MJ, Cohen P, Spiegelman BM. Molecular Mechanisms of Cancer Development in Obesity. *Nat Rev Cancer* (2011) 11(12):886–95. doi: 10.1038/nrc3174
5. Sanchez A, Furberg H, Kuo F, Vuong L, Ged Y, Patil S, et al. Transcriptomic Signatures Related to the Obesity Paradox in Patients With Clear Cell Renal Cell Carcinoma: A Cohort Study. *Lancet Oncol* (2019) 21(2):283–93. doi: 10.1016/S1470-2045(19)30797-1
6. Hakimi AA, Voss MH, Kuo F, Sanchez A, Liu M, Nixon BG, et al. Transcriptomic Profiling of the Tumor Microenvironment Reveals Distinct Subgroups of Clear Cell Renal Cell Cancer: Data From a Randomized Phase Iii Trial. *Cancer Discovery* (2019) 9(4):510–25. doi: 10.1158/2159-8290.CD-18-0957
7. Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* (2019) 179(4):964–83.e31. doi: 10.1016/j.cell.2019.10.007
8. Antonopoulos AS, Sanna F, Sabharwal N, Thomas S, Oikonomou EK, Herdman L, et al. Detecting Human Coronary Inflammation by Imaging Perivascular Fat. *Sci Transl Med* (2017) 9(398):eaal2658. doi: 10.1126/scitranslmed.aal2658
9. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. String V10: Protein-Protein Interaction Networks, Integrated Over the Tree of Life. *Nucleic Acids Res* (2015) 43:D447–52. doi: 10.1093/nar/gku1003
10. Cancer Genome Atlas Research N. Comprehensive Molecular Characterization of Clear Cell Renal Cell Carcinoma. *Nature* (2013) 499(7456):43–9. doi: 10.1038/nature12222
11. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic Rna Interference Reveals That Oncogenic Kras-Driven Cancers Require Tbk1. *Nature* (2009) 462(7269):108–12. doi: 10.1038/nature08460
12. Davidiuk AJ, Parker AS, Thomas CS, Leibovich BC, Castle EP, Heckman MG, et al. Mayo Adhesive Probability Score: An Accurate Image-Based Scoring System to Predict Adherent Perinephric Fat in Partial Nephrectomy. *Eur Urol* (2014) 66(6):1165–71. doi: 10.1016/j.eururo.2014.08.054
13. Thiel DD, Davidiuk AJ, Meschia C, Serie D, Custer K, Petrou SP, et al. Mayo Adhesive Probability Score Is Associated With Localized Renal Cell Carcinoma Progression-Free Survival. *Urology* (2016) 89:54–60. doi: 10.1016/j.urol.2015.10.034
14. Grant RW, Stephens JM. Fat in Flames: Influence of Cytokines and Pattern Recognition Receptors on Adipocyte Lipolysis. *Am J Physiol Endocrinol Metab* (2015) 309(3):E205–13. doi: 10.1152/ajpendo.00053.2015
15. Oikonomou EK, Marwan M, Desai MY, Mancio J, Alashi A, Hutt Centeno E, et al. Non-Invasive Detection of Coronary Inflammation Using Computed Tomography and Prediction of Residual Cardiovascular Risk (the Crisp Ct Study): A Post-Hoc Analysis of Prospective Outcome Data. *Lancet* (2018) 392(10151):929–39. doi: 10.1016/S0140-6736(18)31114-0
16. Janah L, Kjeldsen S, Galsgaard KD, Winther-Sorensen M, Stojanovska E, Pedersen J, et al. Glucagon Receptor Signaling and Glucagon Resistance. *Int J Mol Sci* (2019) 20(13):3314. doi: 10.3390/ijms20133314
17. Lai Y, Zhao Z, Zeng T, Liang X, Chen D, Duan X, et al. Crosstalk Between Vegfr and Other Receptor Tyrosine Kinases for Tki Therapy of Metastatic Renal Cell Carcinoma. *Cancer Cell Int* (2018) 18:31. doi: 10.1186/s12935-018-0530-2
18. de Vivar Chevez AR, Finke J, Bukowski R. The Role of Inflammation in Kidney Cancer. *Adv Exp Med Biol* (2014) 816:197–234. doi: 10.1007/978-3-0348-0837-8\_9
19. Fitzgerald JP, Nayak B, Shanmugasundaram K, Friedrichs W, Sudarshan S, Eid AA, et al. Nox4 Mediates Renal Cell Carcinoma Cell Invasion Through Hypoxia-Induced Interleukin 6- and 8- Production. *PLoS One* (2012) 7(1):e30712. doi: 10.1371/journal.pone.0030712
20. Dirat B, Bochet L, Dabek M, Daviaud D, Dauvillier S, Majed B, et al. Cancer-Associated Adipocytes Exhibit an Activated Phenotype and Contribute to Breast Cancer Invasion. *Cancer Res* (2011) 71(7):2455–65. doi: 10.1158/0008-5472.CAN-10-3323
21. Quail DF, Joyce JA. Microenvironmental Regulation of Tumor Progression and Metastasis. *Nat Med* (2013) 19(11):1423–37. doi: 10.1038/nm.3394
22. Tanaka A, Sakaguchi S. Regulatory T Cells in Cancer Immunotherapy. *Cell Res* (2017) 27(1):109–18. doi: 10.1038/cr.2016.151
23. Landman J, Park JY, Zhao C, Baker M, Hofmann M, Helmy M, et al. Preoperative Computed Tomography Assessment for Perinephric Fat Invasion: Comparison With Pathological Staging. *J Comput Assist Tomogr* (2017) 41(5):702–7. doi: 10.1097/RCT.0000000000000588
24. Bradley AJ, MacDonald L, Whiteside S, Johnson RJ, Ramani VA. Accuracy of Preoperative Ct T Staging of Renal Cell Carcinoma: Which Features Predict Advanced Stage? *Clin Radiol* (2015) 70(8):822–9. doi: 10.1016/j.crad.2015.03.013

25. Siemer S, Lehmann J, Loch A, Becker F, Stein U, Schneider G, et al. Current Tnm Classification of Renal Cell Carcinoma Evaluated: Revising Stage T3a. *J Urol* (2005) 173(1):33–7. doi: 10.1097/01.ju.0000146719.43269.e8

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Wei, Hu, Pan, Wu, Wang, Zhang, Shi, Liu, Zhao, Zhu and Ye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership