Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study

Cheng-Peng Gui, Yu-Hang Chen*, Hong-Wei Zhao*, Jia-Zheng Cao*, Tian-Jie Liu*, Sheng-Wei Xiong, Yan-Fei Yu, Bing Liao, Yun Cao, Jia-Ying Li, Kang-Bo Huang, Hui Han, Zhi-Ling Zhang, Wen-Fang Chen, Ze-Ying Jiang, Ye Gao, Guan-Peng Han, Qi Tang, Kui Ouyang, Gui-Mei Qu, Ji-Tao Wu, Jian-Ping Guo, Cai-Xia Li, Pei-Xing Li, Zhi-Ping Liu, Jer-Tsong Hsieh, Mu-Yan Cai, Xue-Song Li, Jin-Huan Wei, Jun-Hang Luo*

Summary

Background Improved markers for predicting recurrence are needed to stratify patients with localised (stage I–III) renal cell carcinoma after surgery for selection of adjuvant therapy. We developed a novel assay integrating three modalities—clinical, genomic, and histopathological—to improve the predictive accuracy for localised renal cell carcinoma recurrence.

Methods In this retrospective analysis and validation study, we developed a histopathological whole-slide image (WSI)-based score using deep learning allied to digital scanning of conventional haematoxylin and eosin-stained tumour tissue sections, to predict tumour recurrence in a development dataset of 651 patients with distinctly good or poor disease outcome. The six single nucleotide polymorphism-based score, which was detected in paraffin-embedded tumour tissue samples, and the Leibovich score, which was established using clinicopathological risk factors, were combined with the WSI-based score to construct a multimodal recurrence score in the training dataset of 1125 patients. The multimodal recurrence score was validated in 1625 patients from the independent validation dataset and 418 patients from The Cancer Genome Atlas set. The primary outcome measured was the recurrence-free interval (RFI).

Findings The multimodal recurrence score had significantly higher predictive accuracy than the three single-modal scores and clinicopathological risk factors, and it precisely predicted the RFI of patients in the training and two validation datasets (areas under the curve at 5 years: 0·825–0·876 *vs* 0·608–0·793; p<0·05). The RFI of patients with low stage or grade is usually better than that of patients with high stage or grade; however, the RFI in the multimodal recurrence score-defined high-risk stage I and II group was shorter than in the low-risk stage III group (hazard ratio [HR] 4·57, 95% CI 2·49–8·40; p<0·0001), and the RFI of the high-risk grade 1 and 2 group was shorter than in the low-risk grade 3 and 4 group (HR 4·58, 3·19–6·59; p<0·0001).

Interpretation Our multimodal recurrence score is a practical and reliable predictor that can add value to the current staging system for predicting localised renal cell carcinoma recurrence after surgery, and this combined approach more precisely informs treatment decisions about adjuvant therapy.

Funding National Natural Science Foundation of China, and National Key Research and Development Program of China.

Lancet Digit Health 2023; 5: e515–24

Published Online June 29, 2023 https://doi.org/10.1016/S2589-7500(23)00095-X

See **Comment** page e482

*Contributed equally

Department of Urology (C-P Gui MD, Y-H Chen MD, J-Y Li MD, Prof J-H Wei MD, Prof J-H Luo MD), **Department of Pathology** (Prof B Liao MD, Prof W-F Chen MD, Z-Y Jiang MD), **and Institute of Precision Medicine** (Prof J-P Guo PhD, Prof J-H Luo), **First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China; Department of Urology** (Prof H-W Zhao MD, K Ouyang BS, Prof J-T Wu MD) **and Department of Pathology** (Prof G-M Qu MD), **Affiliated Yantai Yuhuangding Hospital, Qingdao University, Yantai, China; Department of Urology, Jiangmen Central Hospital, Jiangmen, China** (Prof J-Z Cao MD); **Department of Urology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China** (T-J Liu PhD, Prof Y Gao MD); **Department of Urology, Peking University First Hospital, Institute of Urology, Peking University, National Urological Cancer Center, Beijing, China** (S-W Xiong MD, Y-F Yu MD, G-P Han MD, Q Tang MD, Prof X-S Li MD); **Department of Pathology** (Prof Y Cao MD, Prof M-Y Cai MD) **and Department of Urology** (K-B Huang MD, Prof H Han MD, Prof Z-L Zhang MD), **Cancer Center, Sun Yat-sen University, Guangzhou, China; School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou, China** (Prof C-X Li PhD, Prof P-X Li PhD); **Department of Internal Medicine and**

Introduction

Kidney cancer is among the top ten most common cancers in both men and women and is estimated to have accounted for 79 000 new cases in the USA in 2022.[1] More than 80% of people with clear cell renal cell carcinoma, which is the major histological subtype, show localised stage I–III disease at first diagnosis, and approximately 30% of these people will relapse after surgical excision.[2,3] Recent studies have shown that a subset of people with high risk of recurrence could benefit from adjuvant immunotherapy or targeted therapy.[4,5]

TNM stage and pathological grade are commonly used to assess the risk of tumour recurrence in patients with localised clear cell renal cell carcinoma after surgery. However, patients with the same TNM stage and pathological grade can have diverse outcomes, and the current staging system is not sufficiently accurate to predict tumour recurrence and guide selection of adjuvant therapy. Refined prognostic models are needed to enable a more targeted approach to the selection of adjuvant therapies. Prognostic models should aim to identify patient subgroups for which the absolute benefits of adjuvant therapy are minimal relative to surgery alone and, at the other end of the spectrum, nominate patients who might benefit from adjuvant therapy because of their high recurrence risk.[6,7] Several clinical risk models, such

Department of Molecular Biology (Prof Z-P Liu PhD)**, and Department of Urology** (Prof J-T Hsieh PhD)**, University of Texas Southwestern Medical Center at Dallas, Dallas TX, USA**

Correspondence to:
Prof Jun-Hang Luo, Department of Urology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China
luojunh@mail.sysu.edu.cn

or

Prof Jin-Huan Wei, Department of Urology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China
weijh23@mail.sysu.edu.cn

or

Prof Xue-Song Li, Department of Urology, Peking University First Hospital, Institute of Urology, Peking University, National Urological Cancer Center, Beijing 100034, China
pineneedle@sina.com

Research in context

Evidence before this study

We searched PubMed with the search terms "renal cell carcinoma", "RCC", "prognostic model", and "multimodal" for articles published in English between Jan 1, 2010, and Aug 28, 2022. We identified several studies in which multimodal prognostic models of renal cell carcinoma were built using histopathological whole-slide images (WSIs) and molecular profile data with The Cancer Genome Atlas (TCGA) dataset. However, the molecular signatures of previous models were developed based on analysis of fresh-frozen specimens in the TCGA set, and none of these models has been validated in multiple independent cohorts.

Added value of this study

In this study, we developed a recurrence scoring system that integrated three modalities—histopathological (the WSI-based score), genomic (the six single nucleotide polymorphism-based score), and clinical (the Leibovich score)—to predict the outcomes of renal cell carcinoma. This multimodal system was validated in independent sets from multicentres of China and the TCGA dataset. Our multimodal system can predict the recurrence risk of localised renal cell carcinoma more accurately than the existing TNM staging system and three established single-modality systems. The recurrence risk of patients with low stage or grade is usually lower than that of patients with high stage or grade; however, our results show that the recurrence risk of the patient subgroup with multimodal system-defined high risk in stage I or II disease was higher than in that with multimodal system-defined low risk in stage III, and that the recurrence risk of the patient subgroup with high risk in grade 1 or 2 was higher than in that with low risk in grade 3 or 4.

Implications of all the available evidence

To our knowledge, this study is the largest in the scientific literature to use a systematic approach to identify prognostic biomarkers for renal cell carcinoma. This assay is directly applicable to routinely available paraffin-embedded tumour tissue and haematoxylin and eosin-stained sections, which makes it easy to translate to clinical application. Our multimodal recurrence scoring system can be a practical and reliable prognostic tool for localised renal cell carcinoma, which can complement the existing staging system to predict recurrence after surgery and support more informed treatment decisions and trials of adjuvant therapy.

as the Leibovich score, and molecular risk models, such as the six single nucleotide polymorphism (SNP)-based score and the 16-gene assay, have been established to add predictive value to the existing staging system of localised clear cell renal cell carcinoma.[8–10]

Recently, in many applications, deep learning has excelled at image interpretation tasks, and it is hypothesised that some of these applications could be useful for retrieving additional information from histopathology images.[11,12] A prognostic model based on histopathological images has been developed that uses deep learning allied to digital scanning of conventional haematoxylin and eosin (H&E)-stained, formalin-fixed, paraffin-embedded (FFPE) tumour tissue sections. The histopathological image-based model stratified patients with stage II and stage III colorectal cancer into sufficiently distinct prognostic groups that the groupings could potentially be used to guide selection of adjuvant treatment.[11]

In this study, we developed a novel recurrence scoring system using a large sample size of localised clear cell renal cell carcinoma that integrated three modalities: clinical, genomic, and histopathological. We validated the predictive accuracy and reproducibility of this multimodal scoring system in independent cohorts, including more than 1600 cases from multicentres of China and The Cancer Genome Atlas (TCGA) dataset.

Methods

Participants

In this retrospective analysis and multicentre validation study, we used FFPE tissue samples from 2332 patients (aged ≥18 years) who underwent resection of a localised, sporadic clear cell renal cell carcinoma between Jan 1, 2006 and June 30, 2016. Patients with stage I–III disease and with available clinicopathological characteristics and follow-up information were included. Exclusion criteria were synchronous or metachronous bilateral renal cell carcinoma or a history of inherited Von Hippel-Lindau disease, neoadjuvant therapy, or adjuvant therapy. The institutional review board at each participating institution approved the retrospective analysis of anonymous patient data.

The training set included 1125 patients from the First Affiliated Hospital of Sun Yat-sen University (437 patients) and Yuhuangding Hospital of Qingdao University (688 patients); the validation set included 1207 patients from the Cancer Center of Sun Yat-sen University (454 patients), Peking University First Hospital (558 patients), and the First Affiliated Hospital of Xi'an Jiaotong University (195 patients). The TNM 2016 staging system was used to classify patients with clear cell renal cell carcinoma.[13] The histopathological grading system used in this study was based on the Fuhrman four-grade scale. Two genito-urinary pathologists (YC and BL) reassessed all samples.

For the TCGA set, clinical data comprised 418 retrospectively identified patients who underwent resection between 1998 and 2013 from 13 medical centres in the USA for localised clear cell renal cell carcinoma. Clinical data, SNP data, and clear diagnostic whole-slide images (WSIs) were downloaded from the Genomic Data Commons Data Portal and are in whole

or in part based on data generated by the TCGA Research Network.[14] Cases with conflicting information were thoroughly re-evaluated and discussed again using all available information to reach a final diagnosis. Clinical features of patients in the training set, the independent validation set, and the TCGA set are described in the table.

WSI-based score

In total, 651 patients from the training set with a distinct outcome, either good or poor, were used as a development set. The digital WSI of each patient was scanned from the representative H&E-stained FFPE tumour tissue sections. We applied deep learning to develop a histopathological modality model, the WSI-based score, to predict the recurrence risk of patients with clear cell renal cell carcinoma. The predictive accuracy of the WSI-based score with the 10× resolution was higher than that with the 40× resolution. The study design is shown in figure 1, and details are available in the appendix (pp 3–8).

Six SNP-based score

As reported in our previous study, we developed a genomic modality model, the six SNP-based score, to assess the recurrence risk of patients with clear cell renal cell carcinoma and to predict prognosis after surgery.[9] The predictive accuracy of the six SNP-based score was significantly higher than that of MET-SNP (rs11762213), which was identified as a prognostic factor of clear cell renal cell carcinoma in previous studies (appendix p 10).[15,16] In the present study, we examined the six SNPs using time-of-flight mass spectrometry (MassARRAY system, Sequenom, San Diego, CA, USA) to analyse extracted DNA from the FFPE samples, as described previously.[9] We then calculated a risk score for each patient based on the six SNP status using the formula: six SNP-based score=(0·1186×rs4479520)–(0·0074×rs4718593) + (0·0072×rs9618567) + (0·0633×rs7934644) –(0·2123×rs7739947) –(0·1466×rs17050001)–0·1650.

Leibovich score

We used a clinical modality model—the Leibovich score, including pathological T stage, regional lymph node status, tumour size, histopathological grade, and tumour necrosis—to assess the recurrence risk of patients with clear cell renal cell carcinoma and to predict prognosis after surgery (appendix p 28).[8]

Multimodal recurrence scoring system

Using Cox regression coefficients, we developed a multimodal recurrence scoring system based on the WSI-based score (with the 10× resolution), the six SNP-based score, and the Leibovich score in the training set. We validated the multimodal recurrence scoring system in the independent validation set and in the TCGA set. The study design is shown in figure 1.

	Training set (n=1125)	Independent validation set (n=1207)	TCGA set (n=418)	Total (n=2750)
Age				
<60 years	704 (62·6%)	713 (59·1%)	196 (46·9%)	1613 (58·7%)
≥60 years	421 (37·4%)	494 (40·9%)	222 (53·1%)	1137 (41·3%)
Sex				
Female	354 (31·5%)	418 (34·6%)	156 (37·3%)	928 (33·7%)
Male	771 (68·5%)	789 (65·4%)	262 (62·7%)	1822 (66·3%)
Stage				
I	658 (58·5%)	783 (64·9%)	247 (59·1%)	1688 (61·4%)
II	193 (17·2%)	176 (14·6%)	56 (13·4%)	425 (15·4%)
III	274 (24·3%)	248 (20·5%)	115 (27·5%)	637 (23·2%)
pT				
T1a	609 (54·1%)	682 (56·5%)	137 (32·8%)	1428 (51·9%)
T1b	52 (4·6%)	129 (10·7%)	112 (26·8%)	293 (10·7%)
T2	204 (18·2%)	203 (16·8%)	57 (13·6%)	464 (16·9%)
T3	260 (23·1%)	193 (16·0%)	112 (26·8%)	565 (20·5%)
pN				
N0/Nx	1108 (98·5%)	1152 (95·4%)	408 (97·6%)	2668 (97·0%)
N1	17 (1·5%)	55 (4·6%)	10 (2·4%)	82 (3·0%)
Tumour size				
<10 cm	1107 (98·4%)	1190 (98·6%)	366 (87·6%)	2663 (96·8%)
≥10 cm	18 (1·6%)	17 (1·4%)	52 (12·4%)	87 (3·2%)
Grade				
1	110 (9·8%)	119 (9·9%)	15 (3·6%)	244 (8·9%)
2	557 (49·5%)	582 (48·2%)	202 (48·3%)	1341 (48·7%)
3	365 (32·4%)	403 (33·4%)	166 (39·7%)	934 (34·0%)
4	93 (8·3%)	103 (8·5%)	35 (8·4%)	231 (8·4%)
Tumour necrosis				
Absent	871 (77·4%)	917 (76·0%)	244 (58·4%)	2032 (73·9%)
Present	254 (22·6%)	290 (24·0%)	174 (41·6%)	718 (26·1%)
Surgical approach				
Partial nephrectomy	577 (51·3%)	632 (52·4%)	136 (32·5%)	1345 (48·9%)
Radical nephrectomy	548 (48·7%)	575 (47·6%)	282 (67·5%)	1405 (51·1%)
Follow-up, months	87 (58–126)	91 (65–115)	43 (21–65)	81 (53–116)

Data are n (%) or median (IQR). TCGA=The Cancer Genome Atlas. pT=pathological T stage. pN=regional lymph node status.

Table: **Baseline characteristics of patients in the multimodal recurrence score assessment set**

Outcomes

The main outcome was recurrence-free interval (RFI), defined as the time from surgery to first renal cell carcinoma recurrence (local disease or distant metastases identified by imaging, biopsy, or physical examination). The secondary outcome was cancer specific survival (CSS), defined as the time from surgery to death from clear cell renal cell carcinoma.

Statistical analysis

This study conformed to the REMARK, TRIPOD reporting, and SAGER guidelines (appendix pp 29–30).[17–19] The Kaplan-Meier method was used to analyse the correlation between variables and RFI and CSS, and the log-rank test was used to assess whether the classifier predicted survival. The Cox regression model was used for multivariable

See Online for appendix

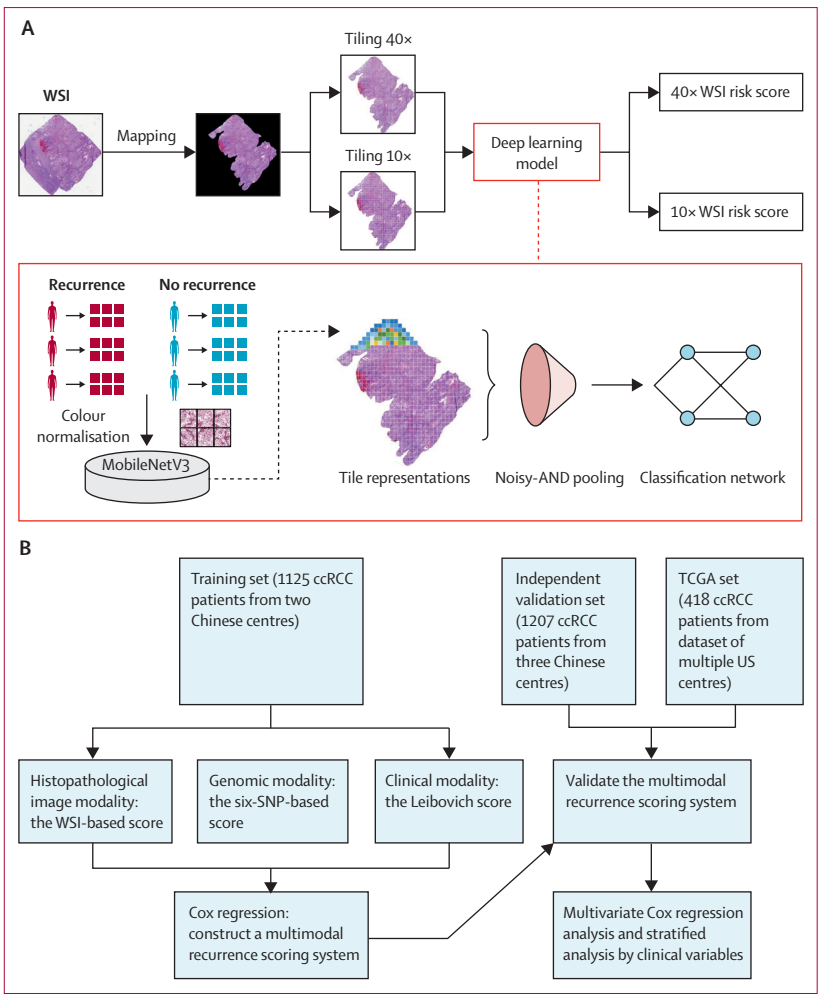


Figure 1: **Study design**
(A) Development of the WSI-based score using deep learning in the development set. (B) Construction and validation of the multimodal recurrence scoring system. ccRCC=clear cell renal cell carcinoma. SNP=single-nucleotide polymorphism. TCGA=The Cancer Genome Atlas. WSI=whole-slide image.

survival analysis, and Cox regression coefficients were used to generate a nomogram.[20] Calibration curves were used to assess whether actual outcomes approximately predicted outcomes for the nomogram.[20] Time-dependent receiver operating characteristic (ROC) curves and areas under the curve (AUCs) at 3, 5, and 7 years were generated to assess the predictive accuracy at different cutoff times.[21] Statistical tests were performed with R software (version 4.1.0). A two-sided p value of less than 0·05 was considered statistically significant.

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

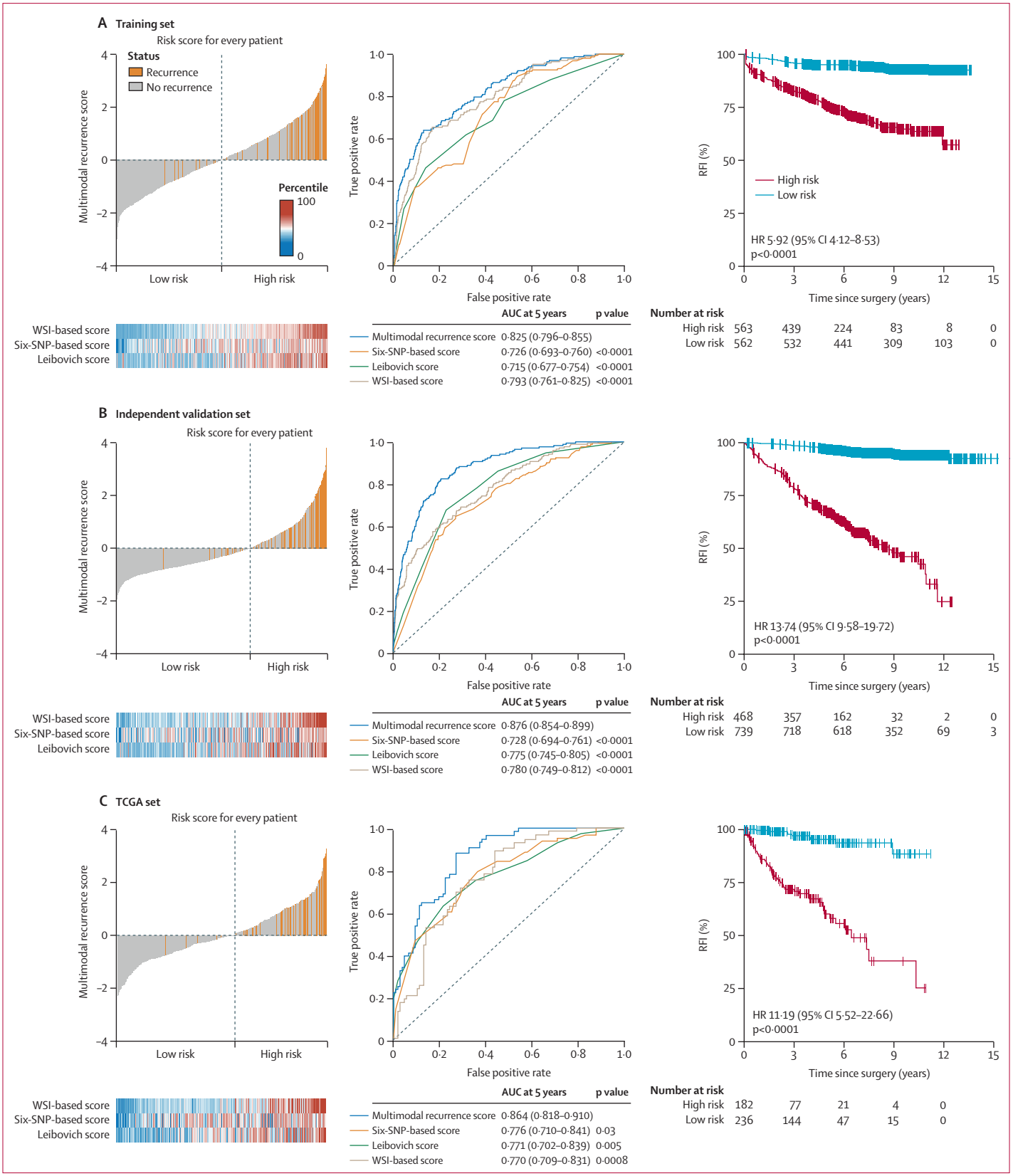
Results

Using the WSI-based score, we calculated a risk score for each of the 1125 patients with clear cell renal cell carcinoma in the training set. When we assessed the distribution of the WSI-based score and recurrence status, patients with lower risk scores generally had lower recurrence rates than patients with higher risk scores (appendix pp 11–12). We also calculated two risk scores of each patient in the training set based on the six SNP-based score and the Leibovich score. Patients with lower risk scores also had lower recurrence rates using these two modalities (appendix pp 11–12). We used a time-dependent ROC curve to describe the predictive accuracy of the three modalities, which determined the AUC at 5 years to be 0·793 for the WSI-based score, 0·726 for the six SNP-based score, and 0·715 for the Leibovich score (figure 2). Next, we developed a multimodal recurrence scoring system combining these three modalities by Cox regression coefficients: multimodal recurrence score= 3·3602×WSI score+1·7601×six SNP-based score+0·2158 ×Leibovich score−1·7850.

For the multimodal scoring system, the AUC at 5 years for predicting tumour recurrence was 0·825, which was significantly higher than that of any single-modality risk model alone, and also significantly higher compared with the clinicopathological risk factors (p<0·05; figure 2 and appendix p 17). Patients in the training set were divided into high-risk (n=563) and low-risk (n=562) groups, with the median risk score (zero) as the cutoff. Compared with patients in the low-risk group, patients in the high-risk group had shorter RFI (hazard ratio [HR] 5·92 [95% CI 4·12–8·53], p<0·0001; figure 2). Patients in the high-risk group also had shorter CSS than those in the low-risk group (6·39 [4·06–10·05], p<0·0001; appendix p 33).

To estimate the reproducibility and validity of the multimodal recurrence scoring system, we tested the classifier in 1207 cases of the independent validation set and in 418 cases of the TCGA set. The risk score for each of these patients was calculated using the same formula as used in the training set. The multimodal recurrence score showed stable predictive accuracy that was similar in the independent validation set (AUC at 5 years=0·876) and the TCGA set (AUC=0·864; figure 2), and the accuracy was significantly higher both than that of any individual single-modality risk model and than the clinicopathological risk factors (p<0·05; appendix pp 13–17). Patients in these two sets were classified into

Figure 2: **Construction of the multimodal recurrence score based on the WSI-based score, six-SNP-based score, and Leibovich score**
Left: risk score distributions of the multimodal recurrence score and patient recurrence status, and heat maps showing the distribution levels of the three modalities. Middle: time-dependent ROC curves of the multimodal recurrence score and three individual modalities. Right: Kaplan-Meier survival analyses for RFI in patients with ccRCC, divided into low-risk and high-risk groups according to risk score. HR and 95% CI were calculated using the Cox proportional hazards model. p values were calculated using the log-rank test. AUC=area under the curve. ccRCC=clear cell renal cell carcinoma. HR=hazard ratio. RFI=recurrence-free interval. ROC=receiver operating characteristics. SNP=single-nucleotide polymorphism. TCGA=The Cancer Genome Atlas. WSI=whole-slide image.


A Training set
Risk score for every patient
Status
Recurrence
No recurrence
Multimodal recurrence score
Percentile
100
0
Low risk
High risk
WSI-based score
Six-SNP-based score
Leibovich score
True positive rate
False positive rate
AUC at 5 years
p value
Multimodal recurrence score 0·825 (0·796–0·855)
Six-SNP-based score 0·726 (0·693–0·760) <0·0001
Leibovich score 0·715 (0·677–0·754) <0·0001
WSI-based score 0·793 (0·761–0·825) <0·0001
RFI (%)
High risk
Low risk
HR 5·92 (95% CI 4·12–8·53)
p<0·0001
Time since surgery (years)
Number at risk
High risk 563 439 224 83 8 0
Low risk 562 532 441 309 103 0
B Independent validation set
Risk score for every patient
Multimodal recurrence score
Low risk
High risk
WSI-based score
Six-SNP-based score
Leibovich score
True positive rate
False positive rate
AUC at 5 years
p value
Multimodal recurrence score 0·876 (0·854–0·899)
Six-SNP-based score 0·728 (0·694–0·761) <0·0001
Leibovich score 0·775 (0·745–0·805) <0·0001
WSI-based score 0·780 (0·749–0·812) <0·0001
RFI (%)
HR 13·74 (95% CI 9·58–19·72)
p<0·0001
Time since surgery (years)
Number at risk
High risk 468 357 162 32 2 0
Low risk 739 718 618 352 69 3
C TCGA set
Risk score for every patient
Multimodal recurrence score
Low risk
High risk
WSI-based score
Six-SNP-based score
Leibovich score
True positive rate
False positive rate
AUC at 5 years
p value
Multimodal recurrence score 0·864 (0·818–0·910)
Six-SNP-based score 0·776 (0·710–0·841) 0·03
Leibovich score 0·771 (0·702–0·839) 0·005
WSI-based score 0·770 (0·709–0·831) 0·0008
RFI (%)
HR 11·19 (95% CI 5·52–22·66)
p<0·0001
Time since surgery (years)
Number at risk
High risk 182 77 21 4 0
Low risk 236 144 47 15 0

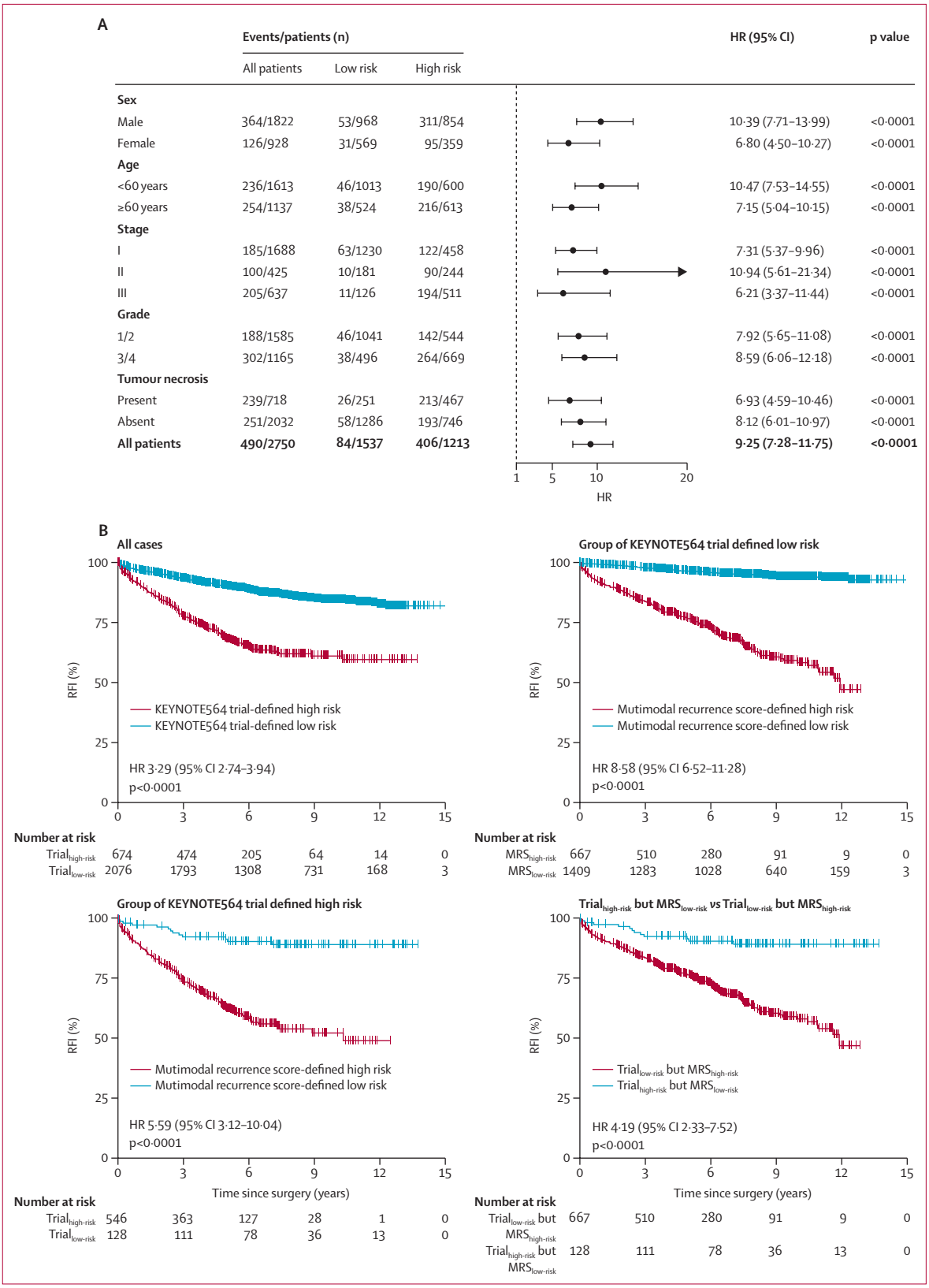


***Figure 3*: HR of RFI for all 2750 patients with clear cell renal cell carcinoma according to the MRS in different subgroups**

(A) HR of RFI for 2750 patients stratified by clinical parameters using univariable Cox regression analysis and forest plot. (B) Kaplan-Meier survival analysis for RFI of the MRS in different subgroups stratified by KEYNOTE564 trial-defined subtypes. HR and 95% CI were computed using the Cox proportional hazards model. p values were calculated using the log-rank test. HR=hazard ratio. MRS=multimodal recurrence score. RFI=recurrence-free interval.

high-risk and low-risk groups, with the same cutoff (risk score=0) as that used in the training set. Patients in the high-risk groups had shorter RFI (HRs 13·74 [95% CI 9·58–19·72] and 11·19 [5·52–22·66], p<0·0001; figure 2) and shorter CSS (HRs 12·73 [8·57–18·91] and 11·79 [4·65–29·90], p<0·0001; appendix p 33) than patients in the low-risk groups in the independent validation and TCGA sets.

When adjusting for clinical variables (age, sex, and stage) by multivariable Cox regression analysis, the multimodal recurrence scoring system was an independent prognostic factor for predicting both RFI and CSS in the training set, independent validation set, and TCGA set (appendix pp 31–34).

An additional survival analysis was performed using the multimodal recurrence scoring system in subsets of patients with different clinical variables. When stratified by clinical variables (age, sex, stage, grade, and tumour necrosis status), the multimodal recurrence score was still a clinically and statistically significant prognostic model for prediction of RFI and CSS (figure 3 and appendix pp 18–22). For example, the recurrence rate in the multimodal system-defined high-risk subgroup was 6·21 times higher (95% CI 3·37–11·44) than in the low-risk subgroup in the patients with stage III clear cell renal cell carcinoma.

The KEYNOTE564 clinical trial defined stage II with grade 4 clear cell renal cell carcinoma and stage III with any grade clear cell renal cell carcinoma as high risk for recurrence and as requiring adjuvant immunotherapy; it also defined stage II with grades 1–3 clear cell renal cell carcinoma and stage I with any grade clear cell renal cell carcinoma as low risk and not requiring adjuvant immunotherapy.[4] The RFI and CSS among patients with KEYNOTE564 trial-defined low-risk disease were significantly longer than among those with KEYNOTE564 trial-defined high-risk disease. Our multimodal system can stratify the KEYNOTE564 trial-defined high-risk group or low-risk group into two subgroups, which is useful because the survival of patients with KEYNOTE564 trial-defined low risk was not always superior to that of patients with KEYNOTE564 trial-defined high risk. The survival of patients with KEYNOTE564 trial-defined low-risk disease but multimodal recurrence system-defined high-risk disease was significantly shorter compared with that of patients with KEYNOTE564 trial-defined high-risk disease but multimodal recurrence system-defined low-risk disease (RFI: HR 4·19 [95% CI 2·33–7·52], p<0·0001; CSS: 4·05 [1·99–8·26], p<0·0001; figure 3 and appendix p 23).

Furthermore, the survival of patients with high-stage and high-grade disease was not always inferior to that of patients with low-stage and low-grade disease. Survival in the subgroup of multimodal recurrence system-defined high risk in stage I or II disease was significantly shorter than that of low risk in stage III disease (RFI: 4·57 [2·49–8·40], p<0·0001; CSS: 4·67 [2·19–9·98], p<0·0001; appendix pp 24–25), and survival in the subgroup of multimodal system-defined high risk in grade 1 or 2 disease was significantly shorter than that of low risk in grade 3 or 4 disease (RFI: 4·58 [3·19–6·59], p<0·0001; CSS: 3·87 [2·55–5·87], p<0·0001; appendix pp 26–27). Thus, our findings suggest that the multimodal recurrence scoring system can provide predictive value that complements prognostic clinical features.

Additionally, we constructed a nomogram that combined the WSI-based score, the six SNP-based score, and the Leibovich score in the training set (figure 4). We validated this nomogram in the independent validation and TCGA sets to provide clinicians with a quantitative method to predict the 3-year, 5-year, and 7-year recurrence-free probabilities in a patient with localised clear cell renal cell carcinoma for individual therapy. Calibration plots showed that the nomogram compared favourably with an ideal model in the training, independent validation, and TCGA sets (figure 4). The AUCs at 3, 5, and 7 years in each set were all higher than 0·80 (figure 4).

Discussion

We conducted a retrospective analysis and multicentre validation study to identify a novel multimodal scoring system that can predict the recurrence risk of localised clear cell renal cell carcinoma more accurately than the existing TNM staging system and three established single-modality models that use clinical, genomic, or histopathological information. Clinical trials of adjuvant targeted therapy in patients with localised renal cell carcinoma define stage III or stage II with grade 4 as high risk; however, our results show that the recurrence risk of the patient subgroup with multimodal system-defined high risk in stage I or II disease was higher than for multimodal system-defined low risk in stage III, and that the recurrence risk of the patient subgroup with multimodal system-defined high risk in grade 1 or 2 was higher than for multimodal system-defined low risk in grade 3 or 4. Therefore, our multimodal system can enable clinicians to make more informed treatment decisions about adjuvant therapy. To our knowledge, this is the largest biomarker discovery project to date in clear cell renal cell carcinoma.

The prognostic model based on genomic signature can add predictive value to the staging system, which can more precisely predict cancer prognosis.[22,23] For example, the 21-Gene Recurrence Score can divide the same clinical stage breast cancer into low-risk and high-risk subgroups with significant differences in RFI, which can guide clinicians regarding adjuvant therapy decisions. The 21-Gene Recurrence Score-defined high-risk subgroup of patients might benefit from adjuvant therapy because of their high recurrence rate, and identifying the low-risk subgroup of patients in the same clinical stage can prevent overtreatment where the absolute benefits of adjuvant therapy are minimal relative

to surgery alone.[7,24] In localised clear cell renal cell carcinoma, the six SNP-based score and the 16-gene assay can provide a more accurate recurrence risk assessment beyond existing staging systems, and both assays, as well as the 21-Gene Recurrence Score, can be conducted using routinely available FFPE tumour tissue,

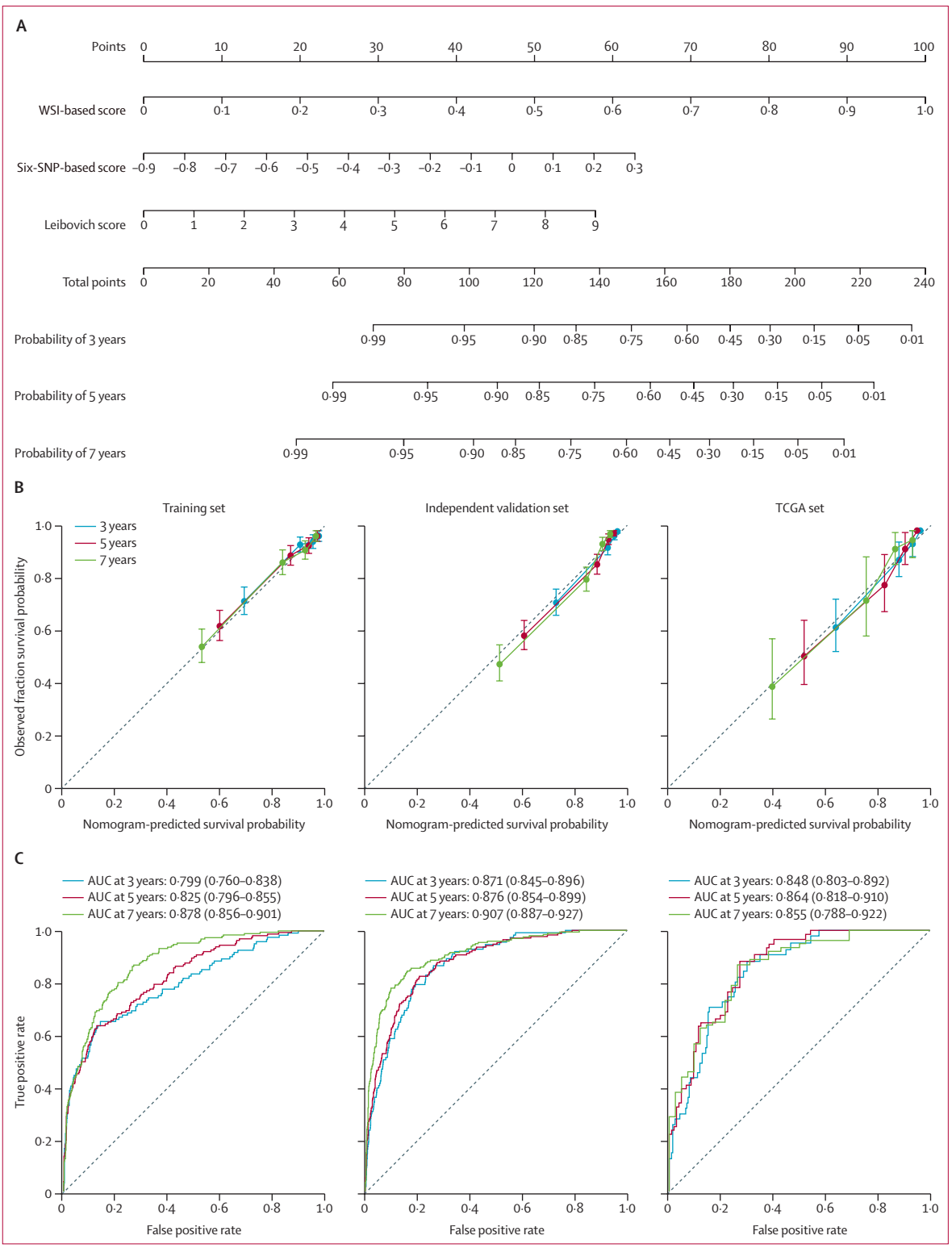


***Figure 4*: Establishing and validating a nomogram that combines the WSI-based score, six-SNP-based score, and Leibovich score to predict RFI**

(A) The position of each variable is found on the corresponding axis, a line is drawn to the Points axis for the number of points, the points from the four variables are added together, and finally a line is drawn from the Total points axis to determine the 3-year, 5-year, and 7-year recurrence-free probabilities at the bottom. (B) Calibration of the nomogram in terms of agreement between predicted and observed 3-year, 5-year, and 7-year outcomes. Model performance is shown relative to the 45° line, representing the performance of an ideal nomogram for which the predicted outcome perfectly corresponds with the actual outcome. (C) Time-dependent ROC curves and AUCs at 3, 5, and 7 years were used to assess the prognostic accuracy of the nomogram. AUC=area under the curve. RFI=recurrence-free interval. ROC=receiver operating characteristics. SNP=single-nucleotide polymorphism. TGCA=The Cancer Genome Atlas. WSI=whole-slide image.

which makes these approaches easy to implement in clinical practice.[9,10,25] Our previous study showed that intratumoural heterogeneity did not hamper the accuracy of the six SNP classifier as a reliable predictive model.[9]

Recent studies have suggested that deep learning could be allied with digital scanning of conventional H&E-stained FFPE tumour tissue sections that potentially use basic morphology to develop markers to predict cancer outcomes that are not readily recognisable by the human eye and that are reproducible.[26,27] This assay is directly applicable to routine H&E-stained sections, which makes it easy to translate to clinical application. Skrede and colleagues constructed a deep learning WSI-based model to predict the CSS of colorectal cancer, and the predictive accuracy of the WSI-based model with 10× resolution was higher than that with 40× resolution.[10] In the present study, we also found that the predictive accuracy of our WSI-based score with 10× resolution was higher than that with 40× resolution in localised clear cell renal cell carcinoma.

In this study, we first developed a recurrence risk model based on the modality of histopathological image analysis using deep learning, and then combined the three modalities—histopathological, genomic, and clinical—into a multimodal risk-scoring system. In construction of a reliable and practical predictive model, the validation process is as important as the development process. The validation process can be conducted in two ways. The first is for the researcher to detect and validate the identified predictive model. The second is to assess and validate the identified predictive model with an external dataset, such as the TCGA dataset, which removes researcher bias. In this study, the multimodal recurrence scoring system is reliable and repeatable because it was validated using both of these approaches, and both validation pathways confirmed similar predictive accuracy of the multimodal system that is significantly higher than that of each single modality and the current staging system.

Our study has limitations that must be considered. First, it is retrospective and should be further validated in a prospective study in multicentre clinical trials. Second, we use manual delineation of tumour areas in our study, which increases the workload of pathologists and is not conducive to large-scale clinical application. In future studies, we will use a convolutional neural network to automate the process for high-throughput clinical applications.

In summary, our multimodal recurrence scoring system is a practical and reliable prognostic tool for localised clear cell renal cell carcinoma, which can complement the existing staging system to predict recurrence after surgery. By identifying patients with localised clear cell renal cell carcinoma at either high or low risk of post-surgery recurrence, our assay can support more informed treatment decisions and trials of adjuvant therapy.

Contributors

J-HL designed the study. All authors obtained and assembled the data. C-PG, Y-HC, C-XL, and P-XL performed the machine learning. C-PG, Y-HC, H-WZ, J-ZC, T-JL, and J-HL analysed and interpreted the data. C-PG, H-WZ, and T-JL verified the underlying raw data. C-PG and J-HL wrote the report, and all authors edited the report and approved the final version. All authors had full access to all the data and had final responsibility for the decision to submit for publication. C-PG and J-HL revised the manuscript. J-HL is the guarantor.

Declaration of interests

We declare no competing interests.

Data sharing

Data will be made available to interested research partners upon reasonable request to corresponding author J-HL or the Institute of Precision Medicine, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, China; the prerequisite for this is a data transfer agreement, approved by the legal departments of the requesting researcher and by all legal departments of the institutions that provided data for the study, and an ethics clearance. All images and patient data from the TCGA cohort used in this study are publicly available at https://portal.gdc.cancer.gov/. We made use of the open-source machine learning frameworks TensorFlow (https://github.com/tensorflow/tensorflow) and MobileNetV3 (https://github.com/xiaolai-sqlai/mobilenetv3) to perform the work. The source code for the deep learning model is available online (Zenodo http://doi.org/10.5281/zenodo.7545427). Moreover, all experimental and implementation details are available in the appendix.

Acknowledgments

This study was supported by grants from the Natural Science Foundation of China (award numbers: 81725016, 81872094, 81972376) and the National Key Research and Development Program of China (award number: 2016YFC0902600). We thank TCGA for providing data.

References

1 Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* 2022; **72:** 7–33.
2 Motzer RJ, Jonasch E, Agarwal N, et al. Kidney cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2022; **20:** 71–90.
3 Stewart SB, Thompson RH, Psutka SP, et al. Evaluation of the National Comprehensive Cancer Network and American Urological Association renal cell carcinoma surveillance guidelines. *J Clin Oncol* 2014; **32:** 4059–65.
4 Choueiri TK, Tomczak P, Park SH, et al. Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. *N Engl J Med* 2021; **385:** 683–94.
5 Ravaud A, Motzer RJ, Pandha HS, et al. Adjuvant sunitinib in high-risk renal-cell carcinoma after nephrectomy. *N Engl J Med* 2016; **375:** 2246–54.
6 Zhang JX, Song W, Chen ZH, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol* 2013; **14:** 1295–306.
7 Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med* 2015; **373:** 2005–14.
8 Leibovich BC, Blute ML, Cheville JC, et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer* 2003; **97:** 1663–71.
9 Wei JH, Feng ZH, Cao Y, et al. Predictive value of single-nucleotide polymorphism signature for recurrence in localised renal cell carcinoma: a retrospective analysis and multicentre validation study. *Lancet Oncol* 2019; **20:** 591–600.
10 Rini B, Goddard A, Knezevic D, et al. A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: development and validation studies. *Lancet Oncol* 2015; **16:** 676–85.
11 Skrede OJ, De Raedt S, Kleppe A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020; **395:** 350–60.
12 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318:** 2199–210.

13 Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 2017; **67:** 93–99.

14 Zhang Z, Hernandez K, Savage J, et al. Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat Commun* 2021; **12:** 1226.

15 Schutz FA, Pomerantz MM, Gray KP, et al. Single nucleotide polymorphisms and risk of recurrence of renal-cell carcinoma: a cohort study. *Lancet Oncol* 2013; **14:** 81–87.

16 Hakimi AA, Ostrovnaya I, Jacobsen A, et al. Validation and genomic interrogation of the MET variant rs11762213 as a predictor of adverse outcomes in clear cell renal cell carcinoma. *Cancer* 2016; **122:** 402–10.

17 Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* 2012; **10:** 51.

18 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162:** 55–63.

19 Heidari S, Babor TF, De Castro P, et al. Sex and gender equity in research: rationale for the SAGER reporting guideline and recommended use. *Res Integr Peer Rev* 2016; **1:** 2.

20 Liang W, Zhang L, Jiang G, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol* 2015; **33:** 861–69.

21 Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56:** 337–44.

22 Tang XR, Li YQ, Liang SB, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. *Lancet Oncol* 2018; **19:** 382–93.

23 Bersanelli M, Travaglino E, Meggendorfer M, et al. Classification and personalized prognostic assessment on the basis of clinical and genomic features in myelodysplastic syndromes. *J Clin Oncol* 2021; **39:** 1223–33.

24 Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010; **11:** 55–65.

25 Sparano JA, Crager MR, Tang G, Gray RJ, Stemmer SM, Shak S. Development and validation of a tool integrating the 21-gene recurrence score and clinical–pathological features to individualize prognosis and prediction of chemotherapy benefit in early breast cancer. *J Clin Oncol* 2021; **39:** 557–64.

26 Courtiol P, Maussion C, Moarii M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med* 2019; **25:** 1519–25.

27 Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* 2019; **16:** e1002730.