Appendix 1 Materials and methods

Patients

To build machine learning models for predicting ADAM17 expression using deep learning features from hematoxylin and eosin (H&E) stained whole-slide images (WSIs), we acquired WSIs of BCa patients from TCGA database. The selection criteria included: (I) confirmed pathological diagnosis of BCa; (II) availability of RNA sequencing data; (III) comprehensive clinical and pathological data; and (IV) high-quality H&E stained WSIs. Exclusion criteria were applied to remove (I) WSIs without evident BLCA lesions; (II) WSIs with substandard scanning quality or those with significant missing values in extracted features; (III) patients lacking RNA sequencing data; and (IV) cases where patients received preoperative treatments.

For model validation, we collected additional H&E-stained slides and RNA sequencing data from BLCA patients treated between September 2017 and May 2024 at two external sites: STPH and The Affiliated Guangdong Second Provincial General Hospital of Jinan University (GD2H). The inclusion and exclusion criteria mentioned above were strictly followed. WSIs (in svs format) were scanned at 40× magnification using an automatic digital slide scanner (KF-PRO-120/005, KFBIO Co., Ltd.). In total, 449 WSIs from 378 TCGA patients, 179 WSIs from 163 patients at STPH, and 53 WSIs from 38 patients at GD2H met the criteria and were included in this study. Patient characteristics are outlined in supplementary information *Table S2*. RNA sequencing data from TCGA, STPH, and GD2H were combined, and FPKM expression values of ADAM17 were batch-corrected using ComBat. Based on the corrected data, patients were stratified into low and high ADAM17 expression groups using the median expression value as the cutoff. Since TCGA provided the largest dataset, its WSIs were randomly divided into a training set and an internal validation set with a 7:3 split. WSIs from STPH and GD2H served as external validation sets. Ethical approval was obtained from the respective committees at STPH (Approval No. 24KT68) and GD2H (Approval No. 2024-KY-KZ-128-01), and informed consent was collected from all participants.

Image annotation and preparation

To develop a deep learning model for identifying tumor and non-tumor regions in bladder cancer (BCa) H&E WSIs, a total of 70 WSIs from the TCGA dataset were randomly selected for model training, with 30 WSIs allocated for internal validation. Additionally, external validation was conducted using 10 WSIs each from the STPH and GD2H datasets. Tumor regions were annotated by an experienced BCa pathologist using QuPath (v0.3.2). The annotated WSIs were divided into 224×224 pixel patches at 20× magnification using the Openslide Python package (v4.0.0), with each patch labeled as tumor or normal tissue. To refine the dataset, edge detection was applied with OpenCV (v4.10.0, threshold 0.02) to remove predominantly blank patches. Subsequently, color normalization was performed on all patches using the Reinhard method in OpenCV to ensure consistency in color representation across datasets, improving model robustness.

Model construction and evaluation

To classify tumor and normal patches, transfer learning was implemented using the Resnet50 model in PyTorch (version 2.4.1). The model training ran for 50 epochs with stochastic gradient descent (SGD), an initial learning rate of 0.1, momentum of 0.9, and weight decay set at 0.001. Cross-entropy loss was the chosen criterion, and model performance was assessed on both an internal validation set and two external validation sets from STPH and GD2H.

The RetCCL model, pre-trained on a diverse set of 22,000 WSIs, was chosen due to its enhanced capacity for feature extraction, particularly through its ability to leverage both intra- and inter-class variation in pathology images. For distinguishing high and low ADAM17 expression groups, two distinct feature extraction approaches based on deep learning were utilized. In the first approach, all background patches were removed using edge detection, after which the RetCCL model, a variant of ResNet50 optimized for contrastive learning, was used to extract 2,048 deep features from the final global average pooling (GAP) layer of both tumor and normal patches. In the second approach, the Resnet50 model trained in this study was applied to isolate tumor patches within the WSIs, and these tumor patches alone were used to derive 2,048 features from the RetCCL model's GAP layer.

For both methods, the extracted 2048 features from each patch were aggregated at the WSI level by calculating summary

 https://dx.doi.org/10.21037/tau-2025-79

statistics such as mean, median, standard deviation (SD), first quartile (Q1), third quartile (Q3), range, and interquartile range (IQR), resulting in a final feature set of 2,048 ×7 variables per WSI. Batch correction for the feature sets across different centers was conducted using ComBat to ensure consistency and mitigate batch effects.

Subsequent to feature extraction, Z-score normalization was applied to the training set features, with the derived parameters saved for use on the internal and external validation sets to maintain consistency. Minimum redundancy maximum relevance (mRMR) analysis was then conducted on the training set to reduce feature redundancy and select the most predictive features for ADAM17 expression prediction. This resulted in the top 30 features being selected from each method, based on the analysis of 315 WSIs from the TCGA training set.

Using the features selected by mRMR, random forest (RF) models were constructed on the training set. Models underwent validation on the internal set as well as the two external cohorts. Hyperparameter tuning was executed through grid search to optimize each model. The grid search varied the number of estimators from 1 to 200 and explored maximum tree depths between 1 and 15. The model with the highest average area under the curve (AUC) and accuracy across internal and external validation sets was considered the optimal model. In cases where multiple models achieved comparable AUC, sensitivity and specificity were evaluated as secondary criteria to ensure balanced performance across low and high ADAM17 expression groups.

Table S1 ADAM17 expression and clinical features in bladder cancer patients based on the TCGA dataset

Group 1	Group 2	Method	Statistic	Difference	95% confidence interval	P value
Normal (n=19)	Tumor (n=406)	WRS test	2,208	0.586	0.242–0.907	0.002
NMIBC (n=22)	MIBC (n=369)	WRS test	2,720	0.462	0.133–0.788	0.009
Papillary (n=132)	Non-papillary (n=269)	WRS test	1.425e+04	0.312	0.124–0.500	0.001
Stage I & II (n=131)	StageIII& IV (n=273)	WRS test	1.555e+04	0.197	0.016–0.381	0.03
Low grade (n=21)	High grade (n=382)	WRS test	2,075	1.013	0.487–1.520	<0.001
PR&CR (n=253)	PD & SD (n=96)	WRS test	9,406	0.340	0.137–0.544	0.001
BMI ≤25 kg/m^2 (n=148)	BMI >25 kg/m^2 (n=208)	WRS test	1.308e+04	0.230	0.047–0.417	0.02
Alive (n =228)	Dead (n=178)	WRS test	1.622e+04	0.305	0.136–0.478	<0.001
Asian (n=44)	African American (n=23)	Dunn's test	3.933	113.800	–	<0.001
Asian	White (n=322)	Dunn's test	4.552	82.230	–	<0.001
African American	White	Dunn's test	–1.300	–31.565	–	0.58

https://dx.doi.org/10.21037/tau-2025-79

Table S2 Characteristics of patients with bladder cancer from multiple datasets

Characteristic	TCGA (n=378)	STPH (n=163)	GD2H (n=38)
Sex, n (%)			
Women	97 (25.7)	29 (17.8)	5 (13.2)
Men	281 (74.3)	134 (82.2)	33 (86.8)
Age, n (%)			
<65 years	137 436.2)	49 (30.1)	18 (47.4)
≥65 years	241 (63.8)	114 (69.9)	20 (52.6)
Pathologic T stage, n (%)			
NMIBC (< pT2)	3 (0.8)	112 (68.7)	22 (57.9)
MIBC (≥ pT2)	375 (99.2)	51 (31.3)	16 (42.1)
Pathological grade, n (%)			
Low	21 (5.6)	34 (20.9)	13 (34.2)
High	354 (93.7)	129 (79.1)	25 (65.8)
NA	3 (0.8)	0 (0)	0 (0)

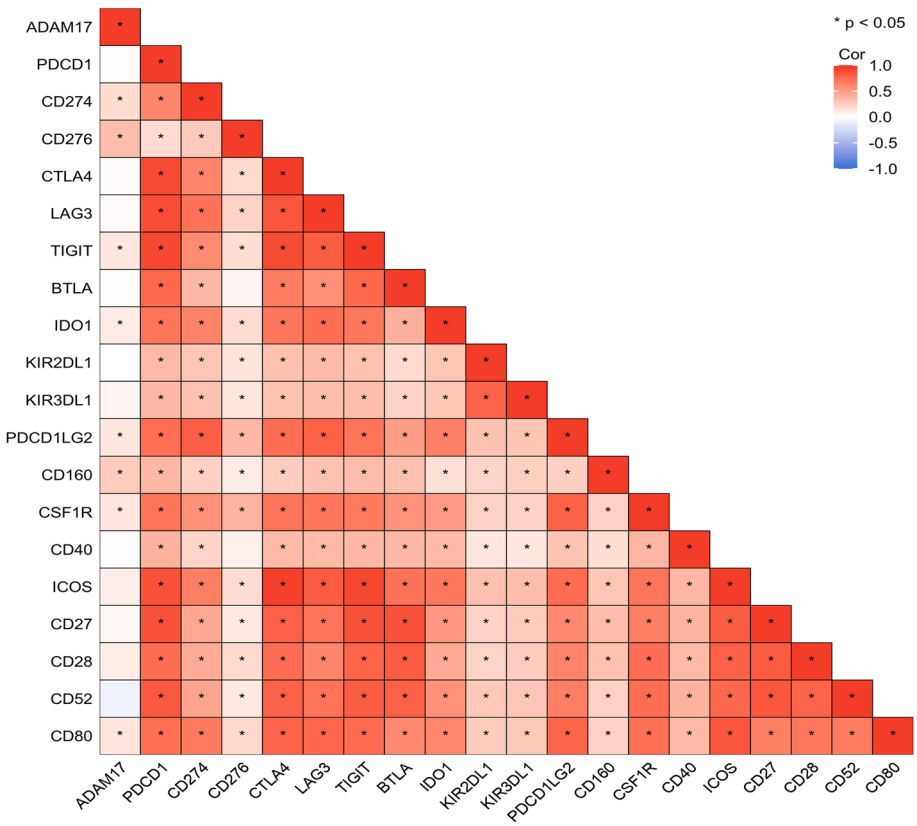

Figure S1 Comparison of the expression correlation between ADAM17 and immune checkpoint genes.

https://dx.doi.org/10.21037/tau-2025-79