Pan-cancer computational histopathology reveals tumor mutational burden status through weakly-supervised deep learning

Siteng Chen\textsuperscript{1*}, Jinxi Xiang\textsuperscript{2*}, Xiyue Wang\textsuperscript{3*}, Jun Zhang\textsuperscript{2*}, Sen Yang\textsuperscript{2}, Junzhou Huang\textsuperscript{2}, Wei Yang\textsuperscript{2}, Junhua Zheng\textsuperscript{4*}; Xiao Han\textsuperscript{2*}

\textsuperscript{1} Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China.
\textsuperscript{2} Tencent AI Lab, Shenzhen 518057, China.
\textsuperscript{3} College of Computer Science, Sichuan University, Chengdu 610065, China.
\textsuperscript{4} Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200135, China

*Equal contributors and co-first authors

\textsuperscript{8} Corresponding authors:
Jun Zhang, Tencent AI Lab, Shenzhen 518057, China. E-mail: junejzhang@tencent.com. Tel: 86-0755-86013388-887069.
Junhua Zheng, Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China. E-mail: zhengjih0471@sjtu.edu.cn. Tel: 86-021-63240090.
Xiao Han, Tencent AI Lab, Shenzhen 518057, China. E-mail: haroldhan@tencent.com. Tel: 86-0755-86013388-779087.

Keywords
Immunotherapy, Tumor mutational burden, Pan-cancer, Deep learning, Histopathology

Acknowledgements We appreciate the partial image data from Clinical Proteomic Tumor Analysis Consortium, the Cancer Genome Atlas, and the Cancer Imaging Archive used in this study.

Contributors Conception and design of study: JHZ, XH and JZ. Acquisition of data: SY, JZH, WY and STC. Analysis and interpretation of data: STC, JXX, and XYW. Writing and review of manuscript: STC, JXX, XYW, JZ, XH and JHZ.

Funding This study was supported by the National Natural Science Foundation of China (81972393).

Competing interests The authors have no conflict of interest.

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Primary data are available from Atlas (https://portal.gdc.cancer.gov/) and the Clinical Proteomic Tumor Analysis Consortium (https://www.cancerimagingarchive.net/). All aggregate data relevant to the study are included in the article or uploaded as supplementary information.
ABSTRACT

Background Tumor mutational burden (TMB) is a potential genomic biomarker that can help identify patients who will benefit from immunotherapy across a variety of cancers. However, TMB detected through whole exome sequencing lacks clinical penetration in low-resource settings due to the high cost, operational complexity, and lengthy time for test. Therefore, a scalable and time-efficient approach for the identification of TMB status with low cost is desirable for patients.

Methods We included whole slide images (WSIs) of 3228 diagnostic slides from the Cancer Genome Atlas and 531 WSIs from the Clinical Proteomic Tumor Analysis Consortium for the development and verification of a pan-cancer TMB prediction model (PC-TMB). We proposed a multiscale weakly-supervised deep learning framework for predicting TMB of seven types of tumors based only on routinely used hematoxylin-eosin (H&E)-stained WSIs.

Results PC-TMB achieved a mean area under curve (AUC) of 0.818 (0.804-0.831) in the cross-validation cohort, which was superior to the best single-scale model. In comparison with the state-of-the-art TMB prediction model from previous publications, our multiscale model achieved better performance over previously reported models. In addition, the improvements of PC-TMB over the single-tumor models were also confirmed by the ablation tests on ×10 magnification. The PC-TMB algorithm also exhibited good generalization on external validation cohort with AUC of 0.732 (0.683-0.761). PC-TMB possessed a comparable survival-risk stratification performance to the TMB measured by whole exome sequencing, but with low cost and being time-efficient for providing a prognostic biomarker of multiple solid tumors. Moreover, spatial heterogeneity of TMB within tumors was also identified through our PC-TMB, which might enable image-based screening for molecular biomarkers with spatial variation and potential exploring for genotype-spatial heterogeneity relationships.

Conclusions PC-TMB was able to predict TMB status of multiple solid tumors accurately based on routinely used H&E pathology slides, which provided a scalable approach for the identification of TMB status and enable personalized treatment decision-making.

INTRODUCTION

The development of immunotherapy has been reported as an important breakthrough for the treatment of cancer. The expression of programmed cell death 1 (PD-1) ligand (PD-L1) in tumor cells,\textsuperscript{1} microsatellite instability, and tumor mutational burden (TMB)\textsuperscript{2,3} have been currently reported to act as effective biomarkers for immune-checkpoint inhibitor (ICI) therapy. However, only TMB has been proved to serve as a candidate biomarker of clinical outcome from immunotherapy for multiple solid tumors.\textsuperscript{4,5} A higher level of TMB is associated with the increased new antigens, which could be easier recognized by the immune system and help benefit from immunotherapy.\textsuperscript{6}

TMB is calculated by counting the number of nonsynonymous mutations across a length of genome sequenced through whole-exome sequencing (WES), which is reported as mutations per coding area of a tumor genome.\textsuperscript{7} In spite of the outstanding performance in predicting the immunotherapeutic responsiveness, TMB is not widely available in clinical applications due to the high cost, operational complexity, and lengthy time for WES.\textsuperscript{8} Even though TMB can also be evaluated from lower-cost panels with targeted genes, potential bias exist due to the limited fraction of the exome sequenced.\textsuperscript{7} Therefore, a scalable and time-efficient approach for the identification of
TMB status with low cost is desirable for patients.

Computational histopathology algorithms, often based on convolutional neural network (CNN), can process and cross-reference large volumes of data to aid in quantifying aberrant cells and tissues. Recent advances in deep learning from hematoxylin-eosin (H&E) stained whole slide image (WSI) have also demonstrated satisfactory performance on capturing phenotypes that are typically not recognized by experienced pathologists. Establishing associations of a digitized WSI with high or low TMB without detailed annotations at the cellular and regional levels is typically considered as weakly-supervised learning problem, and the challenge lies in determining visual clues in slides associated with specific gene status. Currently, related works for TMB prediction from H&E-stained slides only focused on a single cancer, such as stomach adenocarcinoma, colon adenocarcinoma, or lung adenocarcinoma, separately. Since TMB has been proved to be a predictive biomarker for immunotherapy in multiple solid tumors, it comes with an interesting question whether deep learning-based prediction of TMB from H&E-stained slides can be extended from single cancer to multiple solid tumors. The investigation of the pan-cancer model can potentially reduce the work-ups to develop customed models for each cancer type, and also could probably enhance the overall prediction performance by exploiting larger datasets and universal features.

In this study, we aimed to propose a novel weakly-supervised deep learning method with spatial-awareness for predicting TMB of a range of tumors based only on H&E-stained sections and patient-level labels. To our knowledge, this is the first study to propose a state-of-the-art framework for TMB prediction from histopathological images of multiple solid tumors, which might promote the potential utility of these neural networks in exploring biomarkers for immunotherapy.

**METHODS**

**Study design and participants**

In this retrospective study, we developed a weakly-supervised deep learning framework for predicting TMB of a range of tumors from H&E-stained WSIs. The pan-cancer patient cohort was recruited from and the Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) for this study. In addition, an independent pan-cancer cohort was collected from the Clinical Proteomic Tumor Analysis Consortium (CPTAC, https://www.cancerimagingarchive.net/). All recruited patients shall meet the following selection criteria: (i) pathologically diagnosed as a single cancer without other types of malignant tumors; (ii) with corresponding clinical and pathological information; (iii) with access to diagnostic H&E-strained WSIs; (iv) with retrievable TMB or total mutation count data (figure 1).

**Patient-level classification**

TMB of the patient from TCGA was acquired from the cBioPortal for cancer genomics, which was defined as nonsynonymous coding mutations per Megabase (Mb). Since currently there was no consensus in the determined value of TMB-high (TMB-H) and TMB-low (TMB-L), the cut-off value for patient stratification was defined as 10 mutations per Mb (mut/Mb) according to previous reports for this study. Patients with nonsynonymous TMB more than 10 mut/Mb were grouped into TMB-H. The independent validation cohort contained only total mutation count data. It would cause bias by setting the cut-off value as 10 mutations. Therefore, we performed the correlation analysis in the TCGA cohort and found that the total mutation count was highly correlated to TMB.
(related coefficient: 0.98, p < 0.0001, figure S1A). When the cut-off value of TMB was set as 10 mut/Mb, the relative total mutation count value was found to be 279 for the same percentage of TMB-H (figure S1B). So, the cut-off value of total mutation count of TMB-H and TMB-L was set as 279 for the independent validation. After excluding some cancer species with very low percentage of TMB-H and small data volume, we selected 7 types of cancer with a proportional of TMB-H for analysis, including colon adenocarcinoma (COAD), stomach adenocarcinoma (STAD), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), bladder urothelial carcinoma (BLCA), head and neck squamous cell carcinoma (HNSC), and uterine corpus endometrial carcinoma (UCEC). The prevalence of TMB-H by cancer types was shown in Table S1. We performed five-fold cross-validation in the TCGA cohort for the training and internal cross validation of the pan-cancer TMB prediction model (PC-TMB), which was further externally verified in the CPTAC cohort.

**Feature embedding**

The gigapixel WSIs were loaded into memory at a downscaled level and then automated segmented with Otsu’s method to exclude non-tissue regions. After region selection, we cropped the WSIs into tiles of non-overlapping 256 $\times$ 256 pixels within the foreground regions at user-customed magnification level, i.e., $\times$5, $\times$10, and $\times$20 in our scenarios. To preserve informative representation of raw pixels, ResNet50 was pretrained on the large-scale TCGA dataset following the unsupervised contrastive learning paradigm to extract robust and universal features across different tumor types from H&E slides (figure 1). Following patching, we used the pretrained ResNet50 to compute a relatively low-dimensional feature representation for each cropped patch. The use of extracted features, i.e. 2048-dimensional, made it tractable to train the deep learning models with all tiles in a slide (up to 10,000 patches or more in a slide) simultaneously on a GPU.

**Feature convolution and aggregation**

After feature embedding, the following models extracted information from features via graph convolution and attention-based aggregation. The feature embedding from pretrained CNN could describe local morphology of patches. However, single-stream path of 256 $\times$ 256 pixels did not capture adequate spatial context from the tissue micro-environments. Graph based representations could effectively describe the tissue composition by incorporating topology and interactions among entities, i.e. patches from the same slide. To construct the global graph representation of a slide, we saved (x, y) coordinates of cropped patches $X_f$ in raw images to build the adjacency matrix $A_f$ for each slide via fast approximate nearest neighbors k-NN (k=8). A patient-level slide could be denoted as CNN features and graph features $(X_f, A_f)$. Graph convolution network updated a node’s feature vector through aggregating feature information from its neighbors in the graph. We adapt DeepGCN to process the CNN features $X_f$ along with its corresponding graph adjacent matrix $A_f$ following the ordering: normalization, ReLU activation, GraphConv, and addition. Empirically, last step addition was residual learning which was helpful in training deep network. After graph convolution, the raw feature embedding $X_f \in \mathbb{R}^{N\times2048}$ is translated into graph representation $G_f \in \mathbb{R}^{N\times512}$, supposing that the initial embedding dimension is 2048 and the translated dimension was 512 and N was the number of patches in a slide. After graph convolution, the following feature aggregation module was built around the trainable and interpretable attention-based function to aggregate slide level representation into a single vector $V_f$ for classification.
Multi-scale classifier

We aggregated the slide representations and the corresponding class label to produce a single probability per patient per magnification. In this retrospective study, we investigated on pan-cancer TMB prediction and supposed that the cancer type of each slide could be exploited as additional information, providing cancer-specific clues such as visual representation heterogeneity of tumor for deep learning model. The class token was encoded in one-hot vector. Then the attention-weighted feature vector $V_j$ and the class feature $C_j$ were fused by concatenation; the class token features were expanded by feeding into a fully connected layer to match the dimension of image features. Then, the fused features were fed into a classifier and the outputs and the ground-truth labels were used to calculate the cross-entropy loss for a given magnification level. Operating at a higher resolution captures local cellular information but limited the field-of-view due to computational burden and limited the access to global tissue microenvironment information. In contrast, operating at a lower resolution hindered resolvability of cells and access to cellular properties. We aggregated patient-level predictions from different magnifications i.e., $\times 5$, $\times 10$, and $\times 20$ with a multiscale classifier.

Model training

Self-ensemble model training strategy was used to stabilize the training process. Each patch of H&E-stained WSI carried quantitative (not categorical positive or negative) information to the TMBs, with tumor architecture being a continuum. Thus, the supervision between label (i.e., positive or negative mutation burden) and input (i.e., H&E-stained WSI) was considered to be weak and inexact, compared with other classification task in natural image with distinctive labels. To stabilize the training procedure, we saved an exponential moving average version of the on-line training model. The aim was to establish ensemble learning as a backbone to form a solid consensus of the self-ensemble predictions.$^{20}$

Statistical analysis

In this study, receiver operating characteristic curve (ROC) analysis with area under the curve (AUC) was carried out to evaluate the accuracy of the prediction model. Continuous variant between two groups was compared through Mann-Whitney U test, and P value less than 0.05 was defined as significant. Kaplan-Meier analysis and log-rank tests were also applied to assess the survival outcomes among subgroups with TMB-H and TMB-L, measured as hazard ratios (HR) with 95% confidence interval (CI).

RESULTS

Overall performance of the PC-TMB

The analysis pipeline of our study was illustrated in figure 1. A total of 3228 WSIs from the TCGA cohort and 531 WSIs from the independent CPTAC cohort were finally included for analysis. In order to improve the prediction accuracy, the PC-TMB was developed based on a multiscale approach across three magnifications ($\times 5$, $\times 10$, and $\times 20$). As shown in figure 2A, the PC-TMB achieved a mean AUC of 0.818 (0.804-0.831) in five-fold cross-validation of TCGA cohort. The overall performance showed clinically significance of TMB predictions from the PC-TMB, and two highly accurately predicted TMB were COAD and STAD, with AUCs as high as 0.899 and 0.878.
Comparation with the state-of-the-art models

We next compared our PC-TMB with the state-of-the-art models for TMB prediction from previous publications using the same training and test split in the TCGA cohort. Paired observations indicated that our PC-TMB achieved superior performance over previously reported models by a large margin,\(^{12,21,22}\) on COAD (0.899 vs 0.820), LUAD (0.804 vs 0.710), STAD (0.878 vs 0.750), and BLCA (0.757 vs 0.750) with improvements of AUC from 0.5% to 12.8% (table 1).

The PC-TMB model exhibited robustness and high accuracy on different cancer types in predicting TMB status from WSIs.

| Type   | AUC       | Sensitivity | Specificity | Reported AUC |
|--------|-----------|-------------|-------------|--------------|
| COAD   | 0.899(0.864-0.927) | 0.886(0.787-0.949) | 0.801(0.752-0.843) | 0.82 (0.72–0.91) [12] |
| STAD   | 0.878(0.842-0.909)  | 0.842(0.740-0.916)  | 0.749(0.698-0.796)  | 0.75 (0.65–0.84) [12] |
| LUAD   | 0.804(0.767-0.837)  | 0.746(0.673-0.809)  | 0.727(0.677-0.772)  | 0.71 (0.63–0.80) [21] |
| BLCA   | 0.757(0.715-0.796)  | 0.870(0.797-0.924)  | 0.538(0.483-0.593)  | 0.75 (0.68-0.80) [22] |
| HNSC   | 0.790(0.749-0.827)  | 0.750(0.621-0.853)  | 0.702(0.654-0.747)  | /             |
| LUSC   | 0.718(0.675-0.758)  | 0.677(0.601-0.746)  | 0.665(0.609-0.717)  | /             |
| UCEC   | 0.814(0.779-0.845)  | 0.653(0.581-0.720)  | 0.840(0.798-0.876)  | /             |

AUC, area under curve; TMB, tumor mutational burden.

Multi-scale model improves over single-scale counterpart

We further conduct ablation study on multi-scale training against single-scale, arguing that pan-cancer training with multi-scale could greatly improve the model performance. Evaluated by five-fold cross-validation in the TCGA cohort, the prediction model achieved a mean AUC of 0.721 (95% CI 0.705-0.737) on ×5 magnification scale, 0.771 (0.756-0.785) on ×10 magnification, and 0.787 (0.772-0.801) on ×20 magnification (figure 3). The improvement of multi-scale model (PC-TMB, AUC = 0.818) over the best single-scale model (×20 magnification, AUC = 0.787) suggested that there was complementary information that the deep learning algorithm could capture from different levels of granularity in the images, indicating that PC-TMB algorithm might capture more favorable features from different levels of granularity in the WSIs.

Pan-cancer task outperforms single cancer task

One can choose to train an independent model for each cancer type as most previous studies did. Instead, we trained a unified model using data from all cancer types simultaneously, trying to discover novel associations of visual traits across different cancer. To this end, another ablation study was performed to argue the improvements in identifying TMB status by PC-TMB over the single- tumor models. For comparison, we trained 7 individual deep learning model for each cancer type using the same weakly-supervised network on ×10 magnification.

Remarkably, we found that the pan-cancer model PC-TMB outperformed single cancer model on 6 of the 7 cancer types, specifically, COAD (0.874 vs 0.814), LUAD (0.751 vs 0.722), STAD (0.820 vs 0.714), BLCA (0.687 vs 0.659), HNSC (0.734 vs 0.667), and LUSC (0.657 vs 0.559), providing a superior predictive accuracy of the PC-TMB (on ×10 magnification) compared with
respective single-cancer model (table 2). 

**Table 2** Comparation of the performance of the pan-cancer model and respective single-cancer model for TMB prediction based on TCGA dataset

| Type   | Pan-cancer model | Single-cancer model |
|--------|------------------|---------------------|
|        | AUC   | Sensitivity | Specificity | AUC   | Sensitivity | Specificity |
| COAD   | 0.874 | 0.900     | 0.693      | 0.814 | 0.702     | 0.878      |
|        | (0.837-0.905) | (0.805-0.959) | (0.639-0.743) | (0.768-0.855) | (0.577-0.807) | (0.833-0.915) |
| STAD   | 0.820 | 0.829     | 0.689      | 0.757 | 0.714     | 0.747      |
|        | (0.779-0.857) | (0.725-0.906) | (0.635-0.740) | (0.705-0.804) | (0.594-0.816) | (0.686-0.801) |
| LUAD   | 0.751 | 0.728     | 0.652      | 0.705 | 0.722     | 0.593      |
|        | (0.712-0.788) | (0.654-0.793) | (0.600-0.702) | (0.664-0.744) | (0.648-0.788) | (0.539-0.644) |
| BLCA   | 0.687 | 0.683     | 0.639      | 0.642 | 0.659     | 0.578      |
|        | (0.642-0.730) | (0.593-0.764) | (0.584-0.691) | (0.596-0.686) | (0.568-0.742) | (0.522-0.632) |
| HNSC   | 0.734 | 0.567     | 0.800      | 0.656 | 0.667     | 0.580      |
|        | (0.690-0.774) | (0.432-0.694) | (0.757-0.839) | (0.608-0.701) | (0.533-0.783) | (0.527-0.631) |
| LUSC   | 0.657 | 0.594     | 0.687      | 0.533 | 0.559     | 0.526      |
|        | (0.612-0.699) | (0.516-0.669) | (0.632-0.738) | (0.487-0.578) | (0.481-0.635) | (0.469-0.583) |
| UCEC   | 0.773 | 0.751     | 0.707(0.657) | 0.777 | 0.589     | 0.857      |
|        | (0.736-0.807) | (0.684-0.811) | (0.764) | (0.740-0.812) | (0.515-0.659) | (0.816-0.892) |

AUC, area under curve; TMB, tumor mutational burden.

Validation on external cohort

The PC-TMB algorithm also exhibited good generalization on external validation CPTAC cohort (5 cancer types, n=531) with AUC of 0.732 (0.683-0.761, figure 2B). The prediction of TMB status was well replicated in vast majority of cancer types in the external cohort, especially for LUAD [AUC=0.778 (0.687-0.853), p < 0.0001], COAD [AUC=0.719 (0.635-0.793), p < 0.0001], and HNSC [AUC=0.705 (0.601-0.796), p = 0.0019]. A moderate drop in accuracy was found in the external validation, which might result from the potential batch effect and the limited sample size of the validation cohort.

Spatial heterogeneity of TMB through PC-TMB

Spatial heterogeneity of TMB within tumors was also identified through our PC-TMB. We used the proposed model pipeline to predict TMB status of slides from 7 cancer types. The most attended regions recognized by PC-TMB were considered to be associated with TMB-H. Areas with red color of the heatmap represented the regions with predicted TMB-H status (figure 4). The patch-level visualization by our model presented the spatial distributions of diverse mutational burden, which might potentially open avenues for exploring for genotype-spatial heterogeneity relationships.

Prognosis prediction across cancer types

Prognosis prediction across cancer types could still be found in the predicted TMB status through PC-TMB algorithm. Patients with TMB-H status seemed to have favorable survival outcomes when compared with patients with TMB-L status (figure 5A). As shown in figure 5B, patients with different predicted TMB statuses still had distinct prognosis during the follow-up of over ten years,
with HR of 0.761(0.674-0.859), illustrating that our predicted TMB status had a comparable risk stratification performance to the TMB measured by WES, but with low cost and being time-efficient for providing a prognostic biomarker of multiple solid tumors.

**Evaluation of the PC-TMB through stratification analysis**

Except for TMB, mutations in specific genes were also associated with immunotherapy effect, such as DNA damage and repair (DDR), TP53, KRAS, and EGFR which had also been proved to be highly correlated to deep learning-based image features. We next explore whether the performance of our PC-TMB could be independent of specific genetic mutation status. Surprisingly, stratification analysis confirmed the robust prediction performance of PC-TMB in spite of the mutation status of some key genes, including DDR-related genes, TP53, KRAS, and EGFR (figure 6A, C, E, G). The predicted score for TMB status distributed distinctly between patients with TMB-H and TMB-L (figure 6B, D, F, H), which further signified the robustness of our model when applied to different scenarios. In addition, further stratification analysis revealed that our PC-TMB was also independent of tumor grade (figure S2A) and tumor stage (figure S2B).

**DISCUSSION**

The predictive value of TMB for response to ICI therapy has been recognized in multiple cancers. However, how to identify TMB in an appropriate way is still challenging since there are more than 30000 genes in a single cell and the calculation of various mutations in malignancies is intricate. Using the WES technology, a fairly complete landscape of coding mutations could be acquired for the calculation of TMB. Nevertheless, the WES needs demanding samples and is strongly suggested for identifying tumor-specific variants from a tumor sample and matched normal samples, which incurs additional costs and time delays. In addition, larger sizes of selected genes in some locally-accessible panels are still wanted for the normalized TMB evaluation.

In this study, we proposed a novel PC-TMB based on weakly-supervised deep learning strategy for predicting TMB across multiple tumor types from H&E-strained slides, which is routinely used for cancer diagnosis of patients. Taking about 4 minutes to scan the H&E-strained slide and automatically analyze the WSI through our deep learning pipeline, the pathologist can predict the TMB status and evaluate whether the patient is suitable for immunotherapy as soon as he or she makes the pathologic diagnosis.

Unlike previous studies only focusing on specific single cancer species, our PC-TMB could be applied in 7 different types of cancers with satisfying prediction accuracy, especially in COAD and STAD. To our knowledge, this is the first study reporting a TMB prediction model across multiple solid tumors. This result displayed the feasibility of pan-cancer learning-based inference of TMB status across multiple solid tumors directly from histological images, which provided perspectives evidence for generalization of multi-carcinomatous species in genetic prediction.

Furthermore, we identified the improvements in identifying TMB status by PC-TMB over the single-modality tumor models. Critical challenge for predicting TMB status consists in the balance between sensitivity and specificity consisting with WES. The improvements of prediction accuracy in the pan-cancer model might profit from the increased training dataset and the cross talk of image features form multiple solid tumors. Our study provided an important hint for deep image learning based on the connectivity among various cancer species.

Weakly-supervised deep learning strategy was carried out for our prediction pipeline. Unlike
the task of tumor segmentation which could be manually annotated, only slide-level labels for TMB-
H or TMB-L were accessed for our study. It is also impossible to annotate the regions with high TMB for supervised learning. Limited data without annotation has boosted the critical need of data-efficient strategy through weakly-supervised learning.\textsuperscript{31,32} Our framework also provided interpretable visualizations in patch-level to demonstrate the attentions associated with TMB-H within each sub-region of the WSI, making it more practical in clinical practice via elucidating and quantifying TMB in immune microenvironment.

Compared with current WES technology, our PC-TMB also presented the spatial heterogeneity of TMB within tumors. Intratumor heterogeneity of tumor microenvironment plays vital roles in the response to oncotherapy.\textsuperscript{33} Tumor cells with TMB-H status are closely related to higher level of neo-antigens and increased recognition by T cell recognition, which forebodes better ICI reactiveness clinically. Only specific cells in the tumor microenvironment have real value to be deeply tested and studied. It is also uneconomic and impossible to use the entire surgically-removed tumor masses for genetic test. Therefore, the selection of appropriate tumor region for deep genetic test is one of key problems waiting for solution. Our PC-TMB might provide a scalable and time-efficient approach for this requirement.

Deep learning approaches to a single dataset are prone to overfit and should be validated in external populations before clinical deployment. However, the generalizability of the prediction model is limited in previous studies. In this study, our PC-TMB algorithm also exhibited good generalization on external validation cohort, which might attribute to (1) effective feature representation with self-supervised learning performed on TCGA was much more robust than the ubiquitous transfer learning on ImageNet; (2) feature aggregation model using graph neural network and attention pooling to capture instance-level histology features and topological structures in the tumor microenvironment; (3) multi-scale prediction ensemble to take advantage of more complementary information at different resolutions.

In spite of the preliminary application of WSI for predicting TMB in certain tumors, it is not yet applying deep learning for direct prediction of ICI outcome, which might due to the lack of relative clinical data with pathological images. Prediction of indirect outcome indicator might cause bias in primal intentions. Even though there are still challenges for the prediction of ICI reactivity, our study had preliminarily confirmed the great potential of deep learning for the selection of candidates who might respond to ICI treatment based on histopathological images.

Some limitations could also be found in this study. Firstly, the images analyzed in this study were acquired for public databases, which might be affected by the potential population bias. Secondly, only five of seven cancer types were verified in the external validation cohort, which might reduce the stringency of our study to a certain extent. Thirdly, this study is retrospective, which might need further validations in prospective clinical studies.

**CONCLUSION**

In conclusion, our study served as a proof-of-concept for developing multiscale, weakly-supervised deep learning framework for TMB status prediction and paved the way for prospective clinical trials to further assess the efficacy of deep learning-based TMB. Our study also identified the spatial heterogeneity of TMB within each tumor and enable image-based screening for molecular biomarkers with spatial variation and potential exploring for genotype-spatial heterogeneity relationships to facilitate personalized treatment decision-making.
REFERENCES

1. Schoenfeld AJ, Rizvi H, Bandlamudi C, et al. Clinical and molecular correlates of PD-L1 expression in patients with lung adenocarcinomas. *Ann Oncol* 2020; 31(5): 599-608.
2. Karn T, Denkert C, Weber KE, et al. Tumor mutational burden and immune infiltration as independent predictors of response to neoadjuvant immune checkpoint inhibition in early TNBC in GeparNuevo. *Ann Oncol* 2020; 31(9): 1216-22.
3. He B, Dong D, She Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer* 2020; 8(2).
4. Samstein RM, Lee CH, Shoushtari AN, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* 2019; 51(2): 202-6.
5. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med* 2017; 377(25): 2500-1.
6. Chan TA, Yarchoan M, Jaffee E, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* 2019; 30(1): 44-56.
7. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* 2017; 9(1): 34.
8. Steuer CE, Ramalingam SS. Tumor Mutation Burden: Leading Immunotherapy to the Era of Precision Medicine? *J Clin Oncol* 2018; 36(7): 631-2.
9. Mu W, Jiang L, Shi Y, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer* 2021; 9(6).
10. Barisoni L, Lafata KJ, Hewitt SM, Madabhushi A, Balis UGJ. Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol* 2020; 16(11): 669-85.
11. Yu F, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nature Cancer*.
12. Wang L, Jiao Y, Qiao Y, Zeng N, Yu R. A novel approach combined transfer learning and deep learning to predict TMB from histology image. *Pattern Recognition Letters* 2020; 135(1).
13. Jain MS, Massoud TF. Predicting tumour mutational burden from histopathological images using multiscale deep learning. *Nature Machine Intelligence* 2020; 2(6): 356-62.
14. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; 6(269): p11.
15. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; 2(5): 401-4.
16. Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *N Engl J Med* 2018; 378(22): 2093-104.
17. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum Contrast for Unsupervised Visual Representation Learning. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020; 2020.
18. Wang X, Yang S, Zhang J, et al. TransPath: Transformer-Based Self-supervised Learning for Histopathological Image Classification. 2021; Cham: Springer International Publishing; 2021. p. 186-95.
19. Li G, Xiong C, Thabet A, Ghanem B. DeeperGCN: All You Need to Train Deeper GCNs. 2020.
20. Nguyen DT, Mummmadi CK, Ngo T, Nguyen T, Beggel L, Brox T. SELF: Learning to Filter Noisy Labels with Self-Ensembling. 2019.
21. Sadhwani A, Chang HW, Behrooz A, et al. Comparative analysis of machine learning approaches.
to classify tumor mutation burden in lung adenocarcinoma using histopathology images. Sci Rep 2021; 11(1): 16605.

22. Xu H, Park S, Clemenceau JR, et al. Spatial heterogeneity and organization of tumor mutation burden and immune infiltrates within tumors based on whole slide images correlated with patient survival in bladder cancer. bioRxiv 2020: 554527.

23. Ricciuti B, Recondo G, Spurr LF, et al. Impact of DNA Damage Response and Repair (DDR) Gene Mutations on Efficacy of PD-(L)1 Immune Checkpoint Inhibition in Non-Small Cell Lung Cancer. Clin Cancer Res 2020; 26(15): 4135-42.

24. Vadakekolathu J, Laï C, Reeder S, et al. TP53 abnormalities correlate with immune infiltration and associate with response to flotetuzumab immunotherapy in AML. Blood Adv 2020; 4(20): 5011-24.

25. Dong ZY, Zhong WZ, Zhang XC, et al. Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. Clin Cancer Res 2017; 23(12): 3012-24.

26. Lee CK, Man J, Lord S, et al. Checkpoint Inhibitors in Metastatic EGFR-Mutated Non-Small Cell Lung Cancer-A Meta-Analysis. J Thorac Oncol 2017; 12(2): 403-7.

27. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015; 348(6230): 124-8.

28. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. Nat Rev Cancer 2019; 19(3): 133-50.

29. Meléndez B, Van Campenhout C, Rorive S, Remmelink M, Salmon I, D’Haene N. Methods of measurement for tumor mutational burden in tumor tissue. Transl Lung Cancer Res 2018; 7(6): 661-7.

30. Buchhalter I, Rempel E, Endris V, et al. Size matters: Dissecting key parameters for panel-based tumor mutational burden analysis. Int J Cancer 2019; 144(4): 848-58.

31. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 2021; 5(6): 555-70.

32. Azarianpour S, Corredor G, Bera K, et al. Computational image features of immune architecture is associated with clinical benefit and survival in gynecological cancers across treatment modalities. J Immunother Cancer 2022; 10(2).

33. Jia Q, Wu W, Wang Y, et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. Nat Commun 2018; 9(1): 5361.
Figure 1 The analysis pipeline and network architecture of this study.
**Figure 2** The overall performance of the PC-TMB. AUCs were evaluated in five-fold cross-validation in the internal validation cohort (A) and the independent validation cohort (B). AUC, area under curve; TMB, tumor mutational burden.
Figure 3 Prediction performance of the multiscale model across multiple cancer types. AUCs were evaluated in five-fold cross-validation on ×5, ×10 and ×20 magnifications, respectively. AUC, area under curve.
**Figure 4** Spatial heterogeneity of predictions on ×20 magnifications across multiple solid tumors. TMB, tumor mutational burden; pred prob, predicted probability.

**Figure 5** Prognosis prediction across cancer types. Kaplan-Meier survival analysis stratified by actual TMB status (A) and predicted TMB status (B). TMB, tumor mutational burden.
Figure 6 Stratification analysis of TMB prediction. AUCs (A, C, E, G) and violin plots of predicted score from multiscale model (B, D, F, H) for patients stratified by the statuses of DDR, TP53, KRAS, and EGFR (mutant vs WT). AUC, area under curve; TMB, tumor mutational burden. TMB-H, TMB high; TMB-L, TMB low; DDR, DNA damage and repair; WT, wild type.
Supplemental table

**Table S1** The prevalence of TMB-H by cancer types.

| Cancer type | TCGA cohort | CPTAC cohort |
|-------------|-------------|--------------|
|             | Total number | TMB-H | Percentage | Total number | TMB-H | Percentage |
| COAD        | 386         | 70    | 0.18       | 137         | 55    | 0.40       |
| STAD        | 391         | 76    | 0.19       | 107         | 46    | 0.43       |
| LUAD        | 520         | 169   | 0.33       | 98          | 75    | 0.77       |
| BLCA        | 450         | 123   | 0.27       | 110         | 37    | 0.34       |
| HNSC        | 446         | 60    | 0.13       | 93          | 32    | 0.34       |
| LUSC        | 480         | 170   | 0.35       | All         | 545   | 245        | 0.45       |
| UCEC        | 555         | 193   | 0.35       |             |       |            |
| All         | 3228        | 861   | 0.27       |             |       |            |
Supplemental figures

**Figure S1** Identification of the cut-off value in the independent validation cohort. Correction analysis of the total mutation count and tumor mutational burden (a). When the cut-off value of tumor mutational burden was set as 10 mutations per Mb, the relative total mutation count value was found to be 279 for the same percentage of TMB-high (b). $R$, related coefficient.
Figure S2 Stratification analysis of PC-TMB. (A) AUC of predicted score from PC-TMB for patients stratified by tumor grade. (B) AUC of predicted score from PC-TMB for patients stratified by tumor stage. TMB, tumor mutational burden; AUC, area under curve; CI, confidence interval.