# Deep learning approach to predict tumor mutation burden (TMB) and delineate its spatial heterogeneity from whole slide images

Hongming Xu[1], Sunho Park[1], Jean René Clemenceau[1], Nathan Radakovich[1], Sung Hak Lee[2*] & Tae Hyun Hwang[1*]

[1]*Department of Quantitative Health Sciences, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH 44195, USA.*

[2]*Department of Hospital Pathology, Seoul St.Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, 06591 South Korea.*

[*] *To whom correspondence should be addressed (emails): hakjjang@catholic.ac.kr or hwangt@ccf.org*

## Abstract

**Purpose: Tumor Mutation Burden (TMB) is a potential genomic biomarker that could help to identify patients benefiting from immunotherapy across many cancers. Various tissue-based sequencing approaches have been widely used to determine the TMB status. However, the clinical utility of these approaches is often limited, due to time, cost, and tissue availability constraints. These methods could also provide inconsistent TMB status driven by spatial intratumor heterogeneity. Hematoxylin and Eosin (H&E) stained whole slide images (WSI) are routinely used for cancer diagnosis, thus mostly available for cancer patients. We present**

**a deep learning based computational pipeline using WSIs for predicting patient-level TMB status and quantifying its spatial heterogeneity within tumor regions.**

**Experimental Design: In an experiment to predict patient-level TMB status, we used The Cancer Genome Atlas (TCGA) Urothelial Bladder Carcinoma (BLCA) and Lung Adenocarcinoma (LUAD) cohorts. To investigate spatial heterogeneity of TMB status within a tumor and its prognostic utility, we used TCGA BLCA cohort.**

**Results: In an experiment of patient-level TMB predictions, our proposed method achieved the Area Under ROC (AUROC) scores of 0.752 (95% CI, 0.683-0.802) and 0.742 (95%, 0.682-0.794) for TCGA BLCA and LUAD cohorts, respectively, which are better compared to those of state-of-the-art methods. In another experiment to investigate spatial heterogeneity of TMB per patient, we predicted TMB status for each tile in each WSI for patients from TCGA BLCA cohort. We calculated entropy of TMB prediction probabilities in the WSI to determine whether the patient has high or low spatial TMB heterogeneity within the tumor. Kaplan Meier (KM) analysis showed that incorporating spatial heterogeneity of TMB information with patient-level TMB status based on WSIs could improve identification of patient subgroups with distinct survival outcome (a log rank test $P<0.05$).**

**Conclusions: Our proposed deep learning based approach using WSIs can predict patient-level TMB status with good accuracy, sensitivity and specificity compared to state of the art methods. The spatial analysis of TMB heterogeneity could provide a prognostic utility to better select patient subgroups.**

**Code Availability: https://github.com/hwanglab/tcga_tmb_prediction**

# 1 Introduction

Tumor mutational burden (TMB) is a quantitative genomic biomarker that measures the number of mutations within a tumor. TMB level has been shown to be associated with better prognosis and clinical responses to immune-checkpoint inhibitors in various cancer types such as melanoma, lung cancer and bladder cancer [2, 19]. Higher TMB levels are correlated with higher levels of neoantigens which could help the immune system to recognize tumors [1]. Recent studies reported that patients with high TMB status had favorable response to immunotherapy in many cancer types including bladder cancer [15, 18, 19, 32]

Tissue-based DNA sequencing (e.g. Whole exome sequencing (WES), targeted sequencing, etc.) is widely used to assess TMB status. However, due to the limited tissue availability, high costs and time-consuming procedures, the clinical utility of the tissue-based DNA sequencing is limited. Although the blood-based TMB measurement (e.g. liquid biopsies) has recently become available, this approach poses similar challenges to tissue-based approaches to accurately measure TMB [3]. In addition, because of spatial intratumor heterogeneity (ITH) present in a tumor, these approaches often provide inconsistent TMB results [30]. Therefore, there is an urgent unmet need to develop a method to accurately and cost-effectively predict TMB status while addressing the spatial heterogeneity of TMB status within the tumor.

The use of widely available histopathological images poses a promising alternative. Routine histopathological examination is the gold standard for diagnosis and grading of various cancer types in a clinical setting. Recent studies showed that deep learning models that utilize whole

slide images (WSIs) could accurately predict genetic variations present in a tumor. Schaumberg *et al.* (2018) [10] proposed a quantitative model to predict SPOP mutation state in prostate cancer using Hematoxylin and Eosin (H&E) stained WSIs. Their technique first determines a cohort of dominant tumor tiles based on tumor nuclei densities in WSI. Ensembles of residual networks [11] are then trained and integrated to predict SPOP mutation state. Coudray *et al.*,(2018) [12] trained the inception-v3 deep learning model [13] on lung adenocarcinoma (LUAD) WSIs to predict the ten most commonly mutated genes in LUAD. They reported that six of those ten commonly mutated genes, such as STK11, EGFR and FAT1, are predictable from pathology images by using deep learning models. Fu *et al.*, (2019) [24] performed a pan-cancer computational histopathology study, which showed that histological imaging features had significant correlations with a number of driver gene mutations across different cancer types. Kather *et al.*, (2019) [25] performed another pan-cancer computational pathology study, which further evidenced that many gene mutations are predictable from pathology slides with deep transfer learning methods. In addition, Kather et.al., (2019) [28] showed that deep learning models based on WSIs could predict microsatellite instability to guide immunotherapy for patients who are not eligible for genetic or immunohistochemical tests.

Given these studies showing that computational approaches making use of morphological features present in WSIs could reliably predict genetic characterizations present in tumors, we hypothesize that a carefully designed image-based computational method could predict TMB status given a WSI from a patient.

In this work, we develop and evaluate deep-learning based computational pipelines to predict patient and tumor tile-level (i.e., a small region within a WSI) TMB status, and then use the tile-level TMB status to delineate spatial heterogeneity of TMB within WSIs. Our proposed pipeline consists of four modules: 1) tumor detection, 2) representative tile selection, 3) feature extraction from selected tiles using transfer learning, and 4) TMB classification using support vector machines (SVMs) based on image features. To the best of our knowledge, this is the first work to predict bladder cancer patient TMB status, and interrogate TMB spatial heterogeneity and its prognositc utility using WSIs.

In the experiments with TCGA Urothelial Bladder Carcinoma (BLCA) and LUAD cohorts, we first evaluated the performance of our proposed method against state-of-the-art methods, including deep learning and Multiple Instance Learning (MIL) methods, using WSIs to predict patient-level TMB status. Then we applied our proposed model to predict TMB status at the tile level within WSIs for BLCA cohort and applied entropy measurement to evaluate spatial heterogeneity of TMB within WSIs. Identification of patient subgroups based on patient-level TMB status and TMB spatial heterogeneity status indicated that incorporating spatial heterogeneity of TMB could lead better patient stratification.

## 2 Materials and Methods

**Datasets.** A cohort of 386 TCGA BLCA patients (and corresponding clinical information) with 457 diagnostic H&E stained WSIs was downloaded from the TCGA data portal. We selected the

first diagnostic slide image (i.e., with DX1 suffix) if there are multiple diagnostic slide images available for a patient. Based on the percentile of total number single nucleotide variants [2], 386 TCGA BLCA patients were categorized into 3 groups: 128 low, 128 intermediate and 130 high TMB patients. After excluding one high and four low TMB patients, due to severe pen marks on slides, 124 low and 129 high TMB patients were used to train and test a model to predict patient-level TMB high and low status.

Another cohort of 478 TCGA LUAD patients with 541 diagnostic H&E stained WSIs was downloaded from TCGA data portal. In a similar way with TCGA BLCA cohort, TCGA LUAD patients were categorized into 3 TMB levels: 158 low, 159 intermediate and 161 high TMB patients. Due to severe pen markers on slides, 18 low and 4 high TMB patients were excluded from the analysis. Finally, 140 low and 157 high TMB patients were used to train and test a model to predict patient-level TMB high and low status in the LUAD cohort.

**Methods.** We developed a deep learning-based computational pipeline using WSIs to predict patient-level TMB status and delineate spatial heterogeneity of TMB present in tumors. The aim of our approach is to accurately predict and incorporate spatial TMB heterogeneity with patient-level TMB status to identify patient subgroups that could lead to better patient stratification. An illustration of the proposed computational pipeline is provided in Fig. 1.

(1) Tumor Detection. We trained a light-weight convolutional neural network (CNN) model with only about 0.28M trainable parameters to detect tumor regions in the WSI. Given an input image tile ($512\times512$ pixels), the CNN-based tumor detector outputs the probability of belonging to

cancer regions. The prediction map corresponding to the WSI is generated by stitching predicted probabilities for all image tiles. An empirical threshold (e.g., 0.5) is applied on the prediction map to obtain tumor regions. Fig. 2(a) illustrates an example of cancer detection on a WSI. All prediction maps of tumor regions were manually inspected by a pathologist. More details about our trained CNN-based tumor detector are provided in Fig.s1. Results are provided in Fig.s5, s6 and Table s1 in supplementary results.

(2) Representative Tile Selection. To improve computational efficiency of the method to analyze a large size of predicted tumor regions, we selected a subset of representative regions from all predicted tumor regions. We first divided predicted tumor regions into a set of non-overlapping tiles (128×128 pixels) at 2.5× magnification. We then characterized each tumor tile by a 42 dimensional feature vector (i.e., 40 multi-scale local binary pattern features [14] and 2D location of the tumor tile). After that, affinity Propagation (AP) clustering [16] was applied to identify tumor regions containing tiles with similar morphological patterns [31]. The AP clustering simultaneously identified a number of $r$ local tumor regions and their representative tiles $R_j$, where $1 \leq j \leq r$. Figs. 2(b)(c) illustrates AP clustering of tumor tiles on a WSI, where tumor tiles belonging to different clusters are indicated by different color of blocks in the image. Note that there are 56 ($r$ = 56 for this example) representative tiles selected among 490 tumor tiles for the patient slide shown in Fig. 2(b). More details about representative tile selection are described in the supplementary method section 2.

(3) Feature Extraction. We used transfer learning on pre-trained deep learning models to

generate features for selected representative tumor tiles. First, to suppress the influence of color variations, a color deconvolution based method [17] is utilized to normalize tumor tiles into a standard color appearance. Second, Transfer learning on pre-trained Xception [20] model was used to extract features from selected tumor tiles. Given an input tumor tile $R_j$ at $20\times$ magnification ($1024\times1024$ pixels), the transfer learning model outputs a high-level feature representation $V_j$ which is a 2048 dimensional vector. Finally, the feature vector $\overline{V}$ representing the WSI was obtained by integrating features of representative tumor tiles together, i.e., $\bar{V} = \sum_{j=1}^{r} \rho_j V_j$, where $\rho_j = \lambda_j \Big/ \sum_{j=1}^{r} \lambda_j$ and $\lambda_j$ represents the number of tumor tiles belonging to the jth cluster. The feature vector $\bar{V}$ is the weighted mean of features extracted from representative tiles, where each representative tile stands for the major characteristics of tumor tiles within the cluster. See details about feature extraction in the supplementary method section 3.

(4) TMB classification. We trained the Support Vector Machine (SVM) classifier based on features generated from the transfer learning model to predict patient-level TMB status. First, principal component analysis (PCA) was used to reduce the feature dimension to prevent overfitting. Second, feature standardization was performed on each feature component, which ensured its values have zero mean and unit variance. Finally, SVM with radial basis function (RBF) and linear kernels using default parameters were trained to predict patient-level TMB status. See details about TMB classification in the supplementary method section 4.

## 3 Results

In experiments using TCGA BLCA and LUAD cohorts, we first evaluated the performance of our proposed method and baseline methods to predict patient-level TMB status using WSIs. To investigate the prognostic utility of spatial heterogeneity of TMB status within a tumor, we used TCGA BLCA cohort to measure TMB heterogeneity and incorporated spatial TMB heterogeneity status to identify patient subgroups.

**Evaluation on patient-level TMB Prediction for TCGA BLCA and LUAD cohorts.** We first investigated whether the use of either tumor detection, representative tile selection, or color normalization module as well as different transfer learning models could impact the performance of patient-level TMB prediction. Using TCGA BLCA dataset, we ran patient-level TMB prediction experiments by excluding tumor detection (abbreviated as P-E-TD), representative tile selection (abbreviated as P-E-RTS), or color normalization (abbreviated as P-E-CN). We also tested two well-known transfer learning models, Inception-v3 (abbreviated as P-InceptionV3) [13] and Resnet50 (abbreviated as P-Resnet50) [11], in addition to Xception model (abbreviated as P-Xception), to evaluate whether different transfer learning models could impact patient-level TMB prediction performance. After performing principal component analysis to reduce the number of features generated by transfer learning models, we selected the top 100 principal components to train SVM classifiers with linear or RBF kernels to predict patient-level TMB status using leave-one-out cross validation. Fig. 3(a)(b) shows ROC curves of patient-level TMB prediction results using different settings of our pipeline, that is Linear SVM and RBF SVM, respectively (see more

details in Table s2). The P-E-RTS model, which exhaustively uses all tiles within a WSI without selecting representative tiles, and the P-Xception model, which uses the selected representative tiles, showed overall best AUROC values, achieving 0.769 and 0.748 for linear SVM, and 0.753 and 0.752 for RBF SVM, respectively. While both approaches showed good prediction performance, the P-Xception model used the 11,164 selected representative tiles out of 125,358 tiles, which requires significant less computational time (see computational comparison example in Table s3). This indicates that the use of AP clustering to select a set of representative tiles from a WSI could bring computational efficiency and therefore we used the AP clustering module in our proposed pipeline for further experiments. The patient-level TMB prediction performance using Xception model (P-Xception) is much better than those of Inception-v3 (P-InceptionV3) and Resnet50 (P-Resnet50), thus we used Xception model as the transfer learning algorithm in our pipeline for the rest of experiments.

To compare the performance of patient-level TMB prediction of our proposed method and the state-of-the-art methods, we trained our designed CNN model (see Fig.s1 in supplementary methods), VGG16-TL2 [26] and Resnet18 [28], and MIL based deep learning algorithm [29] as baseline models. To train these deep learning models, tumor tiles of each WSI were assigned the same label (e.g., TMB high or low status) as the corresponding patient-level TMB status. The final patient-level TMB prediction was obtained by averaging prediction probabilities of all tumor tiles. In addition, we also extracted local binary pattern (LBP) texture features from representative tumor tiles and made predictions using an SVM classifier with RBF kernel as the baseline. Three-fold cross validation was applied to evaluate baseline deep leaning models, due to computational com-

plexity, and the leave-one-out cross validation was used to evaluate the rest of the methods. Table 1 shows patient-level TMB prediction results in terms of accuracy ($ACC$), specificity ($SPE$), sensitivity ($SEN$) and AUROC values for our proposed method and baseline models. Fig. 3(c) and (d) shows patient-level TMB prediction performance in TCGA BLCA and LUAD, respectively. Overall, the proposed pipeline provides better performance over baseline methods, which achieves from 2% to 5% improvements with respect to AUROC values. Taken together, these results indicate the efficacy of the proposed method to predict patient-level TMB status using WSIs.

**Spatial heterogeneity of TMB status within a tumor and its prognostic utility to identify patient subgroups with distinct overall survival outcome in TCGA BLCA.** We investigated the spatial heterogeneity of TMB status by predicting TMB status of each representative tumor tile within a WSI. We used our proposed pipeline to predict TMB status on selected representative tiles from tumor regions and then assigned predicted TMB status for each representative tile to corresponding tumor regions. Specifically, instead of integrating a set of selected representative tiles from the WSI to predict a patient-level TMB, we used the trained SVM with Linear kernel to predicted TMB status of each selected representative tile within a WSI. Corresponding tumor regions of the representative tumor tiles were assigned with the same TMB prediction status as the predicted TMB status of the representative tumor tile. To determine spatial TMB heterogeneity status (i.e., high or low spatial heterogeneity), we calculated the Shannon entropy [23] of predicted TMB levels of tumor regions within the WSI, i.e., $S = -\sum_k P_k \log_2 (P_k)$, where $P_k$ is the ratio between the number of the $kth$ unique TMB prediction probability and the total number of tumor tiles within the WSI. High entropy value indicates high spatial TMB heterogeneity (e.g., mixture

of predicted TMB high and low regions), while low entropy value indicates low spatial TMB heterogeneity within a tumor (e.g., either TMB high or low status across most of tumor regions within WSIs). High or low entropy status was determined by using the median entropy value from all patients from TCGA BLCA cohort as the threshold. Fig. 4 shows visualization of spatial TMB heterogeneity heatmaps based on tile-level TMB prediction, where red and blue colors indicate predicted TMB high and low status probability, respectively. Fig. 4(a) shows a heatmap of spatial TMB status of TMB high patient based on Whole Exome Sequencing (WES) data. Our WSI-based method correctly predicted the patient-level TMB status as TMB high for this patient. Tile-level TMB prediction and the entropy measurement of the predicted TMB levels of tumor tiles within the WSI showed low spatial TMB heterogeneity. Specifically, the heatmap showed that most tumor regions within the WSI presented TMB high status, while few tumor regions presented TMB low status. Similarly, Fig. 4(b) showed that our WSI-based method correctly predicted the patient level TMB status as TMB low and low spatial TMB heterogeneity for the WSI-based TMB low patient. Fig. 4(c) and (d) showed that while WSI-based patient level TMB status of these two patients agreed with WES-based patient level TMB status, there are mixtures of TMB high and low status within tumor regions. Higher entropy values based on tile-level TMB status indicates higher degree of spatial TMB heterogeneity within WSIs.

To investigate the prognostic utility of spatial TMB heterogeneity status, we used spatial TMB heterogeneity status to select patient subgroups. In experiments using TCGA BLCA cohort, we predicted patient-level TMB high and low status for 368 patients using our proposed WSI-based method. In each patient, we assigned low or high spatial TMB heterogeneity status based

on entropy values using tile-level TMB prediction of tumor regions within WSIs for all patients. We assigned patients with predicted patient-level TMB-high and low spatial TMB heterogeneity into one subgroup and the rest of patients as another subgroup. Then, we generated a Kaplan Meier (KM) plot using overall survival (OS). Fig. 5(a) shows a KM plot for two TMB subgroups indicating that two subgroups have statistically significant OS difference using log-rank test (P = 0.016). In univariate analysis using Chi-square test, the TMB subtypes correlated significantly with differences in tumor stage (P = 0.024), but not age (Age>60 vs others, P = 0.872), sex (P = 0.086), Lymphovascular invasion (P = 0.064) and Inflammatory Infiltrate Response (P = 0.428) (see supplementary Table s4). The patients in patient-level TMB high with low spatial heterogeneity subgroup had more advanced tumor stage. The TMB subtypes did not significantly correlate with known molecular subtypes determined by Reverse Phase Protein Array (RPPA) (P = 0.761) and mRNA subtypes (P = 0.942) from TCGA BLCA study. Multivariable Cox proportional-hazard analyses of cancer stage and the TMB subtypes in relation to the risk of death showed that the TMB subtypes remained statistically significantly correlated with survival. The Hazard Ratio (HR) of the TMB variable is 1.796 (95% CI: 1.18-2.73, P = 0.006), which indicates that the hazard for patients belonging to the group of TMB others is about 1.8 times higher than patients belonging to the group of TMB high & Low spatial TMB heterogeneity (see supplementary Table s5). We also assigned patients into two subgroups as WSI-based, patient-level TMB high or low, without considering spatial TMB heterogeneity. While the subgroup with predicted patient-level TMB-high status tended to have better OS compared to the subgroup with predicted patient-level TMB-low patient subgroup, the log rank test did not show a statistically significant difference (P = 0.072, see

Fig.s7 in supplementary results). Finally, to investigate whether incorporating WSI-based patient-level TMB and spatial TMB heterogeneity with tissue-based TMB testing could improve patient stratification, we divided 126 WES-based TMB-high patients (with tile-level predictions) into two subgroups: 1) WSI-based patient-level TMB-high and low spatial TMB heterogeneity patient subgroup and 2) the rest of WEX-based TMB-high patient subgroup, respectively. Fig. 5(b) showed that WES-based TMB high & WSI-based patient-level TMB high and low spatial TMB heterogeneity patient subgroup have better OS compared to the other subgroup (log rank test P = 0.018). Taken together, these results indicate that incorporating WSI-based patient-level TMB status with spatial TMB heterogeneity information could lead to better patient stratification.

## 4 Discussion

Intratumor heterogeneity is one of key mechanisms driving disease progression and resistance to therapies [5, 21]. Multi-regional tissue-based sequencing from a tumor has shown spatial heterogeneity of mutational signature, mutational burden, T-cell receptor repertoire, etc. [3,4,6,7] and its implication for treatment strategy [8]. While tissue-based sequencing from multiple regions could provide landscape of spatial heterogeneity, it is practically challenging to generate such data, due to high costs, tissue availability, etc.. In this study, we present the computational pipeline based on WSIs to predict patient-level TMB status and investigate spatial heterogeneity of TMB within tumors. We showed that our proposed computational pipeline could achieve overall best performance to predict patient-level TMB compared to other state of the art methods. We also showed that measuring and incorporating spatial heterogeneity of TMB status with patient-level TMB status based

on WSIs or combined with WES-based TMB status could lead to better patient stratification with distinct overall survival outcomes. In particular, patient-level TMB high with low spatial heterogeneity of TMB status was correlated with better overall survival. Visual inspection of selected tumor tiles from WSIs by our pathologist indicates that predicted TMB-high representative tumor tiles from patient-level TMB high WSIs are more enriched with Tumor Infiltrated Lymphocytes (TILs) while showing more high grade tumors (see supplementary Table s6 and Fig.s9). This is consistent with the univariate analysis of TMB subtypes showing a higher portion of high grade tumors in patient-level high TMB and low heterogeneity tumors. Although we observe an enrichment of high graded tumors in this TMB subgroup, the higher presence of TILs in tumors from this subgroup might be one of reasons why this subgroup has better prognosis. To the best of our knowledge, this is the first study to predict spatial TMB heterogeneity status and study its prognostic utility for patient stratification.

There are several limitations and challenges in our study. Due to the limited access of data cohorts, the evaluation of the proposed method and baselines to predict patient-level TMB was limited to TCGA BLCA and LUAD datasets without additional validation cohorts. While we showed overall better performance to predict patient-level TMB status compared with baseline methods, independent cohorts from multiple institutes are needed to evaluate its generalizability. In addition, our evaluations indicated that various deep learning-based prediction models, including end-to-end deep learning models, to predict patient-level TMB status did not show superior performance. Larger and more well-annotated WSI datasets would be needed to better train and improve the performance of deep learning-based prediction models (and thus our computational pipeline too,

since we employ deep learning-based transfer learning models). Finally, we showed that incorporating WSI-based, patient-level TMB status with spatial TMB heterogeneity status could improve patient stratification in TCGA BLCA cohort, thus, it can potentially be used to select patients likely benefit to immunotherapy. However, we could not access WSI datasets from patients treated with immunotherapy to test its utility as a predictive biomarker. Further retrospective and/or prospective studies should be performed to evaluate the utility of WSI-based, patient-level TMB status with spatial TMB heterogeneity status as a predictive biomarker to select patients likely to respond to immunotherapy. It is worth noting that in experiments using TCGA LUAD corhort, WSI-based patient-level TMB status with spatial TMB heterogeneity status was not found to be correlated with OS (see Fig.s8 in supplementary results). However, WES-based TMB status was not significantly correlated with OS either. Recent study integrating multiregion exome and RNA-sequencing (RNA-seq) data with spatial histology to investigate spatial tumor and immune microenvironment (TIME) in LUAD showed that LUAD subgroup with immune cold and low neoantigen burden was significantly correlated with poorer disease free survival [9]. This may indicate that integrating TIME and spatial TMB heterogeneity could lead to improvement to stratify patients in a certain type of cancers.

In summary, this study demonstrates the feasibility of predicting patient-level TMB status and delineating spatial heterogeneity of TMB by using computational models based on histological images. Our spatial TMB heterogeneity analysis shows that patients with more homogeneous TMB high status across regions present better prognosis in bladder cancer. By combining tissue-based TMB high status with image-based TMB high and low spatial TMB heterogeneity status

could further improve patient stratification in bladder cancer. Taken together, integrating image-based TMB prediction and the degree of spatial TMB heterogeneity on WSI yields better patient stratification than tissue-based TMB alone and represents a novel prognostic biomarker. Our computational pipeline is a general model applicable to different types of tumors thus could pave new opportunities to develop rapid and cost-effective biomarkers based on WSIs. Future studies would be required to further refine the presented technique and validate it in prospective patient cohorts with bladder cancer or other cancer types.

## 5 References

1. Brown, S. D. et al. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome research 24.5, (2014): 743-750.

2. Robertson, A. G. et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell, 171.3, (2017): 540-556.

3. Zhang, Yaxiong, et al. "The correlations of tumor mutational burden among single-region tissue, multi-region tissues and blood in non-small cell lung cancer." Journal for immunotherapy of cancer 7.1 (2019): 1-5.

4. Hu, Xin, et al. "Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma." Nature communications 10.1 (2019): 2978.

5. Marusyk, Andriy, Michalina Janiszewska, and Kornelia Polyak. "Intratumor heterogeneity: The rosetta stone of therapy resistance." Cancer cell 37.4 (2020): 471-484.

6. Jamal-Hanjani, Mariam, et al. "Tracking the evolution of nonsmall-cell lung cancer." New England Journal of Medicine 376.22 (2017): 2109-2121.

7. Joshi, Kroopa, et al. "Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer." Nature medicine 25.10 (2019): 1549-1559.

8. Stanta, Giorgio, and Serena Bonin. "Overview on clinical relevance of intra-tumor heterogeneity." Frontiers in medicine 5 (2018): 85.

9. AbdulJabbar, Khalid, et al. "Geospatial immune variability illuminates differential evolution of lung adenocarcinoma." Nature Medicine (2020): 1-9.

10. Schaumberg, A. J. et al. H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer. BioRxiv, (2018): 064279.

11. He, K. et al. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, (2016): 770-778.

12. Coudray, N. et al. Classification and mutation prediction from nonsmall cell lung cancer histopathology images using deep learning. Nature medicine, 24.10, (2018): 1559-1567..

13. Szegedy, C. et al. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, (2016): 2818-2826.

14. Ojala, T. et al. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24.7, (2002): 971-987.

15. Chan, Timothy A., et al. "Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic." Annals of Oncology 30.1 (2019): 44-56.

16. Frey, B. J., & Dueck, D. Clustering by passing messages between data points. Science, 315.5814, (2007): 972-976.

17. Macenko, M. et al. A method for normalizing histology slides for quantitative analysis. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), (2009): 1107-1110.

18. Bandini, Marco, et al. "Predicting the pathologic complete response after neoadjuvant pembrolizumab in muscle-invasive bladder cancer." JNCI: Journal of the National Cancer Institute (2020).

19. Necchi, Andrea, et al. "Updated results of PURE-01 with preliminary activity of neoadjuvant pembrolizumab in patients with muscle-invasive bladder carcinoma with variant histologies." European urology 77.4 (2020): 439-446.

20. Chollet, F. Xception: Deep learning with depthwise separable convolutions. arXiv preprint, (2017): 1610-02357.

21. Failmezger, Henrik, et al. "Topological Tumor Graphs: a graph-based spatial model to infer stromal recruitment for immunosuppression in melanoma histology." Cancer Research 80.5 (2020): 1199-1209.

22. Courtiol, Pierre, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nature medicine, 25.10 (2019): 1519-1525..

23. Jackson, Hartland W., et al. "The single-cell pathology landscape of breast cancer." Nature 578.7796 (2020): 615-620.

24. Fu, Yu, et al. "Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis." bioRxiv (2019): 813543.

25. Kather, Jakob Nikolas, et al. "Pan-cancer image-based detection of clinically actionable genetic alterations." bioRxiv (2019): 833756.

26. Xu, H., et al. Computerized Classification of Prostate Cancer Gleason Scores from Whole Slide Images. IEEE/ACM transactions on computational biology and bioinformatics, (2019).

27. Fabrizio, Federico Pio, et al. Gene code CD274/PD-L1: from molecular basis toward cancer immunotherapy. Therapeutic advances in medical oncology 10 (2018): 1758835918815598

28. Kather, Jakob Nikolas, et al. "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer." Nature medicine 25.7 (2019): 1054-1056.

29. Campanella, Gabriele, et al. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images." Nature medicine 25.8 (2019): 1301-1309.

30. Jia, Qingzhu, et al. "Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer." Nature communications 9.1 (2018): 1-10.

31. Saltz, Joel, et al. "Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images." Cell reports 23.1 (2018): 181-193.

32. Song, Bic-Na, et al. "Identification of an immunotherapy-responsive molecular subtype of bladder cancer." EBioMedicine 50 (2019): 238-245.

**Acknowledgements** Put acknowledgements here.

**Competing Interests** The authors declare that they have no competing financial interests.

**Correspondence** Correspondence and requests for materials should be addressed to Dr.Hwang (email: hwangt@ccf.org).

## List of Figures

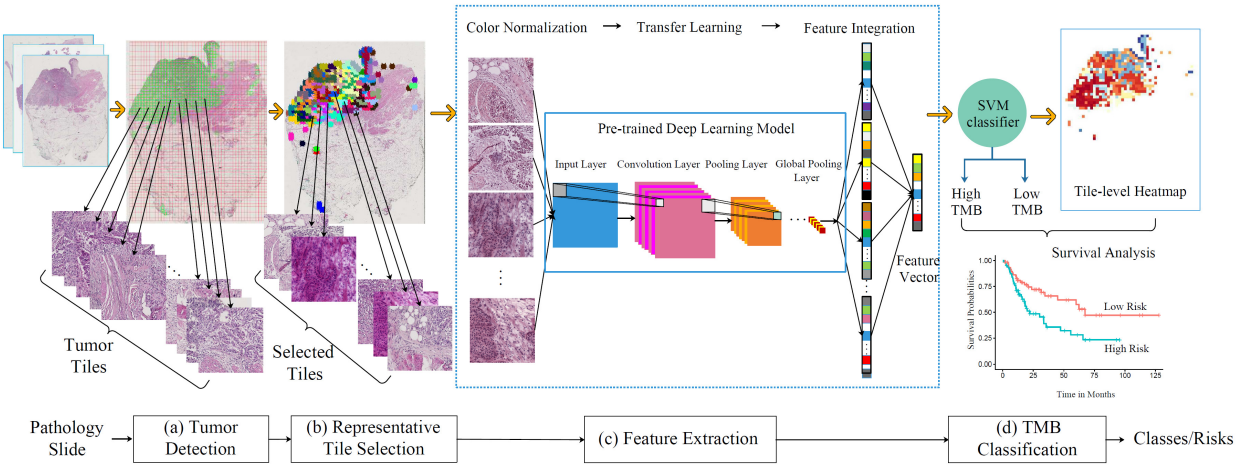



Figure 1: Pipeline of the presented technique.

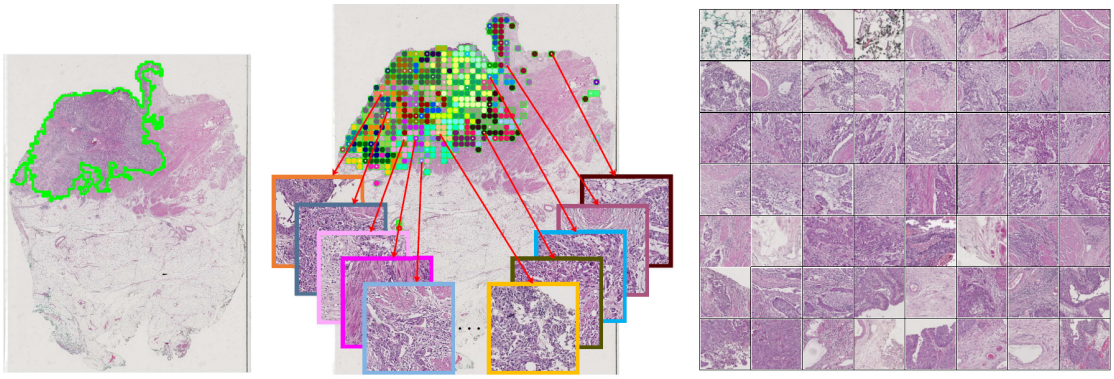



Figure 2: Illustration of tumor detection and AP clustering. (a) Tumor detection result (overlapped green contours). (b) Example of AP clustering on tumor tiles, where tumor tiles belonging to different clusters are indicated by different color of blocks in the image. Several representative tumor tiles indicated by arrows are zoomed-in for better viewing. (c) 56 representative tumor tiles selected by AP clustering for the slide shown in (b).

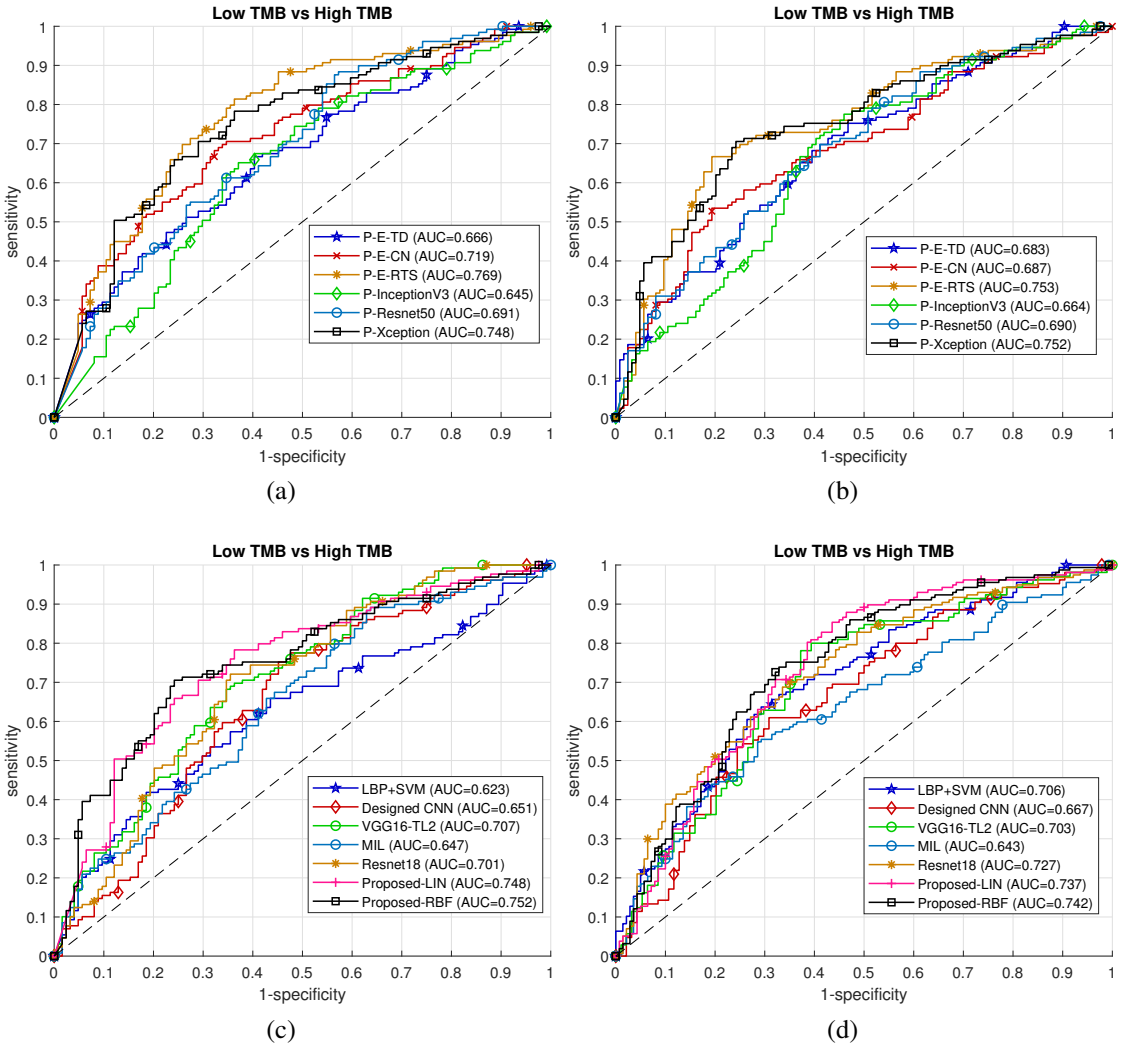



Figure 3: Evaluations on TMB prediction. Ablation study of our method on TCGA BLCA TMB prediction: (a) using SVM with Linear kernel, (b) using SVM with RBF kernel. (c) Baseline comparisons of TCGA BLCA patient-level TMB predictions. (d) Baseline comparisons of TCGA LUAD patient-level TMB predictions. Note that in (c)(d) Proposed-LIN and Proposed-RBF represent the proposed technique using Linear SVM and RBF SVM, respectively.

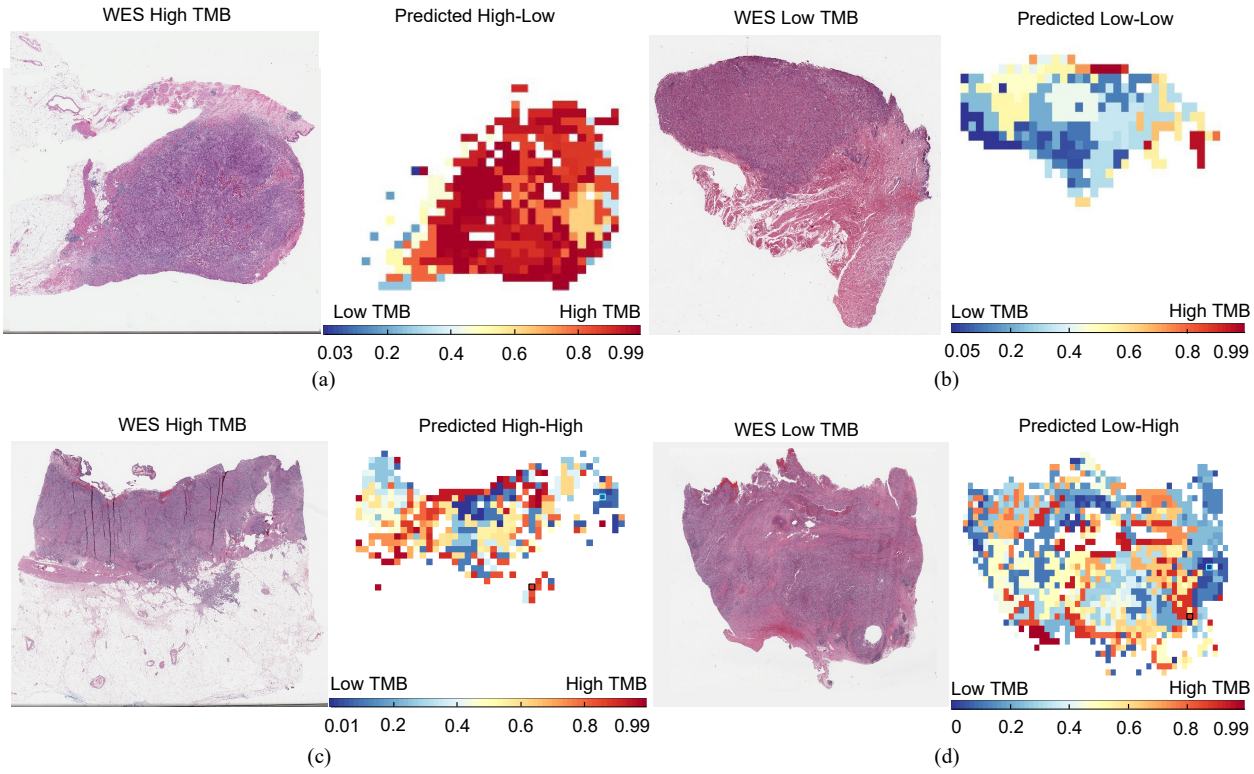
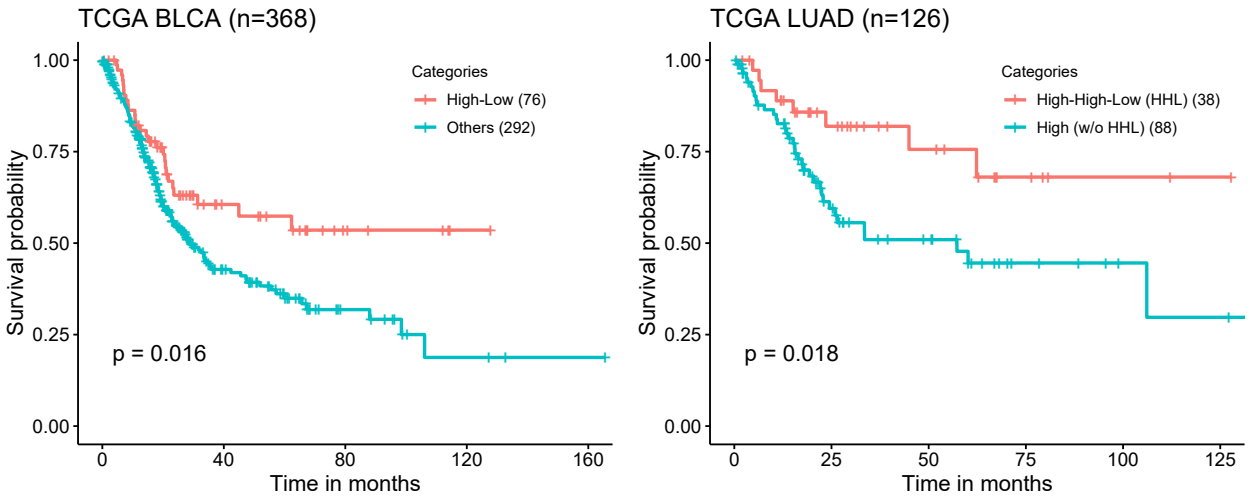



Figure 4: Tile-level TMB prediction visualization. (a) Tissue-based TMB high patient (TCGA-XF-AAN2) was predicted as patient-level TMB high based on our WSI-based method. Tile-level TMB prediction and entropy measurement indicated low spatial TMB heterogeneity (Shannon entropy $S = 4.99$). (b) Tissue-based TMB low patient (TCGA-XF-A9SH) was predicted as patient-level TMB low and low spatial TMB heterogeneity based on our WSI-based method ($S = 4.60$). (c) Tissue-based TMB high patient (TCGA-DK-A3IT) was predicted as patient-level TMB high, while tile-level TMB prediction indicated high spatial TMB heterogeneity ($S = 5.30$). (d) Tissue-based TMB low patient (TCGA-FD-A3B7) was predicted as patient-level TMB low and tile-level high entropy ($S = 6.21$). The median cut-off value for entropy $S$ was 5.19. High entropy indicated high spatial TMB heterogeneity.




(a) (b)

Figure 5: WSI-based TMB subtypes (a) A Kaplan-Meier plot for overall survival according to WSI-based TMB high & low spatial TMB heterogeneity vs other subtypes. (b) A Kaplan-Meier plot for overall survival according to WSI-based TMB high & low spatial TMB heterogeneity vs other subtypes for WES-based TMB high patients.

## List of Tables

Table 1: Comparison of patient-level TMB prediction using different methods. In this table, Proposed-LIN uses SVM classifier with linear kernel, while Proposed-RBF uses SVM classifier with RBF kernel.

| Cohorts | Methods | ACC (%) | SPE (%) | SEN (%) | AUROC (95% CI) |
|---|---|---|---|---|---|
| TCGA-BLCA | LBP+SVM | 60.47 | 64.52 | 56.59 | 0.623 (0.550-0.689) |
| | Designed CNN | 61.66 | 62.10 | 61.24 | 0.651 (0.581-0.741) |
| | VGG16-TL2 [26] | 65.22 | 66.94 | 63.57 | 0.707 (0.639-0.766) |
| | MIL [29] | 58.89 | 58.87 | 58.91 | 0.647 (0.577-0.710) |
| | Resnet18 [28] | 66.80 | 65.32 | 68.22 | 0.701 (0.638-0.765) |
| | Proposed-LIN | 69.57 | 68.55 | **70.54** | 0.748 (0.683-0.802) |
| | Proposed-RBF | **73.12** | **75.81** | **70.54** | **0.752 (0.694-0.810)** |
| TCGA-LUAD | LBP+SVM | 66.67 | **70.00** | 63.69 | 0.706 (0.645-0.763) |
| | Designed CNN | 63.82 | 67.02 | 60.95 | 0.667 (0.583-0.741) |
| | VGG16-TL2 [26] | 69.85 | 62.77 | **76.19** | 0.703 (0.621-0.766) |
| | MIL [29] | 60.27 | 60.00 | 60.51 | 0.643 (0.578-0.698) |
| | Resnet18 [28] | 67.00 | 65.00 | 68.79 | 0.727 (0.666-0.779) |
| | Proposed-LIN | 69.02 | 62.14 | 75.16 | 0.737 (0.671-0.796) |
| | Proposed-RBF | **70.37** | 67.86 | 72.61 | **0.742 (0.682-0.794)** |