Review article

Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images

Jing Hu [a,1], Chuanliang Cui [b,1], Wenxian Yang [a], Lihong Huang [a], Rongshan Yu [a], Siyang Liu [c,d], Yan Kong [b,*]

[a] *Aginome-XMU Joint Lab, School of Informatics, Xiamen University, China*
[b] *Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Renal cancer and Melanoma, Peking University Cancer Hospital & Institute, Beijing, China*
[c] *Guangdong Lung Cancer Institute, Guangdong Provincial Key Laboratory of Translational Medicine in Lung Cancer, Guangdong Provincial People's Hospital & Guangdong Academy of Medical Sciences, Guangzhou, China*
[d] *School of Medicine, South China University of Technology, Guangzhou, China*

ARTICLE INFO

ABSTRACT

Background: Recent studies showed that immune-checkpoint blockade (ICB) has significantly improved clinical outcomes of melanoma and lung cancer patients. However, only a small subset of patients can benefit from ICB. Deep learning has been successfully implemented in complementary clinical diagnosis. The aim of this study is to demonstrate the potential of deep learning to facilitate the prediction of anti-PD-1 response from H&E images directly.

Methods: In this study, 190 H&E slides of melanoma were segmented into 256×256 tiles which were used as the training set for the convolutional neural network (CNN). Additional 54 melanoma and 55 lung cancer H&E slides were collected as independent testing sets.

Findings: An AUC of 0.778(95% CI: 63.8%-90.5%) was achieved for 54 melanoma testing samples with 15(65.2%) responders and 23(74.2%) non-responders correctly classified. We also obtained an AUC of 0.645(95% CI: 49.4%-78.4%) for 55 lung cancer samples.

Interpretation: To our knowledge, this is the first study of using deep learning to determine patients' anti-PD-1 response from H&E slides directly. Our CNN model achieved the state-of-the-art performance and has the potential to screen ICB beneficial patients in routine clinical practice.

Background

Melanoma is one of the most aggressive cancer types originating from melanocytes and tend to metastasize early [1]. Unlike white patients, the subtypes of melanoma most common in Asian patients are acral and mucosal, and they account for up to 58% of all melanoma tumors in that patient population [2]. With the development of cancer treatment during the past decades, immune-checkpoint blockade (ICB) such as anti-PD1 and anti-CTLA4 had demonstrated considerable clinical benefit for various types of cancer [3]. In the KEYNOTE-151 trial, the safety and efficacy of pembrolizumab in Chinese patients were first evaluated with advanced melanoma that progressed following first-line chemotherapy. The objective response rate (ORR) was very low, 16.7% (95% CI, 10%–25.3%), with 15.8% for acral and 13.3% for mucosal melanoma [4]. Furthermore, antitumor efficacy of ICB is also low in lung cancer, which is observed in 20–30% of patients with non-small-cell lung cancer (NSCLC), with most patients not response to ICB [5].

Past research has shown that tumor mutation burden (TMB), microsatellite instability (MSI) and PD-L1 expression may predict ICB response in some cancer types [6–8]. However, no biomarker is currently validated to predict resistance or benefit derived from immunotherapy with clinically significant accuracy. Therefore, the need for effective immunotherapy biomarkers, especially for Asian melanoma patients, is urgent.

In recent years, it was found that histopathological images can be used to provide accurate prediction of several immunotherapy biomarkers. Jakob et al. evaluated H&E slides for 315 samples of STAD, 360 FFPE samples of CRC and 378 snap-frozen samples of CRC from TCGA to classify MSI versus microsatellite stability (MSS) patients. The AUCs for MSI detection were 0.81, 0.84 and 0.77 respectively [9].
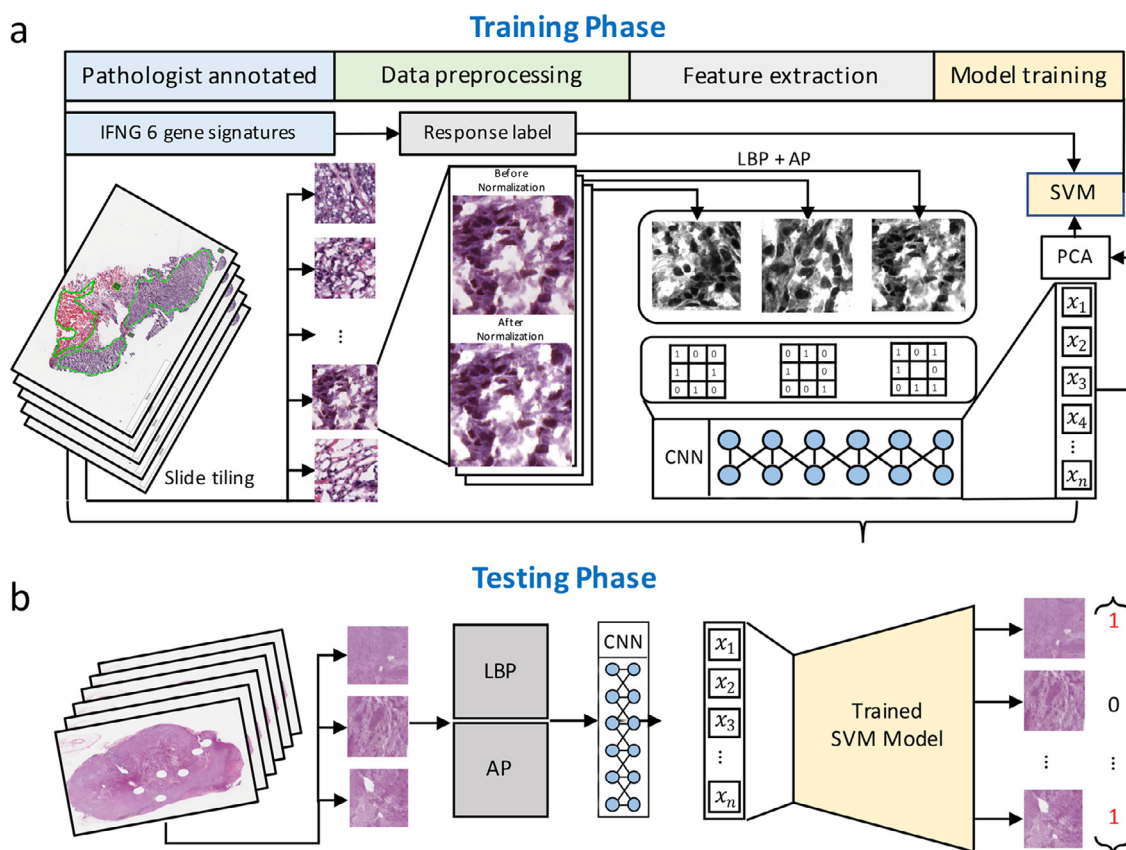
Fig. 1. anti-PD-1 response prediction by H&E histology images. Training phase of the deep learning model. Left, Tumor regions were annotated by two pathologists with green polygon border. Tumor regions were segmented and color normalized for downstream analysis. The multi-scale LBP and AP algorithms were applied on gray-scaled tiles. Right, features were extracted by transfer learning using the Xception model and reduced features were fed into SVM for final classification. Testing phase uses the trained model from the training phase to predict clinical outcomes of unseen samples.

Xu et al. investigated a cohort of 253 patients with bladder cancer from TCGA. Their method achieves an accuracy of 73% and AUC of 0.75 in distinguishing high and low TMB patients [10].

Although it is possible to derive patients' potential to benefit from ICB based on MSI or TMB status, as these biomarkers are only partially associated with ICB response, using them as a surrogate for ICB response prediction adds an intermediate step which decreases the prediction power of histopathological images and therefore constrains their clinical applications. Herein we investigated whether deep learning algorithms can predict anti-PD-1 response directly from H&E images.

In this study, we designed a systematic CNN model which can help determine the likelihood of responding to anti-PD-1 treatment for cancer patients. We tested our model on more than 100 melanoma and lung cancer samples. The model can be generalized to other tumors as well. This work represents the first finding of using transfer learning to determine immunotherapy response on H&E slide samples.

Methods

Study design and patient cohort

We collected a total of 476 patient whole-slide images from TCGA-SKCM database, where 190 patients with top and bottom 20% interferon-gamma (INFG) scores were selected as our training data. Mapping of tumor-infiltrating lymphocytes (TILs) was based on H&E images from 13 TCGA tumors ($n = 1896$) [11] downloaded from GDC.

From March 2016 and December 2017, fifty-four patients from Peking University Cancer Hospital, which were enrolled in four clinical studies of anti-PD-1 monoclonal antibody monotherapy for unresected stage III or stage IV (AJCC Cancer Staging Manual 8th ed.) melanoma ad-

vanced melanoma were chosen according to tumor response (responder vs. non-responders) . Responses were evaluated by investigators using RECIST version 1.1. A non-small-cell lung cancer (NSCLC) cohort from Guangdong Province Cancer Hospital between July 15, 2019 and October 16, 2019 was used as a second validation dataset ($n = 55$ patients). The responses were also evaluated by investigators using RECIST version 1.1. All histology slides were annotated by two board-certificated pathologists.

Image preprocessing and training of the model

Histopathological images are large in size and hard to handle by neural networks directly. Each image is split into small tiles with 256×256 pixels at $20 \times$ magnification using OpenSlidelibrary [12]. The tiles with low cell content information were dropped (>40% of the tiles are background). Then color was normalized by the Macenko's method [13] for each tile. Multi-textural features represented by 4 radius of scales ($r = 2,4,6,8$) LBPs [14] were transformed into 40-dimensional feature vectors. Affinity propagation (AP) algorithm [15], which does not require a predefined number of clusters, was performed subsequently to obtain tens of center tiles. To convert these center tiles to informative features, Xception [16] neural network with ImageNet pre-trained parameters is applied to extract 2048-dimensional feature vectors. The parameters in the neural network were fixed and the last fully connected layer was discarded. The feature vectors of center tiles were summed up, weighted by their corresponding proportion of its cluster. We chose principal component analysis (PCA) to further reduce the dimension of extracted features. The first 20 components are picked according to the percentage of variance explained by total selected components, and used as the input for the final classifier (Fig. 1). Using the reduced features
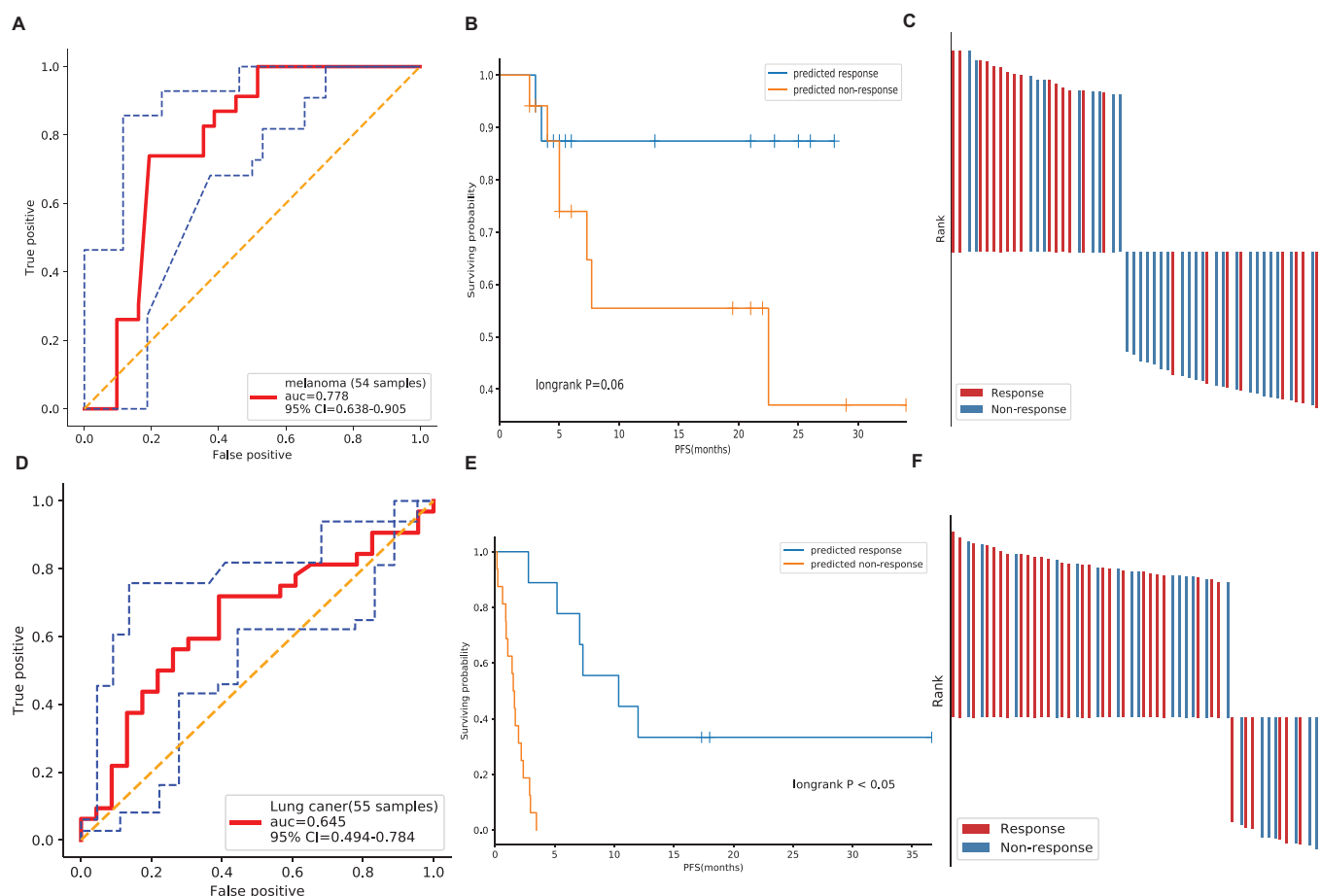
Fig. 2. Prediction performance on the validation datasets (A) Area under the curves (AUC) of melanoma testing set ($n = 54$). (B) Progression-free survival of patients separated by responders and non-responders in melanoma. (C) A waterfall plot of prediction probability score of melanoma samples. (D) AUC curves of lung cancer data set ($n = 55$). (E) Difference in progression-free survival of lung patients in responders and non-responders. (F) A waterfall plot of prediction probability score of lung cancer patients.

of training samples, we train an SVM classifier with Gaussian kernel to predict immunotherapy status. To evaluate the performance and avoid over-fitting, we performed 10-fold cross-validation in the training phase. Finally, $n = 54$ melanoma and $n = 55$ lung cancer patients were analyzed with the previously trained SVM model.

Comparison with TILs and other deep learning models

To calculate the percentage of TILs on each image, we constructed a TILs recognition deep learning model based on VGG-16 [17]. The slides were delineated into 100×100 non-overlapping tiles. The tiles were color normalized and used as training data for CNN model. We then predicted each tile as a TILs or non-TILs tile. The percent of TILs in the whole image was determined and used for calculation of AUC.

Six other deep learning models were evaluated to compare the prediction accuracies. AUCs for all models were calculated on the melanoma dataset only. The area under curve (AUC) was calculated by scikit-learn in Python. Confidence intervals (CIs) at 95% intervals were estimated by 1000 iterations using the bootstrap method.

Results

The ROC curves of testing melanoma and lung cancer datasets are shown in Fig. 2. An AUC of 0.778(95% CI: 63.8%–90.5%) for 54 melanoma testing samples with 15(65.2%) responders and 23(74.2%) non-responders correctly classified (Fig. 2A, C). According to our prediction results, responder group experienced much longer progression-free

survival (log-rank test, $p = 0.06$) compared to the non-responder group (Fig. 2B) indicating histology images coupled with deep learning is an effective ICB biomarker.

We also obtained an AUC of 0.645(95% CI: 49.4%–78.4%) on the lung cancer dataset, which suggests that the deep learning model can be generalized to other cancers (Fig. 2D, F). We observed that progression-free survival of responders was significantly extended compared to that of non-responders (Fig. 2E). It is worth to note that histopathology images from lung cancer patients are core-needle biopsy samples rather than surgery samples as in our training set. This may explain the slightly compromised performance of our model.

The AUC of using TILs for predicting anti-PD-1 response was only 0.58 in the melanoma dataset. This demonstrated the superiority of our deep learning model compared to conventional immune checkpoint biomarkers. Examples of lymphocyte infiltration in responders and non-responders are illustrated in Fig. 3. These results were consistent with T cell staining using IHC in these patients.

Besides, Pearson correlation value was calculated between TMB values from WES and predicted scores from CNN model for $n = 141$ TCGA SKCM samples. The correlation value is 0.05, which suggested that CNN-based model is a completely independent biomarker than molecular-derived signatures such as TMB (supplementary Fig. A).

Discussion

Immunotherapy has changed the landscape of oncology but determining patients who may benefit from the ICBs has remained a serious
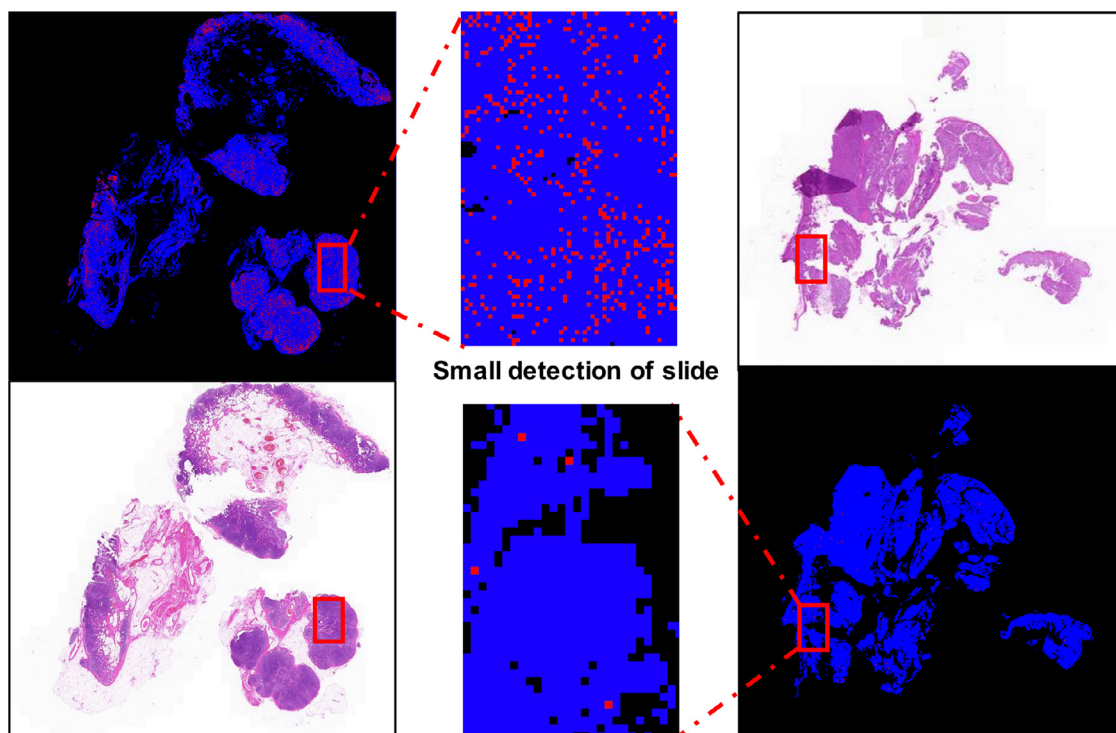
Fig. 3. Examples of TILs from whole slide images of responder and non-responder. Left, a responder example with TILs labeled as red points and tissue regions colored in blue on the masked figure. Right, a non-responder example with TILs labeled as red points and tissue regions colored in blue on the masked figure. Intermediate, randomly selected regions from each slide. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

challenge. Current biomarkers are not sufficiently effective to identify patients who may respond to ICBs. For example, 50% of colorectal cancers (CRCs) with MSI will ultimately progress. Past research shows that high tumor mutation burden was correlated anti-PD-1 response with advanced melanomas in the Caucasian population. However, tumor mutation burden is low for acral and mucosal melanomas [23].

Based on the prediction results of TILs, some patients with melanoma predicted as high density TILs (9.2% TILs in tumor region) are actually non-responders, on the other hand, responders with low density TILs (0.2% TILs in tumor region) exist as well. This reflects the limitation of TILs as a predictive biomarker, due to the complexity nature of the tumor microenvironment. Consequently, an effective biomarker is needed for melanoma and lung cancer patients.

Furthermore, we found that AUCs for Resnet-50, Inception-V3, VGG-19, Nasnet, Desnet and Mobilenet [18-22] are much lower than our proposed CNN model, corresponding AUCs ranging from 0.55 to 0.71 (supplementary Fig. B).

Deep learning has been developed for decades, and has outperformed human in many image classification tasks, for example, predicting clinical grade and stage [24] on several cancer types and gene mutations on non-small cell lung cancer [25]. However, due to lack of enough training data, CNN has limited prediction power. Sample size is truly important for deep learning models to avoid overfitting. With a small training dataset, the model could achieve good AUC score on the training data but poor generalization ability to testing data. For our method, we selected 190 SKCM patients with top and bottom 20% interferon-gamma expression values from the TCGA dataset as our training cohort. We further used principal component analysis (PCA) to reduce overfitting. In this finding, we combined traditional textual features with deep-learning extracted features by transfer learning with pre-trained deep model from ImageNet. 4 radius of scales ($r = 2, 4, 6, 8$) LBPs were transformed to extract information in border regions.

In this study, the training data labels were inferred from INFG levels. INFG secreted by immune cells in the tumor microenvironment causes growth arrest, up-regulates MHC class I expression, contributes to the recruitment of effector cells, causes T-reg fragility and coordinates the process of innate and adaptive antitumor response. Meanwhile, the same INFG signaling compromises antitumor immunity and activates PD-1 activity. INFG induces the expression of PD-L1 through increasing STAT1 signaling and decreasing STAT3 activation. In some studies, strong correlation between INFG and ICB response was reported [26]. In a study of NSCLC and UC, INFG signature is associated with TMB signature [27]. Since the genetic biomarker TMB has been shown to correlated with microsatellite instability, INFG may also correlates to MSI status [28].

In our study, the training data was downloaded from TCGA, which contains mostly European ancestry with cutaneous melanoma while the testing data we collected contains Asian patients with melanoma subtypes dominant in mucosal and acral melanoma. However, the proposed model demonstrates good generalization ability in our evaluation. Previous studies have also demonstrated that interferon-gamma associated gene expression levels play a very similar role in determining response to ICB therapies in both western and Asian melanoma populations, which also supports the idea of generalizing the model to different populations of patients [29]. In the future, it would be interesting to further exam the applicability of the developed model to other cancer types.

Most established biomarkers such as INF-gamma, TIDE [30] and IMPRES [31] require NGS sequencing which is time consuming and expensive. To our knowledge, our deep learning model represents the first finding of using H&E images to determine whether certain patients could benefit from anti-PD-1 immunotherapy. Our method is robust to patient samples from different cancer centers in melanoma and also has the potential to predict immunotherapy response in other cancer types.

Limitations

Despite the superior prediction performance obtained by our method, further studies will be needed to confirm the robustness of our machine learning model and its generalization ability. Considering that there is no immunotherapy response data available on TCGA, in this study, the clinical outcomes in the training dataset were inferred from IFNG scores, where patients with top and bottom 20% INFG values were considered responders and non-responders. The effectiveness of this approach was validated on 54 melanoma patients from BMS (AUC=0.82). We are planning to add more training data from future clinical studies to ensure the extraction of key textural features, and help to optimize and stabilize our prediction model, such that it can eventually be used to perform patient screening in routine clinical practice for immuno-oncology treatments.

Further, many recent studies have shown that intratumoural heterogeneity also palys an important role in shaping anti-tumor immune responses [32]. Tumors with high heterogeneity might escape from the immune surveillance because of the outgrowth of sub-clones. Therefore, in our future work, we would apply the proposed model on tissues from different tumor sites to understand the distributions and variation of TILs at different tumor sites, which would further improve the prediction of the response to immunotherapies.

Conclusion

In summary, our study suggests that deep-learning based model may determine which patients could respond to ICB simply from routine H&E slides. We also discussed the limitations of our current deep-learning based model. Conclusively, deep learning could ultimately enable efficient identification of patients who may most likely benefit from immunotherapies, such as anti-PD-1, in a time-sensitive and cost-effective way.

Author contributions statement

Jing Hu: Methodology, Validation, Formal analysis, Software.
Chuanliang Cui: Conceptualization, Resources, Writing-Original Draft.
Wenxian Yang: Data Curation, Software, Validation.
Lihong Huang: Data Curation, Visulation.
Rongshan Yu: Methodology, Software, Writing-Review& Editing.
Siyang Liu: Resources, Writing-Review & Editing.
Yan Kong: Supervision, Project administration, Writing-Original Draft,Writing-Review& Editing.

Funding

Declaration of Competing Interest

None declared.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2020.100921.

References

[1] C. Rodriguez-Cerdeira, M. Carnero Gregorio, A. Lopez-Barcenas, et al., Advances in Immunotherapy for Melanoma: a Comprehensive Review, Mediators Inflamm. 2017 (2017) 3264217.

[2] Z. Chi, S. Li, X. Sheng, et al., Clinical presentation, histology, and prognoses of malignant melanoma in ethnic Chinese: a study of 522 consecutive cases, BMC Cancer 11 (2011) 85.

[3] M. Binnewies, E.W. Roberts, K. Kersten, et al., Understanding the tumor immune microenvironment (TIME) for effective therapy, Nat. Med. 24 (5) (2018) 541–550.

[4] L. Si, X. Zhang, Y. Shu, et al., A phase Ib study of pembrolizumab as second-line therapy for chinese patients with advanced or metastatic melanoma (KEYNOTE-151), Transl. Oncol. 12 (6) (2019) 828–835.

[5] S. Hwang, A.Y. Kwon, J.Y. Jeong, et al., Immune gene signatures for predicting durable clinical benefit of anti-PD-1 immunotherapy in patients with non-small cell lung cancer, Sci. Rep. 10 (1) (2020) 643.

[6] R. Barroso-Sousa, E. Jain, O. Cohen, et al., Prevalence and mutational determinants of high tumor mutation burden in breast cancer, Ann. Oncol. 31 (3) (2020) 387–394.

[7] R. Cohen, T. Pudlarz, M.L. Garcia-Larnicol, et al., [Localized MSI/dMMR gastric cancer patients, perioperative immunotherapy instead of chemotherapy: the GERCOR NEONIPIGA phase II study is opened to recruitment], Bull. Cancer (2020).

[8] J.F. Gainor, H. Rizvi, E. Jimenez Aguilar, et al., Clinical activity of programmed cell death 1 (PD-1) blockade in never, light, and heavy smokers with non-small-cell lung cancer and PD-L1 expression >/=50, Ann. Oncol. 31 (3) (2020) 404–411.

[9] J.N. Kather, A.T. Pearson, N. Halama, et al., Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer, Nat. Med. 25 (7) (2019) 1054–1056.

[10] Xu H., Park S., Lee S.H., Hwang T.H. Using transfer learning on whole slide images to predict tumor metastasis in bladder cancer patients. *bioRxiv*. 2019:554527.

[11] Saltz J., Gupta R., Hou L., et al. Tumor-infiltrating lymphocytes maps from tcga h&e whole slide pathology images. In: 2018.

[12] A. Goode, B. Gilbert, J. Harkes, D. Jukic, M Satyanarayanan, OpenSlide: a vendor-neutral software foundation for digital pathology, J. Pathol. Inform. 4 (2013) 27.

[13] A. Tam, J. Barker, D Rubin, A method for normalizing pathology images to improve feature extraction for quantitative pathology, Med. Phys. 43 (1) (2016) 528.

[14] F. Riaz, A. Hassan, P. Pimentel-Nunes, E.J.L.D. Libnio, M Tavares Coimbra, How well can the fusion of Gabor filters and local binary patterns help in identifying gastric lesions? Conf. Proc. IEEE Eng. Med. Biol. Soc. 2016 (2016) 1204–1207.

[15] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science 315 (5814) (2007) 972–976.

[16] Chollet F. Xception: deep learning with depthwise separable convolutions. In. *arXiv e-prints*2016.

[17] J. Saltz, R. Gupta, L. Hou, et al., Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images, Cell Rep. 23 (1) (2018) 181–193 e187.

[18] He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *arXiv e-prints*. 2015. https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H. Accessed 10 December 2015.

[19] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision. arXiv e-prints. 2015. https://ui.adsabs.harvard.edu/abs/2015arXiv151200567S. Accessed 1 December 2015.

[20] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv e-prints. 2014. https://ui.adsabs.harvard.edu/abs/2014arXiv1409.1556S. Accessed 1 September 2014.

[21] Zoph B., Vasudevan V., Shlens J., Le Q.V. Learning transferable architectures for scalable image recognition. arXiv e-prints. 2017. https://ui.adsabs.harvard.edu/abs/2017arXiv170707012Z. Accessed 1 July 2017.

[22] Howard A.G., Zhu M., Chen B., et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv e-prints. 2017. https://ui.adsabs.harvard.edu/abs/2017arXiv170404861H. Accessed 1 April 2017.

[23] N.K. Hayward, J.S. Wilmott, N. Waddell, et al., Whole-genome landscapes of major melanoma subtypes, Nature 545 (7653) (2017) 175–180.

[24] G. Campanella, M.G. Hanna, L. Geneslaw, et al., Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, Nat. Med. 25 (8) (2019) 1301–1309.

[25] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, et al., Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning, Nat. Med. 24 (10) (2018) 1559–1567.

[26] L. Ni, J. Lu, Interferon gamma in cancer immunotherapy, Cancer Med. 7 (9) (2018) 4509–4516.

[27] B.W. Higgs, C.A. Morehouse, K. Streicher, et al., Interferon gamma messenger rna signature in tumor biopsies predicts outcomes in patients with non-small cell lung carcinoma or urothelial cancer treated with durvalumab, Clin. Cancer Res. 24 (16) (2018) 3857–3866.

[28] R. Cristescu, R. Mogg, M. Ayers, et al., Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy, Science 362 (6411) (2018).

[29] Kong Y., Xu C., Cui C., et al. Ratio of the interferon-γ signature to the immunosuppression signature predicts anti-PD-1 therapy response in melanoma. bioRxiv. 2020:2020.2004.2018.047852.

[30] P. Jiang, S. Gu, D. Pan, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (10) (2018) 1550–1558.

[31] N. Auslander, G. Zhang, J.S. Lee, et al., Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma, Nat. Med. 24 (10) (2018) 1545–1549.

[32] R. Rosenthal, E.L. Cadieux, R. Salgado, et al., Neoantigen-directed immune escape in lung cancer evolution, Nature 567 (7749) (2019) 479–485.