Report on Spatially Resolved Tumor Purity Maps

Wang Ruining

January 16, 2022

Introduction

Tumors consist of a complex mixture of cells such as cancer cells, normal epithelial cells, stromal cells and infiltrating immune cells. Tumor purity refers to the proportion of tumor cells in tumor tissue [1].

As the promise of personalized medicine, high throughput genomic analysis is indispensable for cancer research [2,3]. Tumor purity influences the sampling and collection of data required for this analysis. Thus, accurate tumor purity estimates are critical for precise pathological assessment and sample selection in high-throughput genomic analysis. If the sample chosen is based on higher level of tumor purity than real values, the false-negative test result may lead to the delay of effective treatment.

When different types of cancer are considered, tumor purity functions as the judgement on immune evasion in gastric cancer [4], patients' survival condition in colorectal cancer [5]. Moreover, therapeutic response in colon cancer and gastric cancer can be predicted by the level of tumor purity [6]. Therefore, accurate tumor purity is also essential to clinical treatment.

Tumor purity estimations

There are two main methods to estimate the tumor purity: percent tumor nuclei estimation and genomic tumor purity inference. The former one needs the counting on percentage of nuclei over a region of interest (ROI) in H&E stained histopathology slides by pathologists. This method is generally adopted with cellular level resolution, but time consumption and differences from different experimenters [7] are main disadvantages. Tumor purity inferred from different genomic data is determined as the gold standard recently, which comes from somatic copy number data, somatic mutations data [6] and some other kinds of genomic data. Genomic tumor purity values are commonly used in genomic analysis and the relationship between tumor purity and clinical variables [6]. It is not suitable when the tumor purity is low. And this method cannot reserve cell level information and there is a lack in reflections on

different values in spatial angle.

Recent years, with the prevalence of convolutional neural network, machine learning methods have been used widely in biomedical area and achieves better performance. An effective tumor purity estimation in spatial diversity by machine learning network has been developed in this study. This network is based on multiple instance learning (MIL) model and characterized by the novel 'distribution' pooling filter. By samples from tumor slides as the input, the prediction values are tested with respect to the standard—genomic tumor purity data. The results correlate significantly with standard values.

This MIL model is utilized to predict tumor purity from H&E stained histopathology slides. The structure is illustrated as the data processing procedure.

- Firstly, inputs are bags of patches cropped from both the top and bottom slides of the tumor samples where the bag labels are the tumor purities. In this step, data are collected in the forms of $\mathcal{D} = \{(X, Y) \mid X \in \mathcal{X} \; and \; Y \in \mathcal{Y}\}$, where X is data patches included in a bag and Y is the bag label—tumor purity.
- Then data experiences feature extraction. The feature extractor is based on ResNet18, a residual learning framework that explicitly reformulates the layers as learning residual functions with reference to the layer inputs. This layer completes the map of $\theta_{feature} : \mathcal{I} \rightarrow \mathcal{F}$. This step extracts feature in J dimension from x_i in a bag.
- After features are extracted in the vector from, these are put into the MIL pooling filter. Mapping in this stage is completed in the form: $\theta_{filter} : \mathbb{R}^{JN} \rightarrow \mathcal{H}$. This procedure, namely 'distribution' pooling, aggregates the features into the bag-level representation space by transform: $h_X = \theta_{filter}(F_X)$. It can generate stronger bag-level representation than normal pooling, *i.e.* max pooling and mean pooling, by estimating marginal feature distributions.
- Finally, the feature in bag-level is transformed into predicted bag label $\hat{Y} \in \mathcal{Y}$ and compared with the true value.

Results and conclusions

This novel MIL model achieves the prediction of tumor purity in fresh-frozen slides of different TCGA cohorts. And it also performs well in H&E stained histopathology slides of formalin-fixed paraffin-embedded (ffpe) sections in the Singapore cohort using transfer learning. The estimations are consistent with standard values, and lower error rate compared with the percent tumor nuclei estimates. Moreover, spatial map of tumor purity is constructed and labeled in forms of the gradation of colors.

During the experiment, it is found that the spatial difference exists in the whole tumor sample. Thus, the effect should be better if two in stead of one slide from top and bottom of samples are collected and input into training to acquire the tumor purity.

This also implies the necessity of spatial map in tumor purity in analysis of why the predictions of pathologists are always higher.

This MIL model learns discriminant features between normal and cancerous samples, and successfully classifies and segments these two kinds of region in sample slides.

References

1. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. Nature communications 6, 1–12 (2015).
2. Schuster, S. C. Next-generation sequencing transforms today's biology. Nature methods 5, 16–18 (2008).
3. Xuan, J., Yu, Y., Qing, T., Guo, L. & Shi, L. Next-generation sequencing in the clinic: promises and challenges. Cancer letters 340, 284–295 (2013).
4. Gong, Z., Zhang, J. & Guo, W. Tumor purity as a prognosis and immunotherapy relevant feature in gastric cancer. Cancer medicine 9, 9052–9063 (2020).
5. Mao, Y., Feng, Q., Zheng, P., Yang, L., Liu, T., Xu, Y., Zhu, D., Chang, W., Ji, M., Ren, L., Wei, Y., He, G., & Xu, J. (2018). Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. Cancer management and research, 10, 3569–3577. https://doi.org/10.2147/CMAR.S171855
6. Oner, U., M., Chen, J., Revkov, E., James, A., Heng, S. Y., Kaya, A. N., Alvarez, J. J. S., Takano, A., Cheng, X. M., Lim, T. K. H., Tan, D. S. W., Zhai, W., Skanderup, A. J., Sung, W. K., Lee, H. K., Obtaining spatially resolved tumor purity maps using deep multiple instance learning in a pan-cancer study, Patterns,2021,100399,ISSN 2666-3899, https://doi.org/10.1016/j.patter.2021.100399.
7. Mikubo, M. et al. Calculating the tumor nuclei content for comprehensive cancer panel testing. Journal of Thoracic Oncology 15, 130–137 (2020).