Report: Obtaining Spatially Resolved Tumor Purity Maps Using Deep Multiple Instance Learning In A Pan-cancer Study

Tumor purification is an essential factor in pathologic since it helps to determinant tumor formation and therapeutic response. Percent tumor nuclei estimates and genomic tumor purity inference are common methods used in the clinic to determine cancer purity but these approaches have some significant limitations. For example, percent tumor nuclei estimates are tedious, time-consuming, and produce inter-observer variability between pathologists. On the other hand, genomic tumor purity inference lacks spatial information of cancer cell location. Hence this study scientists develop a Machine learning model to improve the cancer purity prediction. This model is based on reading H&E stained histopathology slides which is more cost-effective and reduces its manual steps.

Patch-based models and multiple instance learning (MIL) models can be used to predict tumor purity from digital histopathology slides. Although patch-based models are wildly used in different studies [39–44], they require pathologists' pixel level annotation to be trained on patch cropped from a slide using the corresponding patch label determined which is tendinous and expensive. Whereas MIL represents a sample as a bag of patches cropped from the sample's slides and uses a sample-level label as the bag label [45–48].

The novel MIL model they designed consisted of three modules: feature extractor module, MIL pooling filter, and bag-level representation transformation module. And one of the remarkable features is its 'distribution' pooling filter which produces stronger bag-level representations. Also, it uses the neural network in the feature extractor module and bag-level representation transformation module.

This model analyzes data from ten different TCGA cohorts, every cohort has more than 400 patients and a local Singapore cohort, 179 patients. The histopathology slides in each cohort were randomly segregated at the patient level into training, validation, and test sets.

The results of tumor purity predicted by the MIL correlate significantly with genomic tumor purity values, also it produced lower mean absolute error than percent tumor nuclei estimate. Apart from it, the model performs successfully in slides of formalin-fixed paraffin-embedded (ffpe) section which is different than slides of fresh-frozen sections, it preserves morphology better.

On the other hand, the top and bottom slides of a sample generate different tumor purity results, and the degree of spatial variation in tumor purity depend on the types of cancer so the result of cancer purity using both top and bottom slides is better than using only one slide. During the analysis scientists reveal the probable cause of pathologists' high percent tumor nuclei estimates, normally pathologists select high tumor content regions over the slides for percent tumor nuclei estimation. Finally, the novel MIL model learns how to discriminate features between cancerous and normal tissue histology and successfully classifies samples.

Using MIL models to sample selection for molecular analysis, reduce pathologists' workload and decrease inter-observer variability. And spatial information contributes to a better understanding of the tumor microenvironment.

Therefore, its results can be used as prognostic biomarkers to stratify patients. Despite this model performing efficiently, it produces variated results between fresh-frozen and formalin-fixed paraffin-embedded tissue preservation methods. Because it's based on deep learning-based, it performs better with more data.

Reference

1. Viray, H. et al. Automated objective determination of percentage of malignant nuclei for mutation testing. Applied immunohistochemistry & molecular morphology: AIMM/official publication of the Society for Applied Immunohistochemistry **22**, 363 (2014).
   Google Scholar
2. Hamilton, P. W. et al. Automated tumor analysis for molecular profiling in lung cancer. Oncotarget **6**, 27938 (2015).
   PubMedGoogle Scholar
3. Azimi, V. et al. Breast cancer histopathology image analysis pipeline for tumor purity estimation in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (2017), 1137–1140.
   Google Scholar
4. Pei, Z., Cao, S., Lu, L. & Chen, W. Direct cellularity estimation on breast cancer histopathology images using transfer learning. Computational and mathematical methods in medicine **2019** (2019).
   Google Scholar
5. Rakhlin, A. et al. Breast tumor cellularity assessment using deep neural networks in Proceedings of the IEEE/CVF International Conference on Computer Vision Work-shops (2019), 0–0.
   Google Scholar
6. Greene, C. et al. The Potential of Digital Image Analysis to Determine Tumor Cell Content in Biobanked Formalin-Fixed, Paraffin-Embedded Tissue Samples. Biopreservation and Biobanking (2021).
   Google Scholar
7. Quellec, G., Cazuguel, G., Cochener, B. & Lamard, M. Multiple-instance learning for medical image and video analysis. IEEE reviews in biomedical engineering **10**, 213–234 (2017).
   Google Scholar
8. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine **25**, 1301–1309 (2019).
   PubMedGoogle Scholar
9. Tomita, N. et al. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. JAMA network open **2**, e1914645–e1914645 (2019).
   Google Scholar
10. Oner, M. U., Lee, H. K. & Sung, W.-K. *Weakly Supervised Clustering by Exploiting Unique Class Count* in International Conference on Learning Representations (2020).
    Google Scholar