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

Tumor purity, the percentage of cancer cells in the tumor tissue, is the pathological significance in estimating outcomes. For example, Tumor purity is a predictor for colon cancer and gastric cancer[1,2]. An overestimated tumor purity may lead to the result that an actual cancerous patient was tested negative. Therefore, an accurate estimation of tumor purity has significant importance in diagnosis.

There are two main methods to measure tumor purity. The first one is to calculate the percentage of tumor nuclei by counting tumor nuclei over desired histopathology slides stained by H&E which could distinguish nuclei and cytoplasm. It is a basic but tedious approach. The other one is to estimate tumor purity from various genomic data. It is considered as the golden standard to calculate tumor purity nowadays since it reduces the effect of normal cells' contamination[3,4,5]. However, it does not apply to content samples with low tumors[6].

This paper provides cost-effect ways to estimate tumor purity by implementing multiple instance learning models on datasets. Instead of using patch-based models which relies on pathologists' pixel-level annotations, MIL model trains, and tests on bags of samples cropped from histopathology slides which are easily collected[6].

The datasets in this paper are retrieve from ten different cohorts in TCGA and a cohort from East Asia containing H&E stained fresh-frozen section histopathology slides[6] .After cropping the slides from the top and bottom of the samples, all samples are bagged with labels on each bag. After using ResNet18 networks to extract feature of bags of data, distribution pooling filter is implemented to apply marginal distribution on the input feature matrix. Then, another neural network is used to transform the former result to predicted bag label[6].

Reference

1. Mao, Y. *et al.* Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. *Cancer management and research* **10,** 3569 (2018).

2. Gong, Z., Zhang, J. & Guo, W. Tumor purity as a prognosis and immunotherapy relevant feature in gastric cancer. *Cancer medicine* **9,** 9052–9063 (2020).

3. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nature genetics* **45,** 1134–1140 (2013).

4. Akbani, R. *et al.* A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nature communications* **5,** 1–15 (2014).

5. Network, C. G. A. R. *et al.* Comprehensive molecular profifiling of lung adenocarcinoma.
Nature **511,** 543–550 (2014).

6. Mustafa, U O. et al.(July 2021). *Obtaining spatially resolved tumor purity maps using deep multiple instance learning in a pan-cancer study.* Retrieved from *https://doi.org/10.1101/2021.07.08.451443.*