Answer to Q1

Ziyuan Wang

Main point of that paper

In this paper, a machine learning model for predicting tumor purity from H&E stained histopathological sections was developed, thus making predictions consistent with genomic tumor purity values. This approach is less costly and time consuming than genome sequencing.

The image input is regarded as a bag and the $1mm^2$ *regions are considered as instances*

Implementations and Results

The aim of the model

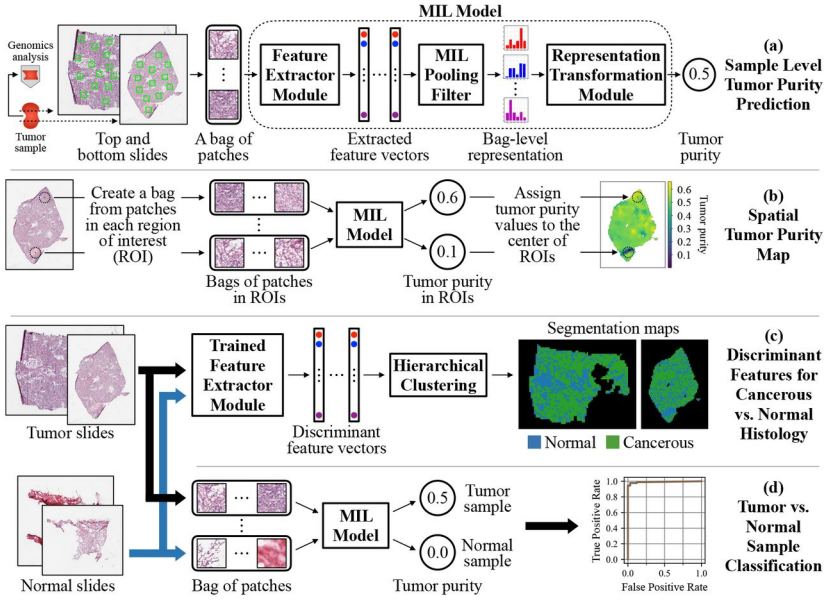


Figure 1: Overview of the paper

- Multi-instance learning (MIL) was carried out through the input sample

image, and the tumor nuclear purity of the sample was predicted by bag-level feature vector in the output layer.
- Obtain a spatial tumor purity map for a slide showing every $1mm^2$ region purity.
- Hierarchical clustering was performed using only weak labels to obtain features that could distinguish cancer tissue from normal tissue.
- Classify samples into tumor vs. normal.

Architecture

Modules

- Feature Extractor module
- MIL pooling filter module
- bag-level representation transformation

ResNet18 model as the feature extractor module and a three-layer multilayer-perceptron as the bag-level representation transformation module.

Unlike max/min-pooling which converts each dimention of extractor features different instances into one value, MIL pooling filter module converts them into a distribution using 21 sample points(Default). In this paper, the performance of this pooling method is better than maximum pooling, minimum pooling and average pooling.

Performance

This tool has a high correlation with the results obtained from transcriptome determination of tumor purity in samples, although there are some outliers.

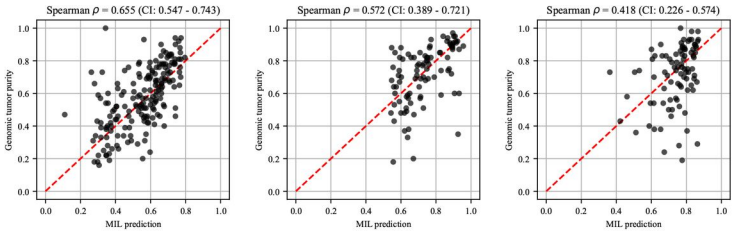
AUC value (0.991) was utilized to evaluate our model perform via, which tumor samples were separated from normal samples in LUAD cohort. Classical image processing and machine learning-based method(Yu et al. 2016) and the DNA plasma-based method(Sozzi et al. 2003) (0.85,0.94 respectively). Other models, such as, the deep learning model of (Coudray et al. 2018) (AUC: 0.993) ans (Fu et al. 2020) (AUC: 0.977 with 95% CI: 0.976 - 0.978). However, there is one concern about the dataset preparation methods of Coudray et al. (2018) and Fu et al. (2020). How they sampled the data made their models' performance illusory

Discussion

Advantages:

- Weak tumor purity labels necessitated a MIL approach. Pixel-level annoations(expensive) can be avoided
- Complement spatial-omics(scRNA-seq) which can be seen from the Fig1.

Source of the error:



(a) BRCA (b) GBM (c) LGG

(d) LUAD (e) LUSC (f) OV

(g) PRAD (h) UCEC (i) LUAD_SG

Figure 2: TCGA dataset benchmark

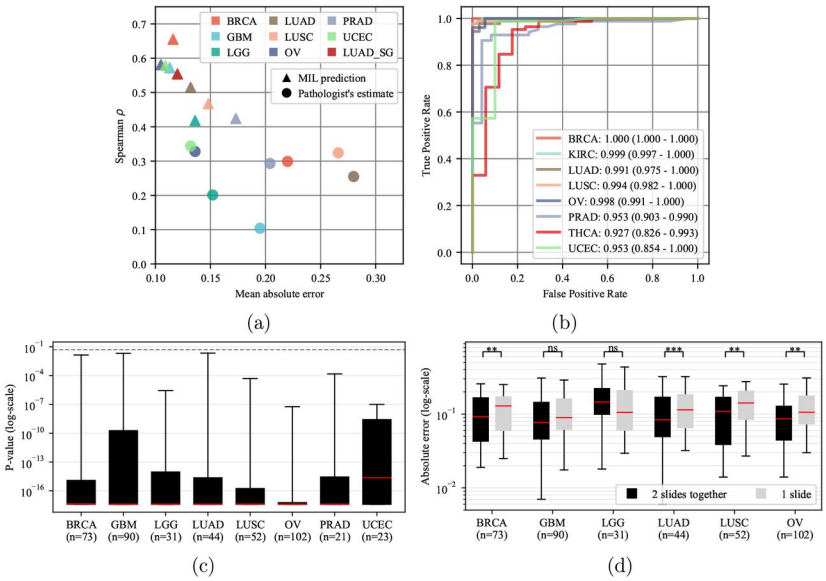


Figure 3: Comparasion with other tools

- Lack of samples.
- Histopathology slides from different areas.
- Some limitation of H&E ained histopathology slides.

Reference

Coudray, Nicolas, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L. Moreira, Narges Razavian, and Aristotelis Tsirigos. 2018. "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning." *Nature Medicine* 24 (10): 1559–67. https://doi.org/10.1038/s41591-018-0177-5.

Fu, Yu, Alexander W. Jung, Ramon Viñas Torne, Santiago Gonzalez, Harald Vöhringer, Artem Shmatko, Lucy R. Yates, Mercedes Jimenez-Linan, Luiza Moore, and Moritz Gerstung. 2020. "Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis." *Nature Cancer* 1 (8): 800–810. https://doi.org/10.1038/s43018-020-0085-8.

Sozzi, Gabriella, Davide Conte, Maria Elena Leon, Rosalia Cirincione, Luca Roz, Cathy Ratcliffe, Elena Roz, et al. 2003. "Quantification of free circulating DNA as a diagnostic marker in lung cancer." *Journal of Clinical Oncology* 21 (21): 3902–8. https://doi.org/10.1200/JCO.2003.02.006.

Yu, Kun Hsing, Ce Zhang, Gerald J. Berry, Russ B. Altman, Christopher Ré, Daniel L. Rubin, and Michael Snyder. 2016. "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features." *Nature Communications* 7 (1): 12474. https://doi.org/10.1038/ncomms12474.