Predicting Prostate Cancer Gene Alteration with Deep Learning on Histopathological Images

Lucas J. Liu[1], Stephanie Harmon[2], Michael Haffner[1], Helen Richards[1], Zhijun Chen[2], Erolcan Sayar[1], Harry Zhang[2], Ruth Etzioni[1]

[1]Fred Hutchinson Cancer Center, Seattle, WA. 98109

[2]Molecular Imaging Branch, National Cancer Institute, NIH, Bethesda, MD, 20892

Abstract:

Identification of genetic alteration is important for effective therapy selection in cancer patients, but routine genomic test is often not feasible due to its high cost. Less costly Hematoxylin and Eosin (H&E)-stained slides used in routine cancer diagnosis have shown their potential to reflect genetic alterations through morphological patterns. This suggests that H&E slides could potentially serve as a cost-effective alternative to genomic tests.

In this study, we present a deep learning framework to predict gene alteration from H&E whole slide images. We analysis 127 H&E images samples from University of Washington clinical biopsy cohort. These samples are derived from metastatic prostate cancer patients, originating from various anatomic sites, including specific locations such as lymph nodes, bone, brain, liver, rectum, lung, and soft tissue (n = 53, 41.7%). Additionally, 74 samples (58.3%) are from an unknown location. We use pretrained foundation models to extract features from high-resolution H&E images and train a multi-tasking prediction model using the extracted features to predict TP53, PTEN, HR, AR and RB1 alterations as well as tumor mutation burden (TMB). We find the heterogenous performances across different genes considered in terms of the receiver operating characteristic curve areas (ROC AUCs). Our method achieves the best ROCAUC of 0.80 for RB1, and the lowest ROCAUC of 0.50 for PTEN, while other alterations have modest performances. These findings suggest that using deep learning to predict mutations from H&E slides can be effective for certain alteration but less so for others. Further research is needed to discover the associations between morphological patterns and specific genetic profiles using deep learning.