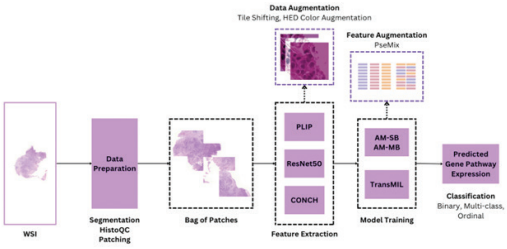

Figure 1: Overview of pipeline

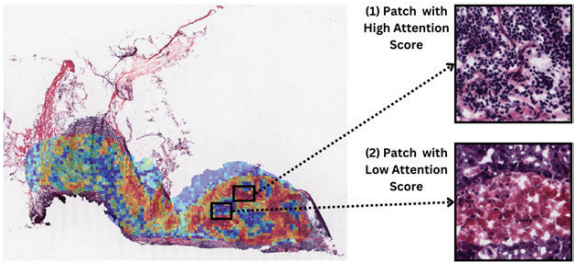

Figure 2: Attention map obtained using the AM-SB architecture for a WSI predicted to have a high degree of T cell-mediated cytotoxicity. Red zones and blue zones indicate high and low attention scores, respectively. At right are two exemplar patches: 1) the high-attention patch illustrates abundant tumor-infiltrating lymphocytes without tumor cells, which are suggestive of high immune activity, and 2) the low-attention region demonstrated areas of tumor necrosis and minimal lymphocytes, consistent with low immune activity.

Reimagining Precision Oncology Through Deep Learning-Enabled Computational Pathology Tools

Albert Kim, MD
Assistant Physician, Mass General Cancer Center;
Assistant Professor, Harvard Medical School
akim46@mgh.harvard.edu

While molecular analysis of tumor tissue has transformed the therapeutic landscape of modern oncology, much work remains to realize the full potential of precision oncology. Existing clinico-genomic biomarkers are insufficient as stand-alone tools given the wide variability between outcomes among patients with similar biomarker profiles. Omics-based studies illustrate that interactions between tumor cells and the tumor microenvironment (TME) dictate therapeutic efficacy. However, most pathology workflows only evaluate tumor cell characteristics within a hematoxylin-and-eosin (H&E) slide and do not reproducibly assess the TME. Therefore, many have proposed incorporating transcriptional profiling into the precision oncology framework to improve treatment selection and prognostication for patients. However, commercial gene expression panels are challenging to develop due to poor RNA quality within paraffin-embedded tissue, thus highlighting a need for alternate methods to quantify TME biology.

To address this unmet clinical need, we applied deep learning (DL) to extract and quantify facets of the TME from H&E whole slide images (WSIs) by recognizing tumor cell morphology and physical co-localization of different cell populations within a tumor. Using a large dataset of paired H&E WSIs of breast cancer core biopsies and bulk RNA-sequencing data, we employed attention-based multiple instance learning to quantify therapeutically relevant tumor and immune phenotypes. We trialed different feature extraction techniques and model architectures to optimize performance (Figure 1). Using binary classification, our models exceeded area under the receiver operating characteristic (AUROC) scores above 0.80 for most gene expression pathways. Attention maps illustrate that our trained models recognized biologically relevant spatial patterns of cell populations from H&E WSIs (Figure 2). This model is now being validated on an external dataset from triple-negative breast cancer patients who received the Keynote-522 regimen.

Our model represents a first step towards developing computational H&E tools that reflect facets of TME biology and have potential to inform selection of more effective treatments for patients of virtually all tumor types. Given the wide availability of H&E slides, our tools fit into existing pathology lab workflows and can serve as a more readily accessible alternative to DNA/RNA sequencing for implementation of precision oncology.