

Weakly-supervised tumor purity prediction from H&E stained slides

Principal Investigator:

- Iman Hajirasouliha, PhD – Weill Cornell Medicine

Co-Principal Investigator:

- Victoria Popic, PhD - The Broad Institute of MIT and Harvard

Abstract: The primary objective of this study is to develop a novel deep learning model to accurately predict tumor purity in a tissue slide. Our approach is a weakly-supervised model based on resnet using attention-MIL. It will also use a multitask approach for predicting tissue type in addition to tumor purity. We will assemble a large dataset from The Cancer Genome Atlas (TCGA) and in-house real samples from the Englander Institute for Precision Medicine (EIPM) for the purpose of training and validation, respectively. To enhance the performance of our models, we will also generate additional training examples using numerous data augmentation techniques and Generative Adversarial Networks (GANs). Tumor purity estimates are critical for tumor sequencing methods and can also provide prognostic information in certain cases.

We will rigorously benchmark and compare our results to previously developed models, pathology derived scores and molecular analysis. Our method will be able to identify the location of heterogeneous purity regions on the slide and classify tumors into low and high purity at several different thresholds that can be cancer-type specific. As a result, our method will enable the identification of candidate tumor regions that can be isolated and used to enrich the tumor sample for improved sequencing.