Introduction

Immunohistochemistry (IHC) is a cell staining technique routinely used by pathologists to diagnose cancers and make prognoses. However, significant inter-rater variability has been reported regarding the annotations of IHC images, specifically in glioma (kappa value = 0.04-0.32). Although methods such as deep learning have been used to analyze IHC images, the concordance of IHC-derived protein expression levels and those measured by reverse-phase protein array (RPPA) were not systematically investigated. Here, we develop deep learning methods to predict the IHC annotations from raw images and correlate the semi-quantitation of protein expression from IHC with the protein expression measurements from RPPA.

Methods

We developed a transfer learning pipeline to predict pathologists’ annotations on IHC images (staining, intensity, quantity, and location) using VGG-16 as the base model. To predict these annotations, the models were trained on 120,000 IHC images from the Human Pathology Atlas stained by 11,938 proteins and tested on 20,000 IHC images stained by 2,343 proteins not presented in the training set. We further gathered the RPPA data from The Cancer Genome Atlas and identified 28 proteins with IHC images, annotations, and RPPA measurements. We computed the correlations between RPPA levels and pathologist’s annotations on the IHC images as well as between the RPPA levels and our model’s predicted annotations.

Results

In a simplified binary classification task, the model correctly predicted high (strong or moderate) and low (low and negative) levels of intensity annotation values with an AUC of 0.886 (Figure 1). We further developed multi-class classification models to predict the detailed levels of intensity (4 classes), quantity (5 classes), staining (4 classes), and location (4 classes) annotated by pathologists, with top-1 accuracies of 0.58, 0.53, 0.56, and 0.59, respectively. For the 28 proteins with both RPPA and IHC data, pathologists’ annotations of intensity and staining levels were negatively correlated with the RPPA measurements (Spearman’s correlation coefficients = -0.039 and -0.164, respectively). The predictions of intensity and staining annotations from our model attained slightly better correlations with RPPA levels (Spearman’s correlation coefficients of 0.014 and -0.038, respectively).

Discussion

Our results suggest that convolutional neural network models can predict the pathologists’ annotations on IHC images with limited accuracy. Large-scale data collection and further optimization are needed to achieve accuracies sufficient for clinical usage. Such quantitative approaches could reduce the subjectivity involved in annotating the images. Interestingly, the pathologists’ annotations did not correlate well with RPPA data, while our model’s predicted annotations presented slight improvement. The low strengths of the correlations highlight the discordance between IHC and RPPA levels. Further investigations are needed to examine the reasons leading to these discordances. The improved correlation may suggest our model detects signals from the image associated with protein levels not captured by the manual pathology evaluation. Our approaches can facilitate the development of reliable methods for objective IHC semi-quantitation.

References