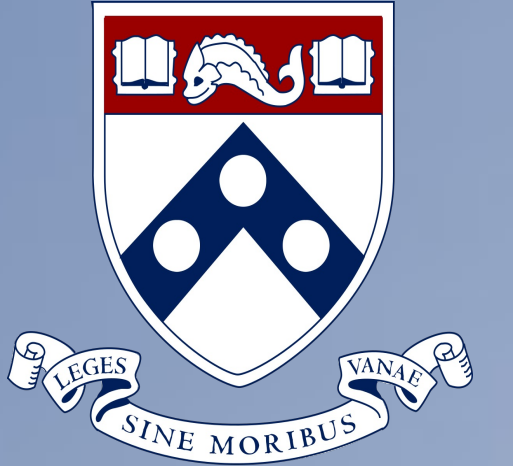




AI-Based Quantification of Cellular Protein Expression Levels

Ashrit Challa¹, Sina Bagheri, MD^{2,3}, Mahsa Kiani, MD^{2,3}, Grace Wu⁴, Chantal Chahine, MD³, Amy Clark, MD⁵, Ira J. Bleiweiss, MD⁶, Lauren Schwartz, MD⁶, Michael D. Feldman, MD, PhD⁶, Elizabeth S. McDonald, MD, PhD^{2,3}

1 Penn Undergraduate Research Mentoring Program (PURM); 2 Penn Breast Cancer Translational Research Group (BCTRG)
 3 Department of Radiology, University of Pennsylvania; 4 Medical Student, Perelman School of Medicine
 5 Division of Hematology Oncology, University of Pennsylvania, 6 Department of Pathology, University of Pennsylvania



Background

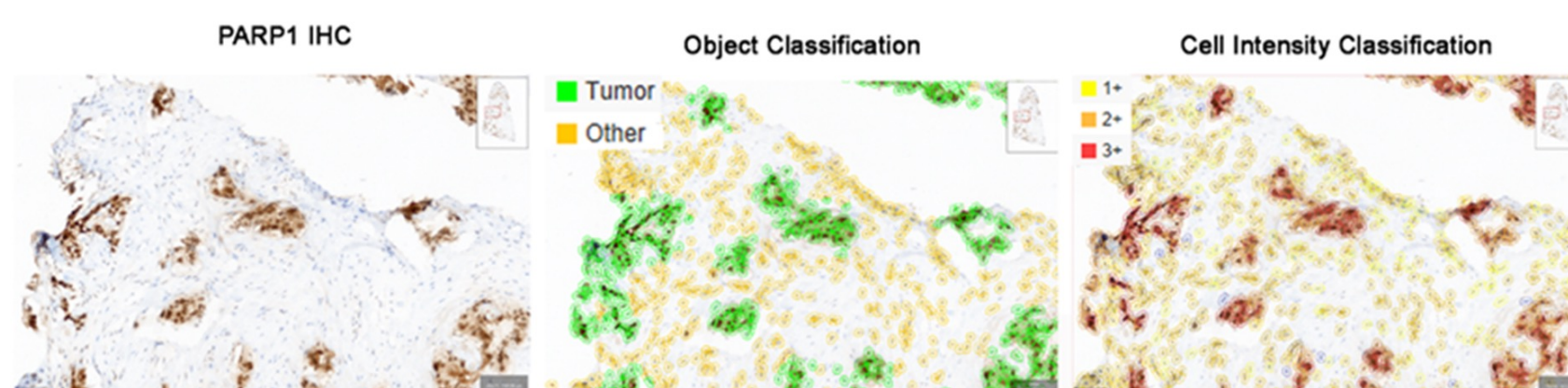
- Using human-collected data, it is possible to create an Artificial Intelligence, or AI, that can process the data in a manner similar to that of humans
- Open-Source software is a software that is free and available for anyone to access, use, and distribute
- QuPath is an Open-Source software that can be used to digitally analyze bioimages, allowing for easy access regardless of financial power, providing equity to the scientific field of bioimage analysis
- The software can be trained to examine a variety of histopathologic slides stained with different techniques such as hematoxylin and eosin, immunohistochemistry, and immunofluorescent staining; it can determine qualitative and quantitative information about that sample

Cell Detection

- In analyzing histological slides, AI algorithms can be built to detect individual cells and their nuclei, all while storing parameters regarding the detections for later use
- Software can be used to calculate an optical density (OD), a measure of light absorbance, for each pixel by adding the OD values of the respective red, green, and blue channels, thereby enabling calculation of the OD for various histological stains such as hematoxylin and DAB (3,3'-Diaminobenzidine)
- Using the computer-generated boundaries, statistical calculations using the OD of each pixel can be automatically quantified, such as the mean or maximum OD of a cell, nucleus, or cytoplasm. In turn, this can relay information about the level of expression for the molecule captured via a DAB stain
- Measurements including the area, perimeter, and nucleus-to-cell area ratio can be calculated and used in later cell classification efforts

Nucleus: Area	128.6543
Nucleus: Perimeter	47.9407
Nucleus: Circularity	0.7034
Nucleus: Max caliper	16.407
Nucleus: Min caliper	10.4896
Nucleus: Eccentricity	0.755
Nucleus: Hematoxylin OD mean	0.3845
Nucleus: Hematoxylin OD sum	199.1806
Nucleus: Hematoxylin OD std...	0.1349
Nucleus: Hematoxylin OD max	0.8925
Nucleus: Hematoxylin OD min	0.0996
Nucleus: Hematoxylin OD ran...	0.7929
Nucleus: DAB OD mean	0.7886
Nucleus: DAB OD sum	408.4843
Nucleus: DAB OD std dev	0.1784
Nucleus: DAB OD max	1.3028

Cell Measurements from a tissue stained via an H-DAB procedure



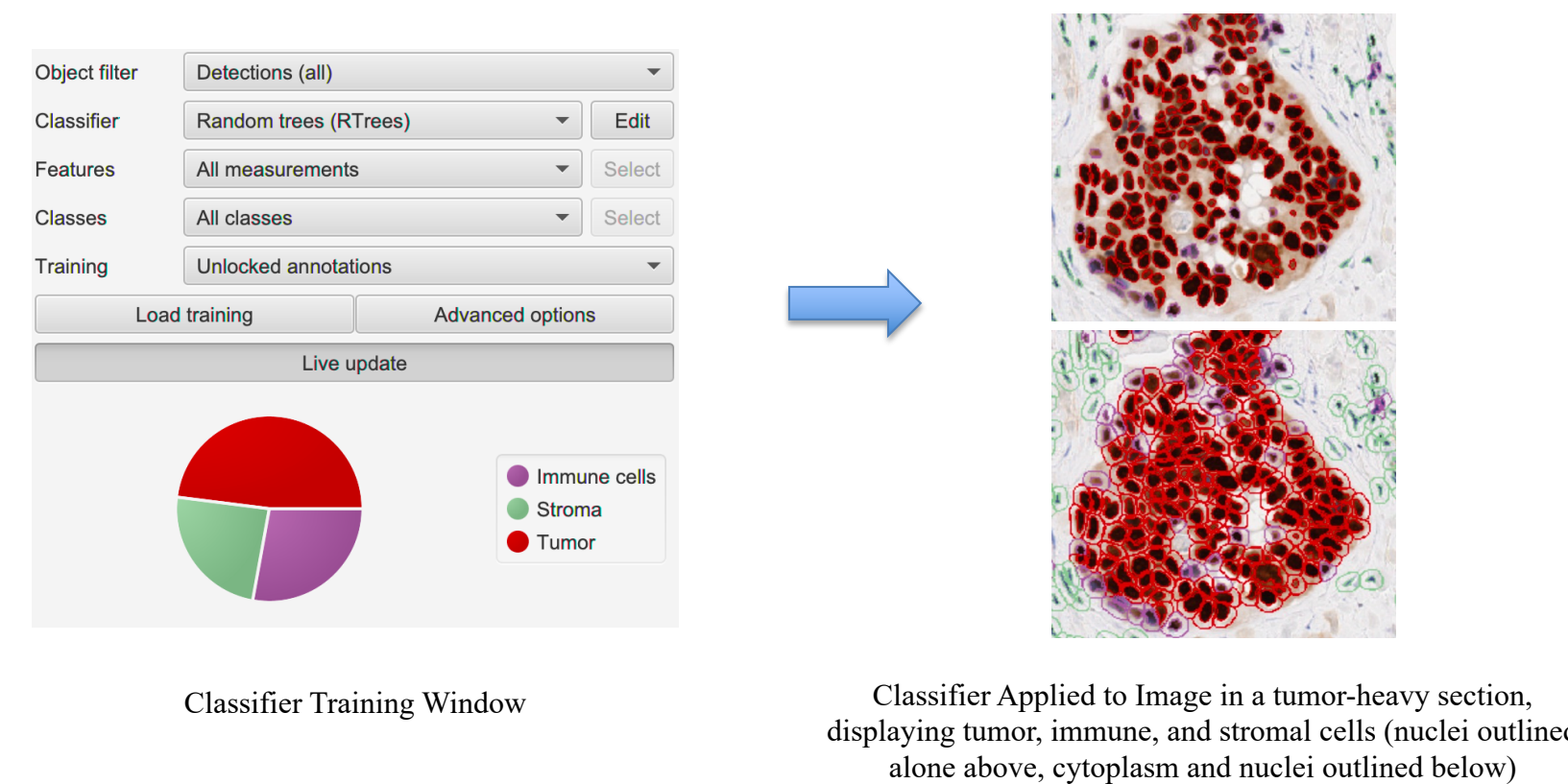
Methodology

Cell Classification

- Cell types in the image can be classified by annotating margins around cells of the same class, such as tumor cells, immune cells, or stromal cells
- Annotation of multiple regions of interest defining each cell class represent data that can be used via AI to train cell-type recognition. The learned algorithms are then applied to the whole image, and over time, increasing information about cell types and expected heterogeneity allows correct cell classification across an image set.

QuPath and AI

- Classifiers consider all measurements of a cell, including optical and geometrical quantities
- While a classifier is active, new annotations during pathology review will lead to cell classifications being updated in real time



Automation

- Subjective thresholds classify cells into the pathology ratings of Negative, +1, +2, and +3, which are gradations of intensity based on protein expression, with a Negative rating correlating to no expression and the +1, +2, and +3 ratings correlating to increasing levels of protein expression
- The ratings allow for calculation of a bioimage's H_q Score, which is based on a pathology metric that considers the relative proportion of each rating across different cell classes and across the whole image.
- The classical H-score of immunoreactivity is obtained by the formula: $3 \times$ percentage of 3+ nuclei + $2 \times$ percentage of 2+ nuclei + percentage of weakly staining, or 1+, nuclei, giving a range of 0 to 300.
- Programming can be used to automate parts of the method such as cell detection, application of the method to multiple images at once, and exportation of relevant quantities and results.
- To apply a classifier to a full dataset or across datasets it is important to have acquired the data with consistent parameters and scanning devices or correct for variability associated with acquisition.

Applications

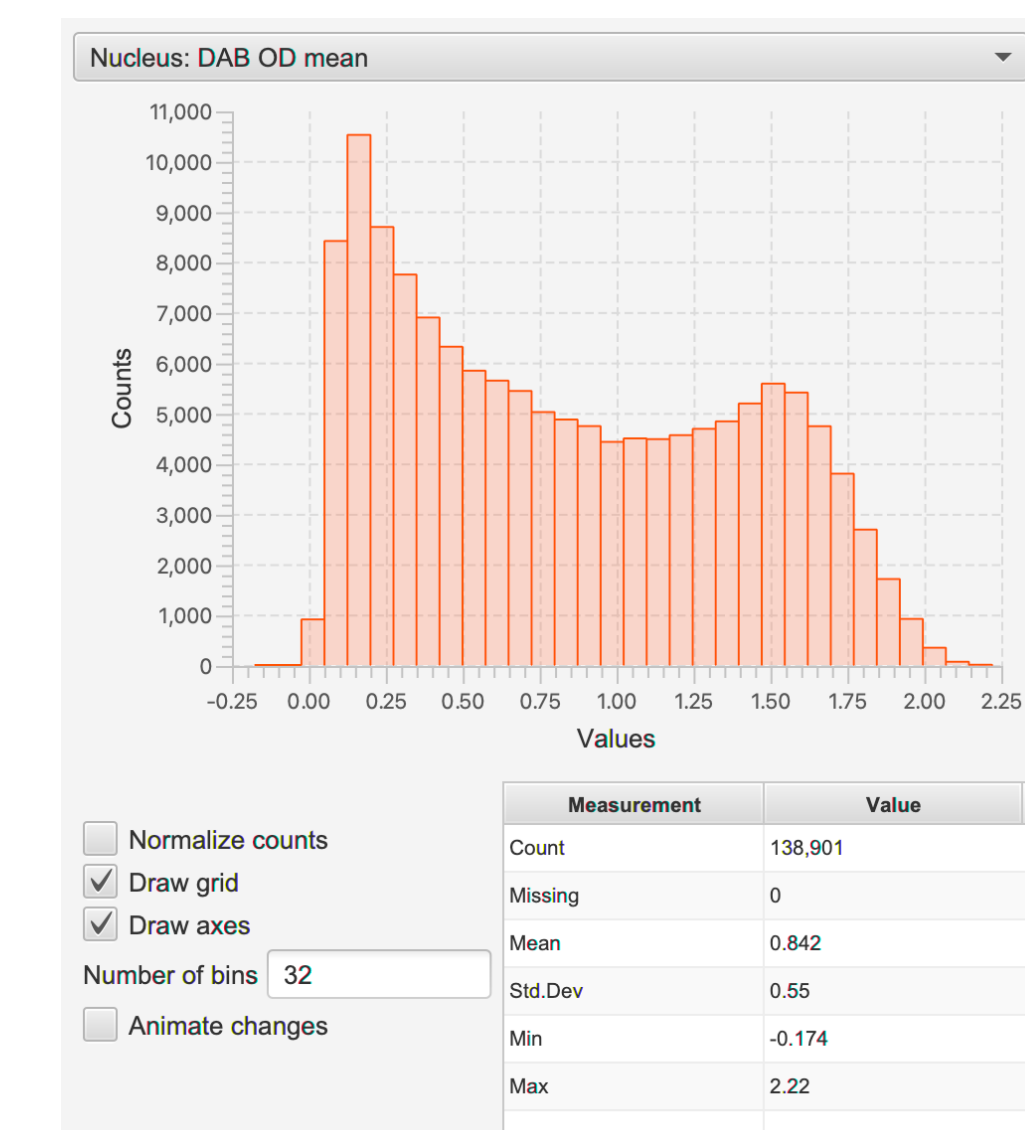
- This method provides a digital alternative, or comparative measure, to manual slide scoring.
- Additionally, this method can be used to determine proteomic information for different cell types and cell compartments via comparison of quantification results for different cell classes.

Conclusions

- We have refined a method of using a digital bioimage analysis platform to quantify protein expression levels
- Protein expression is quantified after cell detection and AI-based cell classification. Subjective thresholds can be applied if classical pathologic metric calculation is desired.
- While this analysis was performed on a subset of human breast cancer pathology slides immunohistochemically stained for the PARP-1 protein, this method can be applied to a variety of stains and a plethora of proteins, thanks to its easy adaptability to the stain type and molecule under examination

Next Steps

- Pre- and post-processing steps will be added to allow for integration of slide quality/stain variability and reduce classification error.
- We will generate separate histograms and datasets related to specific cell classes and compartments for easier visualization of class-specific proteomic differences
- Similar quantification strategies will be applied to multicolor images to allow for digital quantification of multiple proteins simultaneously.



Automatically Generated Histogram of Nucleus DAB OD mean values, which does NOT include separation into distinct cell classes