

# Quantitative Digital Image Analysis and Machine Learning for Staging of Prostate Cancer at Diagnosis

<sup>3</sup>Fangjin Huang\*, <sup>3</sup>Nathan Ing\*, <sup>2</sup>Eric Miller, <sup>3</sup>Hootan Salemi, <sup>1</sup>Michael S. Lewis, <sup>1,2</sup>Isla P. Garraway, <sup>3</sup>Arkadiusz Gertych and <sup>3</sup>Beatrice S. Knudsen

<sup>1</sup>Greater Los Angeles VA Health Care System, <sup>2</sup>UCLA and <sup>3</sup>Cedars- Sinai Medical Center

## Introduction and Objective:

Prostate cancer (PC) with de novo bone metastases at diagnosis (M1) carries a 5-year survival rate of 28% and requires early, aggressive treatment. Clinical assays and the pathology of prostate needle biopsies (PNBX) cannot distinguish primary M1 tumors from high-grade localized (M0) cases. We hypothesized that digital image analysis can be applied to obtain morphologic biomarkers, not recognizable by pathologists. Here we demonstrate how novel software tools that involve Deep Learning frameworks can be used to systematically extract handcrafted and autoencoder features and to build models to predict M1 stage at the time of diagnosis.

## Methods:

A study cohort, nested within a biorepository of 2150 PC patients at the Greater LA VA, consisted of 86 high-grade M0 and 85 M1 cases. Slides were digitized at 40X and 2 pathologists annotated all cancer foci. Approximately 30 image tiles were selected from each case. 62 handcrafted (HC) and 64 autoencoder (AE) features were extracted from nuclei. Feature values were subjected to an equal with binning procedure for normalization. The normalized profile of each primary feature gave rise to 11 secondary features, representing the distribution of the feature within a case. We separated cases into training + testing versus validation groups at a 80:20 ratio. Using a bootstrapping method, we selected the best GLMNET models predicting M0 versus M1 status in the training + testing set and applied them to an independent validation set of cases.

## Results:

After successful conversion of M0 and M1 image tiles to digital nuclear masks and color normalization, ~ 400,000 nuclei were isolated using parameters that enriched for nuclei from cancer cells. A denoising autoencoding neural network was used to generate AE biomarkers for each nucleus. A systematic pipeline of preprocessing, normalization and conversion to case-level secondary features was applied to AE and HC features. The average of 50,000 bootstrapping models resulted in an AUC of 0.82 for the training and an average accuracy of 0.62 for the test cohort. The best 20 models were applied to the independent validation cohort of 25 cases and assigned each case to the M0 versus M1 groups by majority voting. This resulted in an accuracy of 72%.

## Conclusion:

We applied digital imaging technology and machine learning software to AE and HC features in order to predict M0 versus M1 stage from the tumor in PNBXs at diagnosis. Unexpectedly, hidden features in nuclei differed between M0 and M1 cases and succeeded to predict metastatic disease with 72% accuracy. The ultimate goal is to apply this inexpensive approach to develop prediction models of occult metastases and risk of future metastatic progression at the time of diagnosis in all patients with high-grade PC.

**Funding:** DOD PC131996, PCF-Movember GAP1 Unique TMAs Project, Prostate Cancer Foundation (PCF) Creativity Award, Jean Perkins Foundation, NIH/NCI P01 CA098912-09, NIH R01CA131255 and P50CA092131, Stephen Spielberg Team Science Award.