doi:10.1017/S143192762200410X
Microsc. Microanal. 28 (Suppl 1), 2022

Predicting Prostate Cancer Directly from Tissue Images using Deep Learning on Mass Spectrometry Imaging and Whole Slide Imaging Data

Md Inzamam Ul Haque[1], Debangshu Mukherjee[2], Sylwia A. Stopka[3,4], Nathalie Y.R. Agar[3,4,5], Jacob Hinkle[2*] and Olga S. Ovchinnikova[2*]

[1.] The Bredesen Center, University of Tennessee, Knoxville, TN, USA.
[2.] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
[3.] Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
[4.] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
[5.] Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA.
* Corresponding author: hinklejd@ornl.gov, ovchinnikovo@ornl.gov

Prostate cancer (PC) is one of the most common cancers globally and is the second most common cancer in the male population in the US. Integration of mass spectrometry imaging (MSI) and hematoxylin and eosin (H&E) data offers the potential to improve the identification of prostate cancer (PC). In 2015, [1] reported a data fusion framework for MSI and H&E stain microscopy enabling the prediction of a molecular distribution both at high spatial resolution and with high chemical specificity. [2] compared two pansharpening methods, Intensity–Hue–Saturation and Laplacian Pyramid, and demonstrated the latter was more robust for image fusion between MSI and electron microscopy. However, these fusion-based approaches are limited by the fundamental difference between physical mechanisms of image generation and are prone to reconstruction errors. Here, we have demonstrated that a physically constrained model between two different imaging modalities can accurately reconstruct and predict both high spatial resolution images and high spectral resolution mass spectra. We spatially register features obtained through deep learning from whole slide H&E-stained data with MSI data to correlate the chemical signature with the cellular morphology and then use the learned correlation to predict PC from observed H&E images using trained co-registered MSI data.

MSI data consisted of human prostate tissue specimens, cryosectioned and imaged at a pixel size of 120 μm using a 9.4 Tesla SolariX XR FT ICR MS (Bruker Daltonics, Billerica, MA). Corresponding high-resolution annotated H&E images of the same tissue were provided [3]. Raw data were converted to HDF5 format and binary masks for cancerous regions were extracted from the annotated images. Deep features were obtained from whole slide H&E data using the pretrained Resnet-50 model. Then the high-resolution features were downscaled and regridded to co-register with the low-resolution MSI data. Logistic regression has been used to predict PC directly from the H&E data using the learned correlation. Biomarkers for PC were also searched using *m/z* values from a Lipid Maps database.

A qualitative result is shown in Figure 1 which corresponds to the logistic regression results we achieved for two different tissue specimens. If we look closely in Figure 1B and Figure 1F, some secondary regions are revealed with the MSI to label prediction. These regions can be random noise or cancerous regions missing in the original annotation. When we look at the prediction for the second tissue specimen directly from H&E in Figure 1G, the secondary region seen with MSI prediction is revealed more. Interestingly, this region is revealed even more when we predict from predicted MSI components as seen in Figure 1H. For the other tissue specimen in Figure 1, cancerous regions are revealed well for all three predictions.

Figure 2 shows the result of H&E to MSI prediction for two different tissue specimens. We used a linear regression model with regridded H&E features as the input and MSI PCA modes as the labels. Out of 200 PCA components, prediction for the component having the highest R^2 is shown. We can see similar patterns in the predicted MSI images corresponding to the actual MSI images. Prediction for tissue specimen 2 as shown in Figure 2B is slightly better compared to the other tissue specimen. A single pixel is chosen from the cancerous region for each tissue specimen and the mass spectrum is plotted shown in Figure 2C, and Figure 2F. It is evident

that there is an intensity mismatch between the actual and predicted spectrum, but the prediction is clearly able to capture most of the *m/z* peaks in the actual spectrum.

Our deep learning methodology was able to address the limitations in predicting PC from H&E data. We have found that H&E can predict mass spectra somewhat accurately, indicating a correlation between features visible in optical H&E imaging and the chemical information present in MSI. We have also found that PC regions can be predicted reliably from MSI indicating that the mass spectra contain sufficient information for image segmentation. Combining these results, we verified that the overlapping information between modalities matches that needed for segmentation, by predicting cancerous regions directly from H&E as well as from predicted MSI. Moreover, we found two MSI biomarkers corresponding to specific masses that correctly identified the cancerous regions. Our approach shows the feasibility of using readily available H&E data to predict the rich chemical information available in MSI images. Although our training process requires paired data including both H&E and MSI data from the same samples, the resulting trained models could be relevant in clinical settings where only H&E is available. We have also identified some secondary regions in the prediction using MSI. These regions could be errors due to random noise or cancerous regions which were missing in the original pathology annotation. Additional validation of these secondary regions would be useful in future studies, for example using immunohistochemistry (IHC) imaging. By first predicting mass spectra, our method retains the ability to accurately reproduce labels, but adds additional useful chemical information, with relatively little manual training data required. These preliminary results along with a relative lack of MSI in public pathology datasets motivate the collection of larger paired H&E/MSI datasets in the future to support large-scale feature learning efforts for H&E analysis. This methodology lays the groundwork for developing more accurate predictions for PC in patients to improve patient care and trajectories.

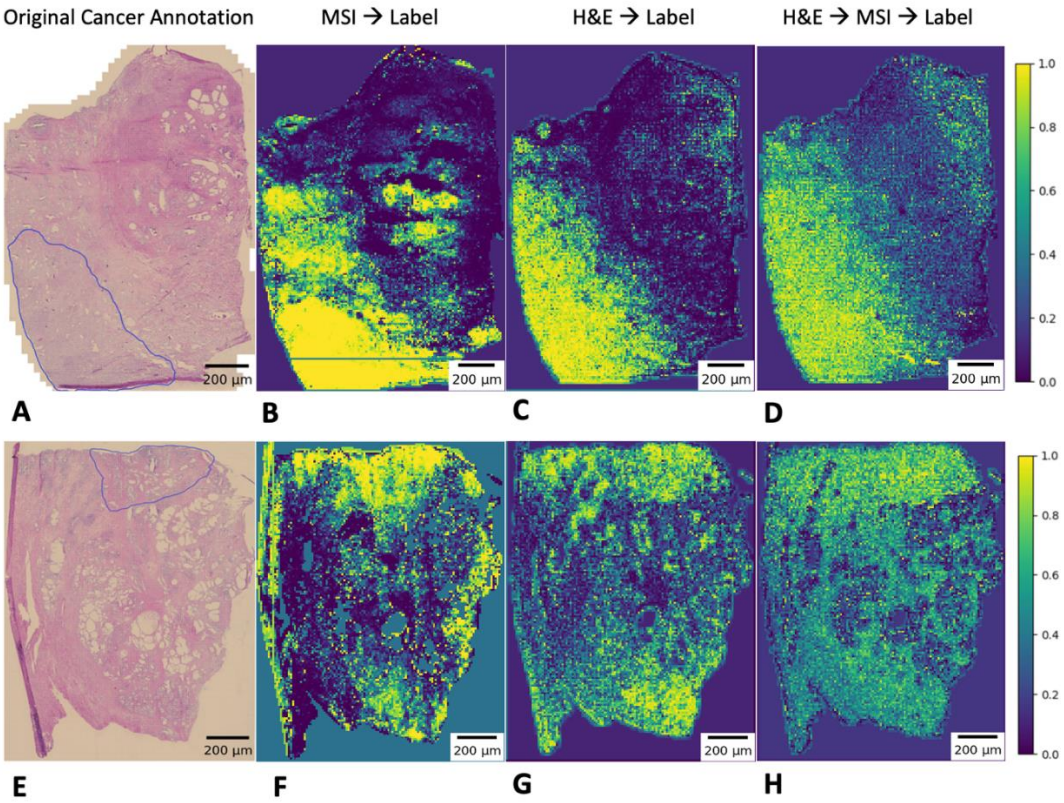

Figure 1. (A), and (**E)** are original cancer annotations for two different tissue specimens. **(B)**, and (**F)** show predicted cancer regions from MSI PCA modes for the two tissue specimens. (**C)**, and (**G)** show predicted cancer labels from regridded H&E features for the two tissue specimens. (**D)**, and (**H)** show prediction of

https://doi.org/10.1017/S143192762200410X Published online by Cambridge University Press

cancer labels from MSI predictions which is predicted from H&E features. All the predictions in this figure are achieved using logistic regression.

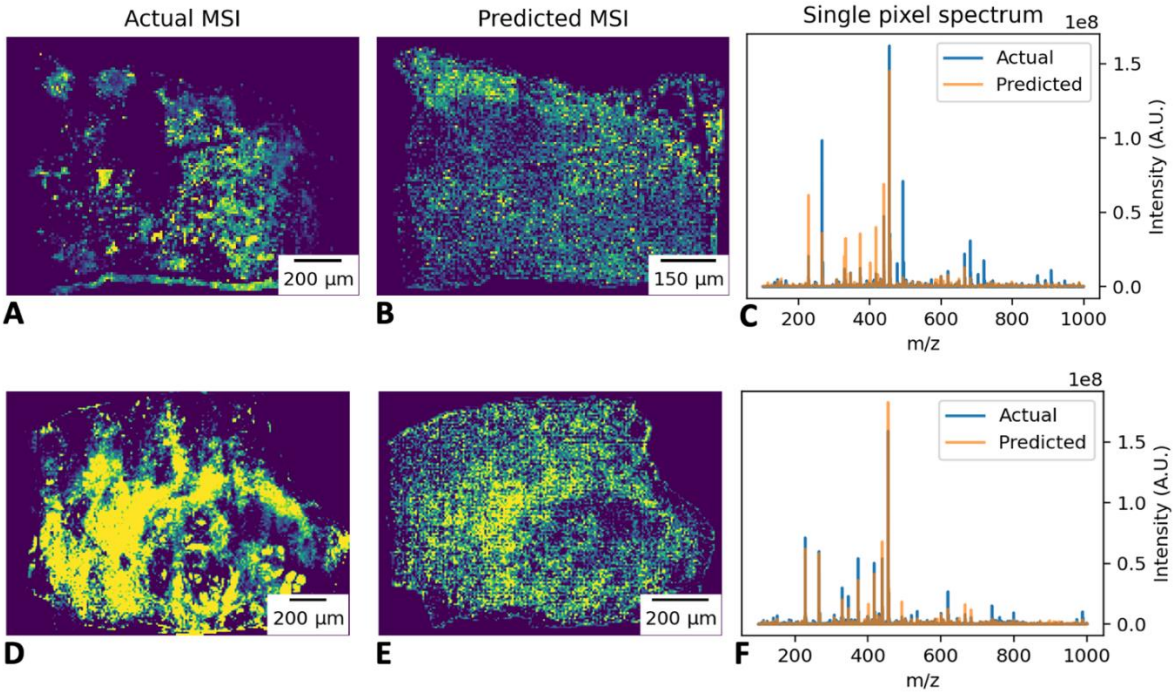

Figure 2. Prediction of MSI directly from regridded H&E features. **(A)**, and **(D)** show the first PCA components of MSI for two different tissue specimens. **(B)**, and **(E)** shows the corresponding predicted MSI PCA component directly from H&E. For the two tissue specimens, spectra of a single pixel from the cancerous region are shown both for actual and predicted images in **(C)**, and **(F)**. Negative values of the predicted spectra are clipped here.

References:

[1] R Van de Plas et al., Nat. Methods **12**(4) (20150, p. 366. doi: 10.1038/nmeth.3296
[2] F Vollnhals et al., Anal. Chem. **89**(20) (2017), p. 10702. doi: 10.1021/acs.analchem.7b01256
[3] EC Randall et al., Mol. Cancer Res. MCR **17**(5), p. 1155. doi: 10.1158/1541-7786.MCR-18-1057

https://doi.org/10.1017/S143192762200410X Published online by Cambridge University Press