AI-powered segmentation and analysis of nuclear morphology predicts genomic and clinical biomarkers in multiple cancer types

Abstract #464

John Abel[1], Suyog Jain[1], Deepta Rajan[1], Ken Leidal[1], Harshith Padigela[1], Aaditya Prakash[1], Jake Conway[1], Michael Nercessian[1], Christian Kirkup[1], Robert Egger[1], Ben Trotter[1], Andrew Beck[1], Ilan Wapinski[1], Michael G. Drage[1], Limin Yu[1], Amaro Taylor-Weiner[1]

[1]PathAI, Boston, MA.

STUDY BACKGROUND

Distortion to the nuclear envelope, such as altered size, shape, and morphology, is a common feature of cancer reflective of the underlying hallmark genomic instability[1,2]. Nuclear morphology is a common visual aid to diagnostic and prognostic pathology. Nuclei can be well-established markers of specific cancers; for example, a clear nucleus ("Orphan Annie Eye") is a known indication for papillary thyroid carcinoma[3]. Nuclear structure changes during mitosis, and distorted nuclei can indicate dysregulated replication processes, genetic mutations that affect stability and function of the nuclear envelope, aneuploidy, and genome instability[1]. Nuclear features have been found to correlate with prognosis in several cancer subtypes.

To enable the use of nuclear morphology in digital pathology, we developed a pan-tissue, deep-learning-based digital pathology pipeline for exhaustive nucleus detection, instance segmentation, and classification on whole-slide hematoxylin and eosin (H&E)-stained pathology images.

PATHAI MODEL

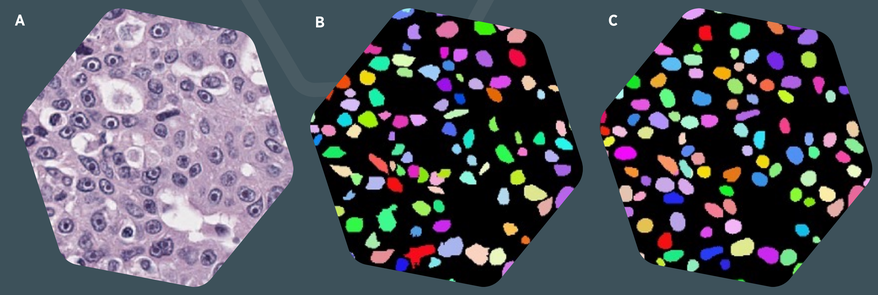


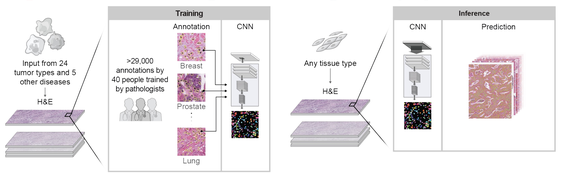
Figure 2. Example of model performance. A) Representative whole slide image patch from a patient with mesothelioma stained with H&E. B) Ground truth nuclei identified manually. C) Nuclei predicted by model. Each color represents a nucleus instance.

METHODS

Samples. Over 29,000 manual nucleus annotations were collected from H&E images from 21 tumor types at 40x and 20x magnification from The Cancer Genome Atlas (TCGA). A proprietary set of H&E-stained tissue biopsies of skin, liver non-alcoholic steatohepatitis, colon inflammatory bowel disease, and kidney lupus was also used.

Machine Learning Model and Deployment. Annotations were used to train an object detection and segmentation model for identifying cellular nuclei. Application of the model to held-out test data, including held-out tissue types, demonstrated performance comparable to models described in the literature (mean Dice score=0.80, aggregated Jaccard index=0.60). Our implementation enabled the application of the model to whole-slide images rather than sampling regions of interest from tissue samples, as is more commonly performed. We deployed our model on primary diagnostic H&E slides from the breast cancer (BRCA, N=892), prostate adenocarcinoma (PRAD, N=392), and lung adenocarcinoma (LUAD, N=426) TCGA cohorts. We extracted interpretable features describing the shape (circularity, eccentricity, area, and major and minor axis dimensions), size, staining intensity (mean and standard deviation), coloring, and texture of each nucleus. Nuclei were assigned as cancer or other cell types using separately trained convolutional neural networks. We used the mean and standard deviation (SD) of each feature across all cancer nuclei to summarize the cancer nuclear morphology on each slide, leading to the identification of nuclear human-interpretable features (nuHIFs).

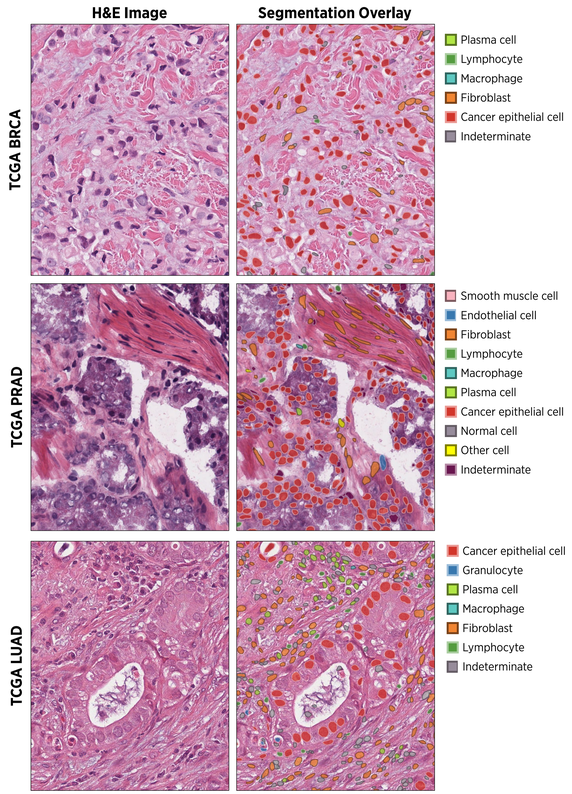
Figure 1. Machine learning model training and deployment.



Predictive Analysis. We used nuclear features to construct interpretable logistic regression classification models for predicting whole genome doubling (WGD), homologous recombination deficiency (HRD[4]), and markers of prognosis. For each dataset, a multivariate logistic regression classification model was used with elastic net regularization. A 67%/33% stratified train/test split was used.

RESULTS

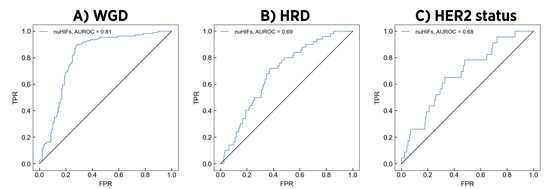
Figure 3. Nuclear segmentation and cell type identification.



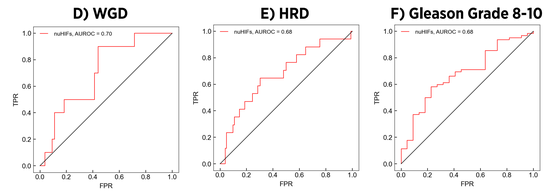
Representative H&E images of breast cancer (TCGA BRCA), prostate adenocarcinoma (TCGA PRAD), and lung adenocarcinoma (TCGA LUAD) are shown at 40X magnification, with and without nuclear segmentation. Colors indicate identified cell types for each image.

RESULTS

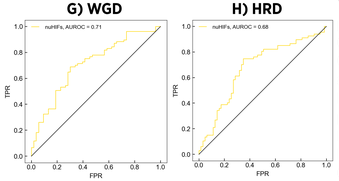
Figure 4. nuHIFs predict features of several cancer subtypes.



TCGA BRCA (N=776). A) nuHIFs predictive of WGD were Mean Cancer Nuclear Eccentricity, SD Cancer Nuclear Major Axis Length, and SD Cancer Nuclear Minor Axis Length (AUROC=0.81). B) nuHIFs predictive of HRD were SD Cancer Nuclear Perimeter, SD Cancer Nuclear Mean Color ("Purple-ness"), and SD Cancer Nuclear Minor Axis Length (AUROC=0.69). C) nuHIFs predictive of HER2 status were SD Cancer Nuclear Minor Axis Length, SD Cancer Nuclear Eccentricity, and SD Cancer Nuclear Solidity (AUROC=0.68).



TCGA PRAD (N=366). D) nuHIFs predictive of WGD were SD Cancer Nuclear Major Axis Length, Mean Cancer Nuclear Grayscale Intensity, and SD Cancer Nuclear Mean Color (AUROC=0.70). E) nuHIFs predictive of HRD were Mean Cancer Nuclear Eccentricity, SD Cancer Nuclear Major Axis Length, and SD Cancer Nuclear Minor Axis Length (AUROC=0.68). F) nuHIFs predictive of Gleason grade 8-10 were SD Cancer Nuclear Grayscale Texture (SD Intensity), SD Cancer Nuclear Mean Color, and Mean Cancer Nuclear Mean Color (AUROC=0.68; N=264).



TCGA LUAD (N=401). G) nuHIFs predictive of WGD were Mean Cancer Nuclear Eccentricity, SD Cancer Nuclear Major Axis Length, and SD Cancer Nuclear Minor Axis Length (AUROC=0.71). H) nuHIFs predictive of HRD were SD Cancer Nuclear Perimeter, SD Cancer Nuclear Mean Color, and SD Cancer Nuclear Minor Axis Length (AUROC=0.68).

CONCLUSIONS

We have developed a powerful pan-tissue approach for nucleus segmentation and featurization on entire whole-slide images. This method enables the construction of predictive models and the identification of features linking nuclear morphology with clinically-relevant prognostic biomarkers, such as WGD and HRD across multiple cancer types, including BRCA, PRAD, and LUAD. These results highlight the potential of machine learning-guided nuclear morphometry as a prognostic tool for cancer pathologists.

CONTACT

Ilan Wapinski
Email: ilan.wapinski@pathai.com

REFERENCES

1. Chow, K., R.E. Factor, and K.S. Ullman. (2012). The nuclear envelope environment and its cancer connections. *Nat Rev Cancer*. **12**(3):196-209.
2. Fischer, E.G. (2020) Nuclear morphology and the biology of cancer cells. *Acta Cytol*. **64**(6):511-519..
3. Hapke, M.R. and L.P. Dehner. (1979). The optically clear nucleus. A reliable sign of papillary carcinoma of the thyroid? *Am J Surg Pathol*. **3**(1):31-38.
4. Telli, M.L., K.M. Timms, J. Reid, et al. (2016). Homologous Recombination Deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin Cancer Res*. **22**(15):3764-3773.

ACKNOWLEDGMENTS

The results shown here are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

This poster template was developed by SciStories LLC. https://scistories.com/

PathAI

AACR Annual Meeting, April 8-13, 2022
New Orleans, Louisiana, USA