Ground Truth Labeling: Can Digital Pathology AI be used to Improve Radiology AI?

Stephanie Harmon, PhD

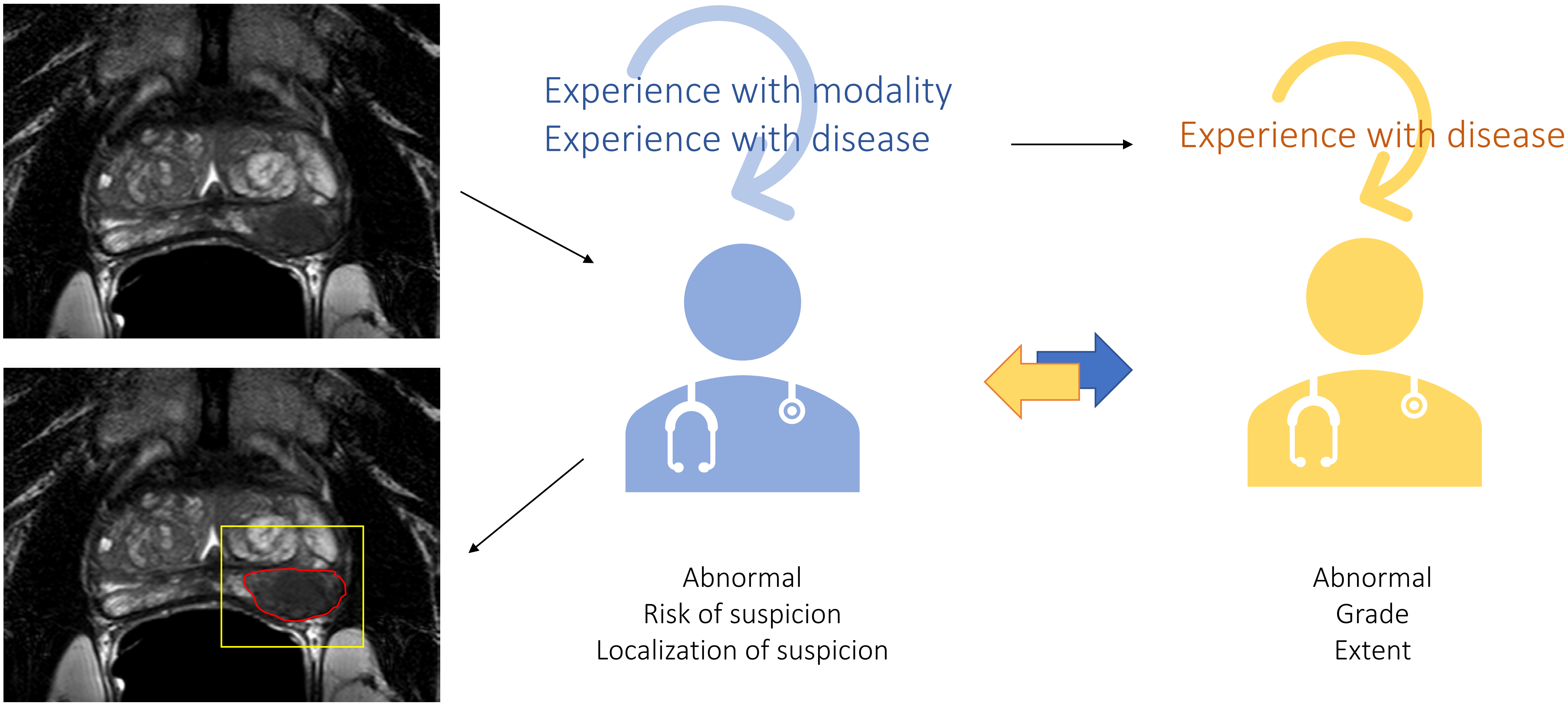
Clinical Research Directorate, Frederick National Laboratory for Cancer Research
sponsored by the National Cancer Institute

Molecular Imaging Program, NCI NIH

MIP
MOLECULAR IMAGING PROGRAM

leidos
Leidos Biomedical Research, Inc.

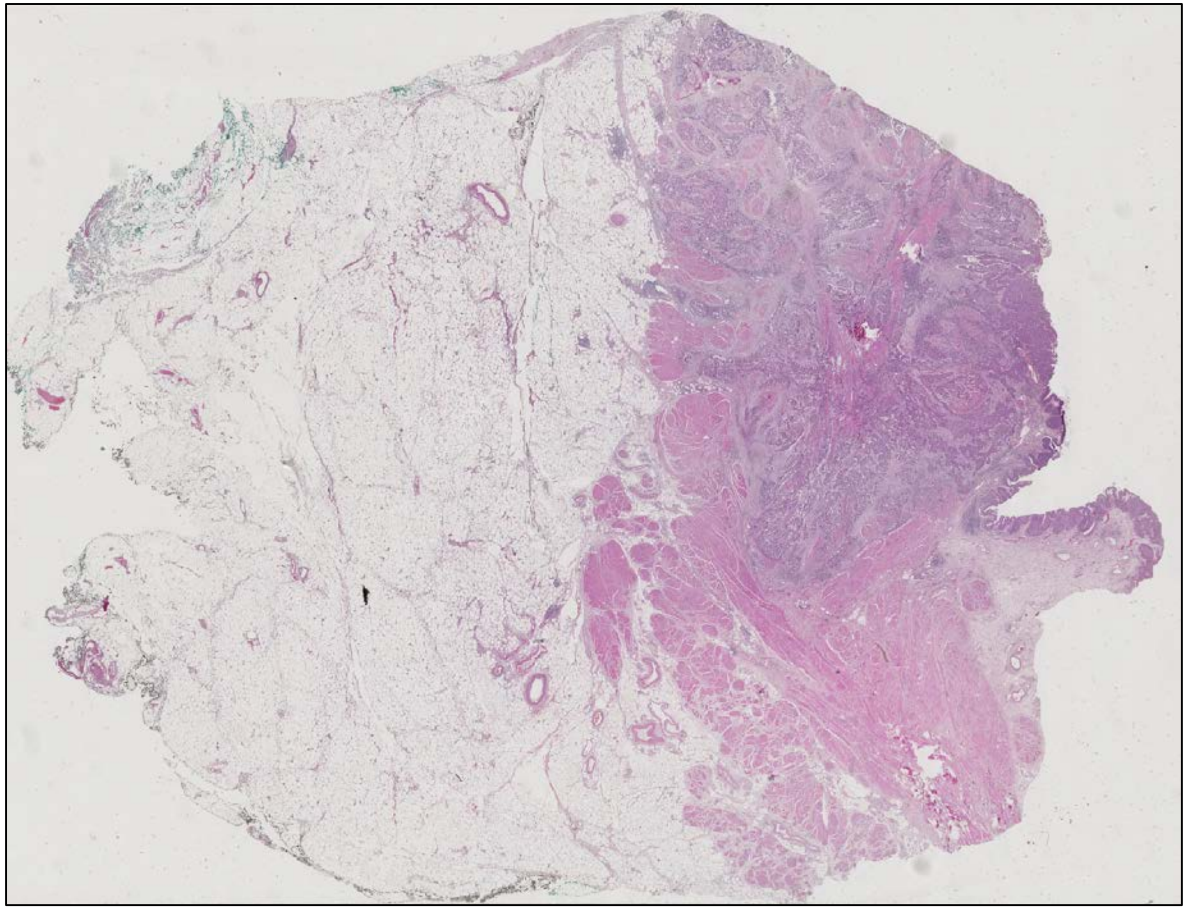
How do we define ground truth?



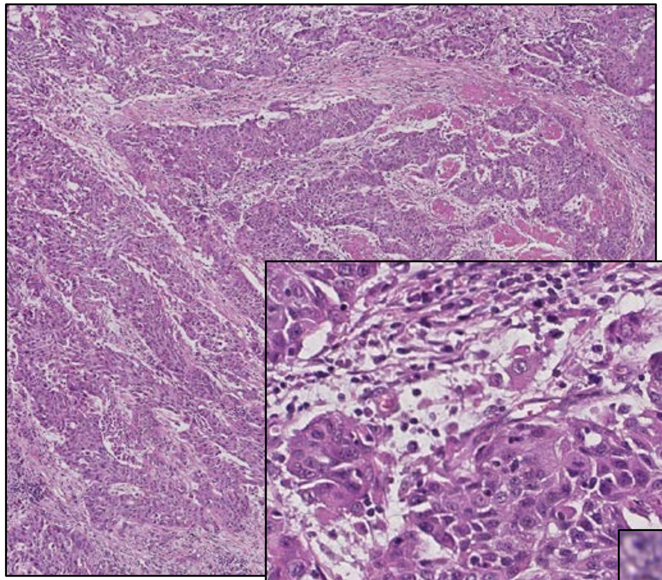
Radiology AI can depend on labeling/experience in both realms

Information at every level

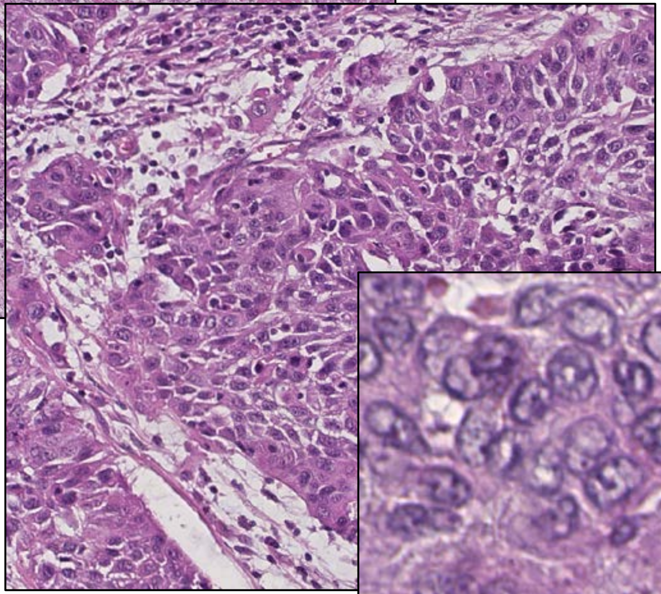
Burden



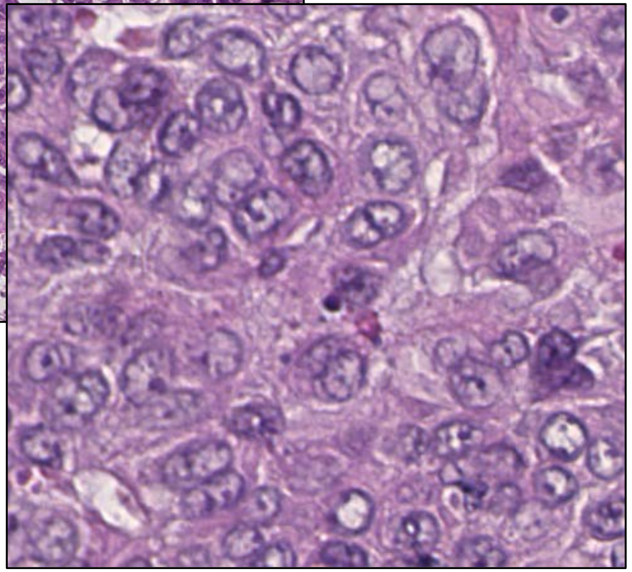
Histology



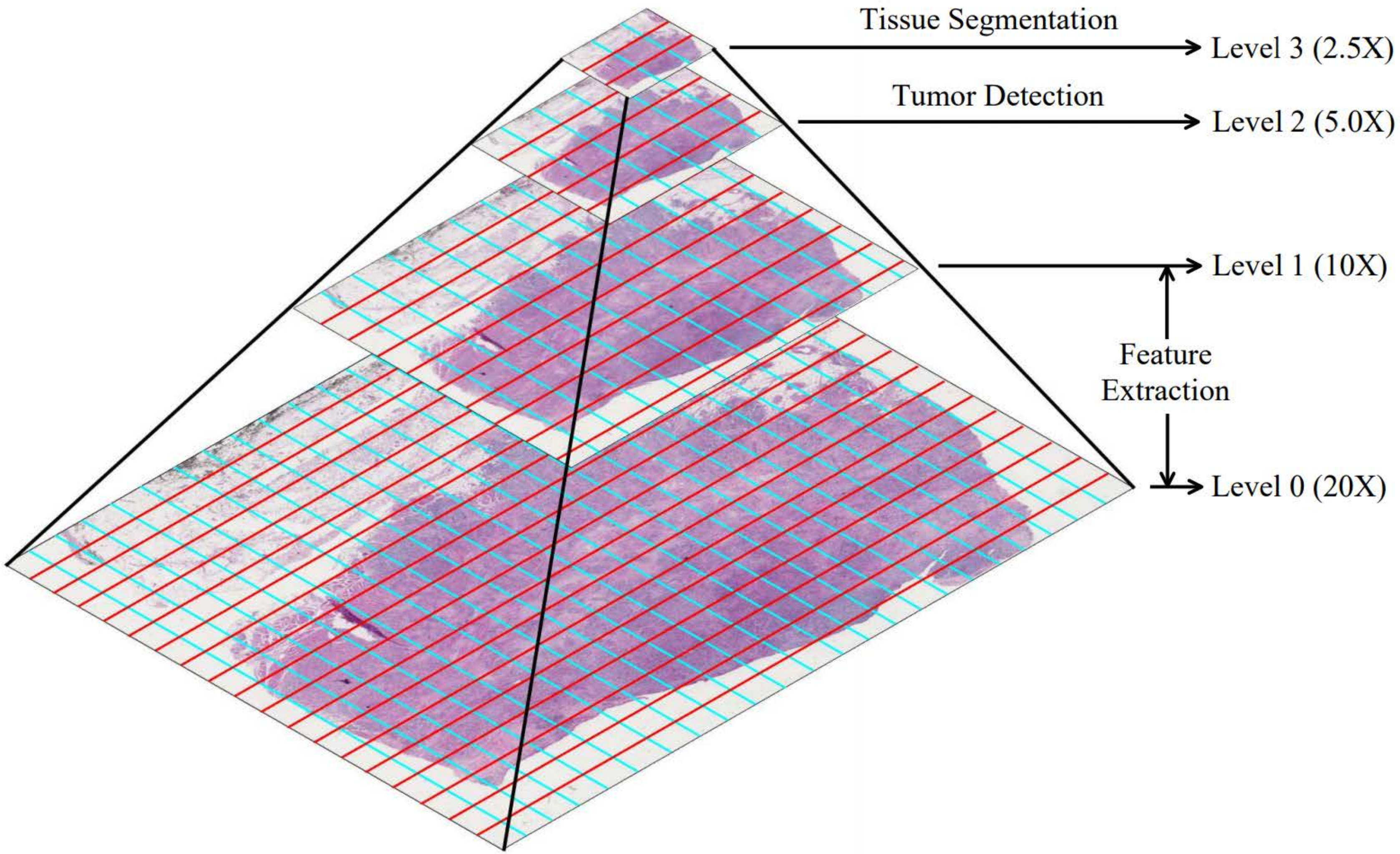
Architecture



Nuclear Grade

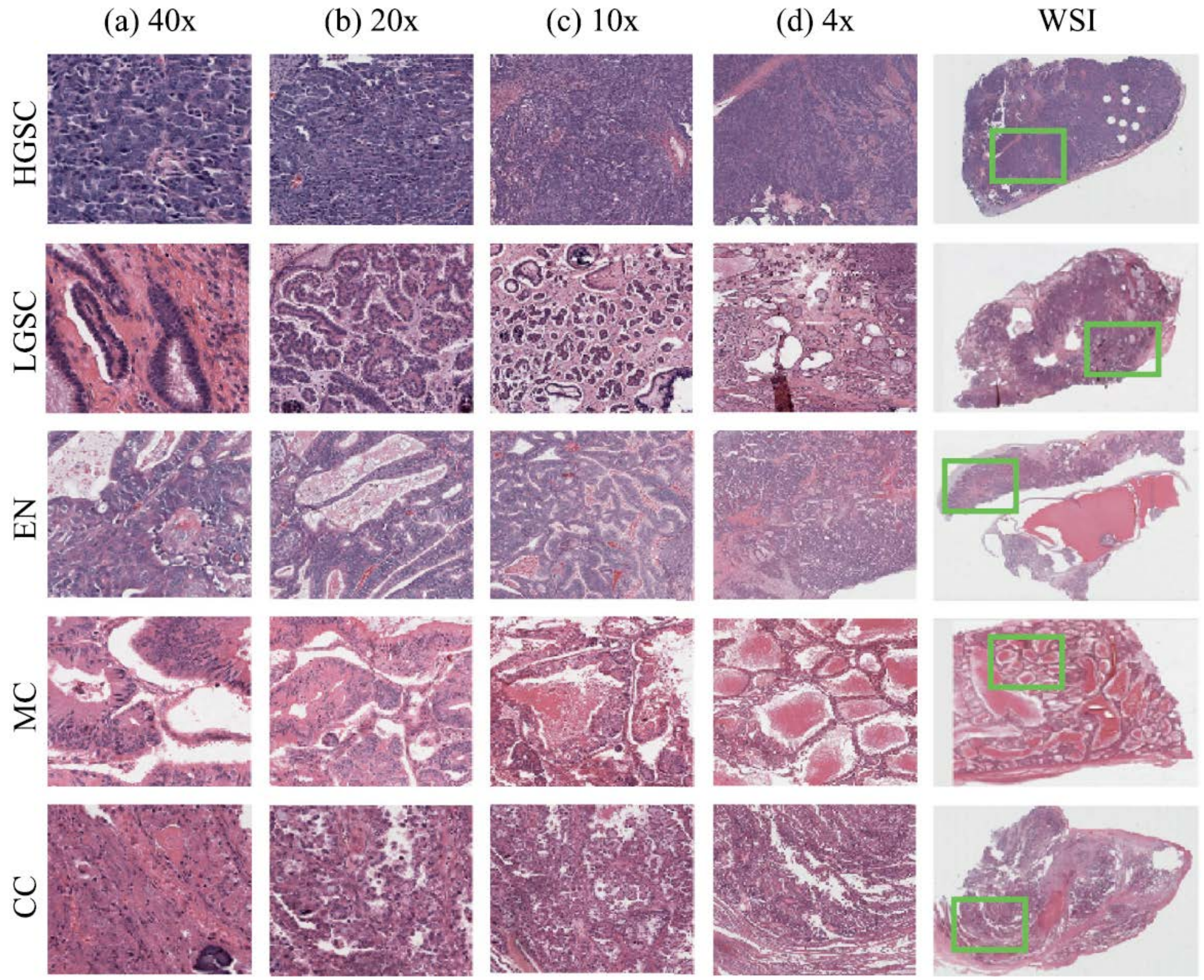


Digital pathology image structure



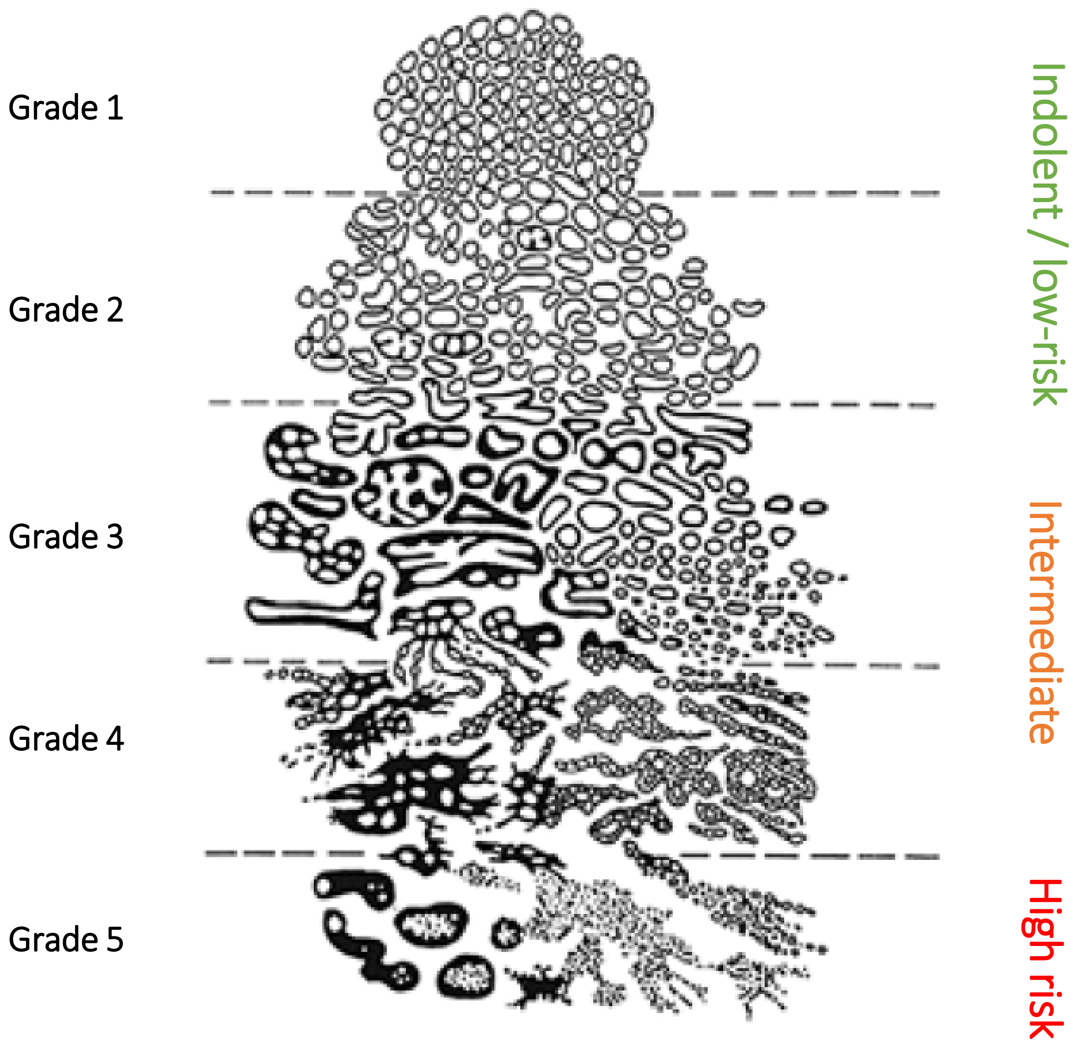
Xu et al bioRxiv 2019

Information at every level



BenTaieb et al Med Img Analysis 2017

Pathological grading of prostate cancer



- **Gleason Grading characterizes risk well on a population-level, but does little to inform on the patient-level**

➢ Poor inter/intra-observer reproducibility

➢ Gleason pattern 4 contain a biologically heterogenous group of cancers

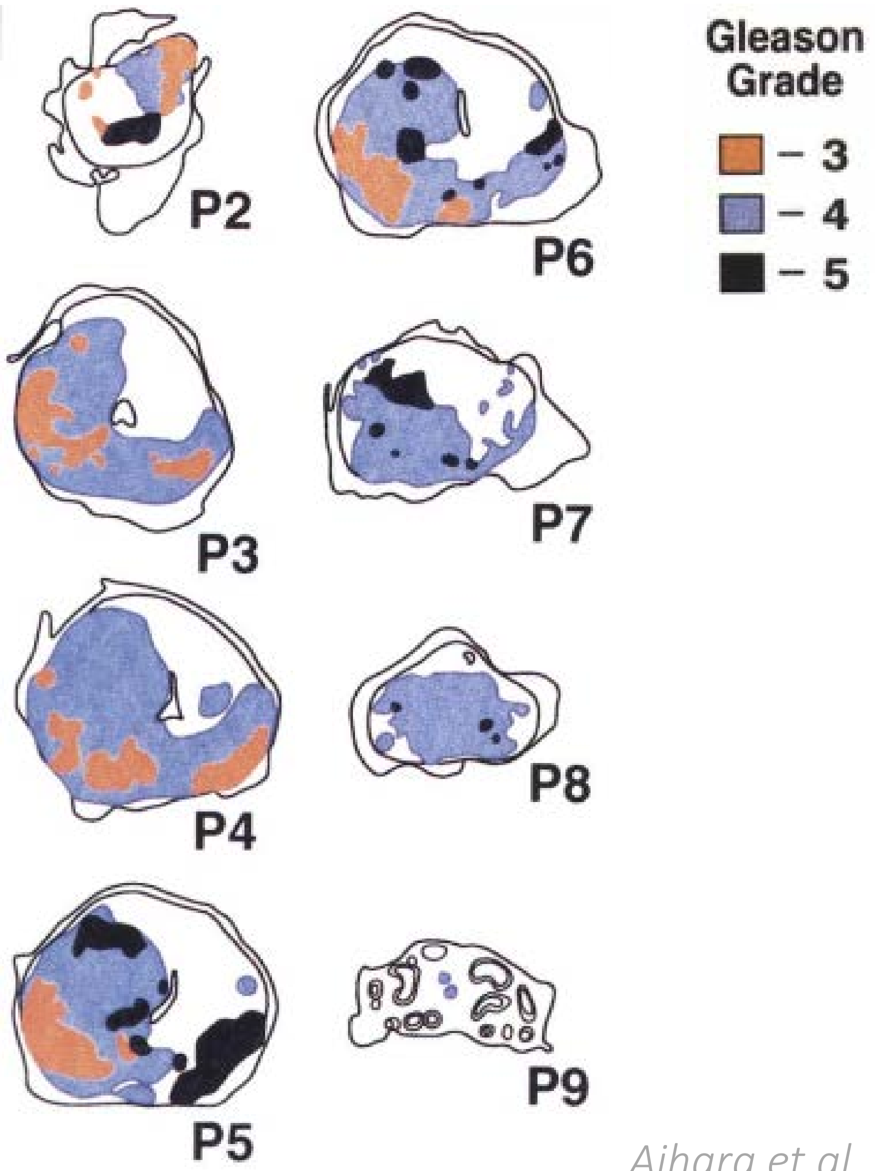
Modified from Epstein et al, A J Surg Path 2016

The devil's in the details: heterogeniety

- 20-30% of men undergoing radical prostatectomy will be "upgraded" to higher grade compared to pre-surgical biopsy *Siddiqui et al, JAMA 2015*

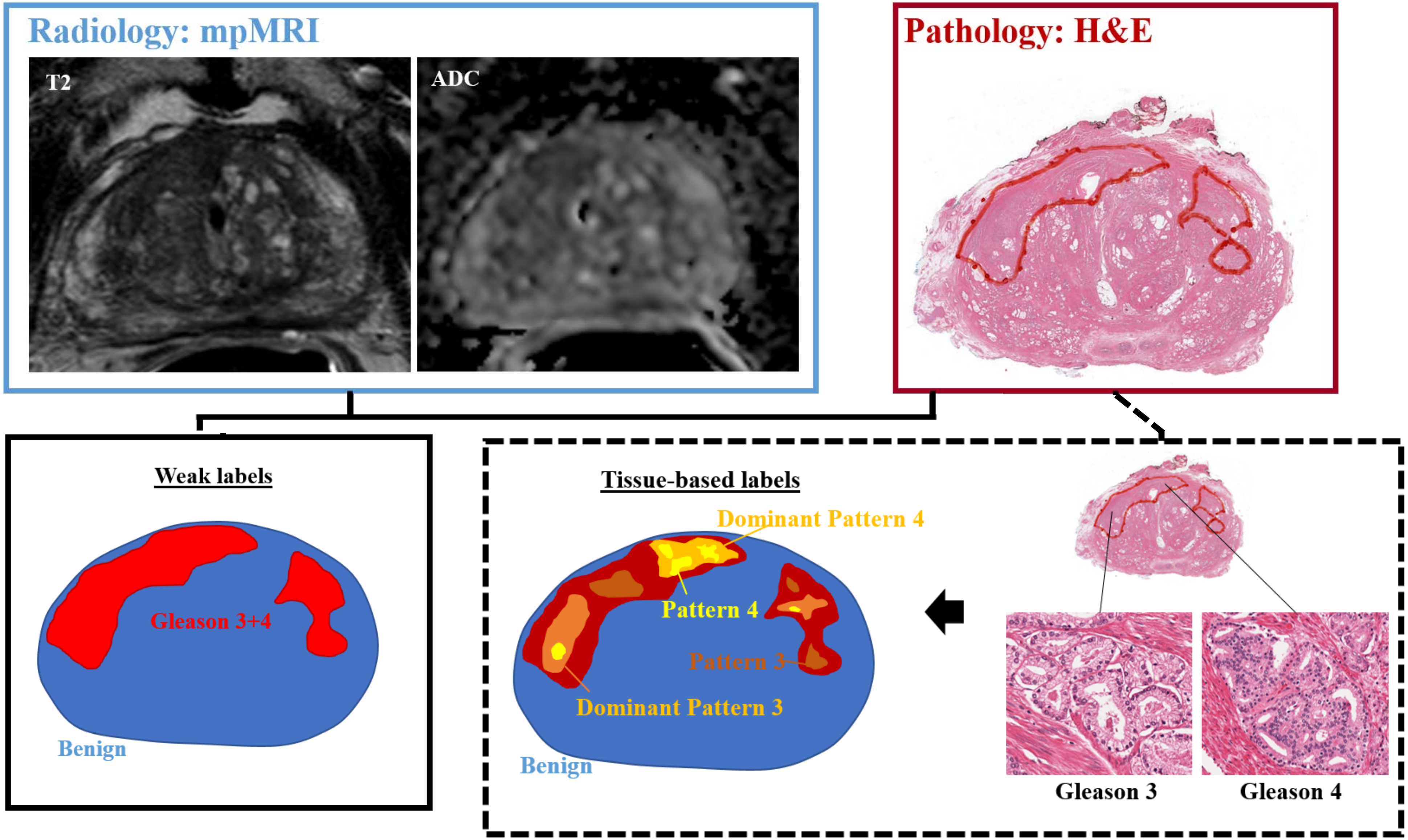
- Sampling bias due to intra-tumoral heterogeneity

  Aihara et al, Urology 1994
  El-Shater et al, Pros Can Pros Dis 2016
  Mesko et al, Am J Clin Onc 2016



Aihara et al, Urology 1994

The devil's in the details: heterogeniety



Harmon et al, DIR 2019

How to improve our labeling for radiology?

- Gland-level pathology annotation is prohibitively slow → can we use AI?

SCIENTIFIC REPORTS

OPEN

Automated Gleason grading of prostate cancer tissue microarrays via deep learning

Received: 22 March 2018
Accepted: 19 July 2018
Published online: 13 August 2018

Eirini Arvaniti[1,5], Kim S. Fricker[2], Michael Moret[1], Niels Rupp[2], Thomas Hermanns[3], Christian Fankhauser[3], Norbert Wey[2], Peter J. Wild[2,4], Jan H. Rüschoff[2] & Manfred Claassen[1,5]

IEEE TRANSACTIONS ON MEDICAL IMAGING 1

Path R-CNN for Prostate Cancer Diagnosis and Gleason Grading of Histological Images

Wenyuan Li, *Student Member, IEEE,* Jiayun Li, Karthik V. Sarma, King Chung Ho, Shiwen Shen, Beatrice S. Knudsen, Arkadiusz Gertych, and Corey W. Arnold

arXiv.org > cs > arXiv:1811.06497

Search or Article ID
(Help | Advanced search)

Computer Science > Computer Vision and Pattern Recognition

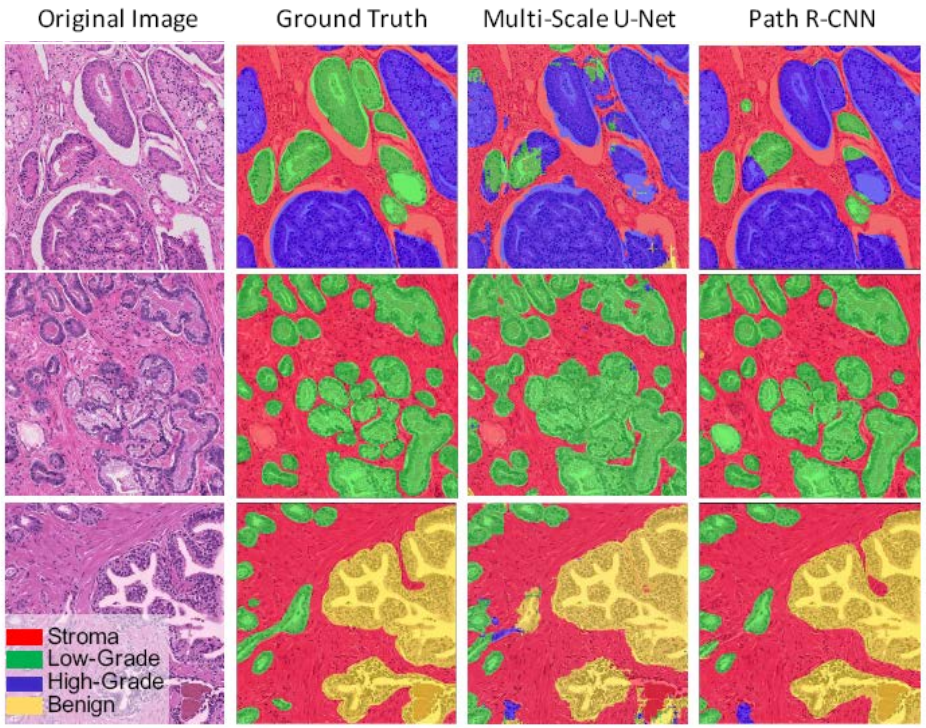
Development and Validation of a Deep Learning Algorithm for Improving Gleason Scoring of Prostate Cancer

Kunal Nagpal, Davis Foote, Yun Liu, Po-Hsuan (Cameron)Chen, Ellery Wulczyn, Fraser Tan, Niels Olson, Jenny L. Smith, Arash Mohtashamian, James H. Wren, Greg S. Corrado, Robert MacDonald, Lily H. Peng, Mahul B. Amin, Andrew J. Evans, Ankur R. Sangoi, Craig H. Mermel, Jason D. Hipp, Martin C. Stumpe

(Submitted on 15 Nov 2018)

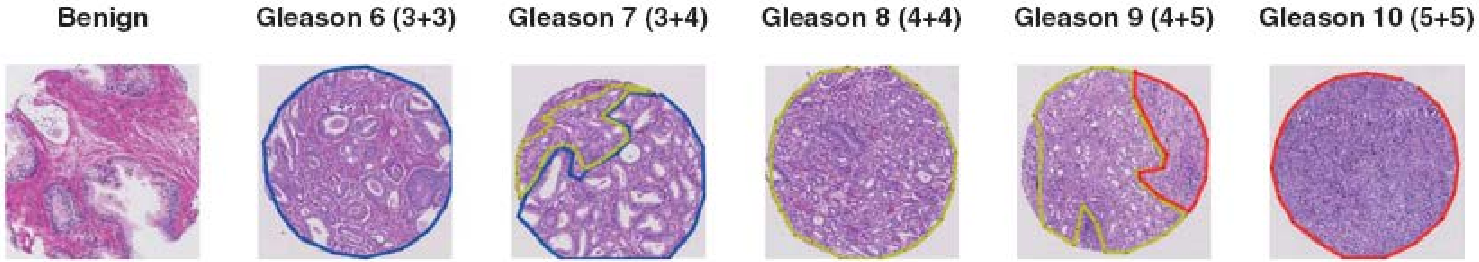
AI for pathologic grading in prostate cancer

Li et al, IEEE TMI 2018

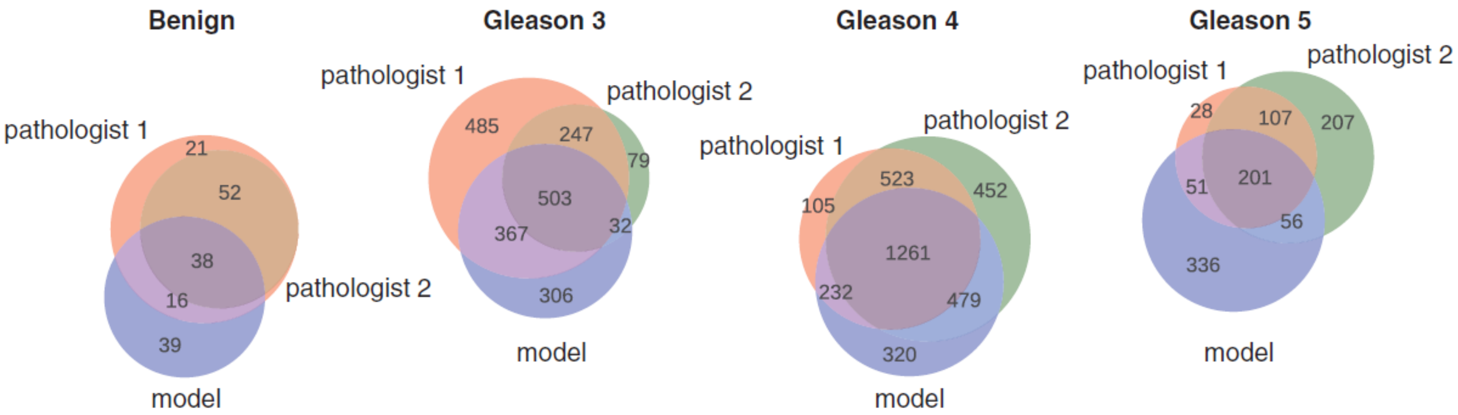


Performance: 71-83% for benign to high grade
Pros: highly trained expert
Cons: small dataset (<1000 tiles from 40 patients)

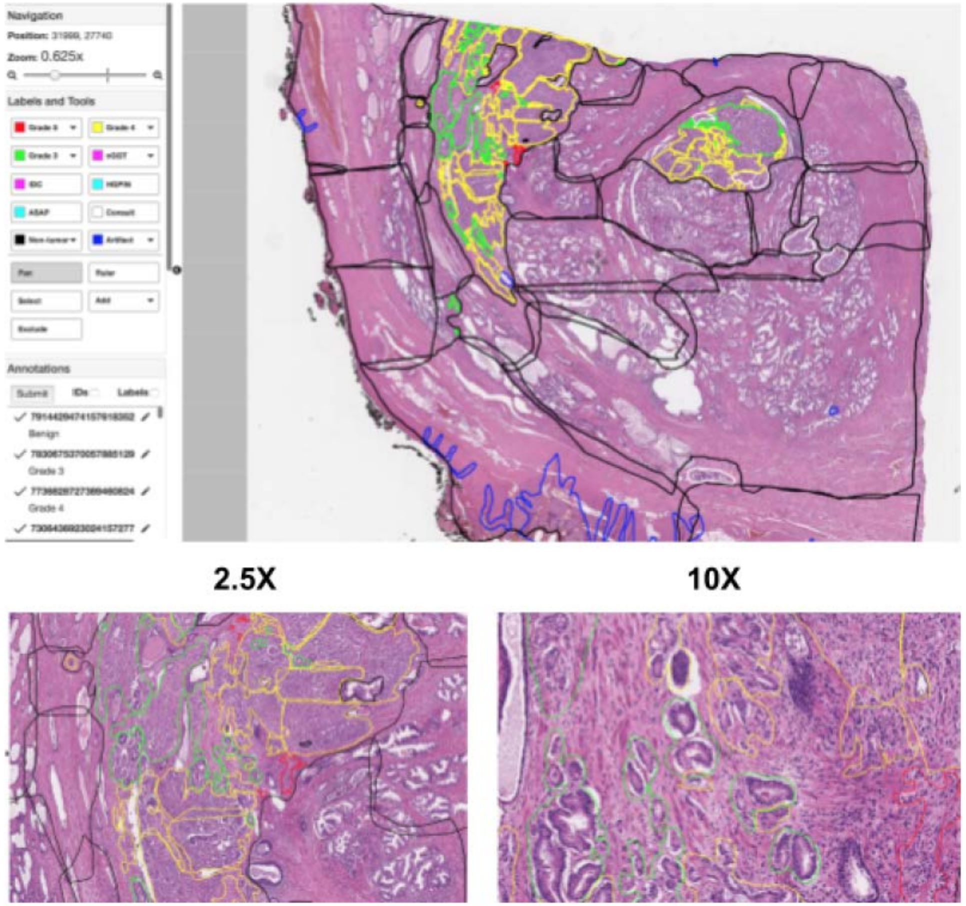
Arvanti et al, Sci Reps 2018

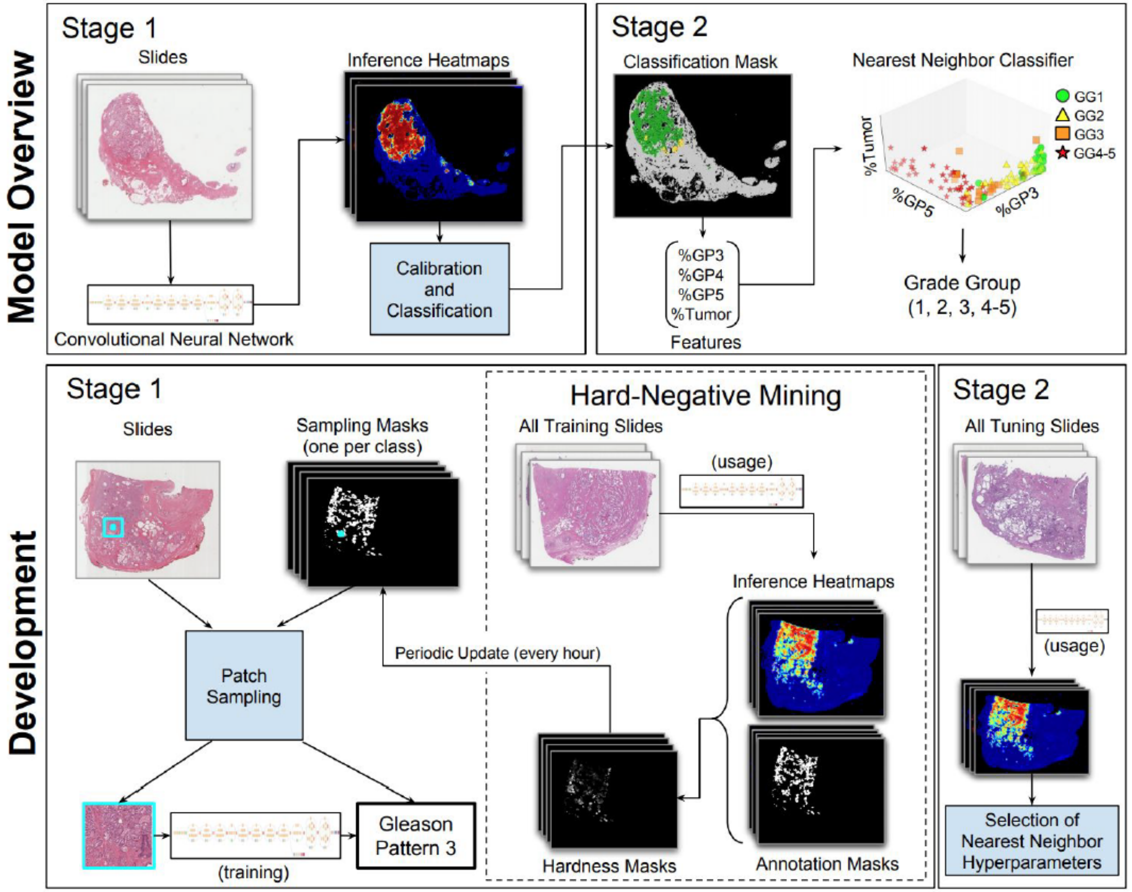


Performance: 70% classification
Pros: large dataset
Cons: lack of tissue heterogeneity in tissue microarrays



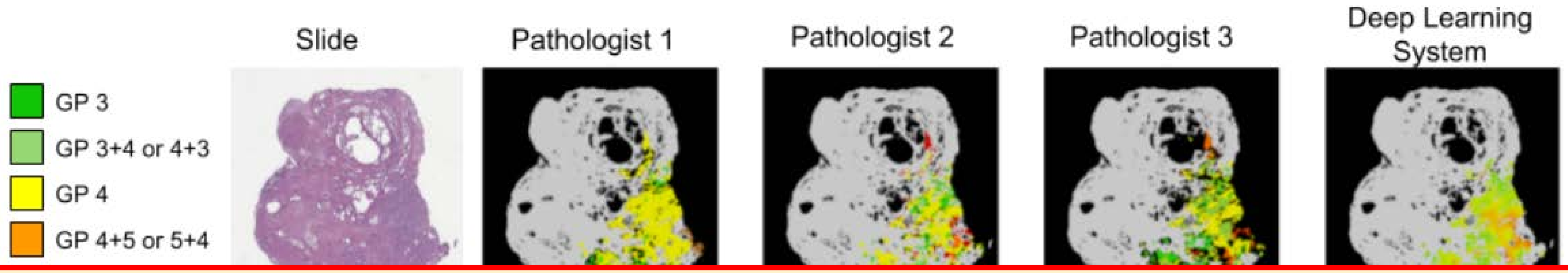
AI for pathologic grading in prostate cancer





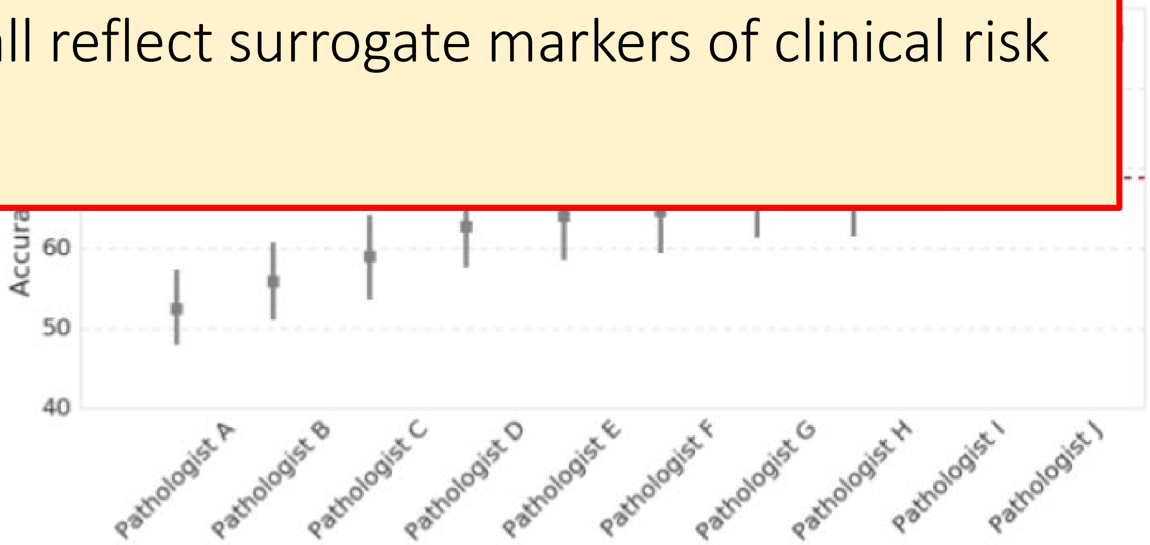
Nagpal et al, arXiv 2018

AI for pathologic grading in prostate cancer



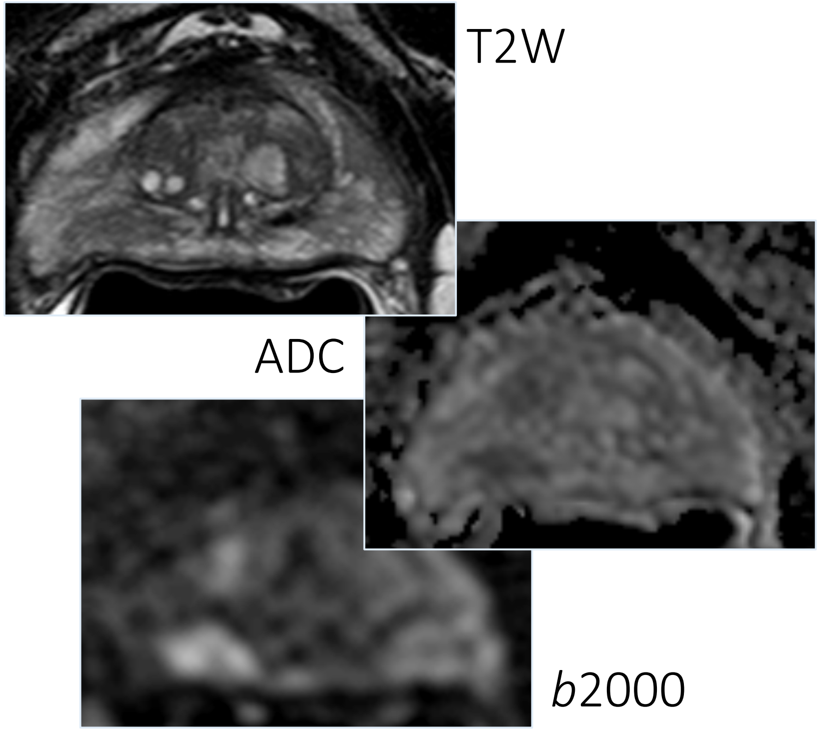
What are we trying to accomplish?

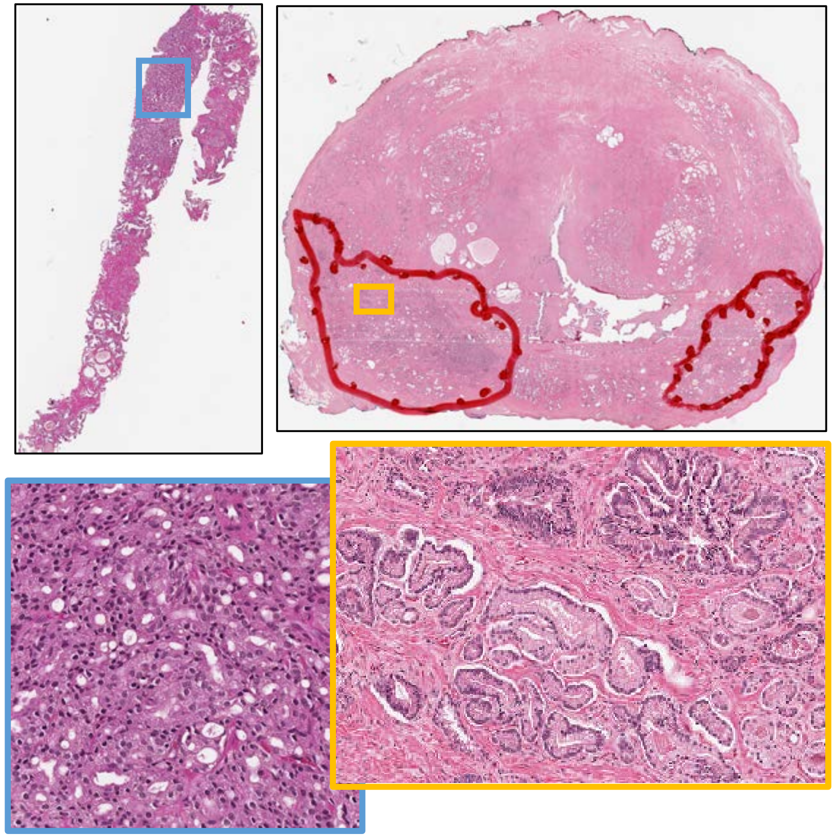
Expert-trained systems all reflect surrogate markers of clinical risk



Nagpal et al, arXiv 2018

The goal is the same: outcome





Detection
Risk Scoring

Biopsy guidance

Staging

Detection
Grading

Clinically-relevant outcomes

Treatment decision
- need for treatment
- type of treatment
- extent of treatment

Treatment response
- surgical outcomes
- disease monitoring

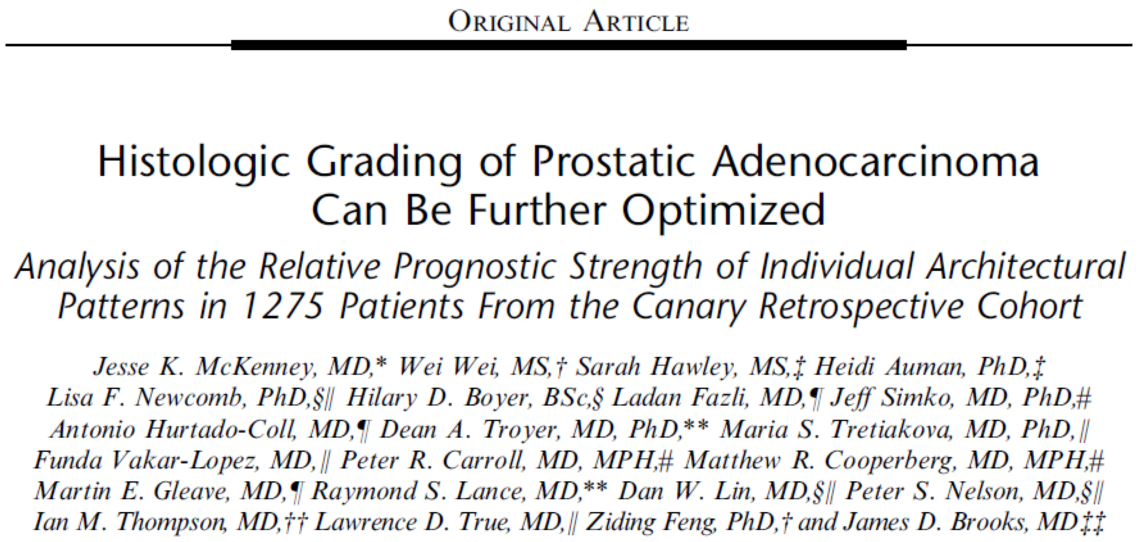
Treatment outcome
- time to failure
- drug sensitivity

Metastatic potential
- time to progression
- detection of mets

Patient outcome
- morbidity
- mortality

Predicting features of poor outcome: with an expert

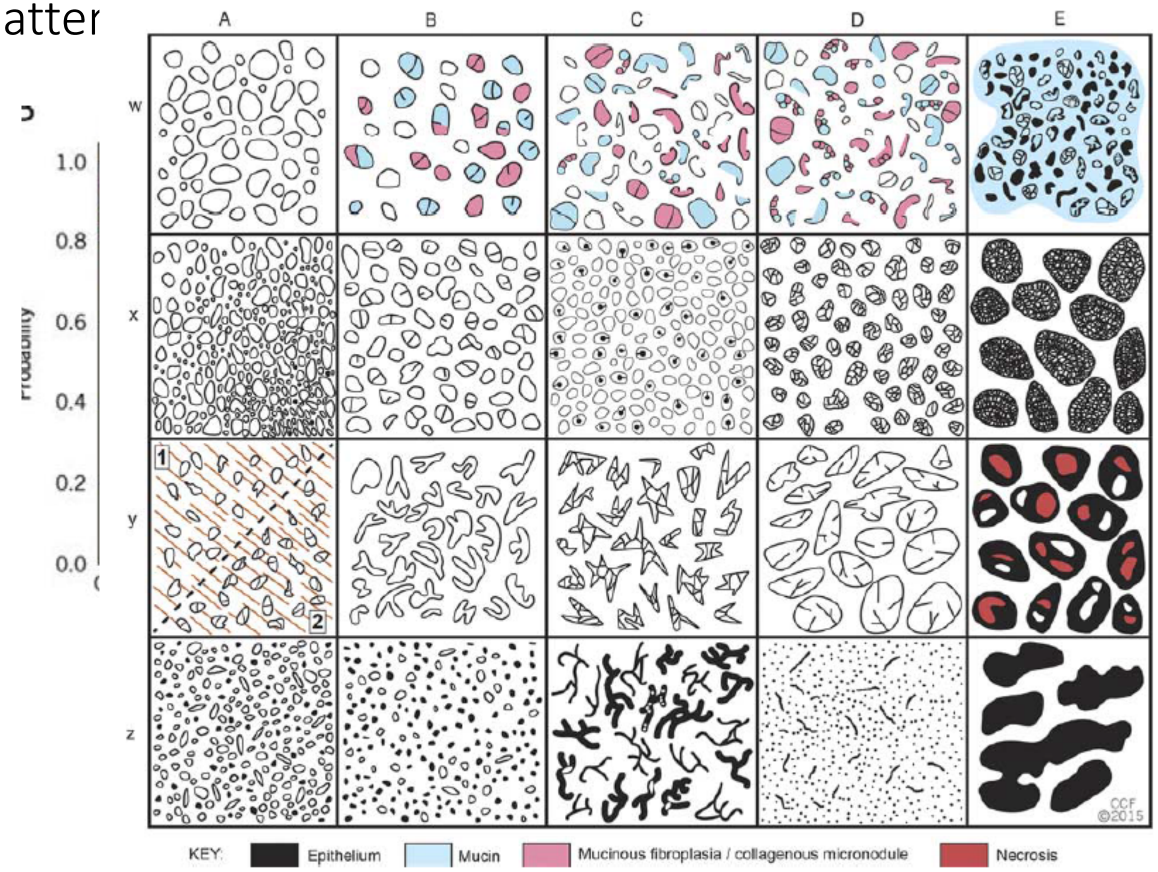
- Gleason score assignment by current consensus guidelines are not entirely optimized for clinical risk stratification
    - ➢ Focal poorly formed gland and cribriform patter


ORIGINAL ARTICLE

Histologic Grading of Prostatic Adenocarcinoma Can Be Further Optimized

Analysis of the Relative Prognostic Strength of Individual Architectural Patterns in 1275 Patients From the Canary Retrospective Cohort

Jesse K. McKenney, MD, Wei Wei, MS,† Sarah Hawley, MS,‡ Heidi Auman, PhD,‡ Lisa F. Newcomb, PhD,§|| Hilary D. Boyer, BSc,§ Ladan Fazli, MD,¶ Jeff Simko, MD, PhD,# Antonio Hurtado-Coll, MD,¶ Dean A. Troyer, MD, PhD,** Maria S. Tretiakova, MD, PhD,|| Funda Vakar-Lopez, MD,|| Peter R. Carroll, MD, MPH,# Matthew R. Cooperberg, MD, MPH,# Martin E. Gleave, MD,¶ Raymond S. Lance, MD,** Dan W. Lin, MD,§|| Peter S. Nelson, MD,§|| Ian M. Thompson, MD,†† Lawrence D. True, MD,|| Ziding Feng, PhD,† and James D. Brooks, MD‡‡*



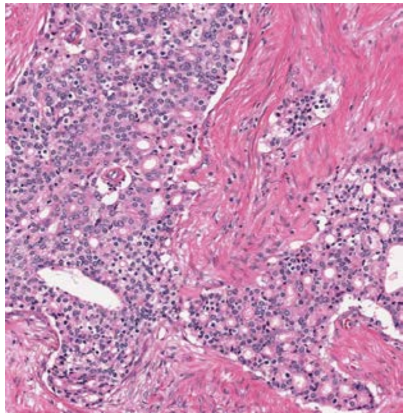
Modified from McKenney et al, A J Surg Path 2016

Predicting features of poor outcome: with an expert

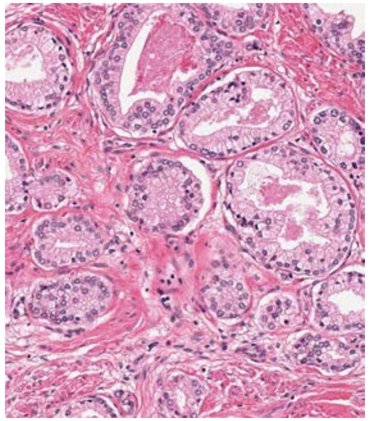
- 5394 local annotations of all histologic architectures (N=21 patients)
- 8 histological patterns independently associated with poor outcome

Sliding-window based binary classification

High risk



Low/Intermediate



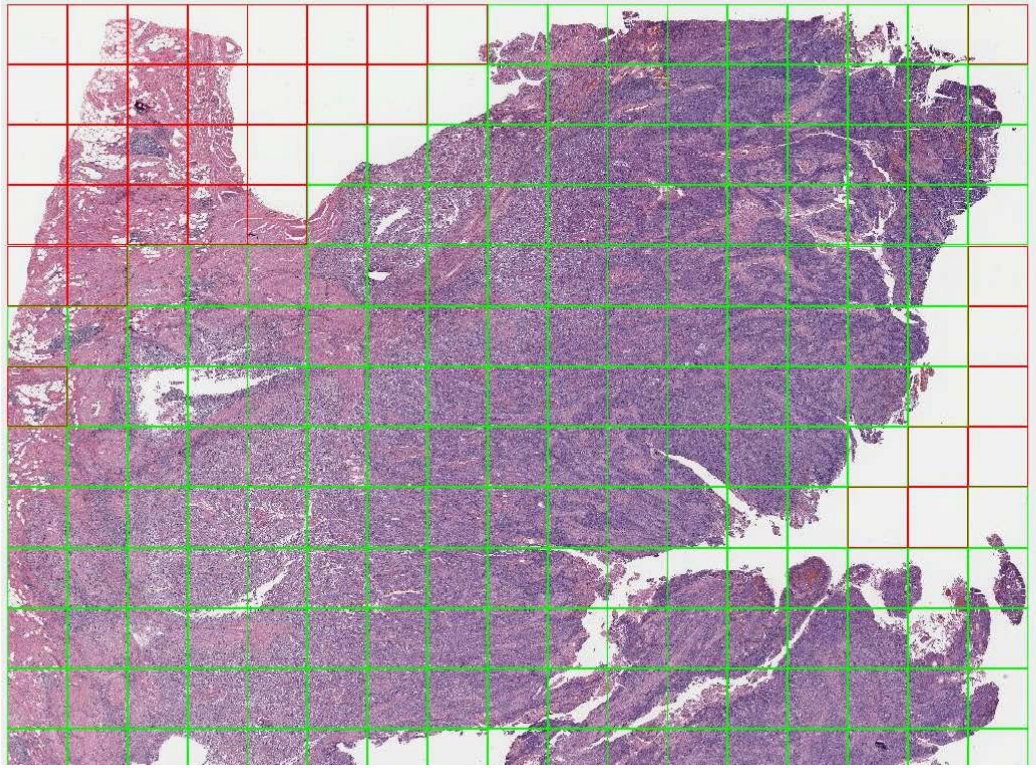
Patch-level validation set accuracy: 95%
Patch-level test set accuracy: 94%

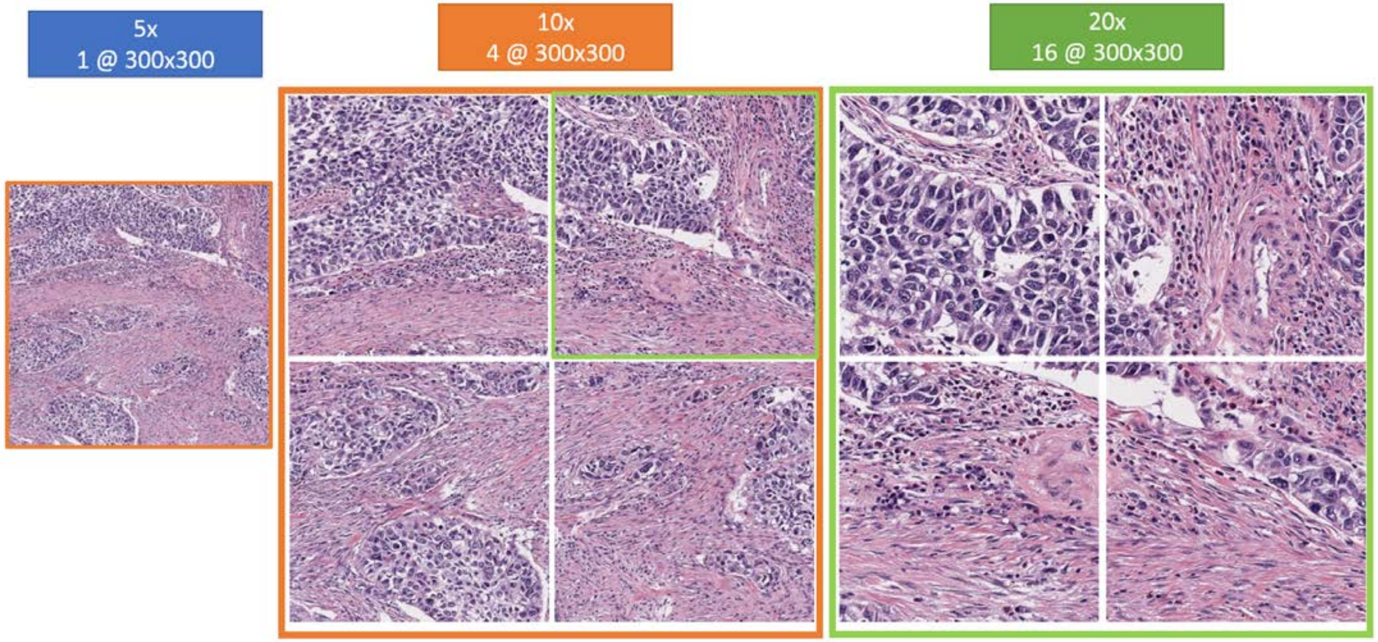
Harmon Turkbey McKenney, unpublished

Direct prediction of poor outcome

– Can we predict metastatic potential directly from the images?

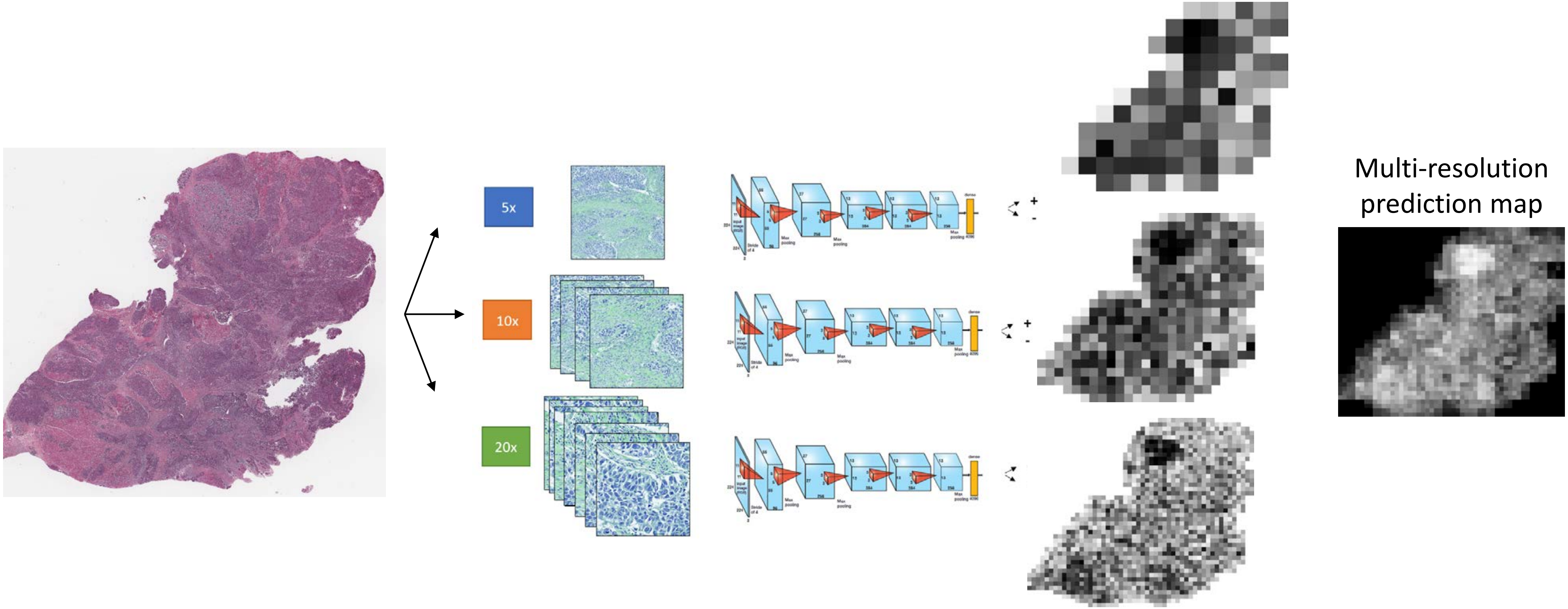
– TCGA: 287 bladder cystectomy samples with digital imaging + pathologic outcomes from 26 centers





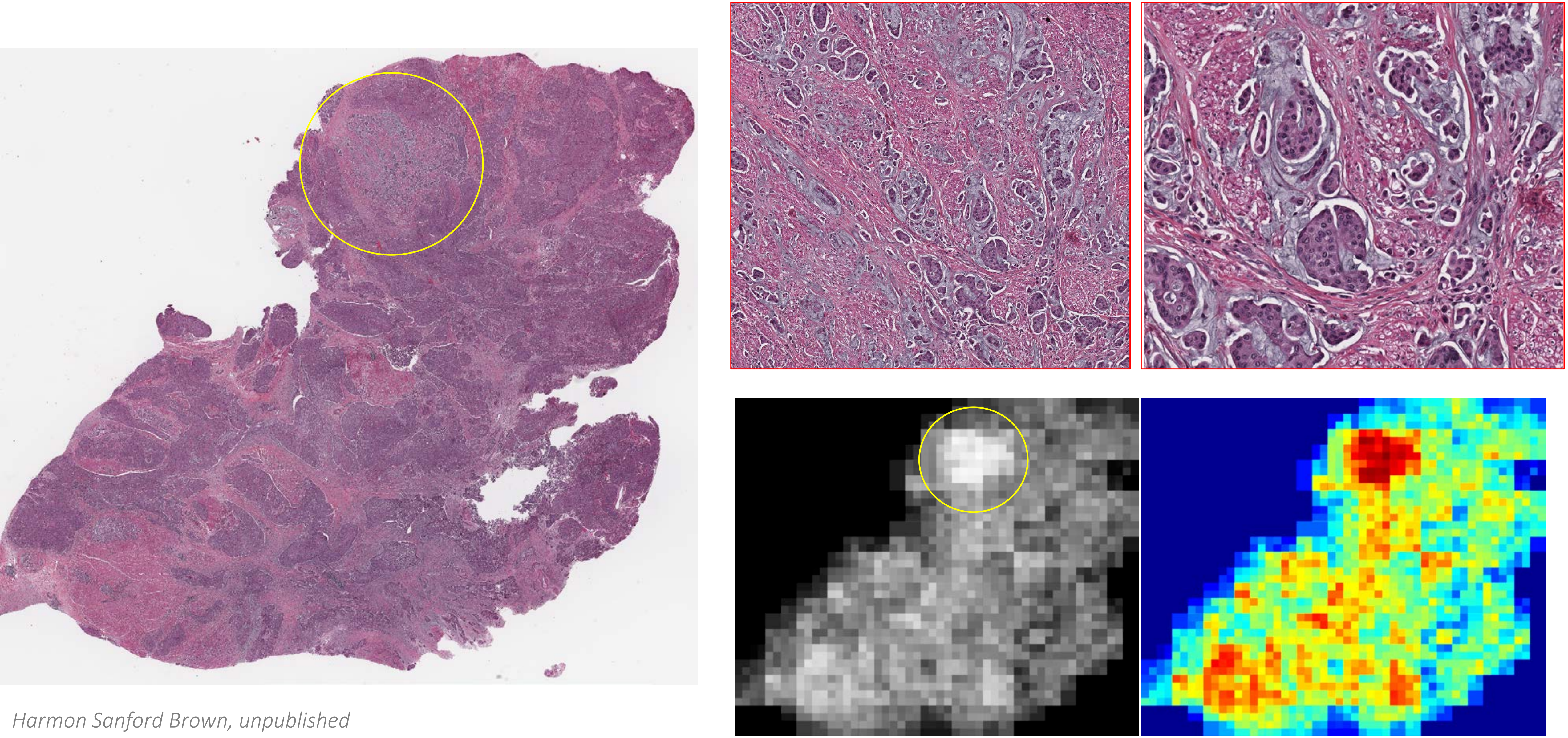
Harmon Sanford Brown, unpublished

Direct prediction of poor outcome



Harmon Sanford Brown, unpublished

Mapping features of poor outcome



Harmon Sanford Brown, unpublished

Multi-modality AI

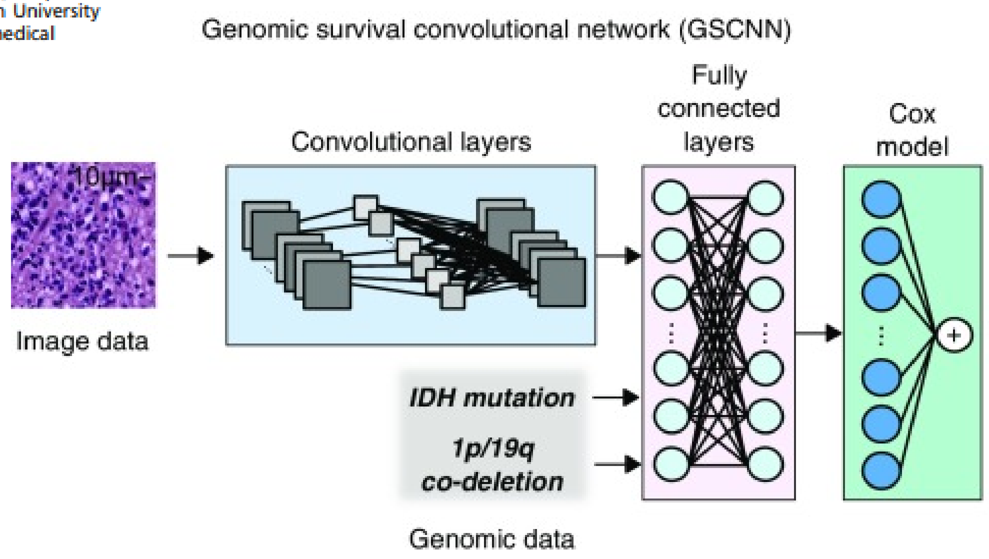
– Ultimately, outcome is a function of several factors
clinical/demographic, histology, genomics, etc...

PNAS

Predicting cancer outcomes from histology and genomics using convolutional networks

Pooya Mobadersany[a], Safoora Yousefi[a], Mohamed Amgad[a], David A. Gutman[b], Jill S. Barnholtz-Sloan[c], José E. Velázquez Vega[d], Daniel J. Brat[e], and Lee A. D. Cooper[a,f,g,1]

[a]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA 30322; [b]Department of Neurology, Emory University School of Medicine, Atlanta, GA 30322; [c]Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106; [d]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322; [e]Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611; [f]Winship Cancer Institute, Emory University, Atlanta, GA 30322; and [g]Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, Atlanta, GA 30322



Big Picture

- Pathology departments are an untapped resource of medical imaging data
- Reproducibility is in the eye of the beholder:
  - AI in clinical medicine is limited by annotations more than data
  - Expert-based systems continually evolve
- The applications are different but the goal is the same: prognosis

Big Picture: **how do we enable the next generation?**

Clinical Education

- The expert knows best
- The system is limited by the impact

Scientific Innovation

- Big data requires big ideas
- Integration of multi-modality information is complex

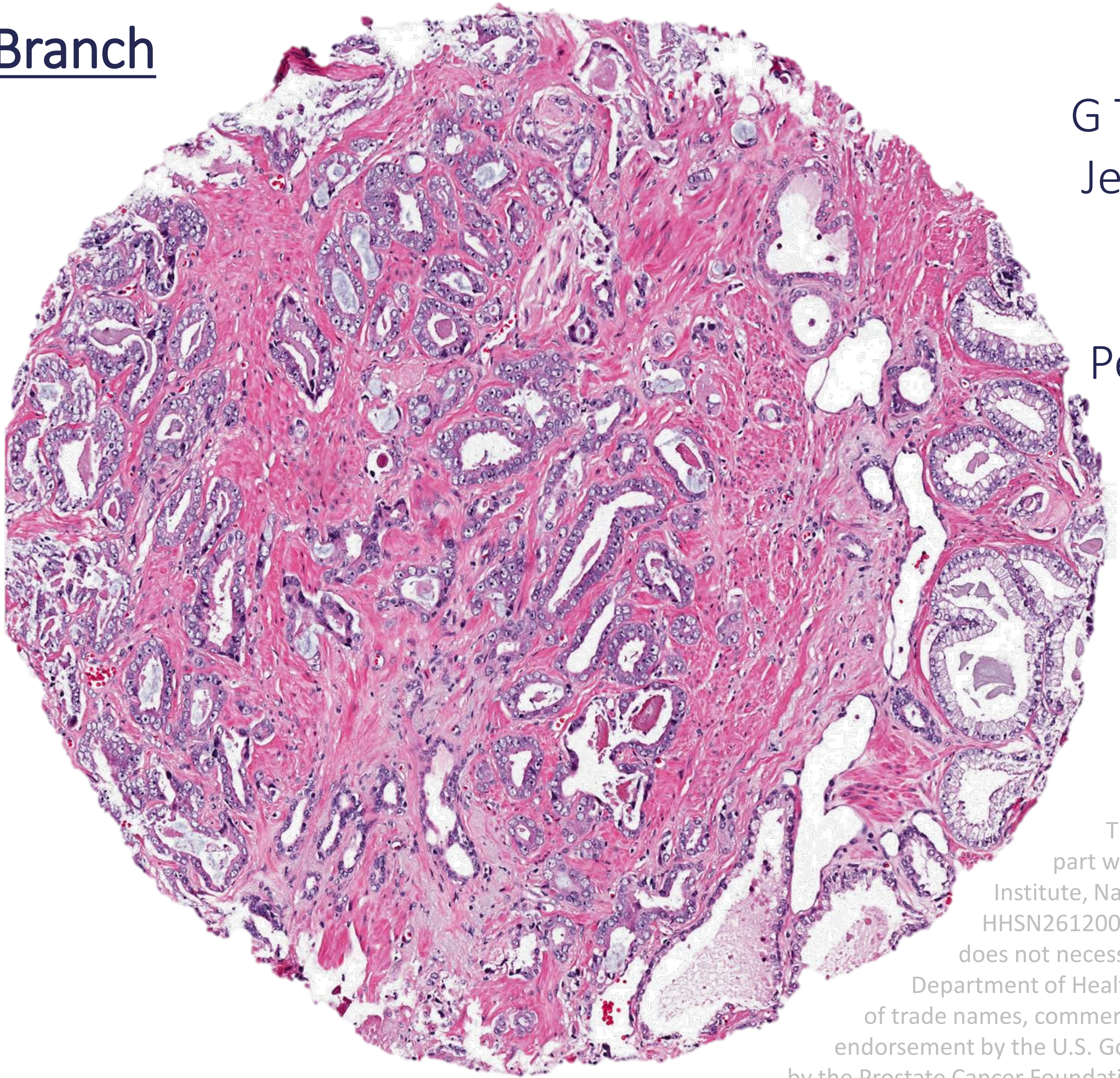
Infrastructure

There is no substitute for data curation

...and outcomes

Molecular Imaging Branch

Peter Choyke
Baris Turkbey
Tom Sanford
Sherif Mehralivand
Jonathan Sackett
Deepak Kesani
Liza Lindenberg
Esther Mena
Steve Adler
Yolanda McKinney
Juanita Weaver
Dagane Daar
Alicia Forest
Phil Eclarinal



Collaborators

G Thomas Brown (NCI)
Jesse McKenney (CCF)
Nathan Lay
Joanna Shih (NCI)
Peter Pinto (NCI UOB)
Brad Wood (NCI IR)
Sheng Xu (NCI IR)
Janet Eary (NCI)

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Harmon is supported in part by the Prostate Cancer Foundation Young Investigator Award.