Beyond Seeing:
AI and the Digital Pathology Revolution

AI For Health Symposium

Washington University St. Louis

18 October 2024

Richard J. Cote, M.D., FRCPath, FCAP
Edward Mallinckrodt Professor and Chair
Department of Pathology and Immunology

Co-Director, Cordell Institute

Chapter

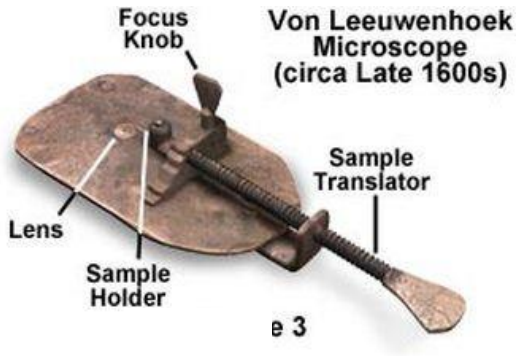
I. Impact of Digital Microscopy and AI on Pathology
II. AI on Digital Images: Doing What Pathologists Can Do
III. Beyond Seeing: Doing What Pathologists Cannot Do
IV. How AI Learns – A Lengthscale Analysis
V. The Challenge of Generalizing AI Algorithms: Stain Variation
VI. What's Next

Chapter I: Impact of Digital Microscopy and AI on Pathology

- Democratizing expertise through digital pathology
- Requirements of Digital Images for Use in AI
- Radical innovation in one of our oldest tools:
  Novel advances in microscopy

The Microscope:
The Greatest Scientific Tool Ever Invented (!,?)

Antonie van Leeuwenhoek

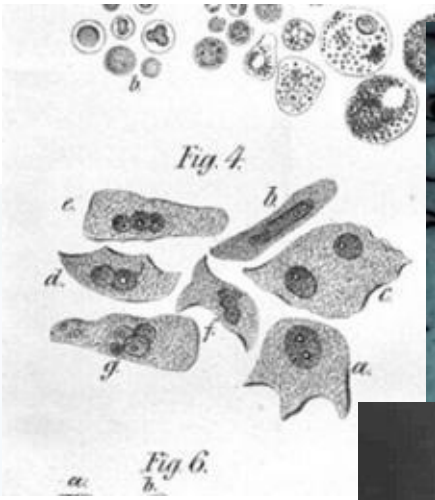


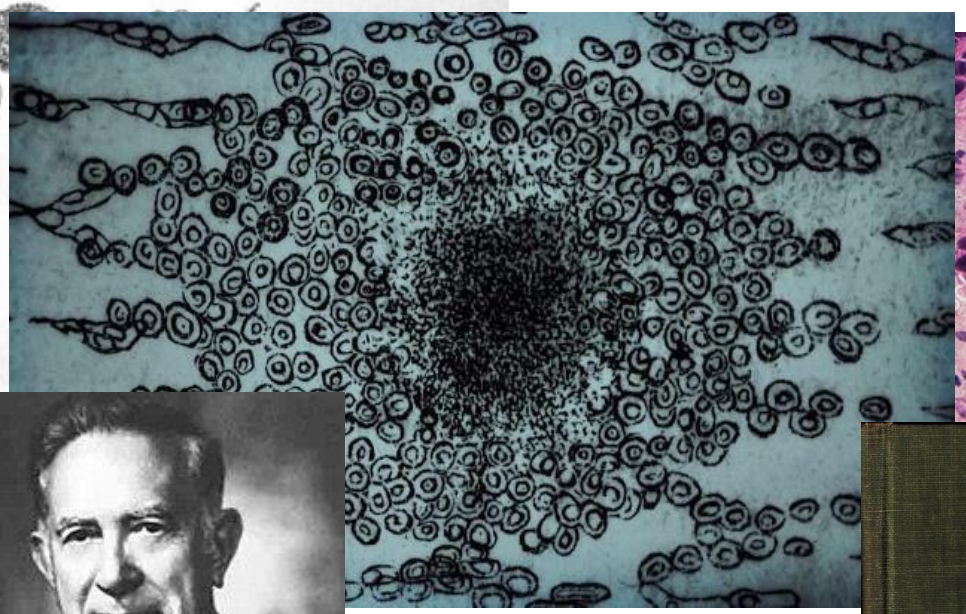
1600's
Von Leeuwenhoek Microscope

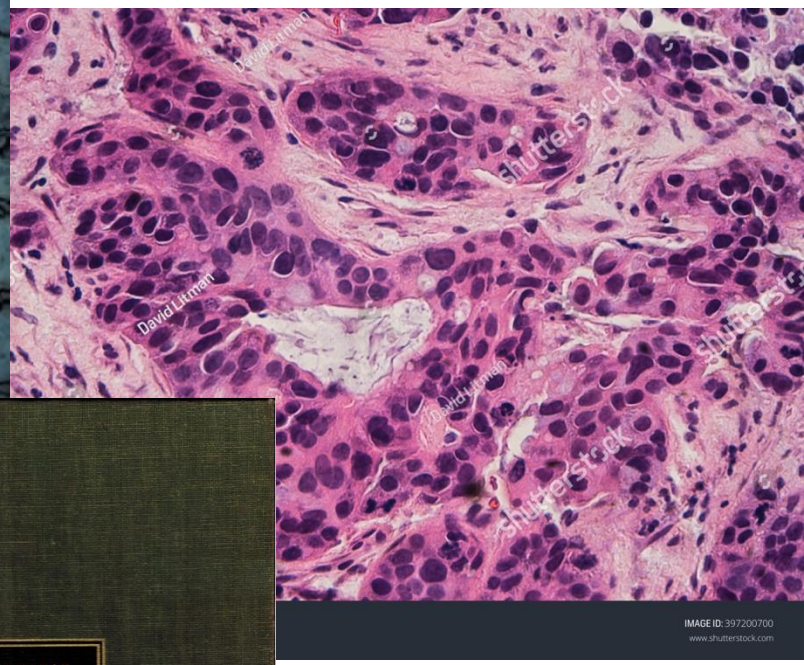
Late 1800's
Zeiss and Abe perfect the microscope

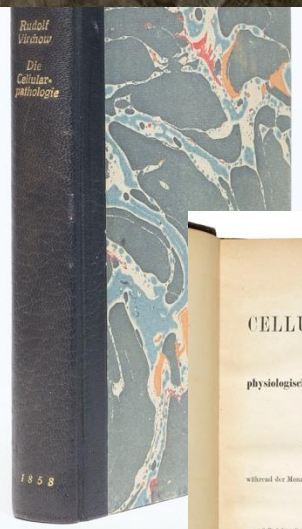
2016
Fourier ptychography microscope

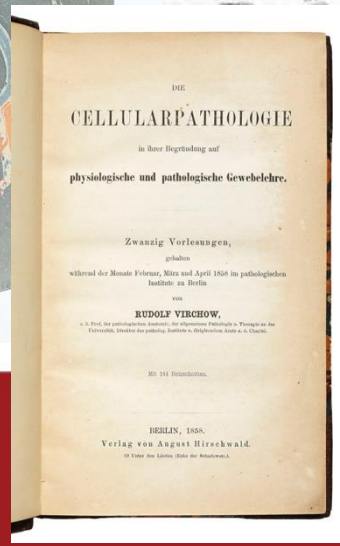
Surgical Pathology: A Morphology Based Science

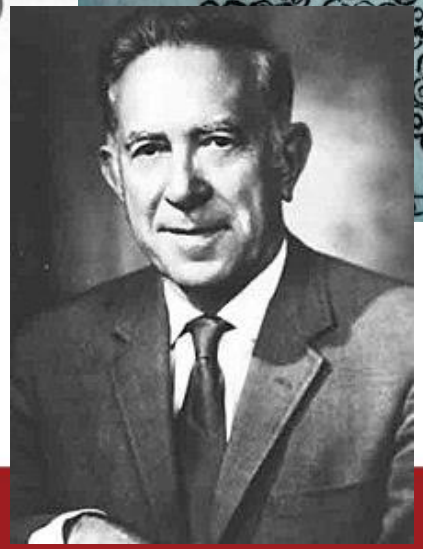


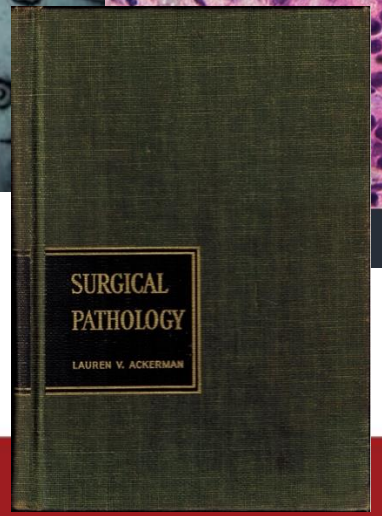










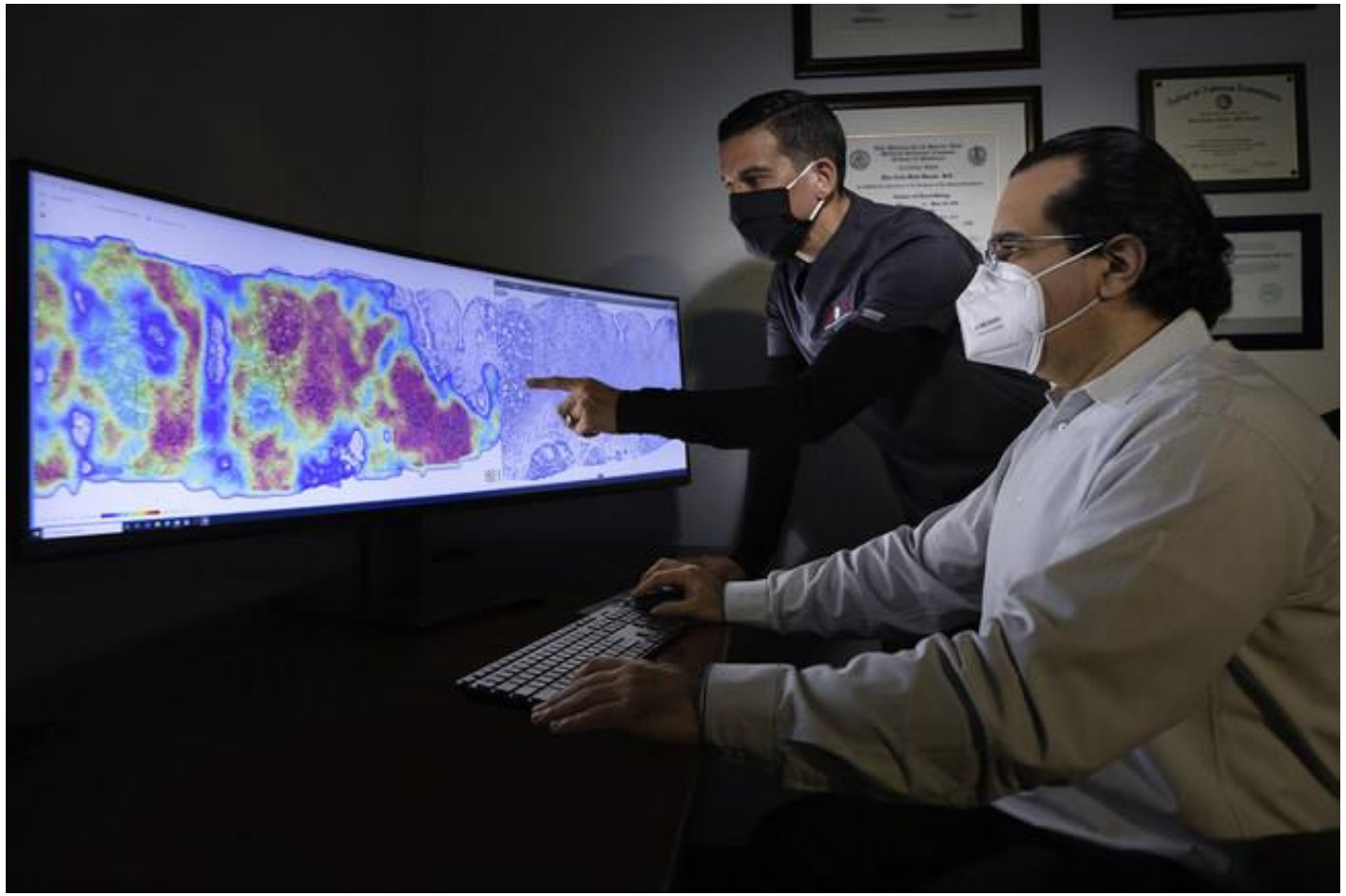


Washington University in St. Louis
SCHOOL OF MEDICINE

Revolutionizing Access: Digital Telepathology

- Quickly scan slides at high power
- Produce high quality digital images
- Transmit images over the cloud to anywhere in the world

- *Provide advanced pathology access to underserved populations*



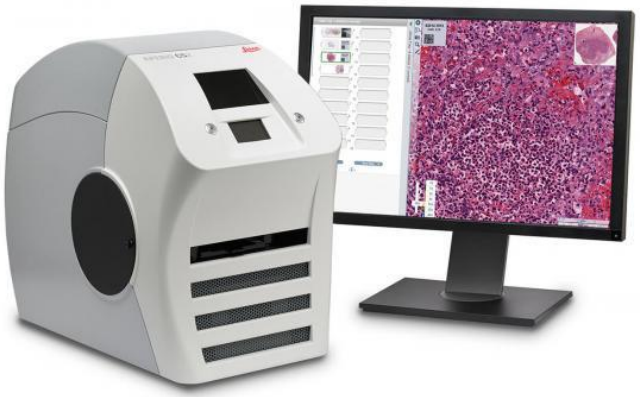
Digital Pathology: Democratizing Expertise

Subspecialty Diagnostic Expertise: Available at limited locations and difficult to access

What happens when a community pathologist or clinician wants to have a Wash U subspecialist review a case?

- Send glass slides by mail
- Wait days/weeks
- Chance of glass slide loss

- *No ability for real time intraoperative consultation*



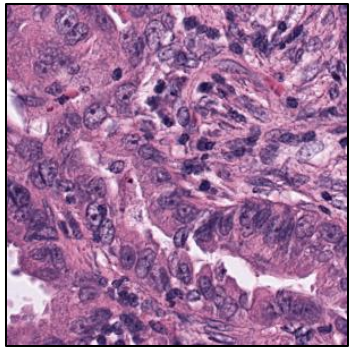
How can we address this?

Digital Pathology

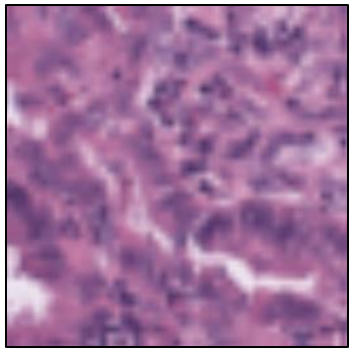
Requirements of Digital Images for AI

Resolution

1 µm

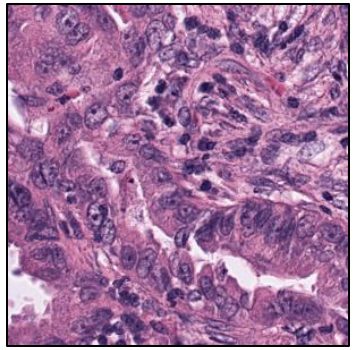


5 µm

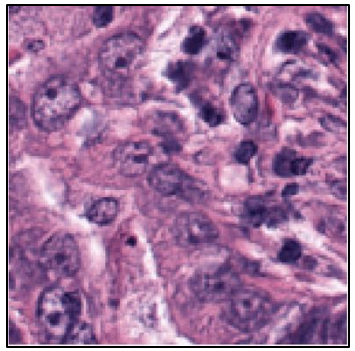


Field-Of-View

120 x 120 µm^2

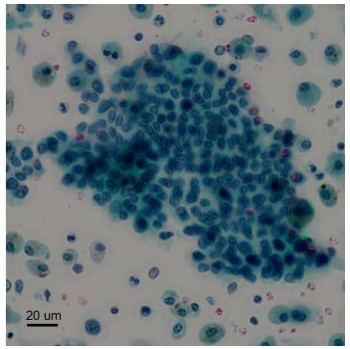


75 x 75 µm^2

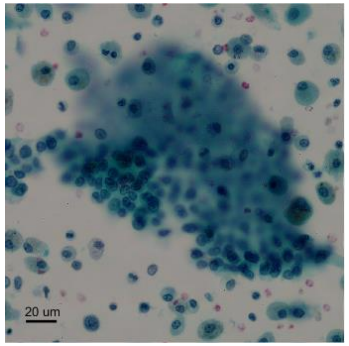


Focus

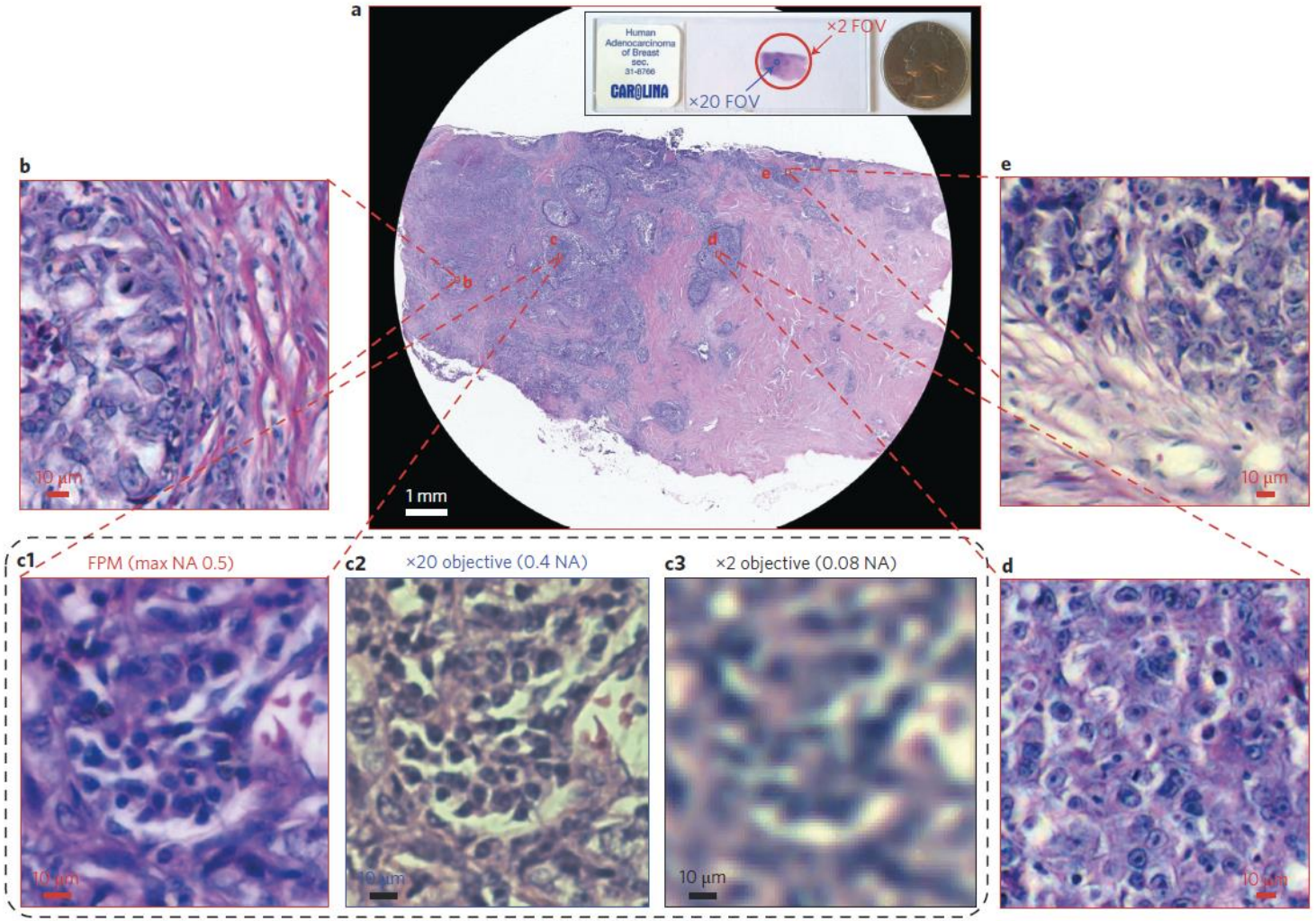
All-In-Focus



Single Plane



Good Microscopy Modalities



For nearly two centuries we have use thin pieces of tissue placed on glass slides with a single light source illuminating the tissue and sending the image to the lens.

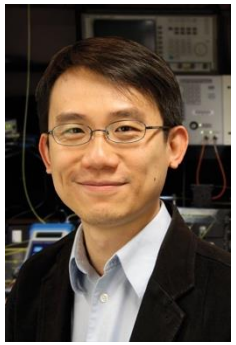
Though these thin tissues look flat, they have tiny aberrations (ripples/waves) that become magnified at higher powers

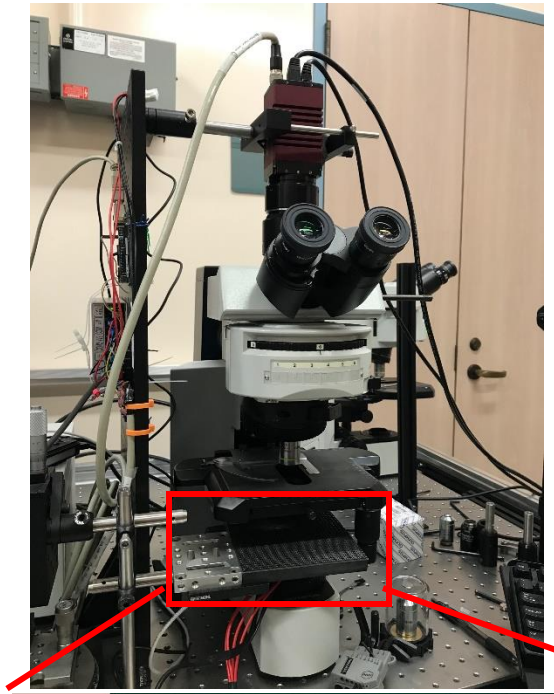
These aberrations make it impossible to obtain an all in focus image from a single focal plan

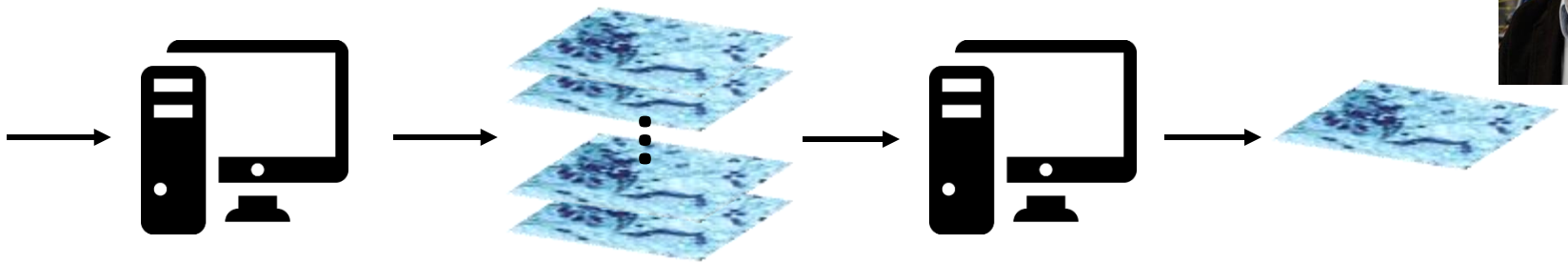
"About 90% of digital slides have defocus issues. For AI, Garbage In = Garbage out."
Philip Nelson, Head of Google Imaging

All-in-focus Fourier Ptychographic Microscope (FPM)

A radical innovation in microscopy







Stack of different focal plane

All-in-focus image

- Illuminated by a LED array from many angles
- Does not require *mechanically scanning* different focal planes

G. Zheng, R. Horstmeyer, and **C. Yang**, *Nature Photonics* **7**(9), 739-745 (2013).
G. Zheng, C. Shen, S. Jiang, P. Song, and **C. Yang**, *Nature Reviews Physics* **3**(3), 207-223 (2021).

Washington University in St. Louis
School of Medicine

Fine Needle Aspirate (FNA) study with Fourier Ptychography

Image generated by conventional microscopy

- ✓ **High Resolution**
- ✓ **Large Field-of-View**
- ✓ **All-In-Focus**
- ✓ **Aberration Correction**

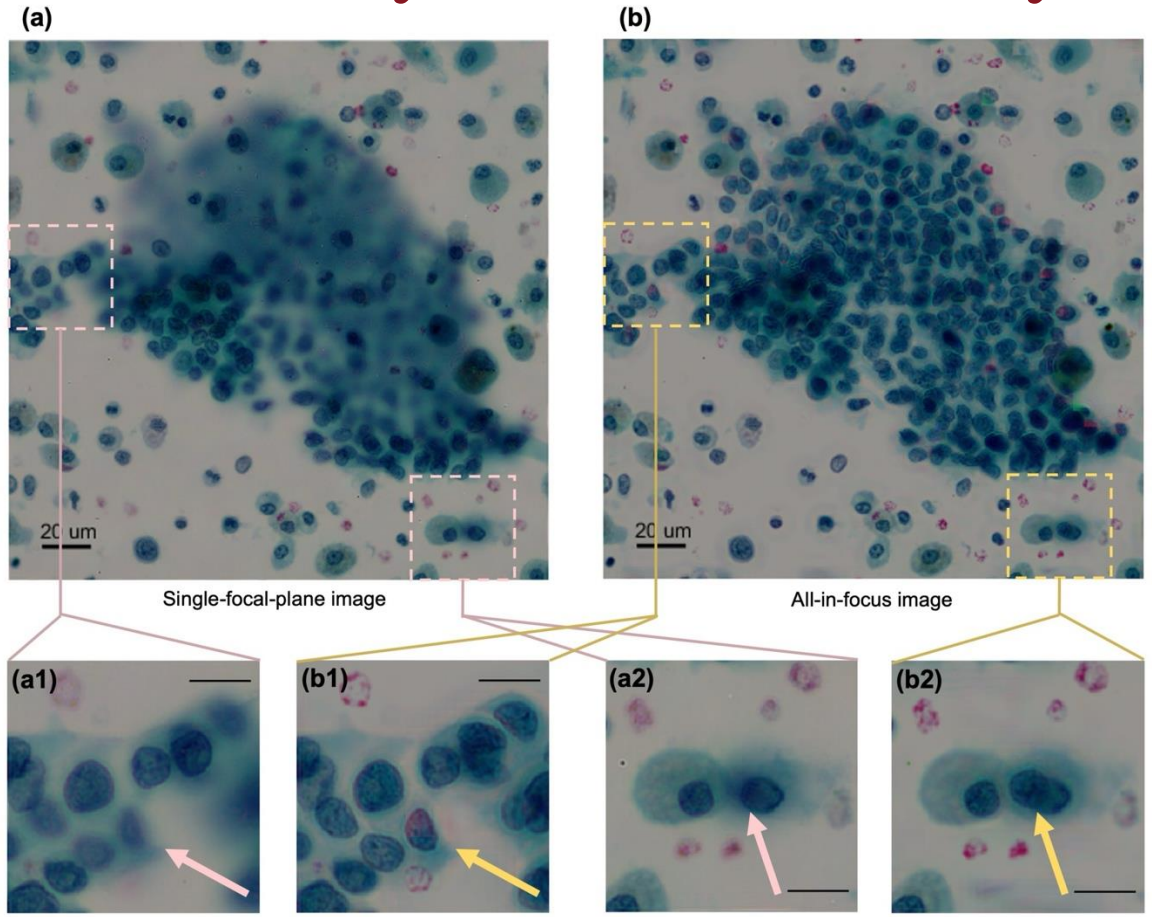


Image generated by all-in-focus FPM

Arrows indicate features discernible in FPM image versus conventional

scale bar = 10 microns

FPM produces all-in-focus images that are well suited for Deep Learning processing.

Liang M, Bernadt C, Wong SBJ, Choi C, Cote R, Yang C. All-in-focus fine needle aspiration biopsy imaging based on Fourier ptychographic microscopy. J Pathol Inform. 2022

Chapter II: Using AI to Do What Pathologists Do: Identifying Objects

Identify and Classify Circulating Tumor Cells (CTC)

Train

Model Samples

Only CTCs*

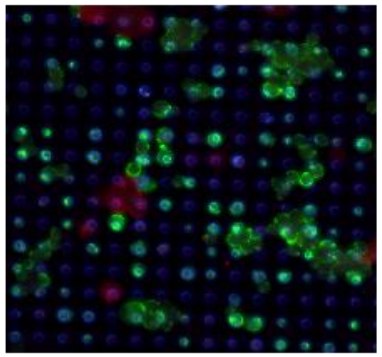
Only CAFs

Only WBCs

Combined

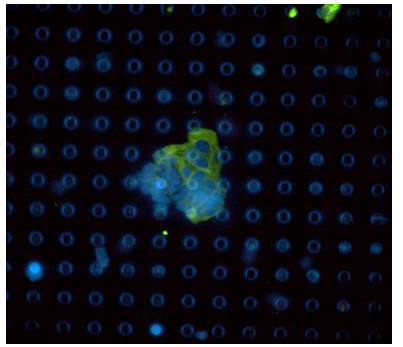
Test 1

Model Sample with tumor, CAF and PBMC combined



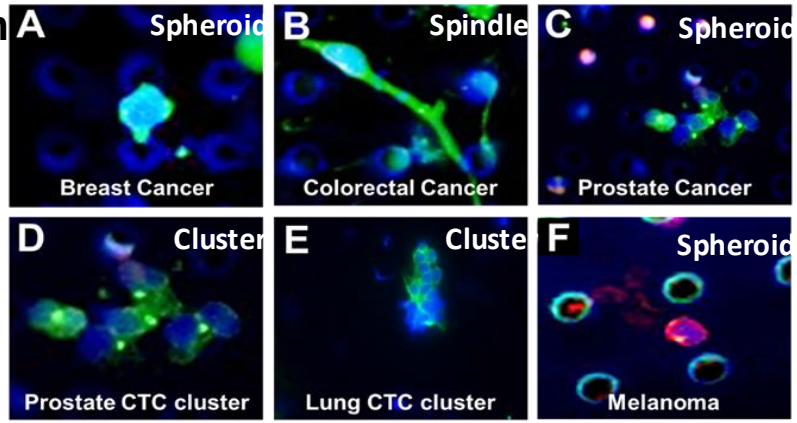
Test 2

Clinical samples



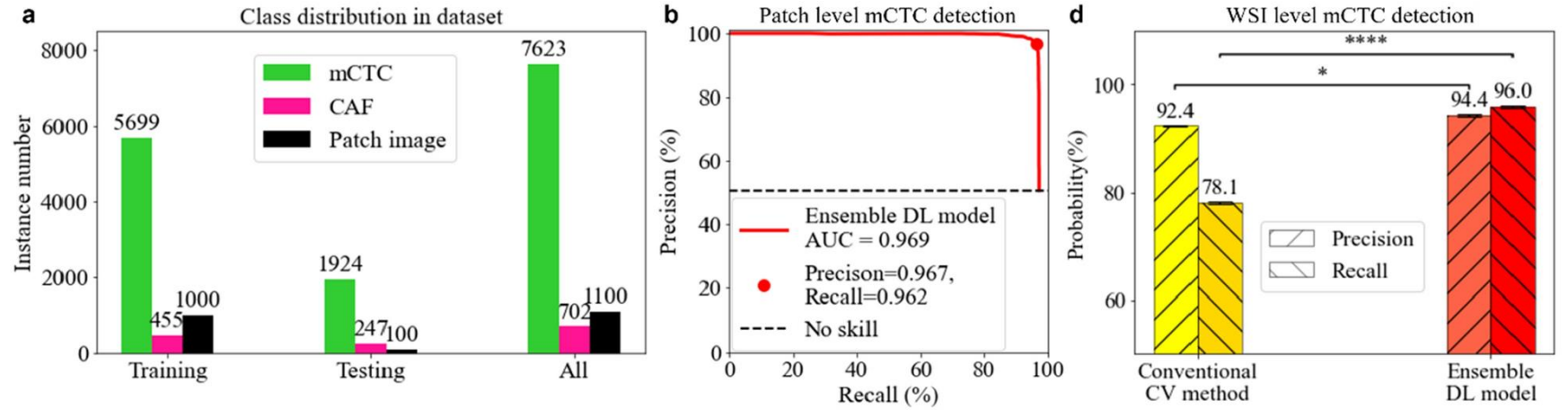
*** Deep Learning must learn to recognize and distinguish multiple CTC m**

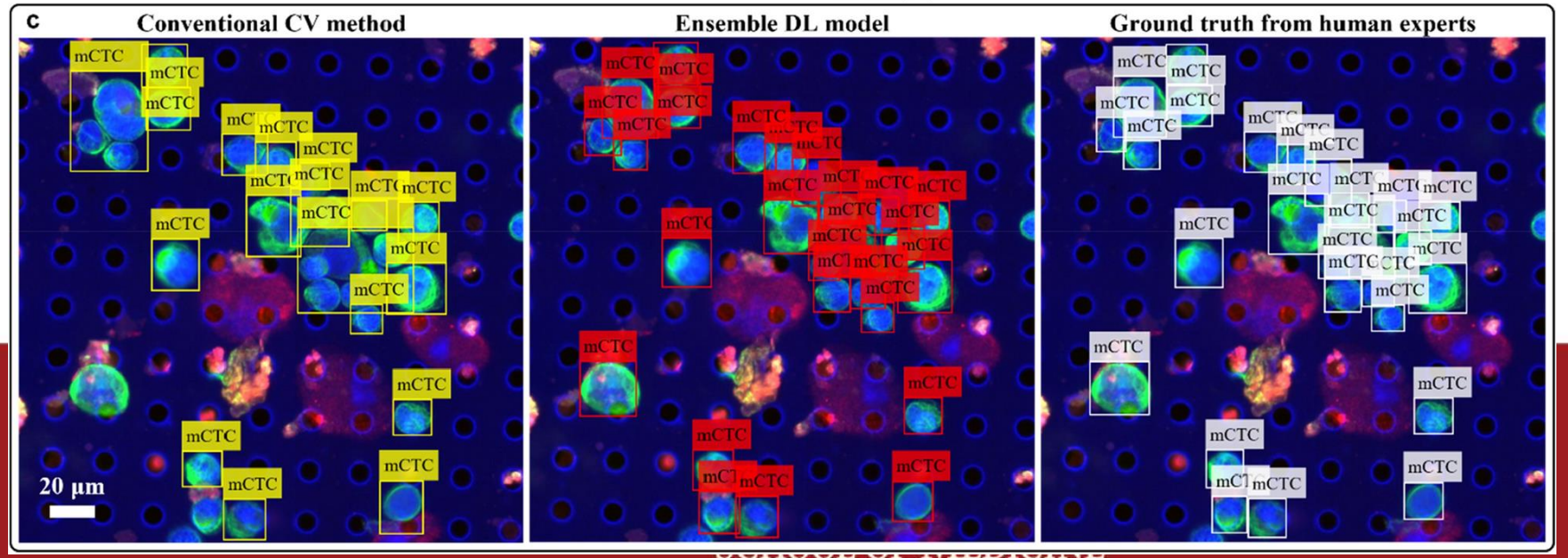
CTC have different appearances. These must be recognized and also distinguished by Deep Learning, and these different morphologies have different biological meaning. For example, clusters of CTC have a worse prognosis, and spindle cell CTC may represent EMT transition



Shen, C., Rawal, S., Brown, R. *et al.* Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning. *Sci Rep* **13**, 5708 (2023). https://doi.org/10.1038/s41598-023-32955-0

Automatic Detection of CTCs

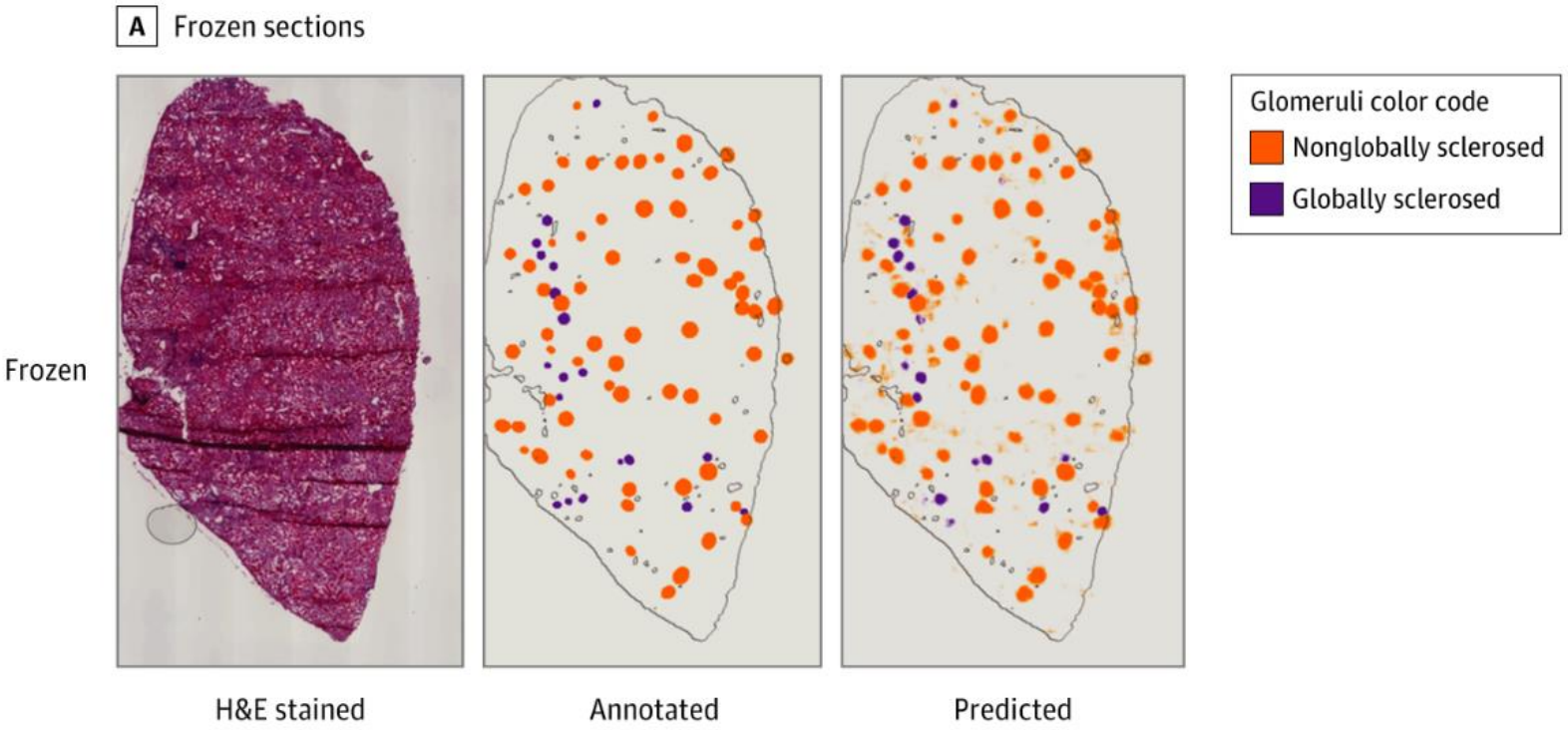




Up to 96% accuracy in correctly identifying CTC and cCAF

Shen, C., Rawal, S., Brown, R. *et al.* Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning. *Sci Rep* **13**, 5708 (2023). https://doi.org/10.1038/s41598-023-32955-0

Automated glomerular analysis using AI

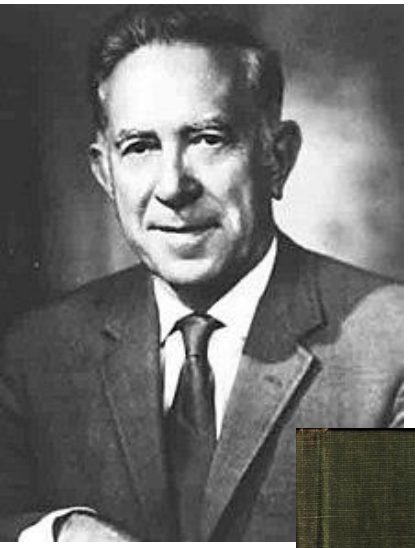


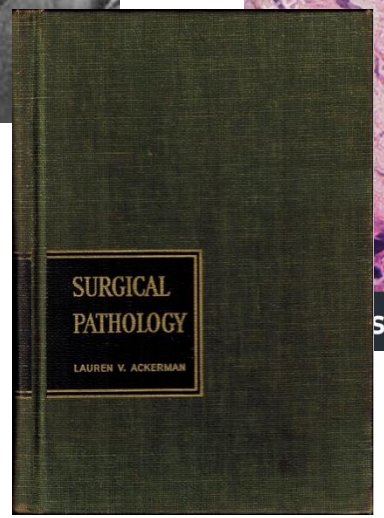
- WU developed artificial intelligence
- Detects non-sclerotic and sclerotic glomeruli (two very similar objects)
- Developed software package to securely deploy the algorithm in the cloud – Trusted Kidney
- ***This technology improves pathologist reproducibility and has potential to significantly decrease the donor kidney discard rate.***

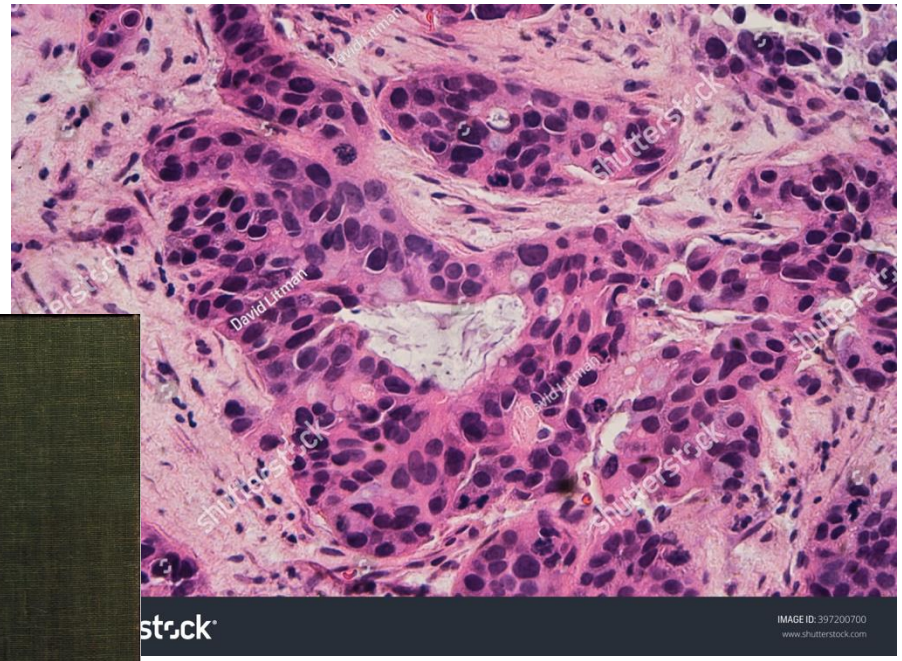
Trusted Kidney can faithfully and reproducibly identify functioning from non-functioning glomeruli, and provide a more accurate percentage than human observation

Joe Gaut
Josh Swamidass

Chapter III: Beyond "Seeing": Doing What Pathologists Can't Do







- Beyond making a diagnosis using the microscope, surgical pathologists have long sought to predict cancer behavior through elucidation of histologic features.
- This began over a century ago when it was increasingly recognized that the less "normal" a cancer looked, the more aggressive its behavior would be.
- This has led to routine histologic grading of tumors
- However, despite active work, grading systems are only able to provide inaccurate estimates of behavior
- Even supplemented by stage and molecular determinants, our ability to accurately predict behavior is inadequate

Can AI predict metastatic progression in early-stage NSCLC based on the original diagnostic histology

i.e., can AI be trained to do something even the best pathologist can't do?

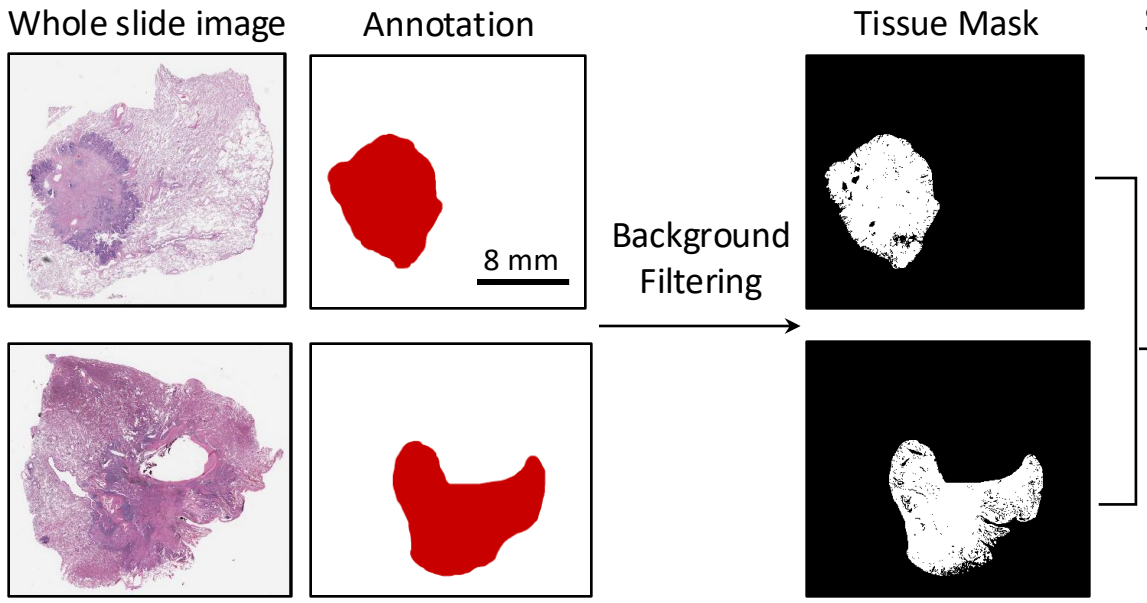
- 158 patients with early-stage NSCLC
- 117 stage I
- 93 with no metastasis (Met -) >5 years
- 65 with brain mets (Met +)
- ***No adjuvant therapy***

H. Zhou*, M. Watson*, et al. *Journal of Pathology* (2024).

		Met⁻ (n=93)	Met⁺ (n=65)
Gender	male	47	27
	female	46	38
Age	Average Age at DX	60 (47-78)	57 (25-73)
Histology	Adenocarcinoma	48	44
	Squamous Cell	32	11
	Large Cell	3	0
	BAC	4	0
	Poorly Differentiated	1	5
	Mixed	5	5
Grade	I	12	4
	II	48	26
	III	25	27
	IV	0	1
	ND	8	7
Stage	**I**	**85**	**32**
	II	3	12
	III	0	9
	IV	0	7
	ND	5	5
Median Follow-up Time (Month)		106	12.2

NSCLC – Deep learning framework: Part I, dealing with limited specimens

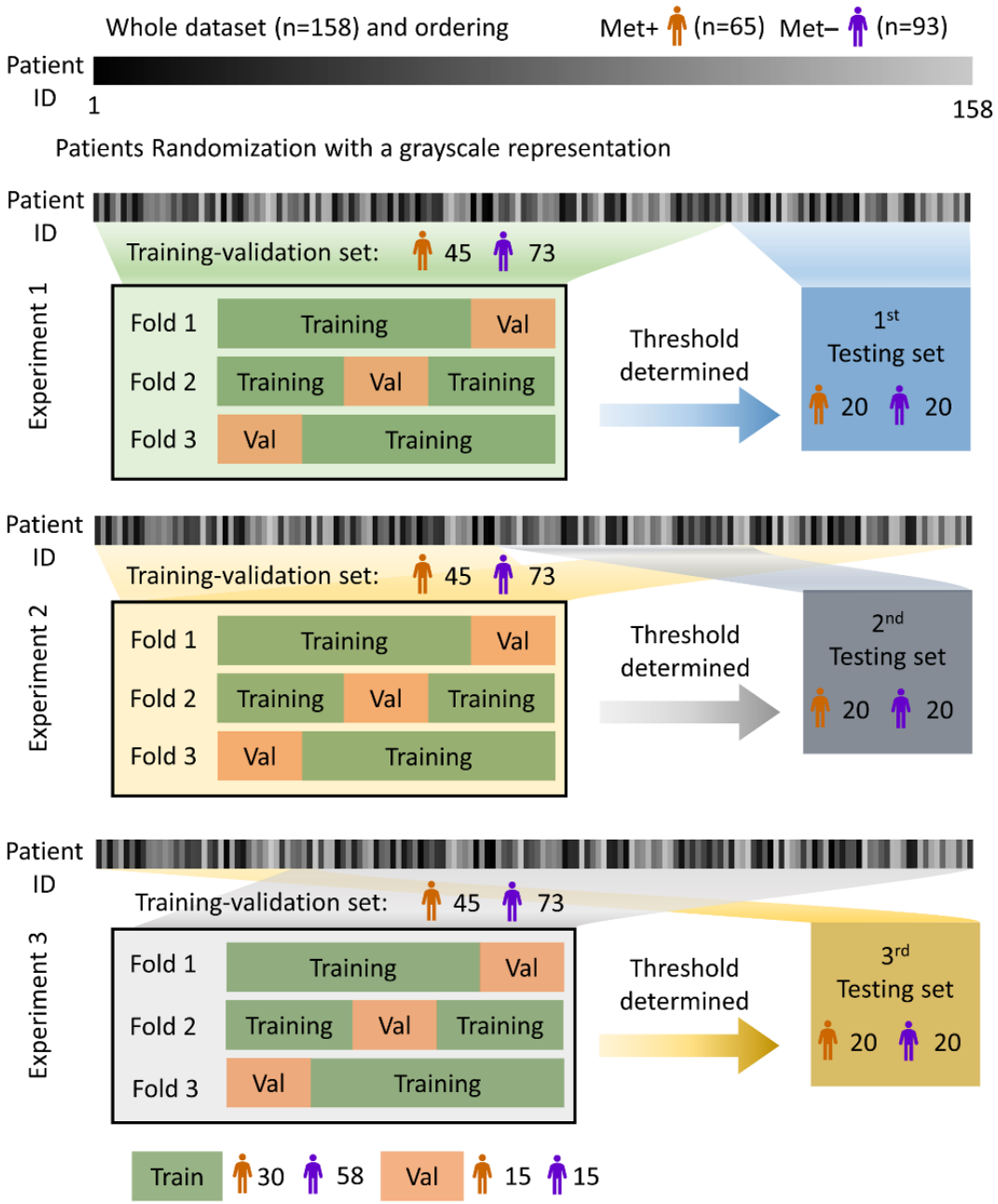
WSI of diagnostic H&E slides from 158 patients with early-stage NSCLC: 65 Met+, 93 Met-



The DL was trained on 1000 different non-overlapping areas (tiles) within each tumor, where each tile was interrogated and used to "teach" the DL. The model then took the "best" prediction for the overall tumor.

H. Zhou*, M. Watson*, et al. *Journal of Pathology* (2024).

NSCLC – Training-Validation-Testing Strategy: Part II, dealing with limited specimens

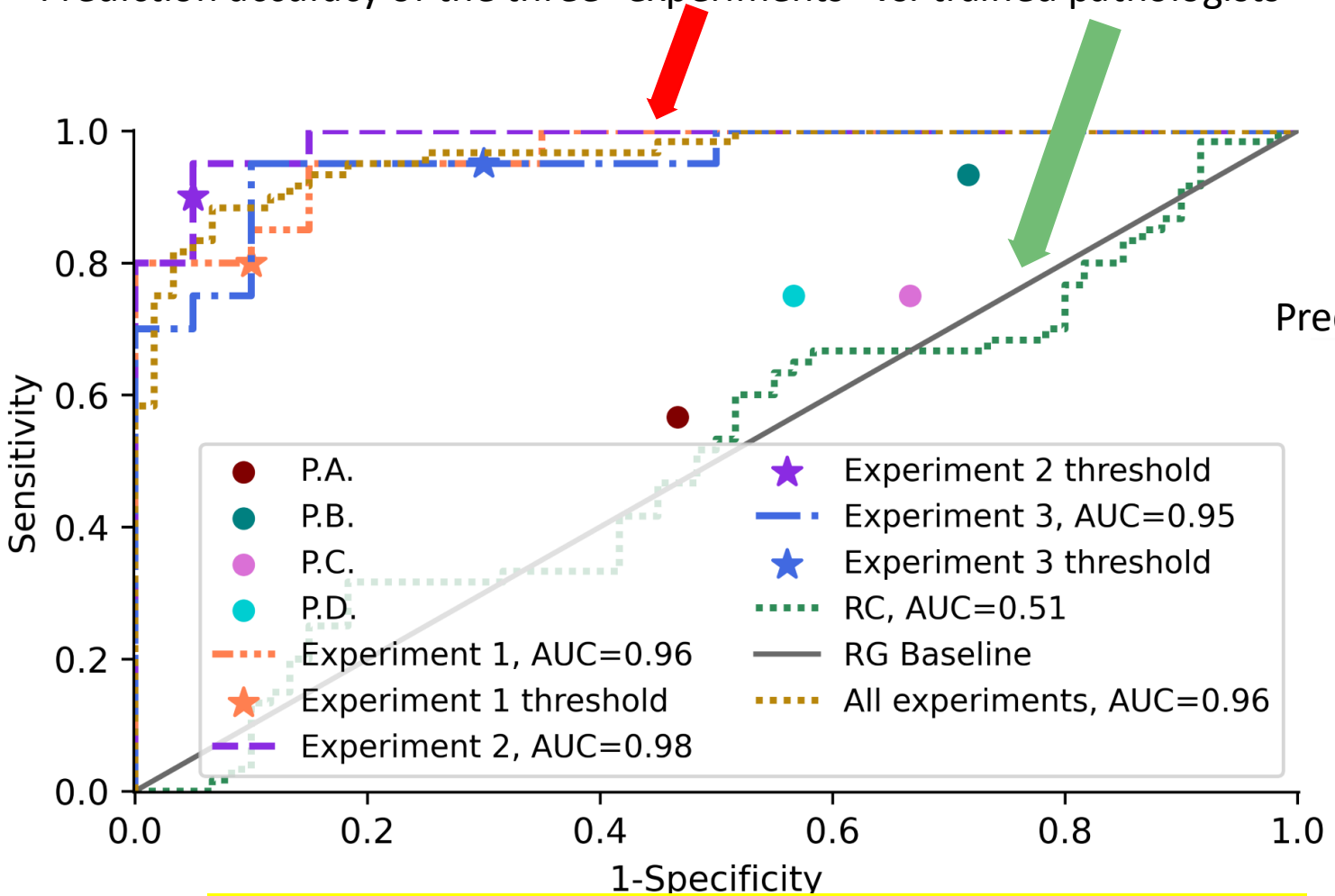


- Due to the limited numbers of patients and events, we implemented a novel method to train, validate and then test the AI algorithm.
- Three “experiments” (i.e., training, validation and testing) were performed.
- For each “experiment” three train/test splits were implemented and the results combined to determine the algorithm.
  - Note that the validation set in the three splits is non-overlapping.
- The algorithm was then tested on a completely separate set of cases (20 Met+, 20 Met-).
  - Note that in each “experiment” the testing/validation set was shuffled, and non-overlapping cases were selected for testing in each of the three experiments (that is, the set of 40 cases was different in each experiment).

H. Zhou*, M. Watson*, et al. *Journal of Pathology* (2024).

Progression risk assessment for NSCLC using deep learning: Comparison of the three "experiments" with expert pathologist review

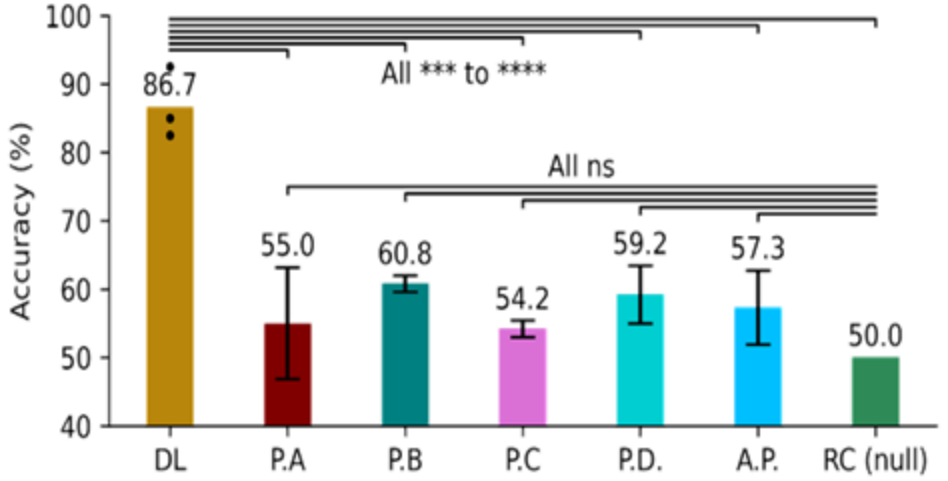
Prediction accuracy of the three "experiments" vs. trained pathologists



Note the high degree of concordance in the three experiments

The prediction accuracy (predicting Met+ vs. Met-) of the DL model was 86%, vs little better than 50% for expert pathologists!

Prediction accuracy of DL model vs expert pathologists



ns: not statistically significant;
***, ****: p-value <0.0001, <0.00001

H. Zhou*, M. Watson*, et al. *Journal of Pathology* (2024).

AI is Highly Predictive of No Progression in Stage 1 NSCLC

Stage I	Low risk	High risk	Row Totals
Met^-	45	9	54
Met^+	2	26	28
Column totals	47	35	82

Table 2: AI prediction of development of progression (Met^+) versus no progression (Met^-) for Stage 1 patients: AI was able to accurately predict which patients do not progress to brain metastasis, with a specificity of 95.7% and negative predictive value (NPV) of 92.9%*. The AI was less accurate at predicting those patients that develop metastases, with a sensitivity of 74.3% and positive predictive value (PPV) of 83.3%.

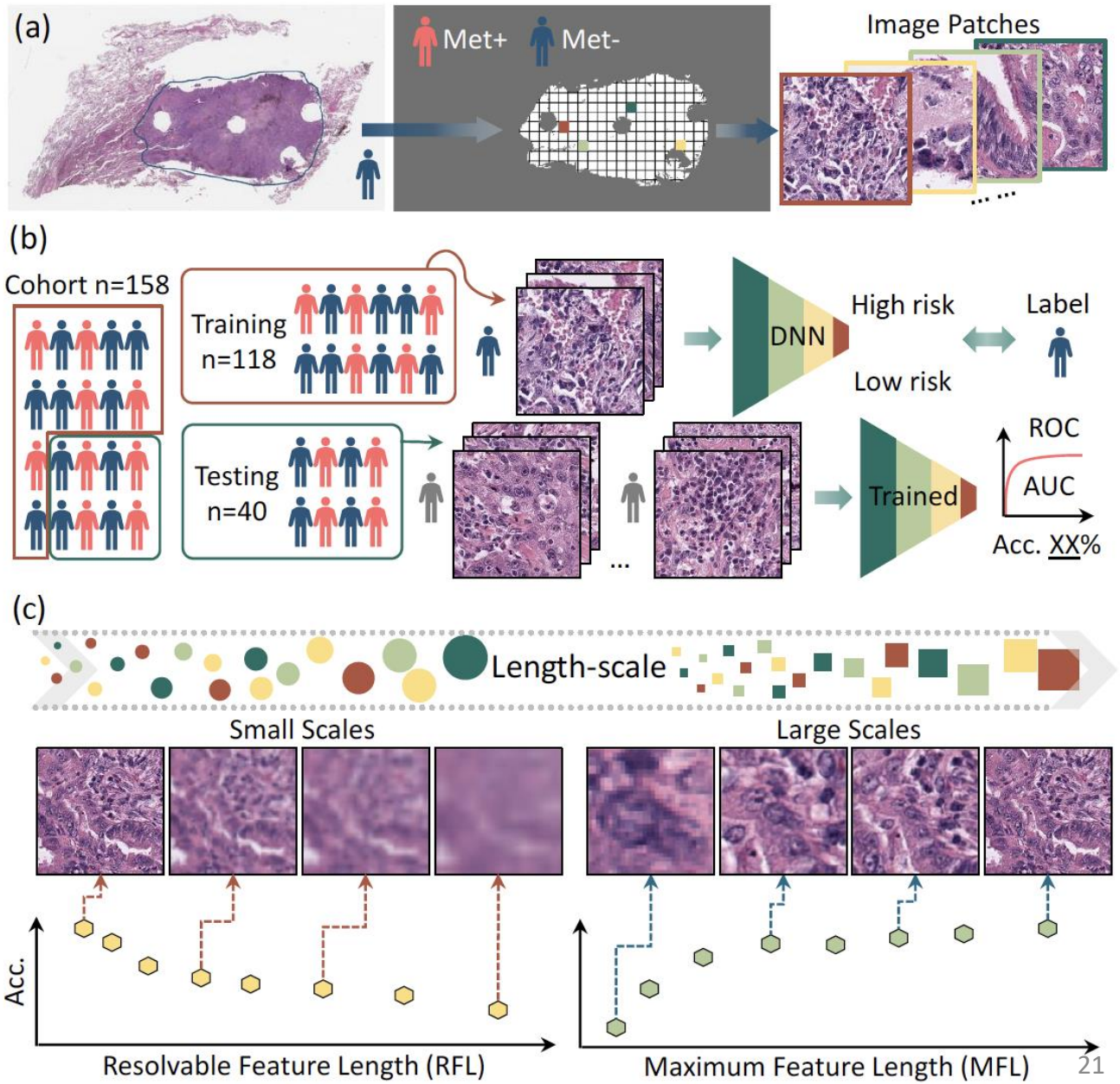
* Predicting Stage I patients who do not go on to metastasis may be the most important management feature, as it would identify patients that do not need and will not benefit from toxic adjuvant therapy. *Recall that this cohort received no adjuvant therapy.*

H. Zhou*, M. Watson*, et al. *Journal of Pathology* (2024).

Chapter IV: How Does AI Learn and What it is Looking at to Learn?

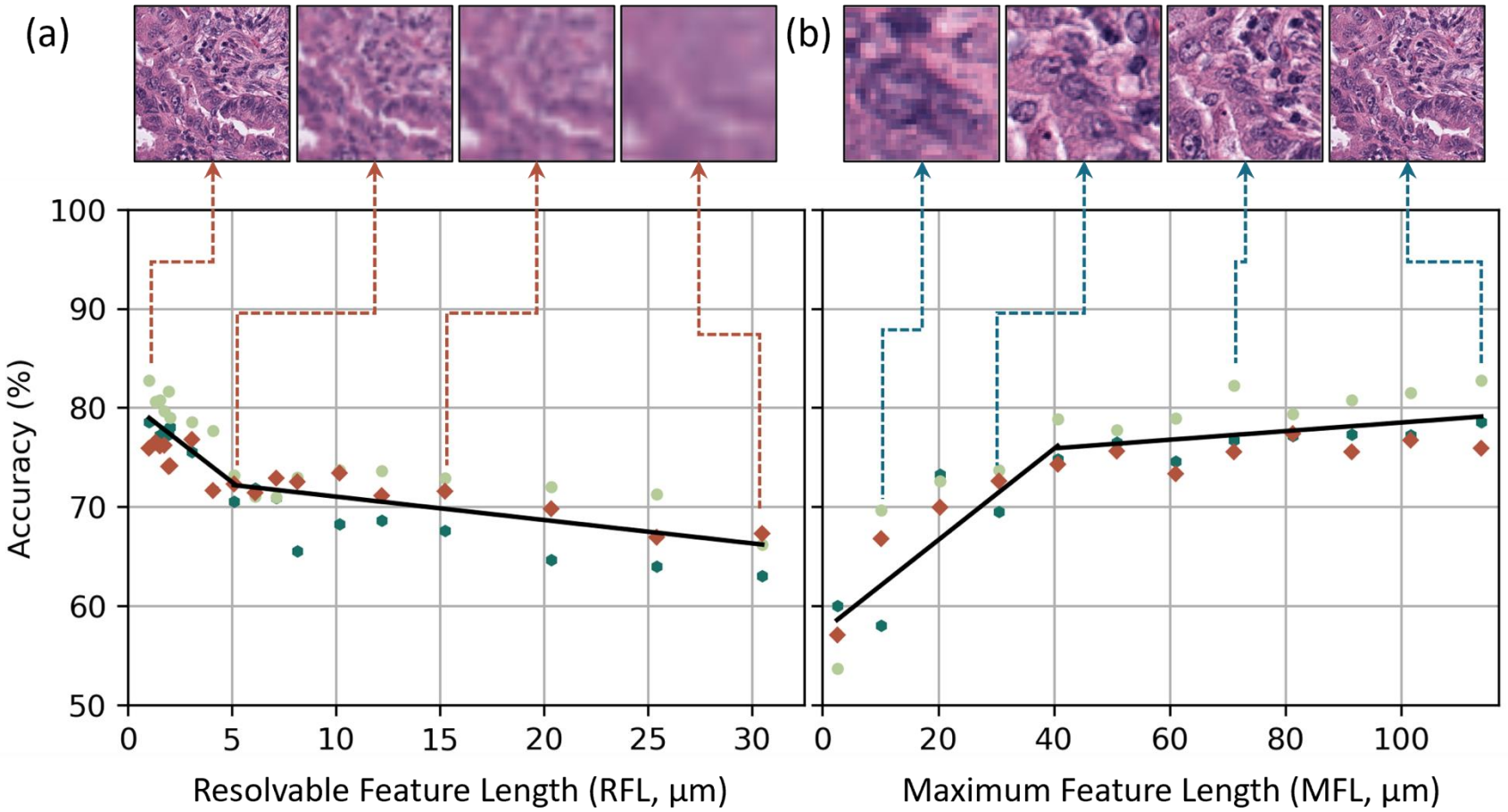
Part 1
Goal: determine the essential features AI needs to optimally learn from histologic images.

The images were manipulated to vary the Resolvable Feature Length (RFL, Resolution) and Maximum Feature Length (MFL, Field of View), and then the AI was trained on the manipulated images



Zhou et. al., Length-scale study in deep learning prediction for non-small cell lung cancer brain metastasis, Scientific Reports 2024 in press

Resolution and Field of View are Important in AI Learning on Histologic Images



At the cellular scale, the predictive power of DNNs progressively increases with higher resolution and significantly decreases when the resolvable feature length exceeds 5 microns. Additionally, DNN uses more macro-scale features associated with tissue architecture and is optimized when assessing visual fields greater than 41 microns.

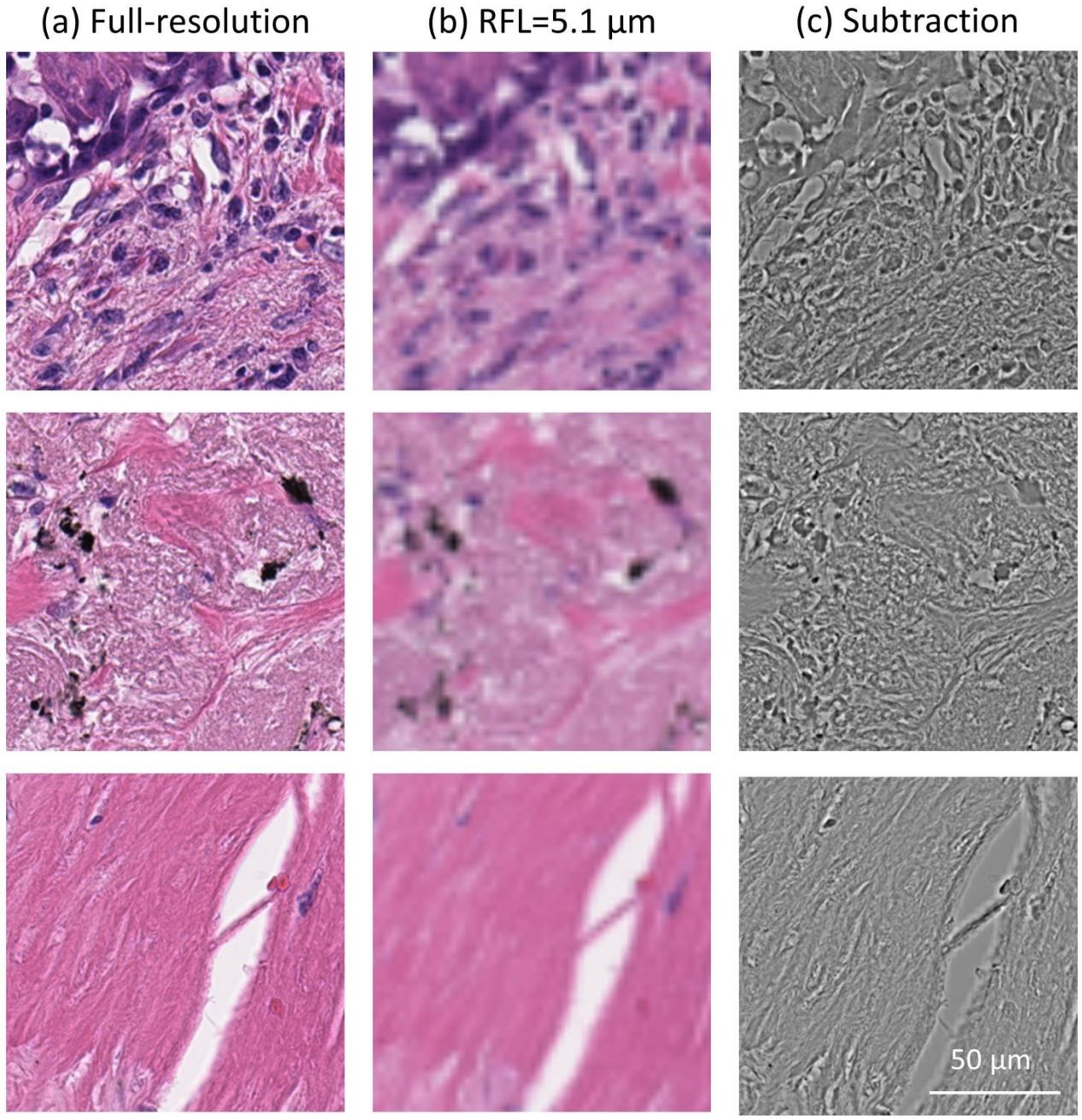
Zhou et. al., Length-scale study in deep learning prediction for non-small cell lung cancer brain metastasis, Scientific Reports 2024 in press

Length-scale study curves for different (a) RFLs and (b) MFLs. The black solid lines are the piecewise linear fittings to the average values of the three experiments.

The Impact of Resolution and Color on AI Learning

RFL 1 micron vs 5.1 microns

- Note that the information that is lost when going from 1 micron to 5 microns is most of the cellular and structural detail (column c)
- However, also note that color information is preserved, including spatial orientation of color features
- Some predictive capacity is preserved even at RFL of 30 microns, indicating the color and its spatial orientation provides useful information to the AI learning process



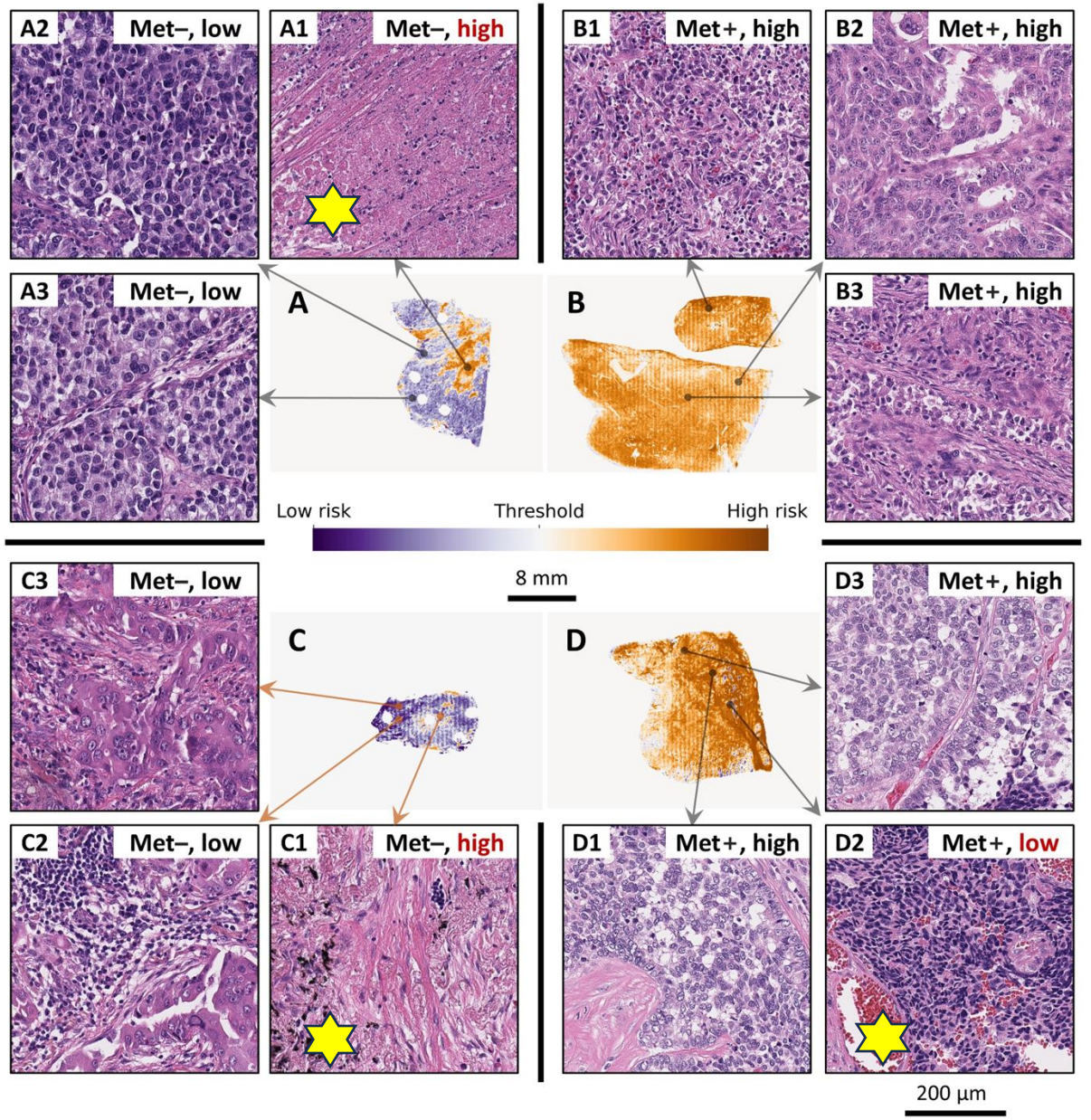
Zhou et. al., Length-scale study in deep learning prediction for non-small cell lung cancer brain metastasis, Scientific Reports 2024 in press

What The AI is Looking At?

Part 2
Goal: Determine what the AI is looking at to learn

The DL was trained on 1000 different non-overlapping areas (tiles) within each tumor, where each tile was interrogated and used to “teach” the DL. The model then took the “best” prediction for the overall tumor. **However, it is notable that not every tile predicted correctly, even if the overall tumor predicted correctly.**

A1, C1, and D2 – incorrectly predicted risk



H. Zhou*, M. Watson*, et al. *Journal of Pathology* (2024).

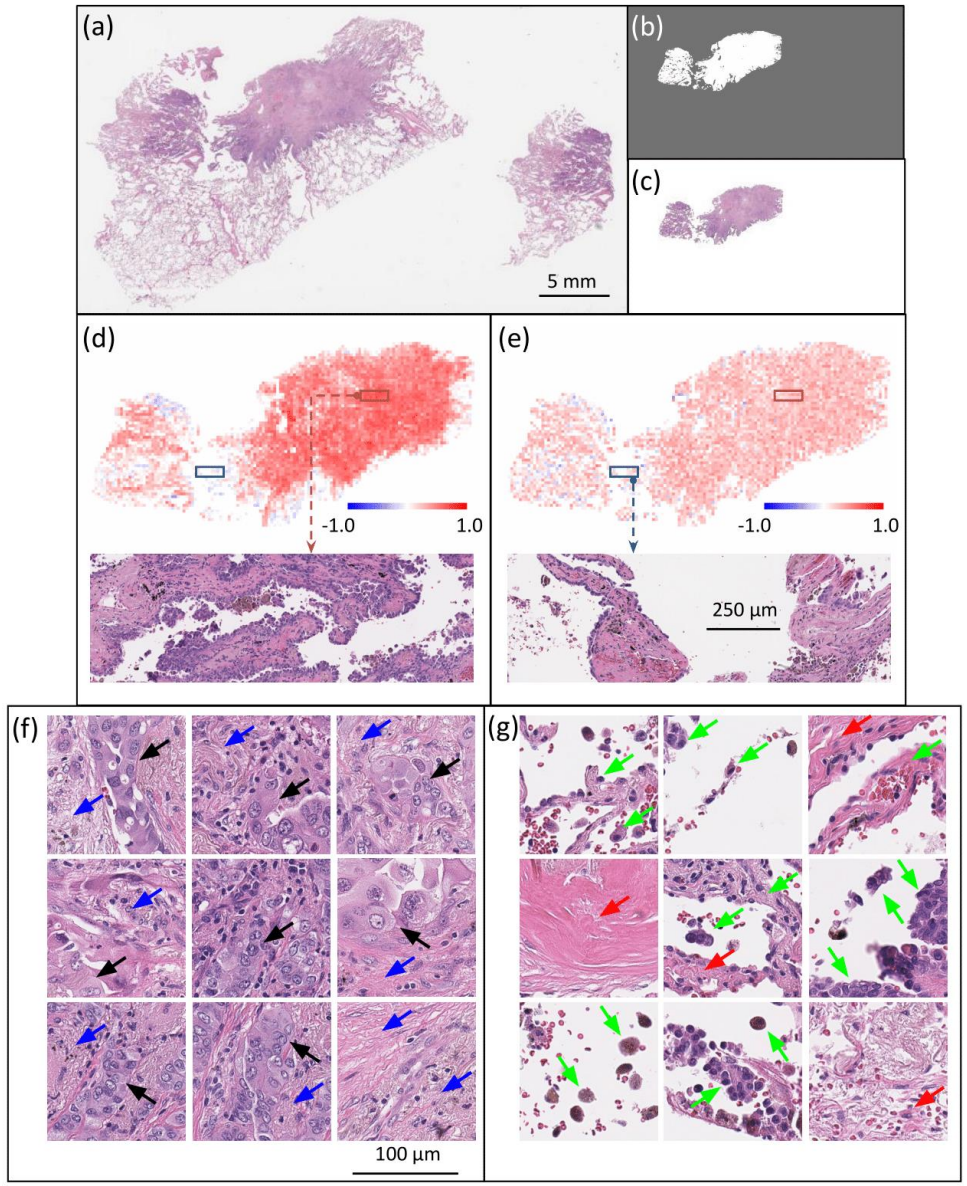
What is AI Looking At to Learn: Tile Level Analysis

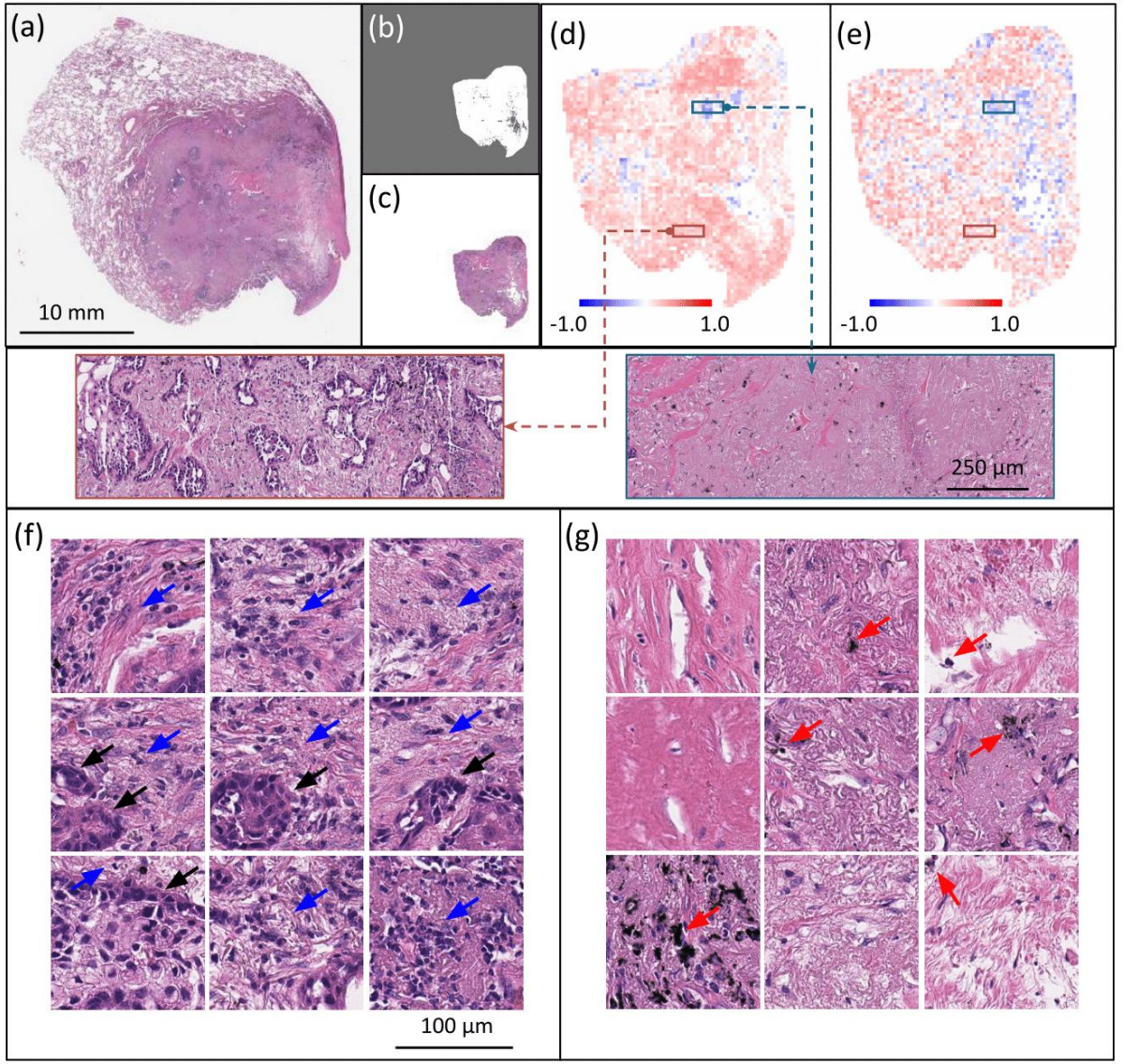
Zhou et. al., Length-scale study in deep learning prediction for non-small cell lung cancer brain metastasis, Scientific Reports 2024 in press

Red = increased prediction accuracy

Clear = no prediction accuracy

Blue = negative prediction accuracy

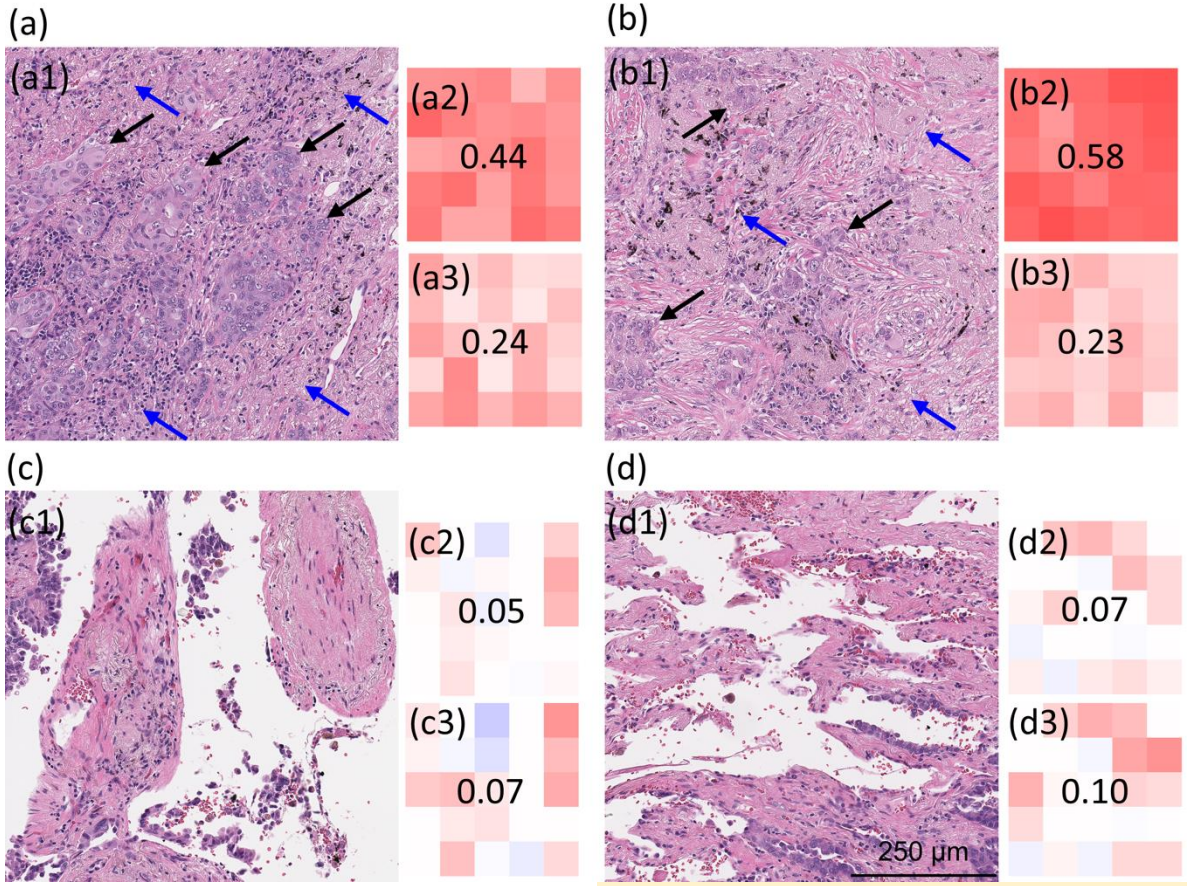




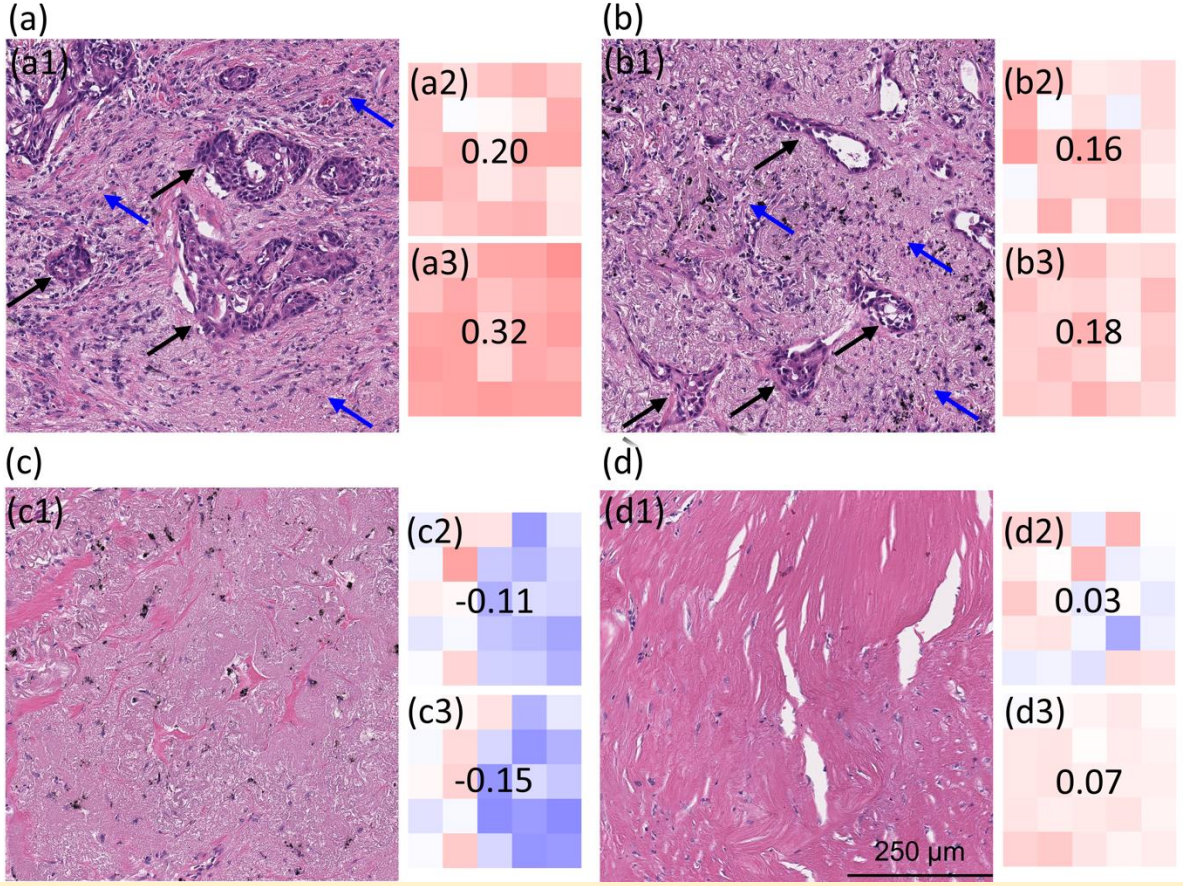
DNN Focuses on Areas with Tumor and TME

Zhou et. al., Length-scale study in deep learning prediction for non-small cell lung cancer brain metastasis, Scientific Reports 2024 in press

Tumor that did not developed metastasis (Met-)



Tumor that developed metastasis (Met+)



Red = increased prediction accuracy

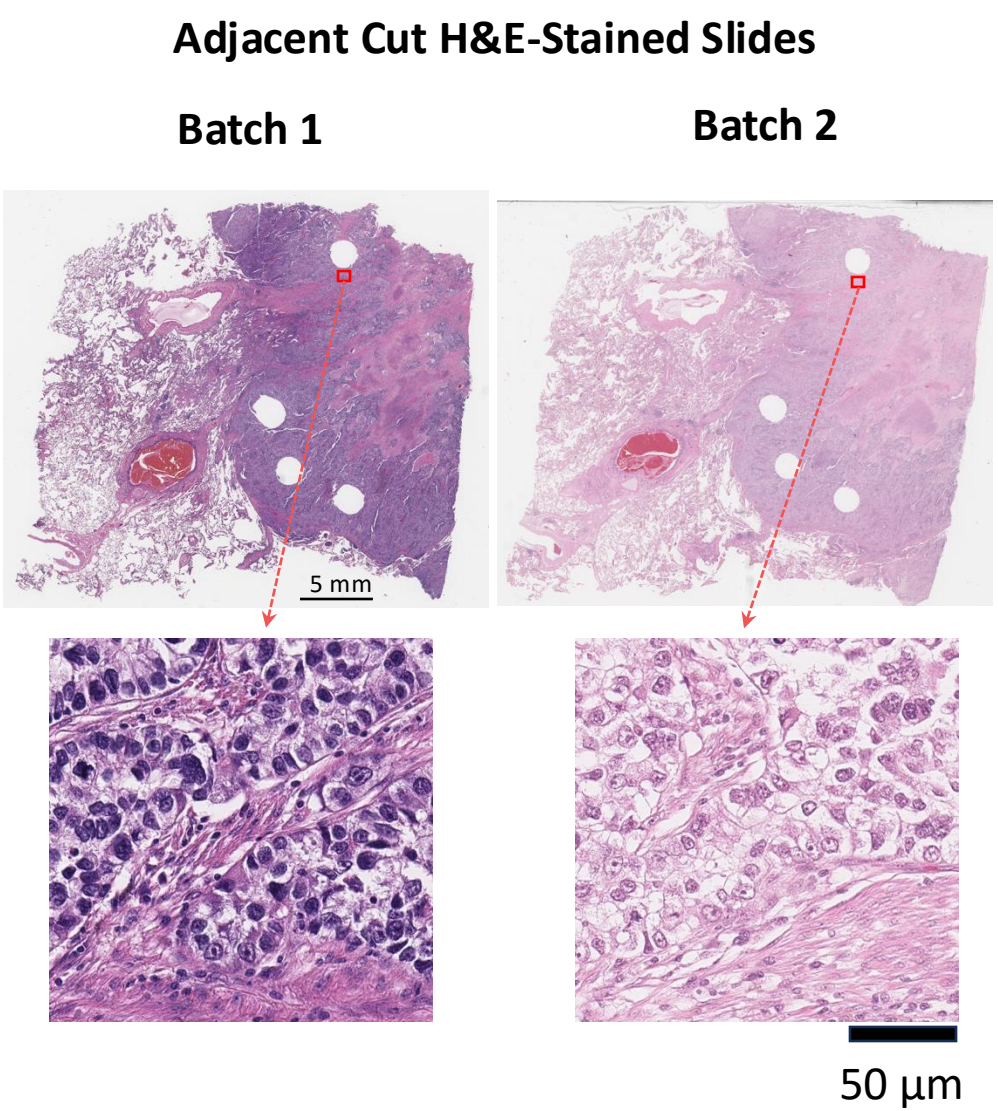
Clear = no prediction accuracy

Blue = negative prediction accuracy

Tiles with high predictive value *always* have tumor (**black arrows**) and TME (blue arrows), whether the prediction is Met+ or Met-.

Tiles with no or negative predictive value *rarely* contained any tumor cells.

Chapter V: The Challenge of Generalizing AI Algorithms: Stain Variation



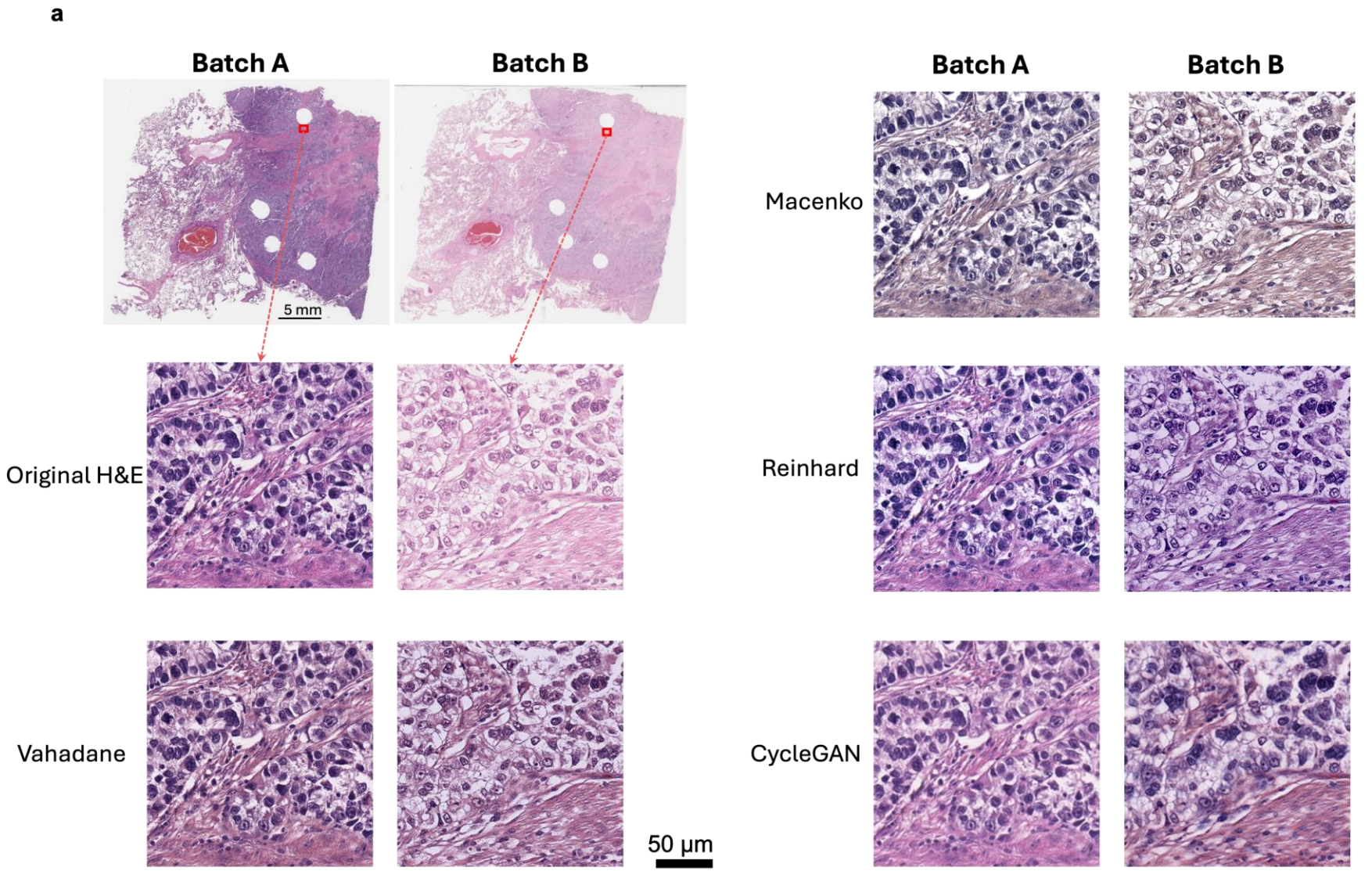
- The inherent variation of H&E vital staining is well recognized in histology
- This is known to be a major problem in AI training on histologic material, and has necessitated such studies to be very large so AI can "train" on these stain variations
- However, when large training sets are not available, such as when specific treatments and outcomes are required, such training sets may not be possible

Goal: Determine the impact of stain/color variation on AI training

Lin et. al., Impact of Stain Variation and Color Normalization for Prognostic Predictions in Pathology, Scientific Reports, in revision

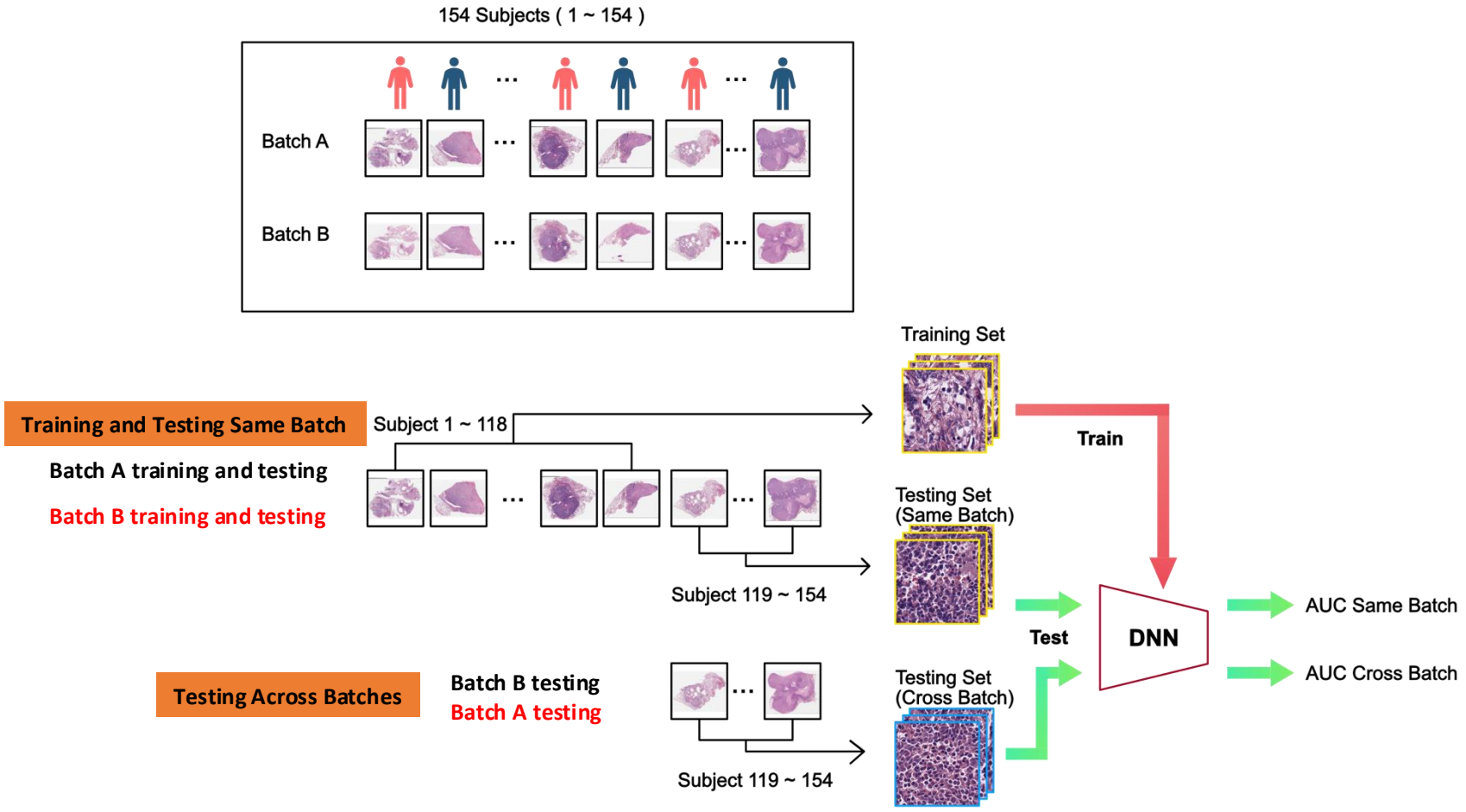
Color Normalization Methods

[1] A. Vahadane et al., IEEE Transactions on Medical Imaging. 2016.
[2] E. Reinhard, et al., 'Color transfer between images', IEEE Computer Graphics and Applications, 2001.
[3] M. Macenko et al., 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009,



Lin et. al., Impact of Stain Variation and Color Normalization for Prognostic Predictions in Pathology, Scientific Reports, in revision

Impact of Color Normalization Schemes and Different DNN Models on AI Training: Cross Batch Testing



Experimental Design:

- Training and Testing in the same batch, per original design
  - Color normalization with different methods across the same batch
- Testing against the other batch (cross batch) using original H&E and the different color normalization methods
- Determine the prediction of metastasis

Lin et. al., Impact of Stain Variation and Color Normalization for Prognostic Predictions in Pathology, Scientific Reports, in revision

Impact of Alternative Color Normalization Schemes on Generalizing the AI Predictive Algorithm

Testing Set	Batch A	Batch B
Original H&E	0.81	0.53
Vahadane	0.96 (p=0.010)	0.60
Reinhard	0.90 (p=0.160)	0.54
Macenko	0.92 (p=0.069)	0.52
Generative Method	0.93 (p=0.069)	0.61

Resnet Train* on Batch A

Testing Set	Batch A	Batch B
Original H&E	0.52	0.74
Vahadane	0.58	0.88 (p=0.033)
Reinhard	0.60	0.85 (p=0.065)
Macenko	0.54	0.90 (p=0.026)
Generative Method	0.52	0.87 (p=0.033)

Resnet Train* on Batch B

Observation 1:
Model predicts well when tested on the same batch of data.
→ Color normalization schemes improve the predictive performance within a batch.

Observation 2:
Model fails to predict across batches, even with normalization schemes.
→ AI failed to generalize in case of variations.

Major Conclusion:
The information needed to predict metastasis is contained within the images, but a predictive AI algorithm trained on one set of slides cannot be generalized to another set of slides despite using a variety of color normalization schemes.

* DNN used in original J Path 2024 study

Impact of Alternative DNN Model on Generalizing the AI Predictive Algorithm

Testing Set	Batch A	Batch B
Original H&E	0.80	0.55
Vahadane	0.92 (p=0.052)	0.60
Reinhard	0.94 (p=0.016)	0.58
Macenko	0.88 (p=0.127)	0.55
Generative Method	0.90 (p=0.127)	0.62

Prov-Giga Train* on Batch A

Testing Set	Batch A	Batch B
Original H&E	0.54	0.77
Vahadane	0.63	0.91 (p=0.021)
Reinhard	0.62	0.87 (p=0.133)
Macenko	0.59	0.90 (p=0.060)
Generative Method	0.57	0.91 (p=0.060)

Prov-Giga Train* on Batch B

Observation 3:
Alternative DNN can be used to develop the metastasis prediction algorithm.

- Color normalization improves the prediction performance within a batch

Observation 4:
Despite using a different DNN model, and applying color normalization, the AI failed to generalize across batches.

Major Conclusion:
The information needed to predict metastasis is contained within the images, but a predictive AI algorithm trained on one set of slides cannot be generalized to another set of slides, despite using color normalization and alternative DNN models.

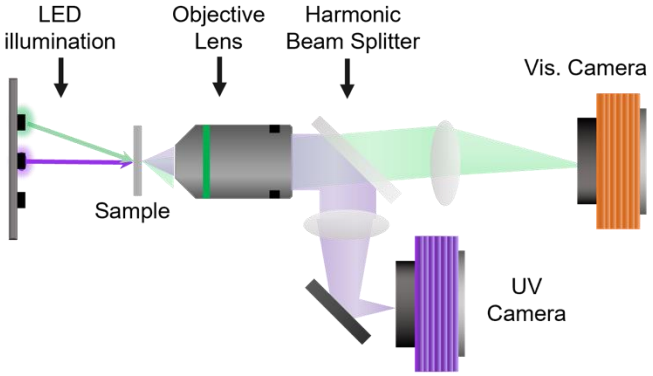
* Alternative DNN

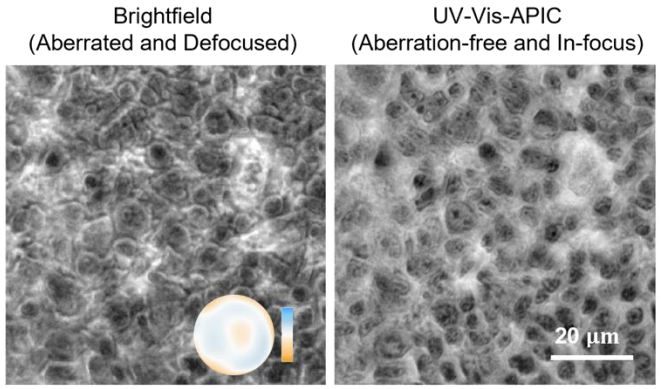
Chapter VI. What's Next

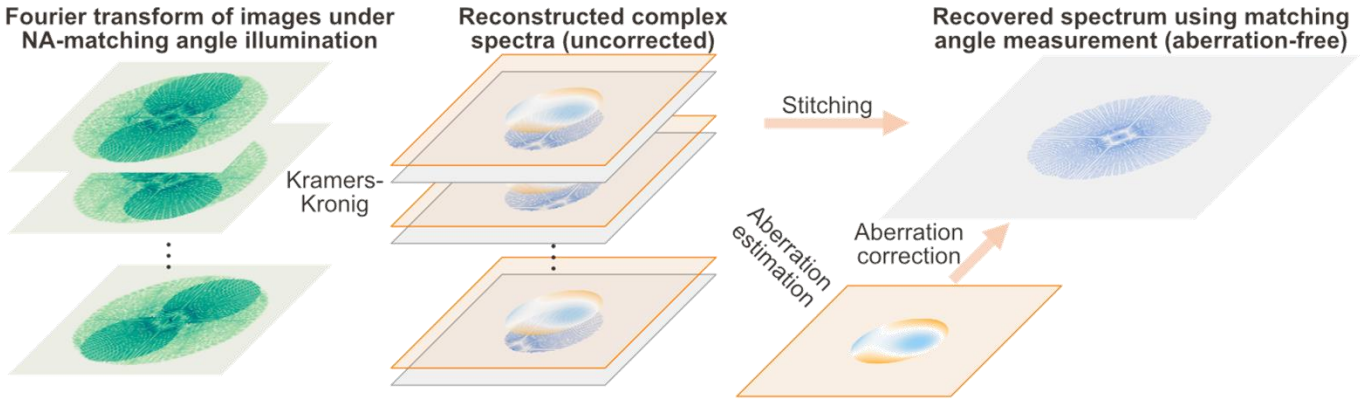
- How do you solve a problem like color?
  - How about eliminating it!
  - *Apologies to Rogers and Hammerstein*

- Other predictive applications
  - Predicting protein and gene expression based on an H&E slide (?!)

Solving Stain Variation: UV and Phase Computational Microscopy Provides Contrast Without Staining







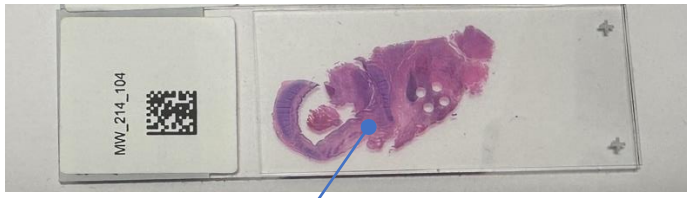
UV: Ultraviolet
VIS: visible light
APIC: **A**ngular **P**tychographic **I**maging with **C**losed-form method, an advanced version of FPM.

Note that cells and substructures, such as nuclei and nucleoli can easily be seen

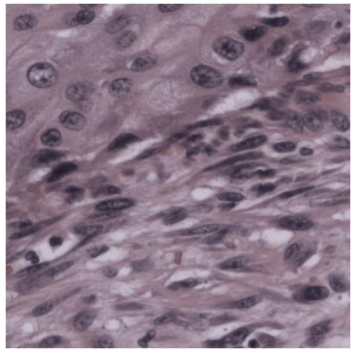
Tissues, cells and substructures have specific absorption patterns. Using unstained tissue sections, we then “image” with UV and visible light, capturing the absorption patterns, and also the refractive index (phase) image, which captures structure. This normally produces an aberrated and defocused image, which is corrected by APIC. Note the ability to easily detect cells, nuclei and nucleoli.

Work in progress.

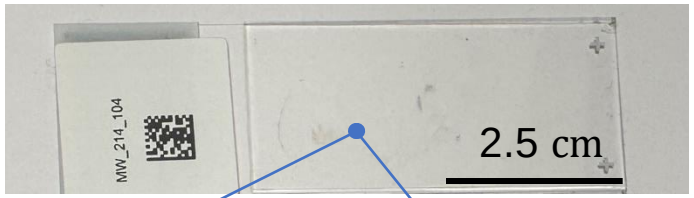
Using Phase Contrast and UV imaging on Unstained Slides: Solving the problem of stain variation?



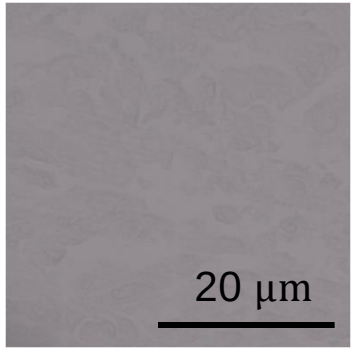
Brightfield microscope



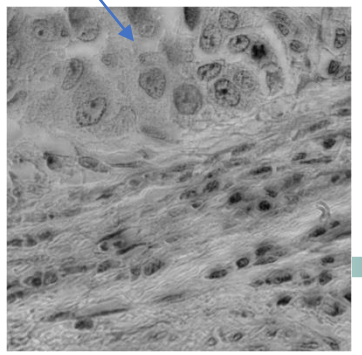
H&E stained tissue



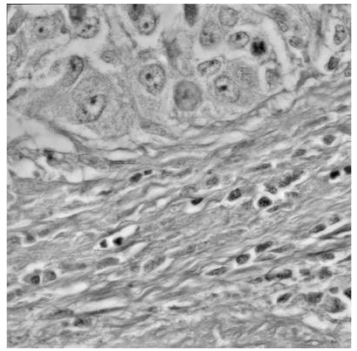
Brightfield microscope



Phase image
unstained tissue



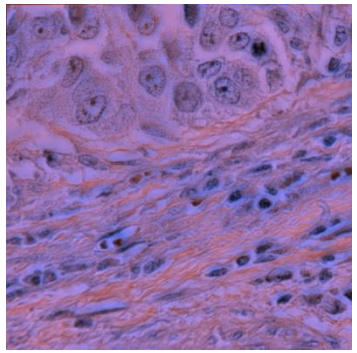
UV channel



Green (vis) channel

Note that UV and visible light produce different absorption spectra, both providing tissue, cell and substructure detail

UV-Vis-APIC

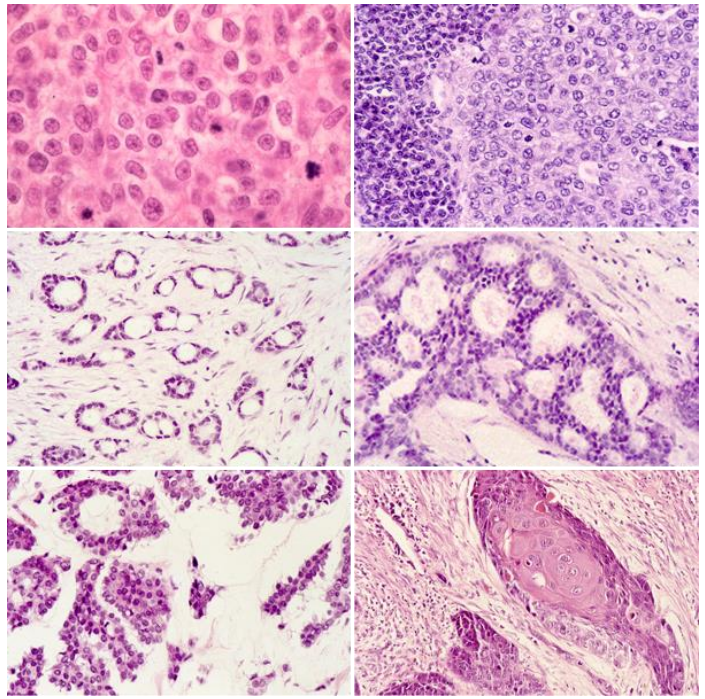


Pseudo-colored

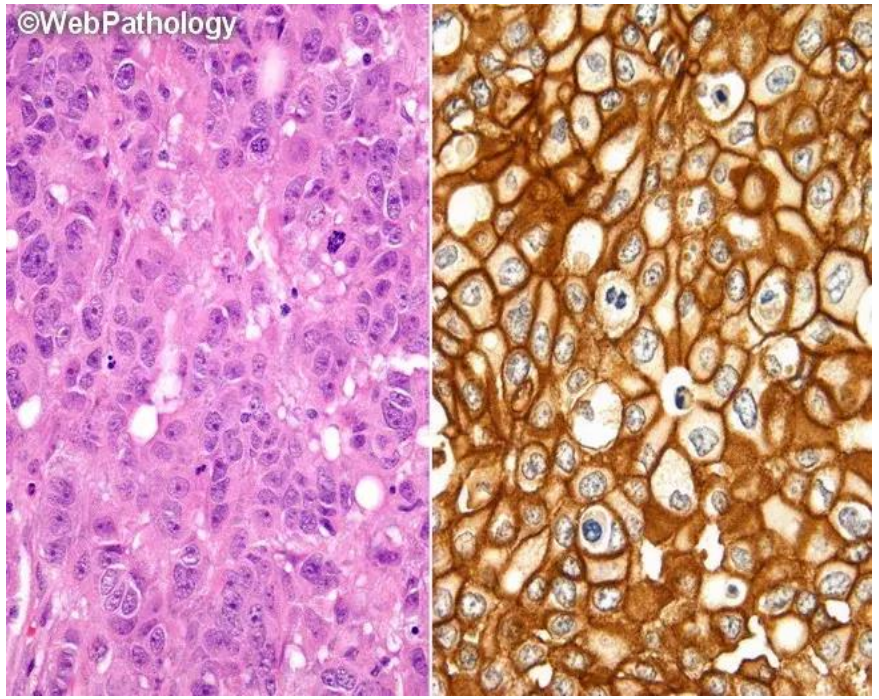
Work in progress.

Can AI Predict Protein and Gene Expression Based on Histology: Morphology Reflects Molecular Biology

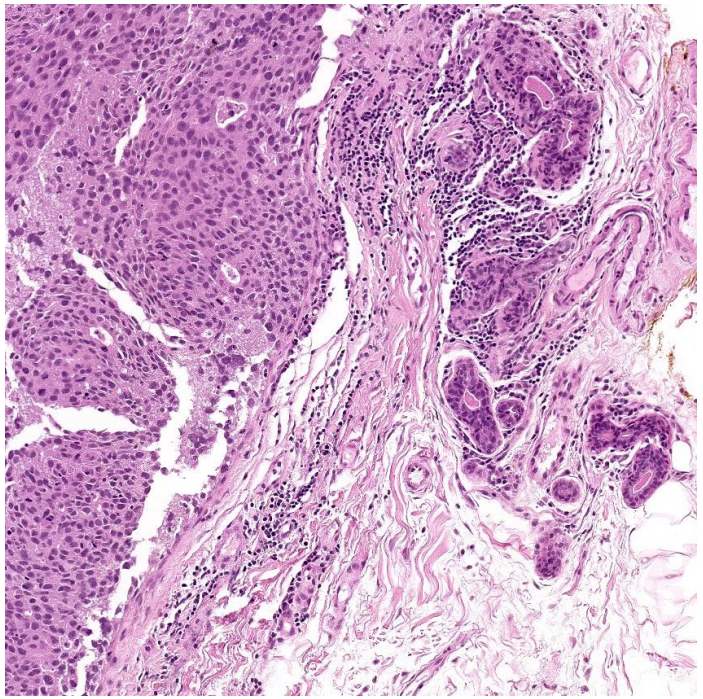
Breast Cancer



ER/PR +



Her2 +



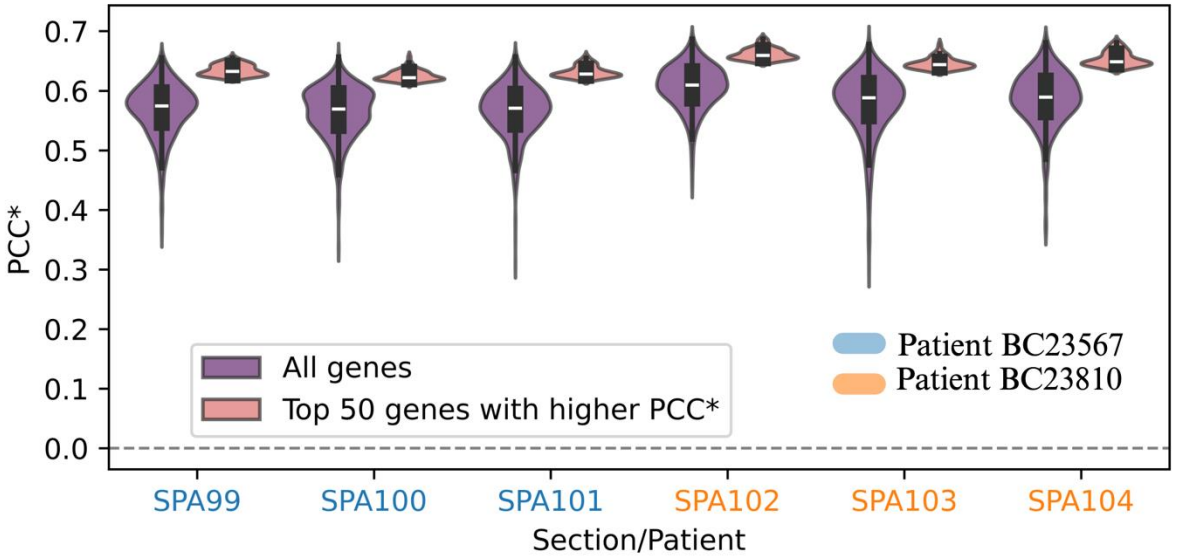
Triple Negative

Toward 3D spatial transcriptomics using deep learning framework

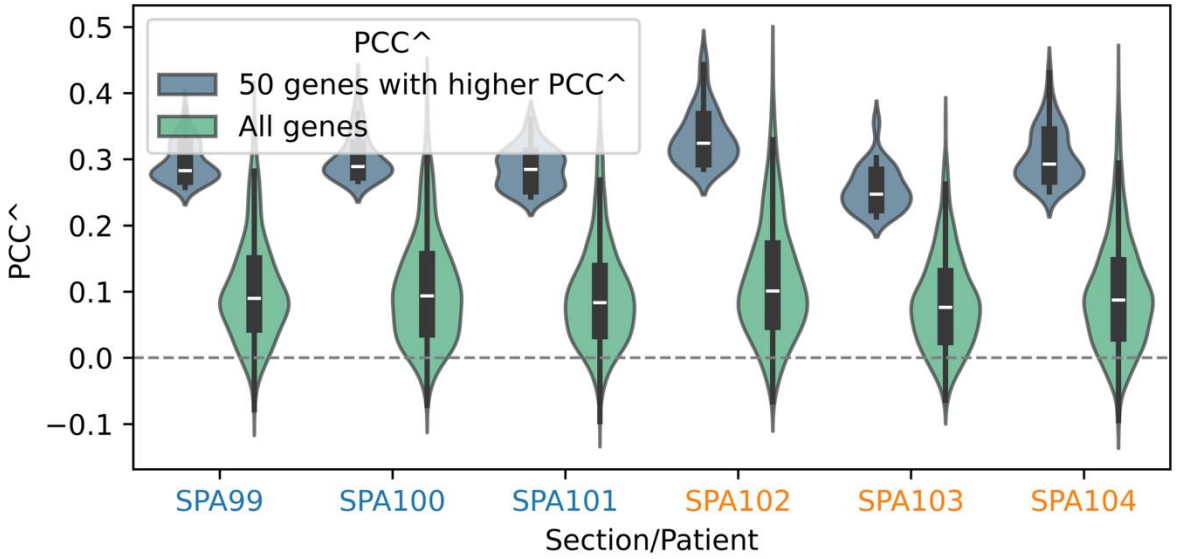
PCC*, how accurately the predicted spatial gene expression matches the actual detected expression at every spot.

PCC^, how accurately the expression of a single gene is predicted across multiple spatial spots in one tissue section

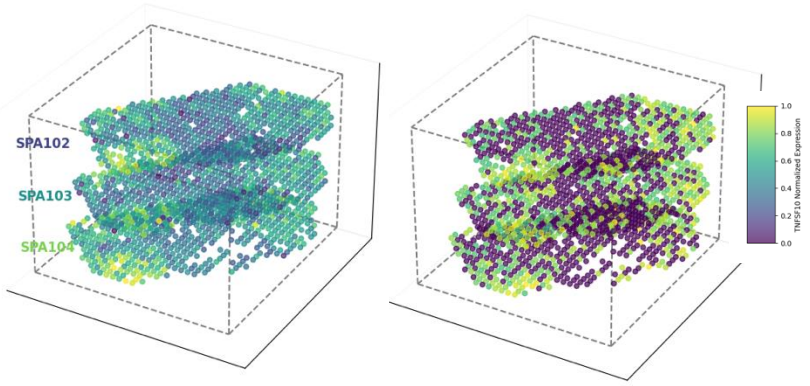
a



b



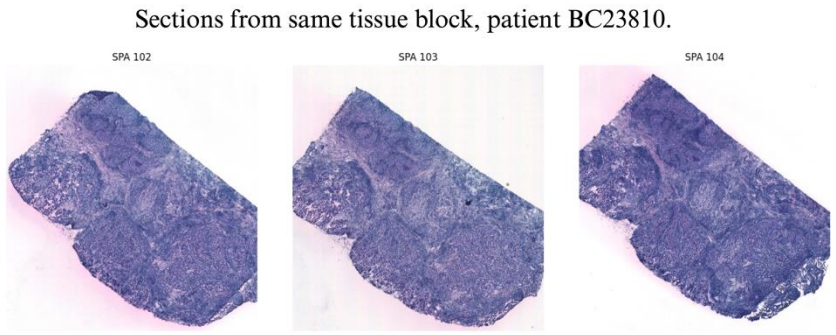
c



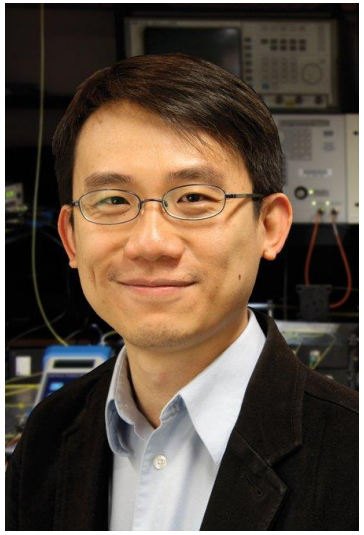
$r_{PCC^{\wedge}} = 0.391$

$p < 0.05$

d



Collaborators and Lab


Yang


Watson


Williams


Govindan


Kepecs


Artyomov


Shen


Zhou


Liang


Lin


Liao


Paul

Thank You