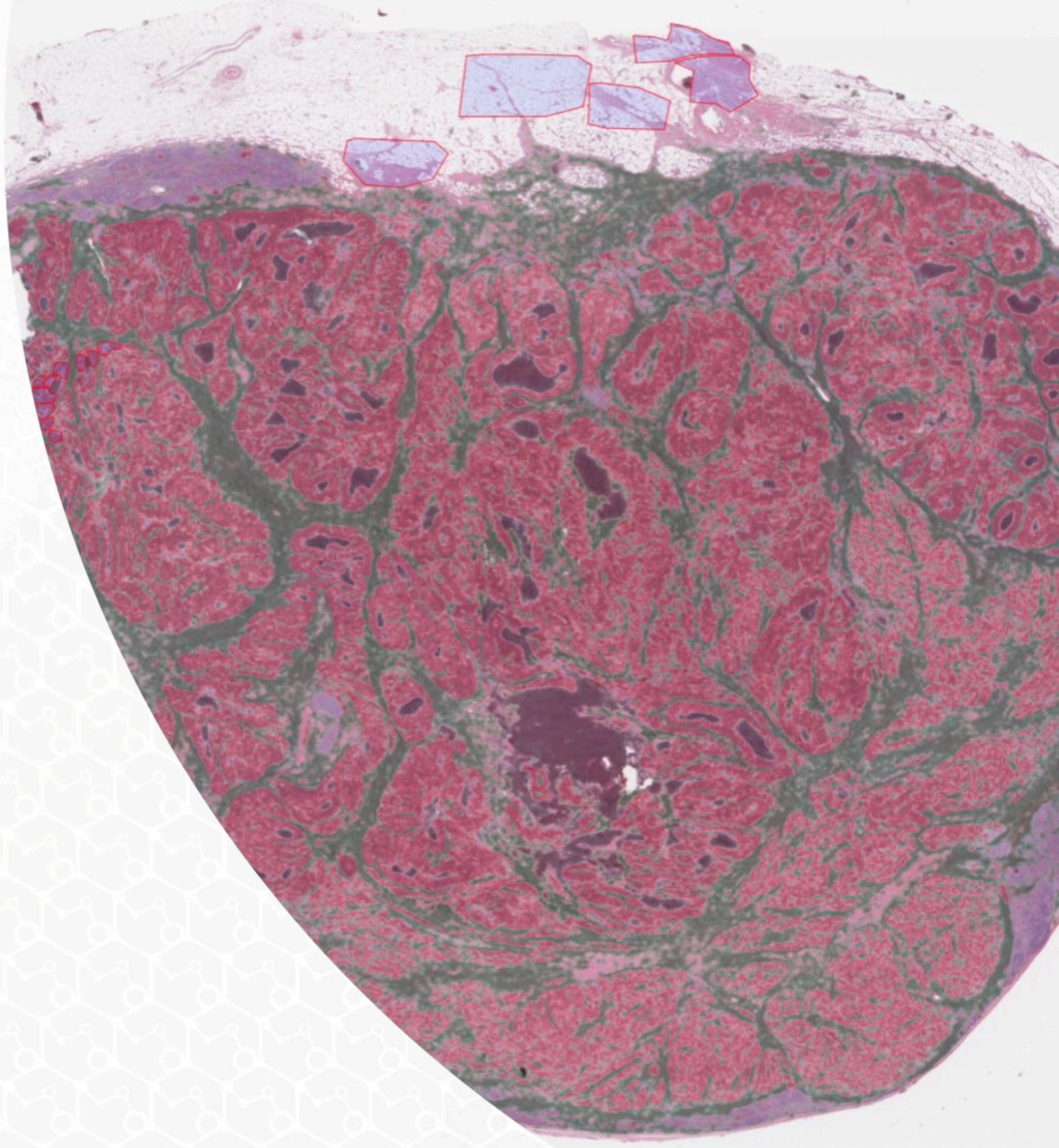
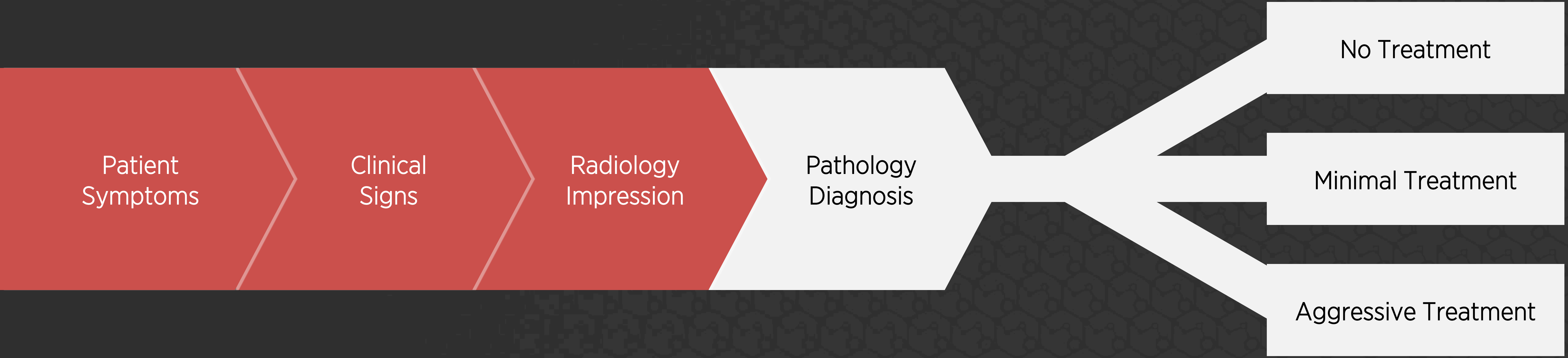
Machine learning for Pathology

Andrew H Beck MD PhD
CEO @ PathAI

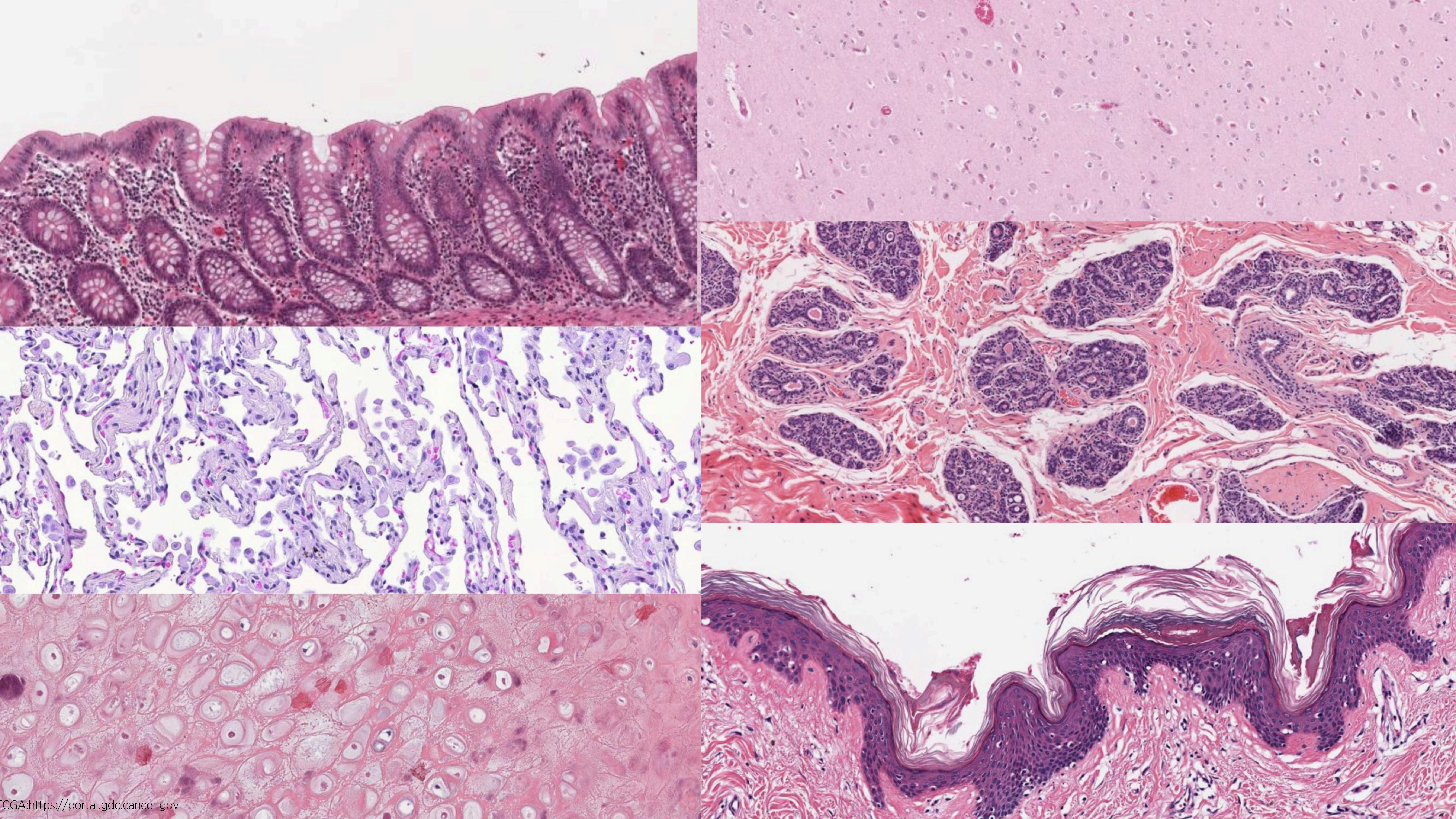


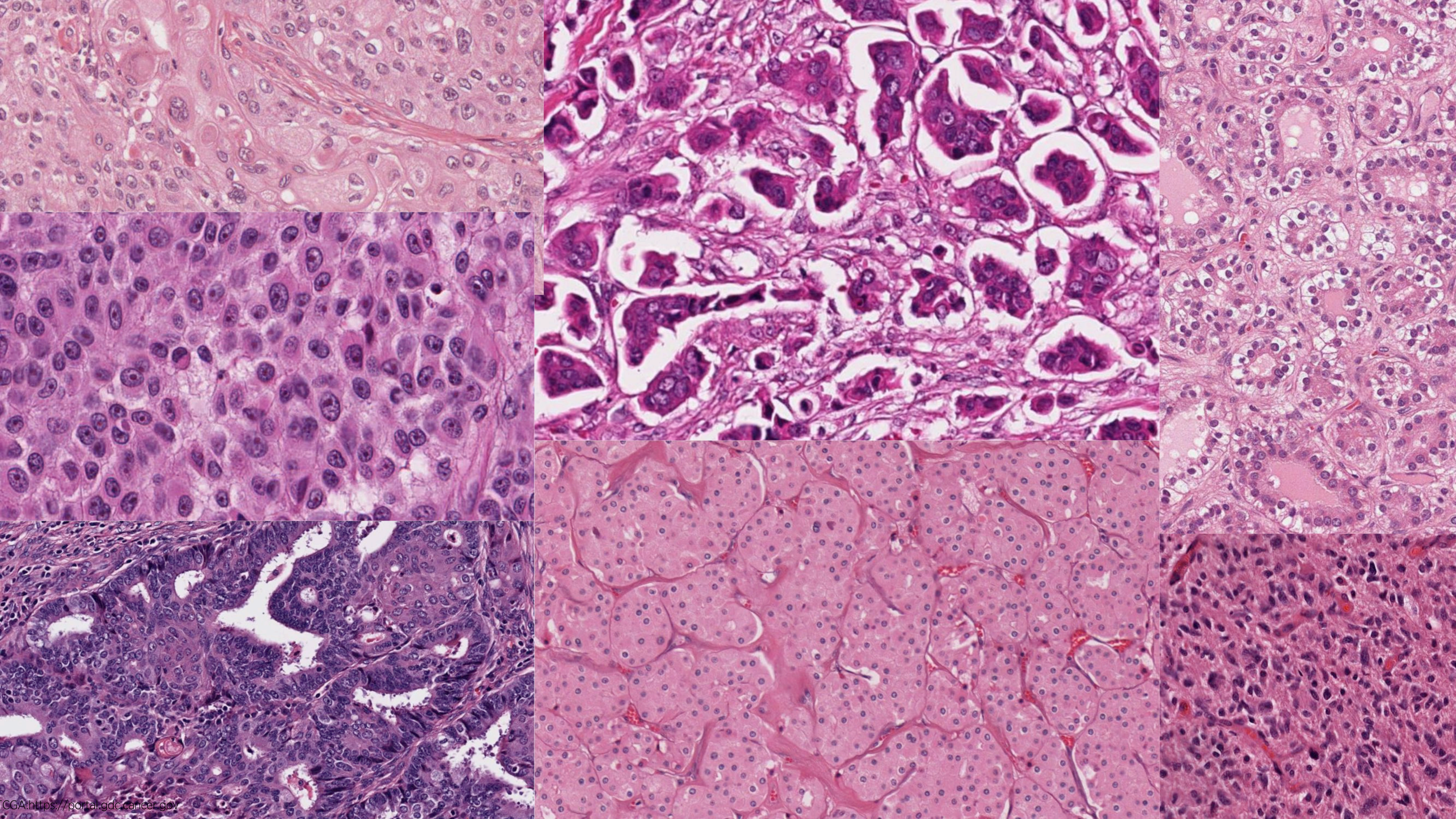
6.S897/HST.956: Machine Learning for Healthcare. MIT.
March 19, 2019

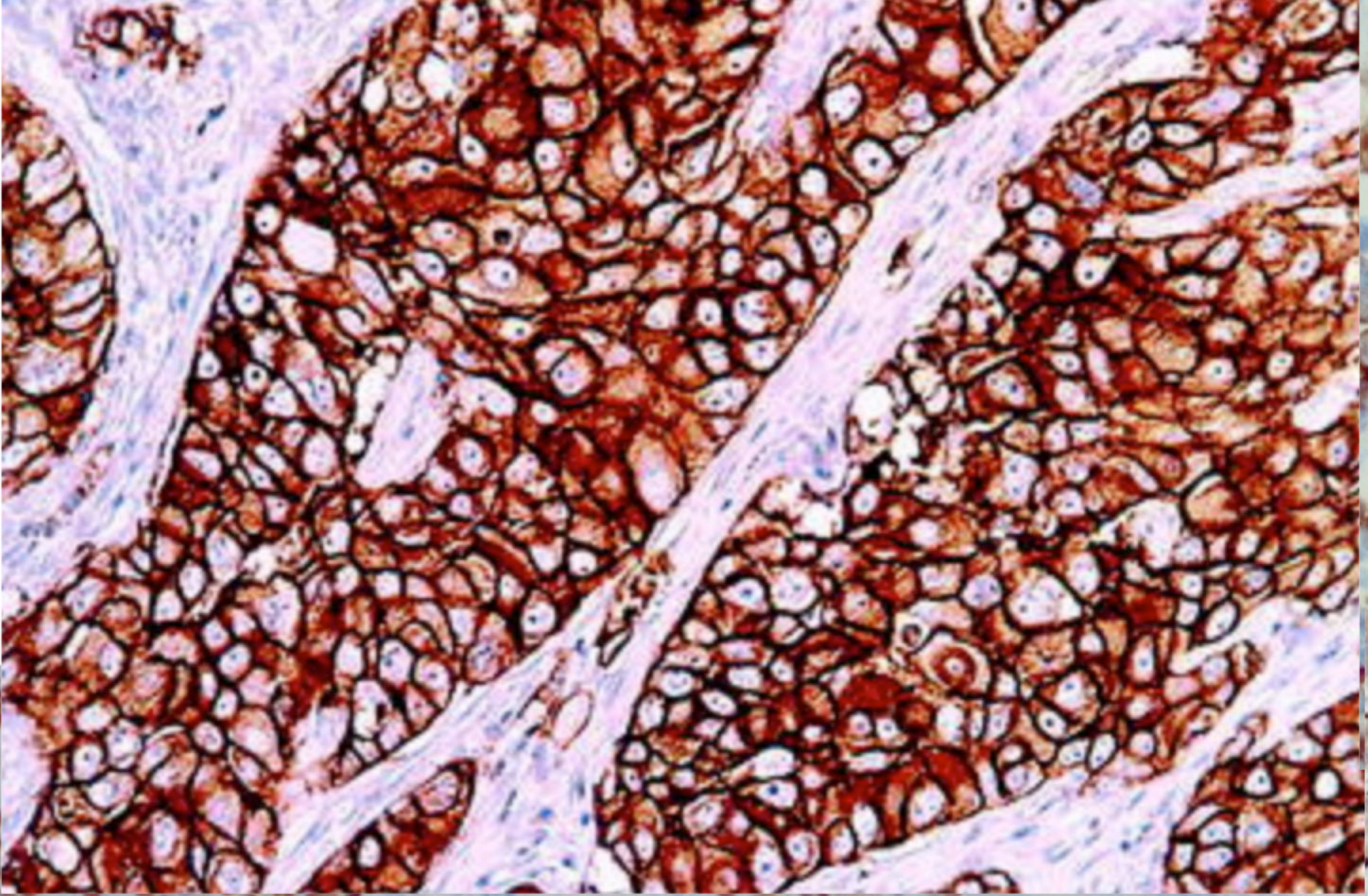
Pathology

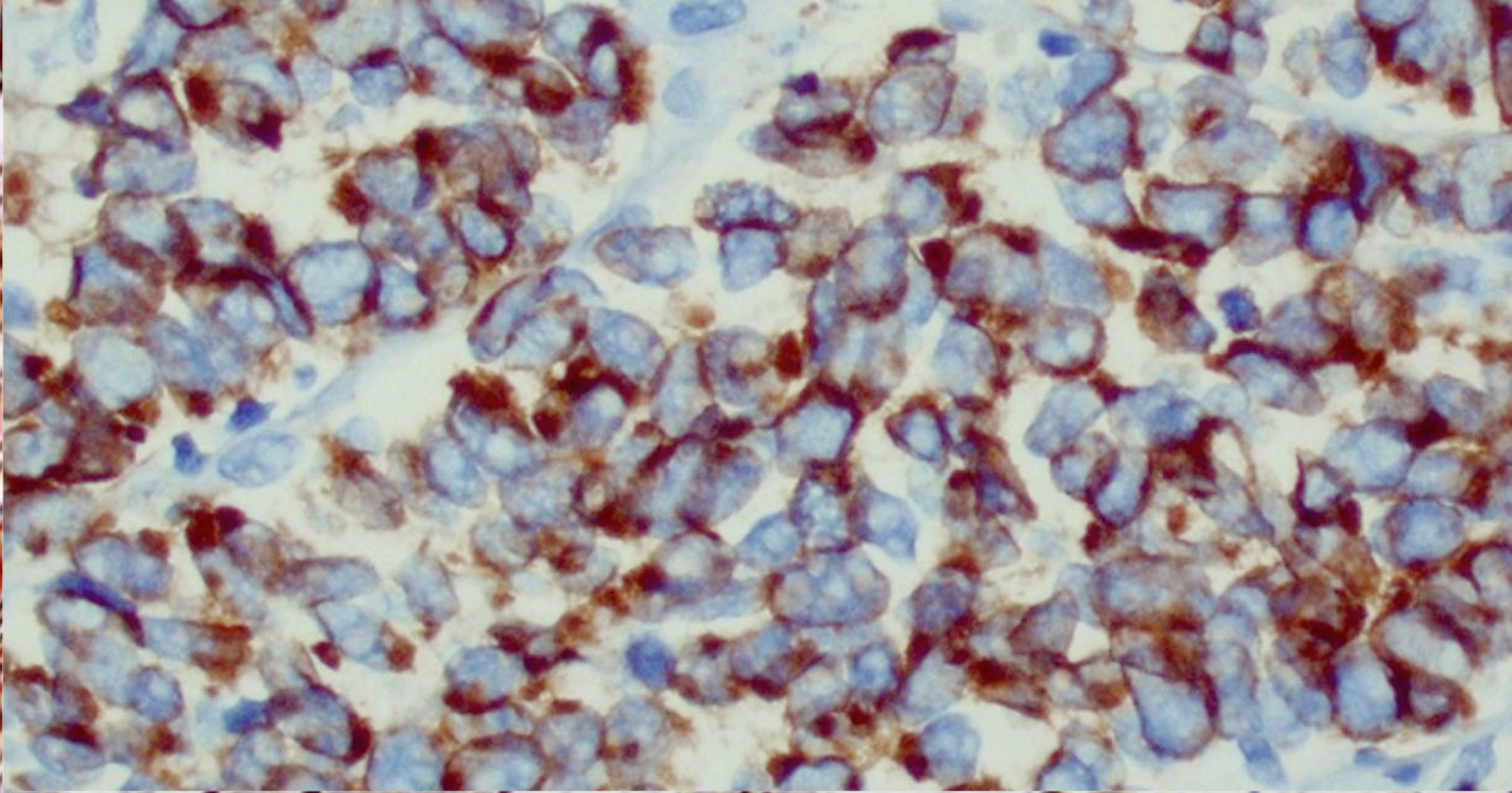


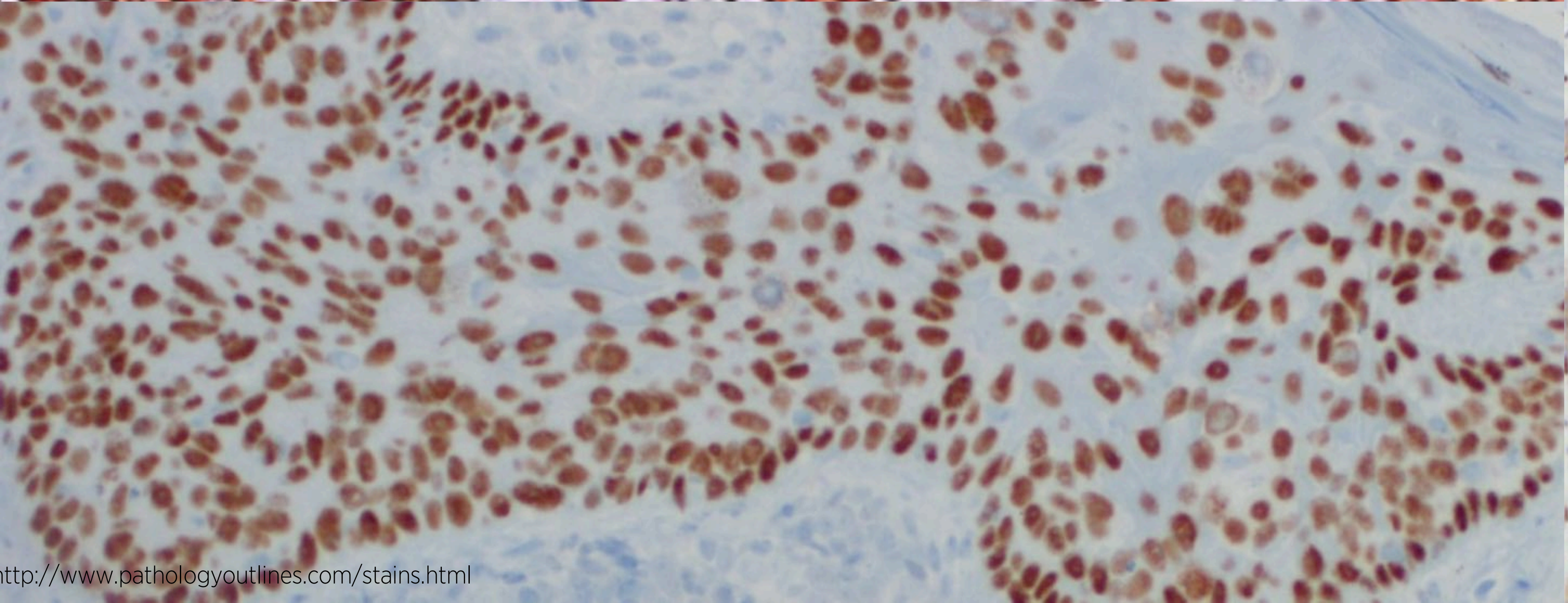
Pathologic diagnosis is a central determinant of therapeutic decisions.

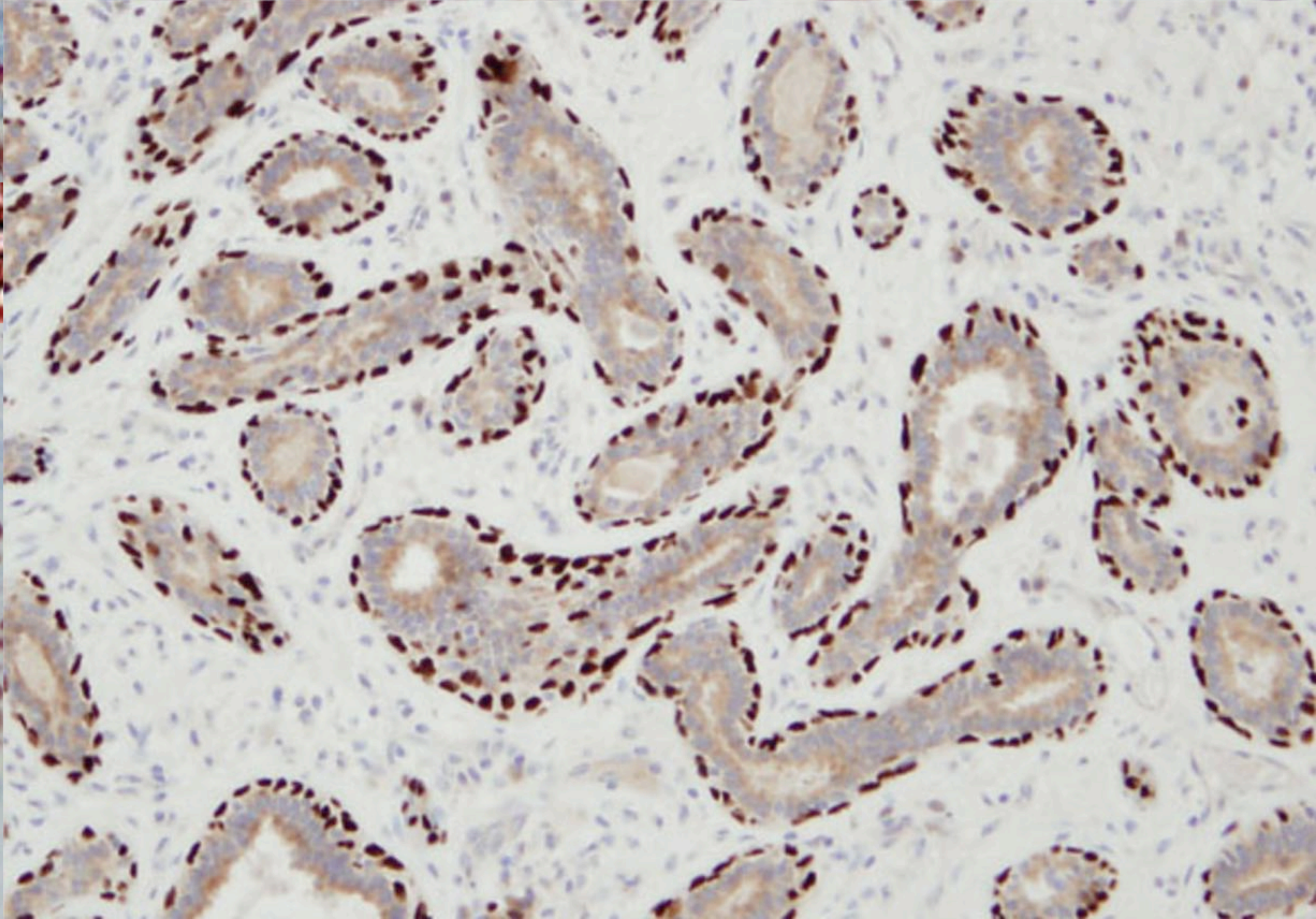

TCGA:https://portal.gdc.cancer.gov


TCGA:https://portal.gdc.cancer.gov

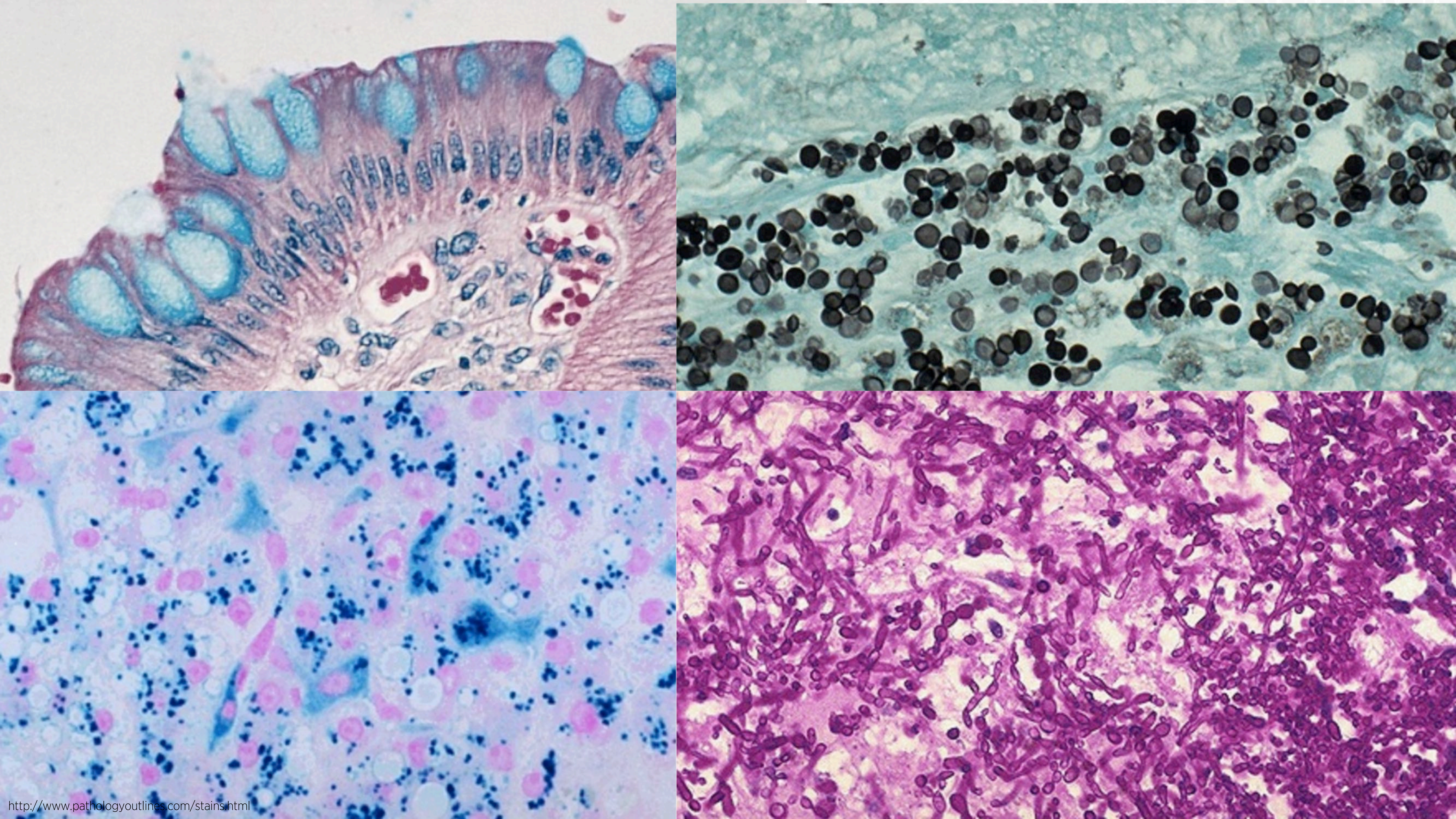








http://www.pathologyoutlines.com/stains.html


http://www.pathologyoutlines.com/stains.html

Emergence of early computational approaches in Pathology (1981)

MORPHOMETRY FOR PROGNOSIS PREDICTION IN BREAST CANCER

SIR,—Some workers have found a correlation between prognosis and microscopical features of the primary tumour in breast cancer[1-3] but in one large prospective study the significance of the nuclear and histological grade for prognosis was weak.[4] Disagreement in grades assigned to the same tumours by different pathologists may range up to 40%,[5,6] and this disagreement may be due to the subjective nature of histopathological assessment. In contrast, the advantages of morphometry are objectivity and high reproducibility.[7]

PERCENTAGE CORRECTLY PREDICTED PROGNOSES

Method	Total (n=78)	Learning set (n=38)	Test set (n=40)
ANS	59	65	54
TNM	64	67	56
Morphometry	87	92	78

Baak et al. Lancet 1981

PathAI
Proprietary & Confidential

Artificial Neural Nets in Quantitative Pathology (1990)

Anal Quant Cytol Histol. 1990 Dec;12(6):379-93. Paperpile

Artificial neural networks and their use in quantitative pathology.

Dytch HE[1], Wied GL.

"It is concluded that artificial neural networks, used in conjunction with other nonalgorithmic <u>artificial intelligence techniques</u> and traditional algorithmic processing, may provide useful software engineering tools for the development of systems in quantitative pathology."

Emergence of Digital Pathology (2000)

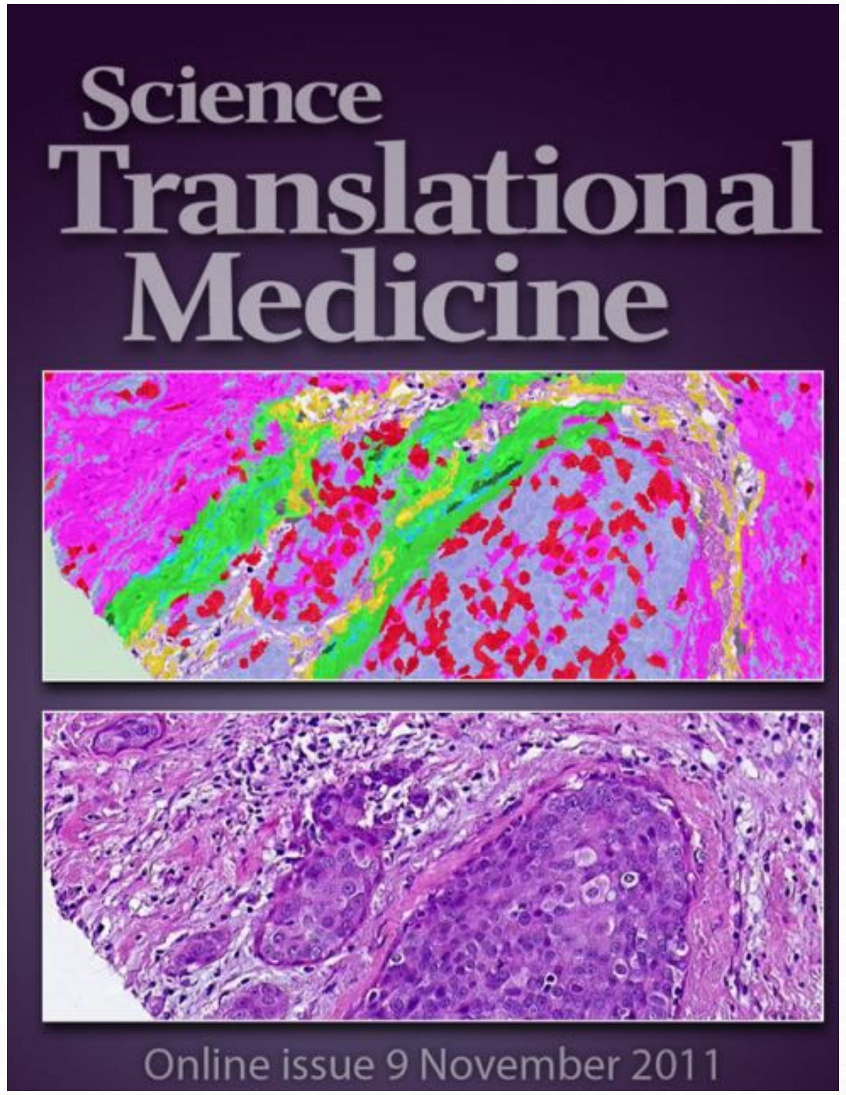
International Journal of Surgical Pathology 8(4):261–263, 2000

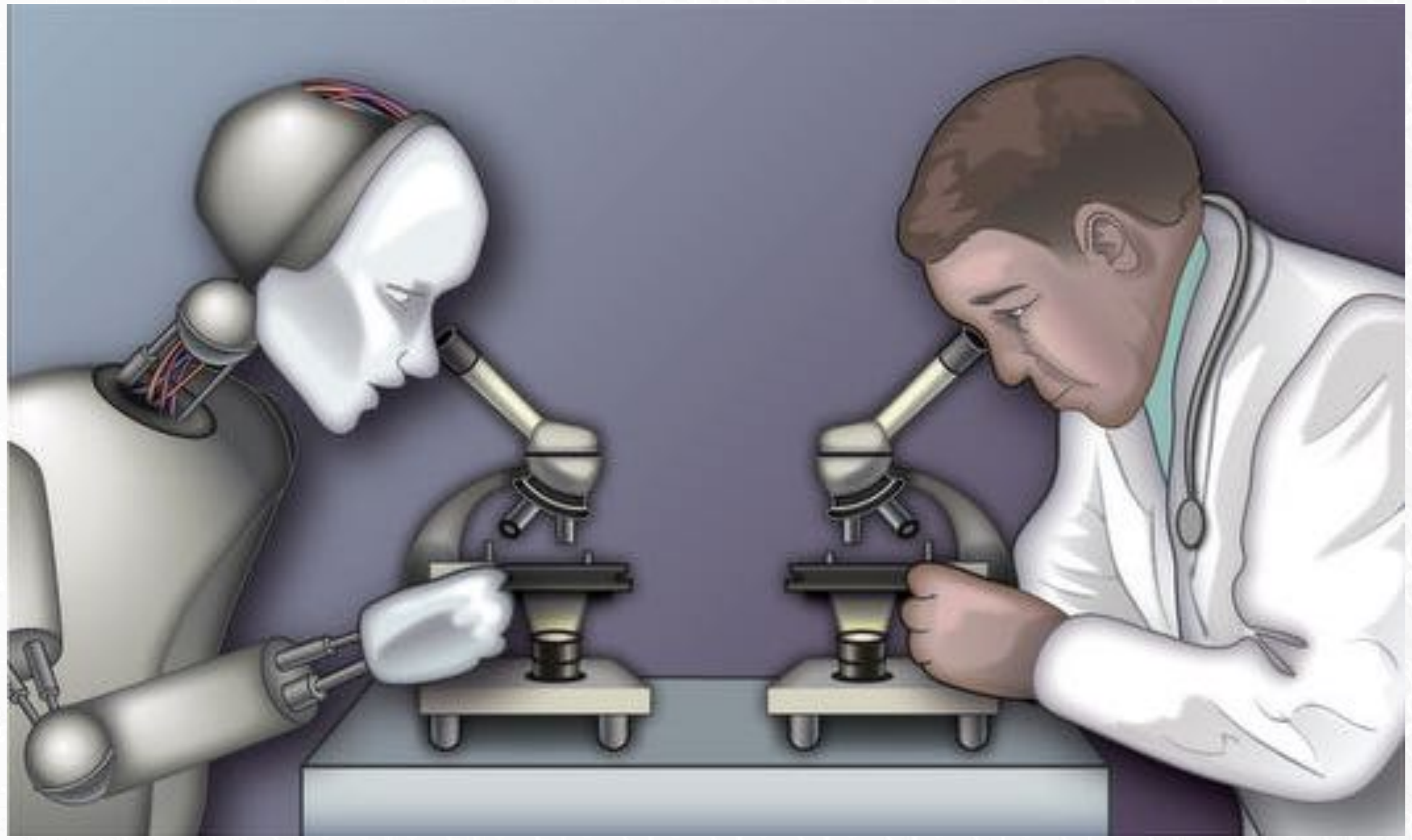
Digital Pathology: Science Fiction?

Mattia Barbareschi,* Francesca Demichelis,† Stefano Forti,†
and Paolo Dalla Palma*

But what will come next? Is it possible to hypothesize that VC will completely substitute our traditional glass slides? Maybe yes, and let us describe the "science fiction" new millennium *digi*tal *path*ology laboratory, which we will call "DIGIPATH."

The emergence of machine learning-based approaches for cancer histopathology

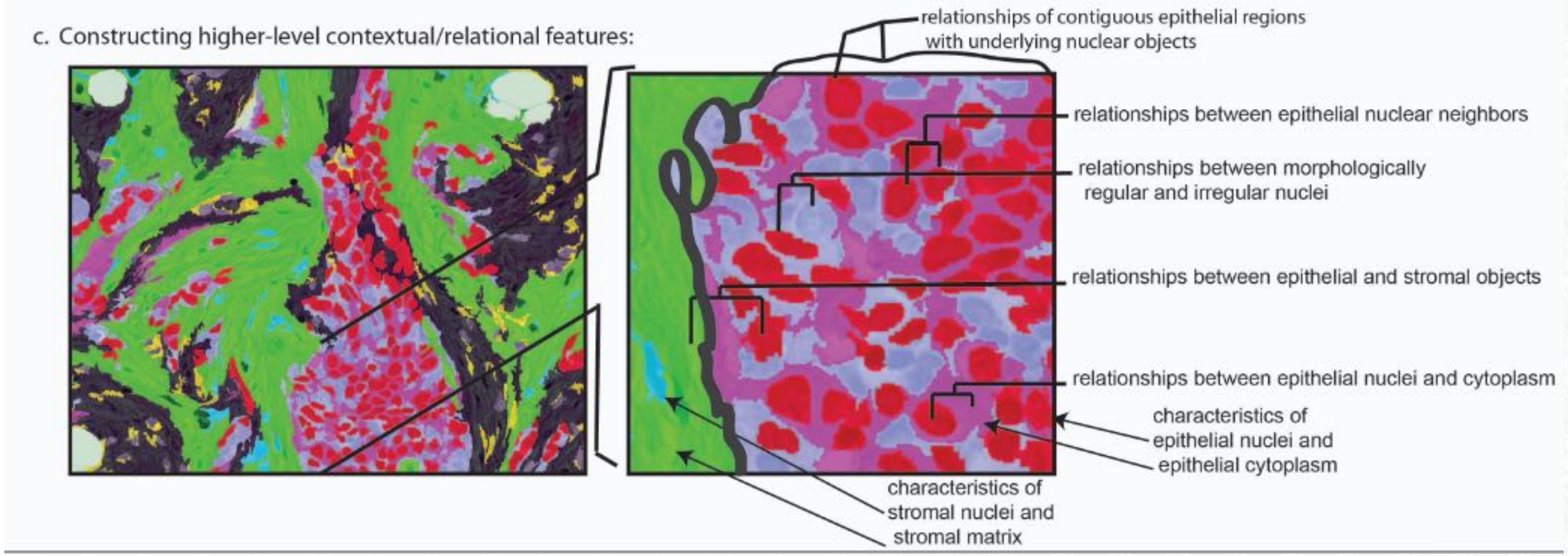




PathAI
Proprietary & Confidential

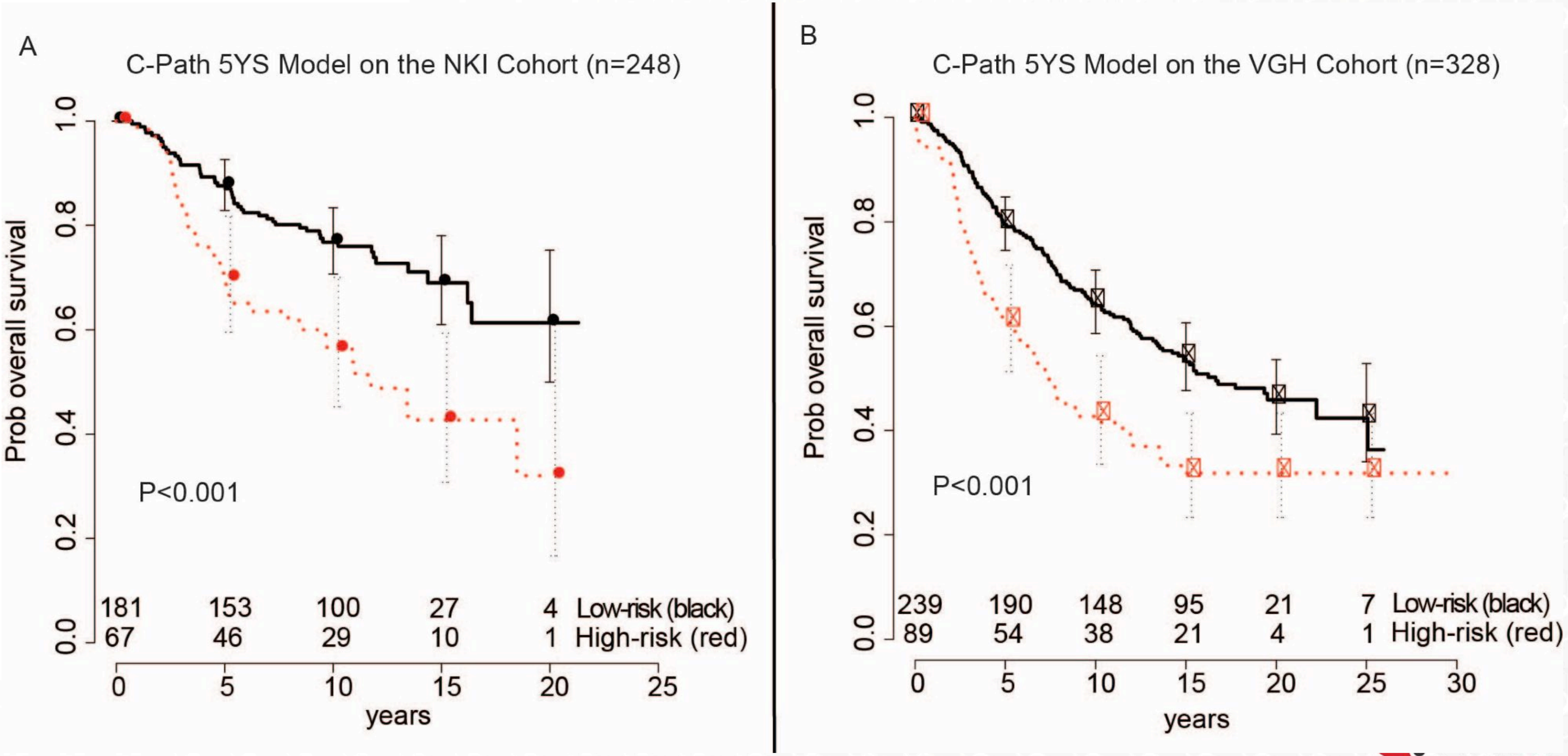
Beck ... Koller. Science Translational Medicine 2011

Extracting a rich quantitative feature set



Beck ... Koller. Science Translational Medicine 2011

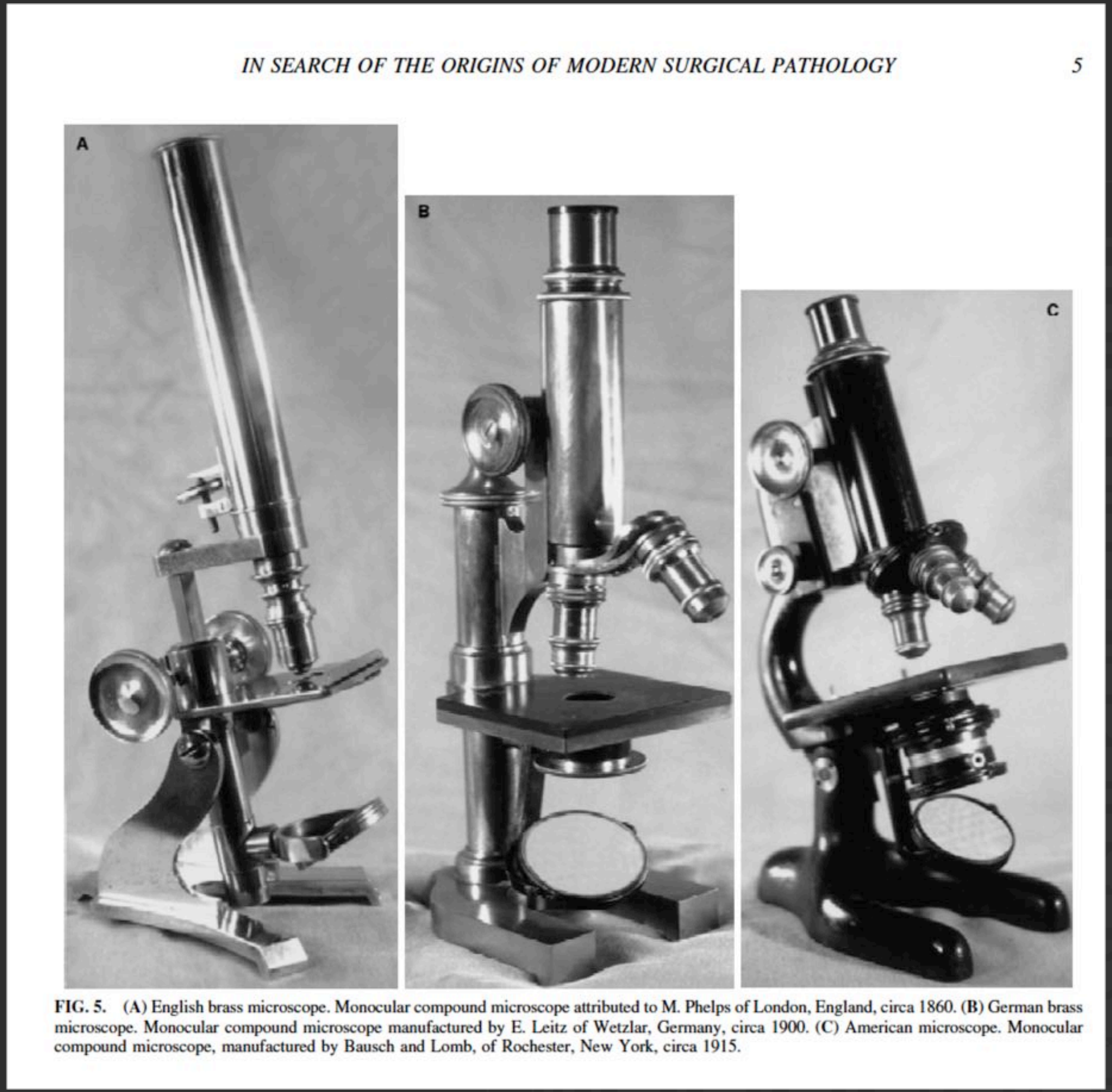
C-Path 5YS Score Significantly Associated with Overall Survival on Both Cohorts



Beck ... Koller. Science Translational Medicine 2011

Even today, the anatomic path lab has been largely unchanged for routine diagnostics

And core technology breakthroughs in routine use are from the 19th century



A

B

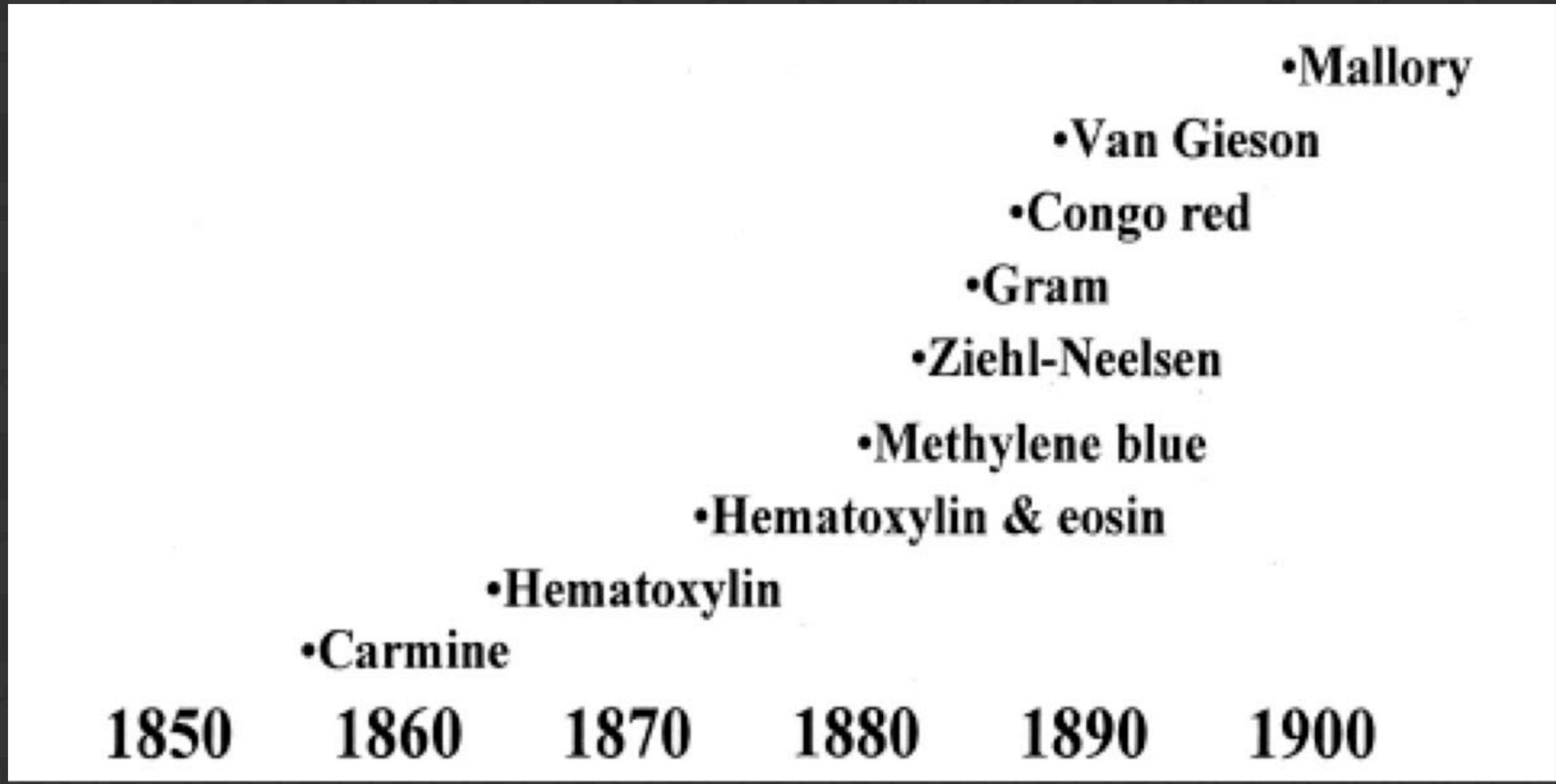
C

FIG. 5. (A) English brass microscope. Monocular compound microscope attributed to M. Phelps of London, England, circa 1860. (B) German brass microscope. Monocular compound microscope manufactured by E. Leitz of Wetzlar, Germany, circa 1900. (C) American microscope. Monocular compound microscope, manufactured by Bausch and Lomb, of Rochester, New York, circa 1915.

Adv Anat Pathol. 2001 Jan;8(1):1-13.

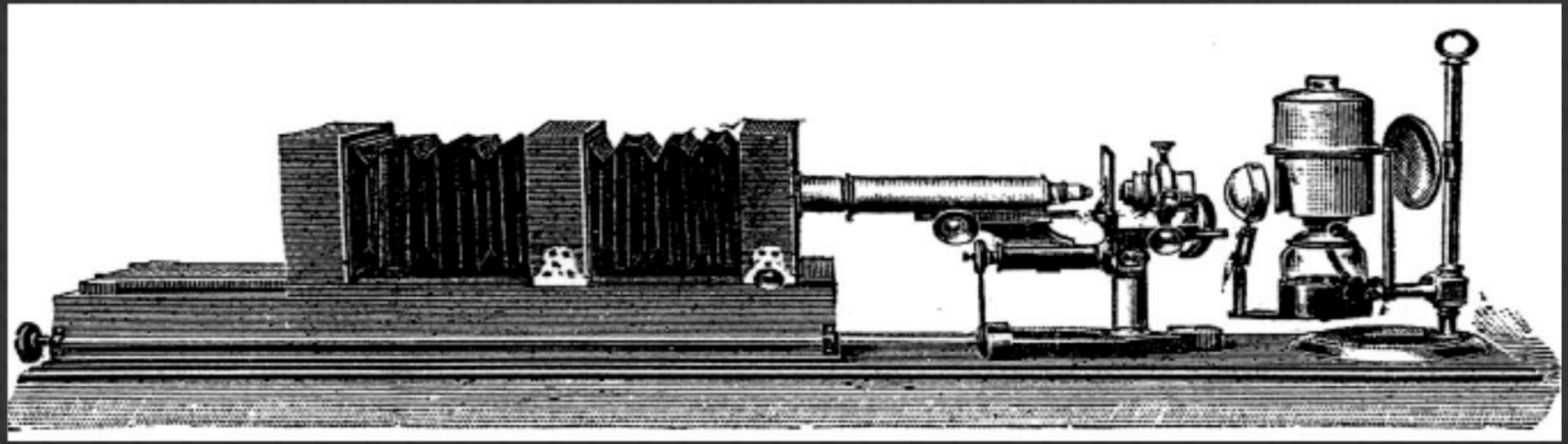
Histochemical Stains

Developed from combinations of analine and natural dyes in the later half of the 19th century

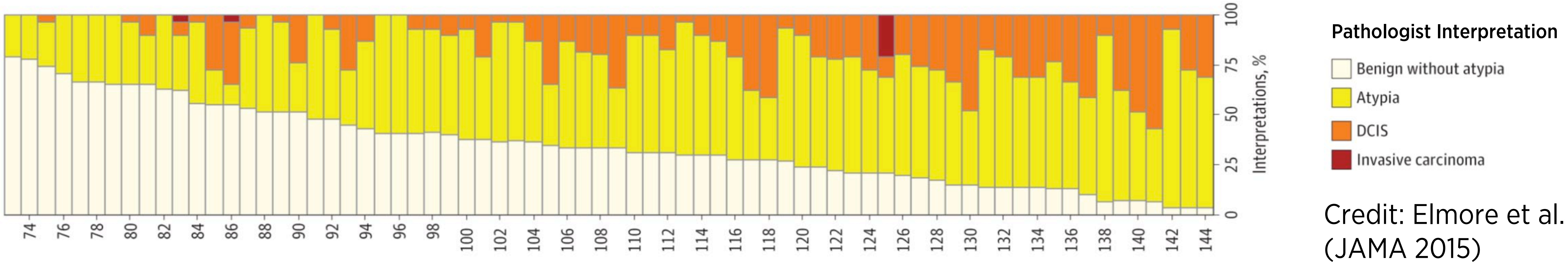


Photomicroscope

Horizontal apparatus with camera, microscope, and light source, 1895.



Discordance among pathologists is common in interpretation of breast biopsies



Credit: Elmore et al. (JAMA 2015)

Phase I Interpretation of Individual pathologist	Phase II Interpretation of Same Individual Pathologist: Benign without atypia	Atypia	DCIS	Invasive	Total	Agreement rates of phase I and II interpretations, % (95% CIs)
Benign without atypia	**947**	137	41	5	1130	84 (81-86)
Atypia	157	**303**	109	2	571	53 (47-59)
Ductal Carcinoma *in situ* (DCIS)	43	94	**792**	14	943	84 (81-87)
Invasive Breast Cancer	8	4	11	**273**	296	92 (88-95)
Total	1155	538	953	294	2940	79 (77-81)

*The same slide was interpreted on two different occasions separated in time by 9 or more months

Pathologists in individual practice setting

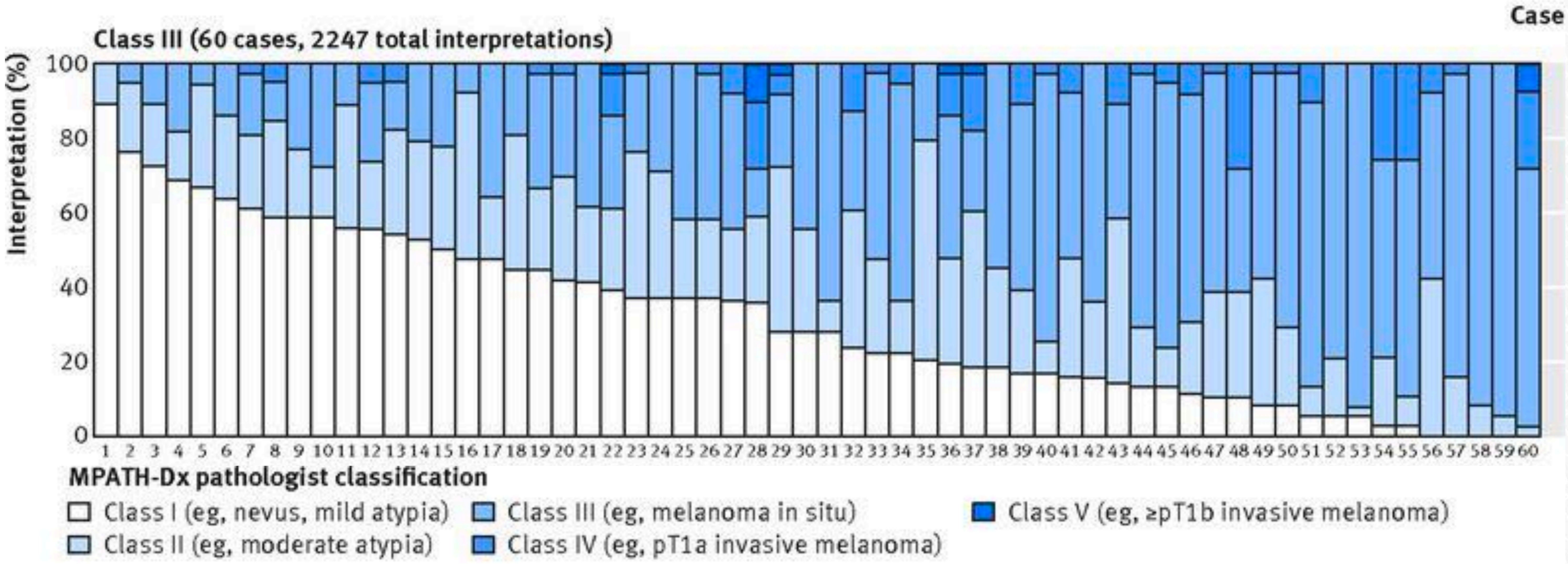
Overall concordance rate of 75% on breast biopsies.

Inter-observer concordance rate of only 48% for a diagnosis of atypia.

Intra-observer concordance is only 79% overall and 53% for atypical lesions

Ref: Jackson SL … Elmore JG. Ann Surg Oncol. 2017 May;24(5):1234-1241.

Discordance among pathologists is common in interpretation of melanocytic neoplasms on skin biopsies



- 187 pathologists interpreted skin lesion biopsies, resulting in an overall discordance of 45%
- 118 pathologists read the same samples 8 months apart, and had an intraobserver discordance of 33%

BMJ 2017;357:j2813 | doi: 10.1136/bmj.j2813

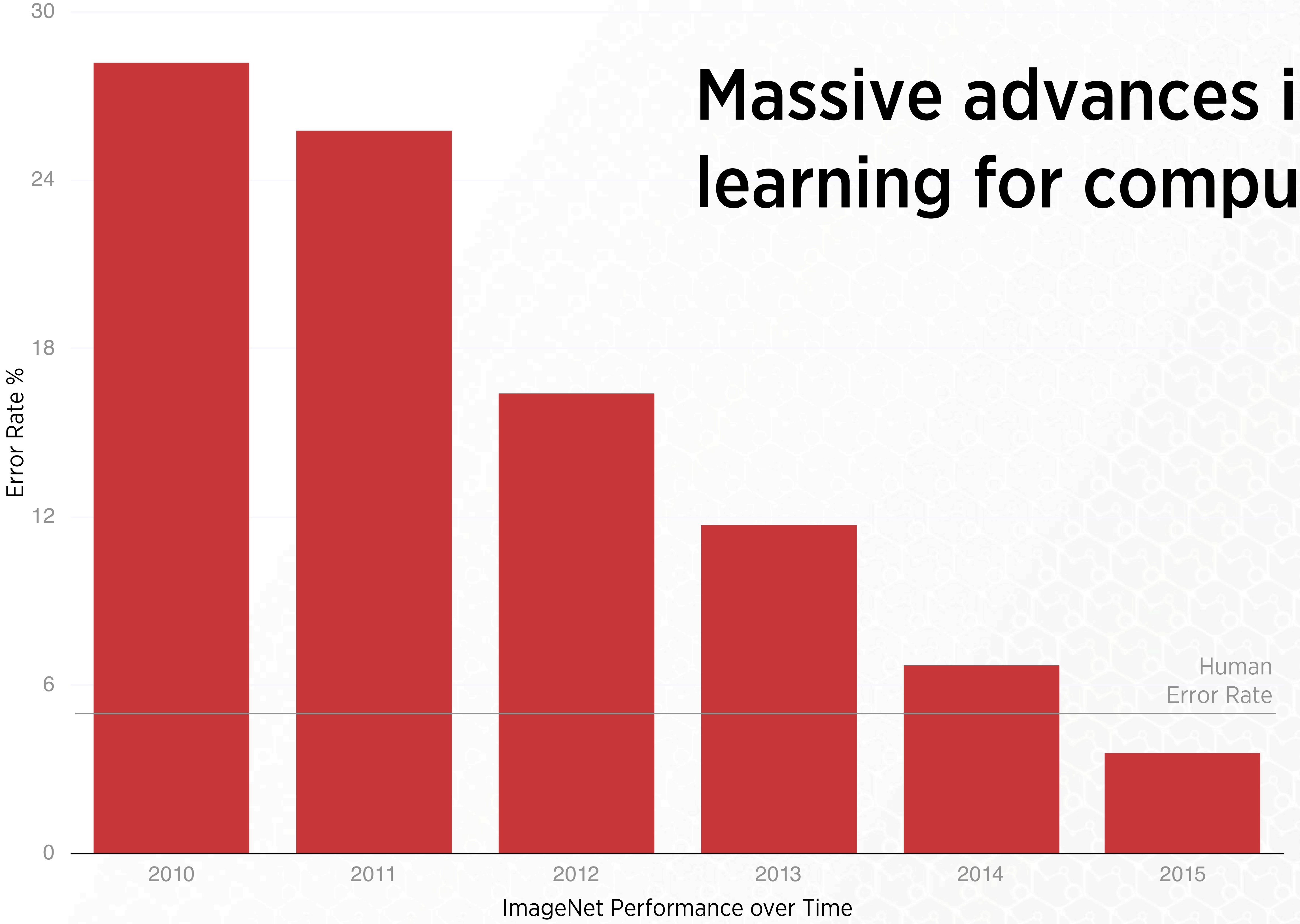
Discordance rates across a broad set of specimen types is fairly high with little improvement over past several decades

Table 3. Summary of Studies That Express a Discrepancy or Major Discrepancy Rate				
	Discrepancy Rates, %		Major Discrepancy Rates, %	
Study Type	No. of Studies	Median (25th–75th Percentile)	No. of Studies	Median (25th–75th Percentile)
All studies	116[c]	18.3 (7.5–34.5)	78[d]	5.9 (2.1–10.5)
Surgical pathology	84[e]	18.3 (7.5–37.4)	63[f]	6.3 (1.9–10.6)
Cytology	[illegible]	24.8 (17.4–38.8)	11[h]	4.3 (2.8–7.5)
Both	[illegible]	9.1 (6.7–15.8)	11[j]	5.9 (3.3–8.7)
Multiorgan	[illegible]	9.1 (3.8–18.7)	42[l]	3.9 (1.1–7.4)
Single-organ[a]	73[m]	25.2 (14.0–43.7)	36[n]	8.0 (3.7–15.8)
Internal[b]	35[o]	10.9 (3.8–17.6)	22[p]	1.2 (0.30–3.1)
External	79[q]	23.0 (10.6–40.2)	56[r]	7.4 (4.6–14.7)

High Error Rates

Arch Pathol Lab Med. 2016 Jan;140(1):29-40.

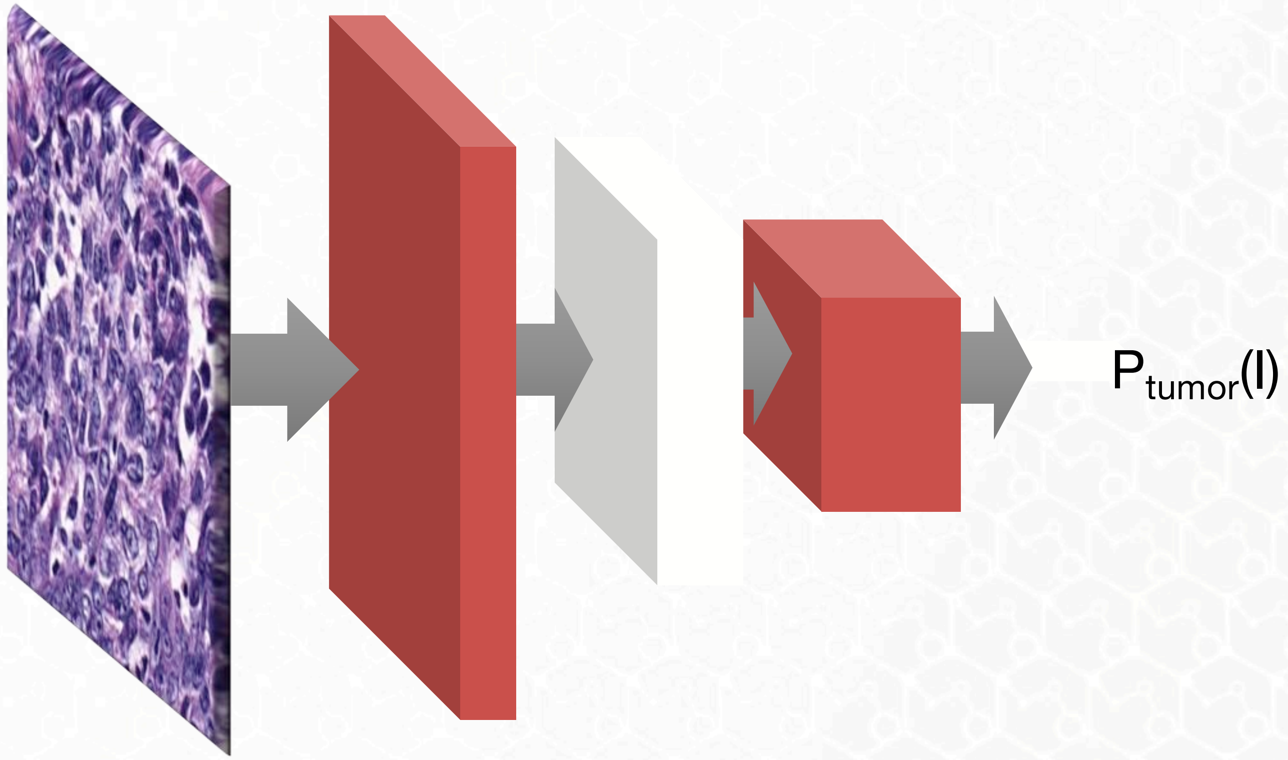
Massive advances in deep learning for computer vision…



PathAI
Proprietary & Confidential

What does AI mean at PathAI?

- Models which learn how to make decisions and predictions by recognizing patterns in data.
- These can be traditional machine learning models or, more commonly, deep convolutional neural networks.



The human defines the data, the data defines the algorithm.

Traditionally, the human defines the algorithm

What can AI do for pathology?

A (somewhat) *practical* treatment

- Exhaustive – the model is tireless and is not distracted
- Quantitative – the model is reproducible and objective
- Efficient – massive parallelization for speedy processing
- Exploratory - learn relationships in a purely data-driven manner

What AI *can't* do for pathology

Replace pathologists!

A diagnosis/detection example:

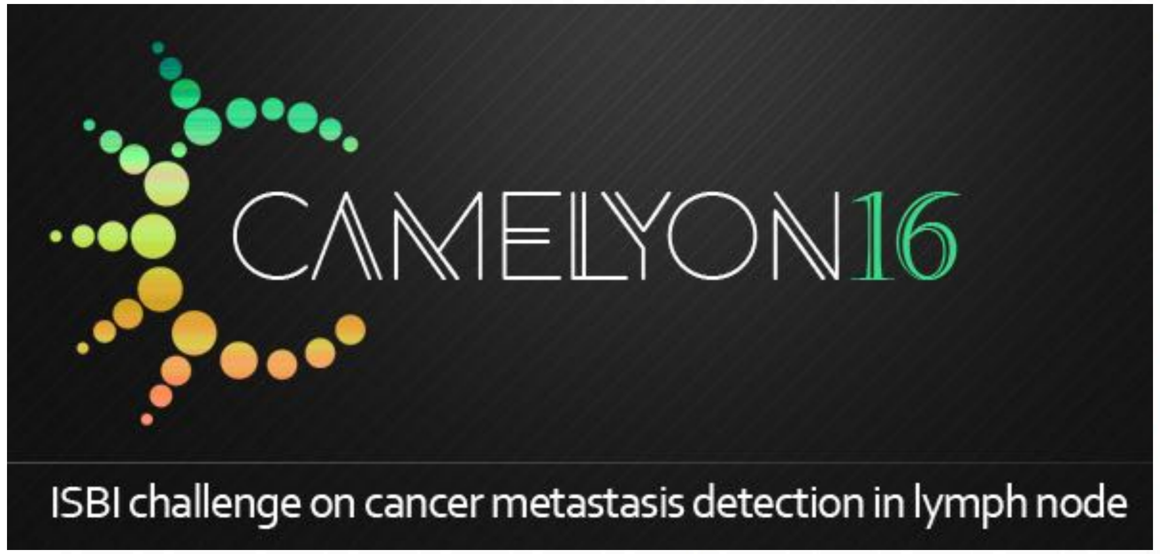
Breast cancer metastases

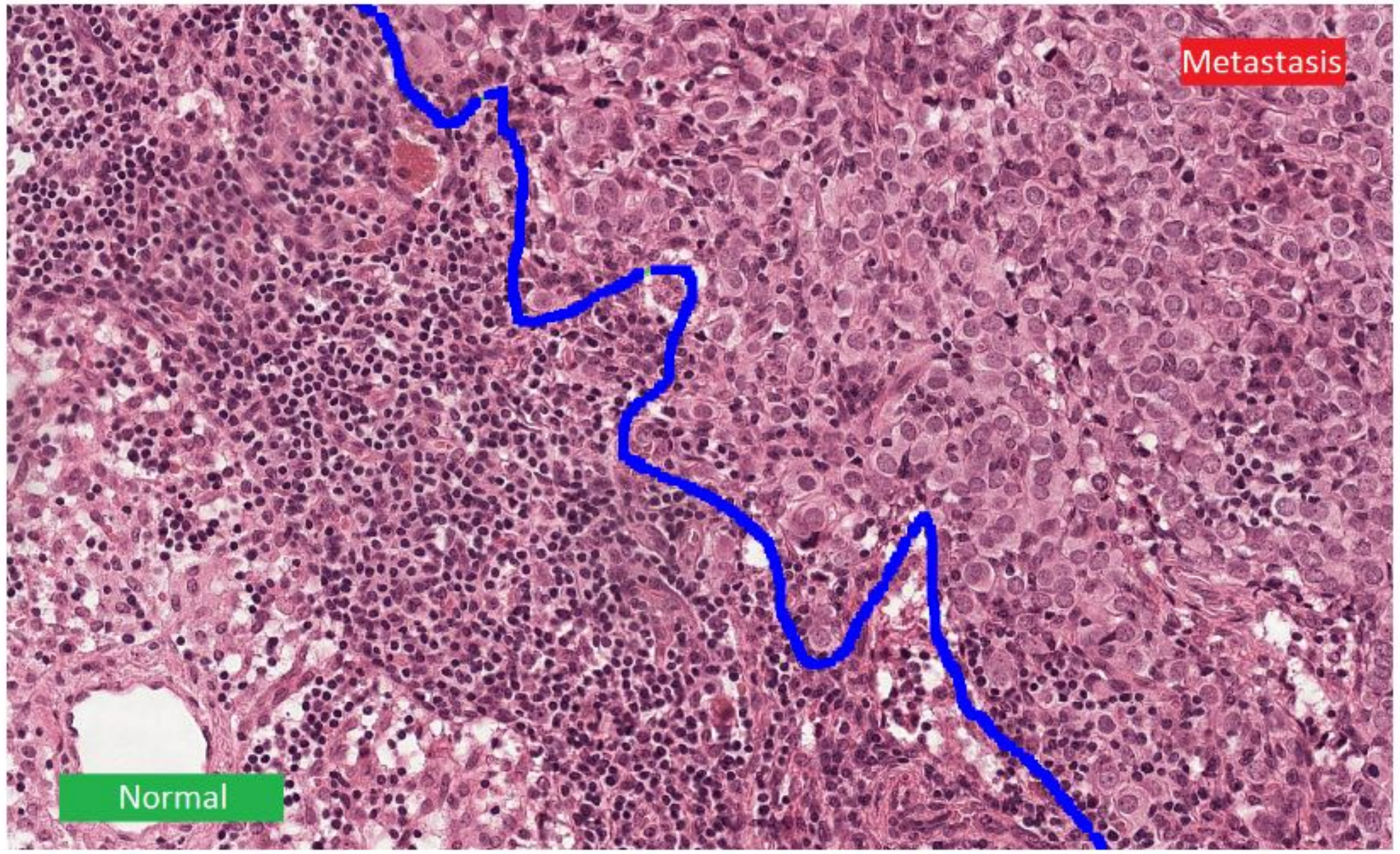
- After a primary mass discovered, lymph nodes are biopsied
- Pathologists check these for metastases
- Non-zero failure rate: a retrospective study found a 24% disagreement rate[1]

[1]Vestjens JHMJ, Pepels MJ, de Boer M, et al. Relevant impact of central pathology review on nodal classification in individual breast cancer patients. *Ann Oncol*. 2012;23(10):2561-2566.

The data - CAMELYON

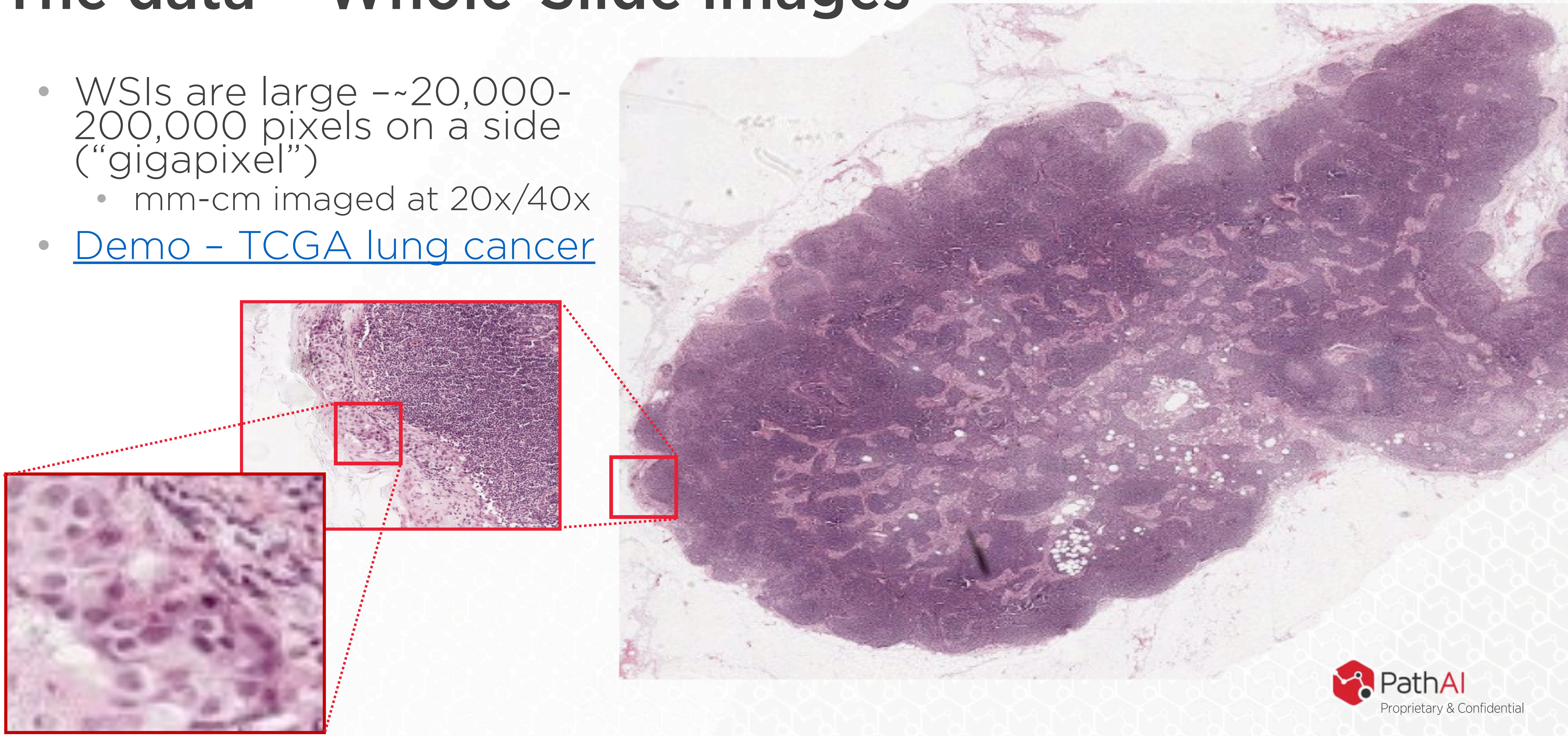
- H & E stained, Formalin-Fixe Paraffin-Embedded (FFPE)
  - 270 training slides, 129 test
- Annotated by a panel





The data – Whole-Slide Images

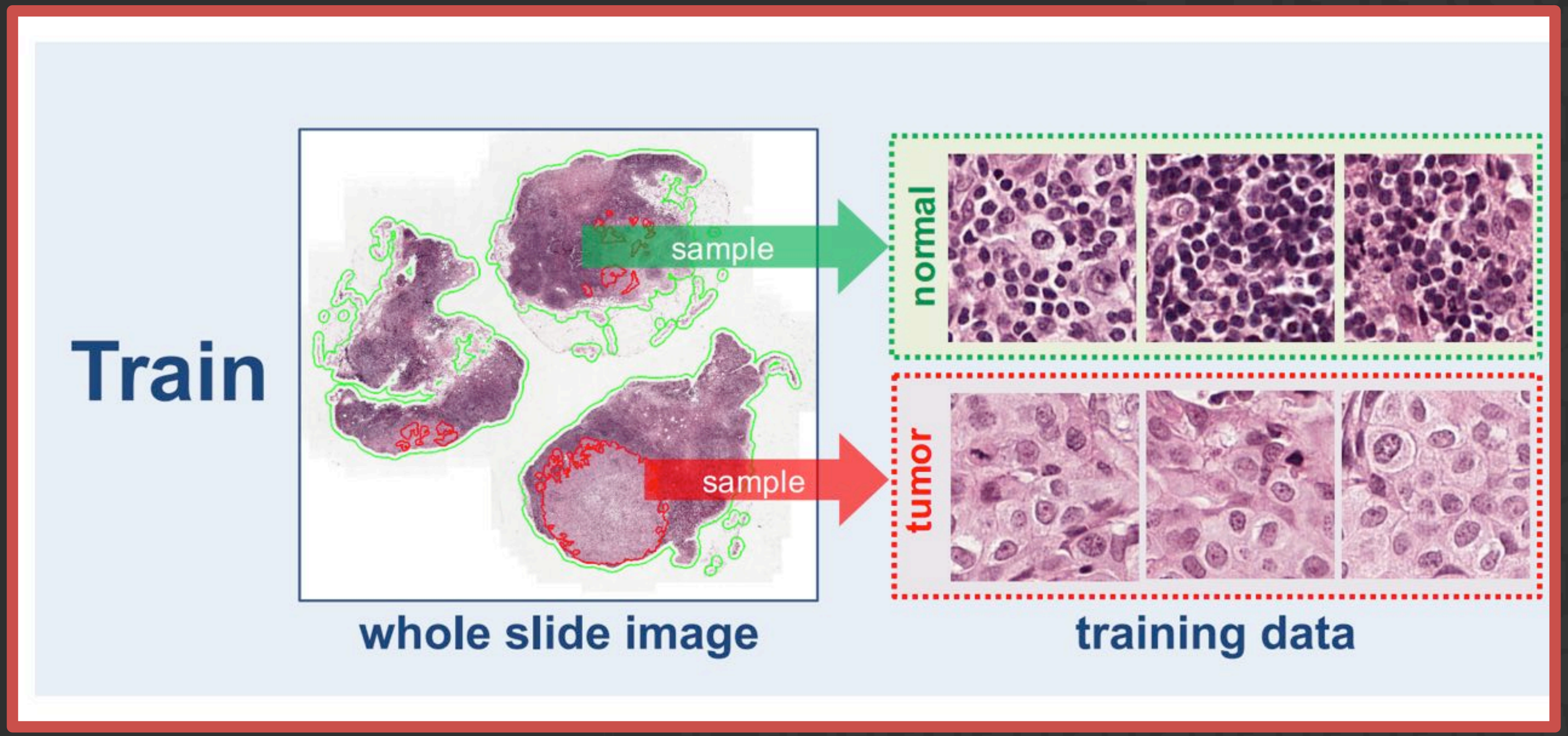
- WSIs are large –~20,000-200,000 pixels on a side ("gigapixel")
  - mm-cm imaged at 20x/40x
- [Demo – TCGA lung cancer]

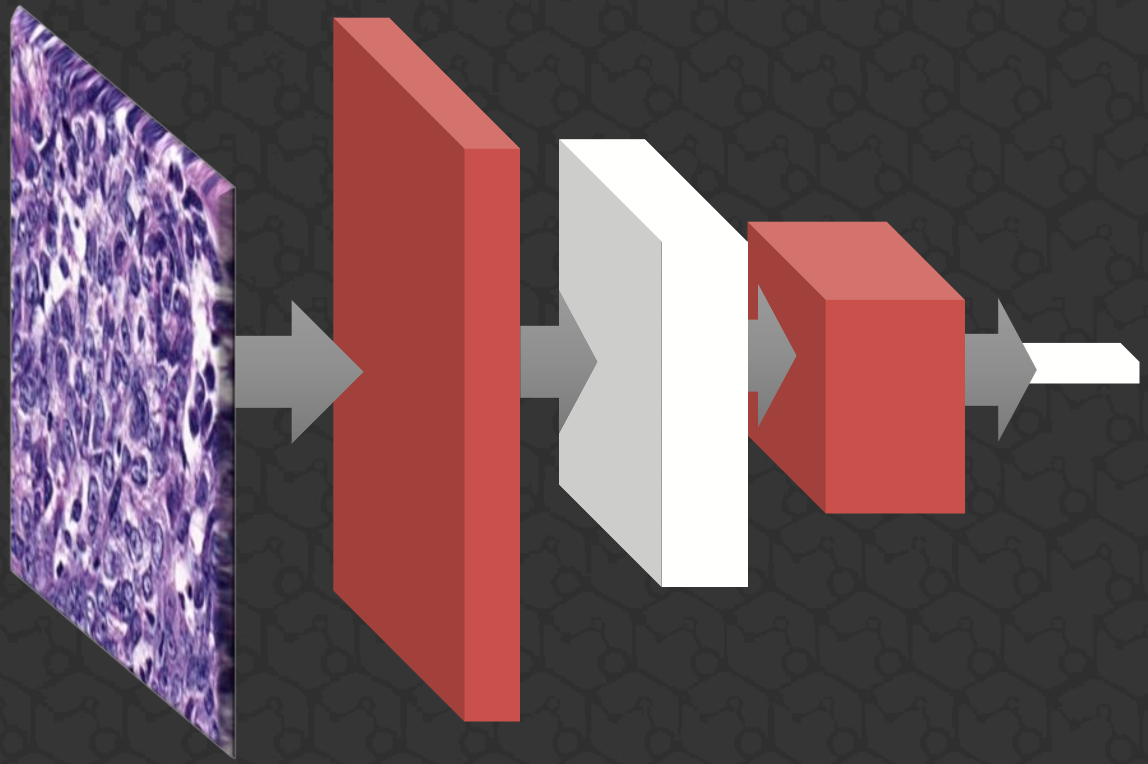


PathAI
Proprietary & Confidential

Approach

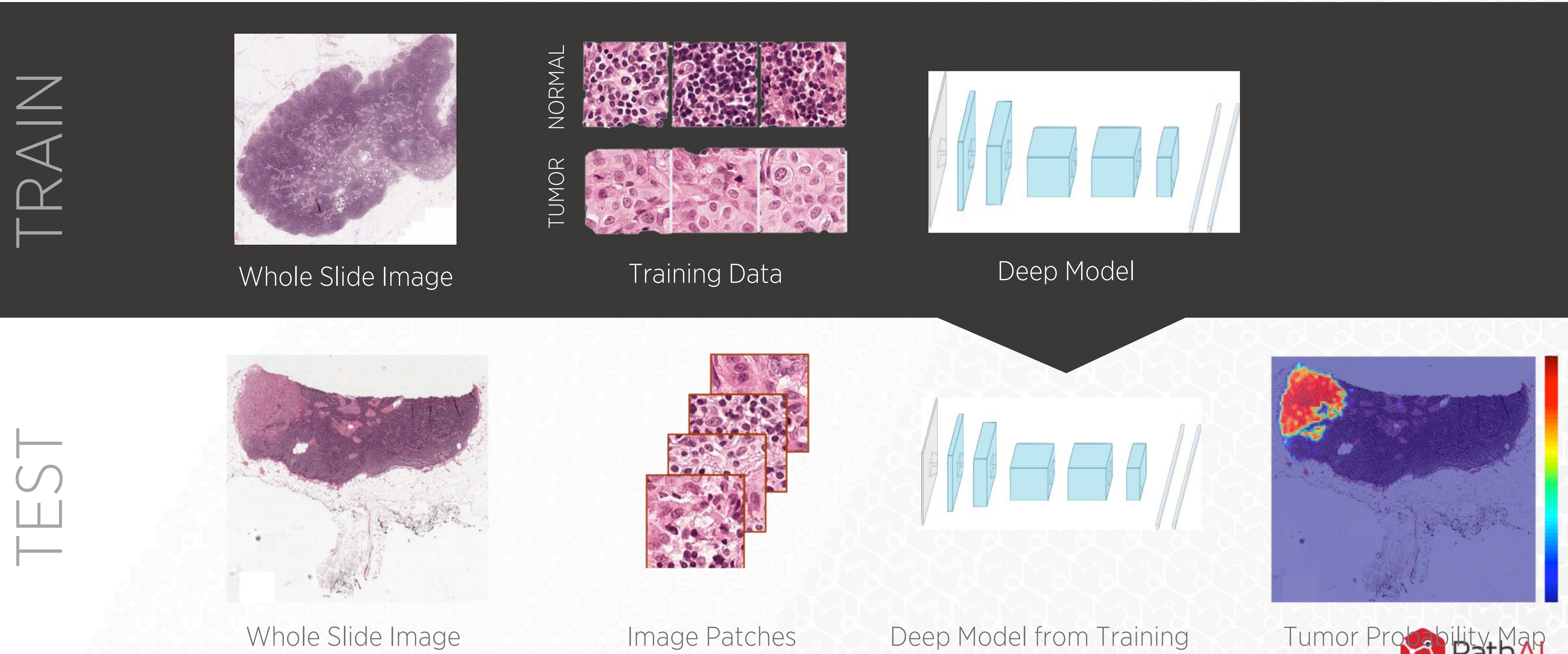
- Standard image classification approach needs a twist for WSIs: sampling





Successfully applied deep learning approach to pathology

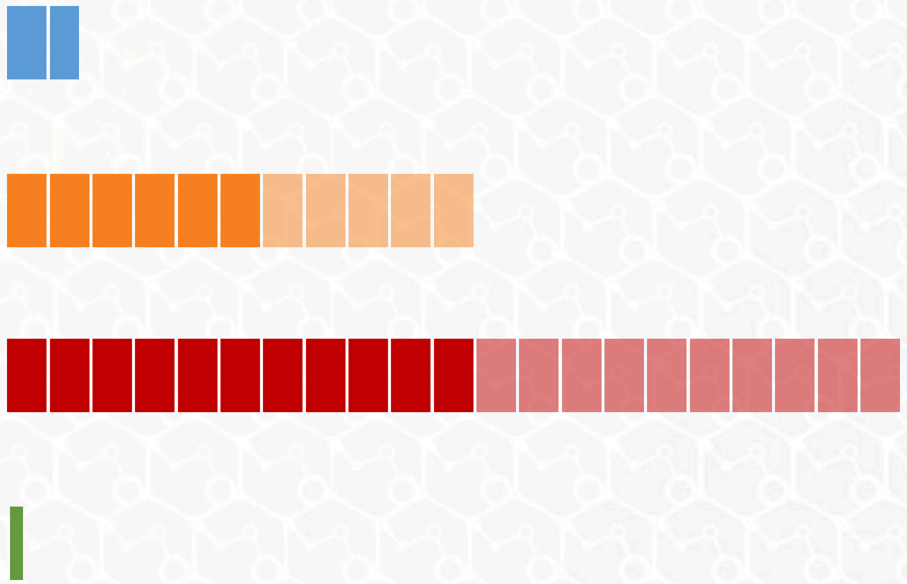
Our team won the Camelyon challenge in 2016, demonstrating outstanding initial performance in pathology



Wang, D., Khosla, A., ... Beck, A.H., 2016. Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718. JAMA. 2017 Dec 12;318(22):2199-2210.

Deep learning model outperforms human pathologists in the diagnosis of metastatic cancer

	Error Rate (1-AUC)
Pathologists in competition	3.5%
Pathologists in clinical practice[1]	13 – 26%
Pathologists on micro-metastasis[2]	23 – 42%
Deep learning model	0.65%



[1] n=12 [2] Small tumors

References: Wang, Khosla, … Beck (2016) https://arxiv.org/abs/1606.05718 Camelyon16 (JAMA, 2017)

Caveats and considerations

- Real world data vs. competition data

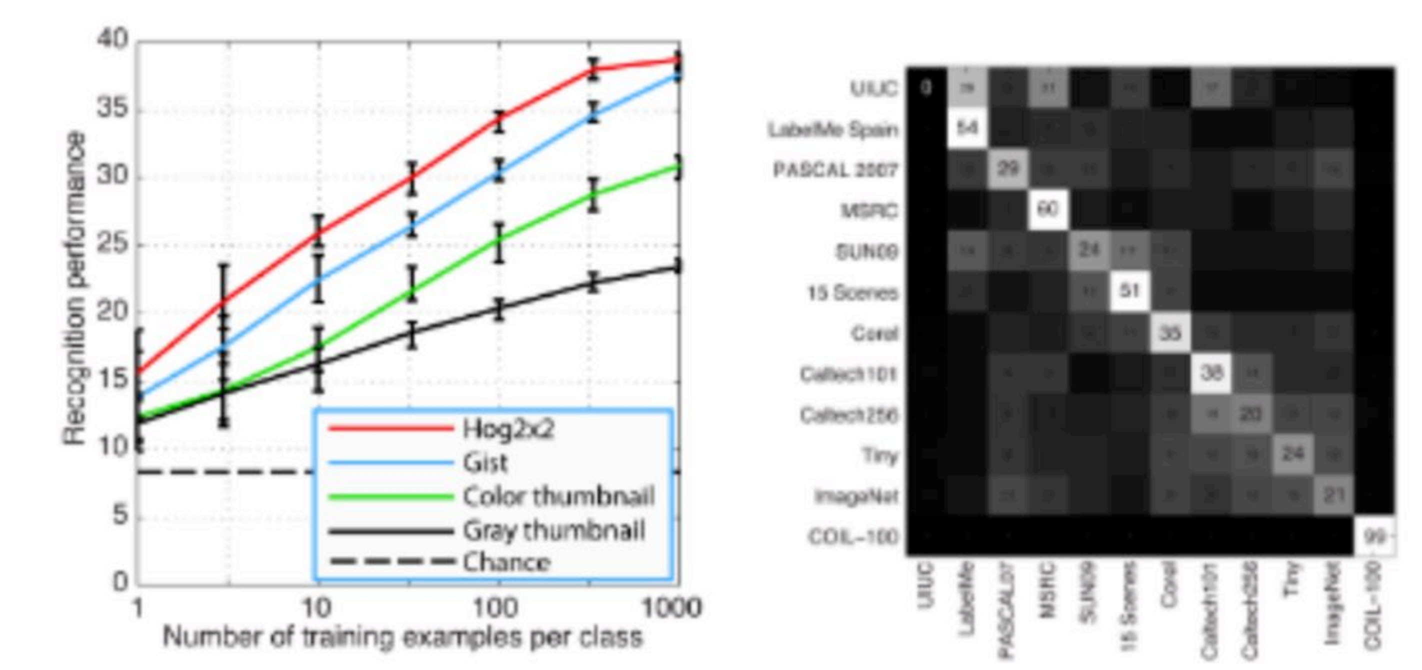
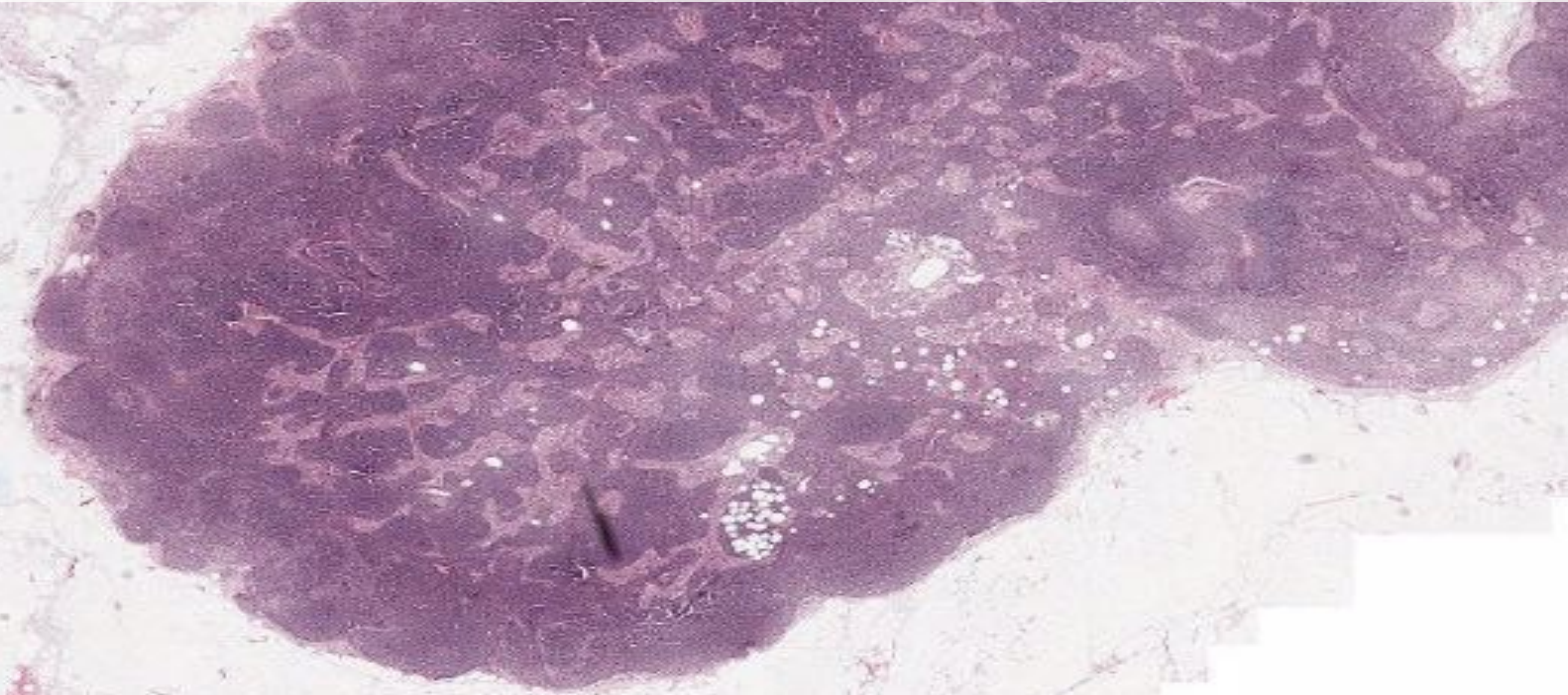


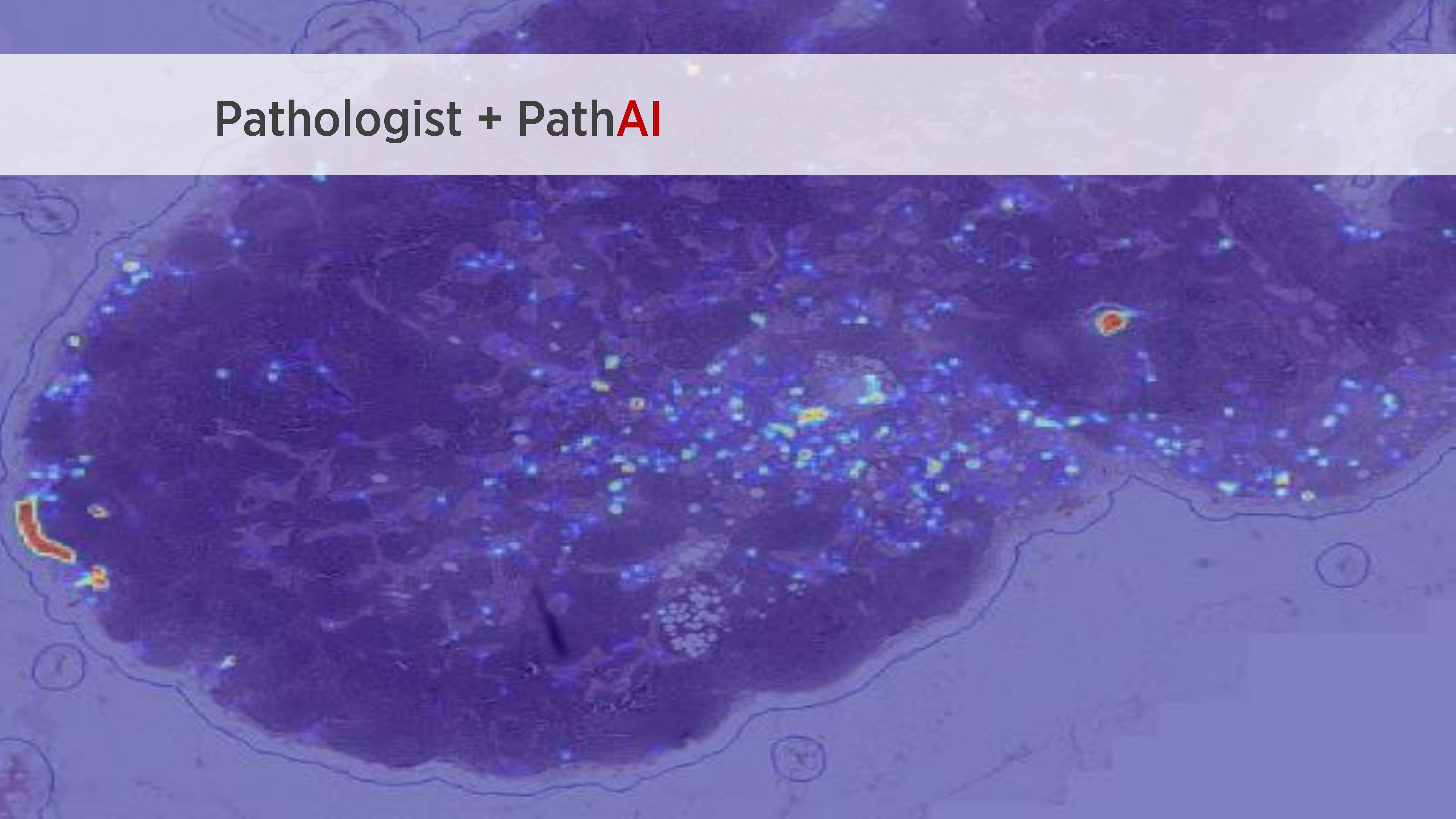
Figure 2. Computer plays *Name That Dataset*. Left: classification performance as a function of dataset size (log scale) for different descriptors (notice that performance does not appear to saturate). Right: confusion matrix.

Torralba & Effros, 2011

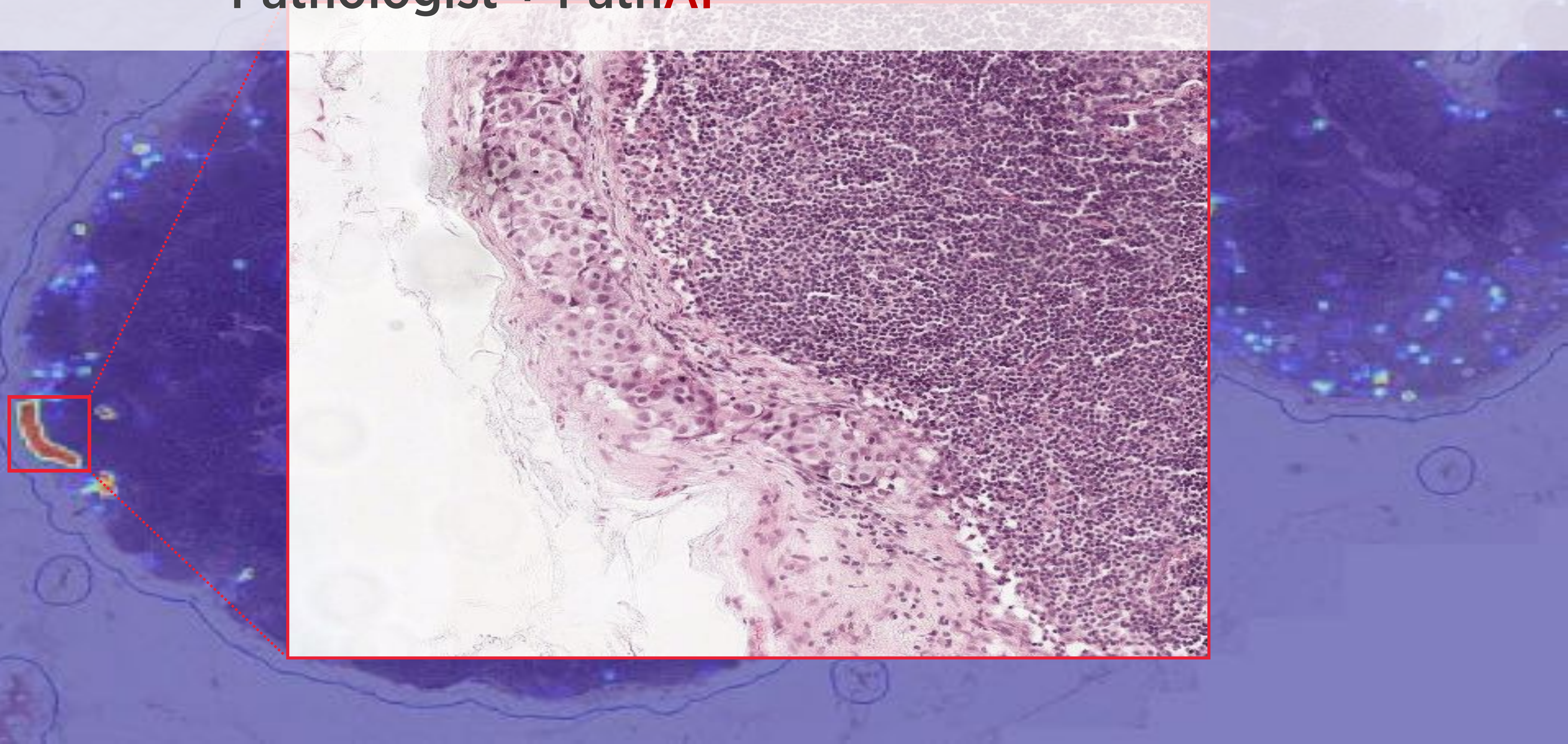
Pathologist + PathAI



Pathologist + PathAI

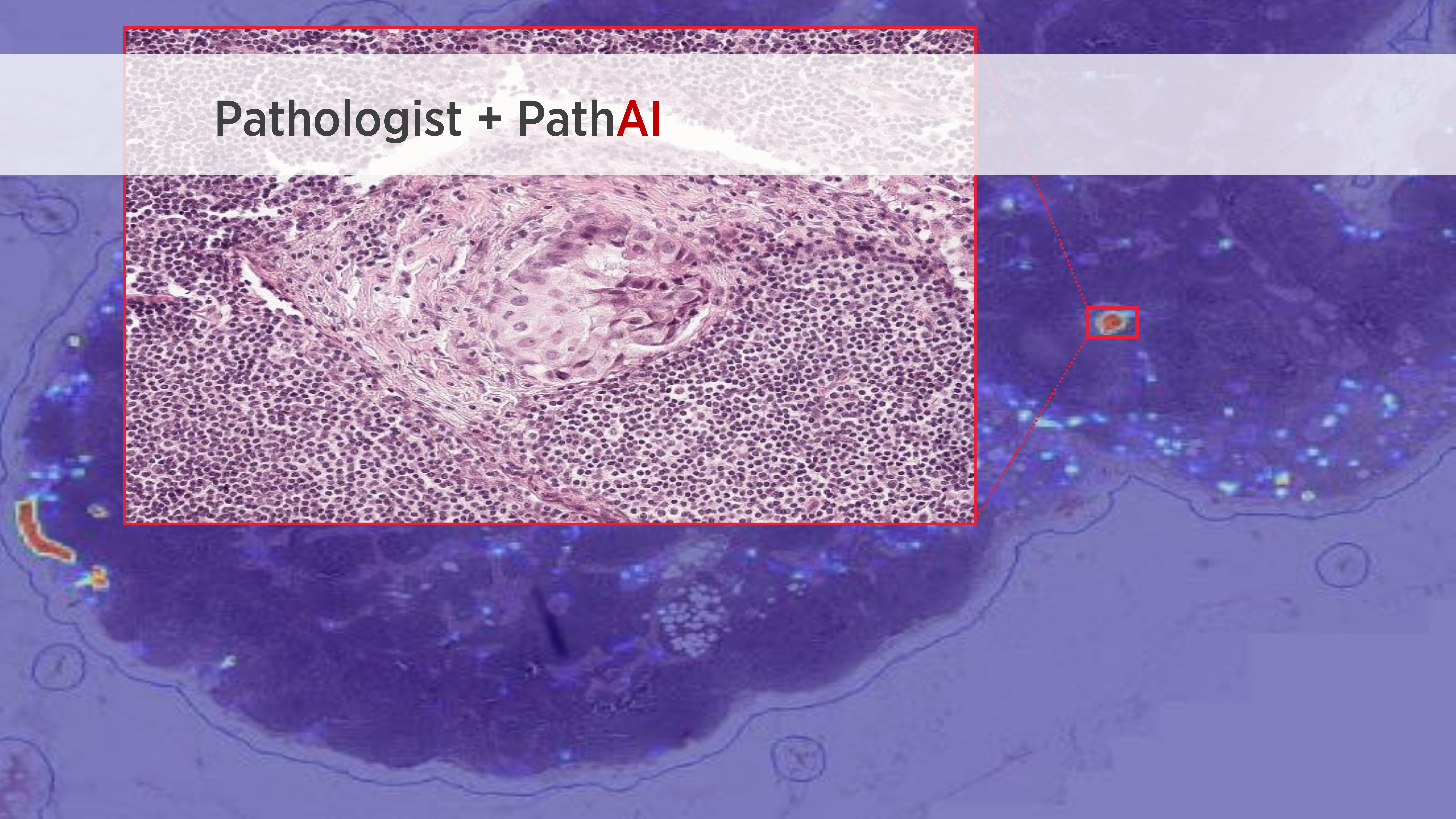


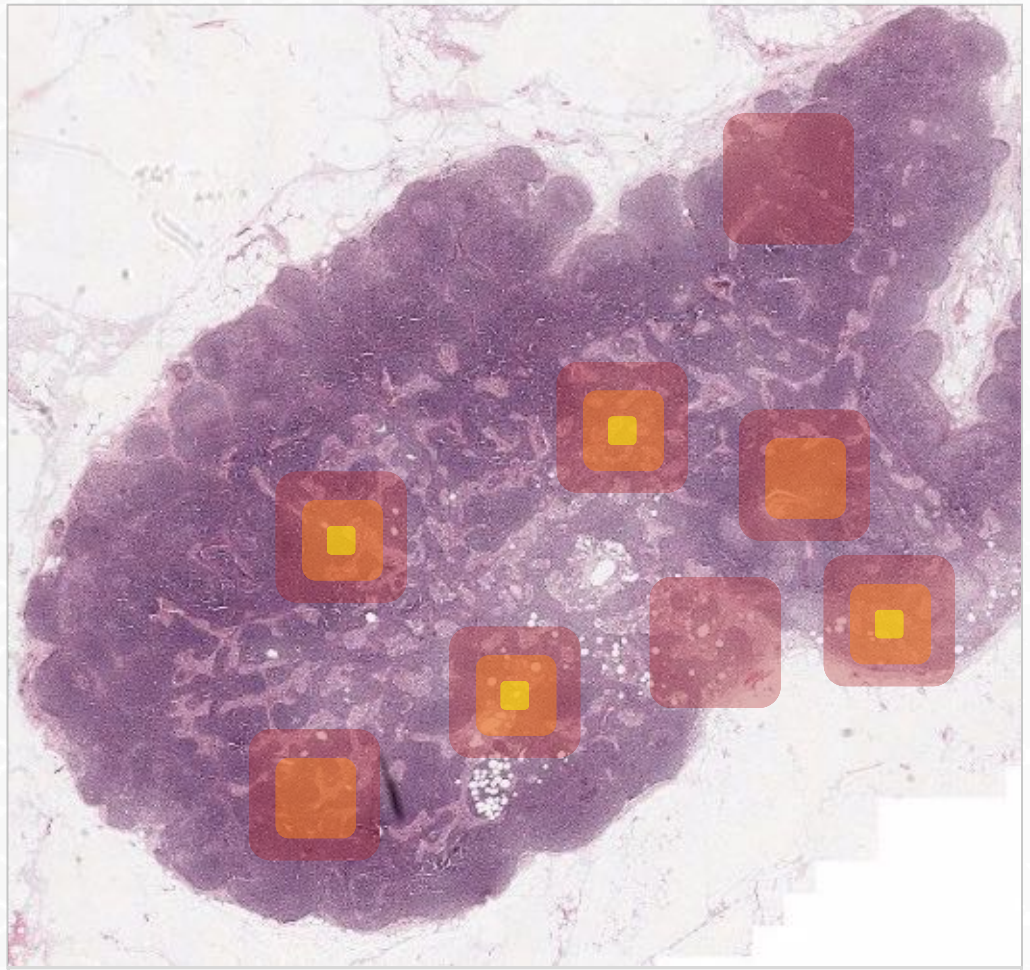
Pathologist + PathAI



Pathologist + PathAI

Pathologist + PathAI



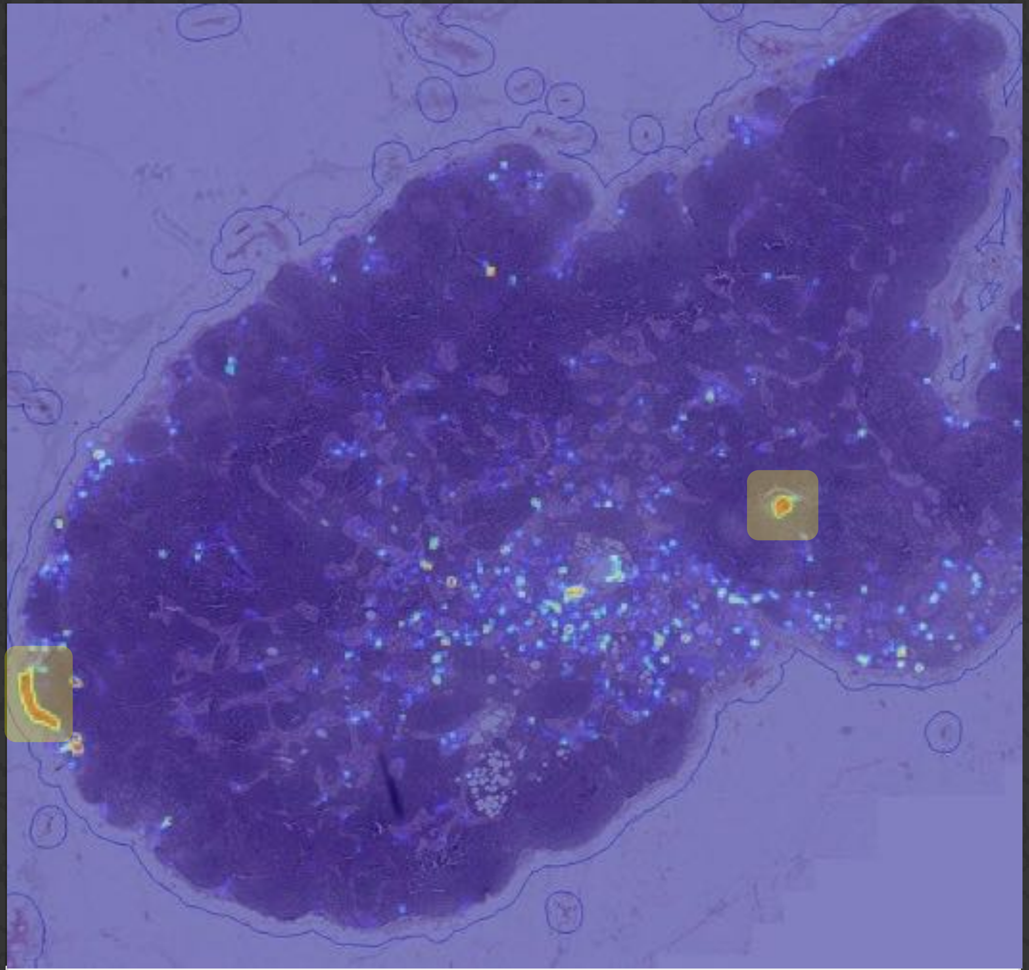


Pathology Report

Patient: **John Doe**
Diagnosis:
Size:

pTNM staging:
# of Pos LN:
# of Neg LN:

Time per slide: 1 – 10 minutes
Accuracy: ~85%
Reproducibility: Low



Pathology Report Confirm

Patient: **John Doe**
Diagnosis: Met. Cancer
Size: 2.3mm

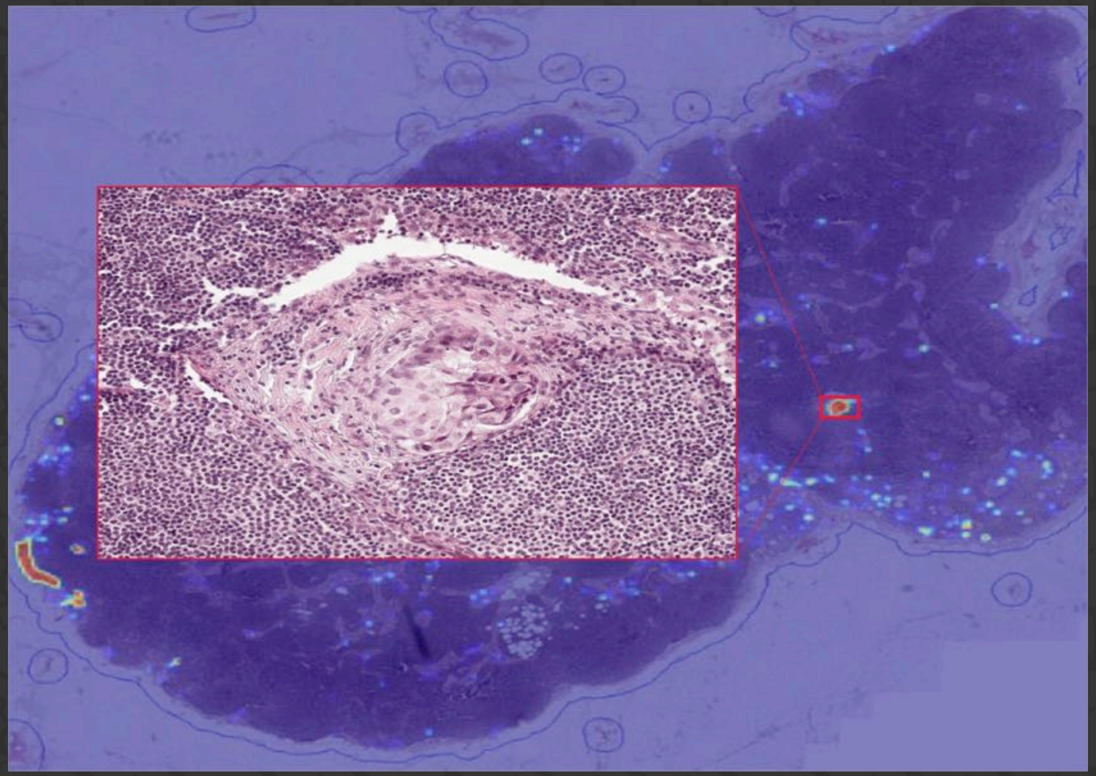
pTNM staging: pT2N1MX
# of Pos LN: 1
# of Neg LN: 4

Time per slide: 10– 60 seconds
Accuracy: >99.5%
Reproducibility: High

PathAI

Why is this a good application for AI?

- Exhaustive analysis is beneficial
  - Large volume
- Local image data necessary and sufficient
- Interpretability: Heatmaps & simple models provide insight into how the *patient-level* prediction was made
- *Required accuracy is high*



A predictive example:

Precision immunotherapy

- Some cancers express immune-inhibitory ligands, activating immune “checkpoints”
- “checkpoint inhibitors” mask these signals, unleashing the immune system

2018 Nobel Prize in Medicine Awarded to 2 Cancer Immunotherapy Researchers

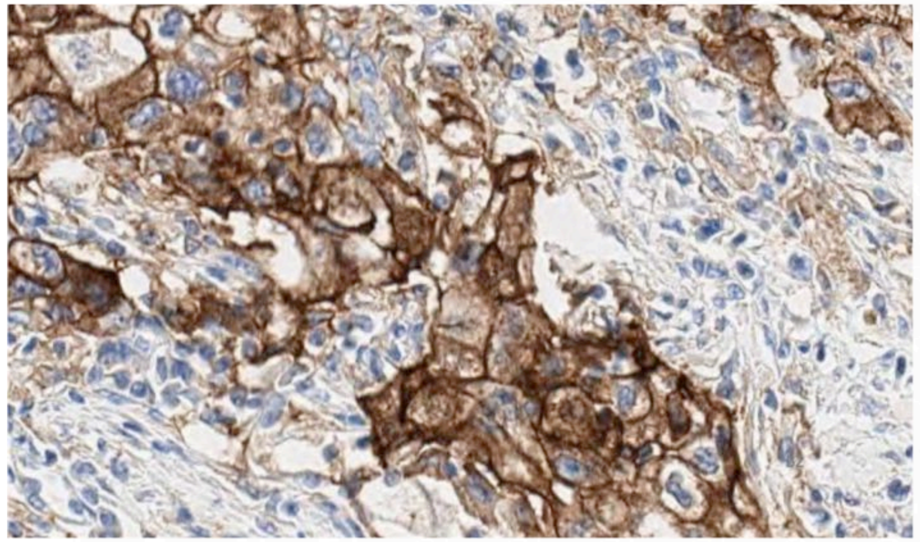
The Nobel Prize for Physiology and Medicine was awarded to James P. Allison, left, and Tasuku Honjo on Monday for their work on cancer research. Jonathan Nackstrand/Agence France-Presse — Getty Images

By The New York Times

Oct. 1, 2018

A predictive example:
Precision immunotherapy

- Response rate is low, but some fraction of patients are essentially "cured"
- PD-L1 expression is somewhat indicative of response



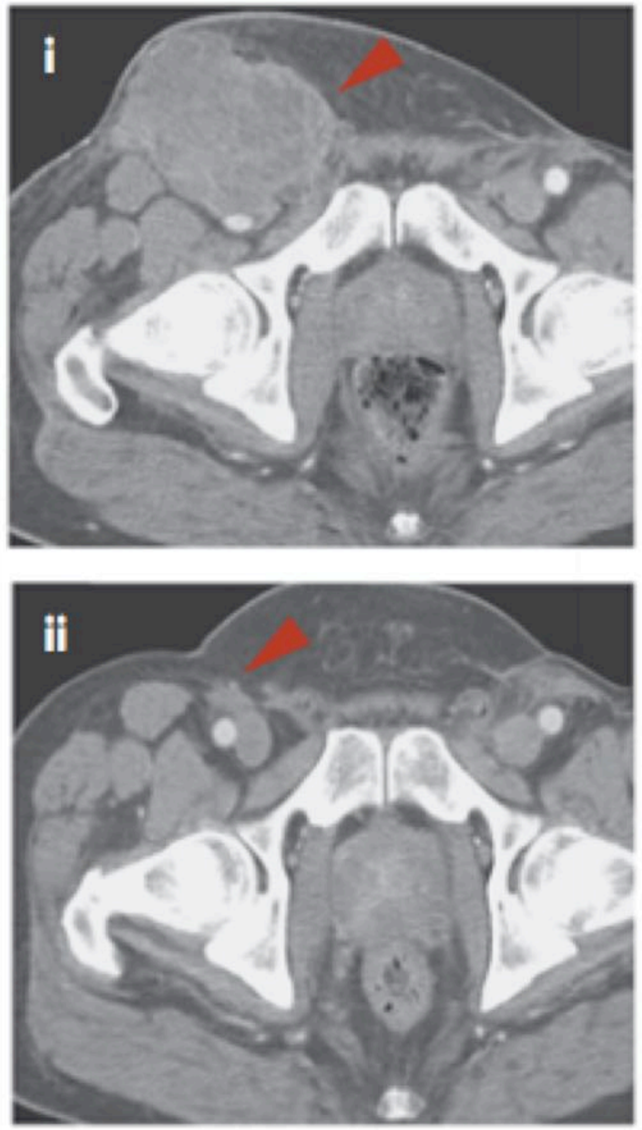
The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 28, 2012 VOL. 366 NO. 26

Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer

Suzanne L. Topalian, M.D., F. Stephen Hodi, M.D., Julie R. Brahmer, M.D., Scott N. Gettinger, M.D., David C. Smith, M.D., David F. McDermott, M.D., John D. Powderly, M.D., Richard D. Carvajal, M.D., Jeffrey A. Sosman, M.D., Michael B. Atkins, M.D., Philip D. Leming, M.D., David R. Spigel, M.D., Scott J. Antonia, M.D., Ph.D., Leora Horn, M.D., Charles G. Drake, M.D., Ph.D., Drew M. Pardoll, M.D., Ph.D., Lieping Chen, M.D., Ph.D., William H. Sharfman, M.D., Robert A. Anders, M.D., Ph.D., Janis M. Taube, M.D., Tracee L. McMiller, M.S., Haiying Xu, B.A., Alan J. Korman, Ph.D., Maria Jure-Kunkel, Ph.D., Shruti Agrawal, Ph.D., Daniel McDonald, M.B.A., Georgia D. Kollia, Ph.D., Ashok Gupta, M.D., Ph.D., Jon M. Wigginton, M.D., and Mario Sznol, M.D.

Patient with Melanoma

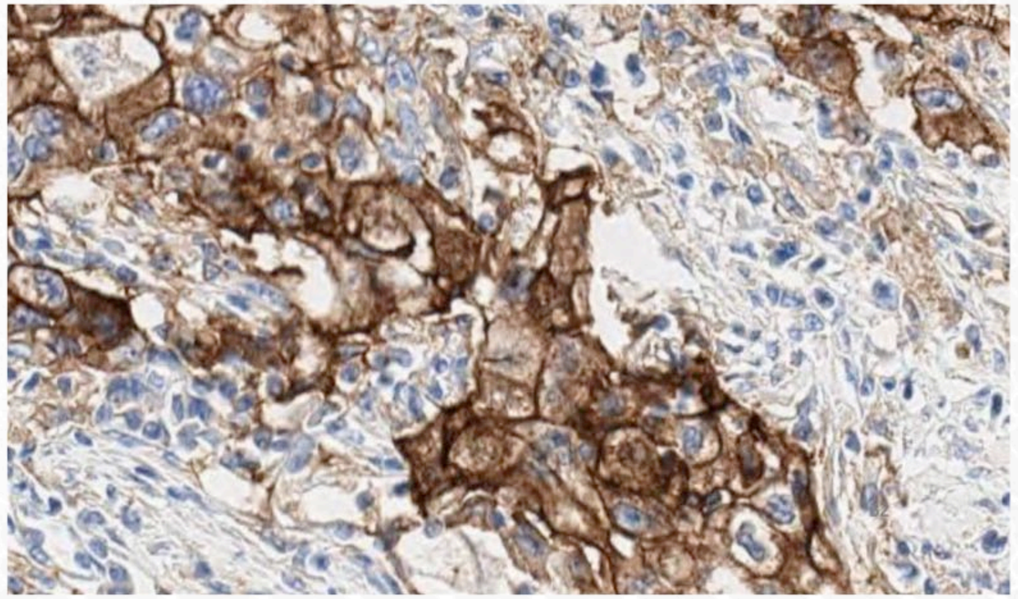


CONCLUSIONS

Anti–PD-1 antibody produced objective responses in approximately one in four to one in five patients with non–small-cell lung cancer, melanoma, or renal-cell cancer; the adverse-event profile does not appear to preclude its use. Preliminary data suggest a relationship between PD-L1 expression on tumor cells and objective response. (Funded by Bristol-Myers Squibb and others; ClinicalTrials.gov number, NCT00730639.)

N ENGL J MED 366;26 NEJM.ORG JUNE 28, 2012

Manual interpretation of PD-L1 IHC is highly variable



PDL1 manual IHC scores on immune cells are unreliable

Table 2. ICC for the Pathologist Scores and Concordance Statistics

	Antibody, ICC (95% CI)				
Cells[a]	22c3	28-8	SP142	E1L3N	Summary, Mean (SD)
Tumor cells	0.882 (0.873-0.891)	0.832 (0.820-0.844)	0.869 (0.859-0.879)	0.859 (0.849-0.869)	0.86 (0.02)
Immune cells	0.207 (0.190-0.226)	0.172 (0.156-0.189)	0.185 (0.169-0.203)	0.229 (0.211-0.248)	0.19 (0.03)

Abbreviation: ICC, intraclass correlation coefficient.

[a] N = 90.

Rimm et al. (JAMA Oncol; 2017)

Manual scoring of PD-L1 is variable

...and not always predictive

ORIGINAL ARTICLE

Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer

Julie Brahmer, M.D., Karen L. Reckamp, M.D., Paul Baas, M.D., Lucio Crinò, M.D., Wilfried E.E. Eberhardt, M.D., Elena Poddubskaya, M.D., Scott Antonia, M.D., Ph.D., Adam Pluzanski, M.D., Ph.D., Everett E. Vokes, M.D., Esther Holgado, M.D., Ph.D., David Waterhouse, M.D., Neal Ready, M.D., Justin Gainor, M.D., Osvaldo Arén Frontera, M.D., Libor Havel, M.D., Martin Steins, M.D., Marina C. Garassino, M.D., Joachim G. Aerts, M.D., Manuel Domine, M.D., Luis Paz-Ares, M.D., Martin Reck, M.D., Christine Baudelet, Ph.D., Christopher T. Harbison, Ph.D., Brian Lestini, M.D., Ph.D., and David R. Spigel, M.D.

RESULTS

The median overall survival was 9.2 months (95% confidence interval [CI], 7.3 to 13.3) with nivolumab versus 6.0 months (95% CI, 5.1 to 7.3) with docetaxel. The risk of death was 41% lower with nivolumab than with docetaxel (hazard ratio, 0.59; 95% CI, 0.44 to 0.79; P<0.001). At 1 year, the overall survival rate was 42% (95% CI, 34 to 50) with nivolumab versus 24% (95% CI, 17 to 31) with docetaxel. The response rate was 20% with nivolumab versus 9% with docetaxel (P=0.008).

0.47 to 0.81; P<0.001). The expression of the PD-1 ligand (PD-L1) was neither prognostic nor predictive of benefit. Treatment-related adverse events of grade 3

prognostic nor predictive of benefit. Treatment-related adverse events of grade 3
or 4 were reported in 7% of the patients in the nivolumab group as compared with 55% of those in the docetaxel group.

PathAI