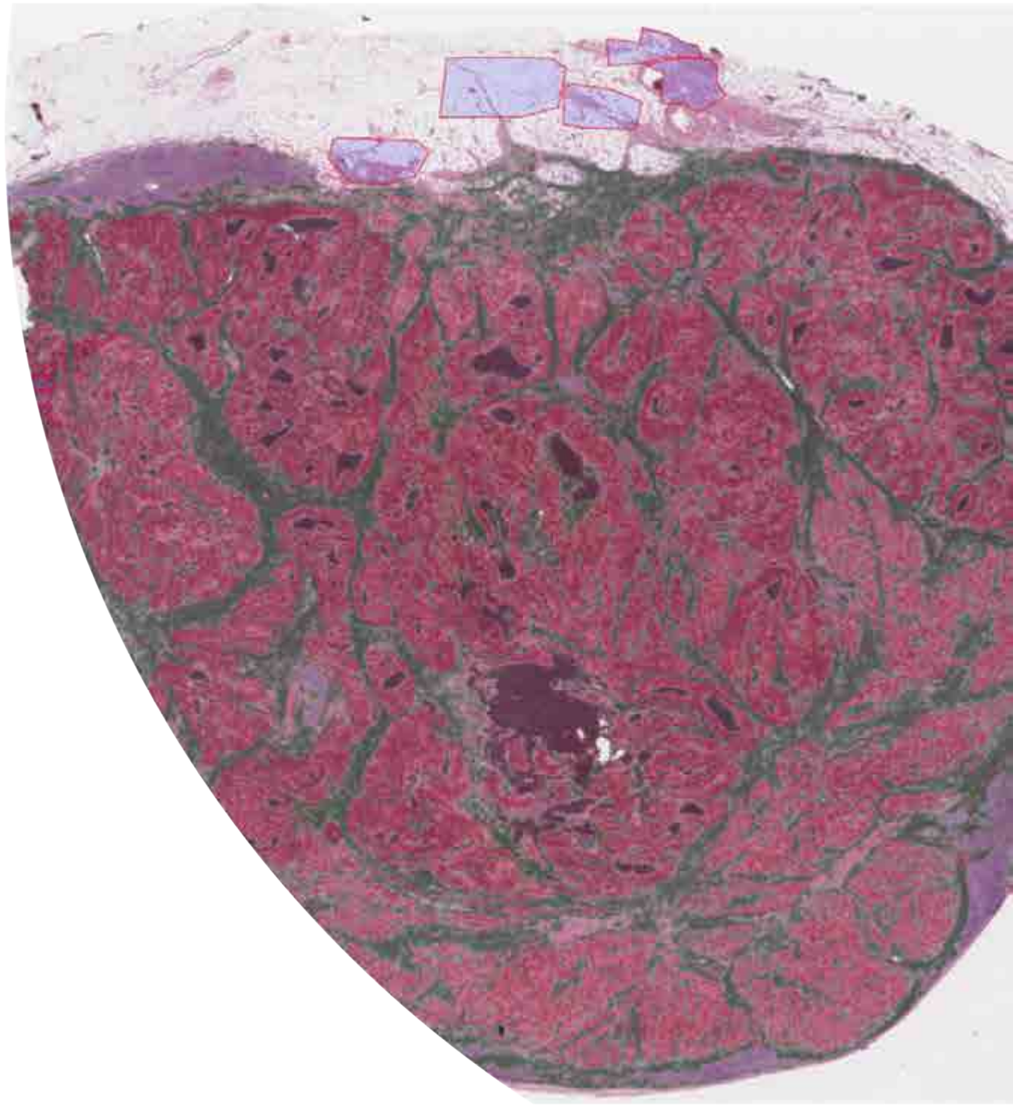
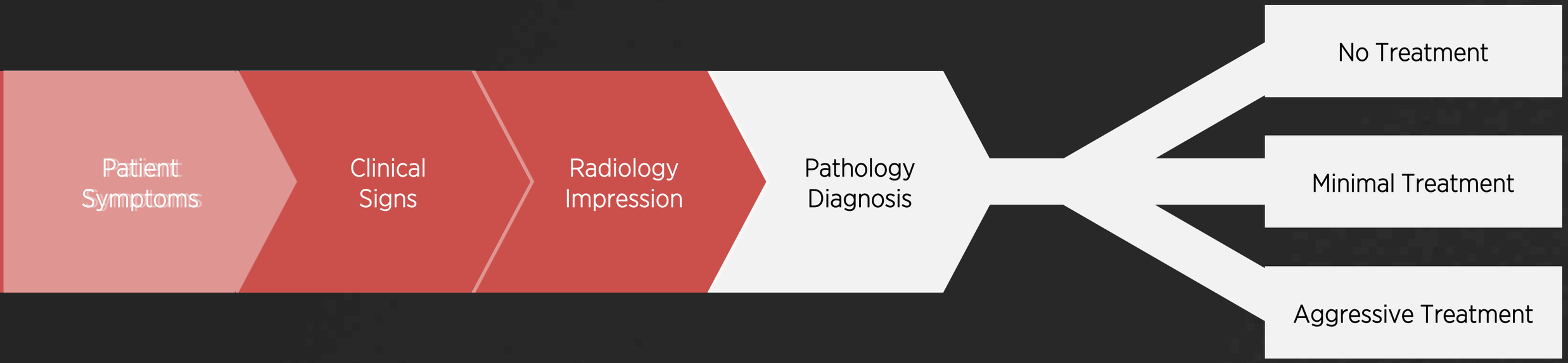
Machine learning for Pathology

Andrew H Beck MD PhD
CEO @ PathAI

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



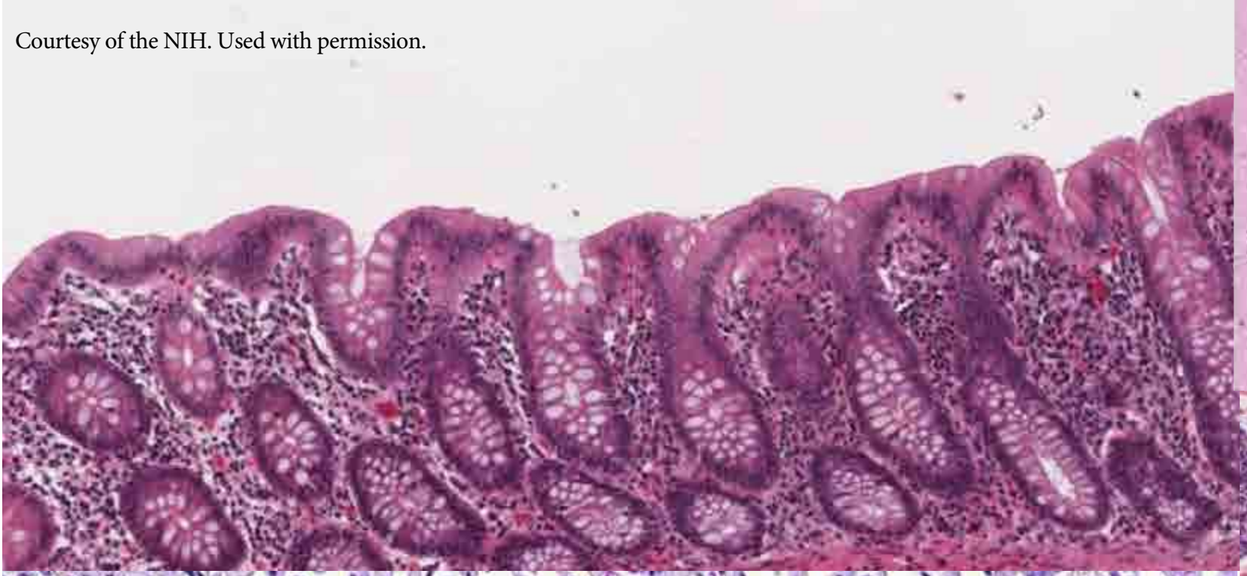
Pathology

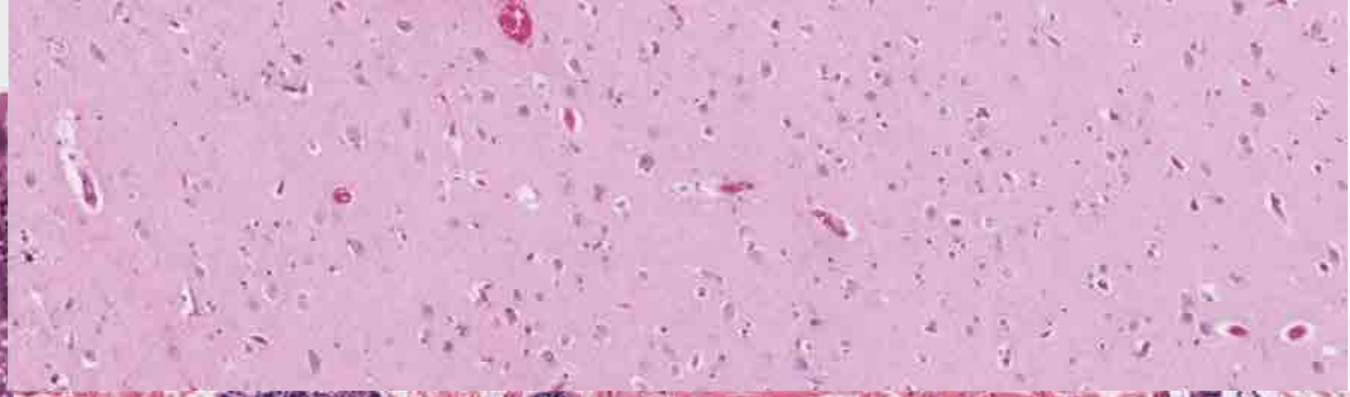


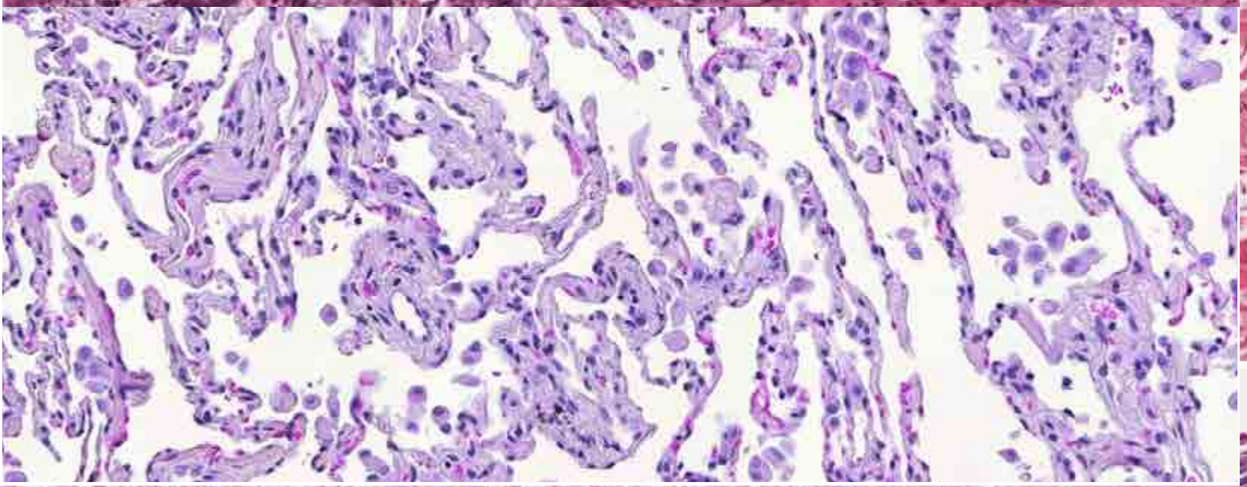
Pathologic diagnosis is a central determinant of therapeutic decisions.

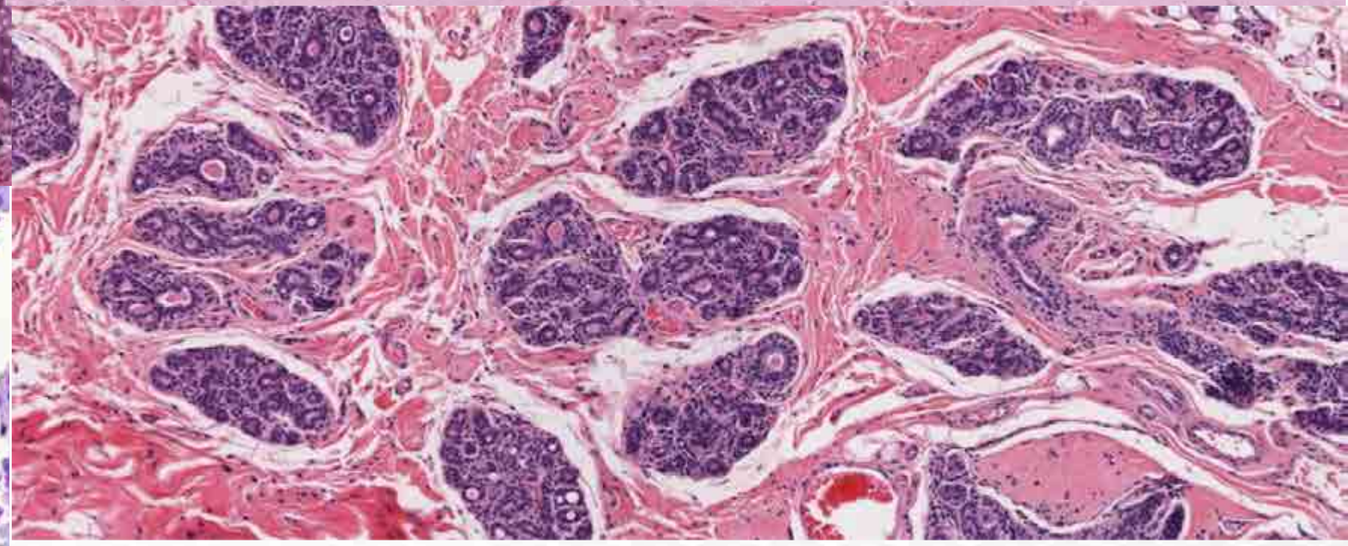
PathAI

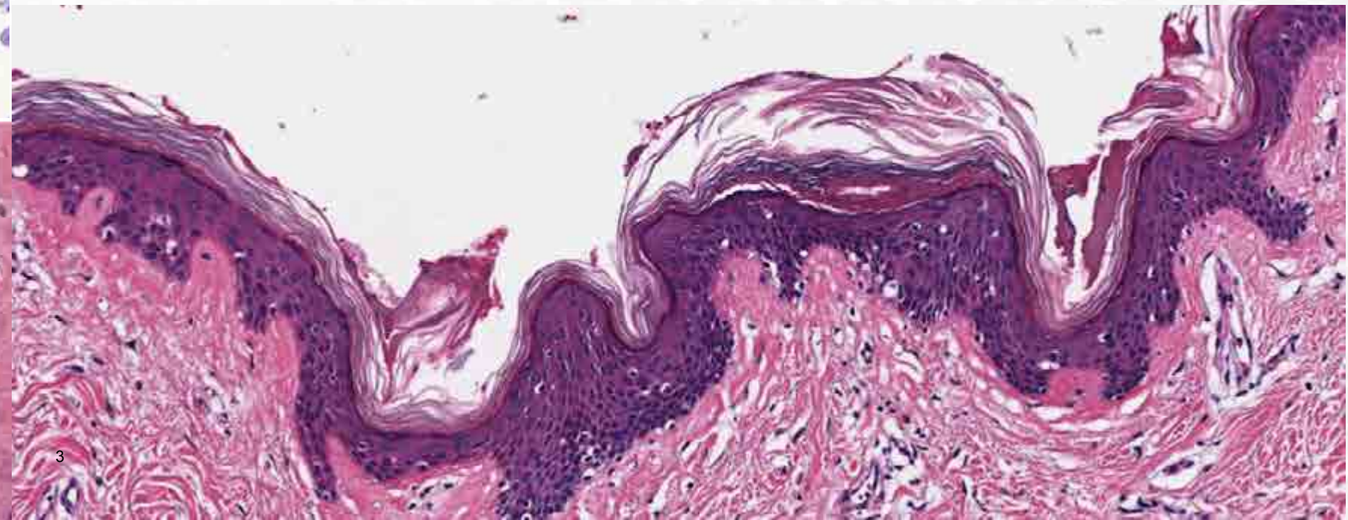
Courtesy of the NIH. Used with permission.





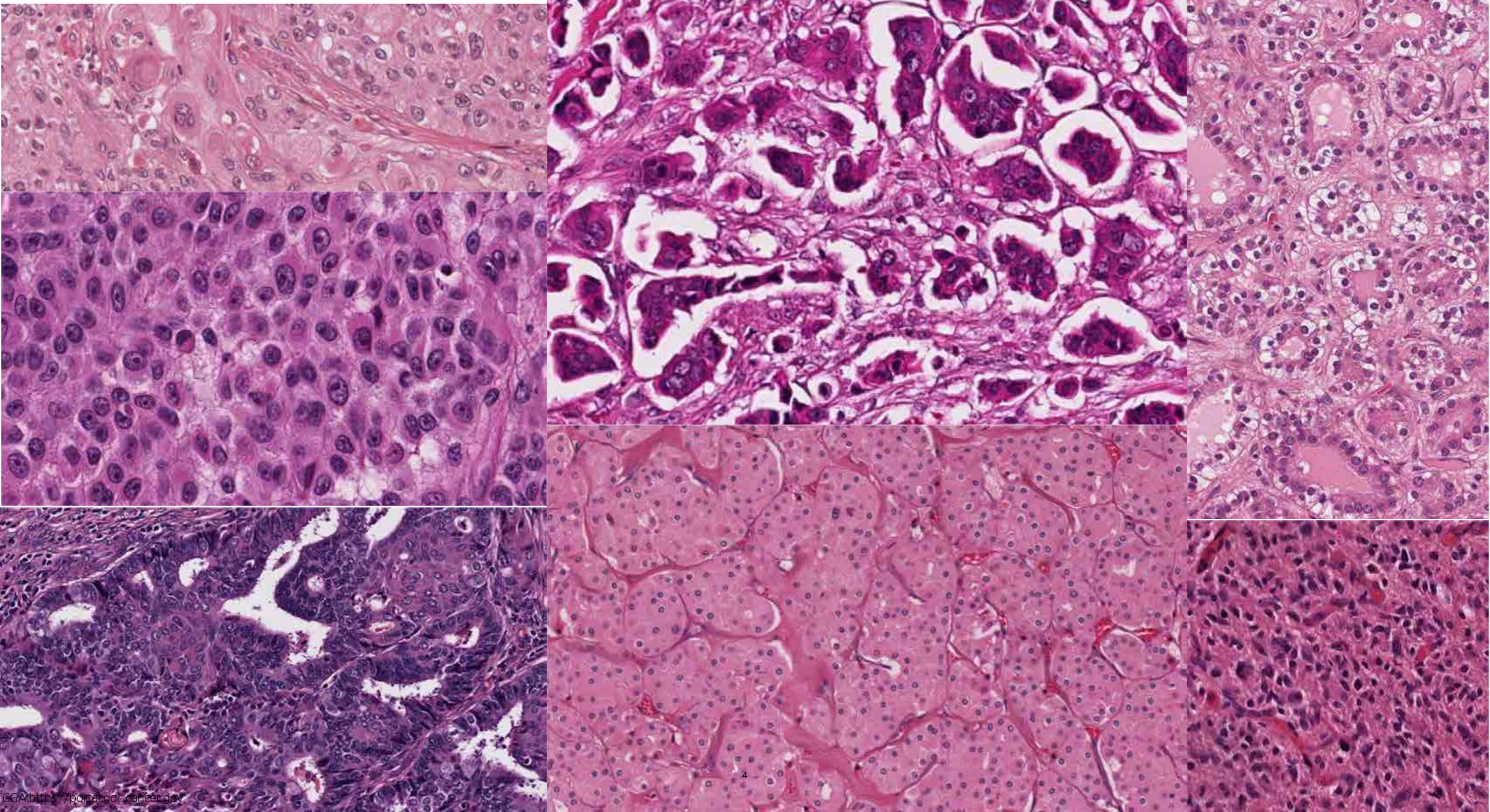






CGA:https://portal.gdc.cancer.gov

Courtesy of the NIH. Used with permission.



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

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 artificial intelligence techniques and traditional algorithmic processing, may provide useful software engineering tools for the development of systems in quantitative pathology.”

Emergence of Digital Pathology (2000)

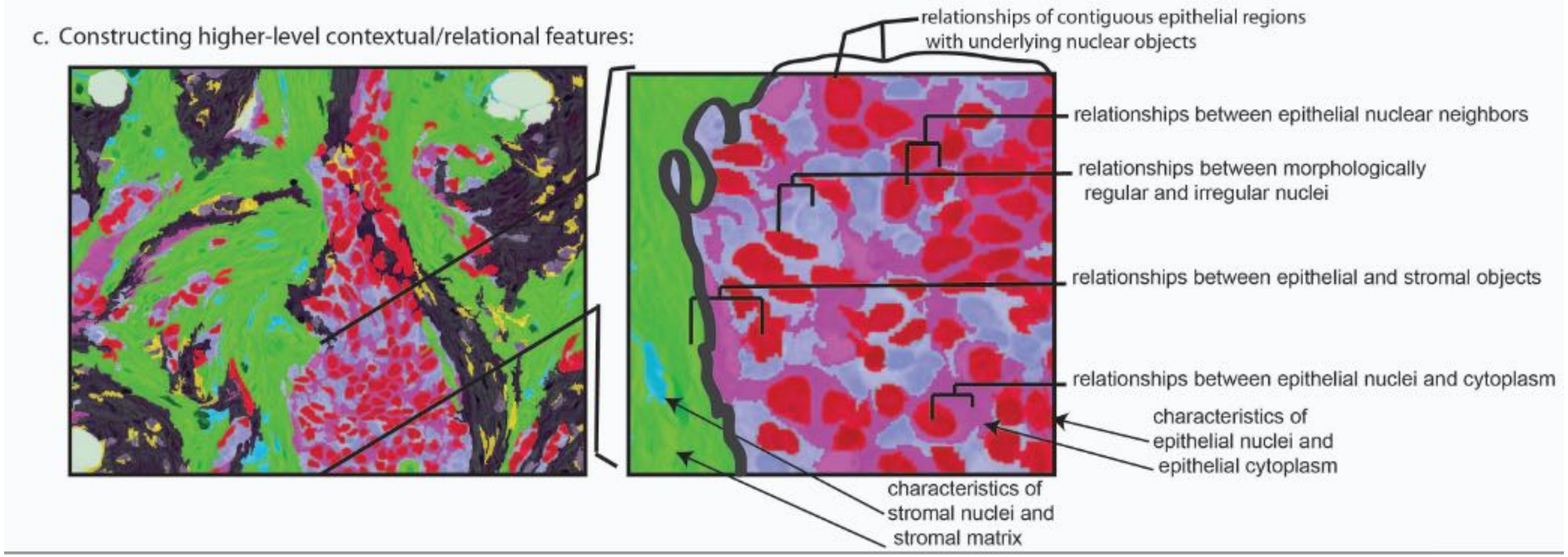
International Journal of Surgical Pathology 8(4):261–263, 2000

Digital Pathology: Science Fiction?

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

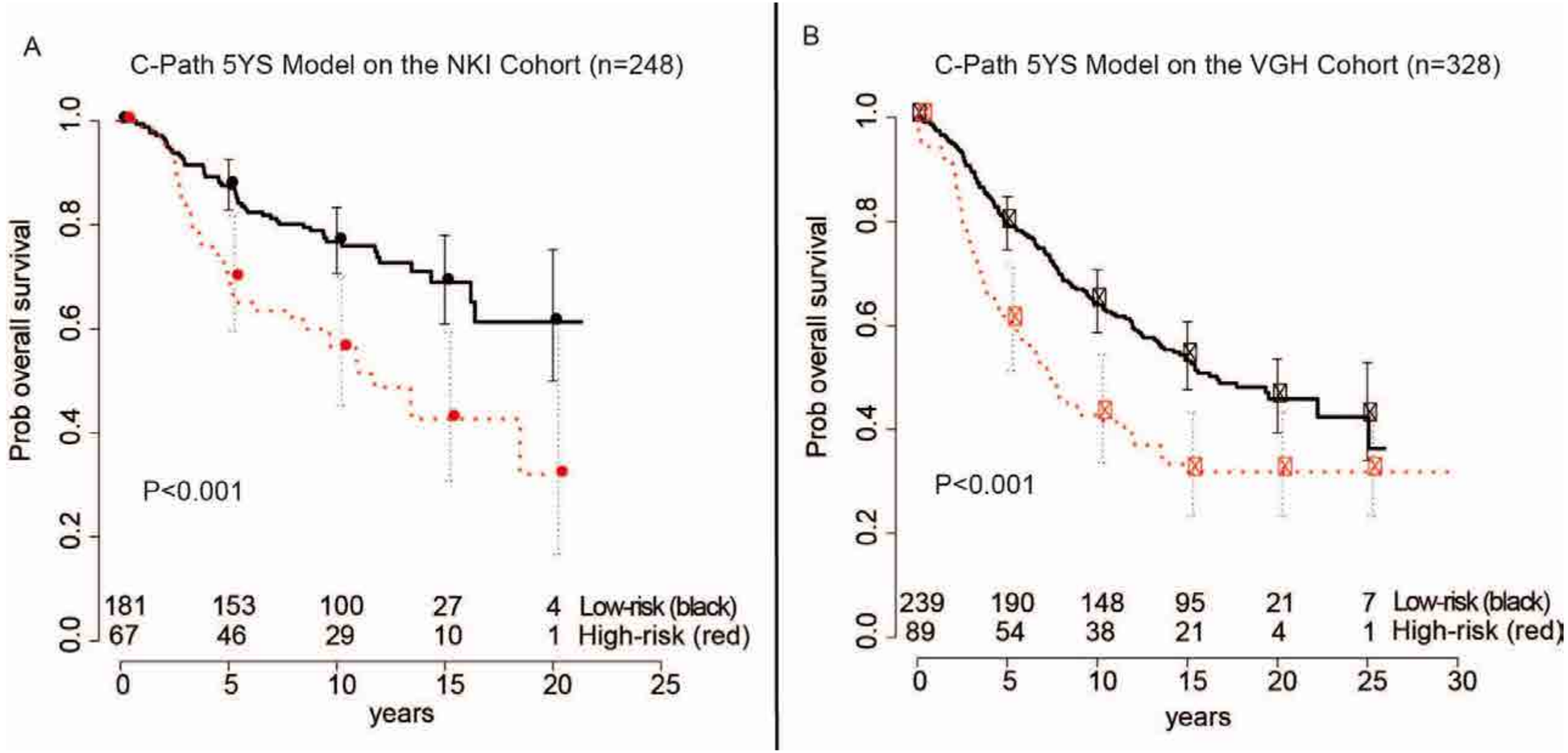
But what will come next? Is it possible to hypothesize that VC will completely substitute our traditional glass slides? Maybe yes, and let us describe the "science fiction" new millennium *digi*tal *patho*logy laboratory, which we will call "DIGIPATH."

Extracting a rich quantitative feature set



Beck ... Koller. Science Translational Medicine 2011

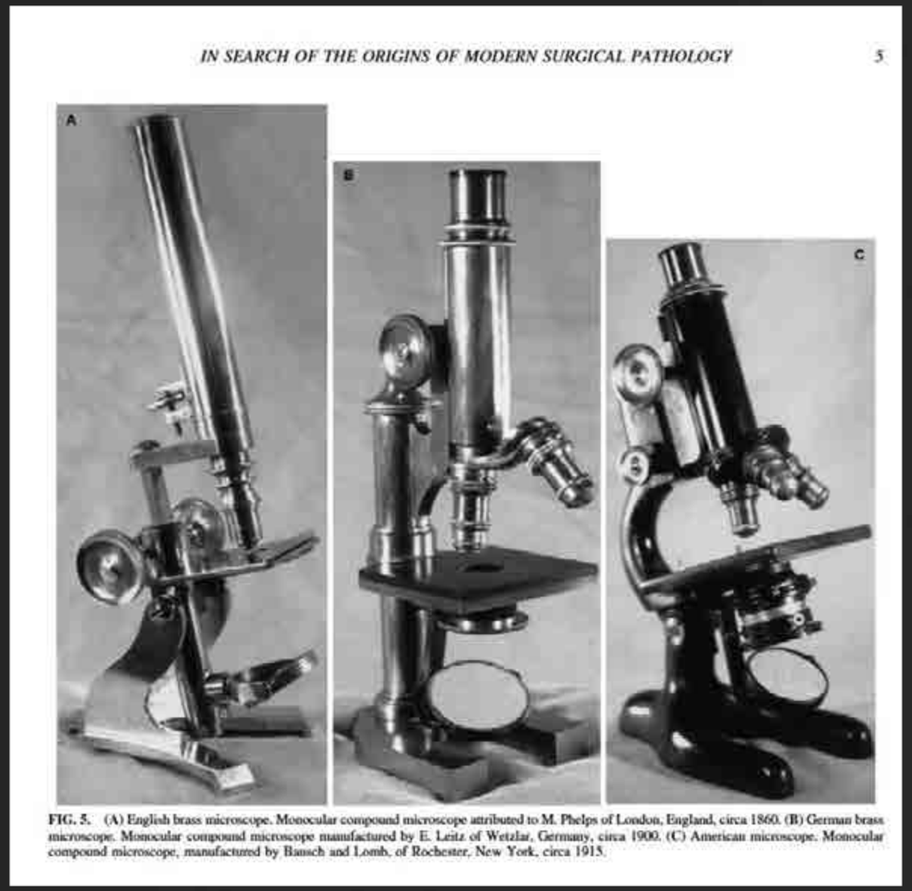
C-Path 5YS Score Significantly Associated with Overall Survival on Both Cohorts



Beck … Koller. Science Translational Medicine 2011

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

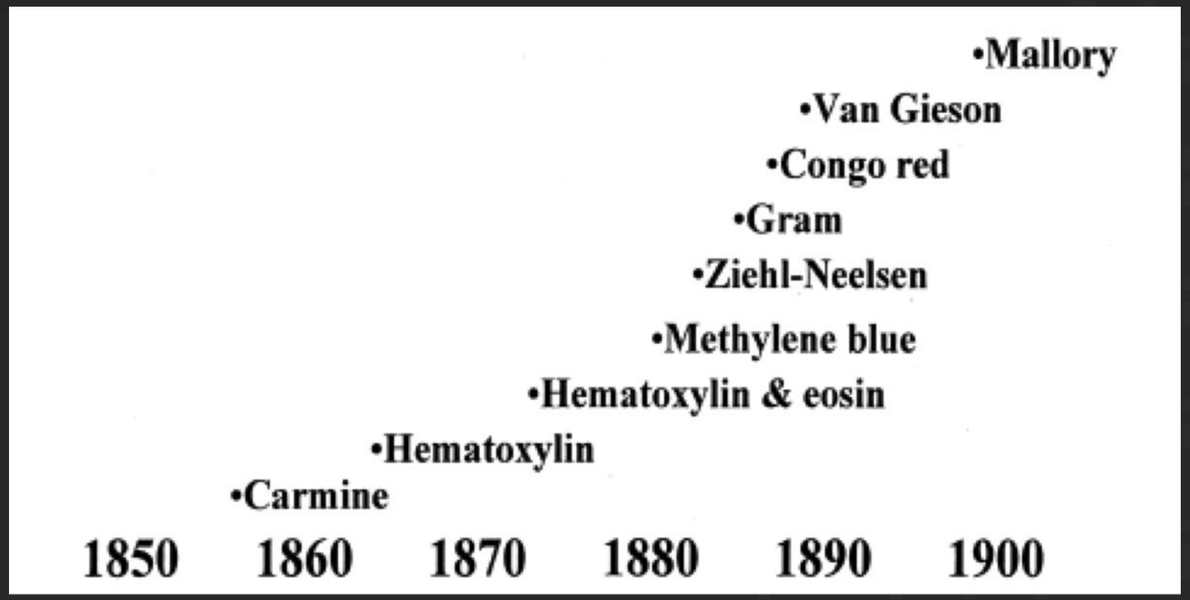
And core technology breakthroughs in routine use are from the 19th century



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

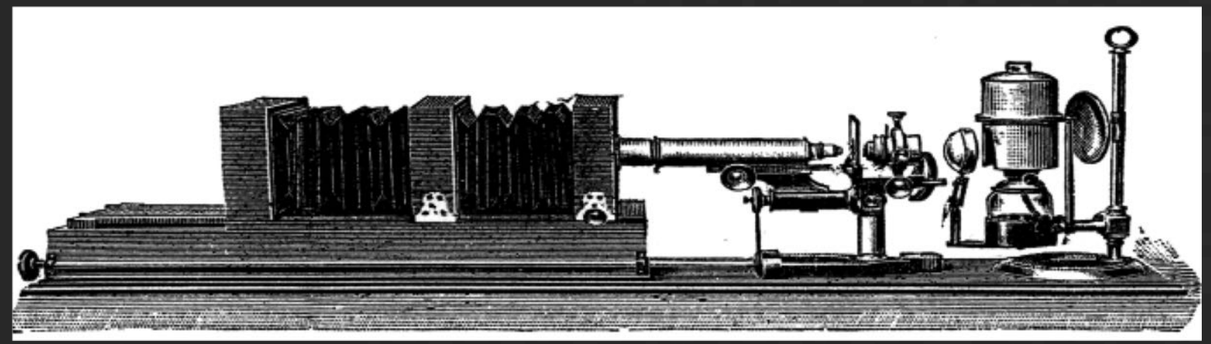
Histochemical Stains

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

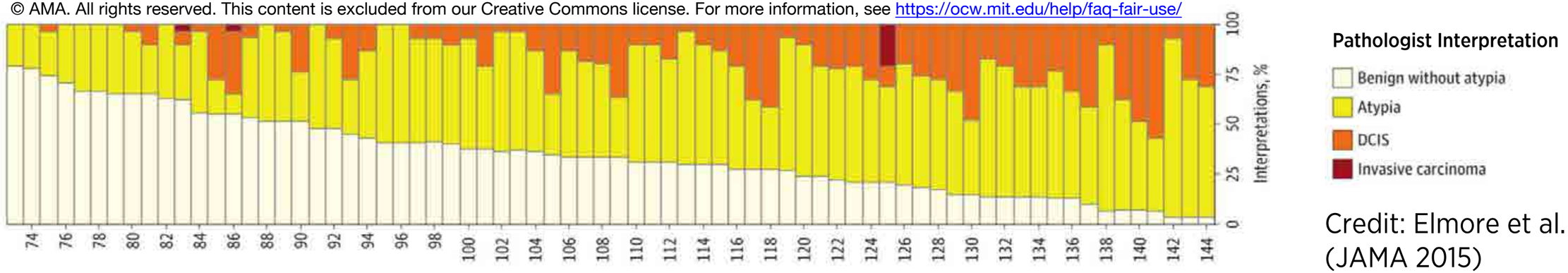


Photomicroscope

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 II Interpretation of Same Individual Pathologist					
Phase I Interpretation of 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

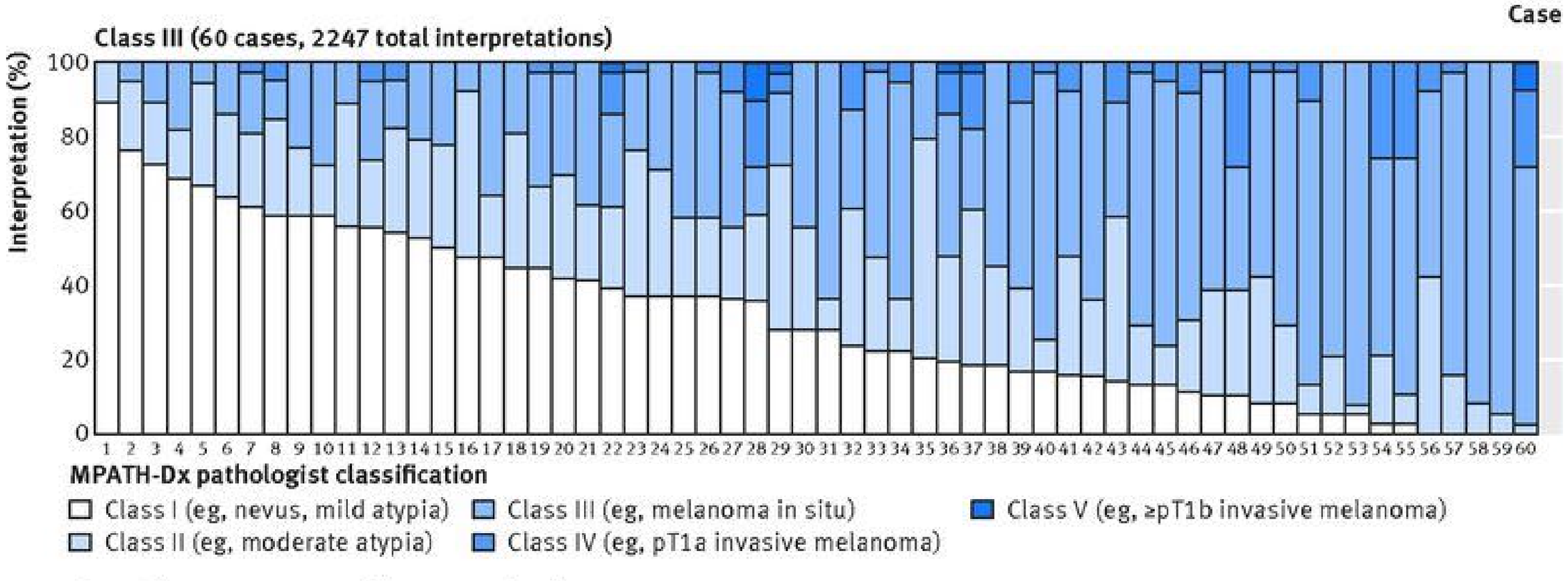
Overall concordance rate of 75% on breast biopsies.

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.

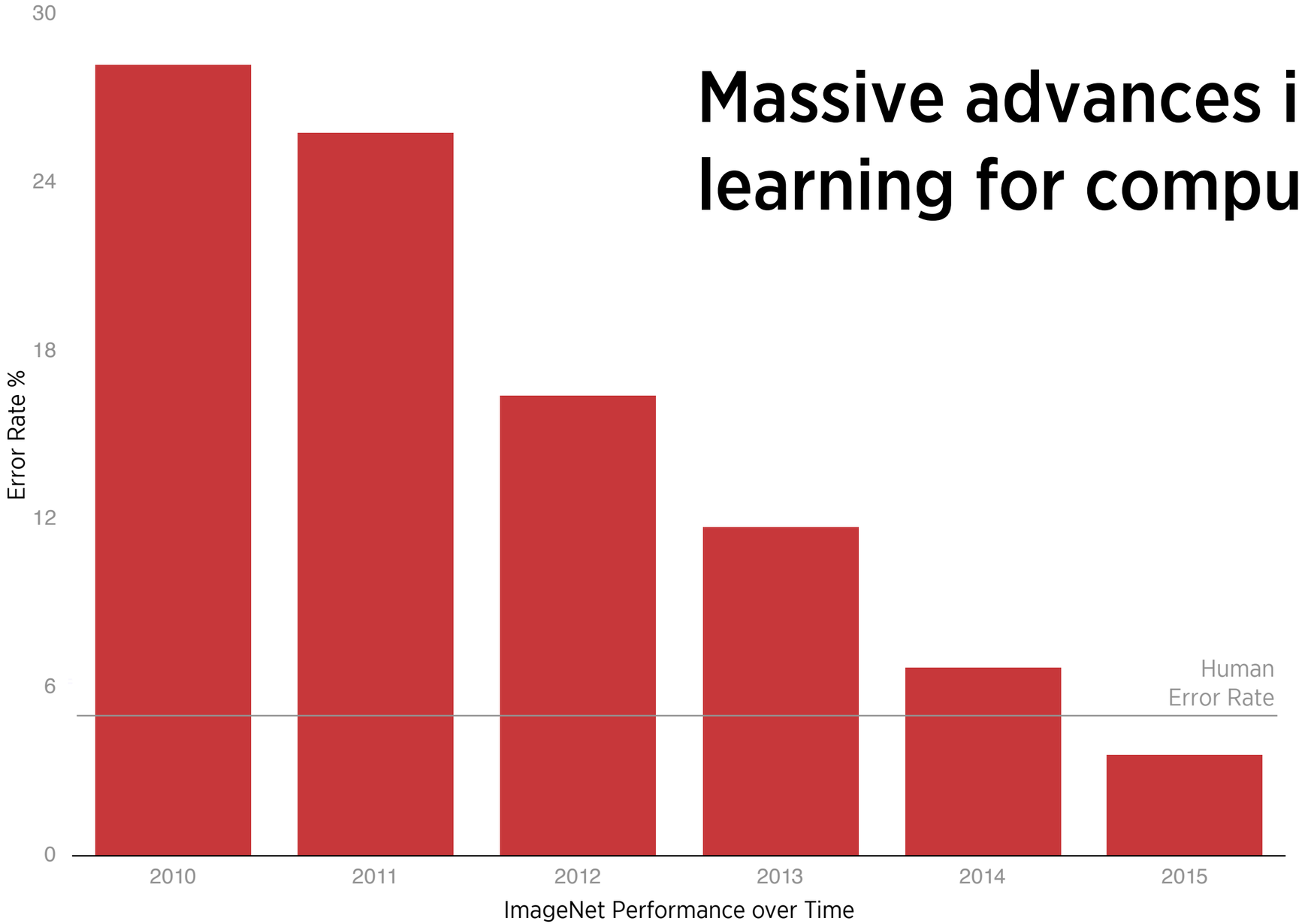
Discordance among pathologists is common in interpretation of melanocytic neoplasms on skin biopsies



- 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%

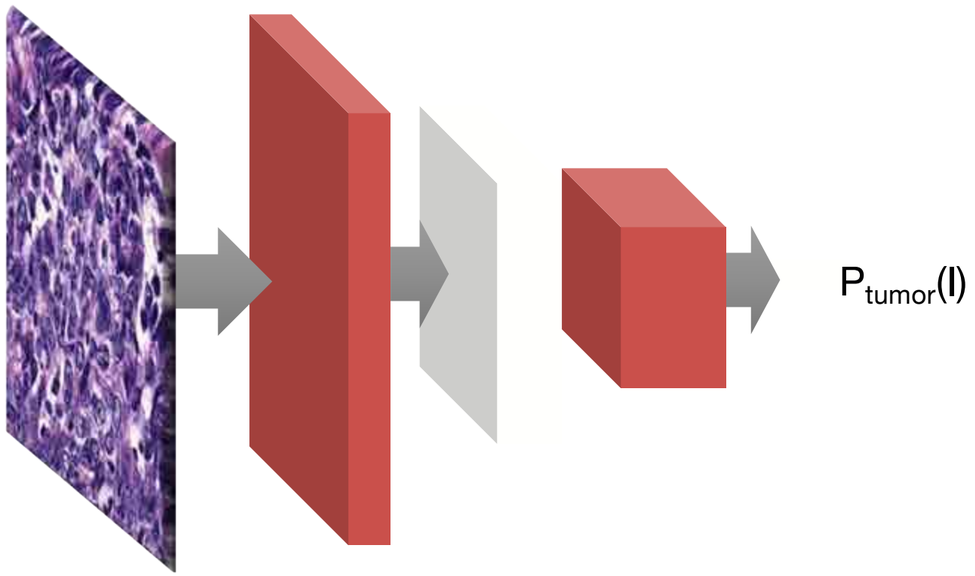
BMJ 2017;357:j2813 | doi: 10.1136/bmj.j2813

Massive advances in deep learning for computer vision…



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!

PathAI

A diagnosis/detection example:

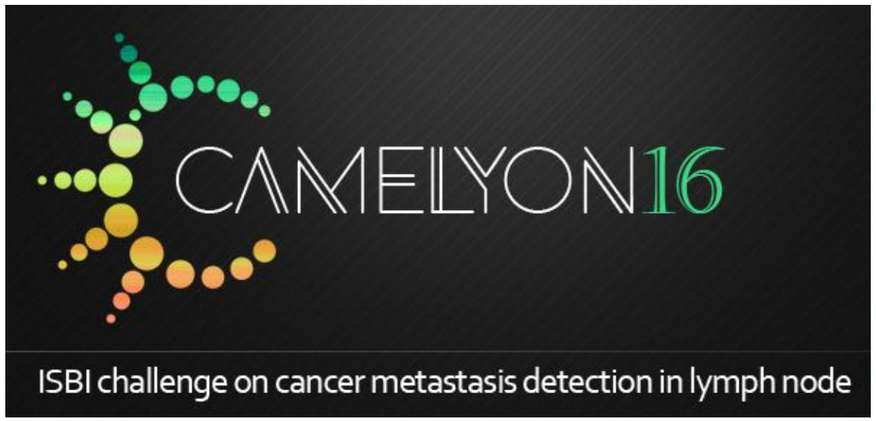
Breast cancer metastases

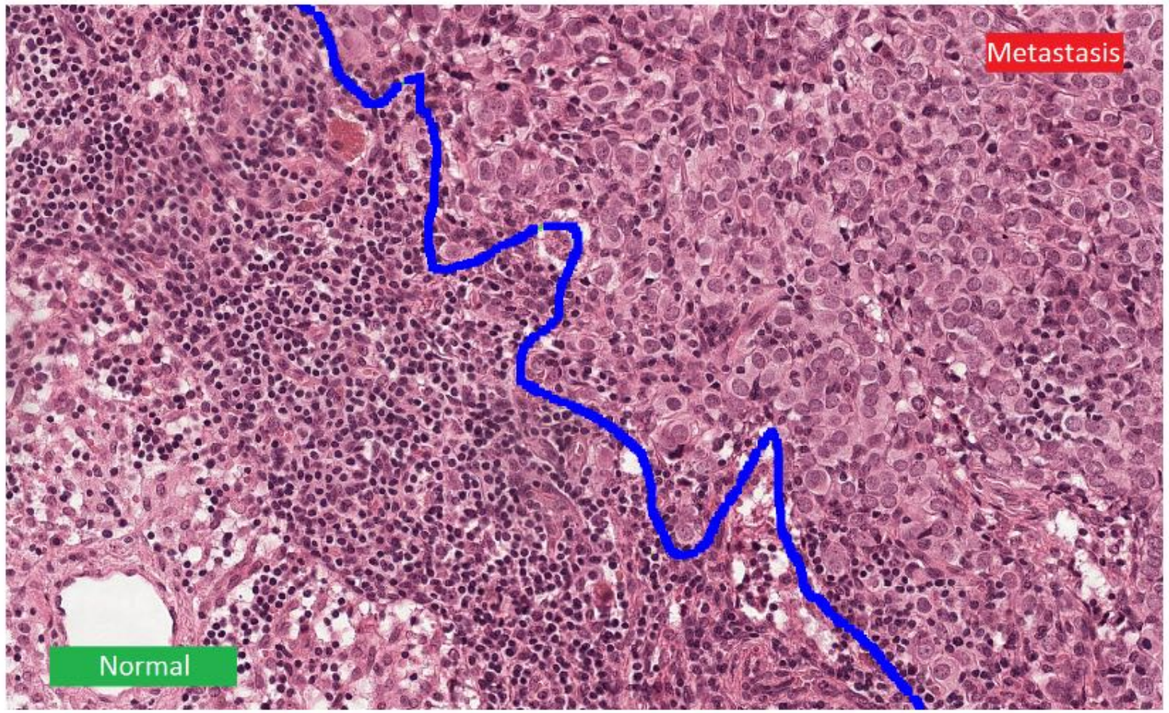
- After a primary mass discovered, lymph nodes are biopsied
- Pathologists check these for metastases
- Non-zero failure rate: a retrospective study found a 24% disagreement rate[1]

[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

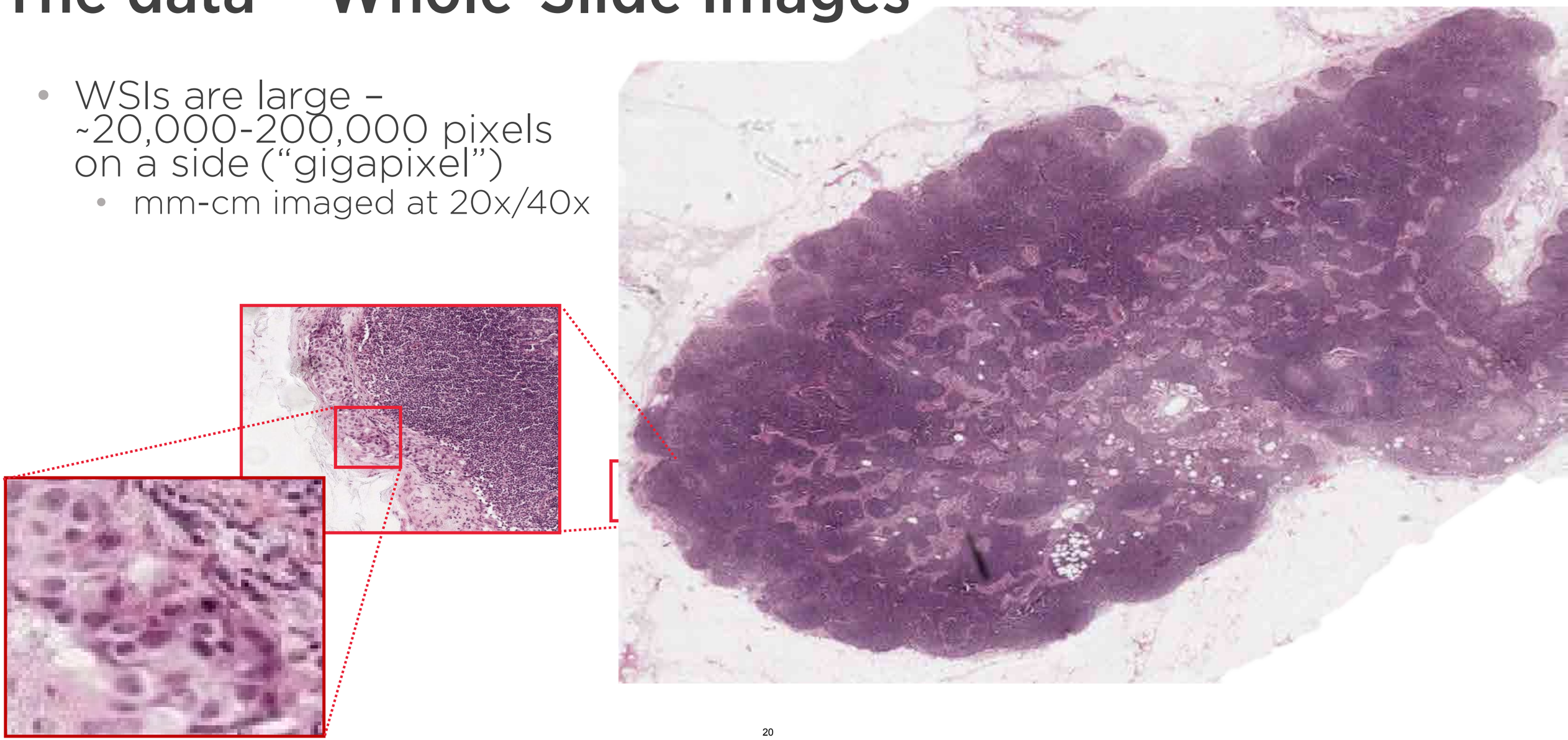
- H & E stained, Formalin-Fixe Paraffin-Embedded (FFPE)
  - 270 training slides, 129 test
- Annotated by a panel





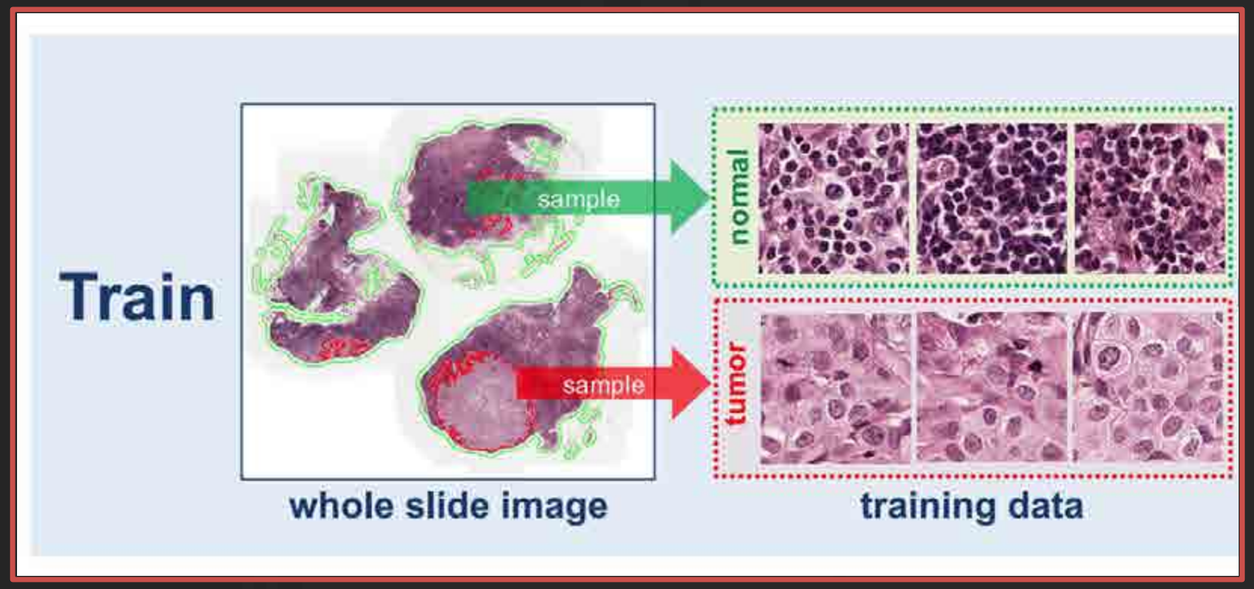
The data – Whole-Slide Images

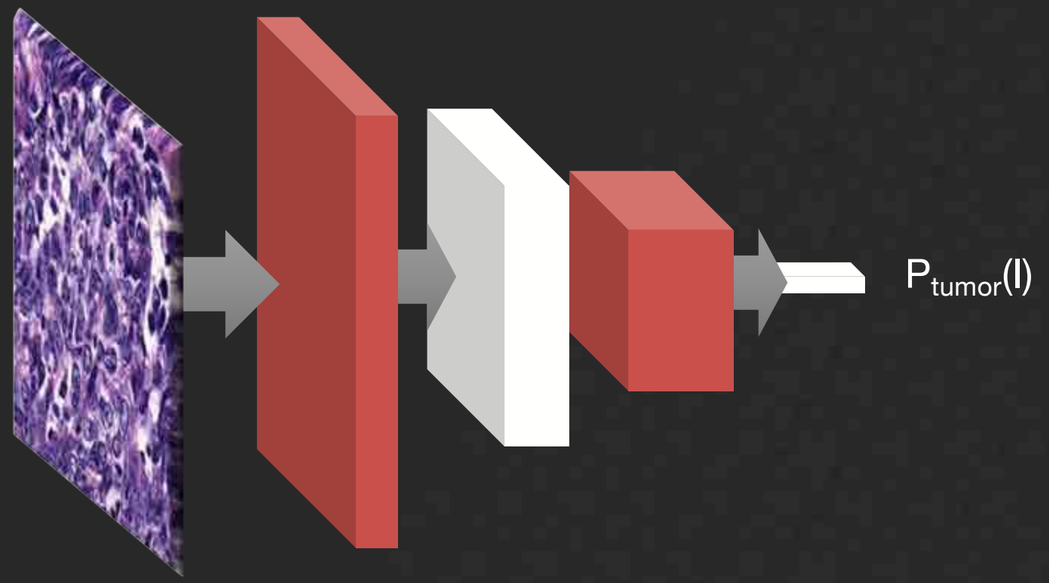
- WSIs are large – ~20,000-200,000 pixels on a side ("gigapixel")
  - mm-cm imaged at 20x/40x



Approach

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



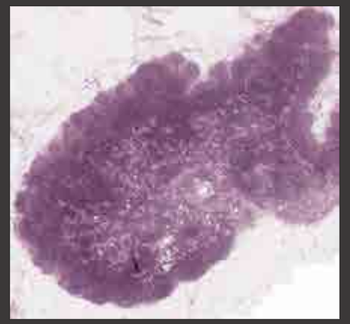


PathAI

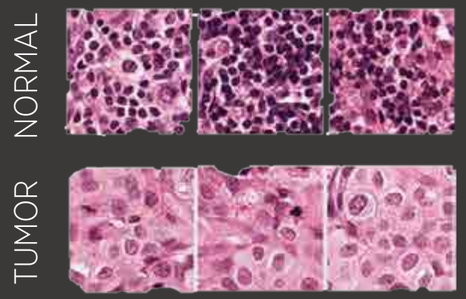
Successfully applied deep learning approach to pathology

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

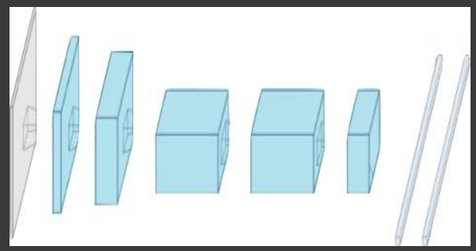
TRAIN



Whole Slide Image

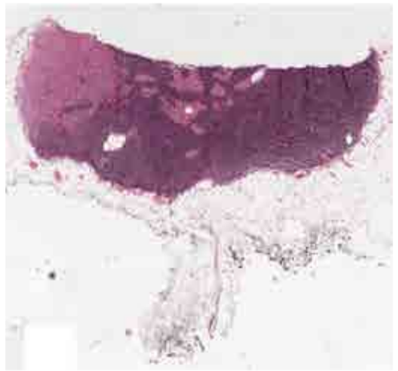


Training Data



Deep Model

TEST



Whole Slide Image

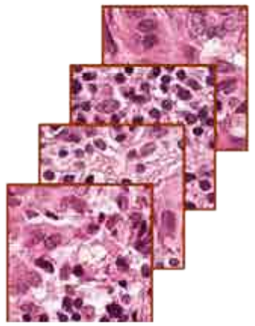
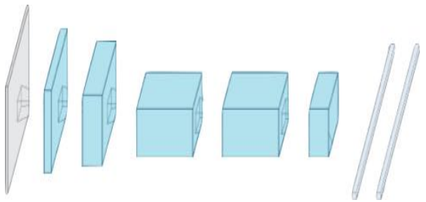
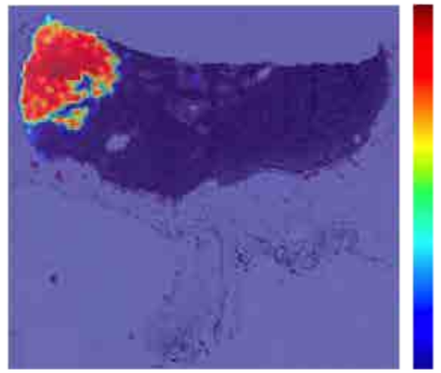


Image Patches



Deep Model from Training

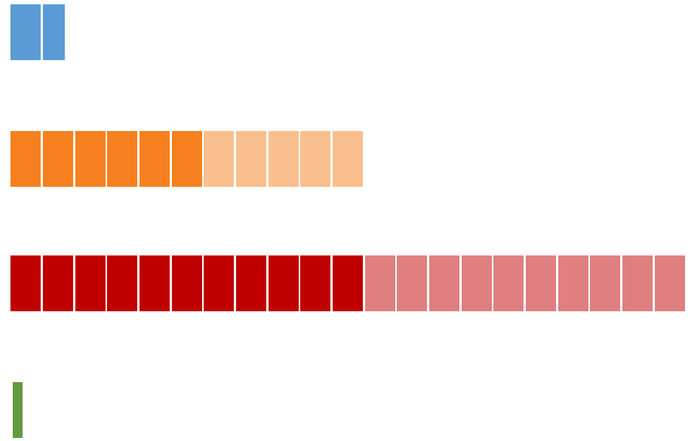


Tumor Probability Map

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

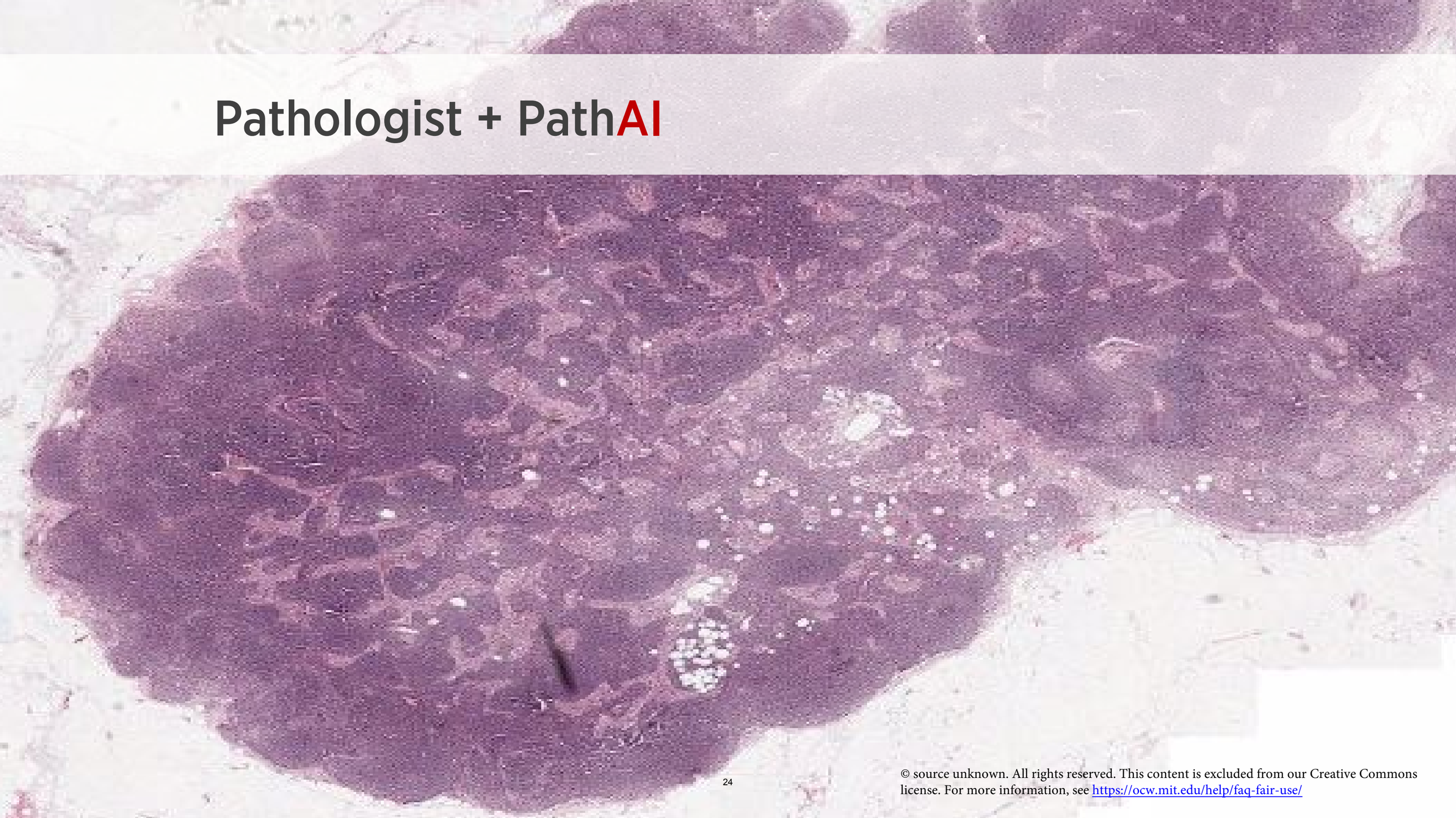
	Error Rate (1-AUC)
Pathologists in competition	3.5%
Pathologists in clinical practice[1]	13 – 26%
Pathologists on micro-metastasis[2]	23 – 42%
Deep learning model	0.65%



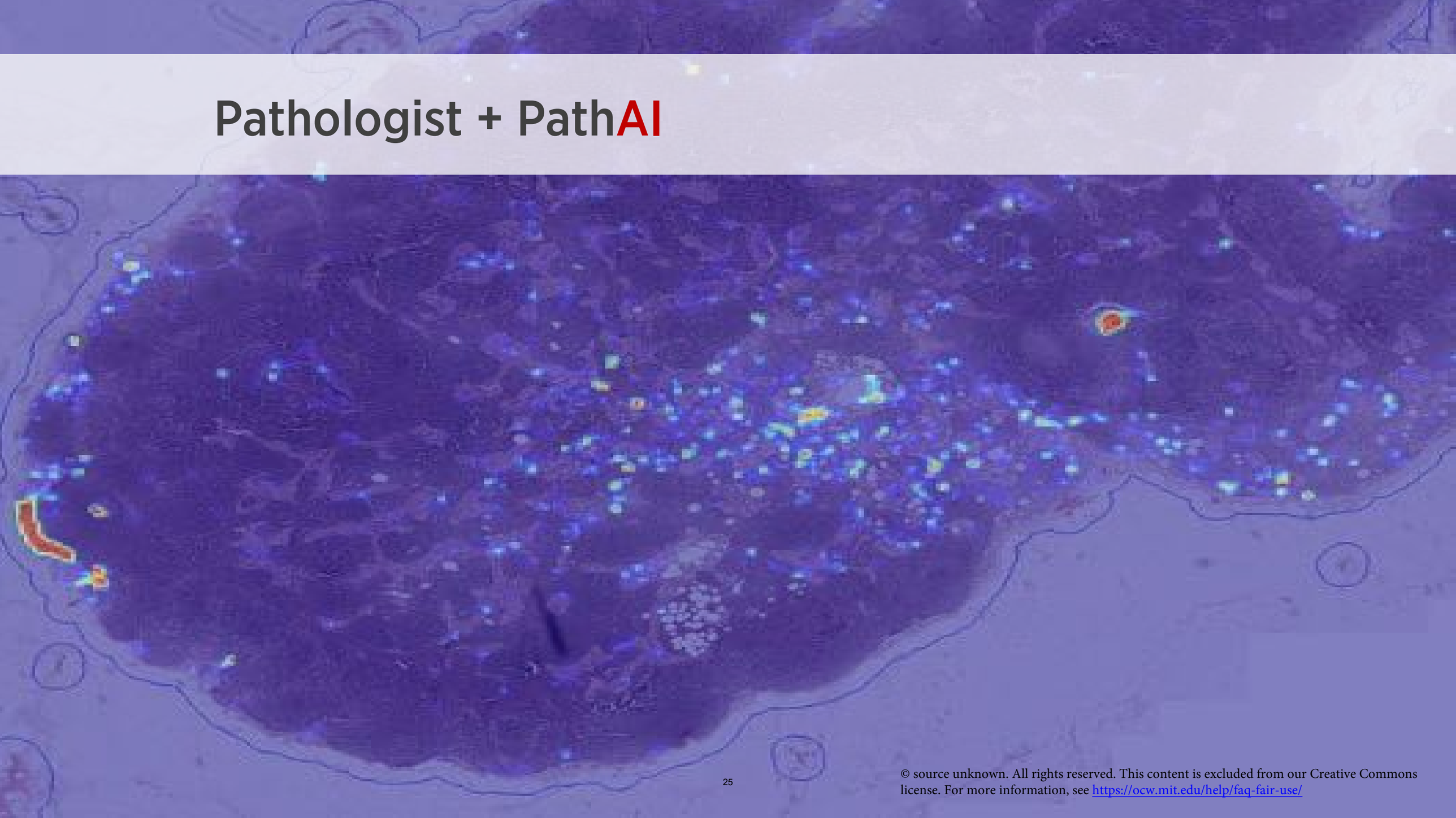
[1] n=12 [2] Small tumors

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

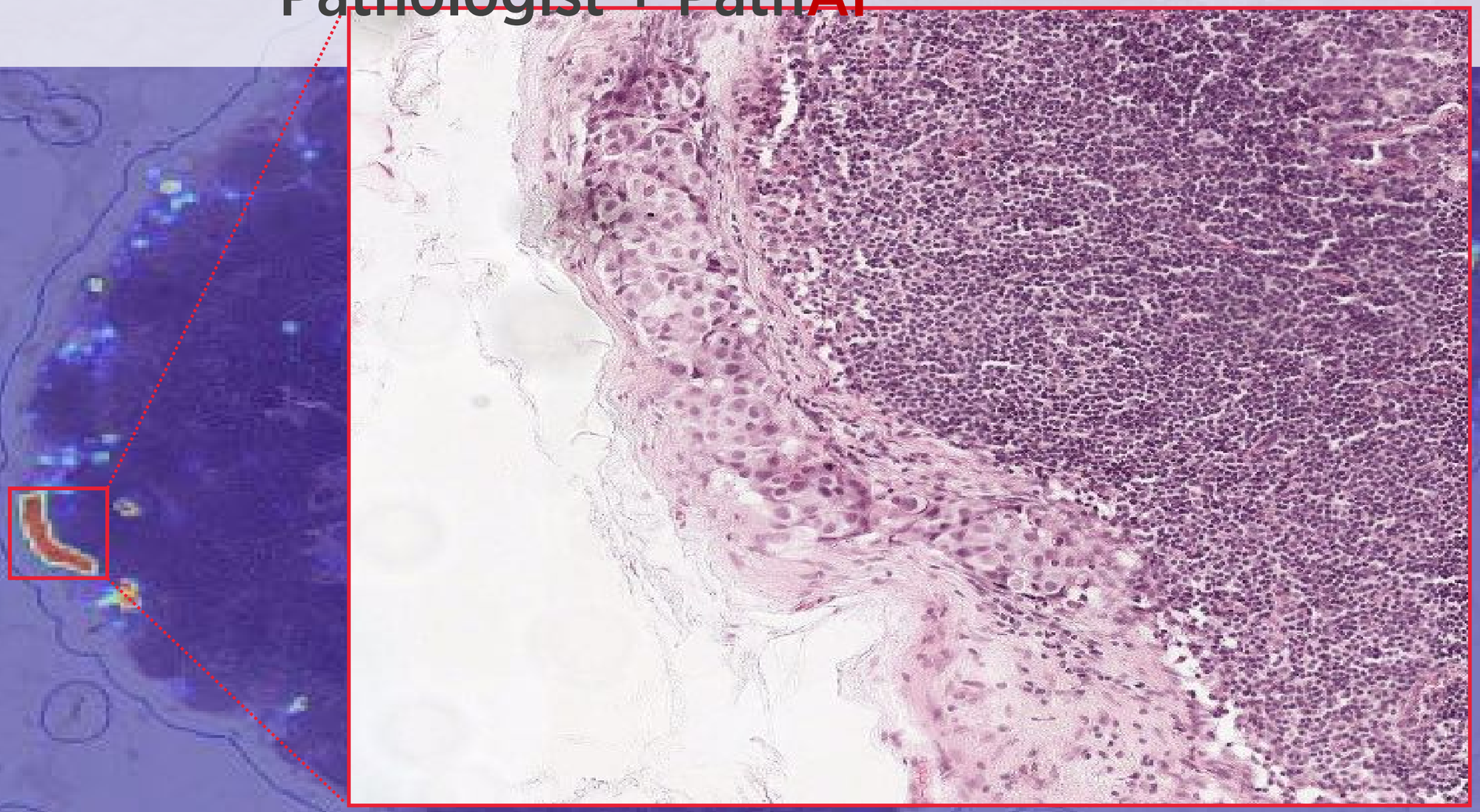
Pathologist + PathAI



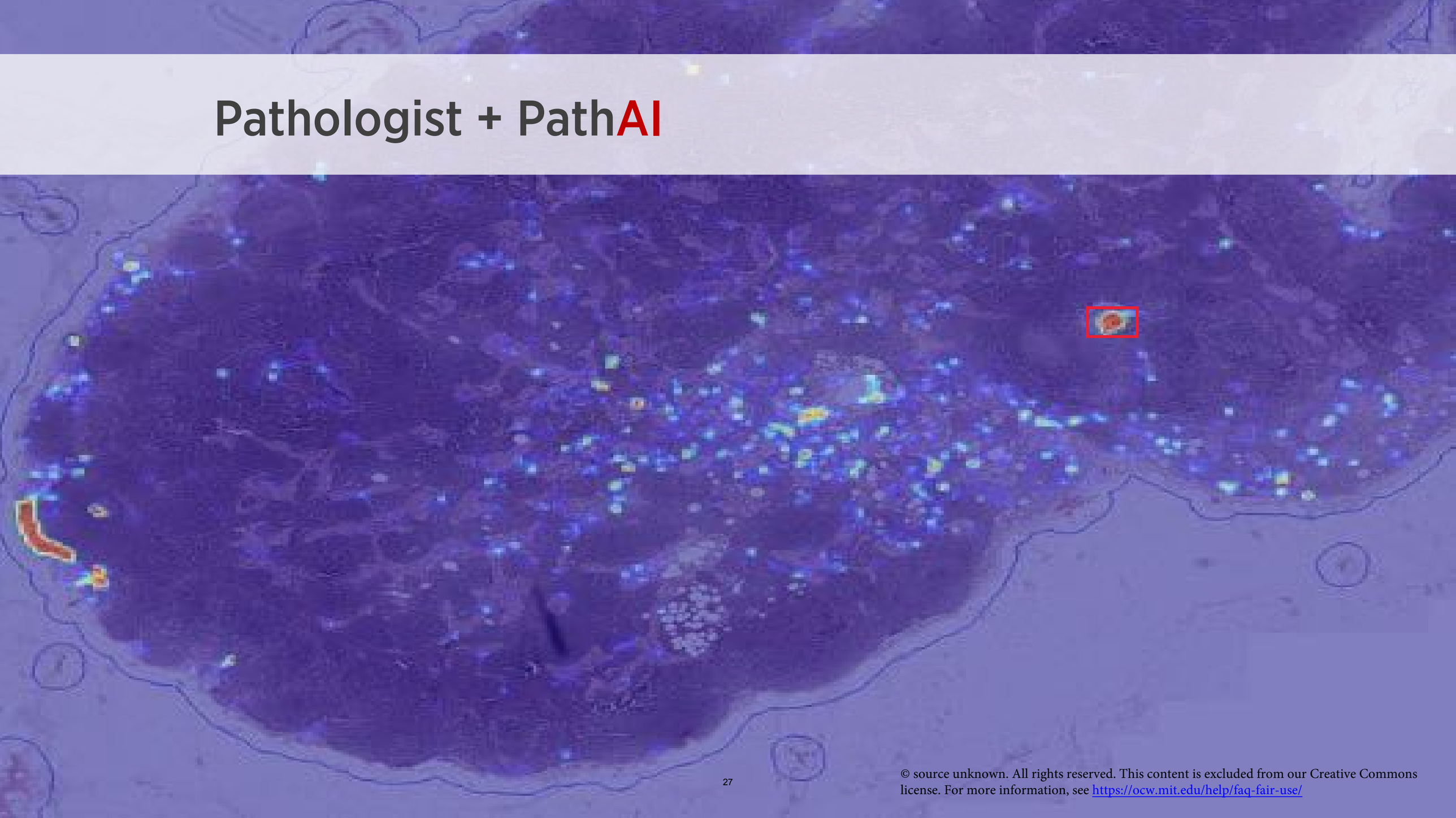
Pathologist + PathAI



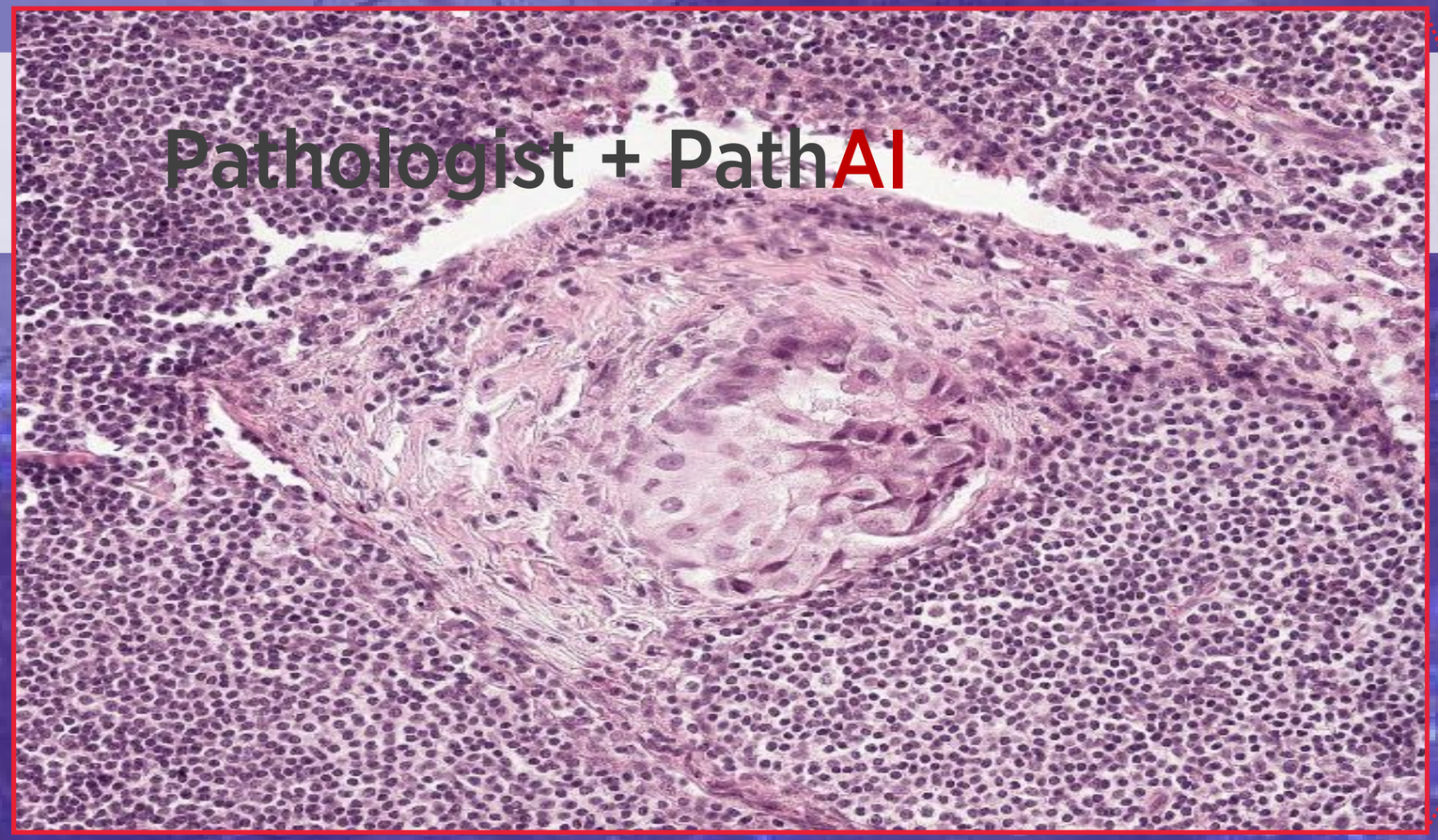
Pathologist + PathAI

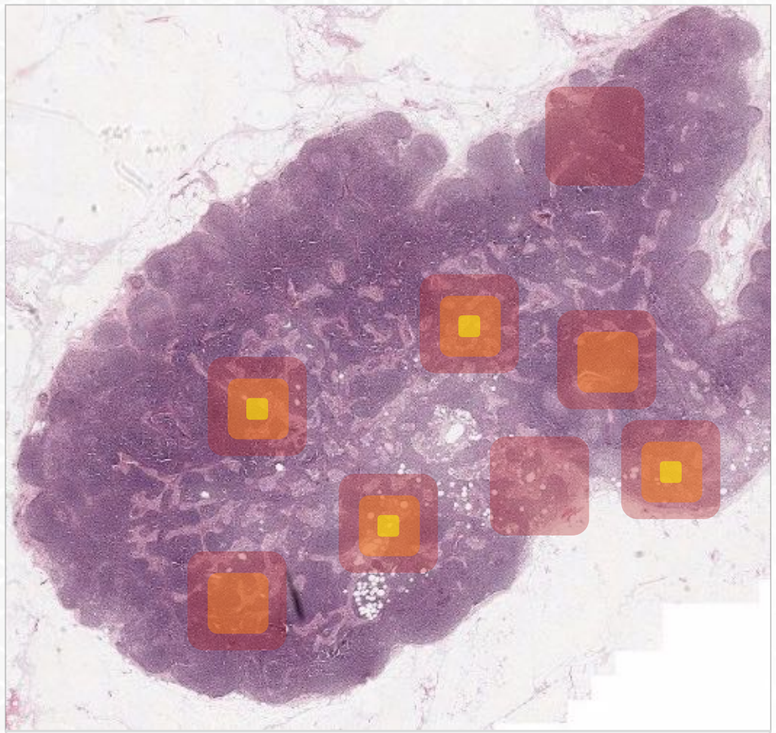


Pathologist + PathAI



Pathologist + PathAI



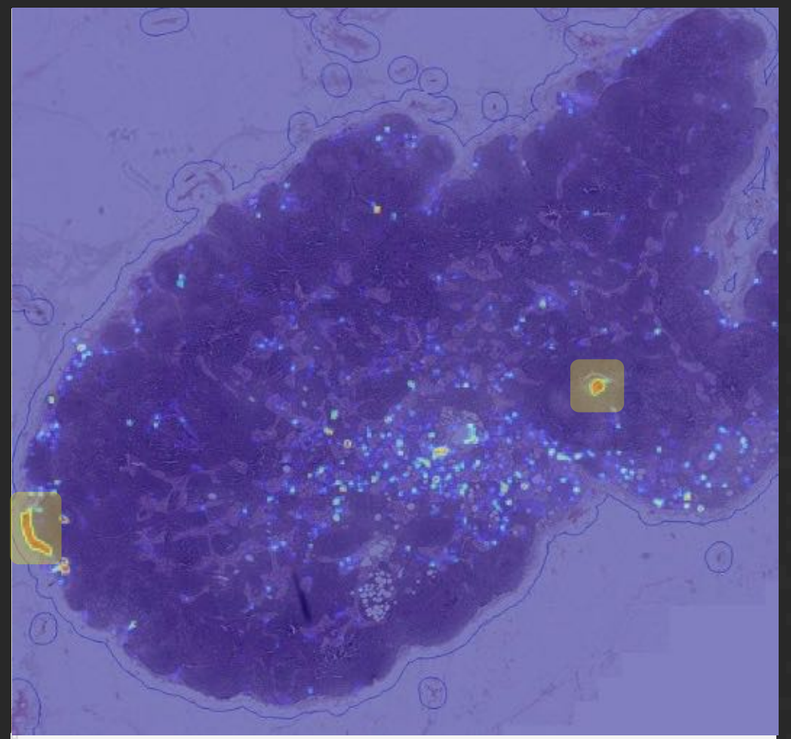


Pathology Report

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

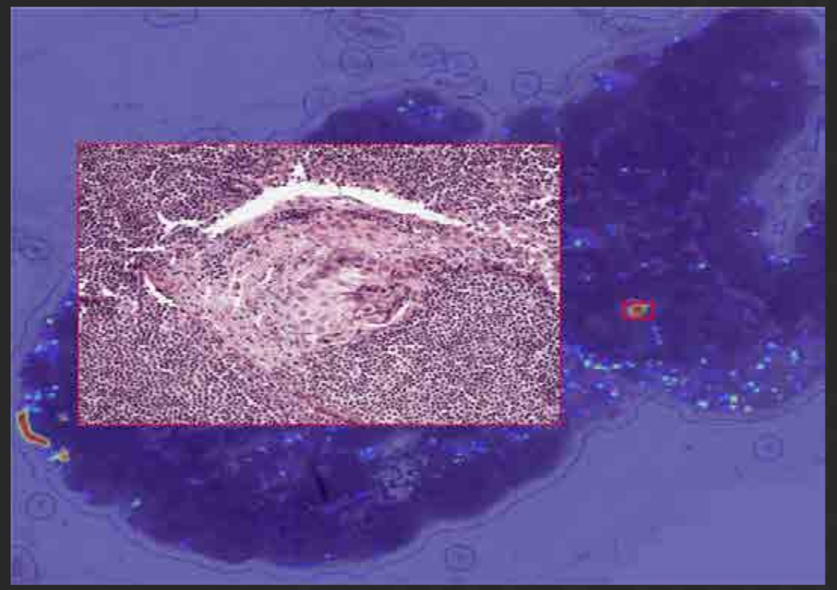
Time per slide: 10– 60 seconds
Accuracy: >99.5%
Reproducibility: High

PathAI

Proprietary & Confidential

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

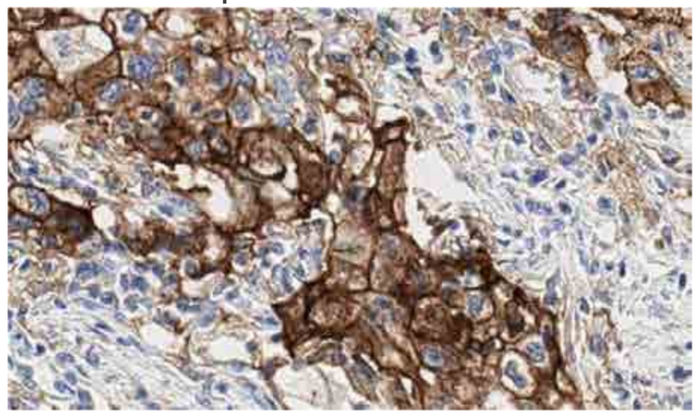
- Some cancers express immune-inhibitory ligands, activating immune “checkpoints”

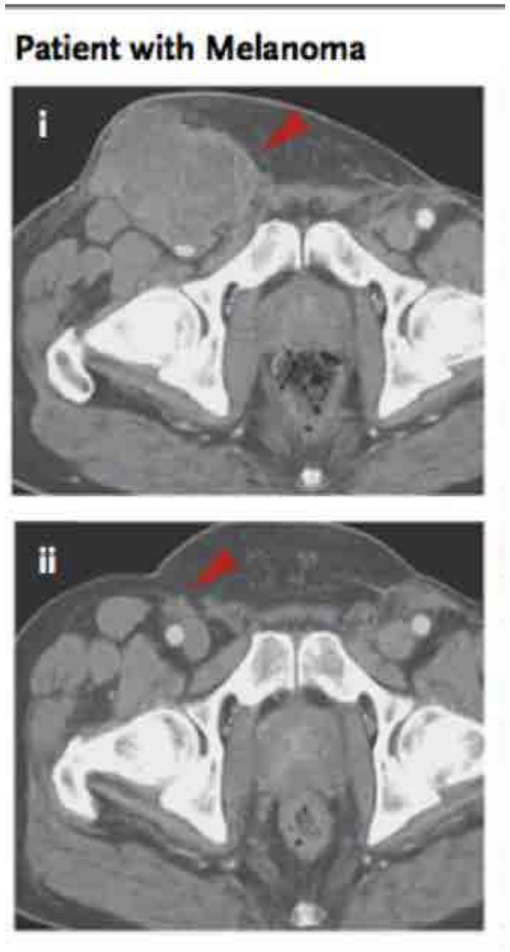
- “checkpoint inhibitors” mask these signals, unleashing the immune system

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.

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.)

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.

Can we do better?

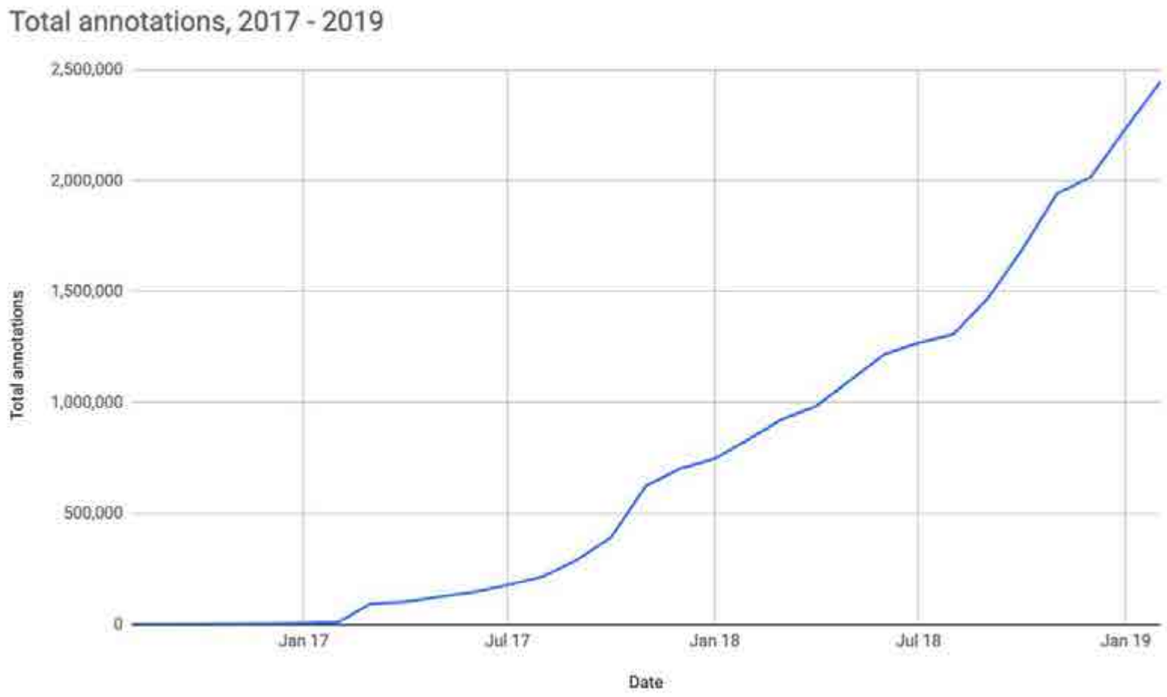
- Deep learning is data hungry
  - Need 10s of thousands of precise cell annotations

First, we need the data

PathAI

Board-certified training data

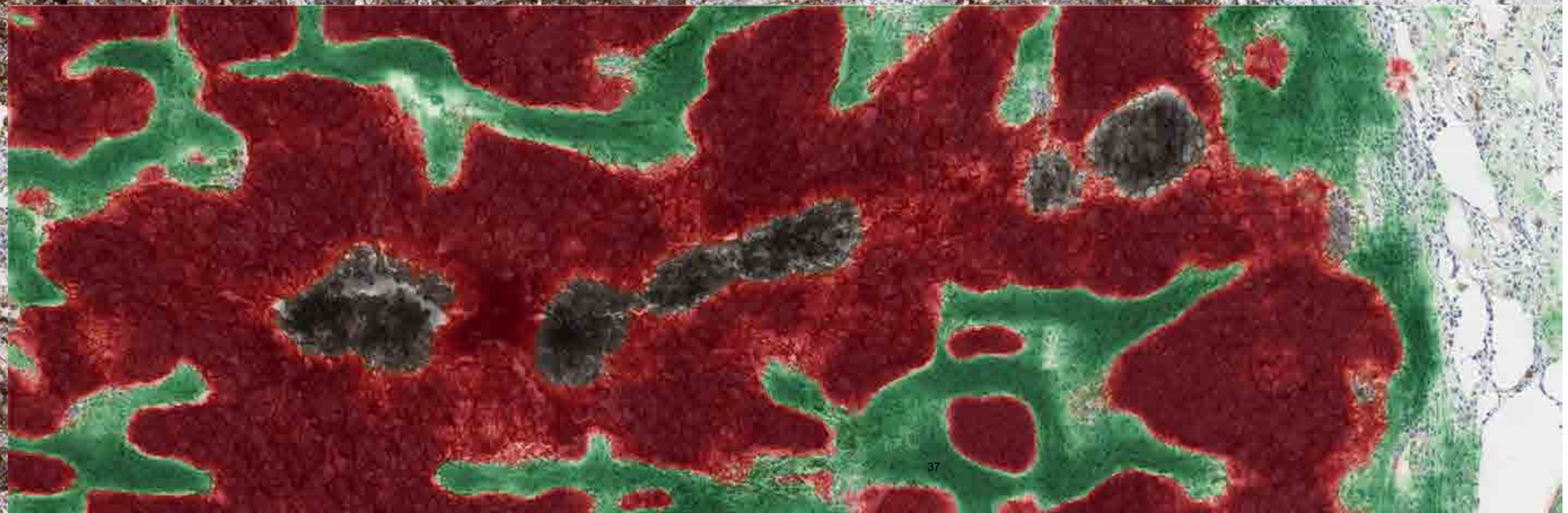
Working with pathologists around the country to generate high-quality annotations



>2.5M
notations

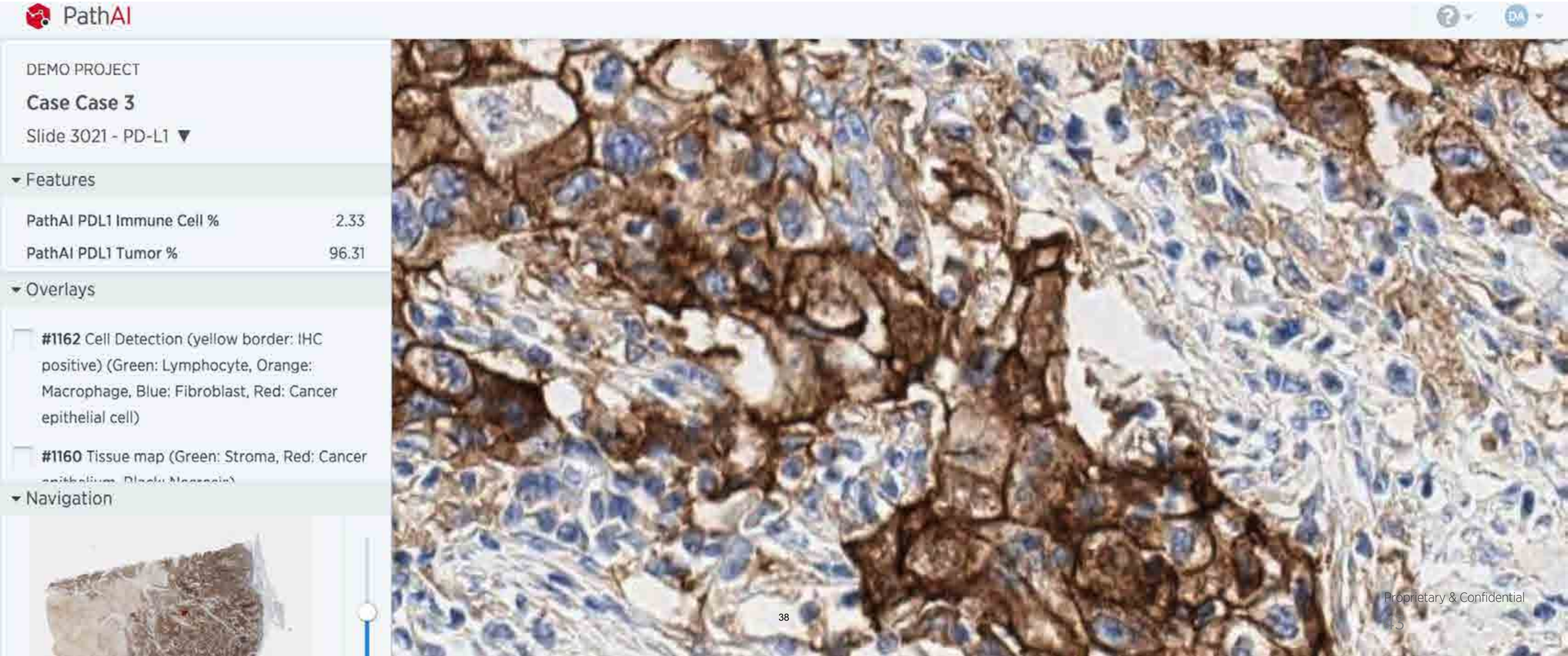
Automatic and exhaustive regions of interest

tumor and relevant stroma



IHC expression difficult to detect

on immune and tumor cells



Exhaustive automated classification

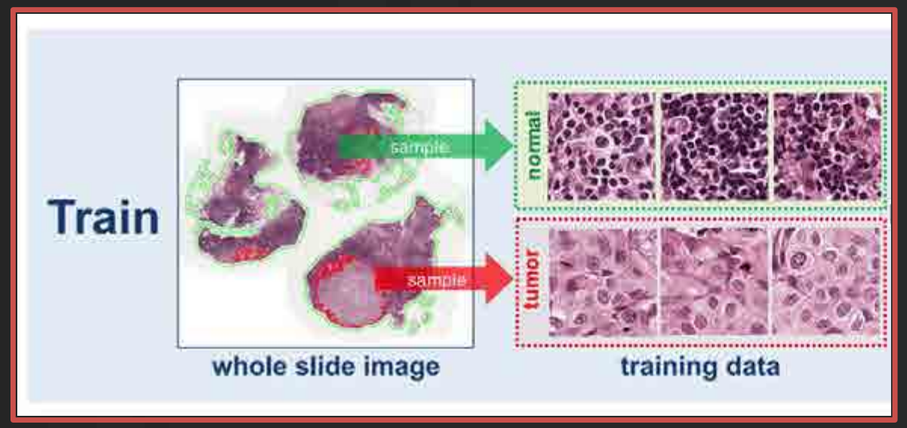
Cell type and cellular IHC positivity classification

Quantitative and reproducible PD-L1 scoring

- Manual review: few hundred cells over a few arbitrary high-power fields of view
- Automated analysis: exhaustive classification of 10k-1million cells

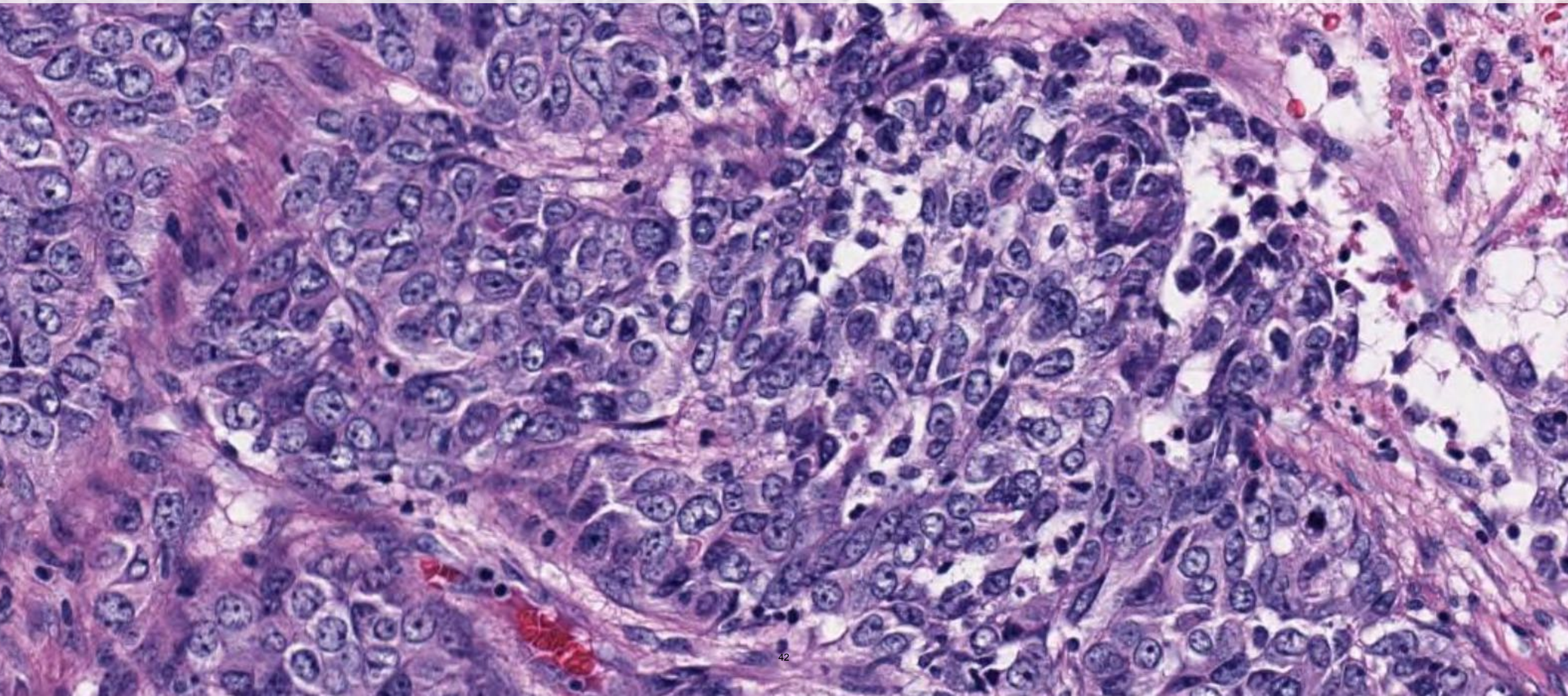
Taking it further
From quantitative assay to patient prediction

- PD-L1 scoring alone reduces billions of pixels to 1-2 numbers.
- Can we identify additional relevant information?
  - Using data from randomized controlled clinical trials
- However: Millions of patches, *hundreds* of patients



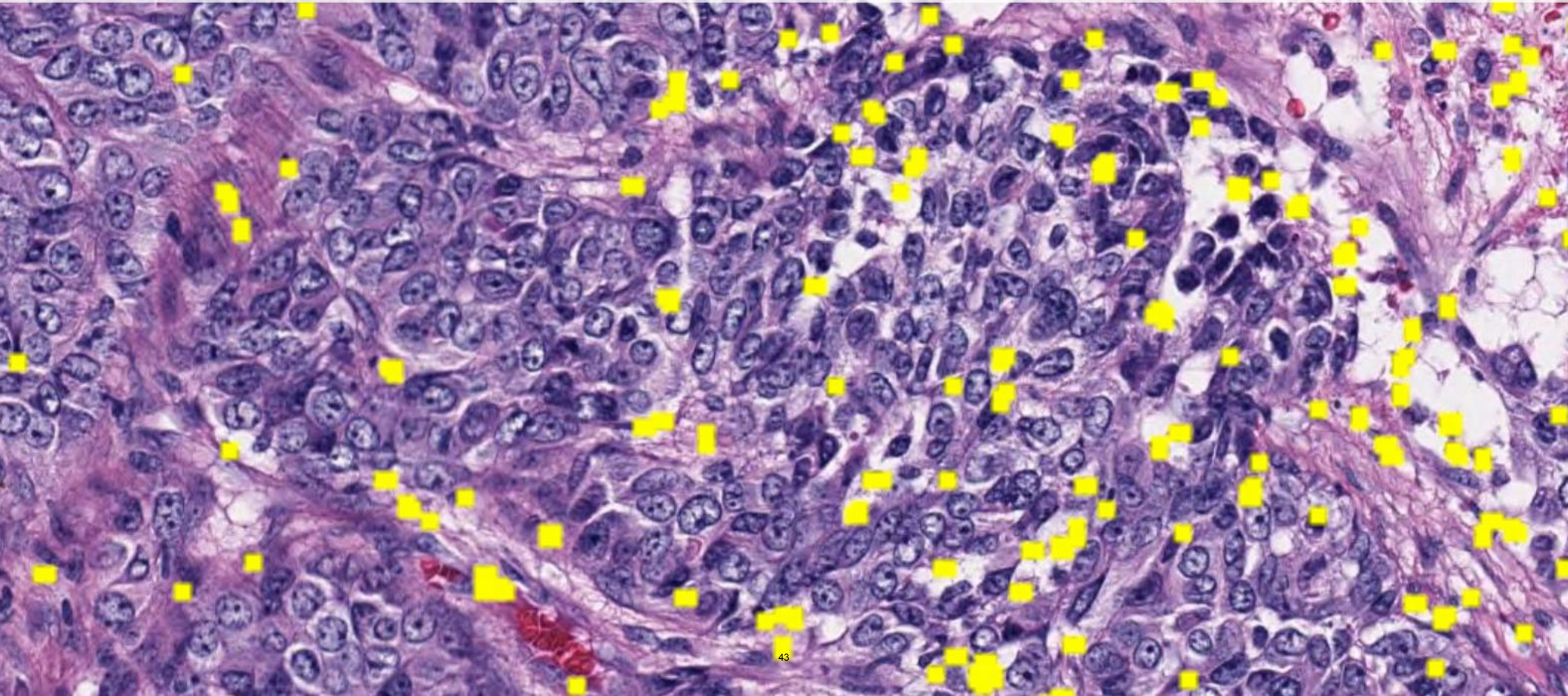
Predictive features guided by biomedical priors

H & E slide matching PD-L1 slide



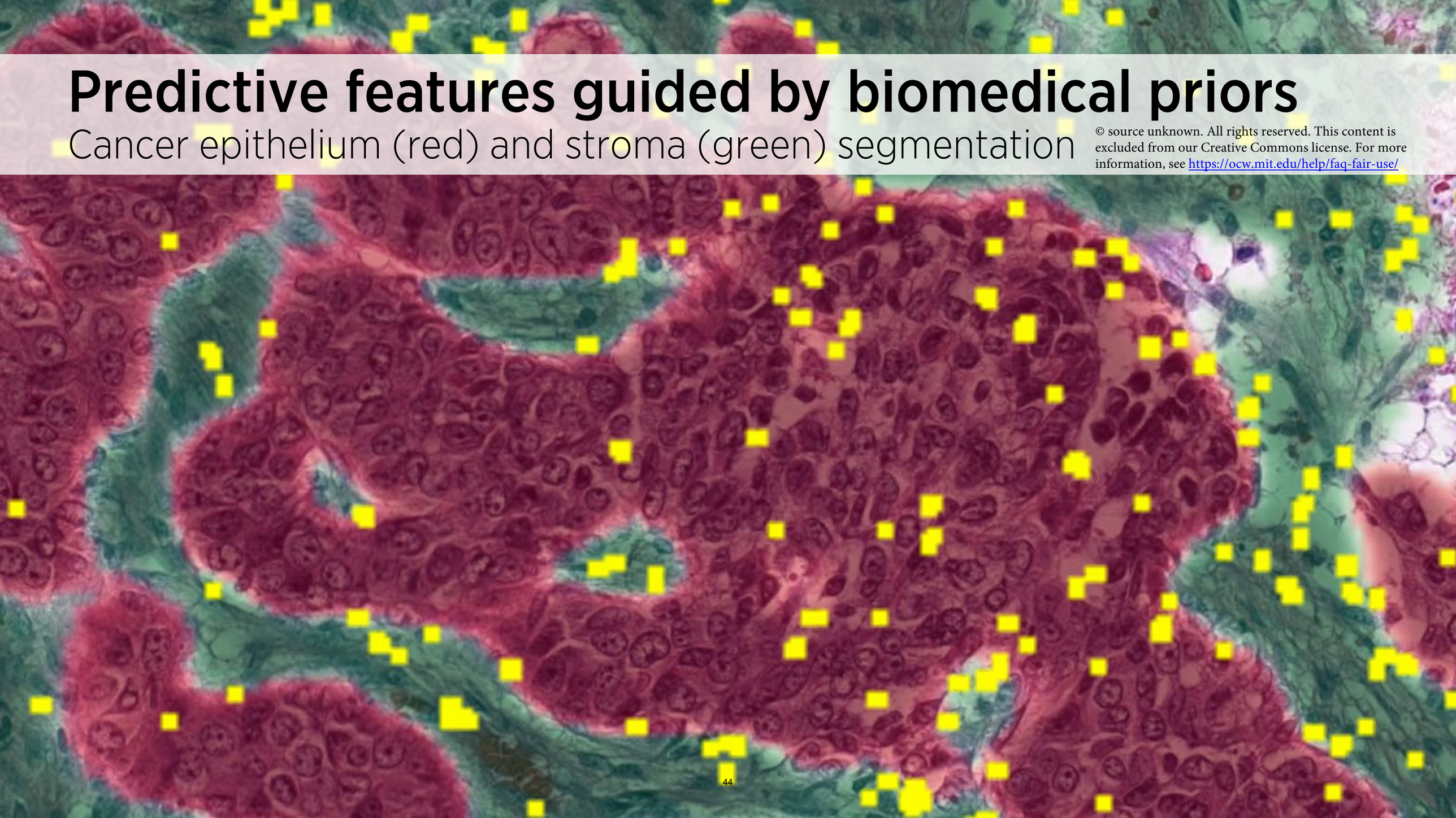
Predictive features guided by biomedical priors

Immune cell (lymphocyte) detection



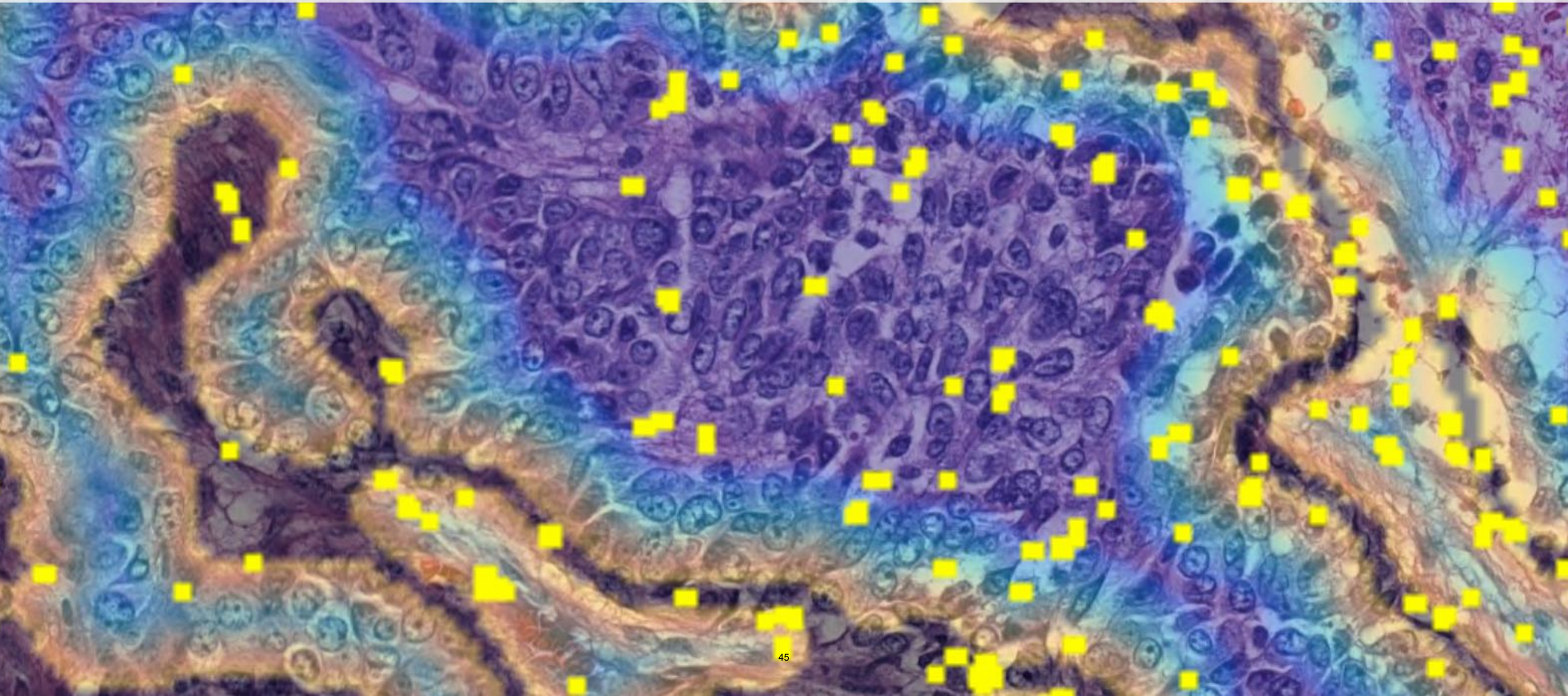
Predictive features guided by biomedical priors

Cancer epithelium (red) and stroma (green) segmentation

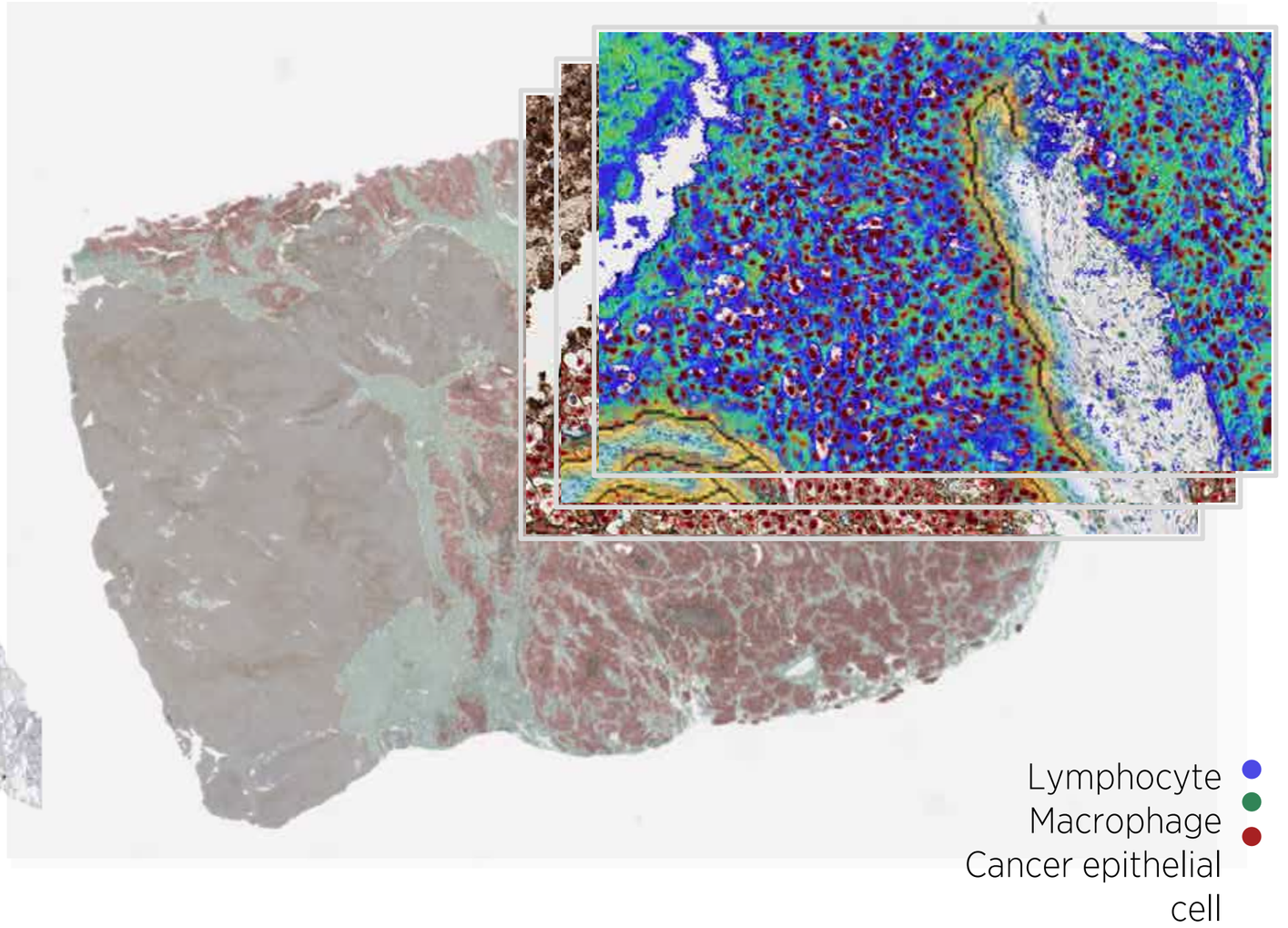


Predictive features guided by biomedical priors

Epithelial-stromal interface definition



Cell-type specific, tissue context-aware IHC-quantification



Data-driven identification of pathological phenotypes associated with drug response

Total number of macrophages in epithelial/stroma interface (80um)

Total number of macrophages in epithelial/stroma interface (120um)

Total number of macrophages in invasive margin (250um)

Total number of lymphocytes in epithelial/stromal interface on H&E stain

Total number of plasma cells in epithelium on H&E stain

Total number of plasma cells in stroma on H&E stain

Tumor (epithelium + stroma) area on H&E stain

Total number of plasma cells in epithelial/stroma interface (40um)

Total number of plasma cells in epithelial/stroma interface (80um)

Area (mm^2) of epithelial/stroma interface (80um) target positive cancer cells on target stain

Area (mm^2) of epithelial PDL-1 positive macrophages on target stain

Necrosis area on target stain

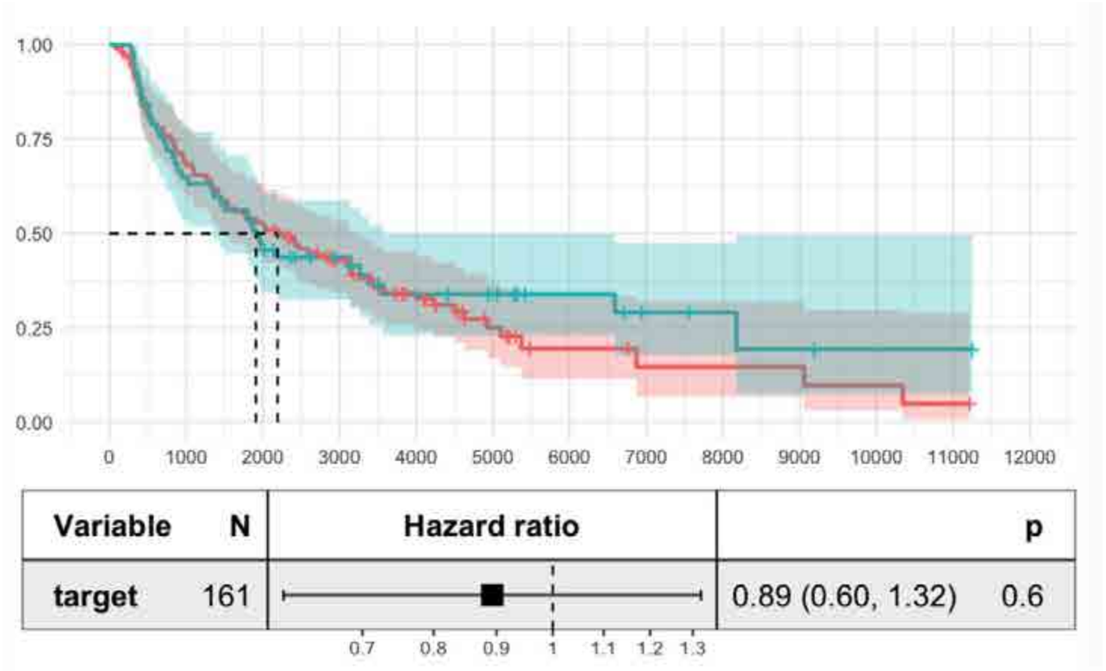
Proportion of tumor infiltrating lymphocytes engaged by target positive macrophages

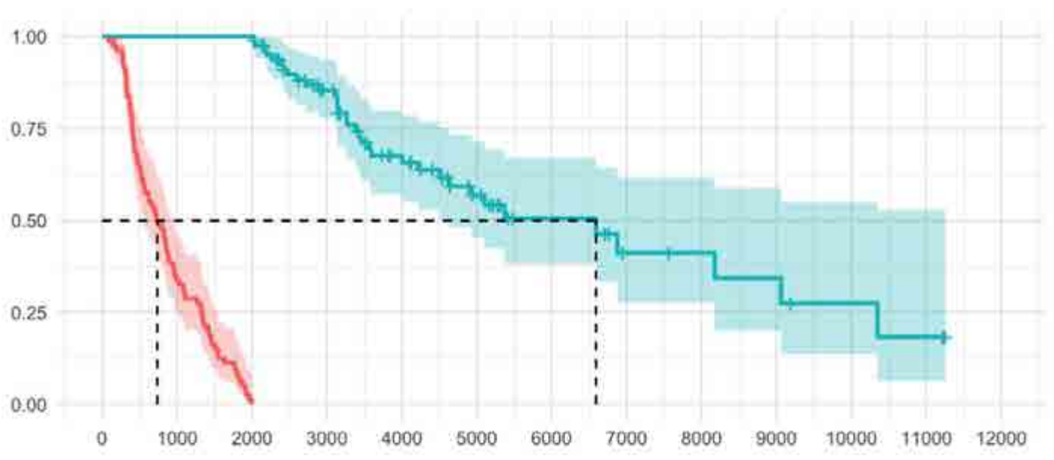
Stroma area on target stain

Tissue area on target stain

Multivariate models predictive of IO response

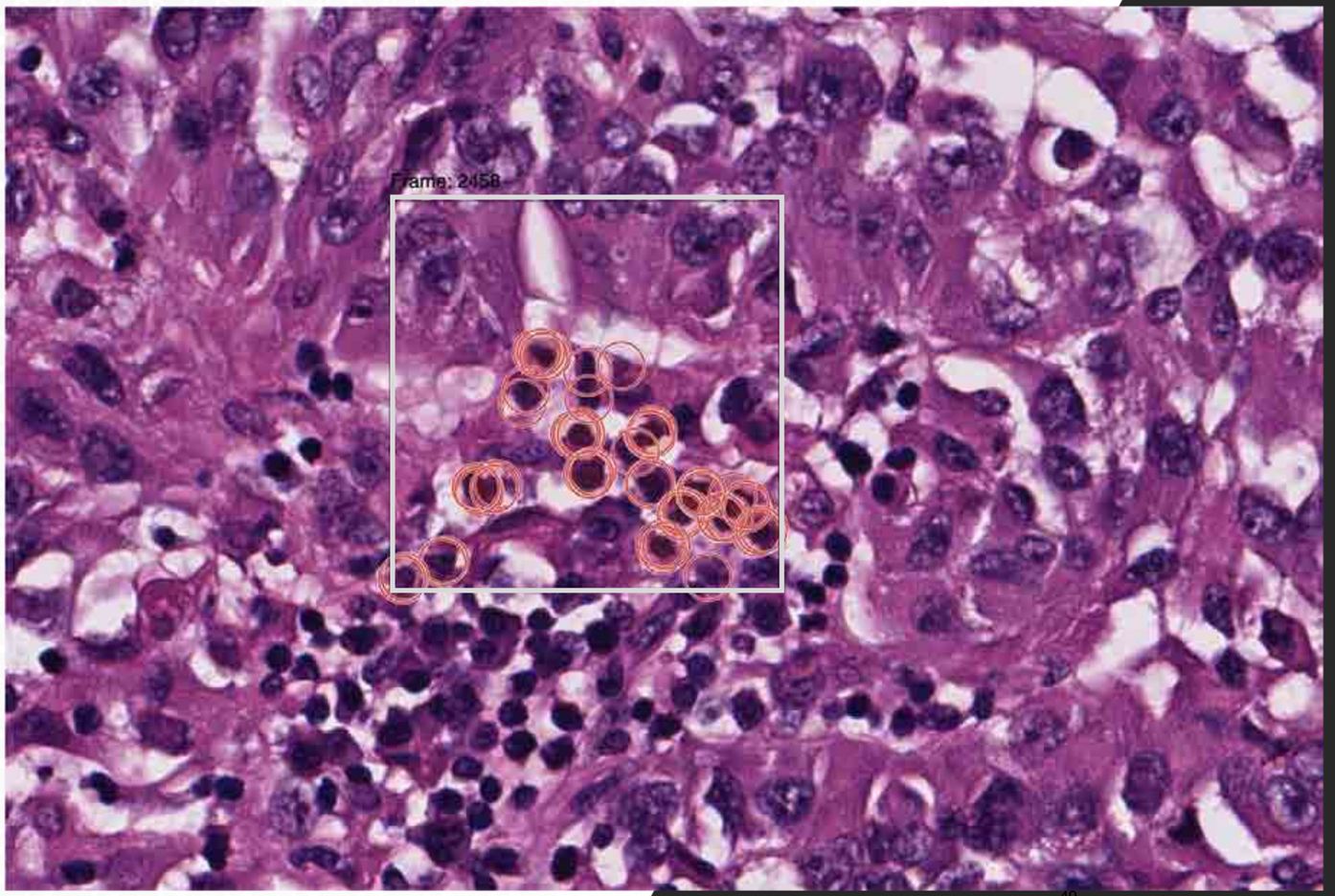
- Low *n*, interpretability and measures of uncertainty valuable:
  - No deep learning (gasp!)
- Feature importance/selection in these models can provide disease insight
  - *Now we're doing things pathologists can't rather than automating / improving what they already can*





Note: KM curves for illustration only

How do we know these features are *correct?*

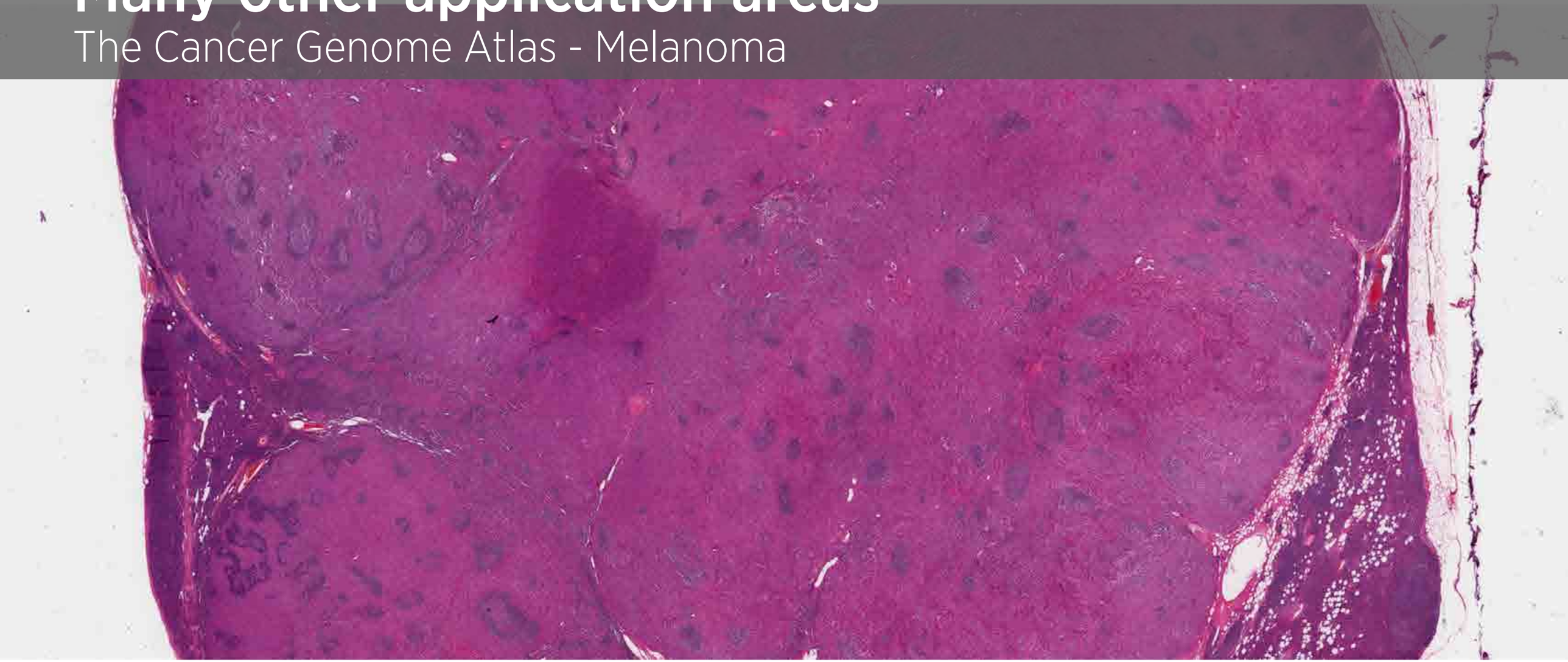


Frames

Validation by exhaustive consensus

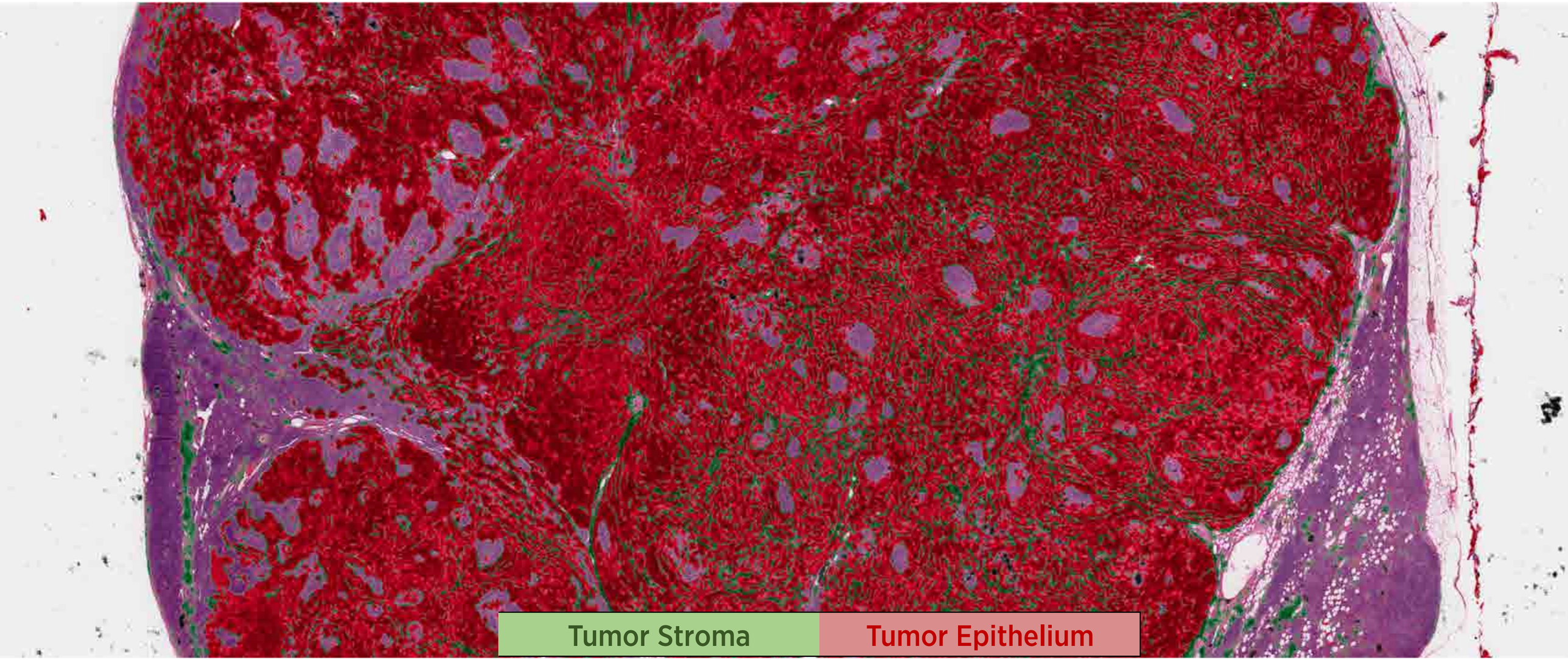
Many other application areas

The Cancer Genome Atlas - Melanoma



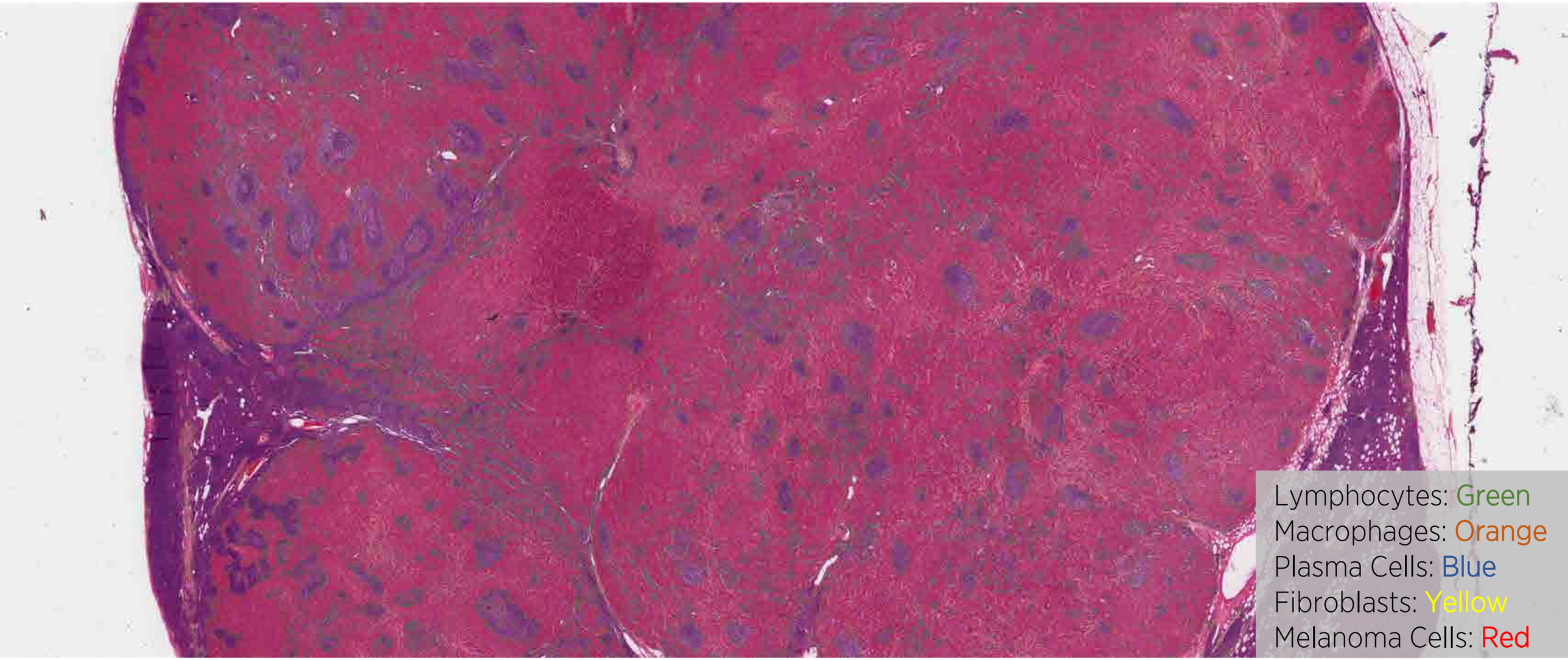
TCGA-EE-A2GL, Malignant Melanoma

Melanoma Tissue Map



TCGA-EE-A2GL, Malignant Melanoma

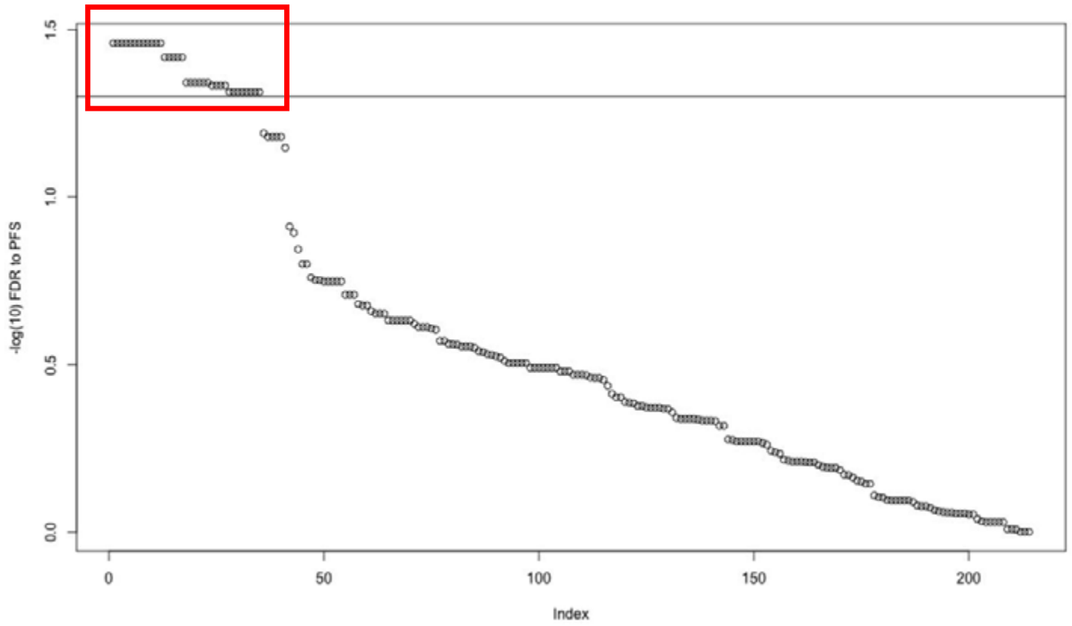
Melanoma Cell Map



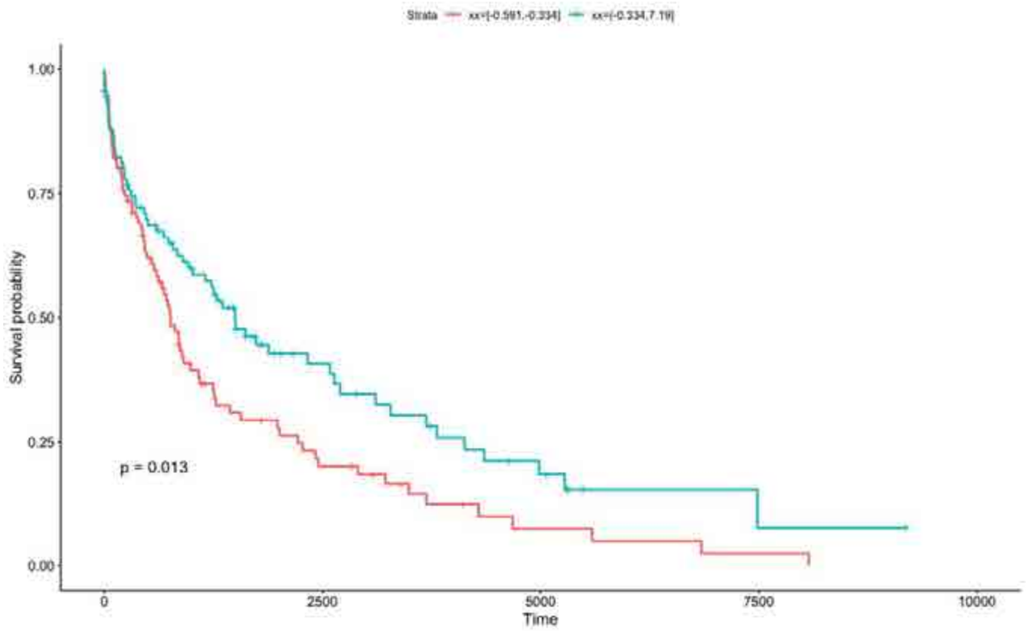
TCGA-EE-A2GL, Malignant Melanoma

Exhaustive analysis of cellular features in TCGA to enable data-driven identification of pathological predictors of survival in malignant melanoma

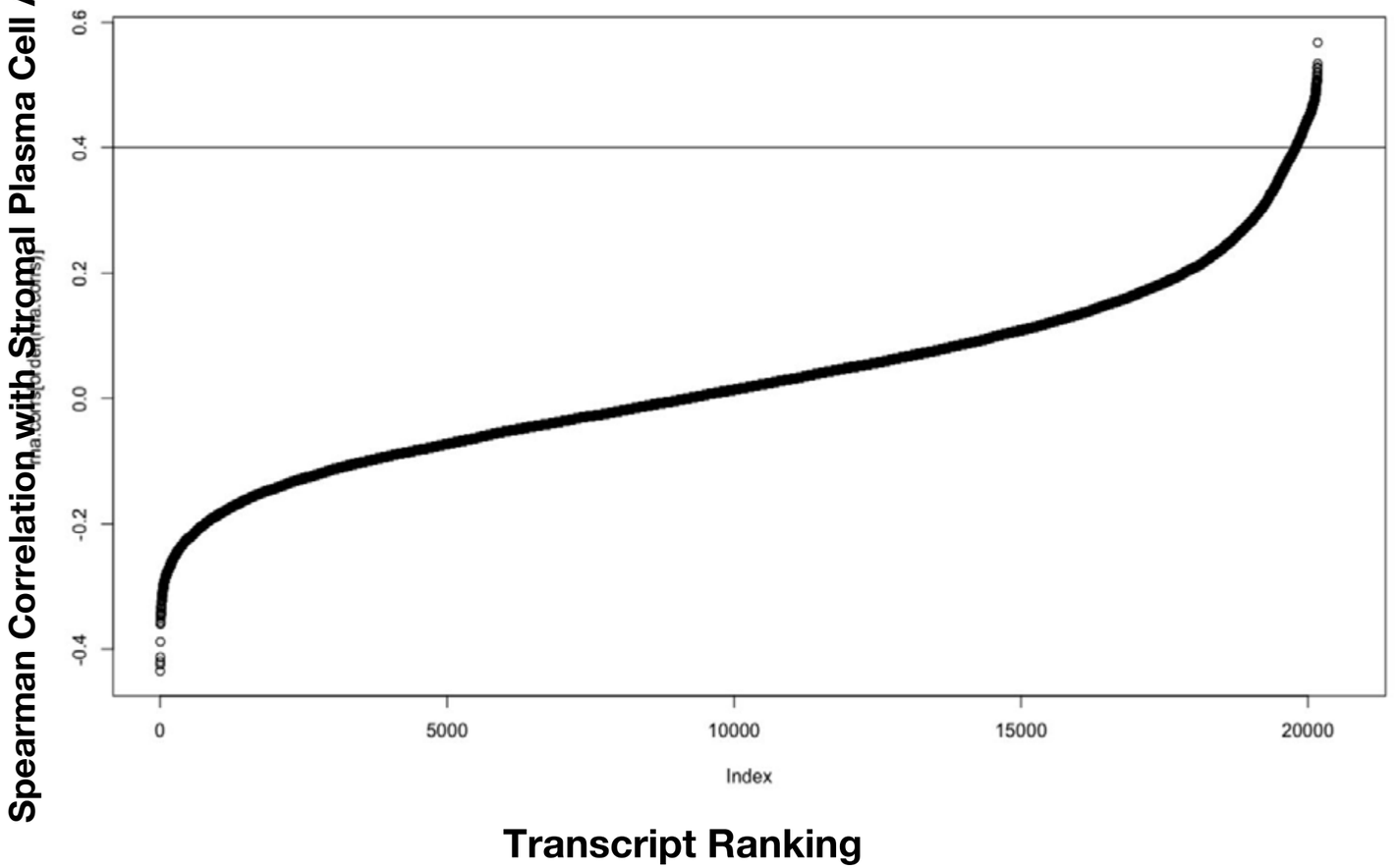
Pathological phenotypes with FDR < 5% for association with Progression Free Survival



Increased area of stromal plasma cells associated with improved survival in melanoma



Data-driven identification of transcriptional signature underlying stromal area of plasma cells in melanoma



Top-ranking transcripts associated with stromal area of plasma cells

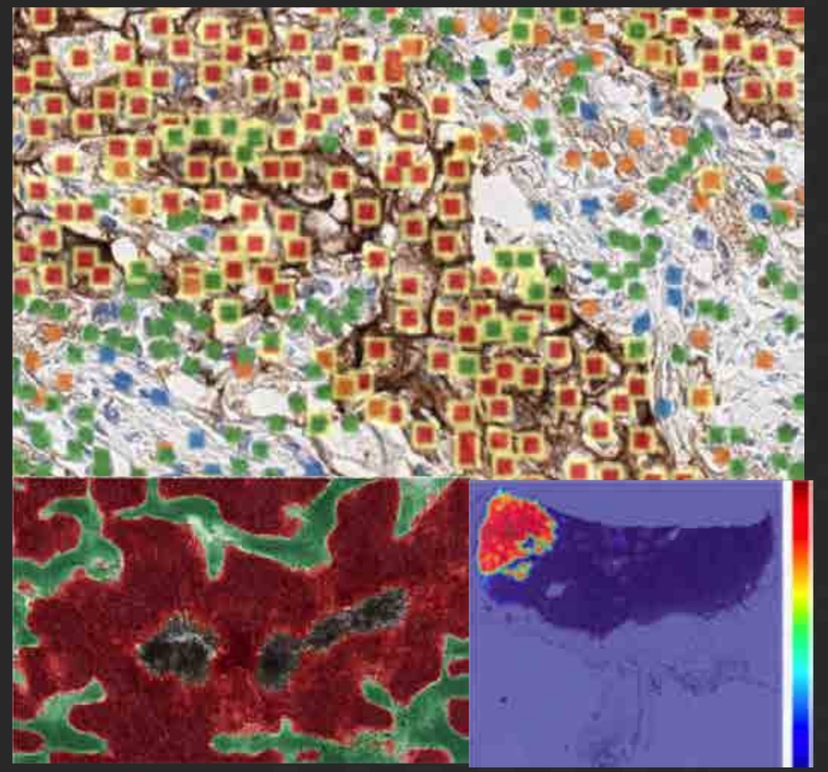
Gene	Correlation
REC8	0.57
GPR174	0.53
CD38	0.53
LAX1	0.53
TOX	0.53
AKAP5	0.53
C8orf80	0.52
JSRP1	0.52
IGJ	0.52
TNFRSF17	0.51
EAF2	0.51

Stromal plasma cell area RNA signature strongly enriched for immune genes

Gene Set Name	Description	FDR q-value
REACTOME_IMMUNE_SYSTEM	Genes involved in Immune System	7.62E-57
REACTOME_ADAPTIVE_IMMUNE_SYSTEM	Genes involved in Adaptive Immune System	6.02E-42
PID_TCR_PATHWAY	TCR signaling in naive CD4+ T cells	4.24E-30
REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	Genes involved in Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	6.07E-26
KEGG_PRIMARY_IMMUNODEFICIENCY	Primary immunodeficiency	7.98E-24
PID_IL12_2PATHWAY	IL12-mediated signaling events	9.27E-24
PID_CD8_TCR_PATHWAY	TCR signaling in naive CD8+ T cells	9.27E-24
KEGG_CELL_ADHESION_MOLECULES_CAMS	Cell adhesion molecules (CAMs)	3.00E-22
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	Cytokine-cytokine receptor interaction	6.38E-22
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	Intestinal immune network for IgA production	3.37E-21
REACTOME_TCR_SIGNALING	Genes involved in TCR signaling	3.24E-20
REACTOME_PD1_SIGNALING	Genes involved in PD-1 signaling	3.44E-19
REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY	Genes involved in Costimulation by the CD28 family	5.48E-19

Another AI plus: scalability

- Same pipeline for any solid tumor type
  - Contrast to traditional approach: hand-crafted algorithms.



PathAI for Immuno-oncology

PathAI platform has been applied to:

- Non-small cell lung cancer (Adenocarcinoma)
- Non-small cell lung cancer (Squamous Cell Carcinoma)
- Small cell carcinoma of the lung
- Urothelial carcinoma of the bladder
- Head and neck squamous cell carcinoma
- Melanoma
- Breast cancer
- Prostate cancer
- Colon cancer

>30 IO-IHC biomarkers studied

IHC images processed
10,000+

Number of Annotations
2.5 Million+

PDL1 IHC cells classified
1 Billion+

In 2018, PathAI classified ~15x the number of cells that all US pathologists could perform in a year

PathAI

Extensive Slide Search & Data Standardization

Slides Search

Filter Images
Choose criteria

TCGA

TCGA

Any case

Any stain

Any group

Original file name

Overlays:
Yes No Either

Annotations:
Yes No Either

30872 matching images Clear filters

Case TCGA-OR-A5J1
Frozen

Case TCGA-OR-A5J1
H & E

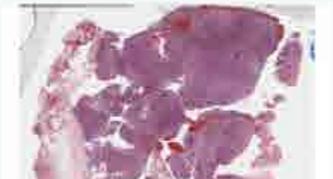
Case TCGA-OR-A5J1
H & E

Case TCGA-OR-A5J1
H & E

Case TCGA-OR-A5J1
H & E

Case TCGA-OR-A5J1
H & E

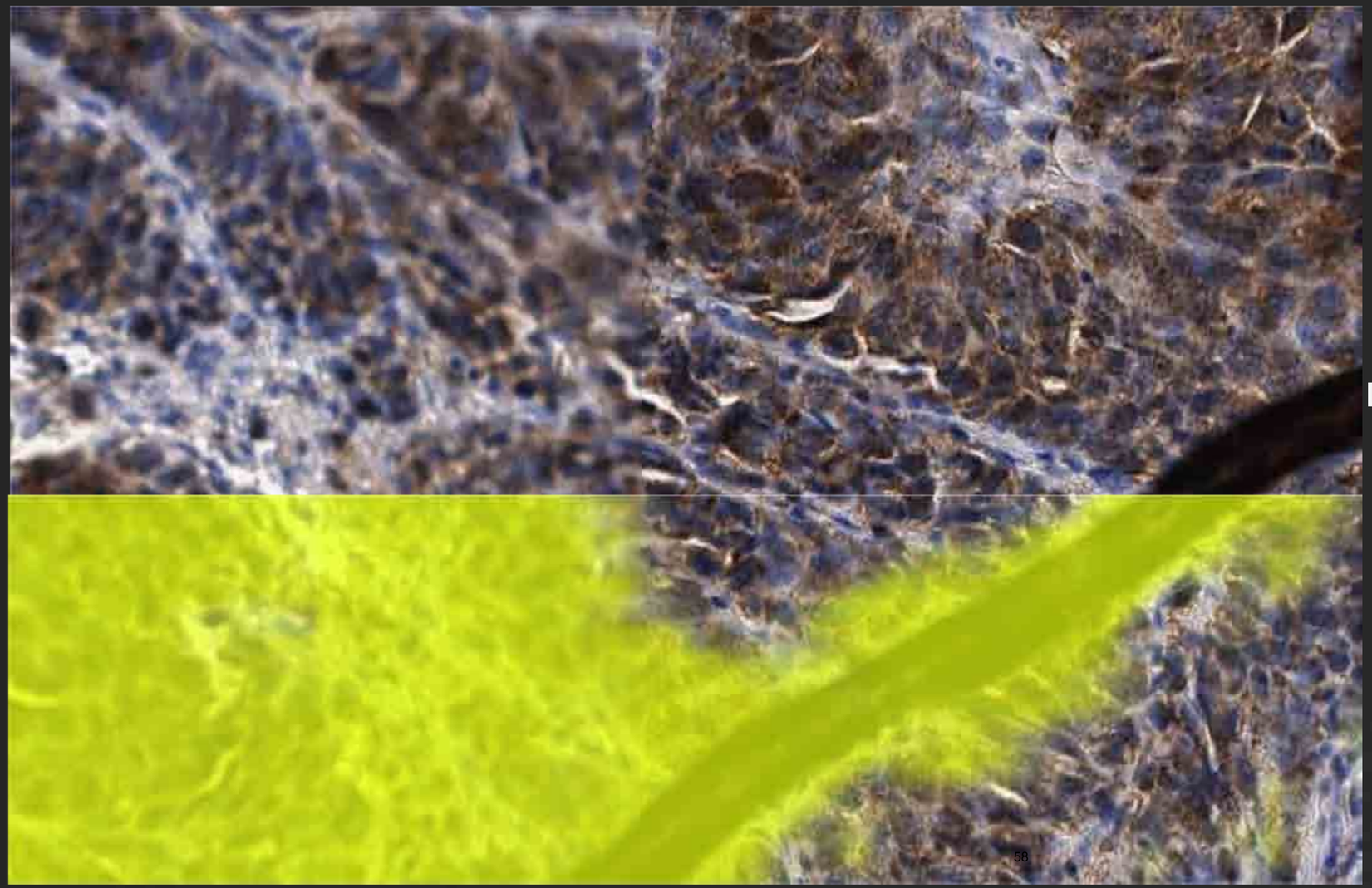
Case TCGA-OR-A5J2
Frozen



Case TCGA-OR-A5J2
H & E

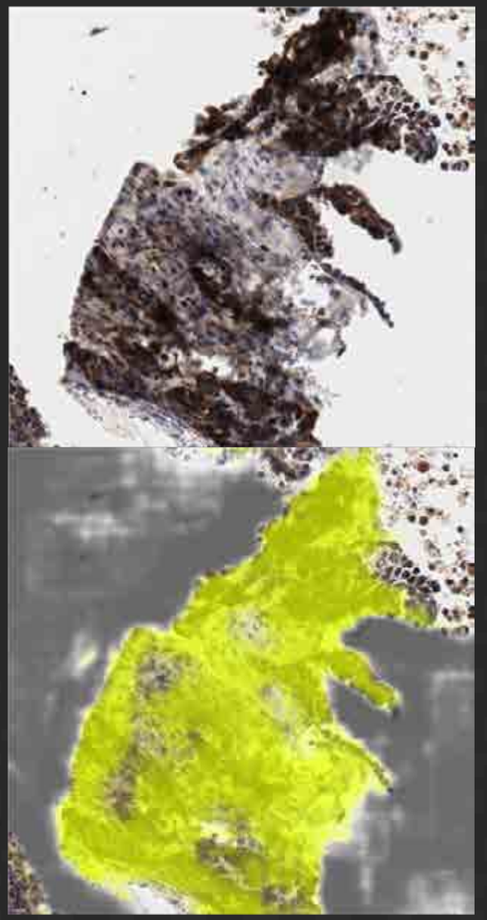
Automated quality control

Blurred areas



Debris

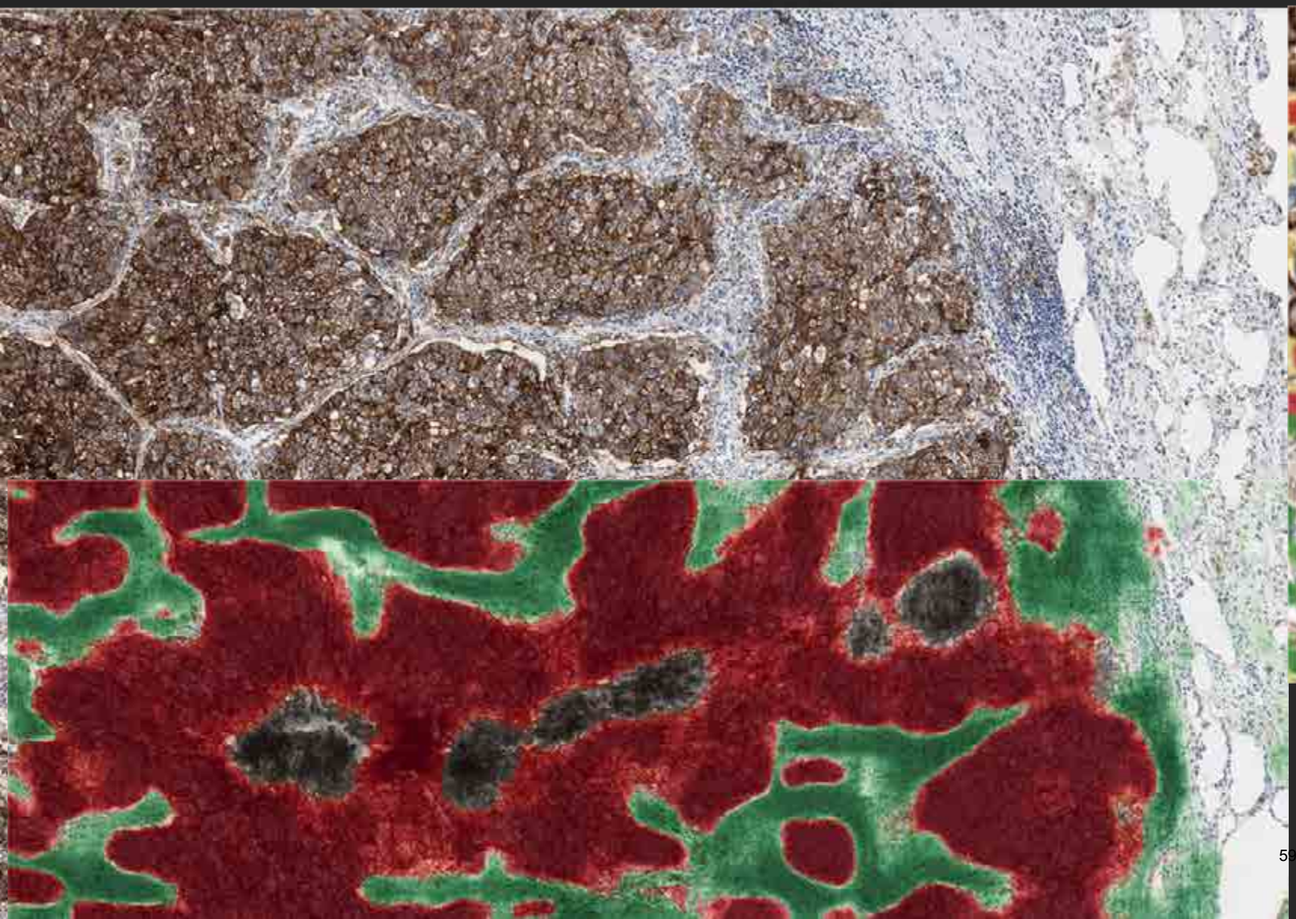
Folded / damaged tissue



PathAI

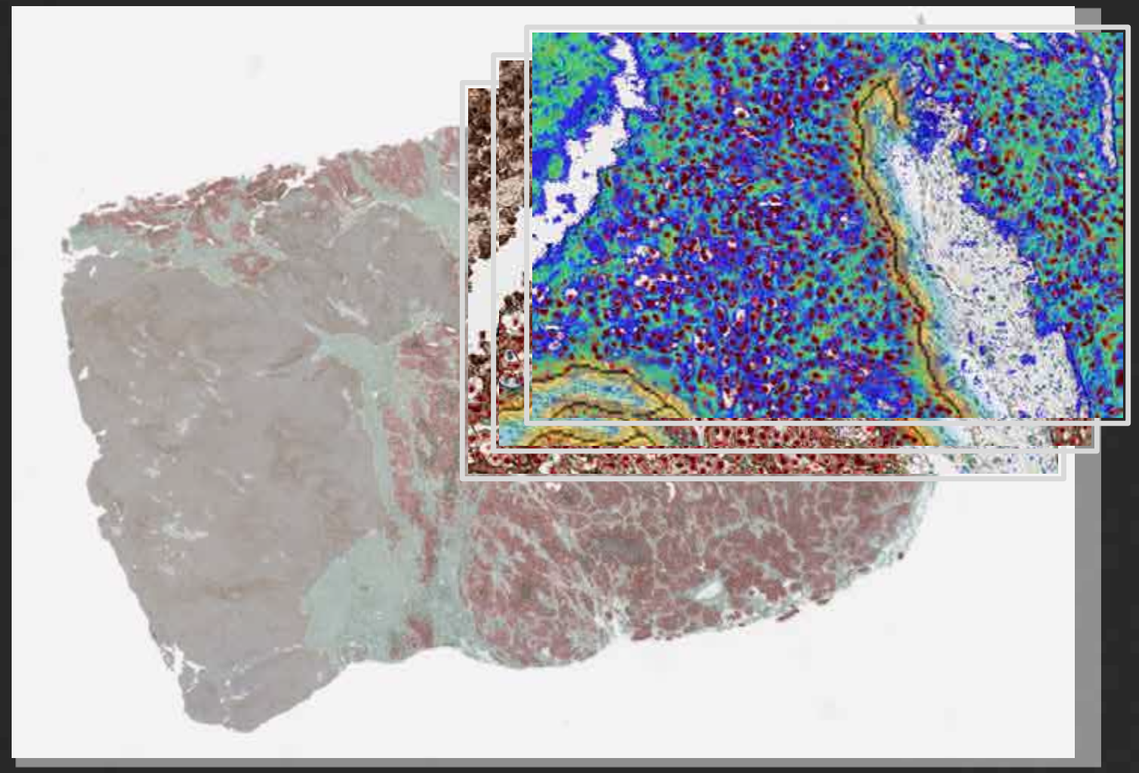
Annotate, train and deploy task-specific models

- Determined by partner needs

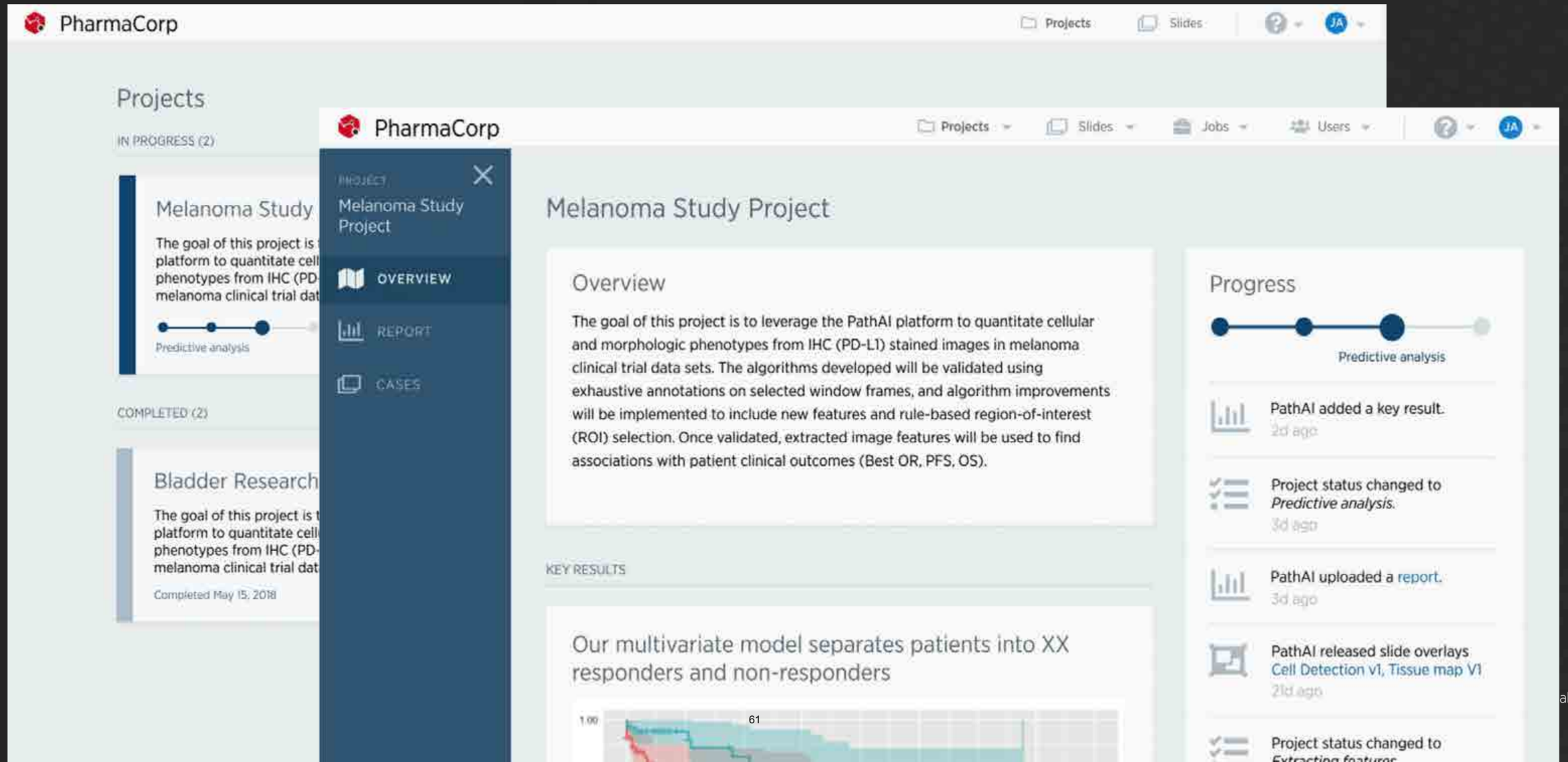


Interpretable feature extraction

- Hypothesis & data driven



Interactive Reports & Live Project Progress



The PathAI Deep Learning Process

Whole-Slide Images + Data

Transmit training data securely to the PathAI cloud

Annotations

Network of board-certified pathologists to provide ground truth consensus

Deep Learning Analysis

Cell detection, tissue & region classification

Deep Learning Feature Analysis

Over 200 relevant features extracted, measured and analyzed

Assay Validated

Identified features of significance reduced to practice

Assay Deployed

Analyze samples, quantified & visual results delivered

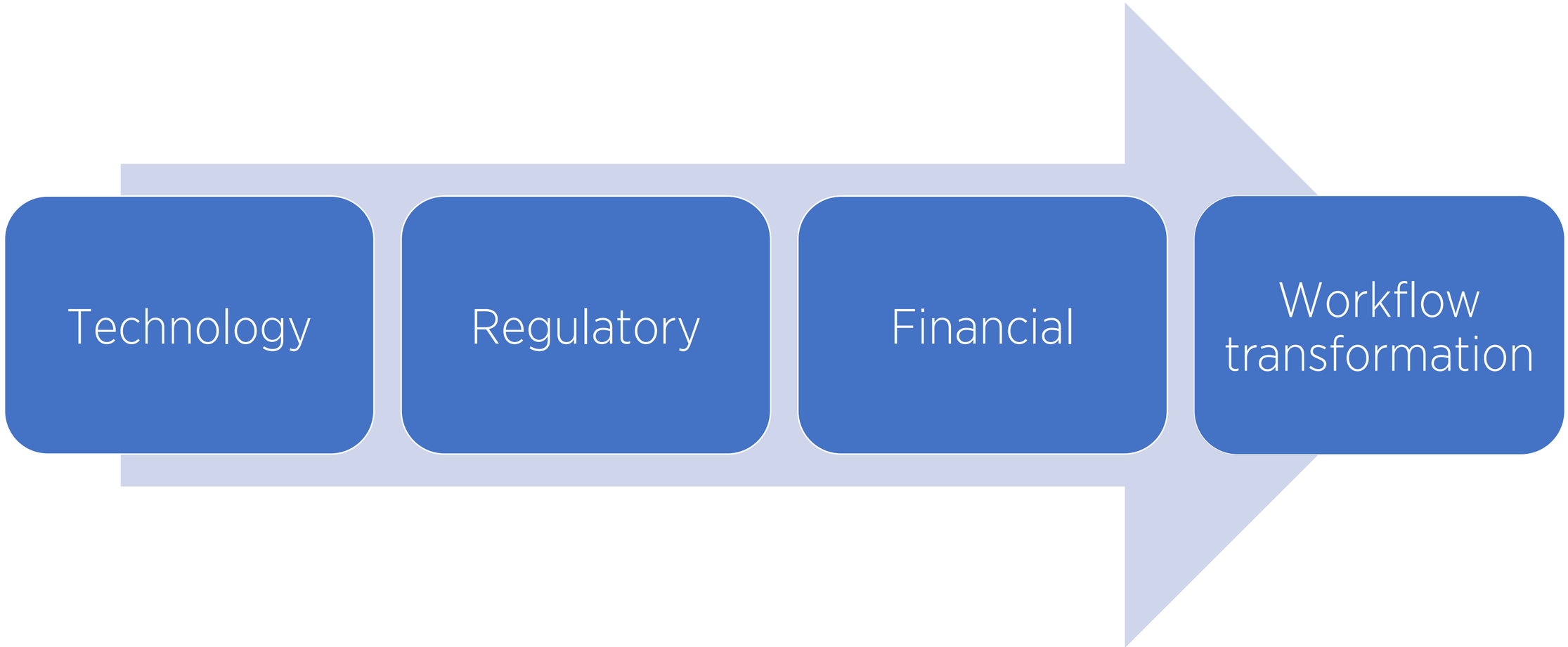
We can execute process in 4 – 8 weeks for new assays

PathAI

AI in medicine
Some closing thoughts

- ML in the real world:
  - Building the right dataset is 75% of the challenge
- Modern ML: engineering and empirical science
  - Rigorous validation is key
- Ideas and algorithms vs. teams and infrastructure

Core challenges and road ahead



Key Takeaways

- Researchers have been working on AI for pathology for ~30 years
- In the past 5 years, advances in:
  - Availability of digital data
  - Access to large-scale computing resources
  - Major algorithmic advances (e.g., Deep CNNs)
- AI works extremely well when these 3 factors are all available and fails when they are not

Key Takeaways

- AI-powered pathology is broadly applicable across all image-based tasks in pathology and enables integration with other structured data types (e.g., ‘Omics)
- As AI and digital pathology are incorporated into clinical workflow, they will offer significant operational and efficiency advantages
- AI will drive improvements in the accuracy and predictiveness of pathology leading to research advances and improved care for patients

“In the Future…” (1987)

- “Integrated information systems, patient care management by exception, decision support tools, and, in the future, "artificial intelligence" assists can all be expected to become staples of pathology practice, especially impacting those pathologists who choose to be responsive to the new practice milieu of medical information science.”

“Using the computer to optimize human performance in health care delivery. The pathologist as medical information specialist.”
(Arch Pathol Lab Med. 1987)

6.S897 / HST.956 Machine Learning for Healthcare
Spring 2019