Mahmood Lab
AI for Pathology

@AI4Pathology

Multimodal, Generative, and Agentic AI for Pathology

Faisal Mahmood, Ph.D.

Associate Professor, Harvard Medical School
Department of Pathology, BWH and MGH
Cancer Data Science Program, Dana Farber Cancer Center
Broad Institute of Harvard and MIT

faisalmahmood@bwh.harvard.edu

www.mahmoodlab.org

HARVARD MEDICAL SCHOOL · BRIGHAM AND WOMEN'S HOSPITAL · Dana-Farber Cancer Institute · BROAD INSTITUTE

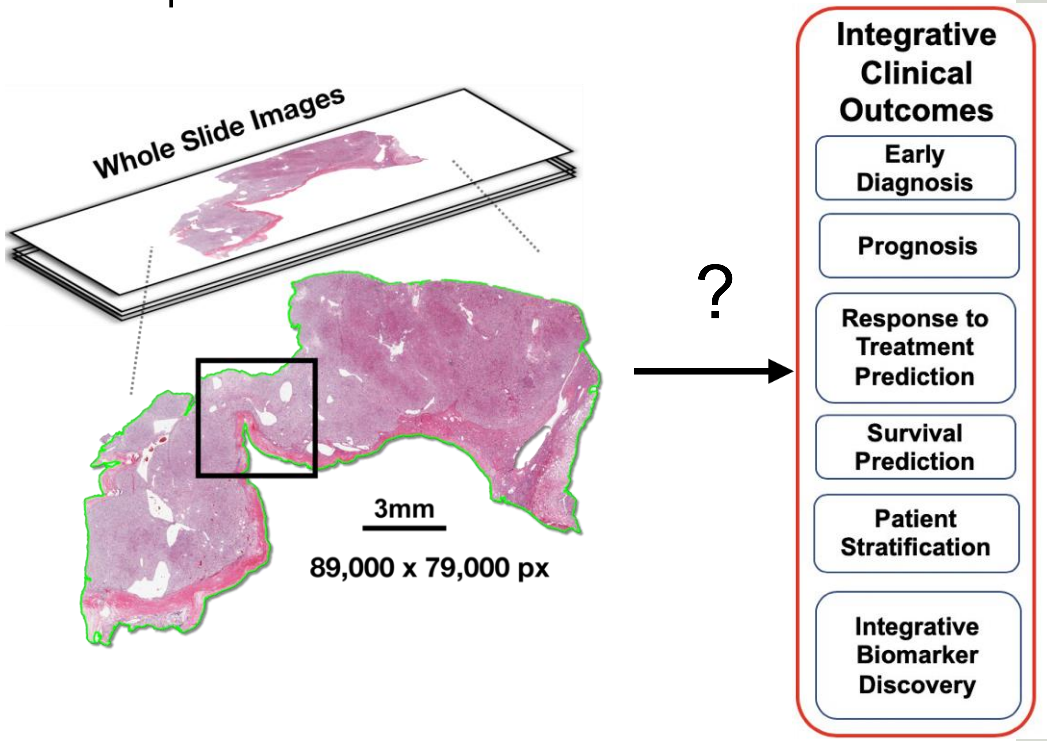
Outline

- **Weakly Supervised Models for Pathology**
    - CLAM (Nature BME, 2021)
    - Cancers of Unknown Primary (Nature, 2021)
    - Cardiac Allograft Rejection (Nature Medicine, 2022)
- **Multimodal Data Integration**
    - Pan-cancer, fusing histology and genomics (Cancer Cell, 2022)
- **Foundation Models**
    - Vision centric foundation model (Nature Medicine, 2024)
    - Vision-language foundation model (Nature Medicine, 2024)
- **Generative AI for Pathology**
    - PathChat (Nature, 2024)
- **Transitioning from 2D to 3D Pathology**
    - TriPath (Cell, 2024)
- **Bias and Fairness**
    - Do foundation models reduce model bias? (Nature Medicine, 2024)

Problem Formulation

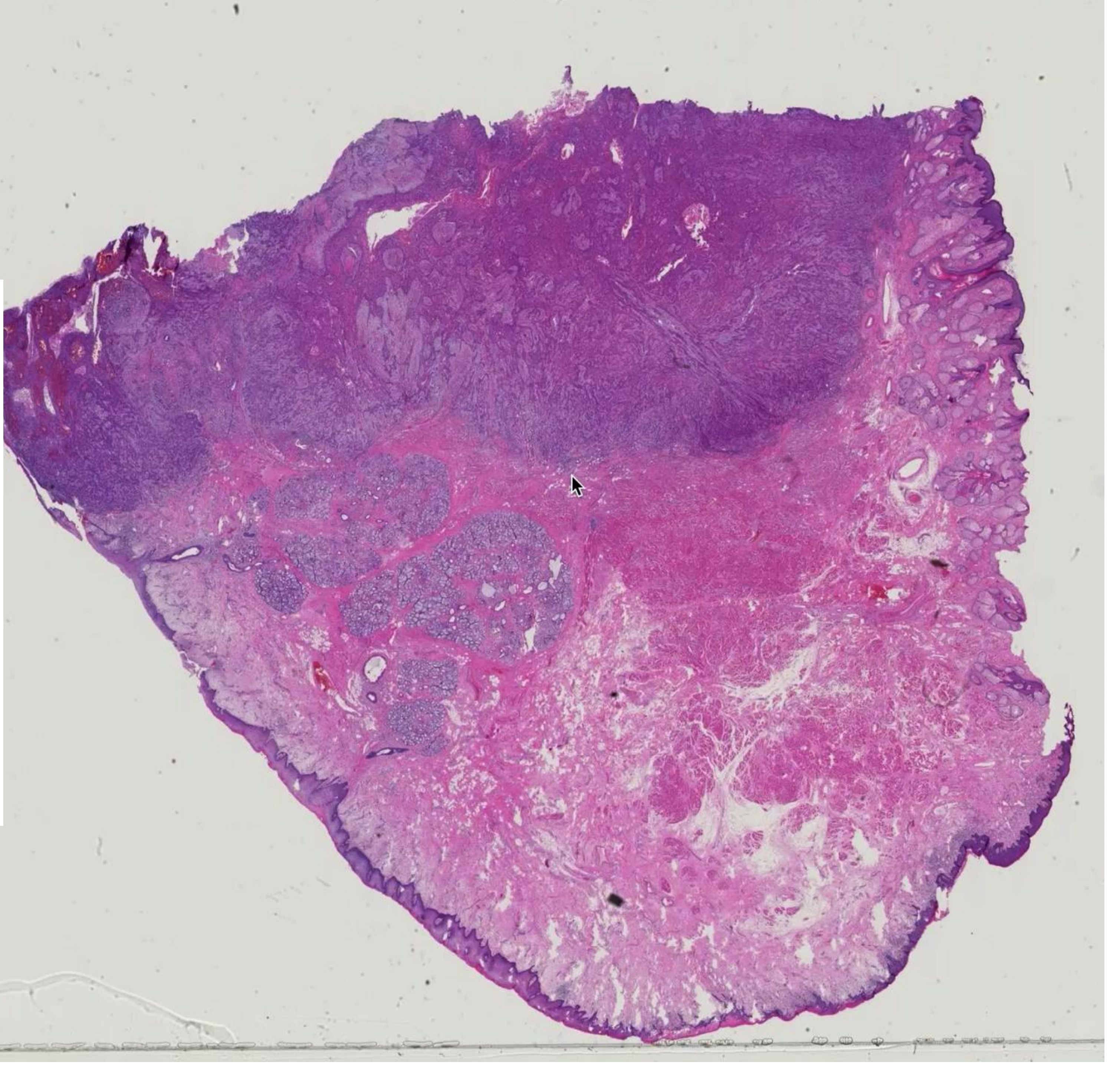
Slide-Level Task: Given ~150K × 150K image (*e.g.* – Whole-Slide Image or WSI), predict:

- Cancer stage / subtype
- Survival outcome
- Response-to-treatment



Integrative Clinical Outcomes

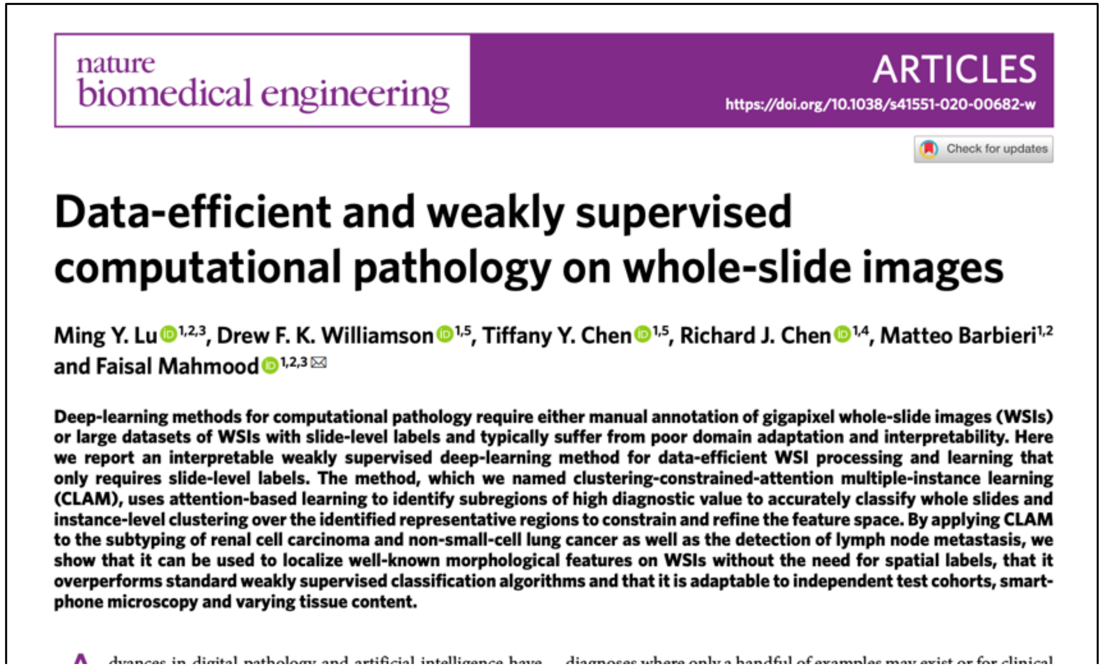
- Early Diagnosis
- Prognosis
- Response to Treatment Prediction
- Survival Prediction
- Patient Stratification
- Integrative Biomarker Discovery



CLAM Workflow

CLAM

- Weakly supervised learning from histology whole slide images.
- Adapts Attention Based Multiple Instance Learning for Computational Pathology.
- Used pre-trained feature encoders instead of end-to-end training.
- Easy to use codebase.



nature biomedical engineering

ARTICLES

https://doi.org/10.1038/s41551-020-00682-w

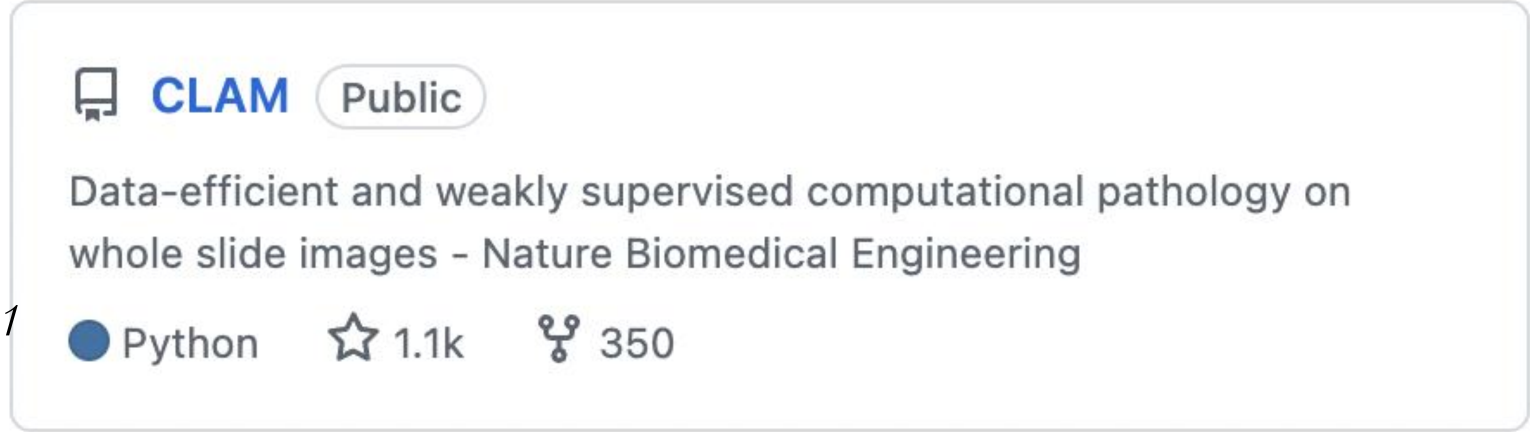
Check for updates

Data-efficient and weakly supervised computational pathology on whole-slide images

Ming Y. Lu[1,2,3], Drew F. K. Williamson[1,5], Tiffany Y. Chen[1,5], Richard J. Chen[1,4], Matteo Barbieri[1,2] and Faisal Mahmood[1,2,3] ✉

Deep-learning methods for computational pathology require either manual annotation of gigapixel whole-slide images (WSIs) or large datasets of WSIs with slide-level labels and typically suffer from poor domain adaptation and interpretability. Here we report an interpretable weakly supervised deep-learning method for data-efficient WSI processing and learning that only requires slide-level labels. The method, which we named clustering-constrained-attention multiple-instance learning (CLAM), uses attention-based learning to identify subregions of high diagnostic value to accurately classify whole slides and instance-level clustering over the identified representative regions to constrain and refine the feature space. By applying CLAM to the subtyping of renal cell carcinoma and non-small-cell lung cancer as well as the detection of lymph node metastasis, we show that it can be used to localize well-known morphological features on WSIs without the need for spatial labels, that it overperforms standard weakly supervised classification algorithms and that it is adaptable to independent test cohorts, smart-phone microscopy and varying tissue content.

Advances in digital pathology and artificial intelligence have diagnoses where only a handful of examples may exist or for clinical



CLAM Public

Data-efficient and weakly supervised computational pathology on whole slide images - Nature Biomedical Engineering

● Python ☆ 1.1k 350

(Nature Biomedical Engineering, 2021

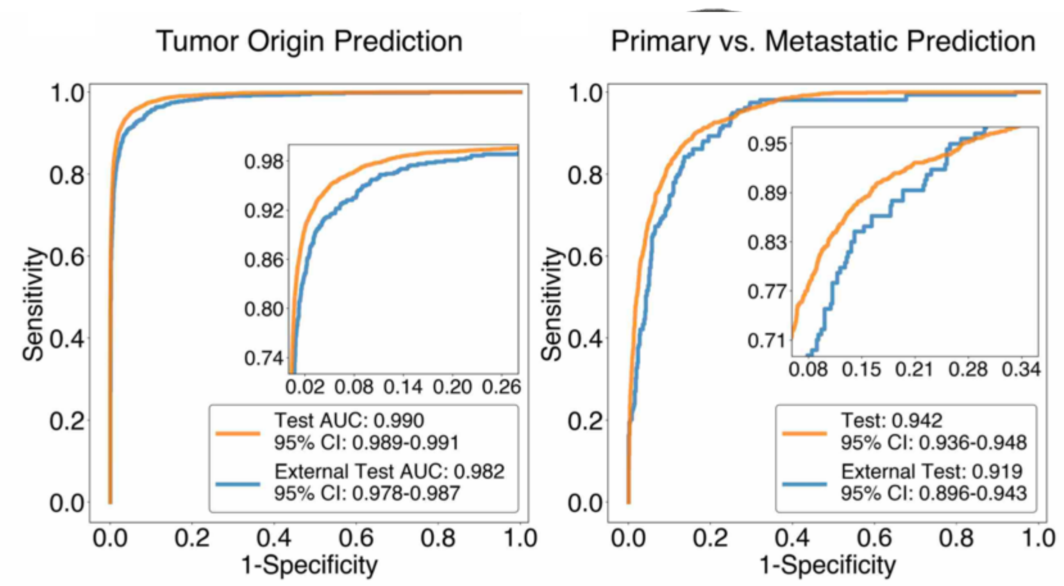
Mahmood Lab
AI for Pathology

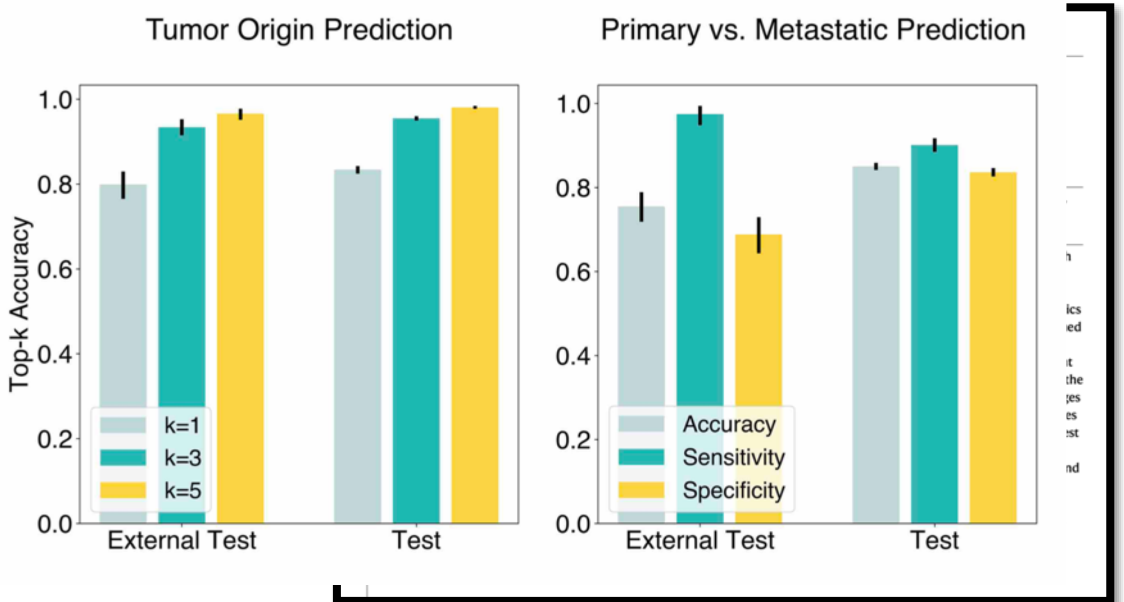
Cancers of Unknown Pri

Cancers where a primary origin can not be determined.

- 1-2% of all cancers.
- **30,270** cases expected to be diagnosed in the
- Median survival **2.7-16 months**.
- 2-year survival rate: **20-25%**
- CUP patients undergo a complete workup of clinical, radiological, endoscopy, molecular tes an attempt to determine origin.

Can we use H&E whole slides to deter origins for cancers of unknown primary



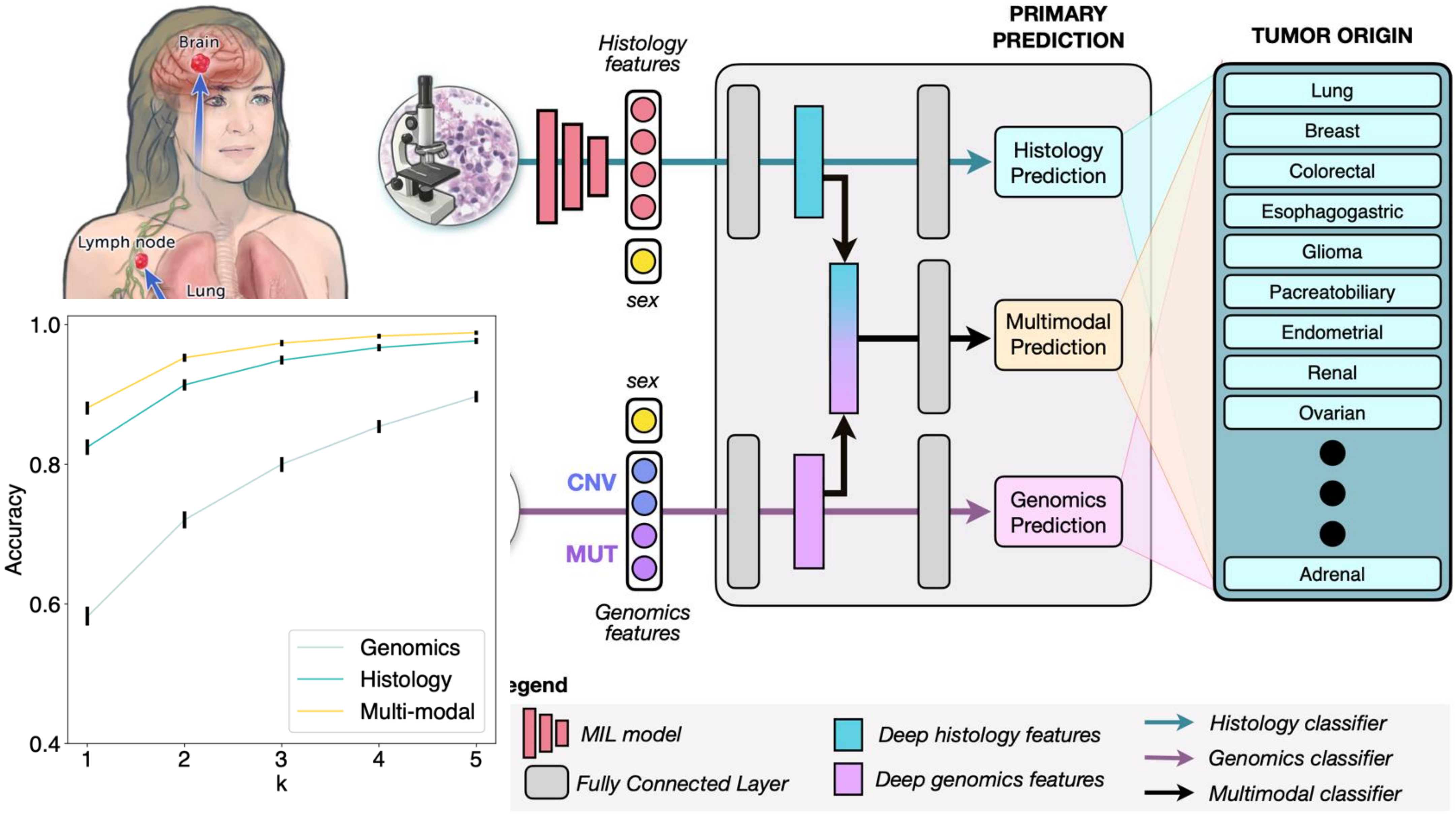


Mahmood Lab
AI for Pathology *(Nature, 2021)*

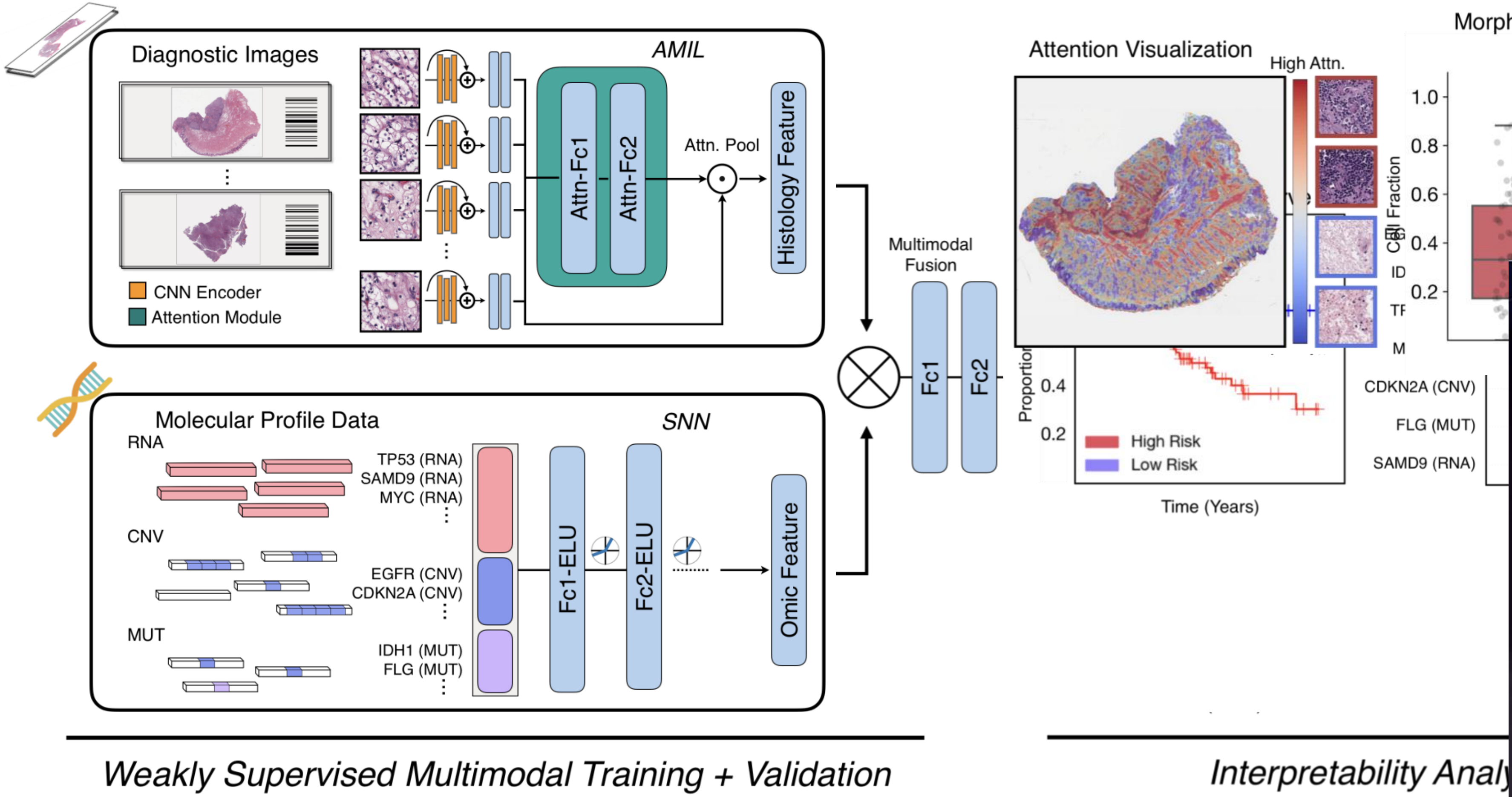
HARVARD MEDICAL SCHOOL BRIGHAM AND WOMEN'S HOSPITAL Dana-Farber Cancer Institute

(Nature BME, 2024 – in press)

Integrating histology + genomics for origin prediction



PORPOISE: Overview (http://pancancancer.mahmoodlab.org)



Network Architecture

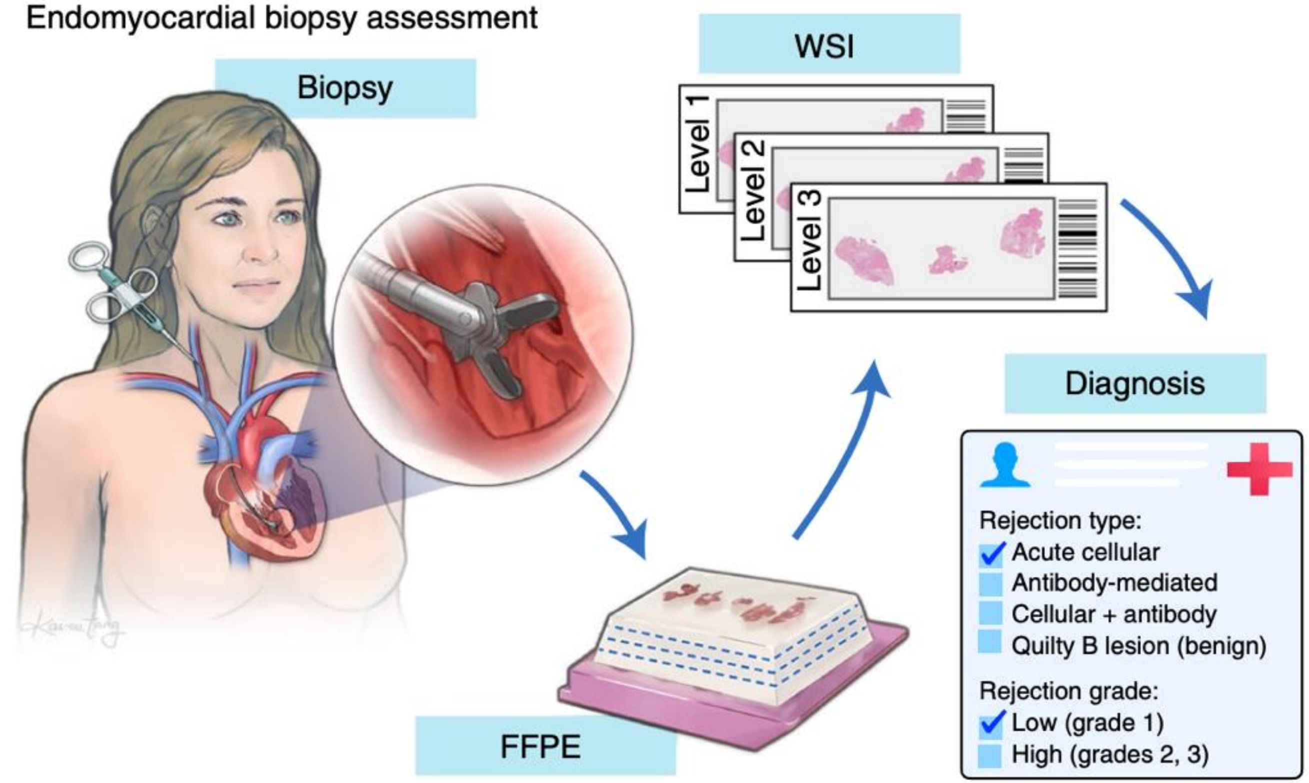
- Unimodal branch for WSIs using CLAM / ABMIL
- Unimodal branch for Mut+CNV+RNA using SNN
- Multimodal Fusion via Kronecker Product

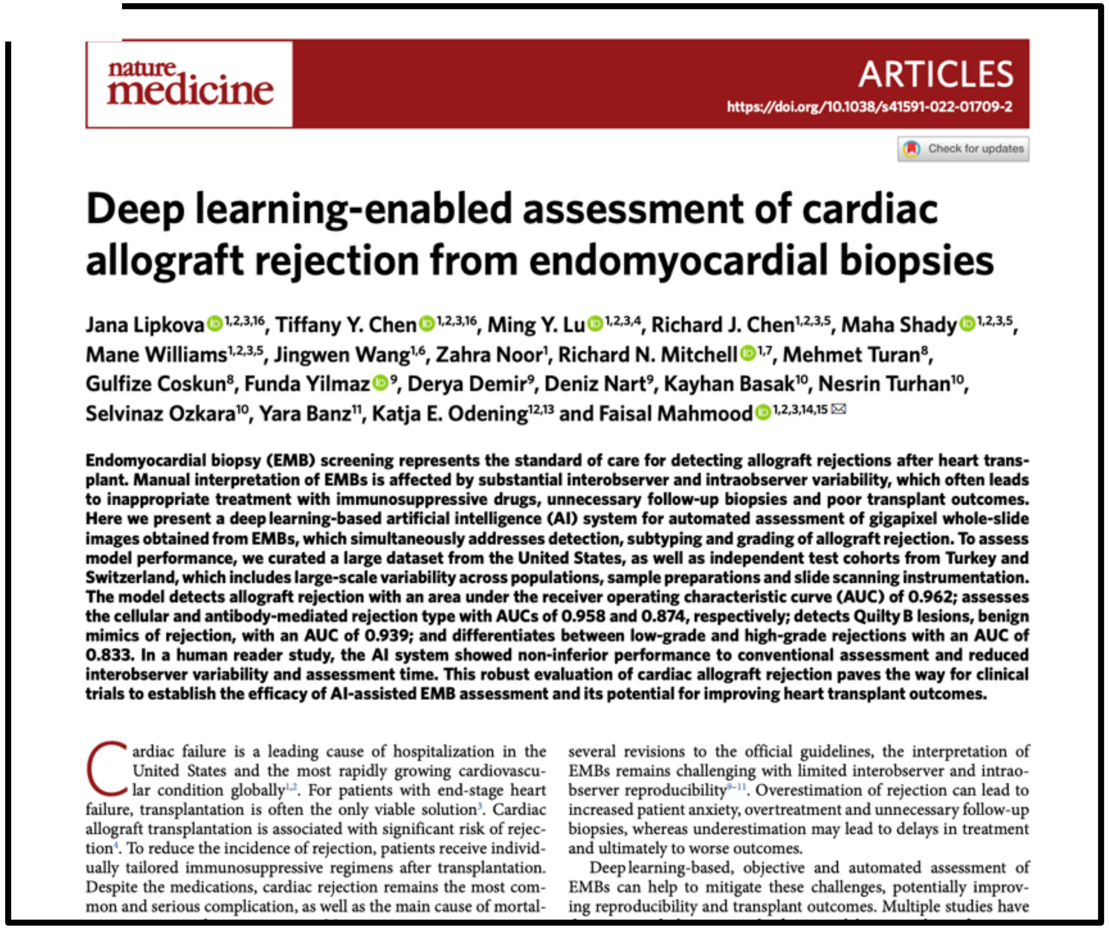
Interpretability Strategy

- Integrated Gradients fo
- Attention Weights + HIF f

Chen *et al.*, Cancer Cell 2022

Endomyocardial Biopsy Assessment




nature medicine

ARTICLES

https://doi.org/10.1038/s41591-022-01709-2

Check for updates

Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies

Jana Lipkova[1,2,3,16], Tiffany Y. Chen[1,2,3,16], Ming Y. Lu[1,2,3,4], Richard J. Chen[1,2,3,5], Maha Shady[1,2,3,5], Mane Williams[1,2,3,5], Jingwen Wang[1,6], Zahra Noor[1], Richard N. Mitchell[1,7], Mehmet Turan[8], Gulfize Coskun[8], Funda Yilmaz[9], Derya Demir[9], Deniz Nart[9], Kayhan Basak[10], Nesrin Turhan[10], Selvinaz Ozkara[10], Yara Banz[11], Katja E. Odening[12,13] and Faisal Mahmood[1,2,3,14,15] ✉

Endomyocardial biopsy (EMB) screening represents the standard of care for detecting allograft rejections after heart transplant. Manual interpretation of EMBs is affected by substantial interobserver and intraobserver variability, which often leads to inappropriate treatment with immunosuppressive drugs, unnecessary follow-up biopsies and poor transplant outcomes. Here we present a deep learning-based artificial intelligence (AI) system for automated assessment of gigapixel whole-slide images obtained from EMBs, which simultaneously addresses detection, subtyping and grading of allograft rejection. To assess model performance, we curated a large dataset from the United States, as well as independent test cohorts from Turkey and Switzerland, which includes large-scale variability across populations, sample preparations and slide scanning instrumentation. The model detects allograft rejection with an area under the receiver operating characteristic curve (AUC) of 0.962; assesses the cellular and antibody-mediated rejection type with AUCs of 0.958 and 0.874, respectively; detects Quilty B lesions, benign mimics of rejection, with an AUC of 0.939; and differentiates between low-grade and high-grade rejections with an AUC of 0.833. In a human reader study, the AI system showed non-inferior performance to conventional assessment and reduced interobserver variability and assessment time. This robust evaluation of cardiac allograft rejection paves the way for clinical trials to establish the efficacy of AI-assisted EMB assessment and its potential for improving heart transplant outcomes.

Cardiac failure is a leading cause of hospitalization in the United States and the most rapidly growing cardiovascular condition globally[1,2]. For patients with end-stage heart failure, transplantation is often the only viable solution[3]. Cardiac allograft transplantation is associated with significant risk of rejection[4]. To reduce the incidence of rejection, patients receive individually tailored immunosuppressive regimens after transplantation. Despite the medications, cardiac rejection remains the most common and serious complication, as well as the main cause of mortal-

several revisions to the official guidelines, the interpretation of EMBs remains challenging with limited interobserver and intraobserver reproducibility[9–11]. Overestimation of rejection can lead to increased patient anxiety, overtreatment and unnecessary follow-up biopsies, whereas underestimation may lead to delays in treatment and ultimately to worse outcomes.

Deep learning-based, objective and automated assessment of EMBs can help to mitigate these challenges, potentially improving reproducibility and transplant outcomes. Multiple studies have

(Nature Medicine, 2022)

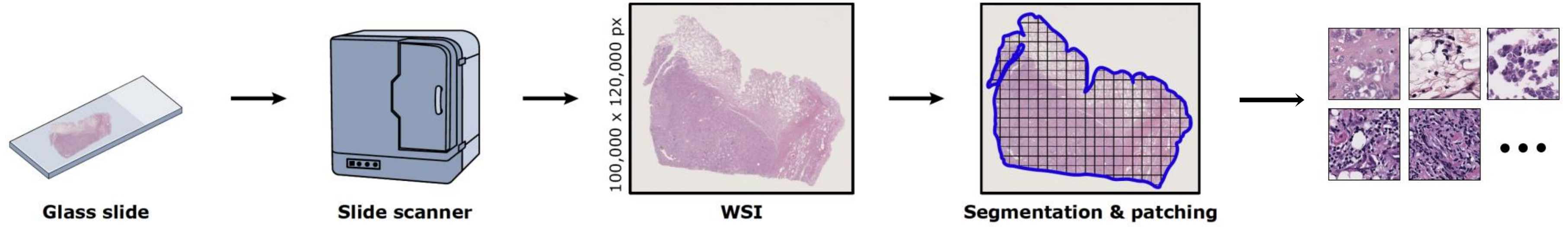
MIL Frameworks

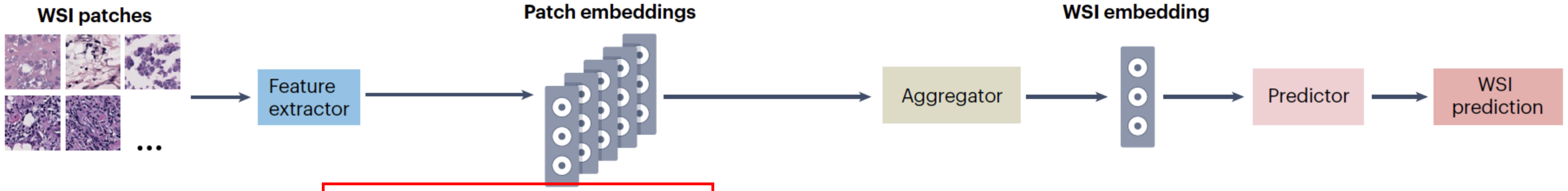
For patient j, **input**: WSI X^j **target**: clinical endpoint Y^j

- Lung cancer subtype Y^j = {Lung squamous cell carcinoma, Lung adenocarcinoma}
- Gene mutation Y^j = {wildtype, mutated}

Problem formulation

Classification $P\left(Y^j = k \middle| X^j\right) = \frac{\exp(w_k(X^j))}{\sum_{k=1}^{K} \exp(w_k(X^j))}$ ⇒ (Multinomial) Logistic regression!





Input patching

$X^j \Rightarrow \{x_1^j, \dots, x_N^j\}$

$N > 10{,}000$

$x_i^j \in \mathbb{R}^{256 \times 256 \times 3}$

⇒

Feature extraction

$z_i^j = f(x_i^j)$, for $i = 1, \dots, \mathrm{N}$

f: neural network

$z_i^j \in \mathbb{R}^{768}$ (for ViT)

⇒

Feature aggregation

$\tilde{z}^j = g(\{z_i^j\}_{i=1}^N)$

g: max / average / NN

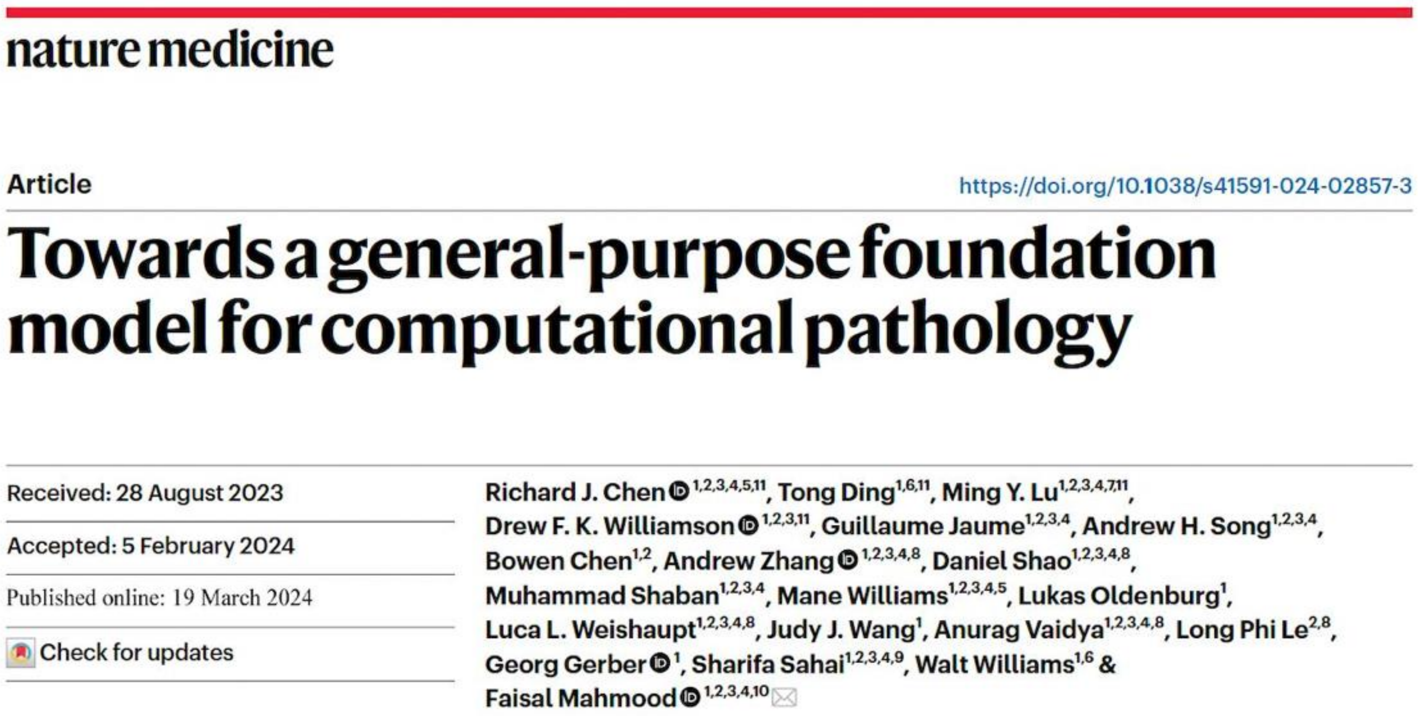
⇒

Prediction

$\hat{Y}^j = h_k(\tilde{z}^j)$

Why do we need foundation models for pathology?

- **Foundation models are generic models capable of generally encoding data into meaningful representations.**
- **Can be applied to many downstream tasks with minimal data (rare diseases, clinical trials etc.)**
- **Ideal for multi-task, multi-tissue models.**
- **Not necessarily meant to completely replace supervised, task specific models.**



nature medicine

Article

https://doi.org/10.1038/s41591-024-02857-3

Towards a general-purpose foundation model for computational pathology

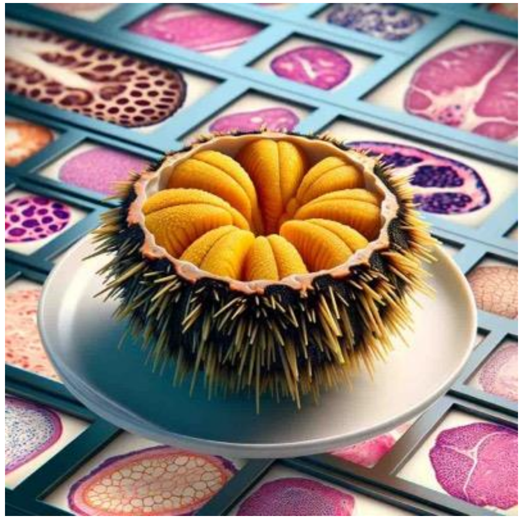
Received: 28 August 2023

Accepted: 5 February 2024

Published online: 19 March 2024

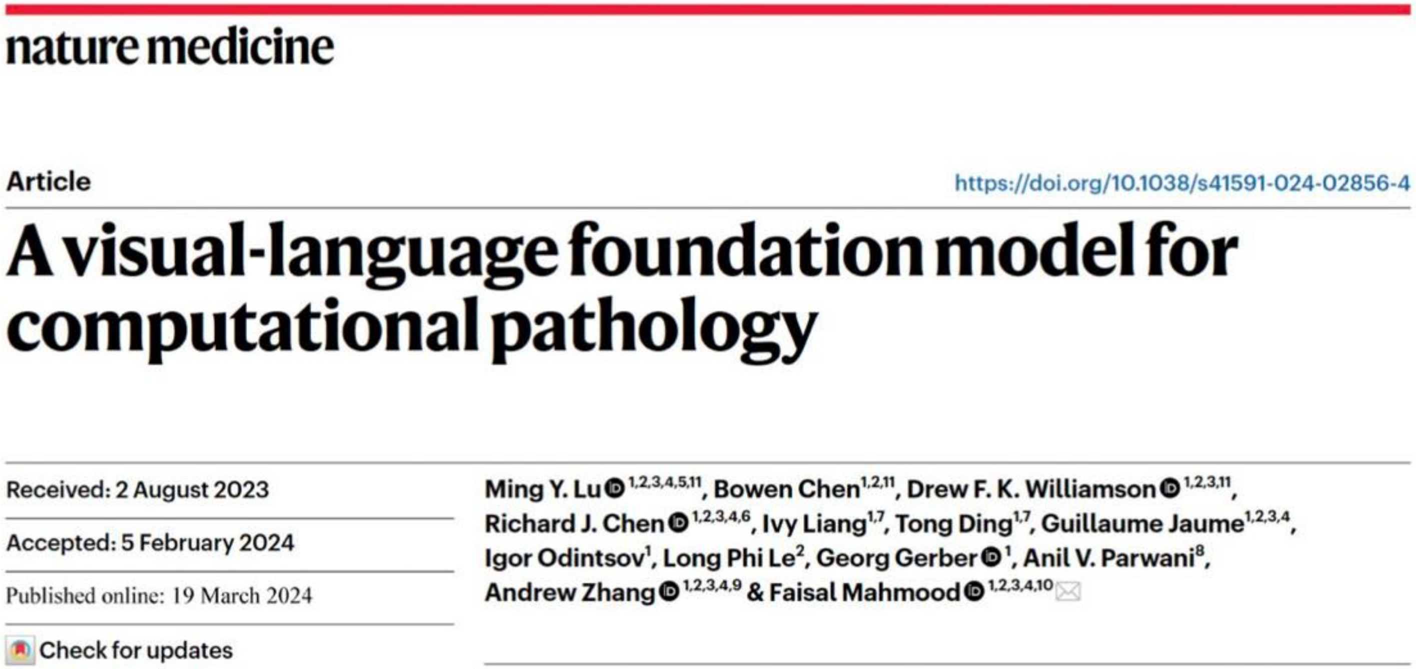
Check for updates

Richard J. Chen[1,2,3,4,5,11], Tong Ding[1,6,11], Ming Y. Lu[1,2,3,4,7,11], Drew F. K. Williamson[1,2,3,11], Guillaume Jaume[1,2,3,4], Andrew H. Song[1,2,3,4], Bowen Chen[1,2], Andrew Zhang[1,2,3,4,8], Daniel Shao[1,2,3,4,8], Muhammad Shaban[1,2,3,4], Mane Williams[1,2,3,4,5], Lukas Oldenburg[1], Luca L. Weishaupt[1,2,3,4,8], Judy J. Wang[1], Anurag Vaidya[1,2,3,4,8], Long Phi Le[2,8], Georg Gerber[1], Sharifa Sahai[1,2,3,4,9], Walt Williams[1,6] & Faisal Mahmood[1,2,3,4,10] ✉



UNI

http://github.com/mahmoodlab/UNI



nature medicine

Article

https://doi.org/10.1038/s41591-024-02856-4

A visual-language foundation model for computational pathology

Received: 2 August 2023

Accepted: 5 February 2024

Published online: 19 March 2024

Check for updates

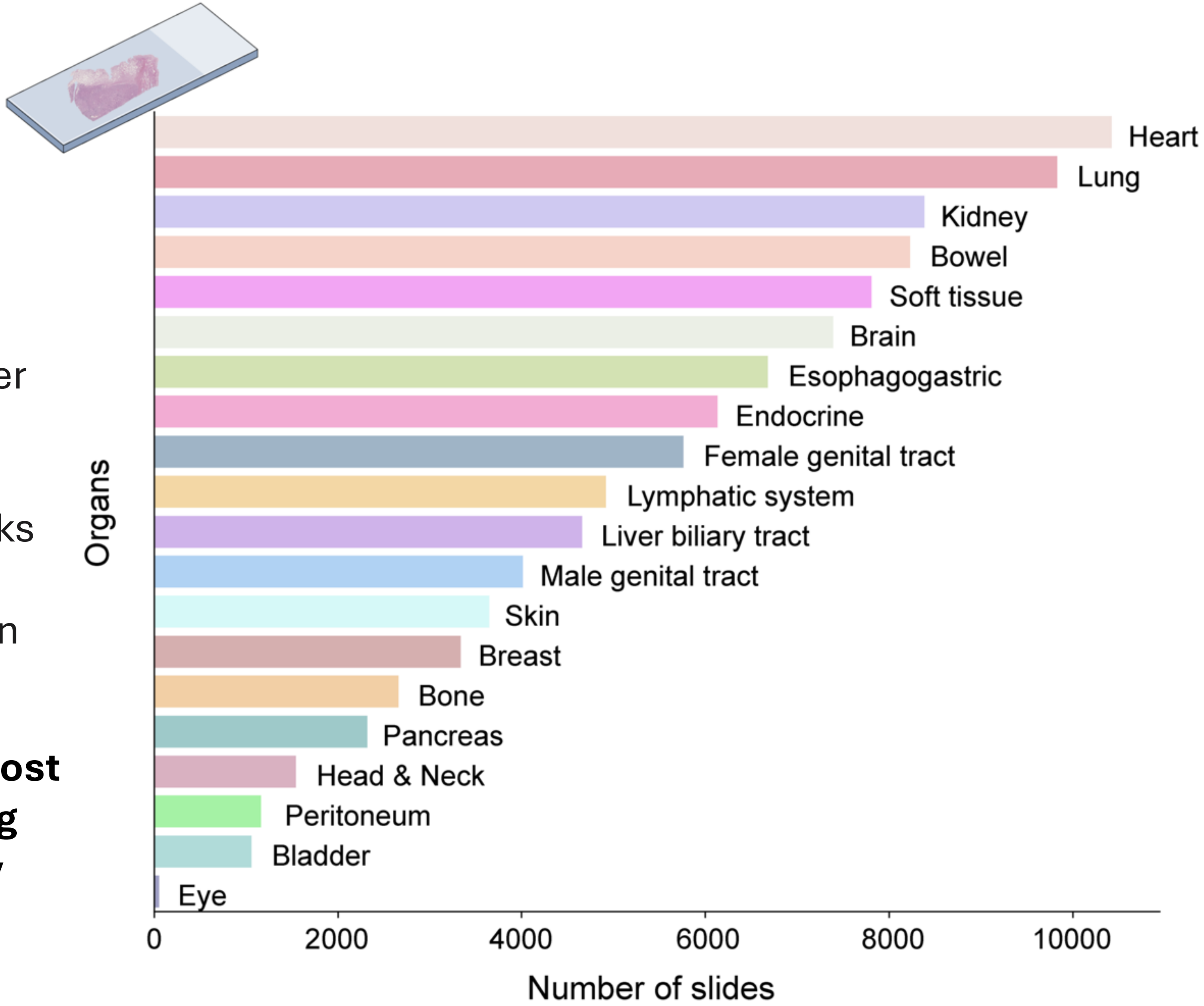
Ming Y. Lu[1,2,3,4,5,11], Bowen Chen[1,2,11], Drew F. K. Williamson[1,2,3,11], Richard J. Chen[1,2,3,4,6], Ivy Liang[1,7], Tong Ding[1,7], Guillaume Jaume[1,2,3,4], Igor Odintsov[1], Long Phi Le[2], Georg Gerber[1], Anil V. Parwani[8], Andrew Zhang[1,2,3,4,9] & Faisal Mahmood[1,2,3,4,10] ✉

CONCH

http://github.com/mahmoodlab/CONCH

UNI: Mass-100K - 100K WSIs for large-scale vision SSL pretraining

- 100 million patches sampled across 100,000+ WSIs
- 380+ unique OncoTree Codes and other disease labels
- WSIs from commonly used benchmarks (e.g. TCGA) are not included to avoid data leakage in downstream evaluation

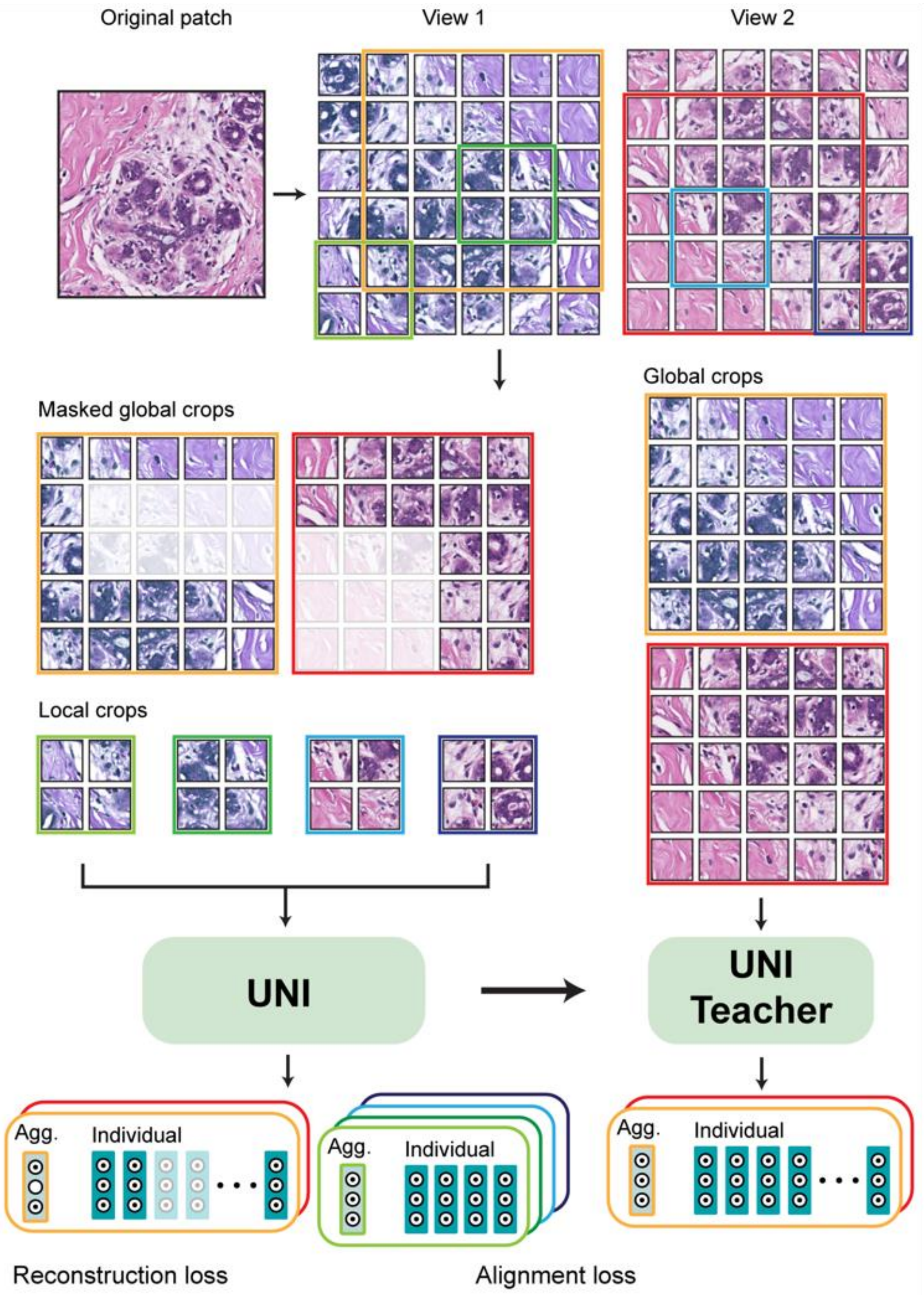


Mass-100K represents the largest and most diverse SSL pretraining dataset including neoplastic, infectious and inflammatory diseases.

(Nature Medicine, 2024)

UNI: Pretraining via DINOv2

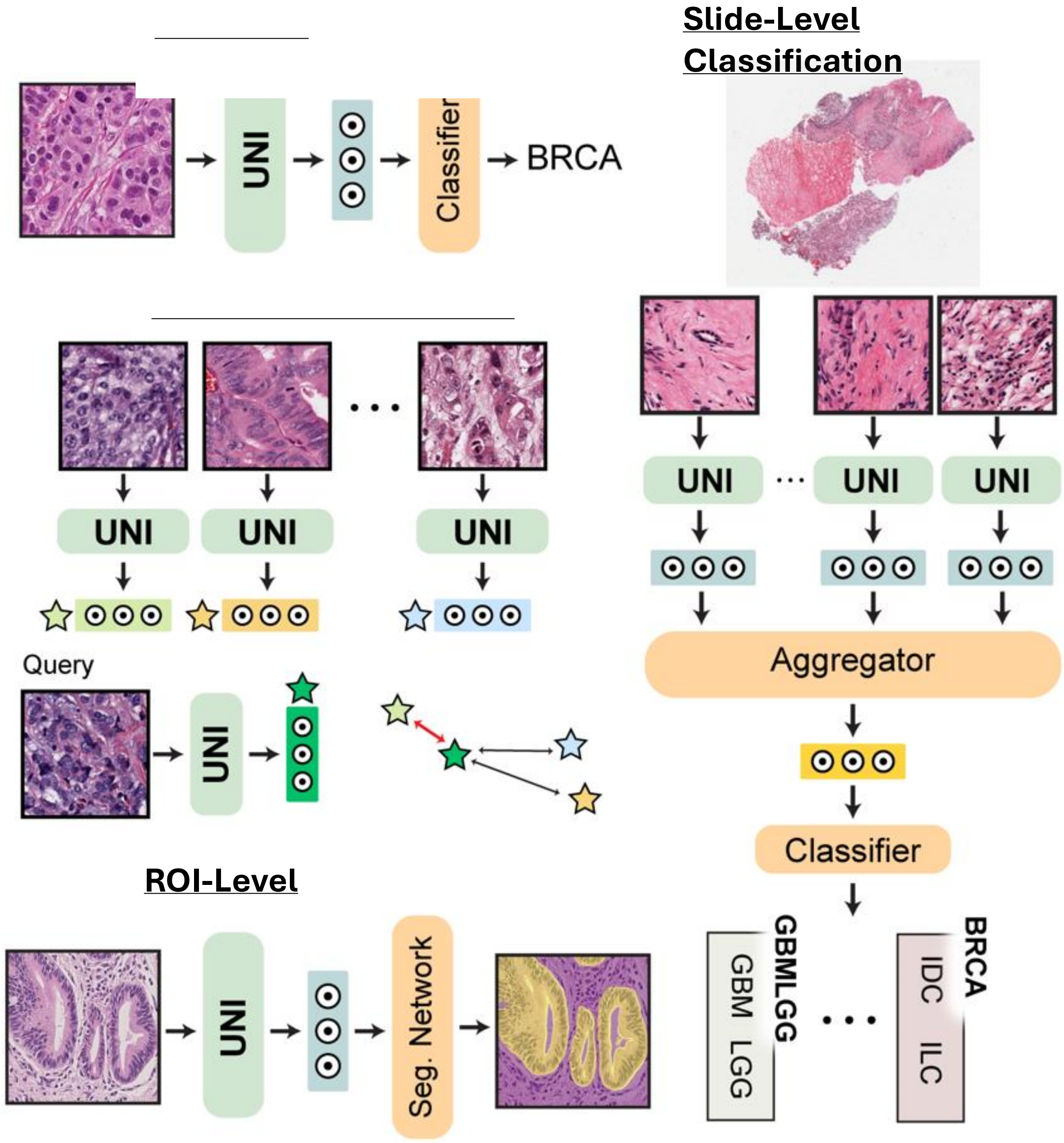
- Dino v2 SSL pretraining recipe combing masked image modeling and self-distillation
- 4 x 8 A100 GPUs for multi-node training of ViT-L on Mass-100K for up to 125,000 iterations
- Compare against SOTA SSL encoders + baseline:
  - CTransPath (Wang *et al.* 2022)
  - REMEDIS (Azizi *et al.* 2023)
  - ResNet50 (transfer from ImageNet)

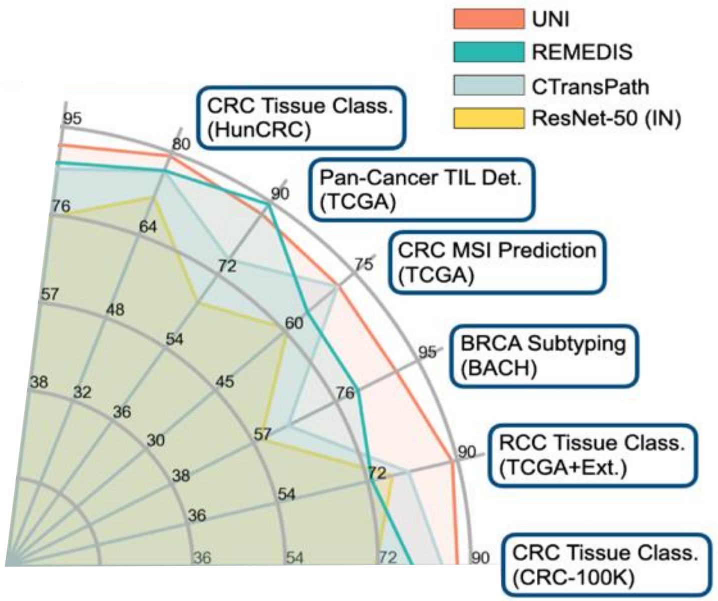


(Nature Medicine, 2024)

UNI: Overview of Tasks

(Nature Medicine, 2024)



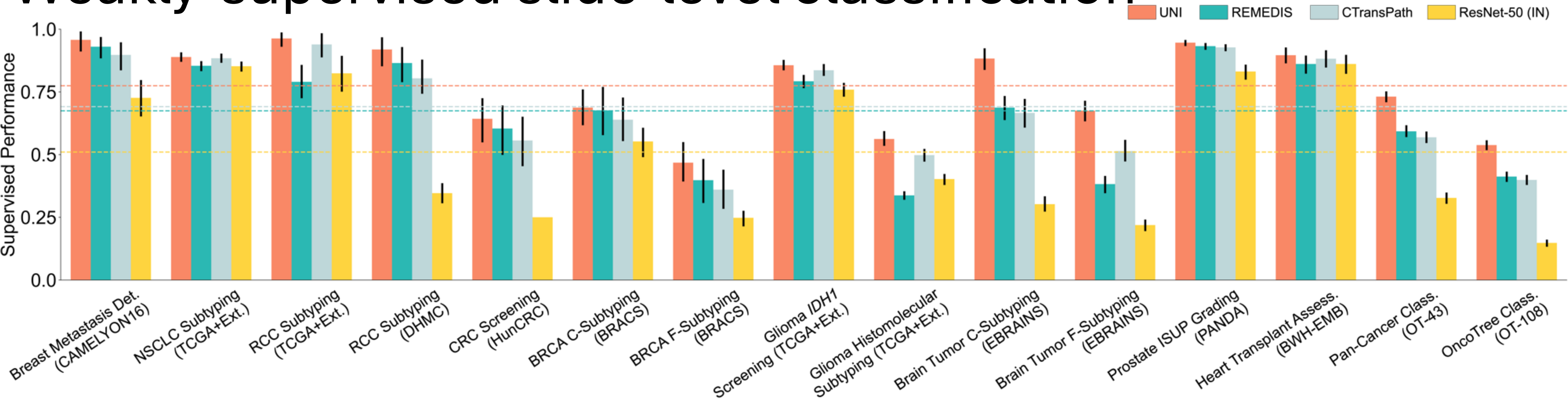


UNI outperforms other pretrained encoders on 33 clinical tasks in anatomical pathology.

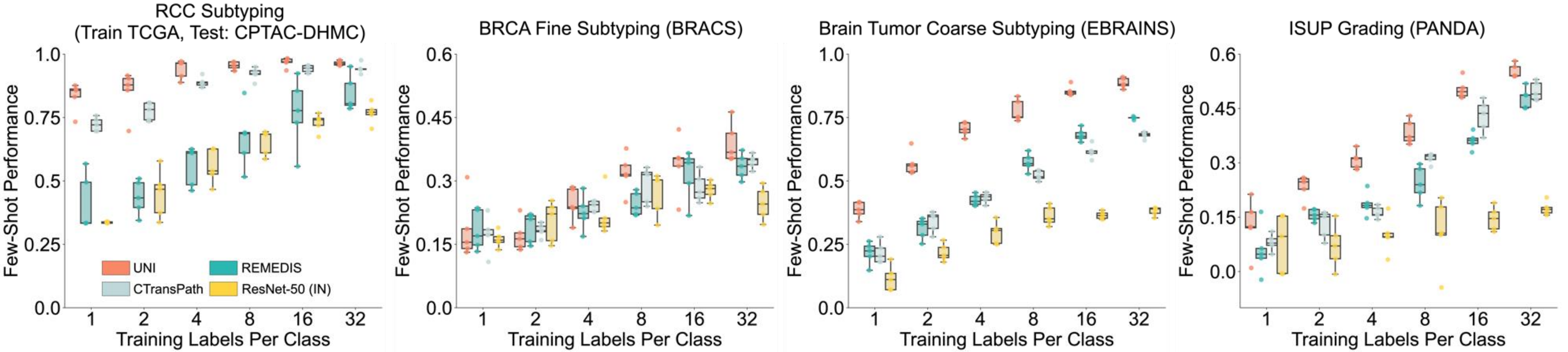
HARVARD MEDICAL SCHOOL BRIGHAM AND WOMEN'S HOSPITAL Dana-Farber Cancer Institute BROAD INSTITUTE

Weakly-supervised slide-level classification

(Nature Medicine, 2024)



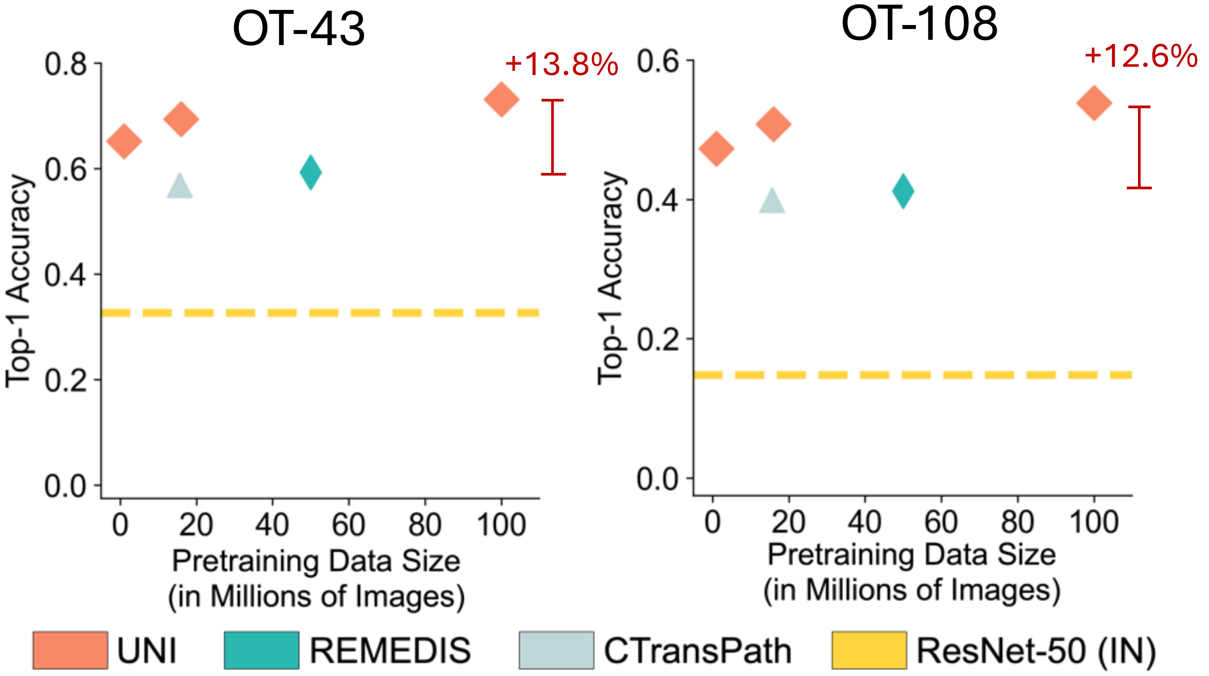
Few-shot classification:



UNI: OT-43/108 - A new large-scale subtyping benchmark

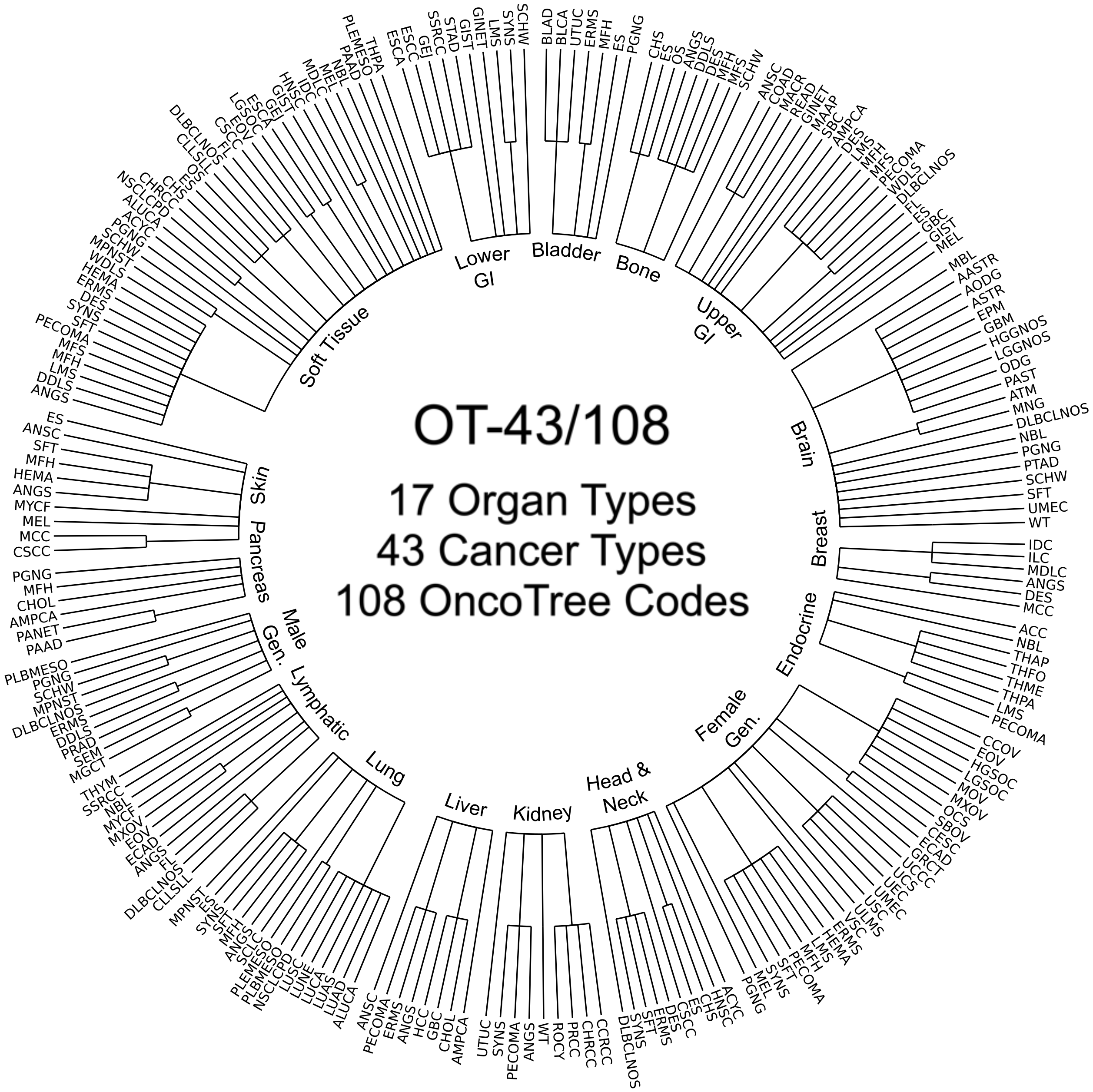
(Nature Medicine, 2024)

- OncoTree-43 (OT-43): 43-way cancer type classification
- OncoTree-108 (OT-108): 108-way OncoTree Code (cancer subtype) classification
- Challenging, large, representative benchmark for assessing performance of SSL pretrained encoders



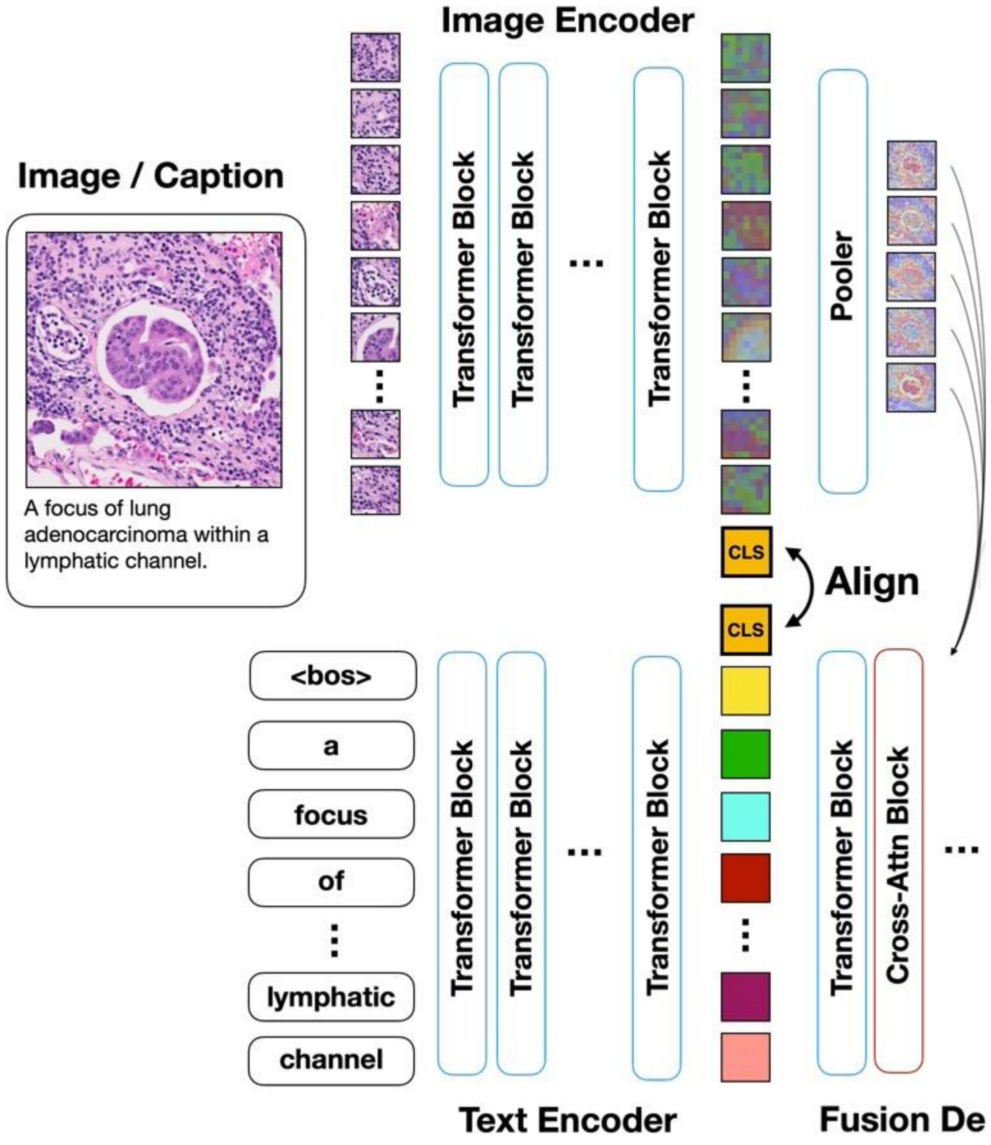
OT-43

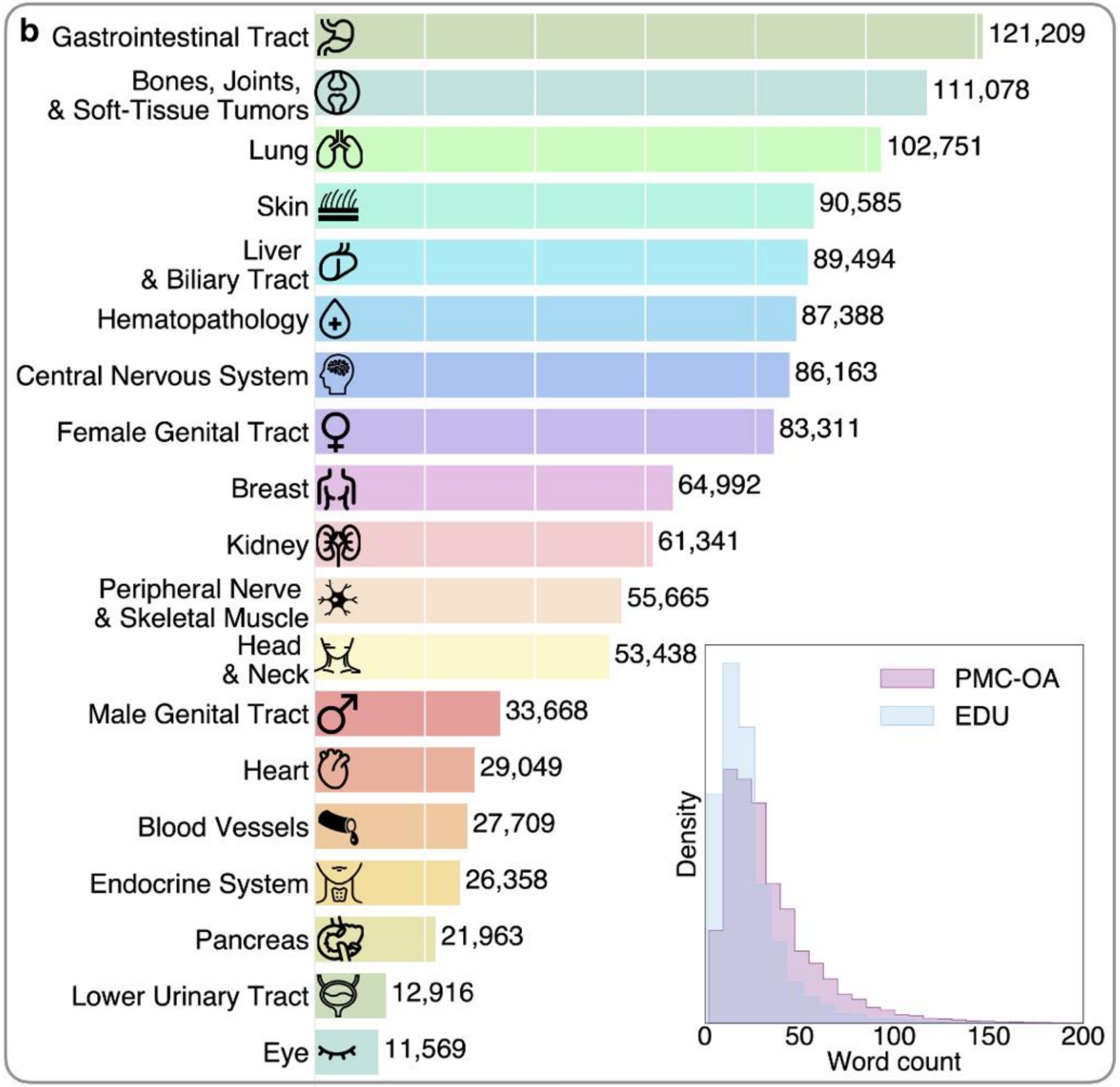
OT-108



CONCH: CONtrastive learning from Captions for Histopathology

(Nature Medicine, 2024)





l,

atch

/

g)

(Nature Medicine, 2024)

Zeroshot classification

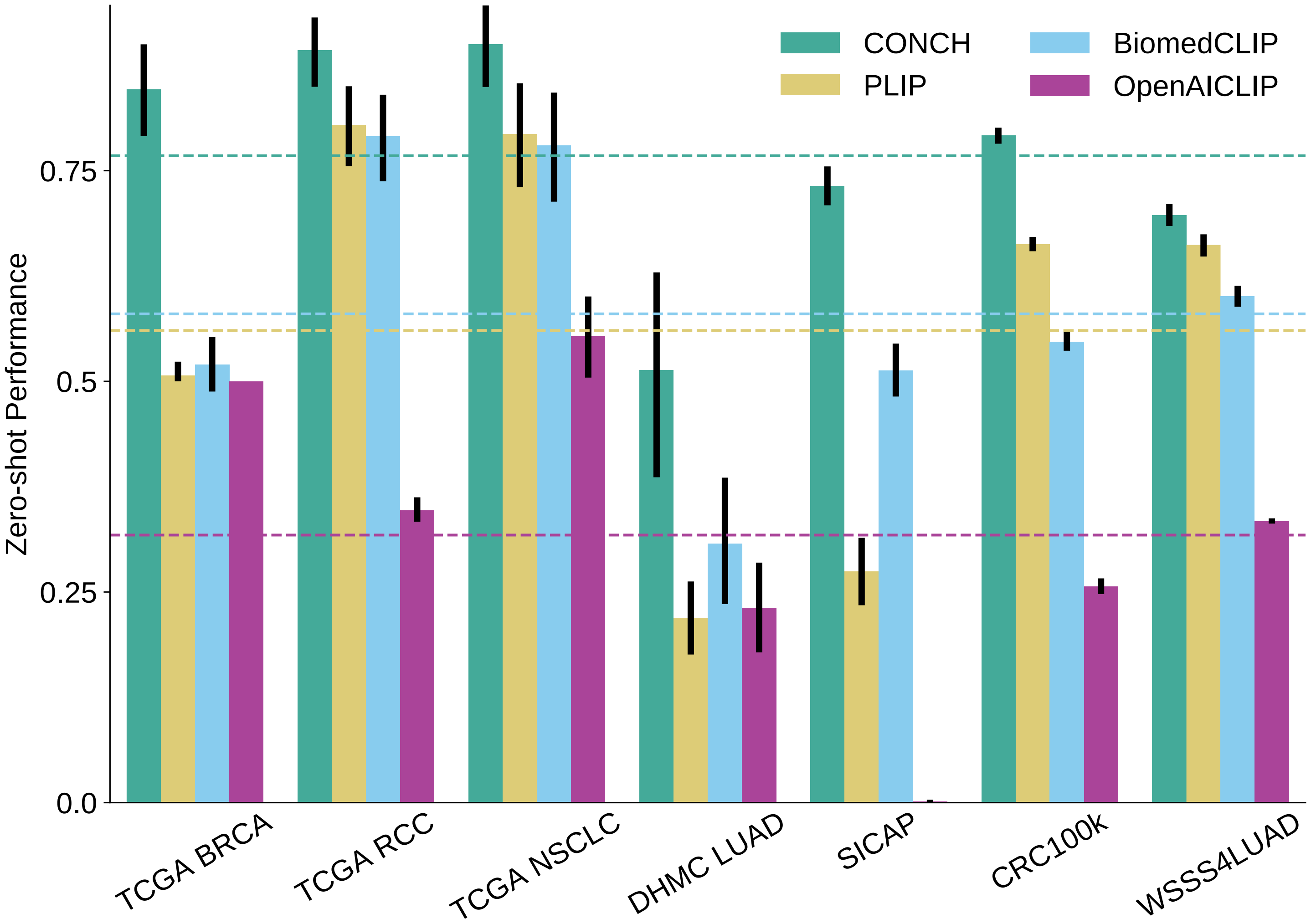
Zero shot classification through prompting!

4 Slide-level benchmarks (using MI-Zero):

- TCGA BRCA subtyping
- TCGA RCC subtyping
- TCGA NSCLC subtyping
- DHMC LUAD pattern classification

3 Patch-level benchmarks (using CLIP-style zeroshot):

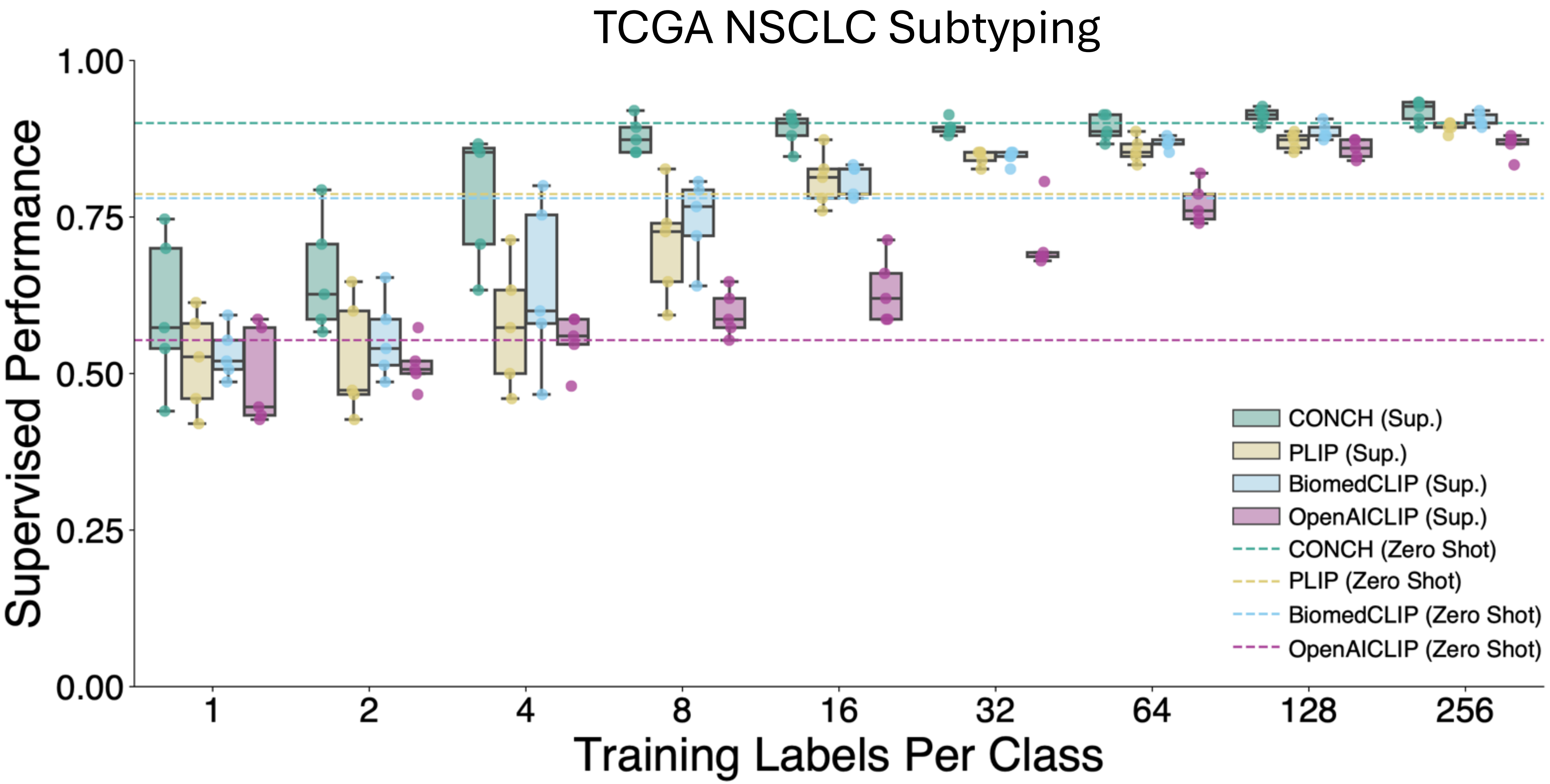
- SICAP gleason grading
- CRC100k tissue type classification
- WSSS4LUAD tissue type classifcation



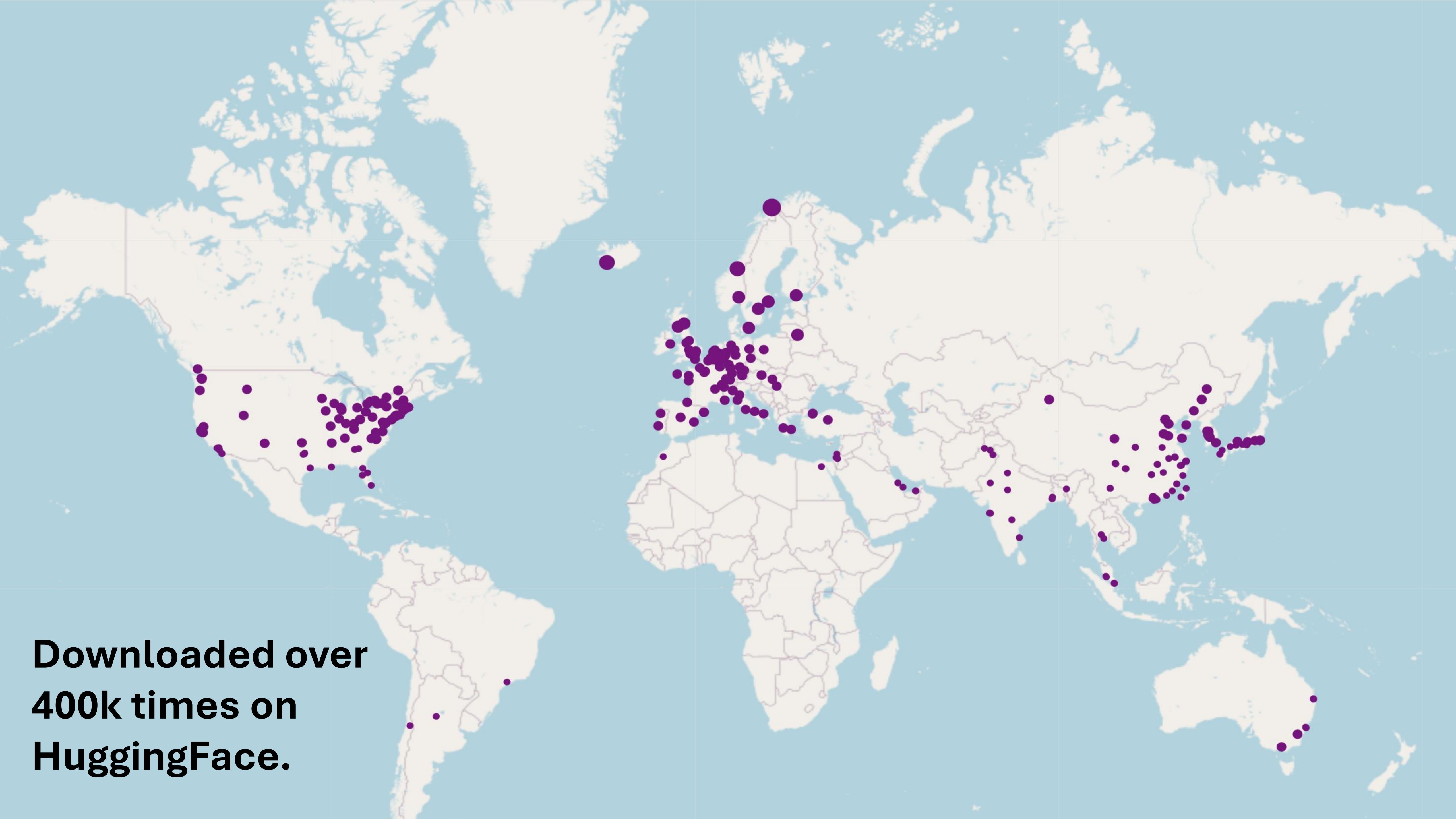
Mahmood Lab
AI for Pathology

(Nature Medicine, 2024)

Fewshot classification



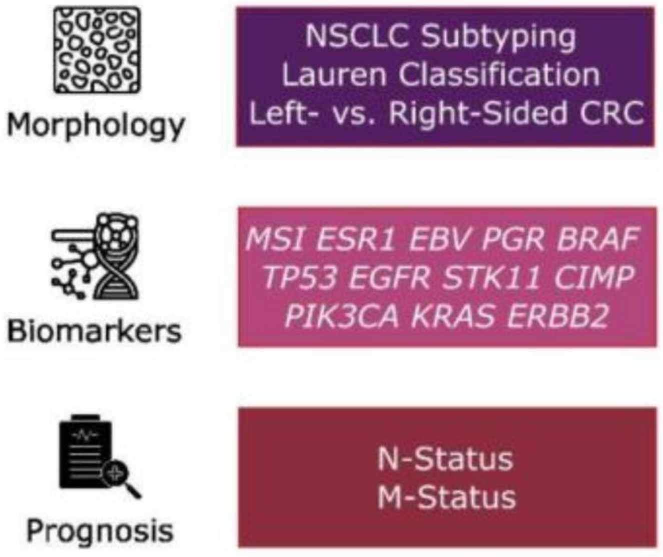
1. **CONCH zeroshot is a strong baseline for classification, competitive with supervised few-shot learning by SOTA visual language encoders.**
2. **CONCH image encoder is more label efficient and often requires few labels to reach competitive performance**



Downloaded over 400k times on HuggingFace.

UNI and CONCH External Validation: Biomarker Assessment and FM Comparisons

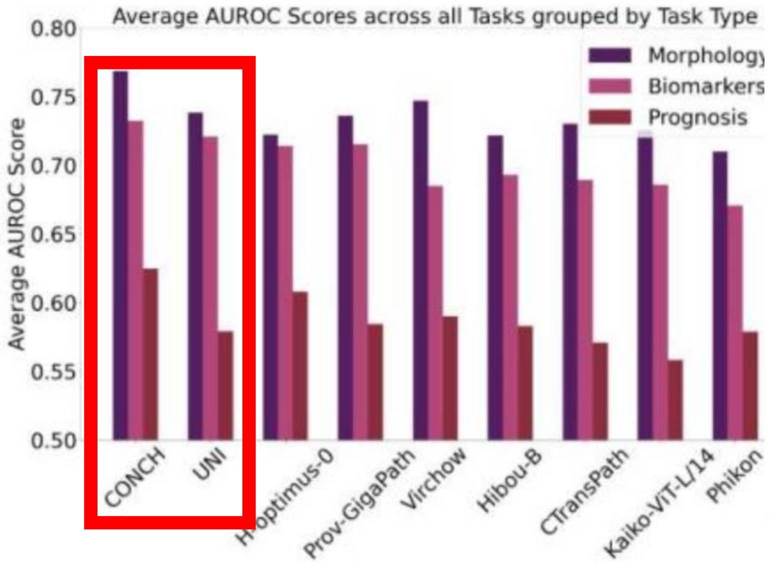
- CONCH and UNI continues to be the top-2 SOTA ROI foundation models across 31 clinical tasks spanning morphological subtyping, biomarker prediction, and cancer prognosis

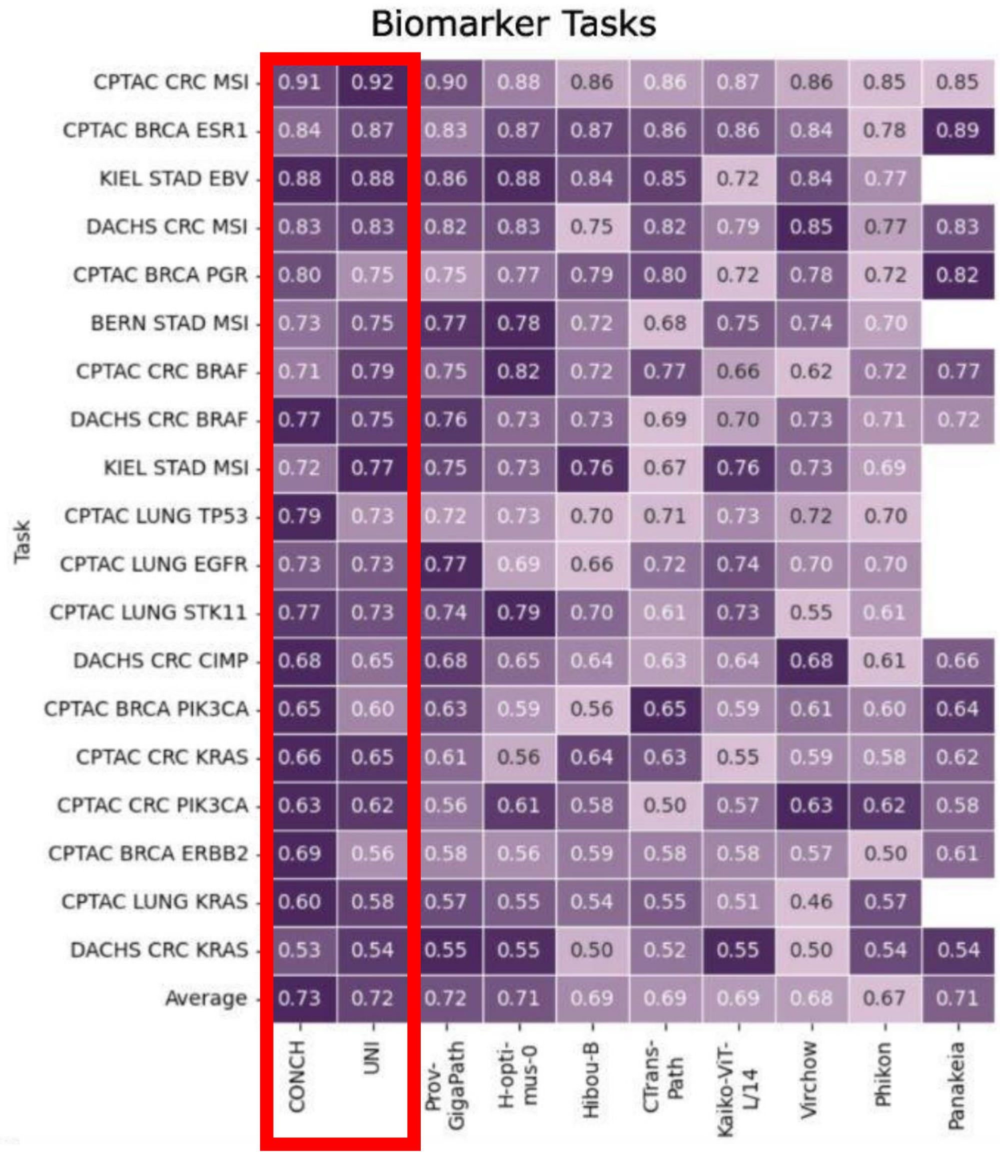


NSCLC Subtyping
Lauren Classification
Left- vs. Right-Sided CRC

MSI ESR1 EBV PGR BRAF TP53 EGFR STK11 CIMP PIK3CA KRAS ERBB2

N-Status
M-Status


Average AUROC Scores across all Tasks grouped by Task Type


Biomarker Tasks

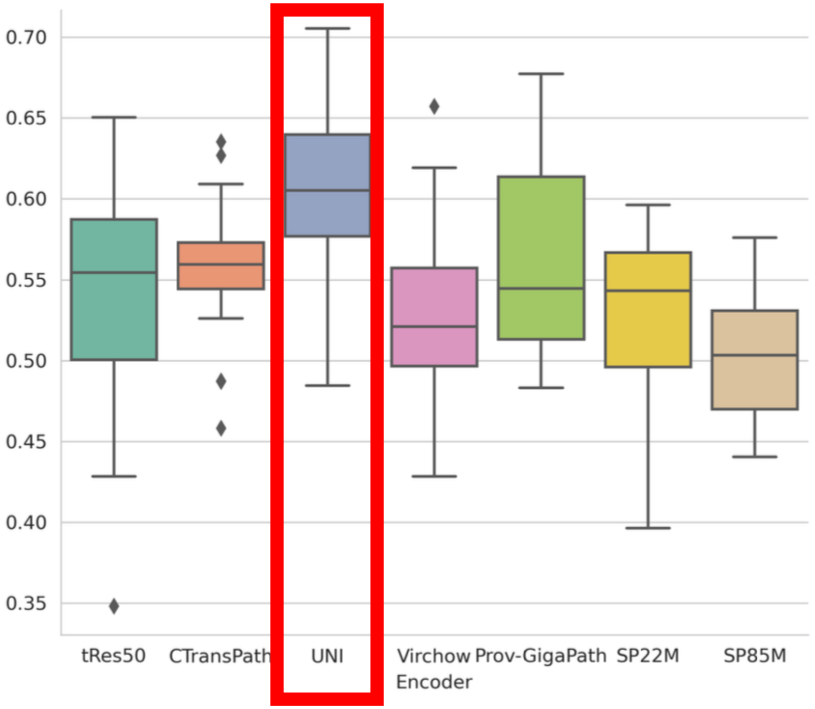
Task	CONCH	UNI	Prov-GigaPath	H-opti-mus-0	Hibou-B	CTrans-Path	Kaiko-ViT-L/14	Virchow	Phikon	Panakeia
CPTAC CRC MSI	0.91	0.92	0.90	0.88	0.86	0.86	0.87	0.86	0.85	0.85
CPTAC BRCA ESR1	0.84	0.87	0.83	0.87	0.87	0.86	0.86	0.84	0.78	0.89
KIEL STAD EBV	0.88	0.88	0.86	0.88	0.84	0.85	0.72	0.84	0.77	
DACHS CRC MSI	0.83	0.83	0.82	0.83	0.75	0.82	0.79	0.85	0.77	0.83
CPTAC BRCA PGR	0.80	0.75	0.75	0.77	0.79	0.80	0.72	0.78	0.72	0.82
BERN STAD MSI	0.73	0.75	0.77	0.78	0.72	0.68	0.75	0.74	0.70	
CPTAC CRC BRAF	0.71	0.79	0.75	0.82	0.72	0.77	0.66	0.62	0.72	0.77
DACHS CRC BRAF	0.77	0.75	0.76	0.73	0.73	0.69	0.70	0.73	0.71	0.72
KIEL STAD MSI	0.72	0.77	0.75	0.73	0.76	0.67	0.76	0.73	0.69	
CPTAC LUNG TP53	0.79	0.73	0.72	0.73	0.70	0.71	0.73	0.72	0.70	
CPTAC LUNG EGFR	0.73	0.73	0.77	0.69	0.66	0.72	0.74	0.70	0.70	
CPTAC LUNG STK11	0.77	0.73	0.74	0.79	0.70	0.61	0.73	0.55	0.61	
DACHS CRC CIMP	0.68	0.65	0.68	0.65	0.64	0.63	0.64	0.68	0.61	0.66
CPTAC BRCA PIK3CA	0.65	0.60	0.63	0.59	0.56	0.65	0.59	0.61	0.60	0.64
CPTAC CRC KRAS	0.66	0.65	0.61	0.56	0.64	0.63	0.55	0.59	0.58	0.62
CPTAC CRC PIK3CA	0.63	0.62	0.56	0.61	0.58	0.50	0.57	0.63	0.62	0.58
CPTAC BRCA ERBB2	0.69	0.56	0.58	0.56	0.59	0.58	0.58	0.57	0.50	0.61
CPTAC LUNG KRAS	0.60	0.58	0.57	0.55	0.54	0.55	0.51	0.46	0.57	
DACHS CRC KRAS	0.53	0.54	0.55	0.55	0.50	0.52	0.55	0.50	0.54	0.54
Average	0.73	0.72	0.72	0.71	0.69	0.69	0.69	0.68	0.67	0.71

Neidlinger et al., Benchmarking foundation models as feature extractors for weakly-supervised computational pathology. arXiv, 2024.

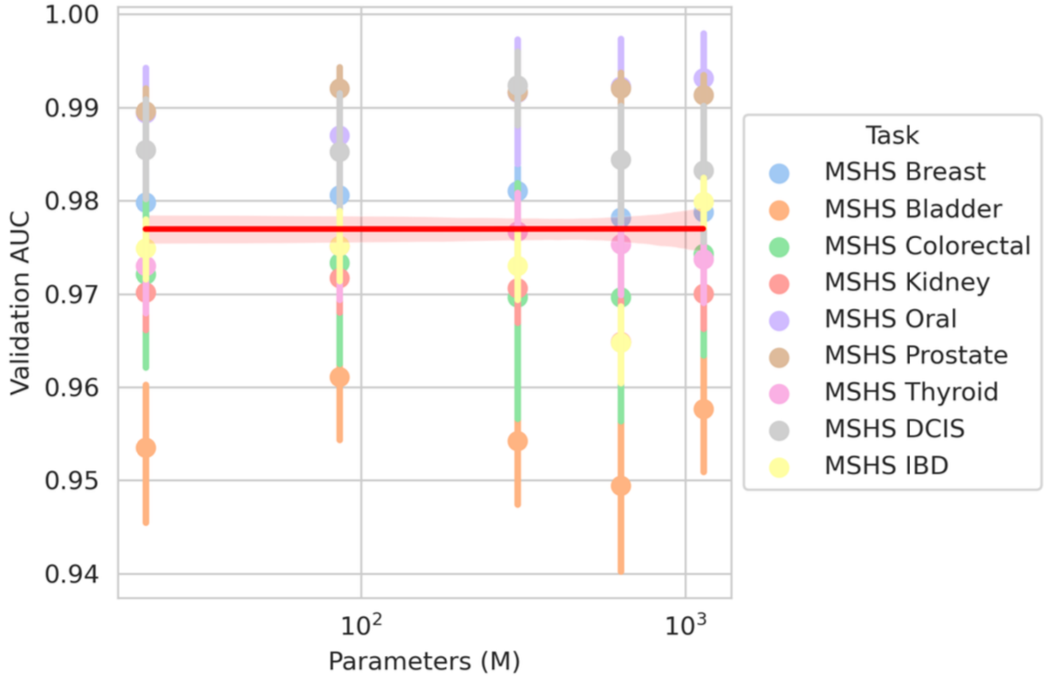
UNI and CONCH External Validation: Performance Efficiency Assessment

- UNI continues to be the SOTA ROI foundation model on 9 disease detection and 11 biomarker prediction tasks.
- Diminishing performance gains found with larger models.
- UNI performance on NSCLC IO is attributed to training diversity

<u>MSKCC NSCLC IO Prediction</u>

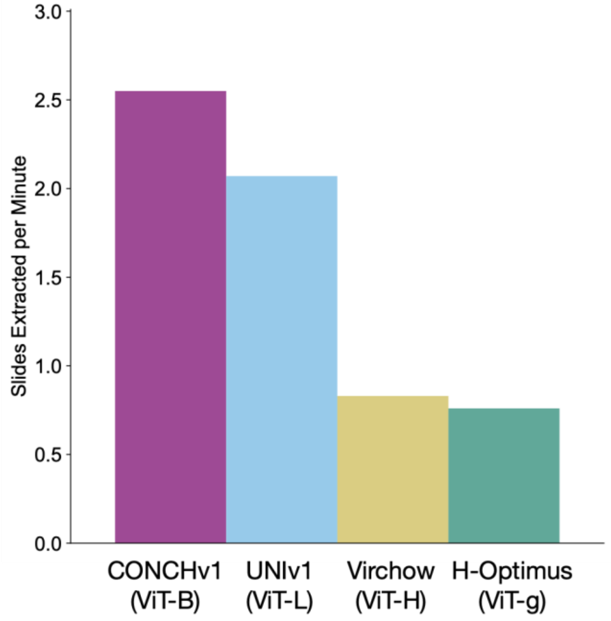


<u>AUC Performance vs. Parameter Efficiency</u>

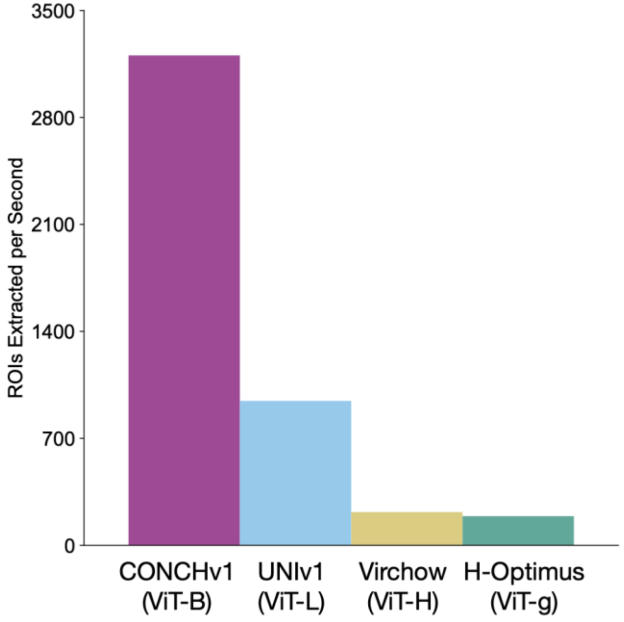


Campanella et al., A Clinical Benchmark of Public Self-Supervised Pathology Foundation Models. arXiv, 2024.

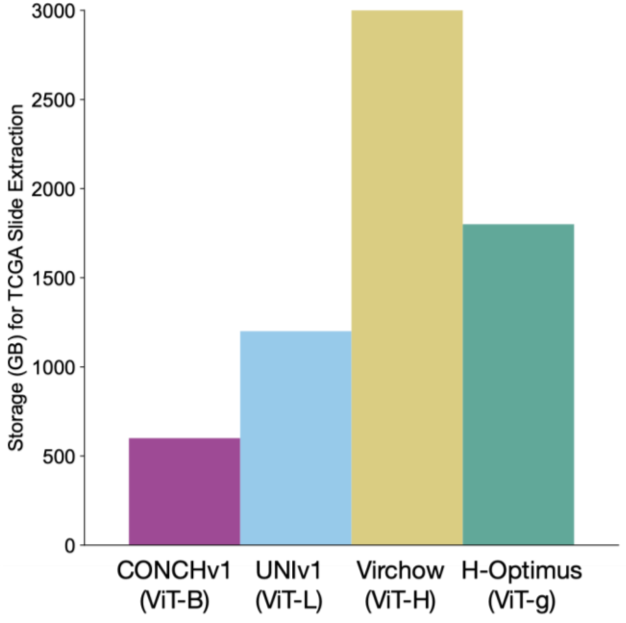
UNI and CONCH

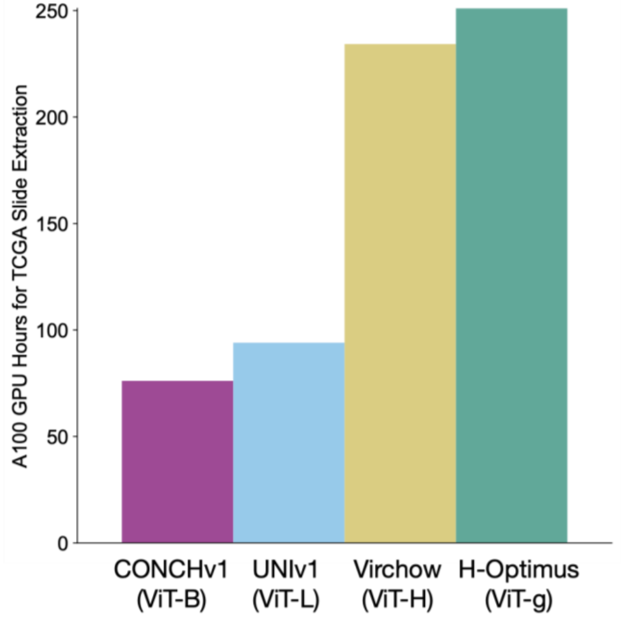

Slide Feature Extraction Speed

* Includes IO and CPU-GPU comm


Patch Feature Extraction Speed

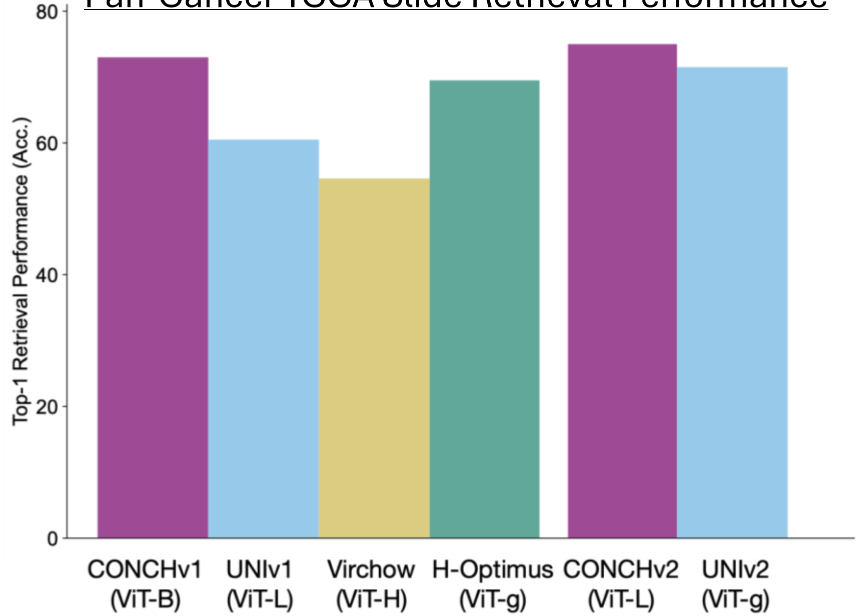
* Excludes IO and CPU-GPU comm


Storage Cost

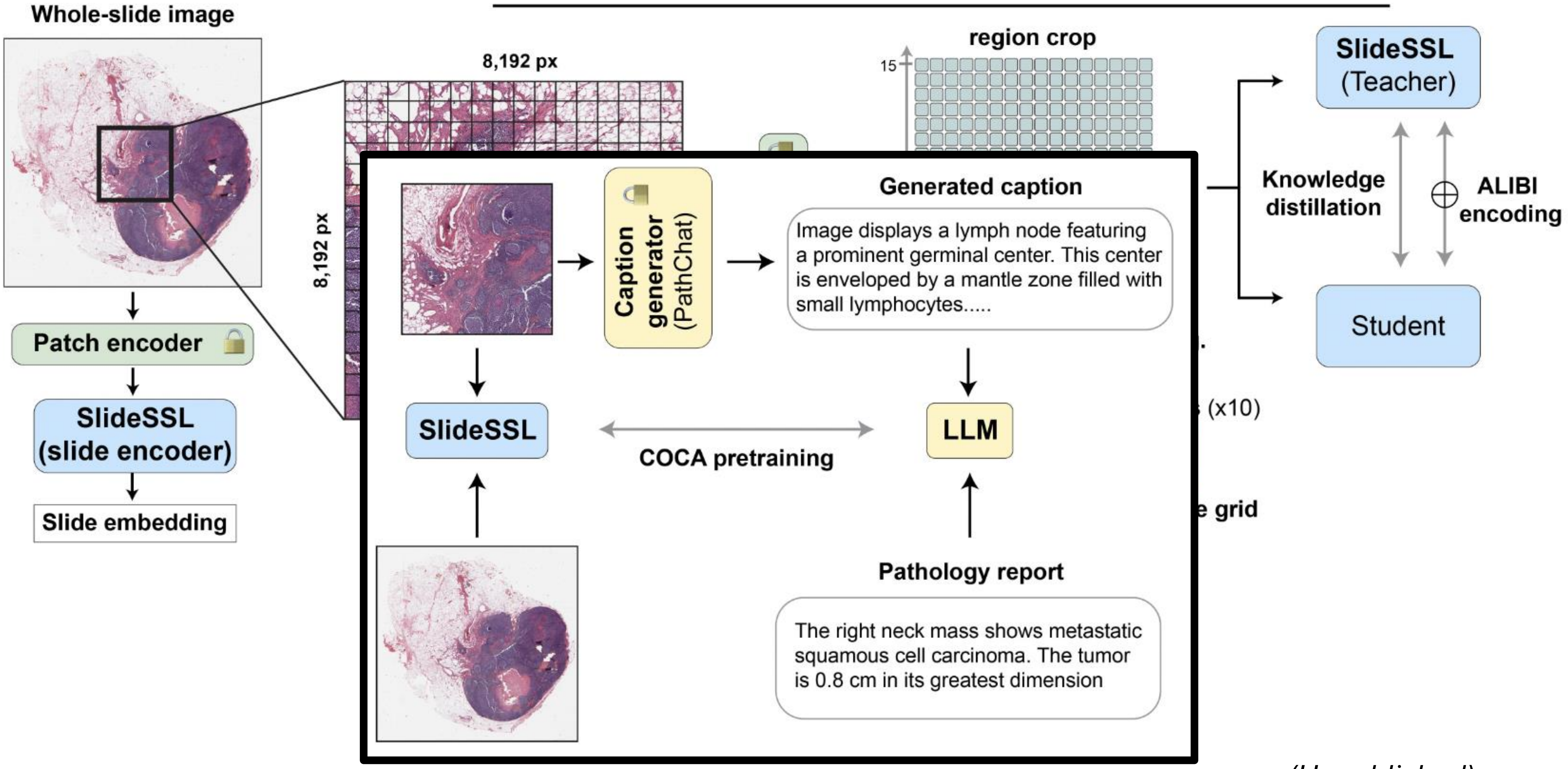

Extraction Time Cost

- Non-overlapping [256 x 256] patch feature extraction from 11,661 WSIs in the TCGA
- Approx. 13,353 tissue patches per WSI (155.7M tissue patches in total)
- A100 80GB SXM4 with PanFS HPC Storage

- 32-class ROI-level pan-cancer tissue retrieval evaluation in the TCGA


Pan-Cancer TCGA Slide Retrieval Performance

Slide level SSL



Whole-slide image

Patch encoder

SlideSSL (slide encoder)

Slide embedding

8,192 px

8,192 px

region crop

15

SlideSSL (Teacher)

Knowledge distillation

ALIBI encoding

Student

(x10)

e grid

Caption generator (PathChat)

Generated caption

Image displays a lymph node featuring a prominent germinal center. This center is enveloped by a mantle zone filled with small lymphocytes.....

SlideSSL

COCA pretraining

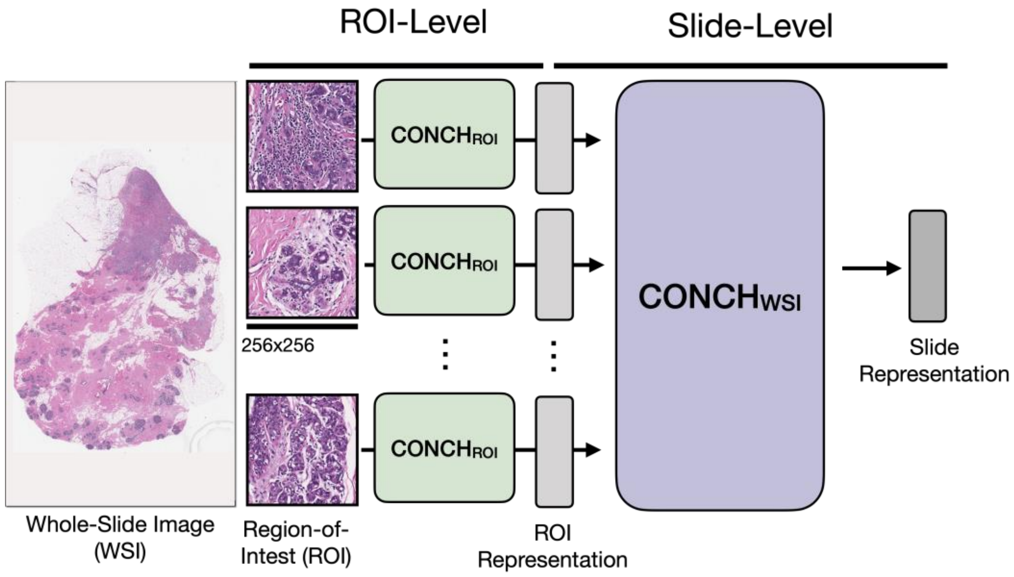
LLM

Pathology report

The right neck mass shows metastatic squamous cell carcinoma. The tumor is 0.8 cm in its greatest dimension

(Unpublished)

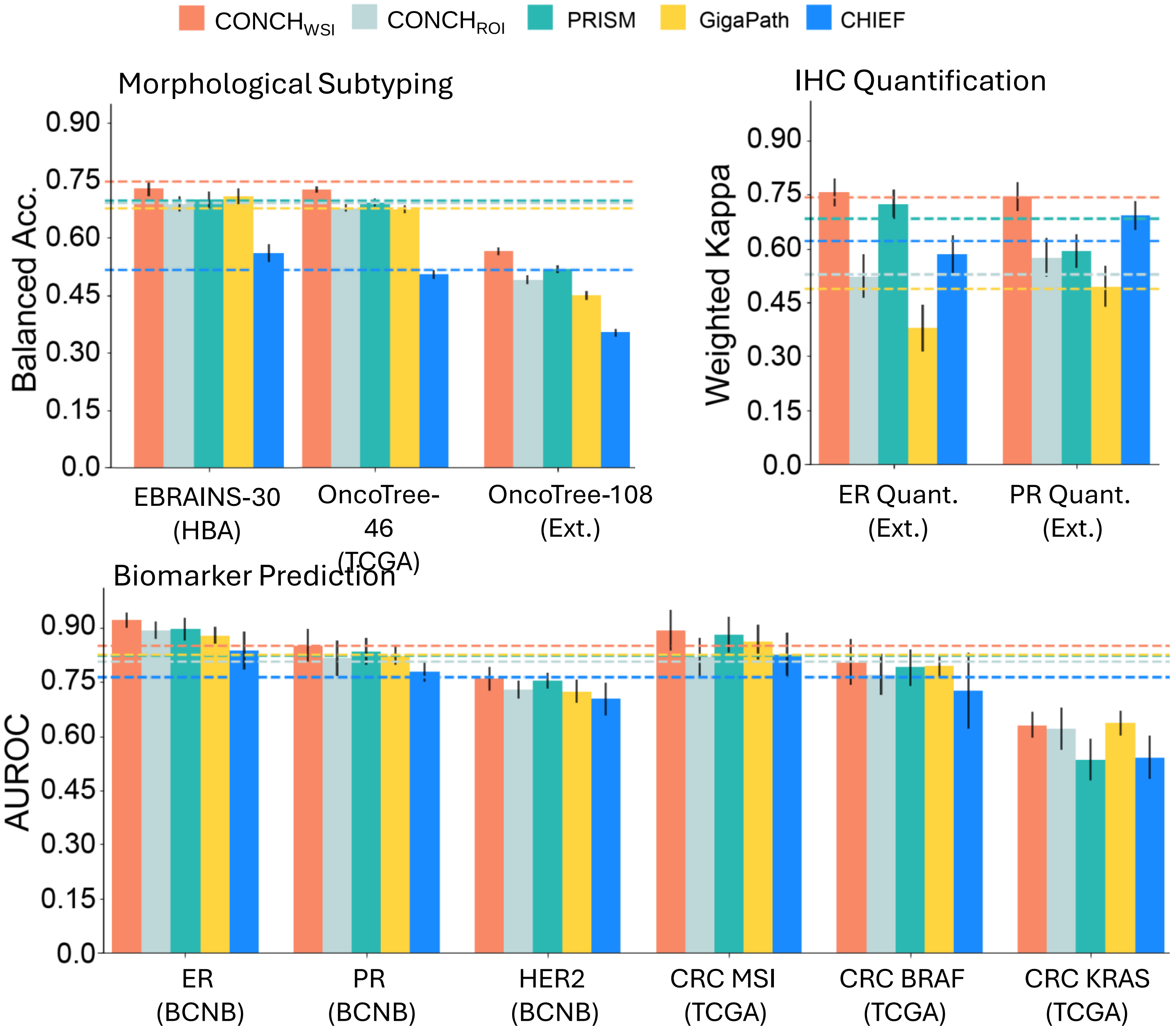
Slide$_{SSL}$: A Vision-Language Slide Foundation Model for CPath



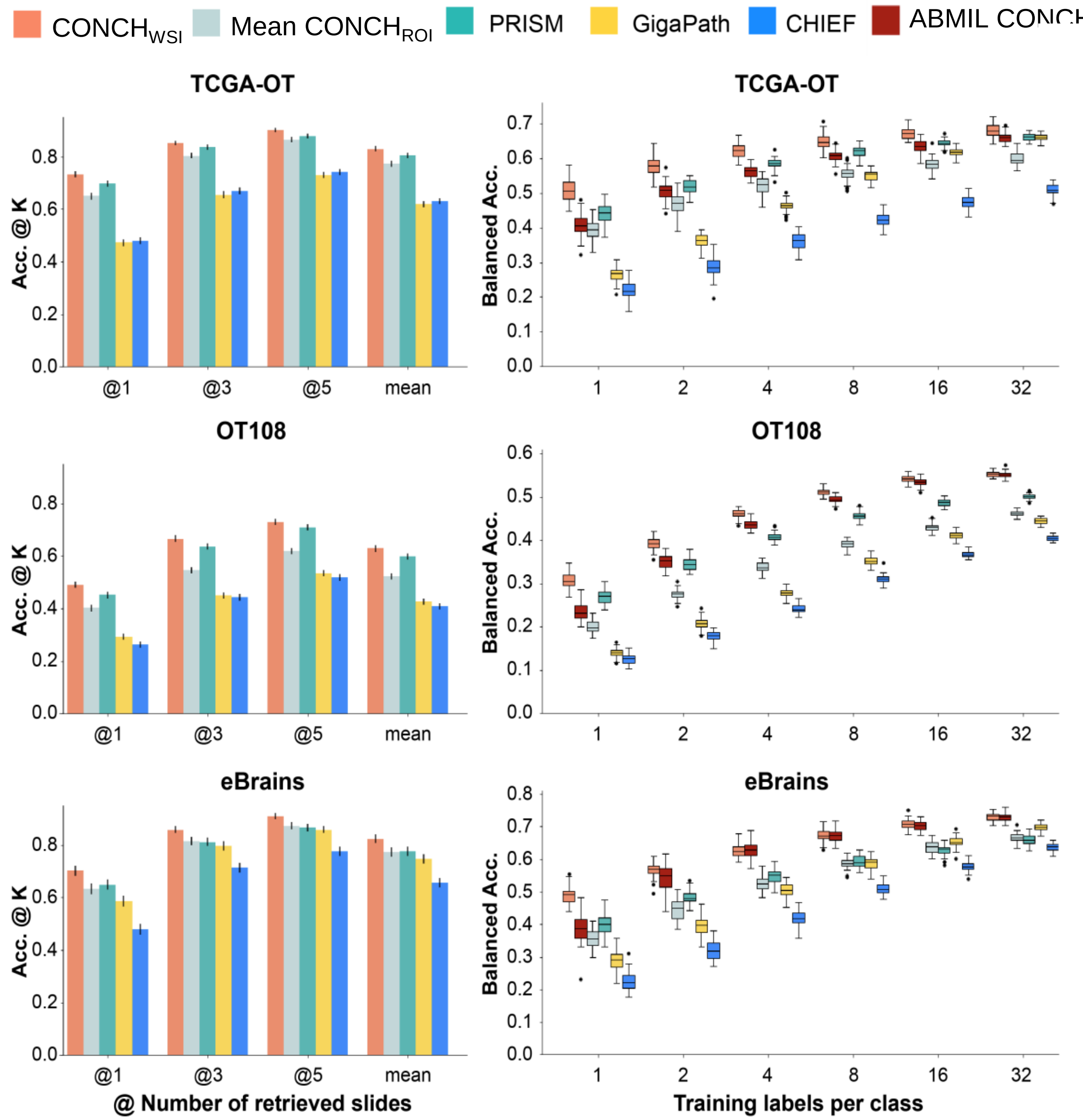
- A vision-language foundation model that scales the capabilities of CONCH to the whole slide level

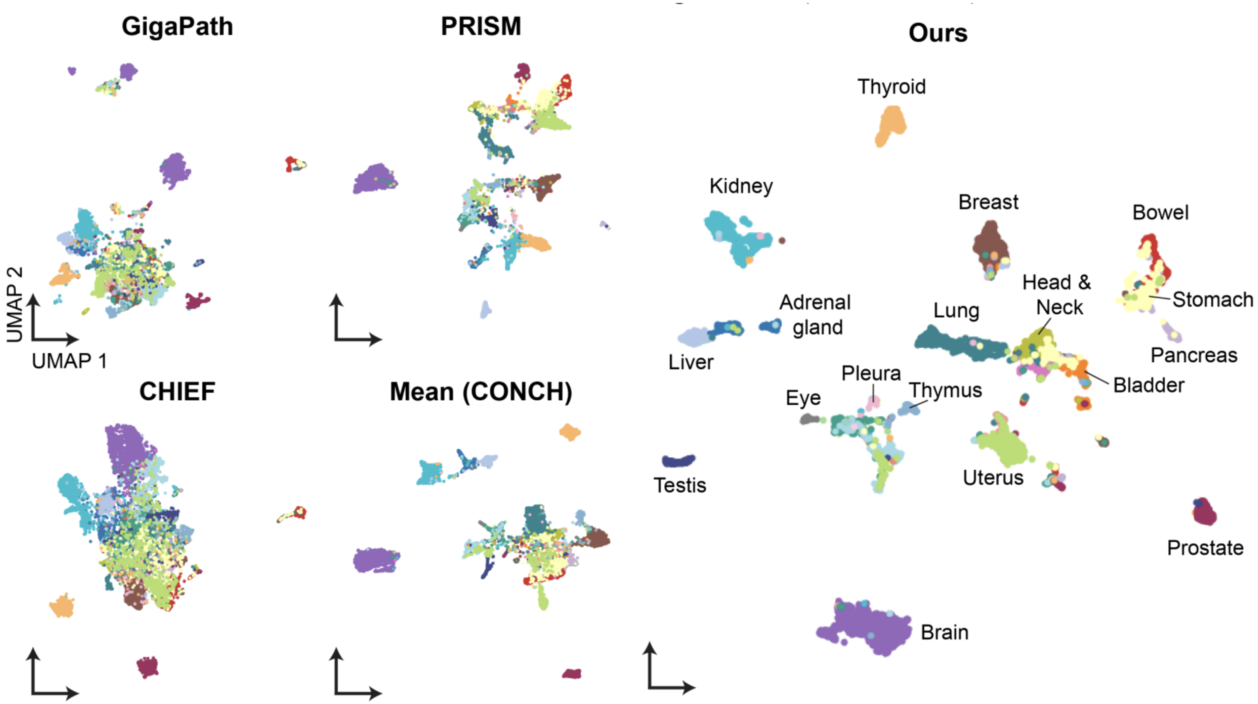
- Over 400K WSIs paired with synthetic pathology reports created by PathChat

- No weakly-supervised MIL needed in directly extracting powerful slide features



Slide$_{SSL}$: Few-Shot Performance and Human Pathology Atlas Development

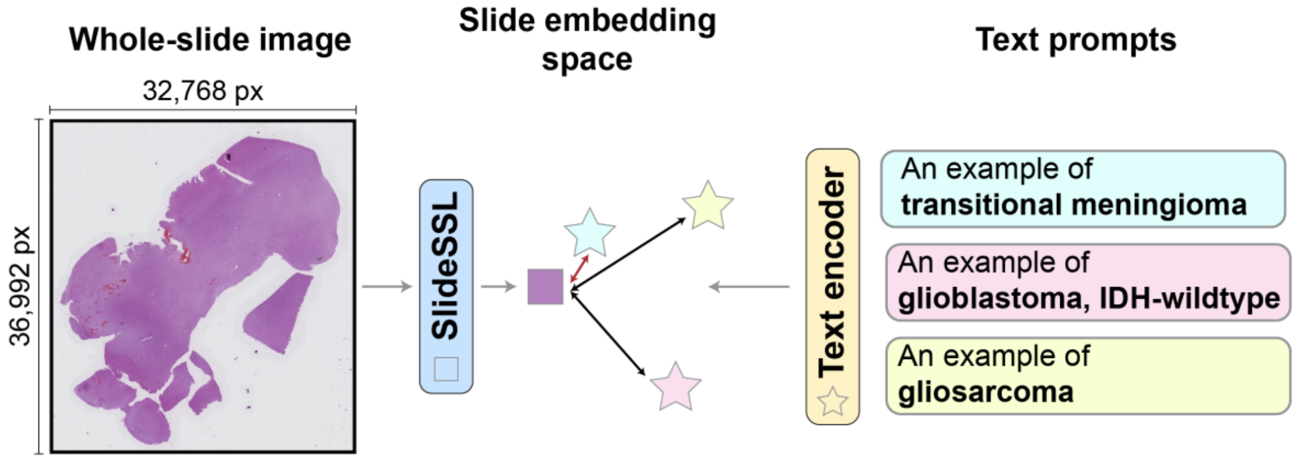




- CONCH$_{WSI}$ is the only slide foundation model that can outperform MIL and mean pooling baselines

- Save cost on embedding stores by saving slide features (instead of patch features)

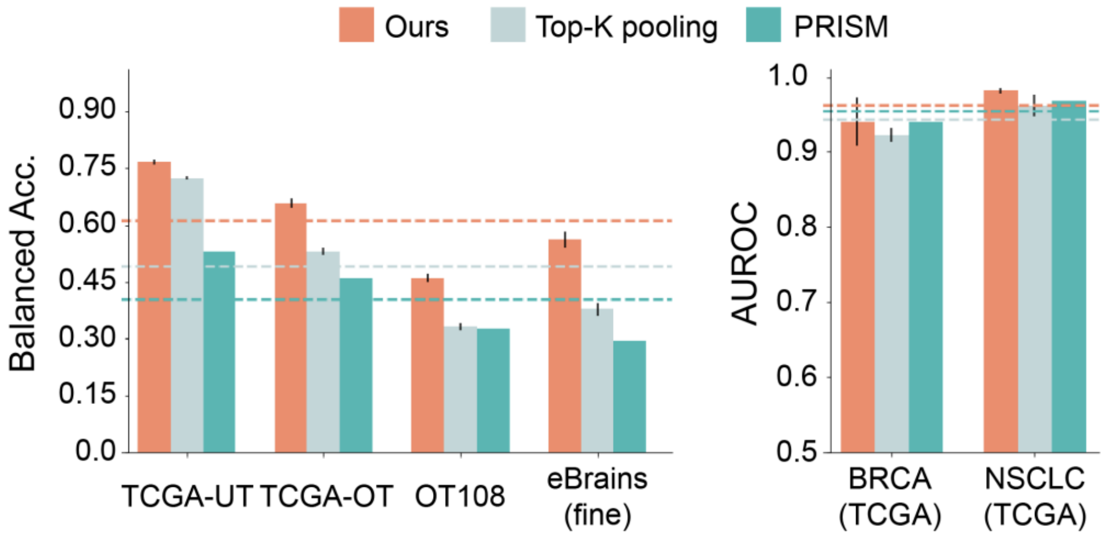
Slide$_{SSL}$: Zero-Shot Slide Classification and Report Generation

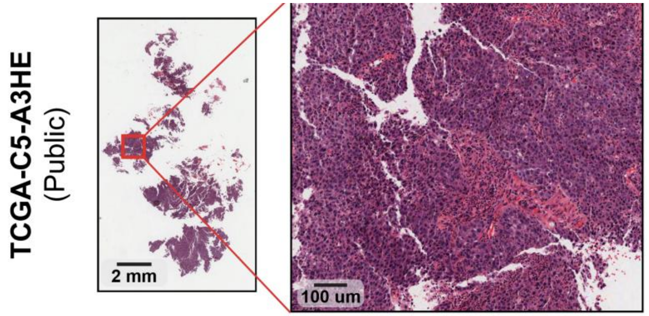


An example of **transitional meningioma**

An example of **glioblastoma, IDH-wildtype**

An example of **gliosarcoma**





Clinical report

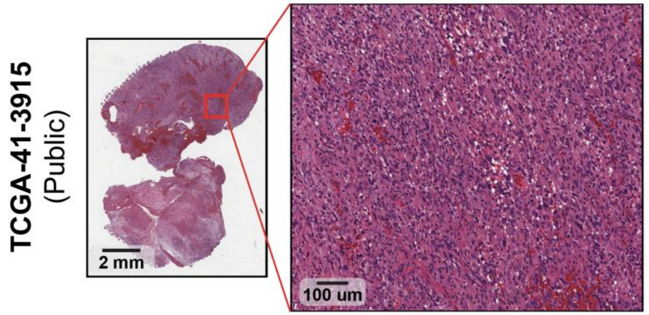
The slide from the cervix uteri shows a poorly differentiated squamous cell carcinoma.

Ours (generated)

microscopic examination of the provided slide confirms a diagnosis of squamous cell carcinoma originating from the uterine cervix.

PRISM (generated)

Diagnosis: High-grade serous carcinoma of unspecified organ.



Clinical report

The slide from the brain shows a glioblastoma multiforme (Astrocytoma WHO Grade IV), characterized by multinucleated malignant astrocytic cells with hyperchromasia, nucleomegaly, frequent mitotic activity, neovascularity, and large areas of necrosis.

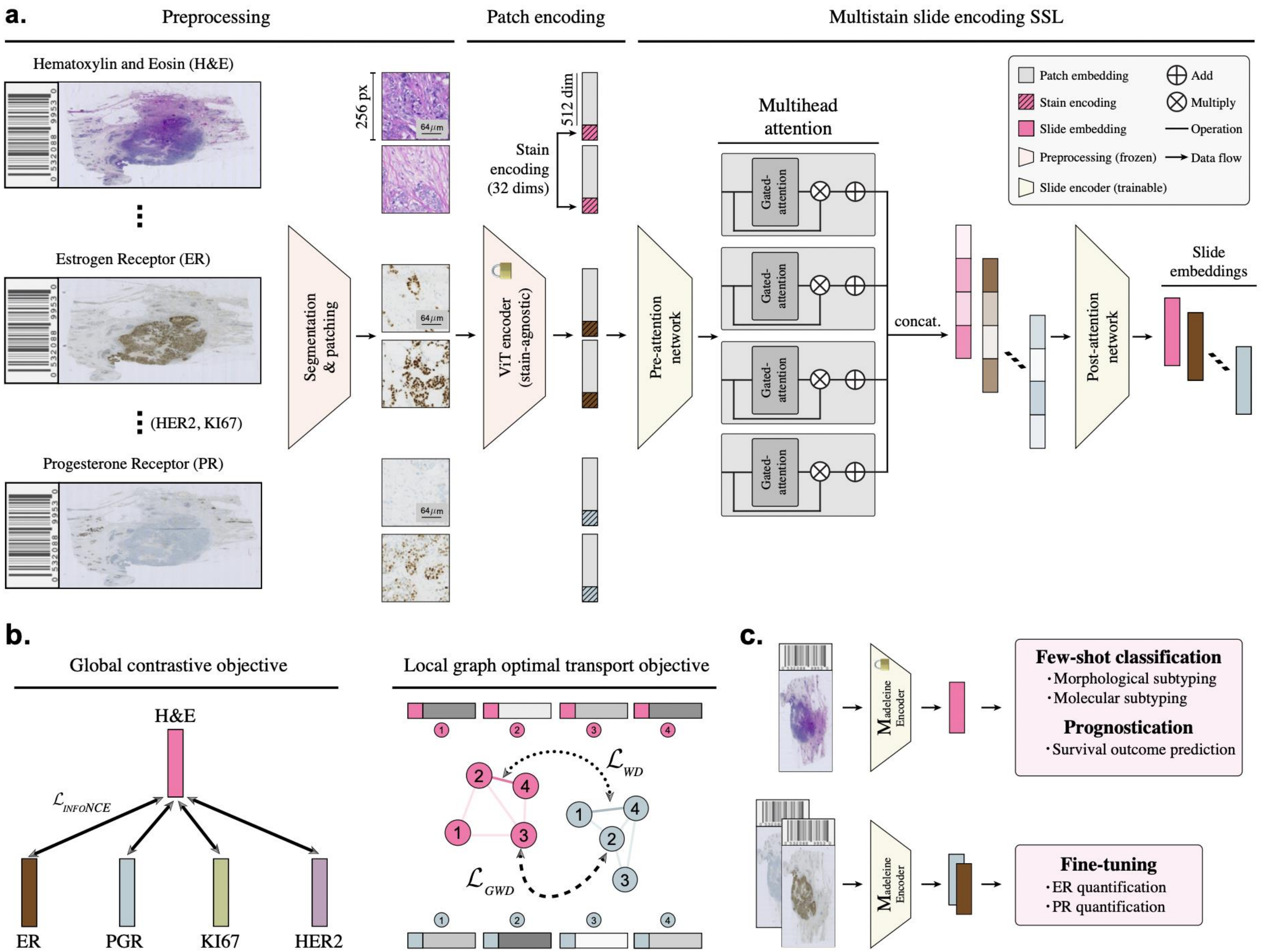
Ours (generated)

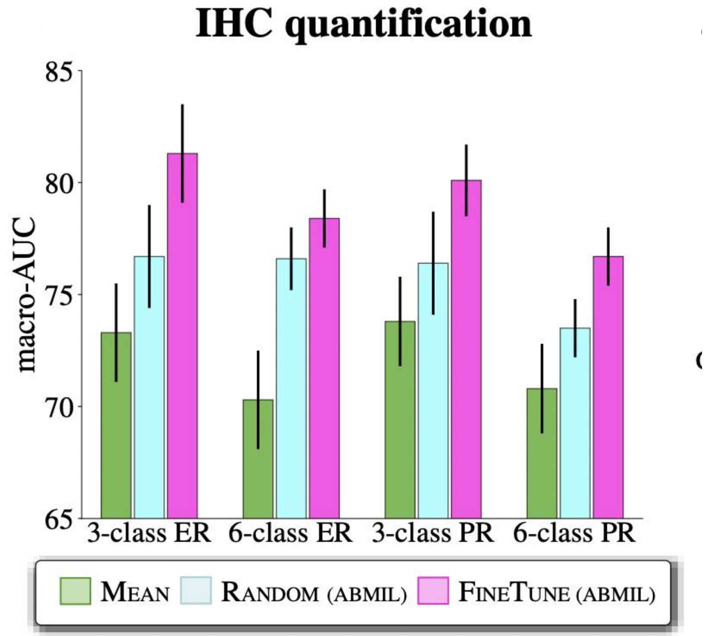
Microscopic analysis of the brain biopsy indicates a high-grade glioblastoma, classified as world health organization (w.h.o.) grade iv (icd-0 9440/3), characterized by high cellularity, marked atypical features, frequent mitotic activity, extensive vascular proliferation, and significant areas of necrosis.

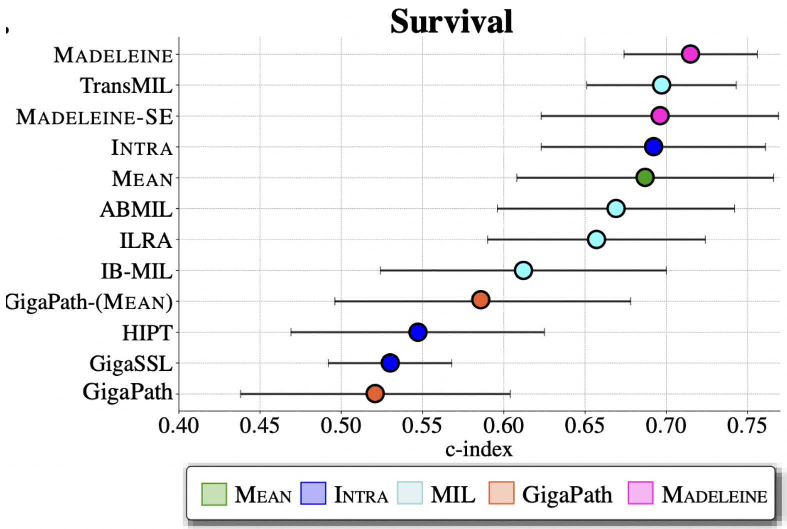
PRISM (generated)

Diagnosis: High-grade myxofibrosarcoma in left frontal brain.

MADELINE: Contrasting HE with IHCs, Special Stains



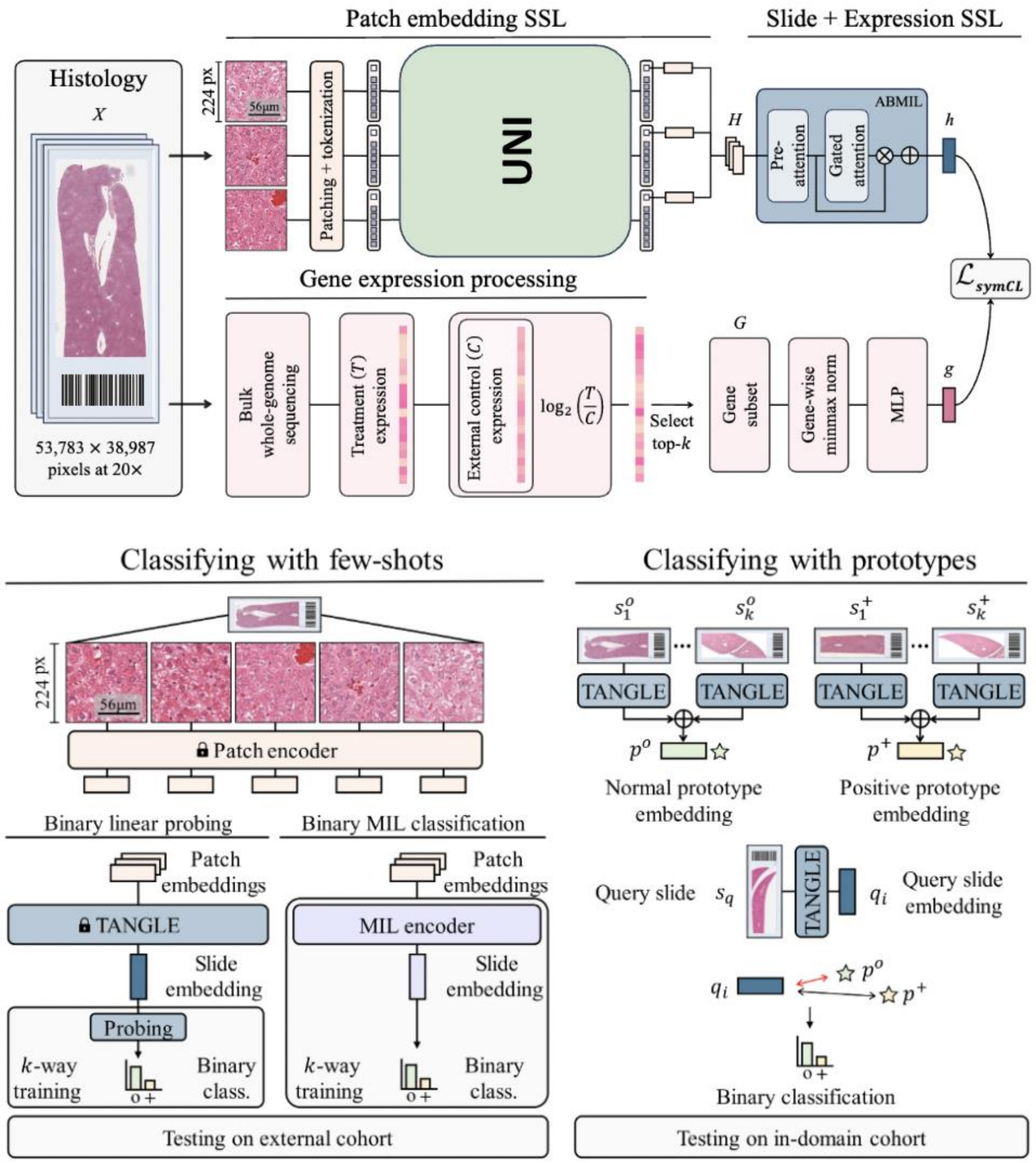

IHC quantification


Survival

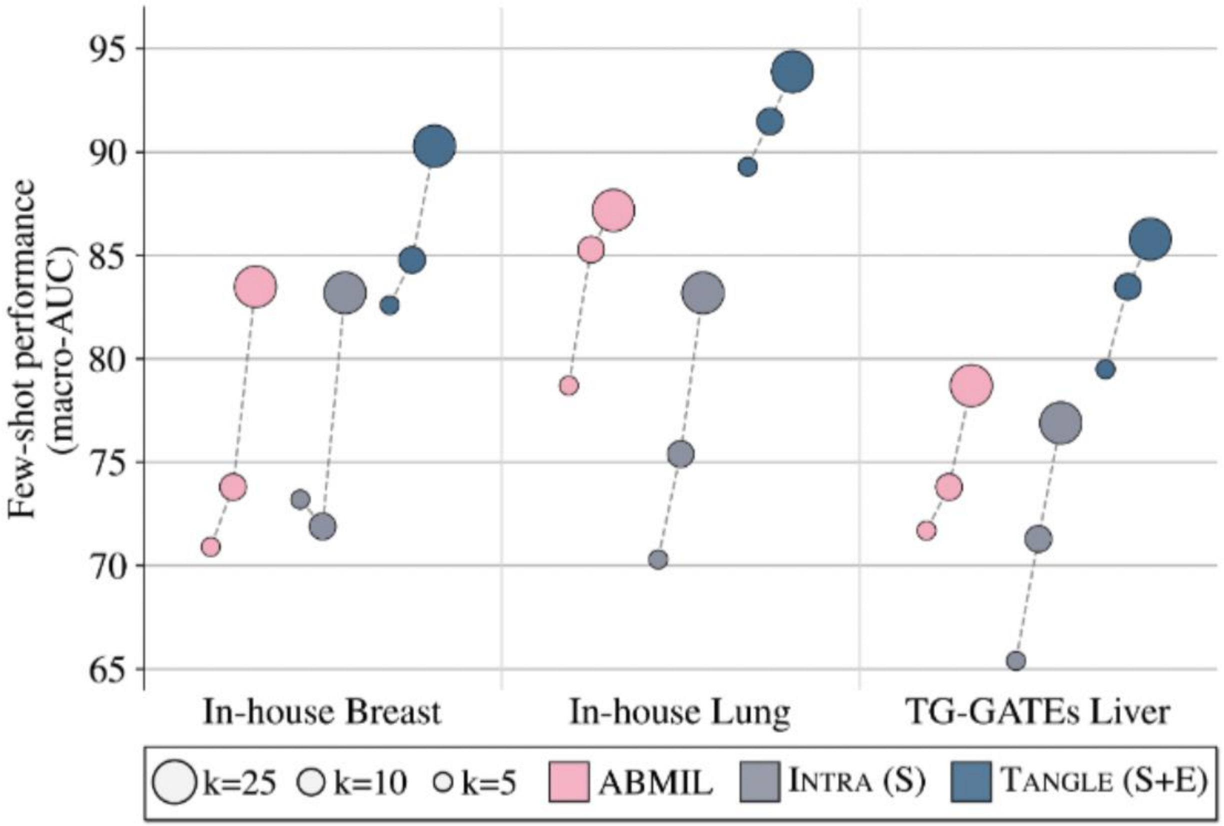
(ECCV, 2024)

TANGLE: A Slide-Level Foundation Model with H&E + Transcriptomics

Slide-Level Pretraining

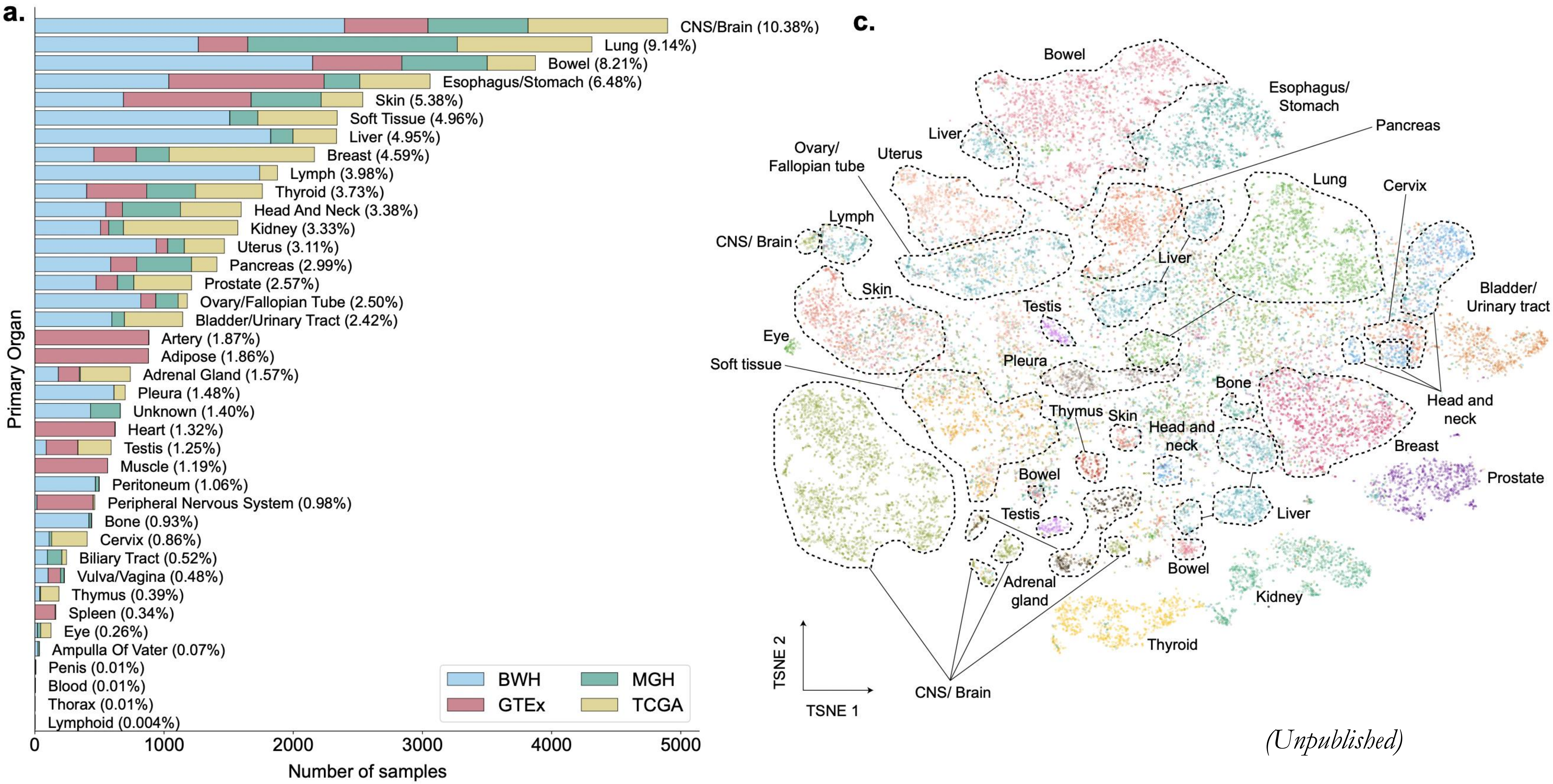


Few-Shot Slide Classification

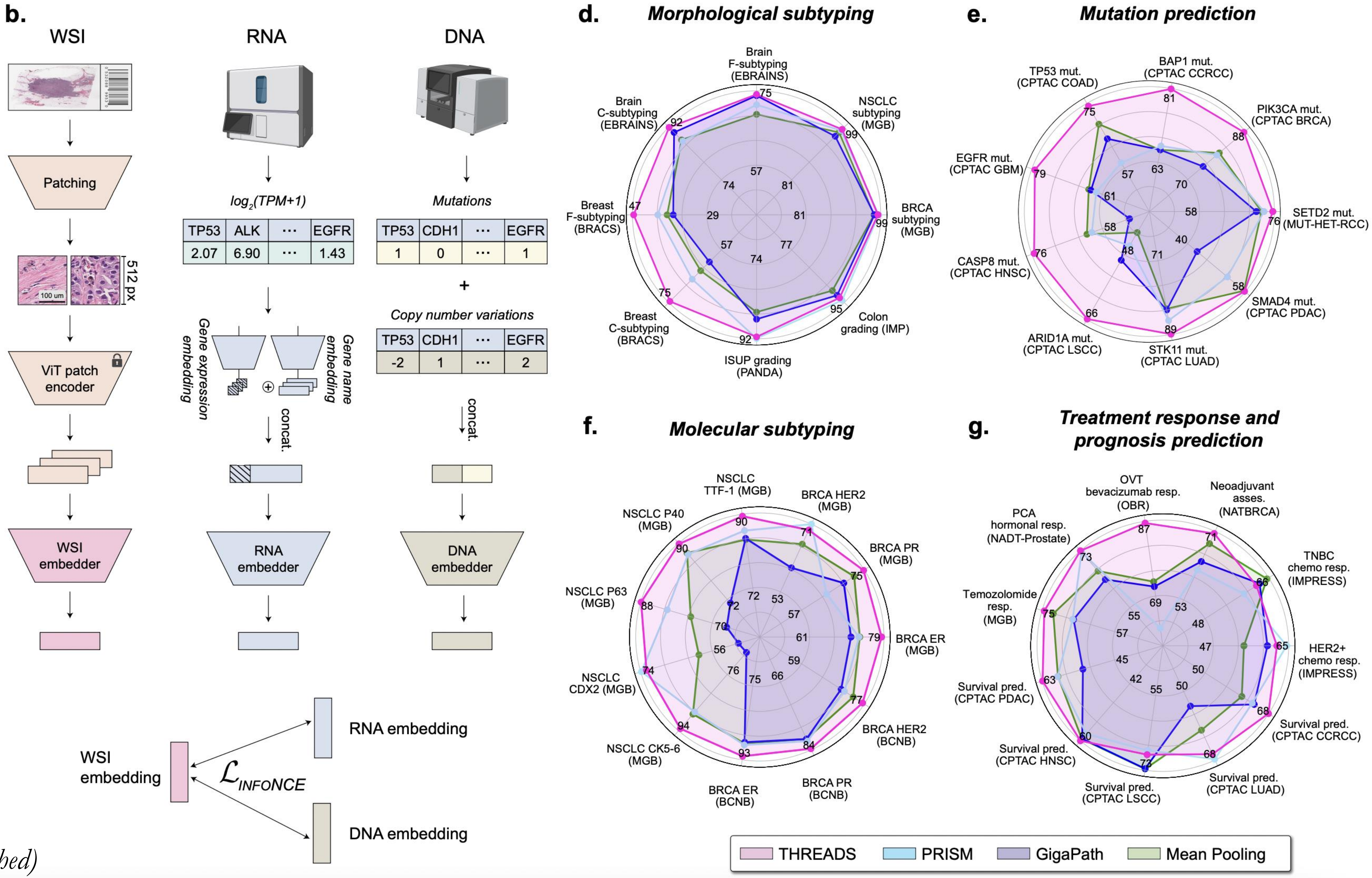


(CVPR, 2024)

THREADS: A contrastive foundation model with Histology + Genomics

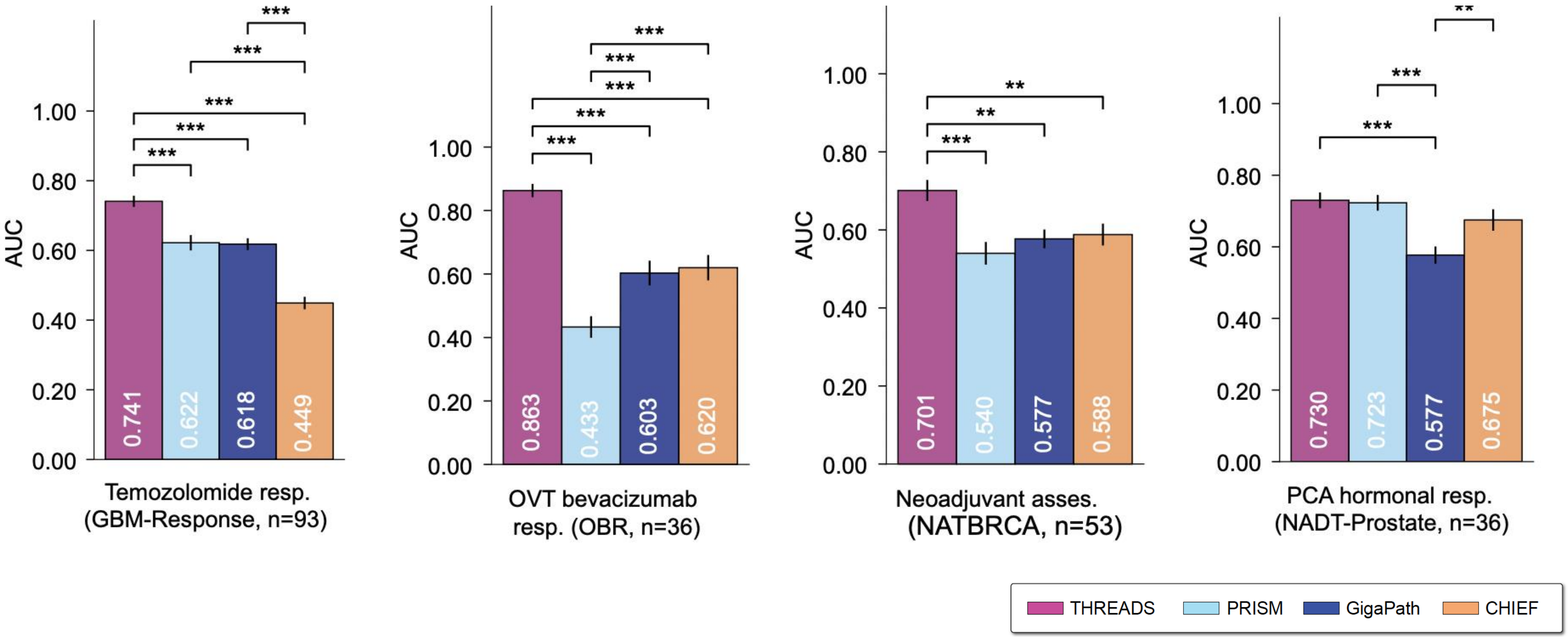


(Unpublished)



(Unpublished)

Treatment Response Tasks using THREADS

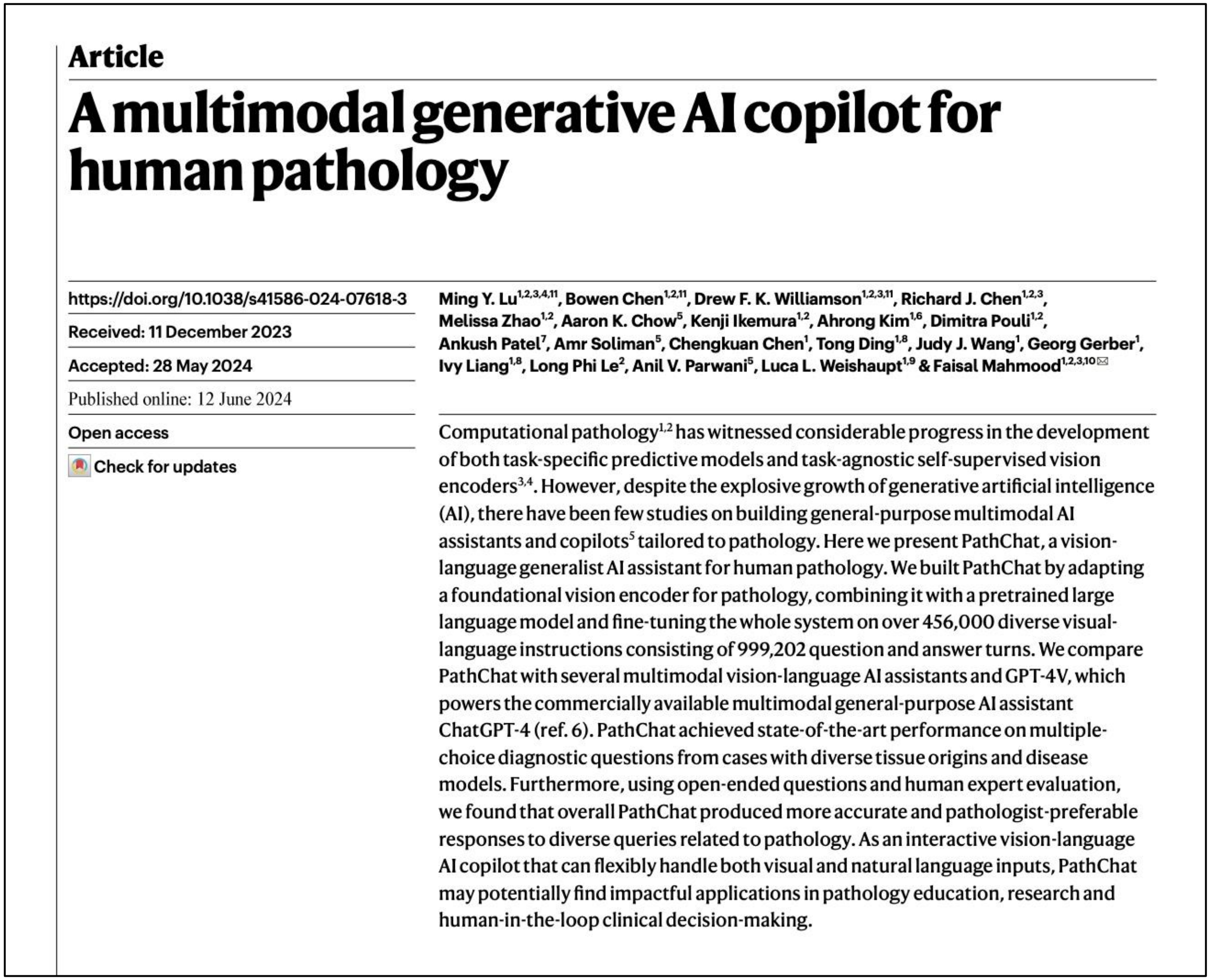


(Unpublished)

Generative AI for Pathology

What do we need to build a universal multimodal chatbot for anatomic pathology?

- **A visual centric pathology foundation model.**
- **A vision-language foundation model.**
- **A large instruction dataset using with pathology images, questions and responses.**
- **Robust evaluation.**



Article

A multimodal generative AI copilot for human pathology

https://doi.org/10.1038/s41586-024-07618-3

Received: 11 December 2023

Accepted: 28 May 2024

Published online: 12 June 2024

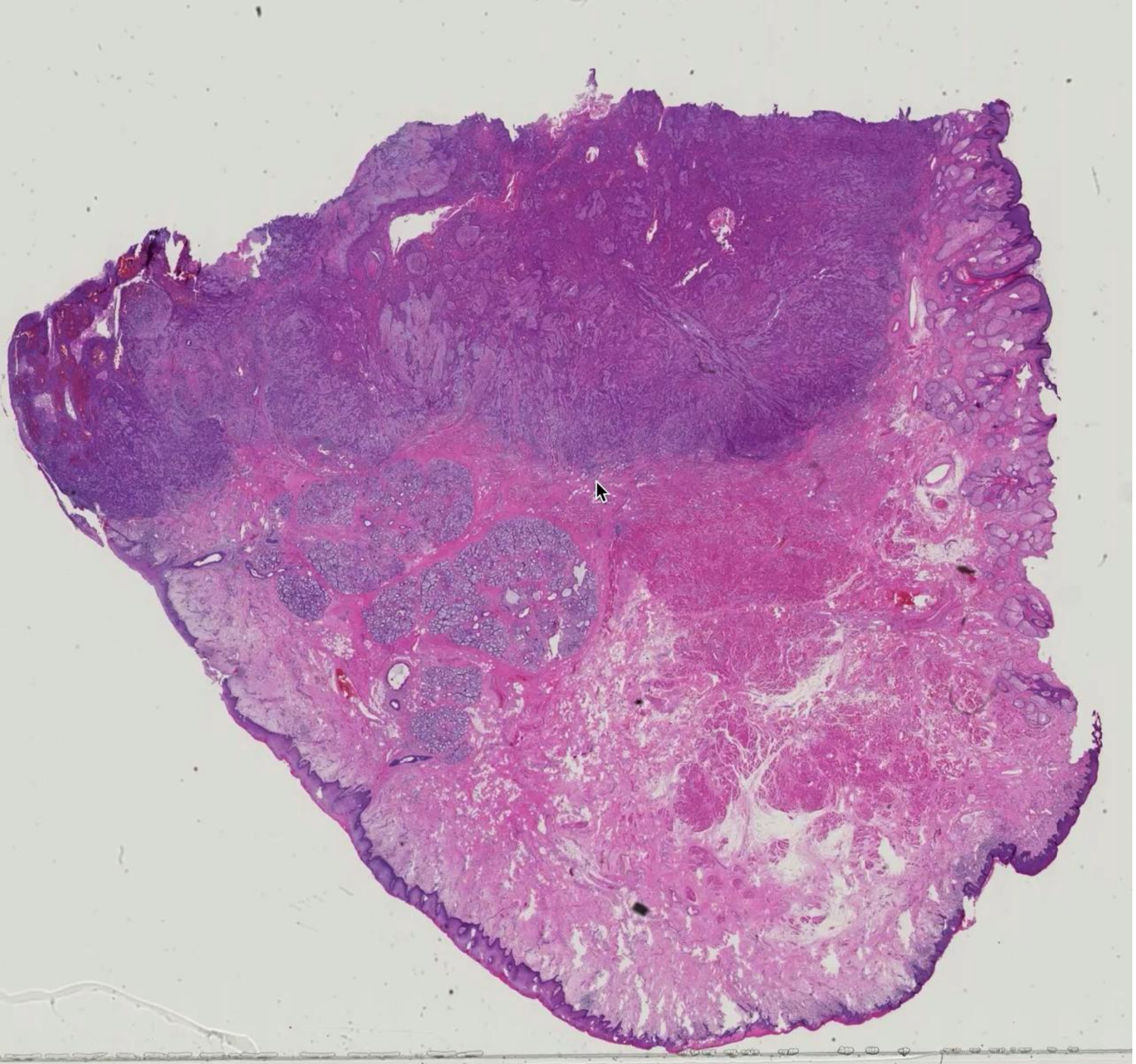
Open access

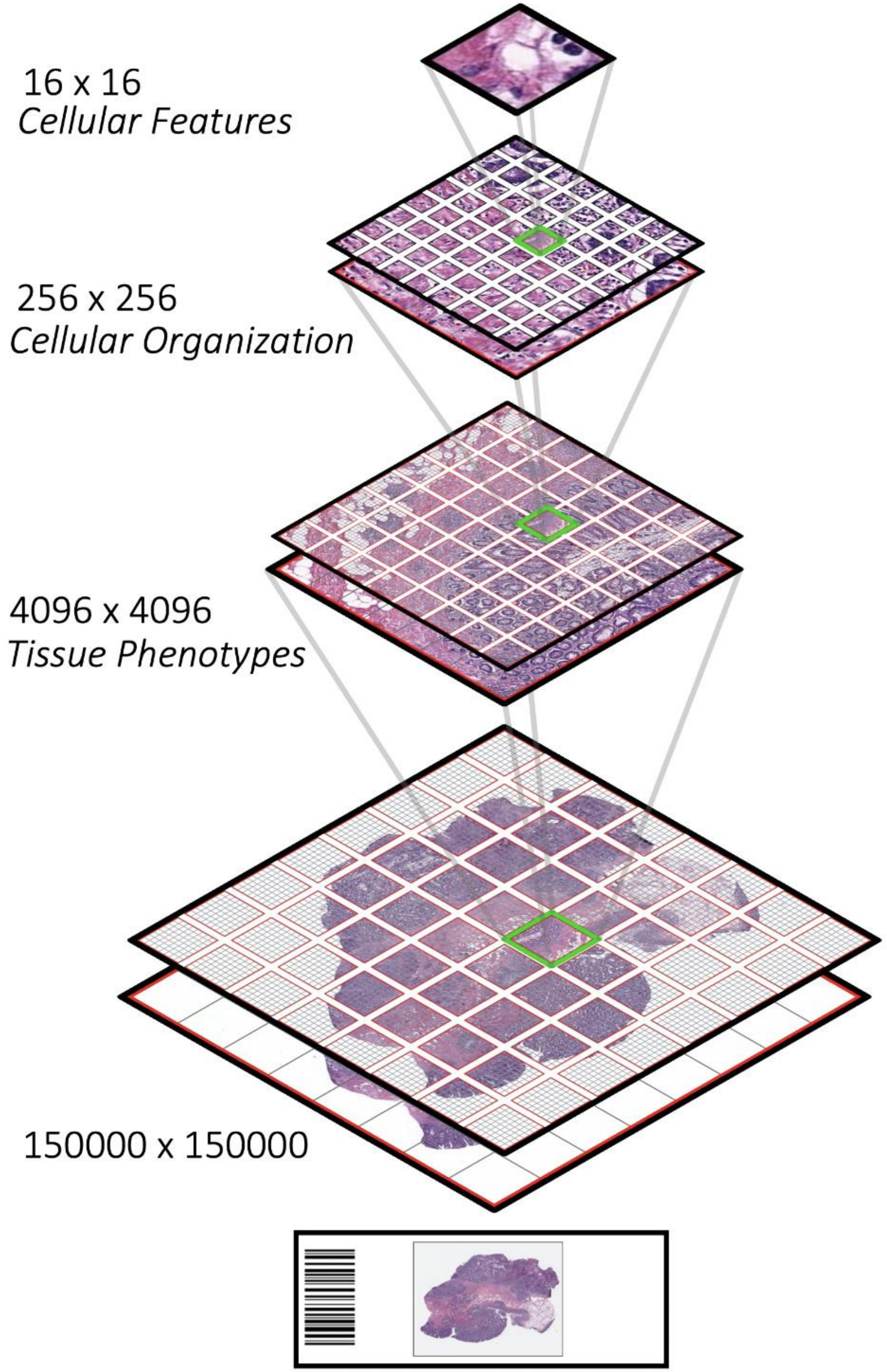
Check for updates

Ming Y. Lu[1,2,3,4,11], Bowen Chen[1,2,11], Drew F. K. Williamson[1,2,3,11], Richard J. Chen[1,2,3], Melissa Zhao[1,2], Aaron K. Chow[5], Kenji Ikemura[1,2], Ahrong Kim[1,6], Dimitra Pouli[1,2], Ankush Patel[7], Amr Soliman[5], Chengkuan Chen[1], Tong Ding[1,8], Judy J. Wang[1], Georg Gerber[1], Ivy Liang[1,8], Long Phi Le[2], Anil V. Parwani[5], Luca L. Weishaupt[1,9] & Faisal Mahmood[1,2,3,10] ✉

Computational pathology[1,2] has witnessed considerable progress in the development of both task-specific predictive models and task-agnostic self-supervised vision encoders[3,4]. However, despite the explosive growth of generative artificial intelligence (AI), there have been few studies on building general-purpose multimodal AI assistants and copilots[5] tailored to pathology. Here we present PathChat, a vision-language generalist AI assistant for human pathology. We built PathChat by adapting a foundational vision encoder for pathology, combining it with a pretrained large language model and fine-tuning the whole system on over 456,000 diverse visual-language instructions consisting of 999,202 question and answer turns. We compare PathChat with several multimodal vision-language AI assistants and GPT-4V, which powers the commercially available multimodal general-purpose AI assistant ChatGPT-4 (ref. 6). PathChat achieved state-of-the-art performance on multiple-choice diagnostic questions from cases with diverse tissue origins and disease models. Furthermore, using open-ended questions and human expert evaluation, we found that overall PathChat produced more accurate and pathologist-preferable responses to diverse queries related to pathology. As an interactive vision-language AI copilot that can flexibly handle both visual and natural language inputs, PathChat may potentially find impactful applications in pathology education, research and human-in-the-loop clinical decision-making.

(Nature, 2024)



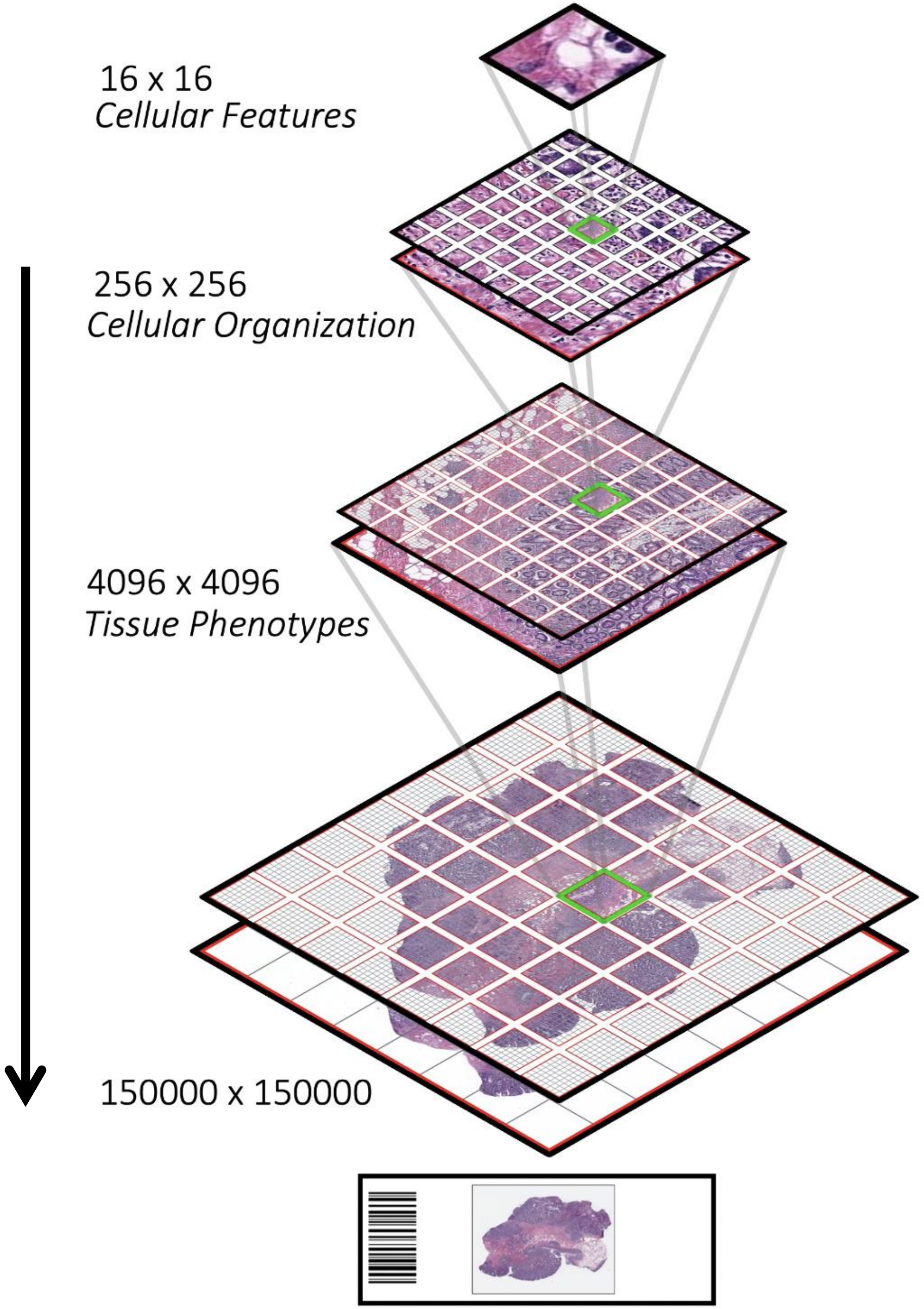


16 x 16
Cellular Features

256 x 256
Cellular Organization

4096 x 4096
Tissue Phenotypes

150000 x 150000



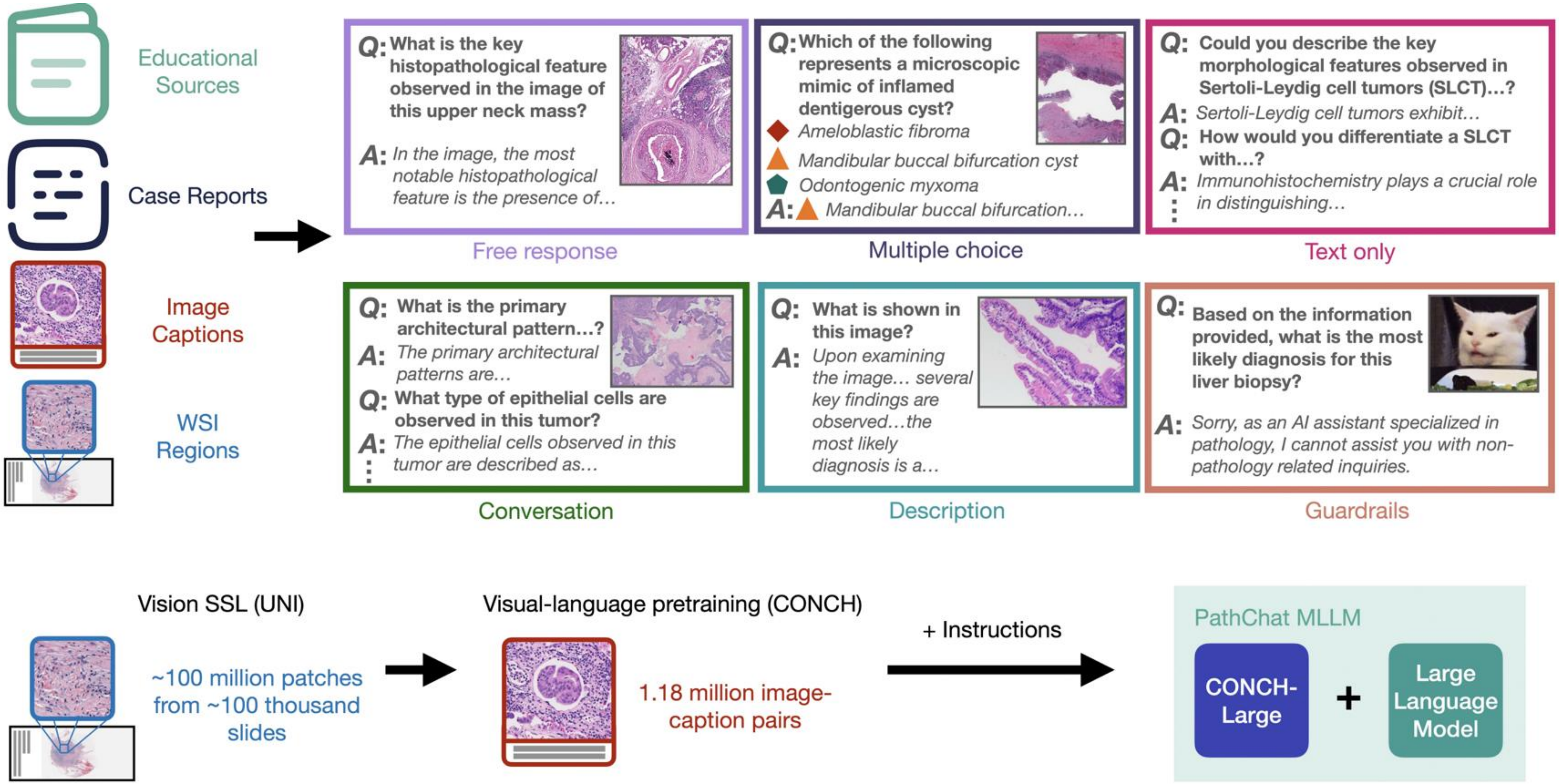
Fine grained understanding of pathology regions at the cellular leads to slide level and patient level descriptions.

Pathology reports are at the level of the slide or patient and don't have fine grained morphologic details.

We need fine grained morphologic details at the level of cellular organization and tissue phenotypes to have a close relation between text-image pairs which can be used to train PathChat.

(Nature, 2024)

Building PathChat



Q: What is the key histopathological feature observed in the image of this upper neck mass?

A: *In the image, the most notable histopathological feature is the presence of…*

Free response

Q: Which of the following represents a microscopic mimic of inflamed dentigerous cyst?

◆ *Ameloblastic fibroma*

▲ *Mandibular buccal bifurcation cyst*

⬟ *Odontogenic myxoma*

A: ▲ *Mandibular buccal bifurcation…*

Multiple choice

Q: Could you describe the key morphological features observed in Sertoli-Leydig cell tumors (SLCT)…?

A: *Sertoli-Leydig cell tumors exhibit…*

Q: How would you differentiate a SLCT with…?

A: *Immunohistochemistry plays a crucial role in distinguishing…*

⋮

Text only

Q: What is the primary architectural pattern…?

A: *The primary architectural patterns are…*

Q: What type of epithelial cells are observed in this tumor?

A: *The epithelial cells observed in this tumor are described as…*

⋮

Conversation

Q: What is shown in this image?

A: *Upon examining the image… several key findings are observed…the most likely diagnosis is a…*

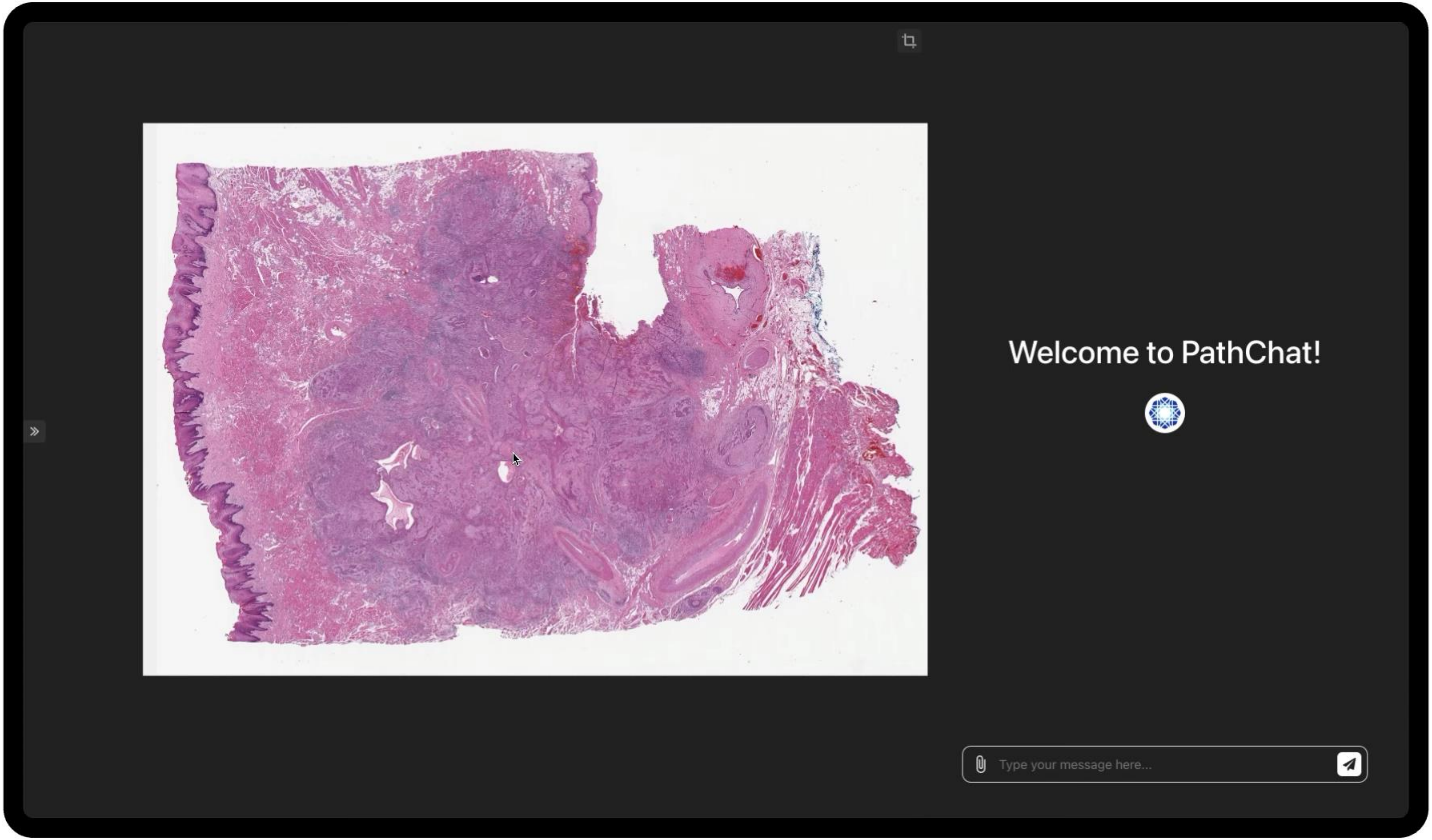
Description

Q: Based on the information provided, what is the most likely diagnosis for this liver biopsy?

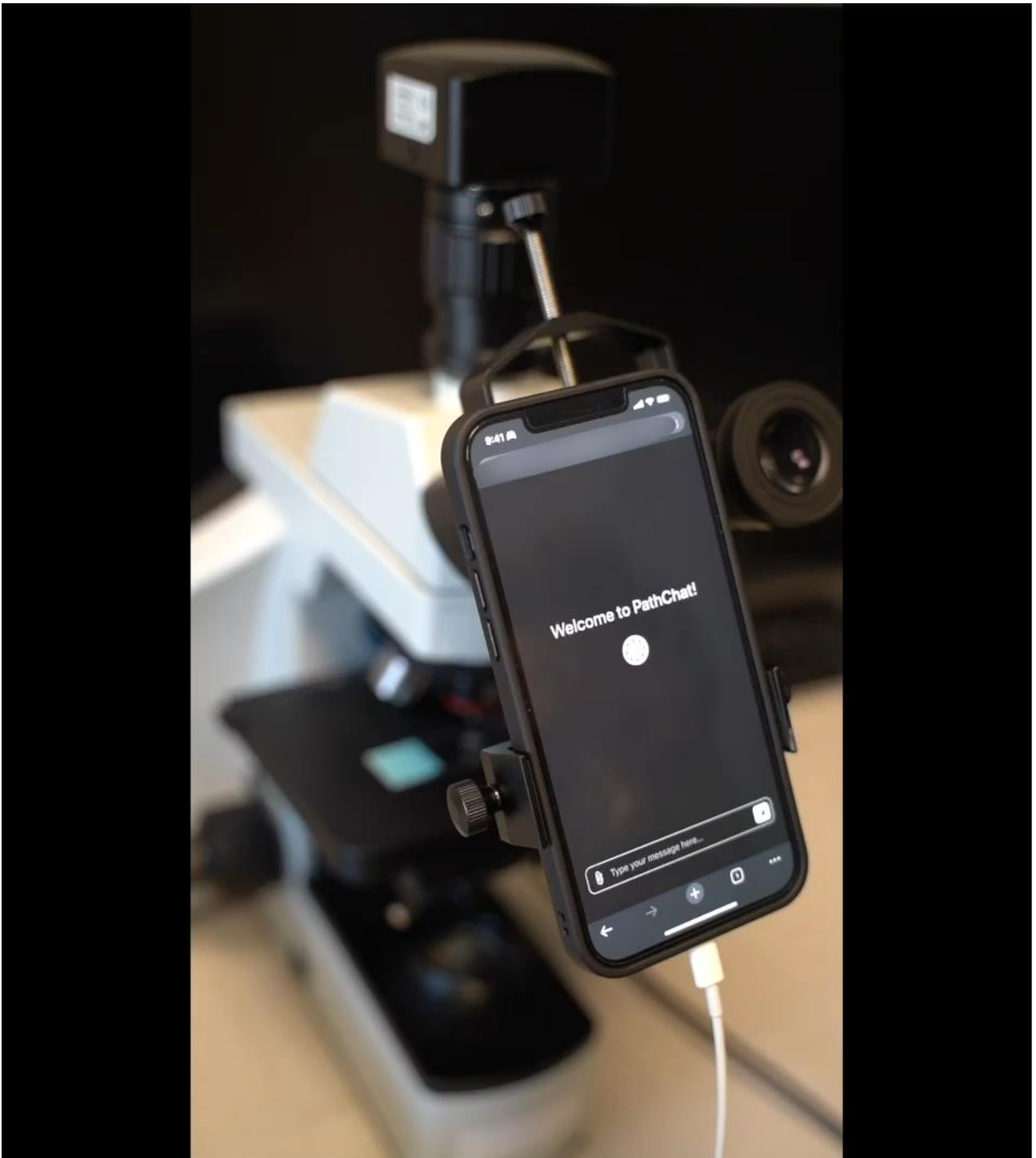
A: *Sorry, as an AI assistant specialized in pathology, I cannot assist you with non-pathology related inquiries.*

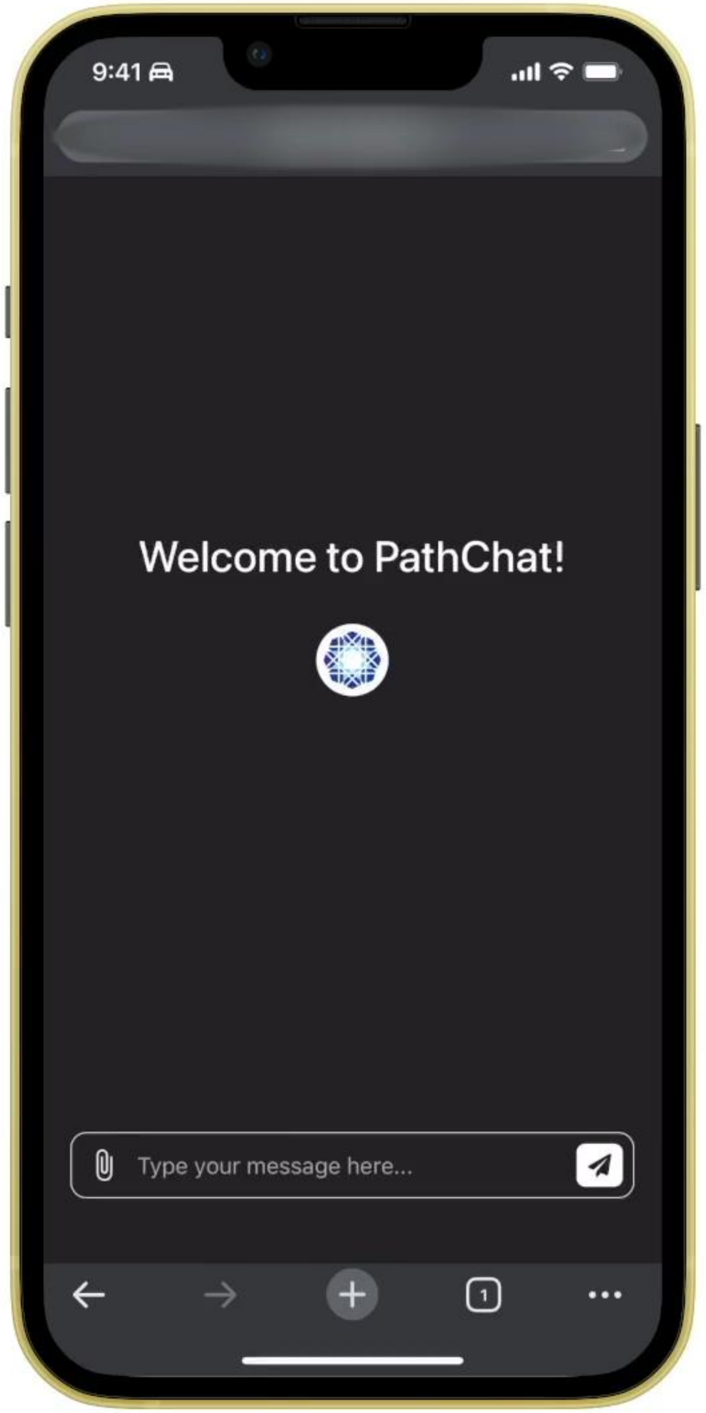
Guardrails

Vision SSL (UNI)
~100 million patches from ~100 thousand slides
Visual-language pretraining (CONCH)
1.18 million image-caption pairs
+ Instructions
PathChat MLLM
CONCH-Large
+
Large Language Model

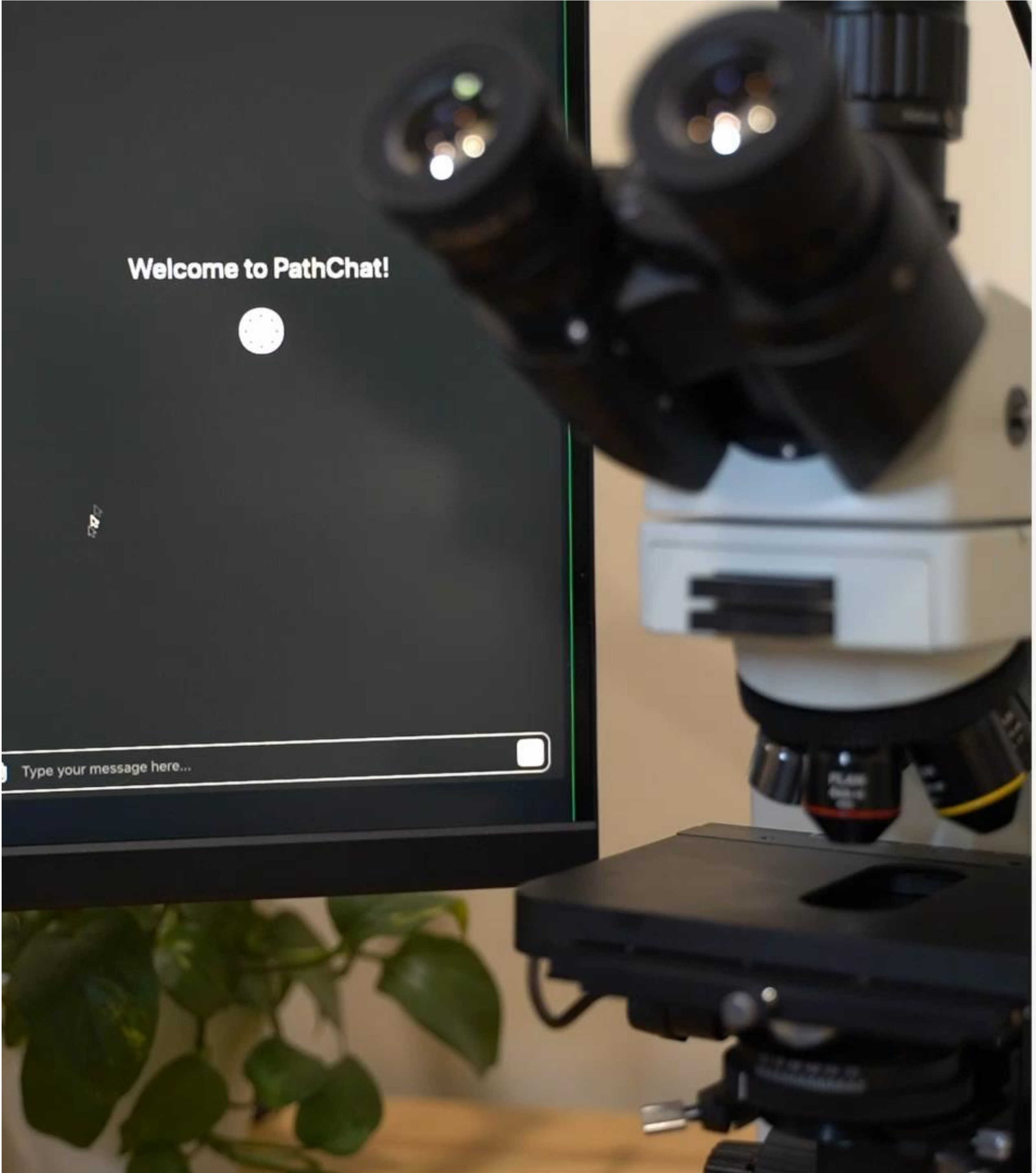


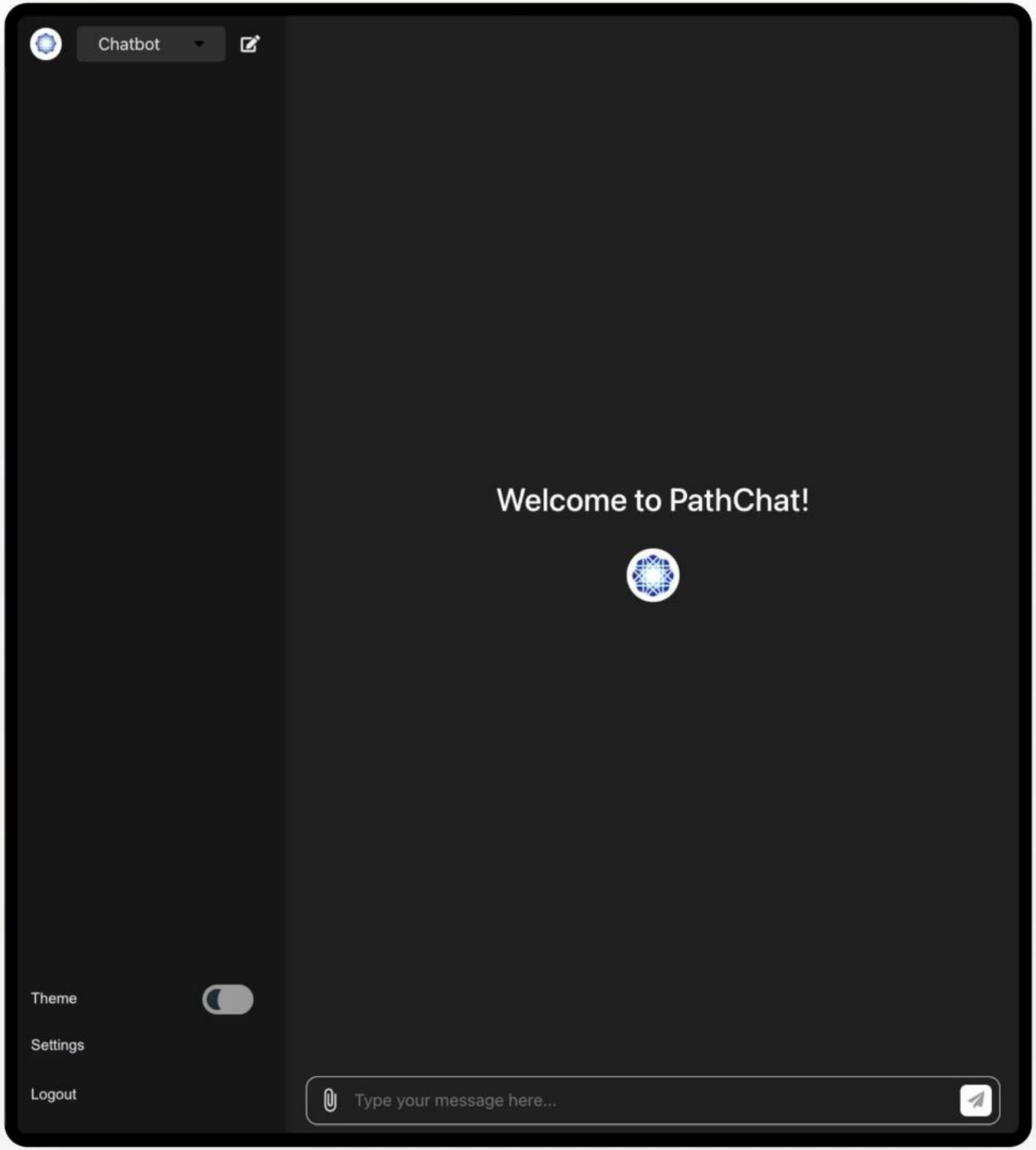
Welcome to PathChat!



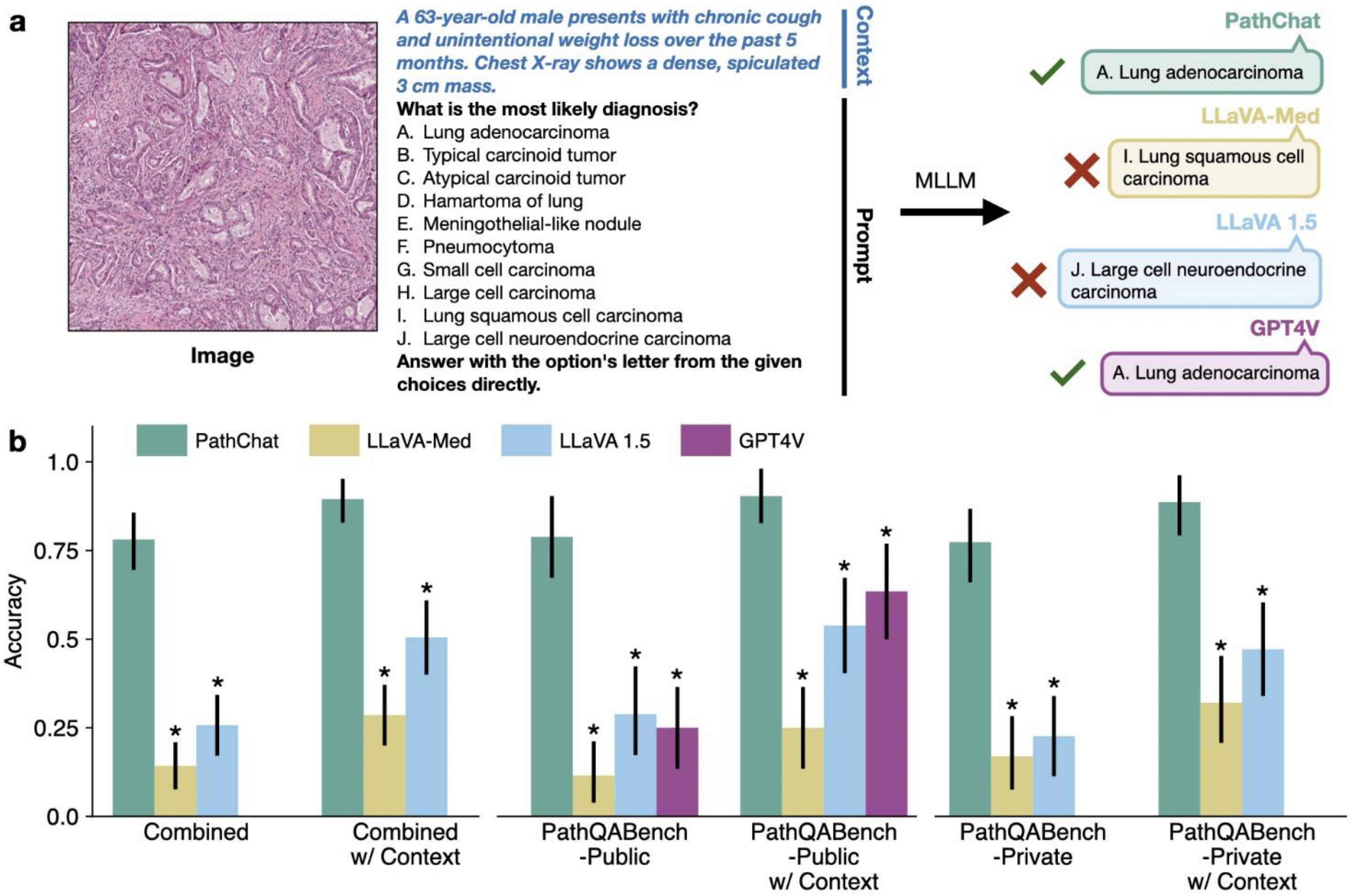


Welcome to PathChat!


Welcome to PathChat!
Type your message here...


Chatbot
Welcome to PathChat!
Theme
Settings
Logout
Type your message here...

(Nature, 2024)



A 63-year-old male presents with chronic cough and unintentional weight loss over the past 5 months. Chest X-ray shows a dense, spiculated 3 cm mass.

What is the most likely diagnosis?

A. Lung adenocarcinoma
B. Typical carcinoid tumor
C. Atypical carcinoid tumor
D. Hamartoma of lung
E. Meningothelial-like nodule
F. Pneumocytoma
G. Small cell carcinoma
H. Large cell carcinoma
I. Lung squamous cell carcinoma
J. Large cell neuroendocrine carcinoma

Answer with the option's letter from the given choices directly.

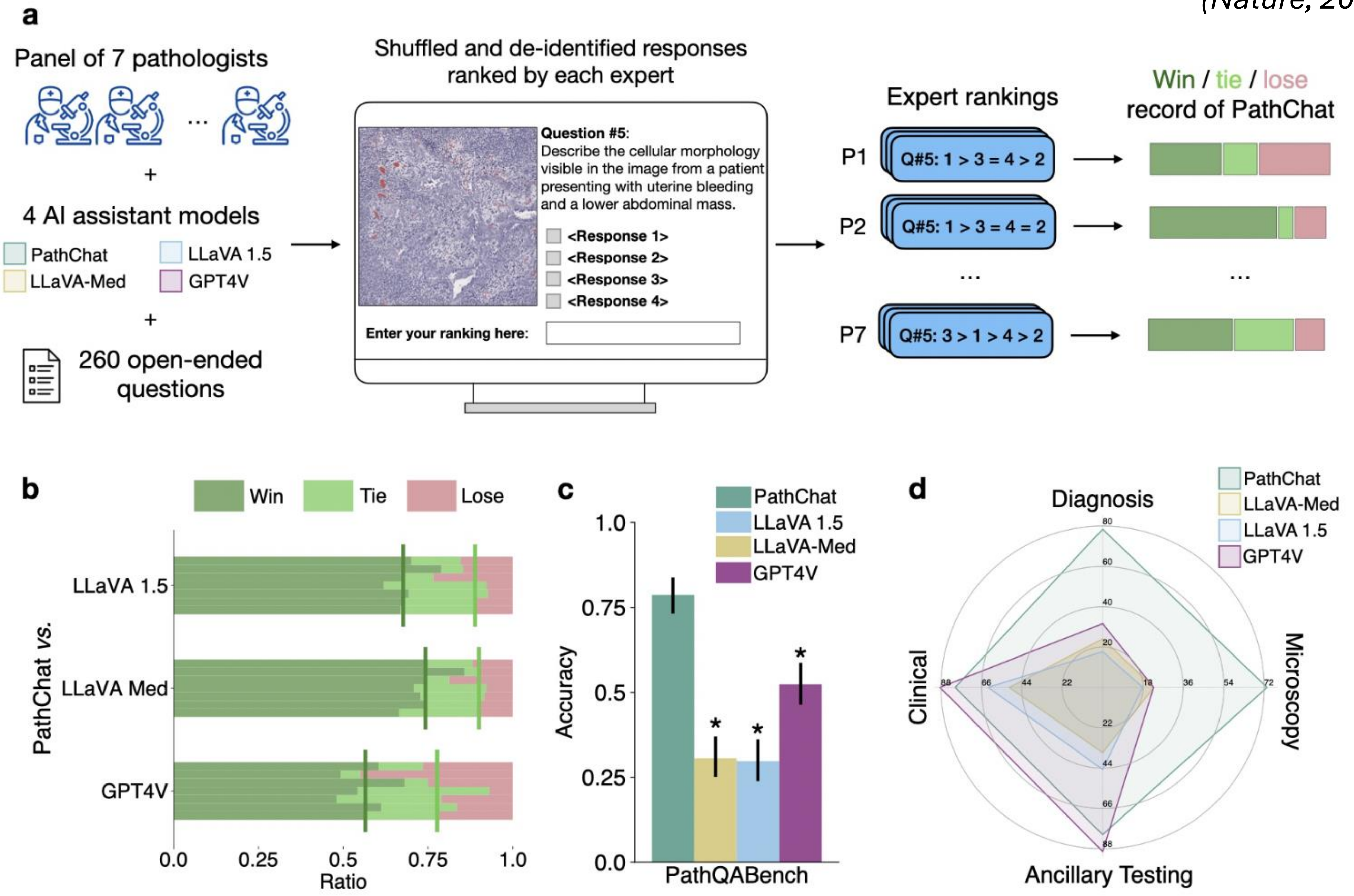
PathChat: ✓ A. Lung adenocarcinoma

LLaVA-Med: ✗ I. Lung squamous cell carcinoma

LLaVA 1.5: ✗ J. Large cell neuroendocrine carcinoma

GPT4V: ✓ A. Lung adenocarcinoma

(Nature, 2024)

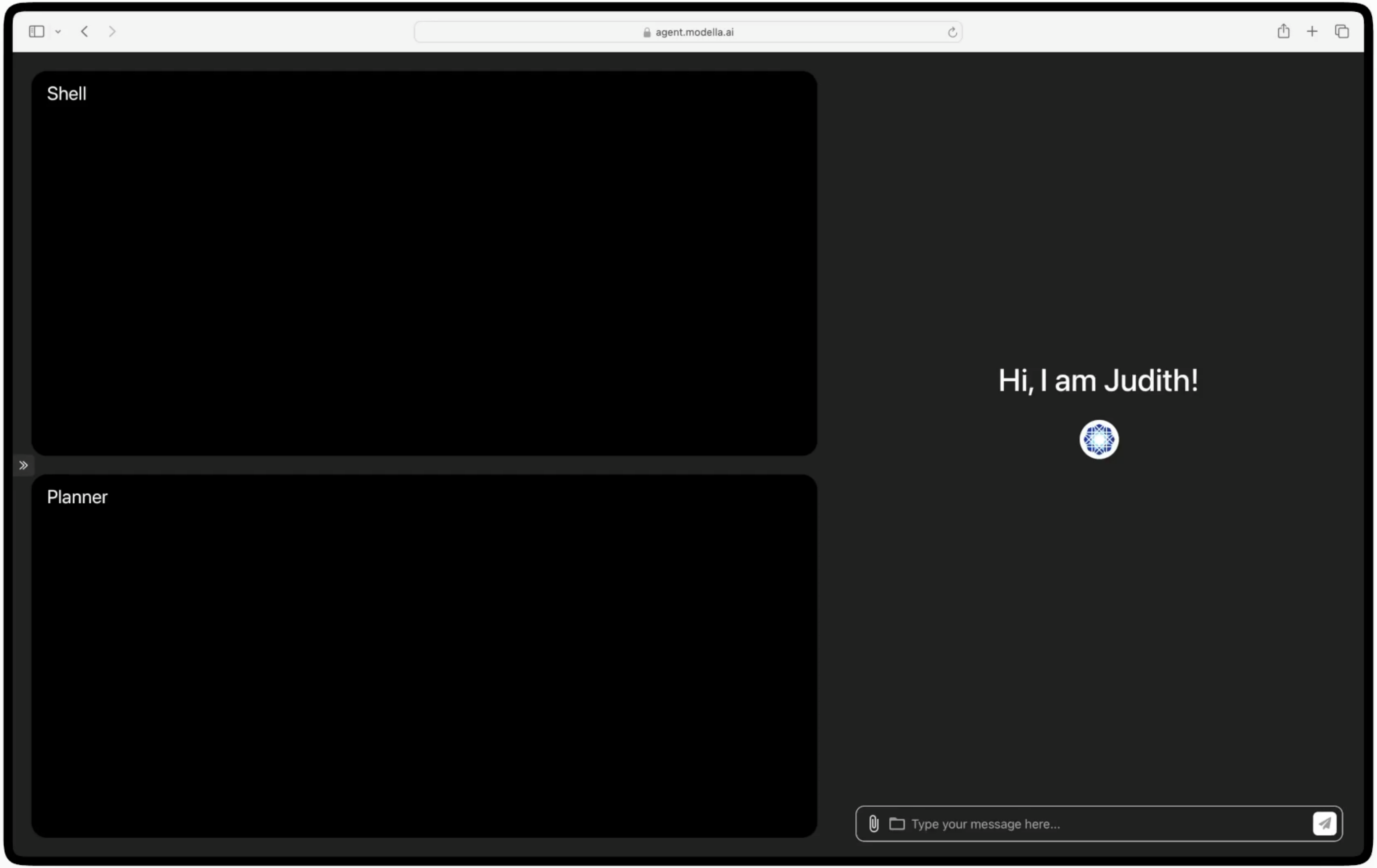


Question #5: Describe the cellular morphology visible in the image from a patient presenting with uterine bleeding and a lower abdominal mass.

AI Agent for Computational Pathology - Preview

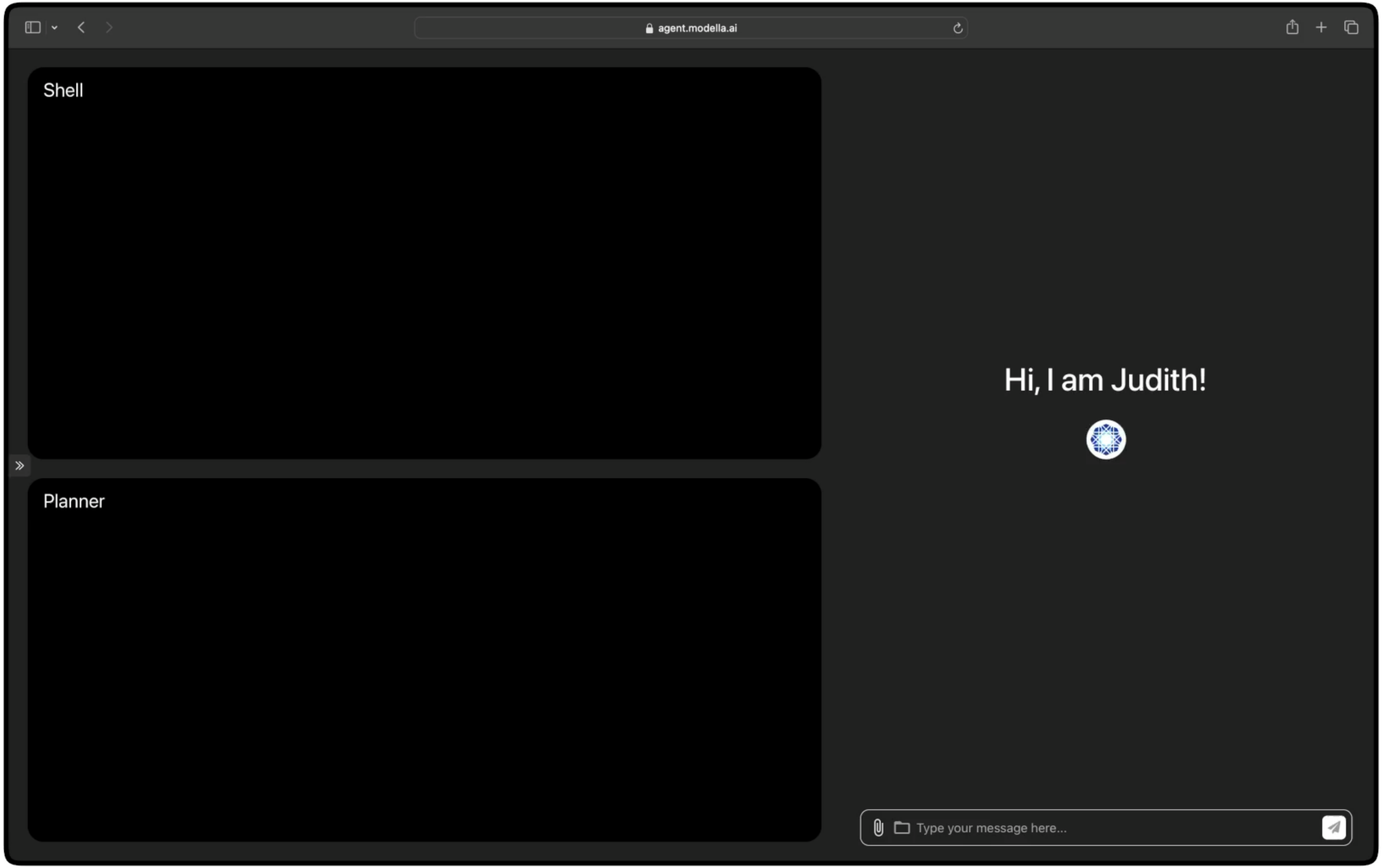
- AI agents do things for you!
- **What if AI agents could do all biomedical data analysis for you?**
- ***What if an AI agent could develop, assess, and explain AI models for pathology?***
- ***What if an AI agent could write code, run experiments and test hypothesis?***
- ***What if an AI agent could continuously run in the background attempting to find common morphologic features across patient cohorts and correlate with outcome?***

(Unpublished)



Hi, I am Judith!

Unpublished

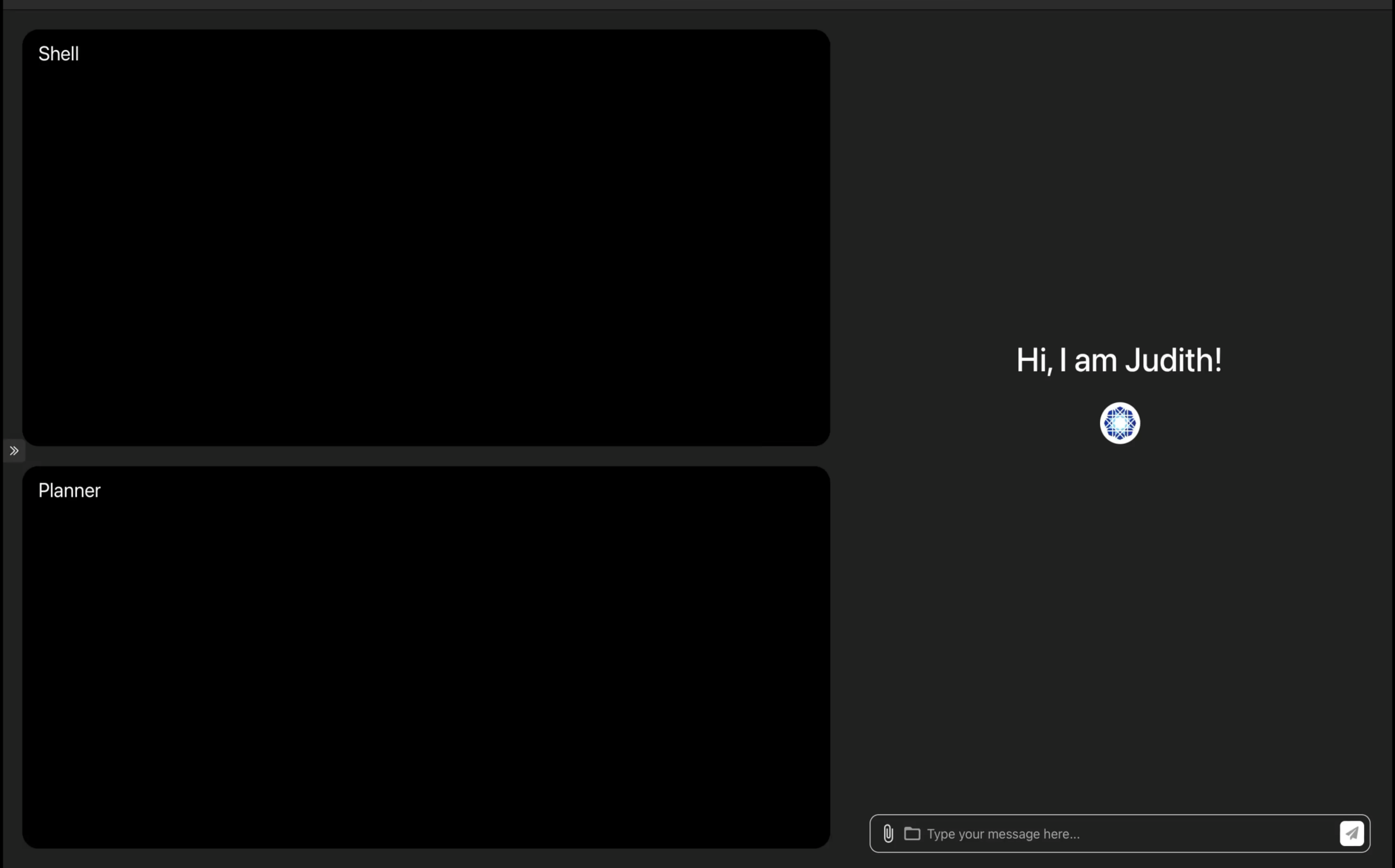


Shell

Planner

Hi, I am Judith!

Unpublished



Hi, I am Judith!

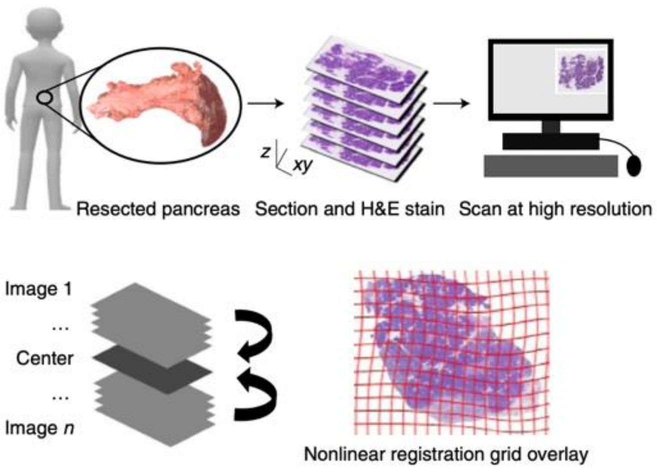
Motivation Transitioning from 2D to 3D pathology

Human tissue is inherently 3D

=> Current clinical practice - microscopic analysis of thinly-sliced 2D tissue section

Active development of 3D tissue imaging modality

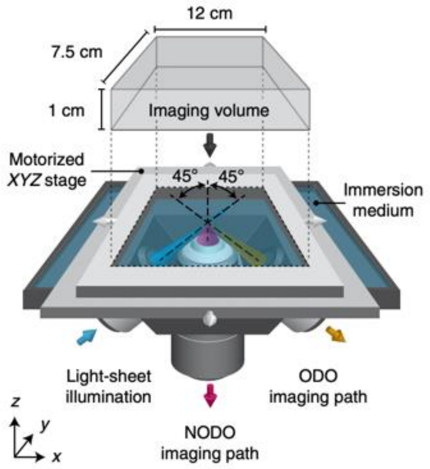
CODA – serial sectioning & registration



Kieman A. et al., *Nature Methods*, 2022

Microcomputed tomography (microCT)

Open-top light-sheet microscopy (OTLS)

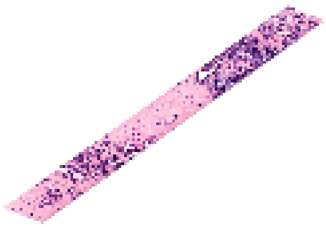


Glaser K. et al., *Nature Methods*, 2022

▶ Infeasible for pathologists to manually examine 3D data

▶ There does not exist AI pipeline to process the volumetric data

Whole volume > Portion of volume

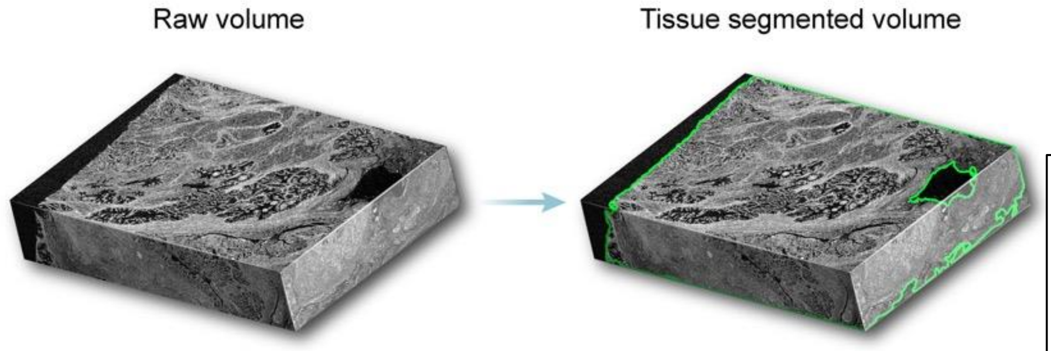


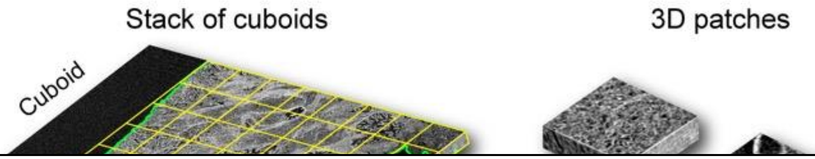
TriPath

Whole-block AI-based computational pipeline

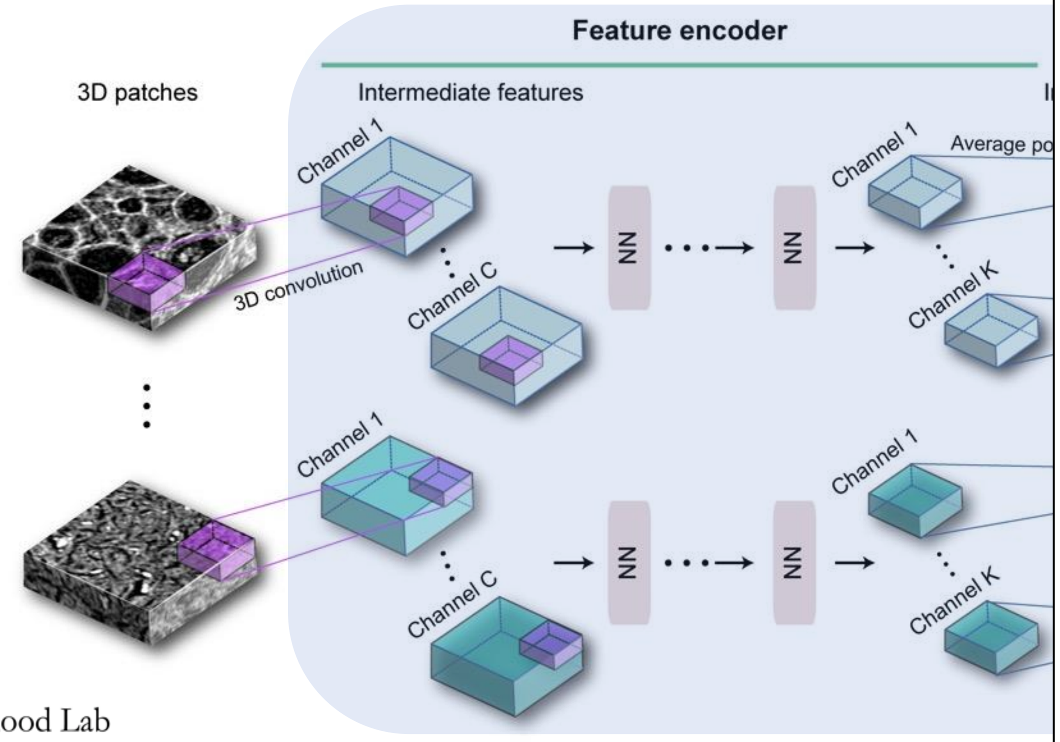
(Cell, 2024)

Data preprocessing

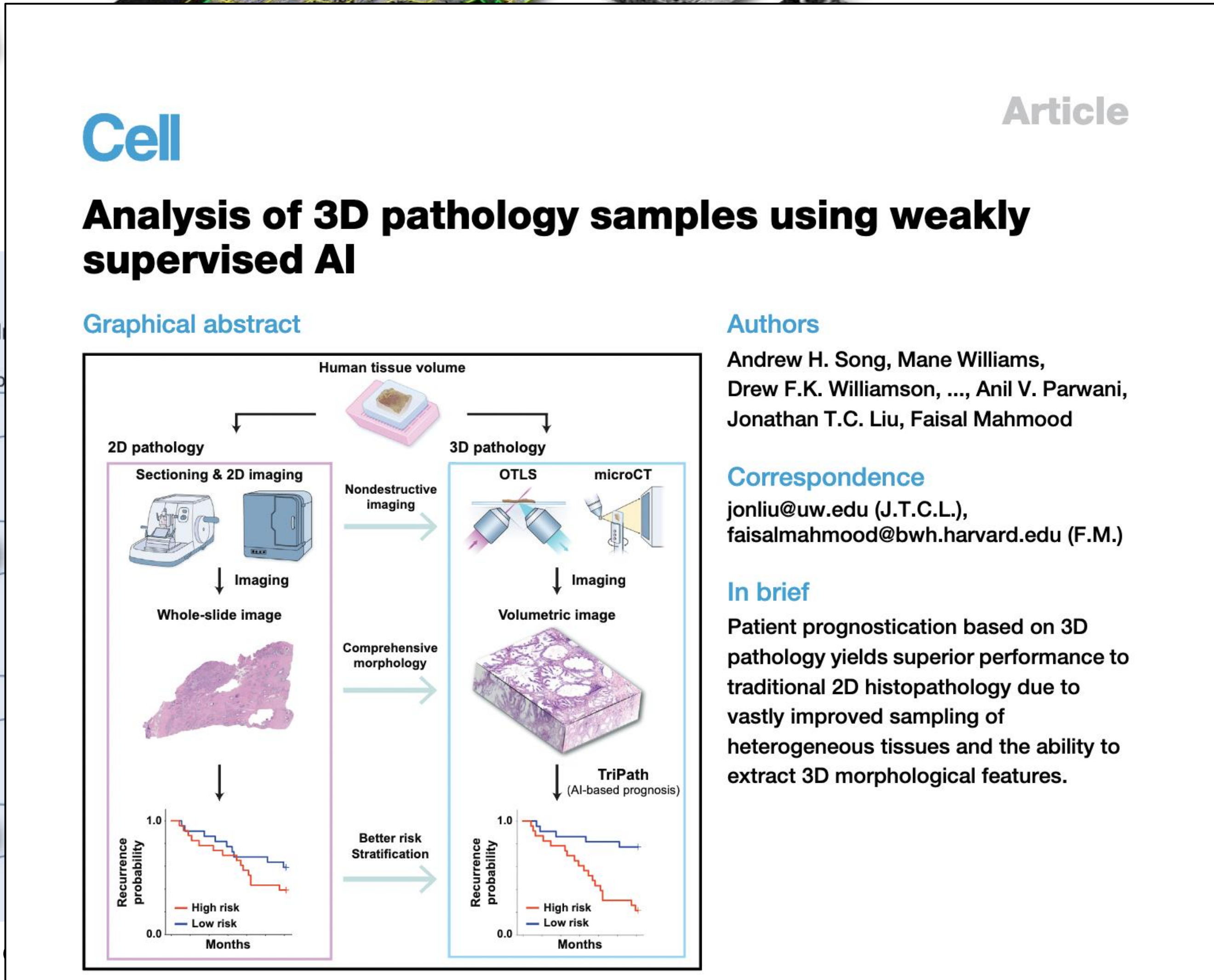




AI-based Computational processing



AI pipelin



Cell

Article

Analysis of 3D pathology samples using weakly supervised AI

Graphical abstract

Authors

Andrew H. Song, Mane Williams, Drew F.K. Williamson, ..., Anil V. Parwani, Jonathan T.C. Liu, Faisal Mahmood

Correspondence

jonliu@uw.edu (J.T.C.L.), faisalmahmood@bwh.harvard.edu (F.M.)

In brief

Patient prognostication based on 3D pathology yields superior performance to traditional 2D histopathology due to vastly improved sampling of heterogeneous tissues and the ability to extract 3D morphological features.

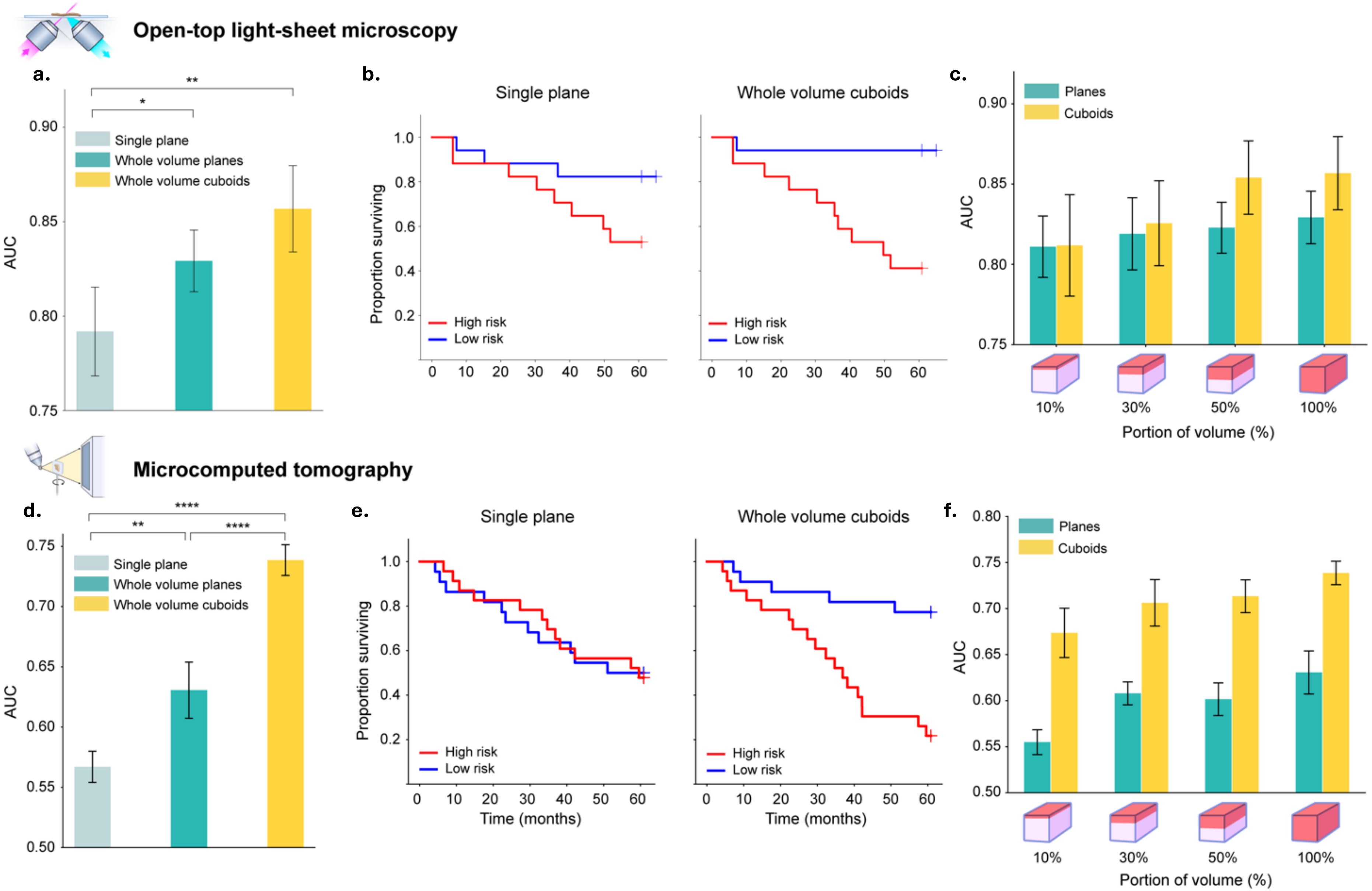
Performance

(Cell, 2024)

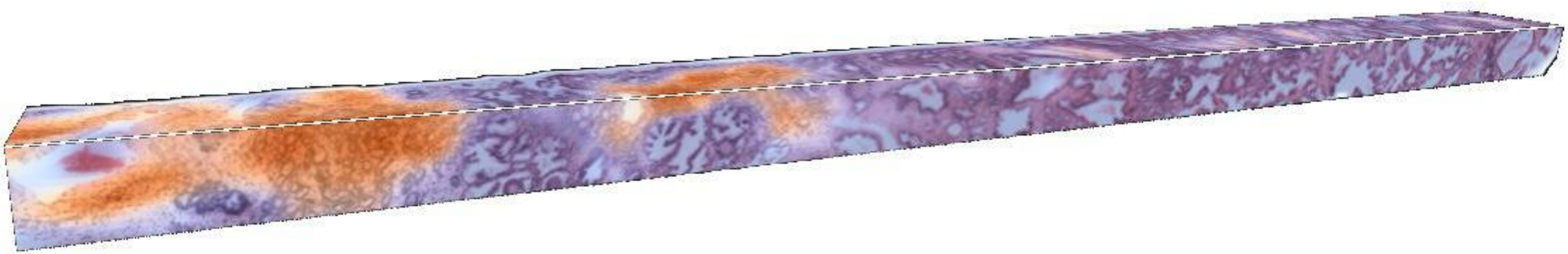
🟦 **Single**
(single 2D slice + 2D AI)

🟩 **Whole volume planes**
(Whole volume + 2D AI)

🟨 **Whole volume cuboids**
(Whole volume + 3D AI)

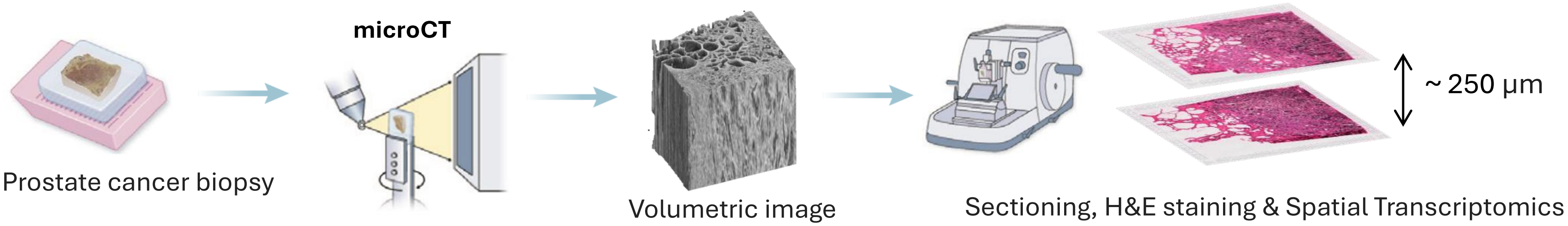


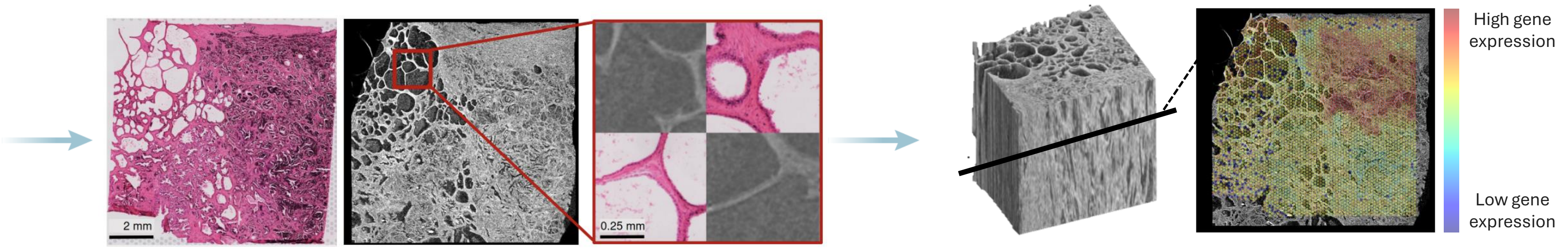
(Cell, 2024)



AI-driven 3D Spatial Transcriptomics (?)

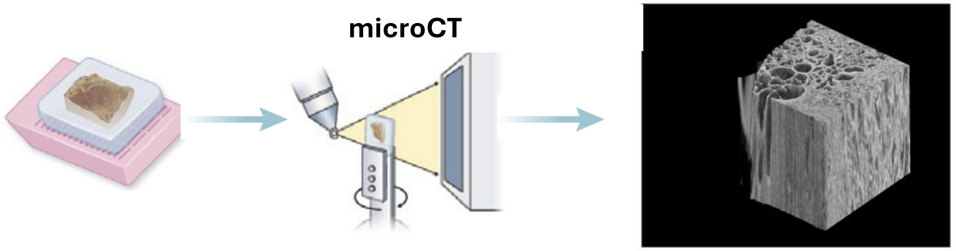
Generating the training/validation data ...





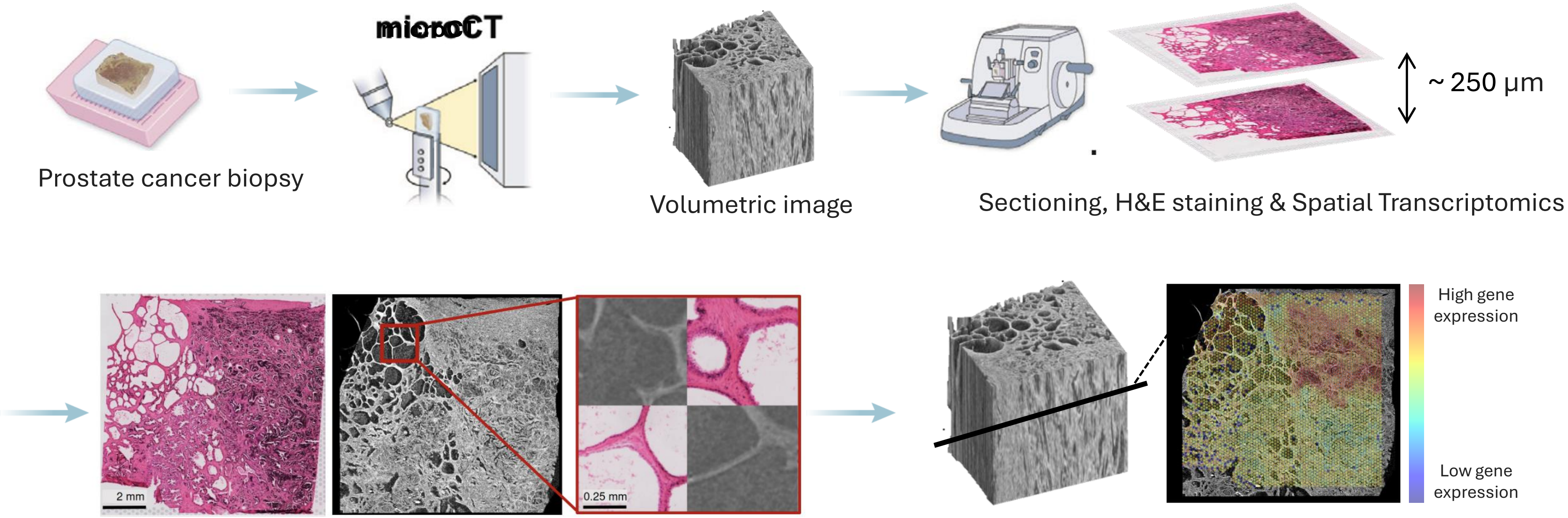
2D-3D registration for microCT – ST alignment

After model training...



AI-driven 3D Spatial Transcriptomics (?)

Generating the training/validation data ...



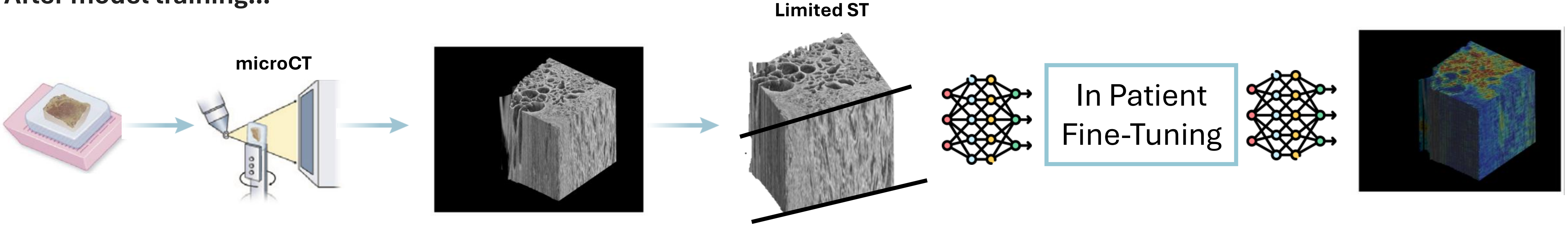
Prostate cancer biopsy

Volumetric image

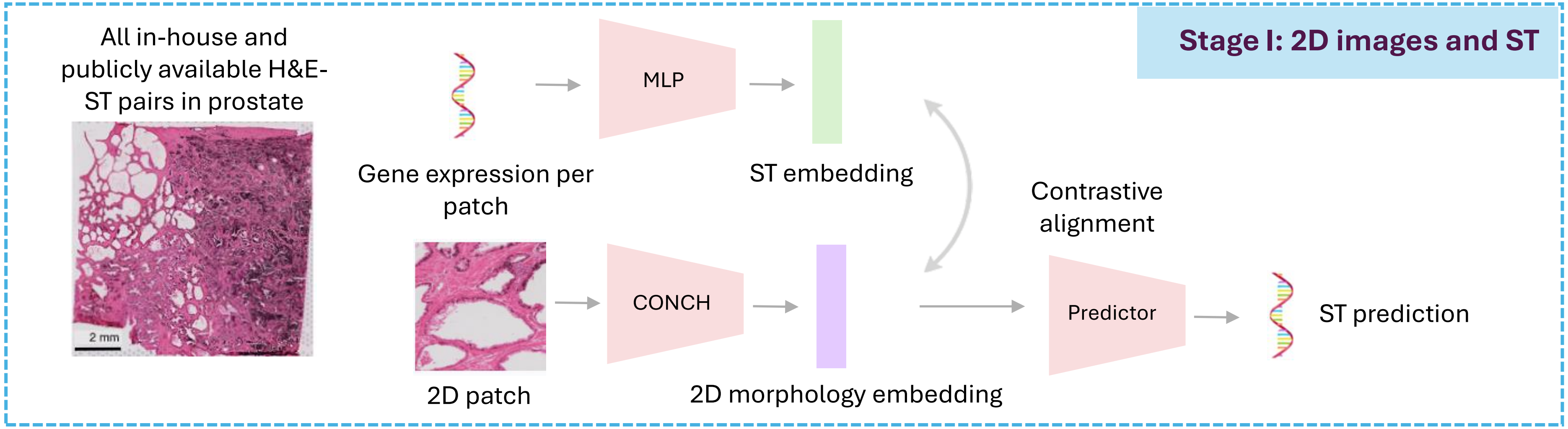
Sectioning, H&E staining & Spatial Transcriptomics

2D-3D registration for microCT – ST alignment

After model training...

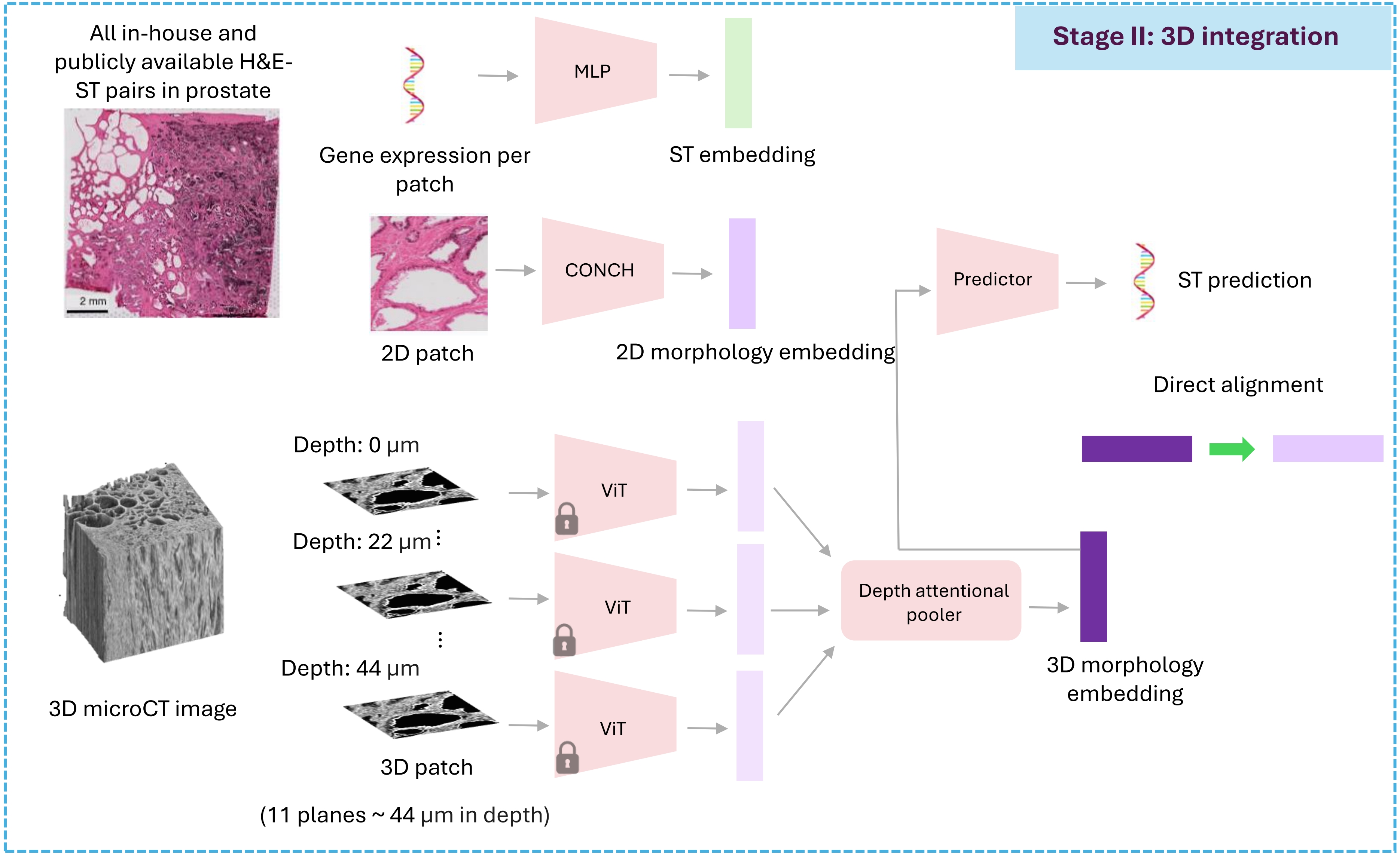


AI-driven 3D Spatial Transcriptomics (?)


Stage I: 2D images and ST
All in-house and publicly available H&E-ST pairs in prostate
2 mm
Gene expression per patch
MLP
ST embedding
Contrastive alignment
2D patch
CONCH
2D morphology embedding
Predictor
ST prediction

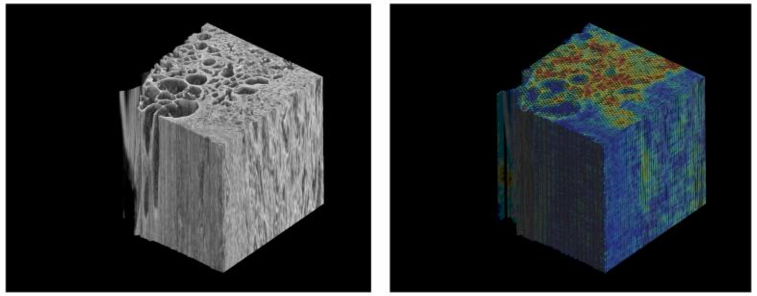
Mahmood Lab
AI for Pathology

AI-driven 3D Spatial Transcriptomics (?)



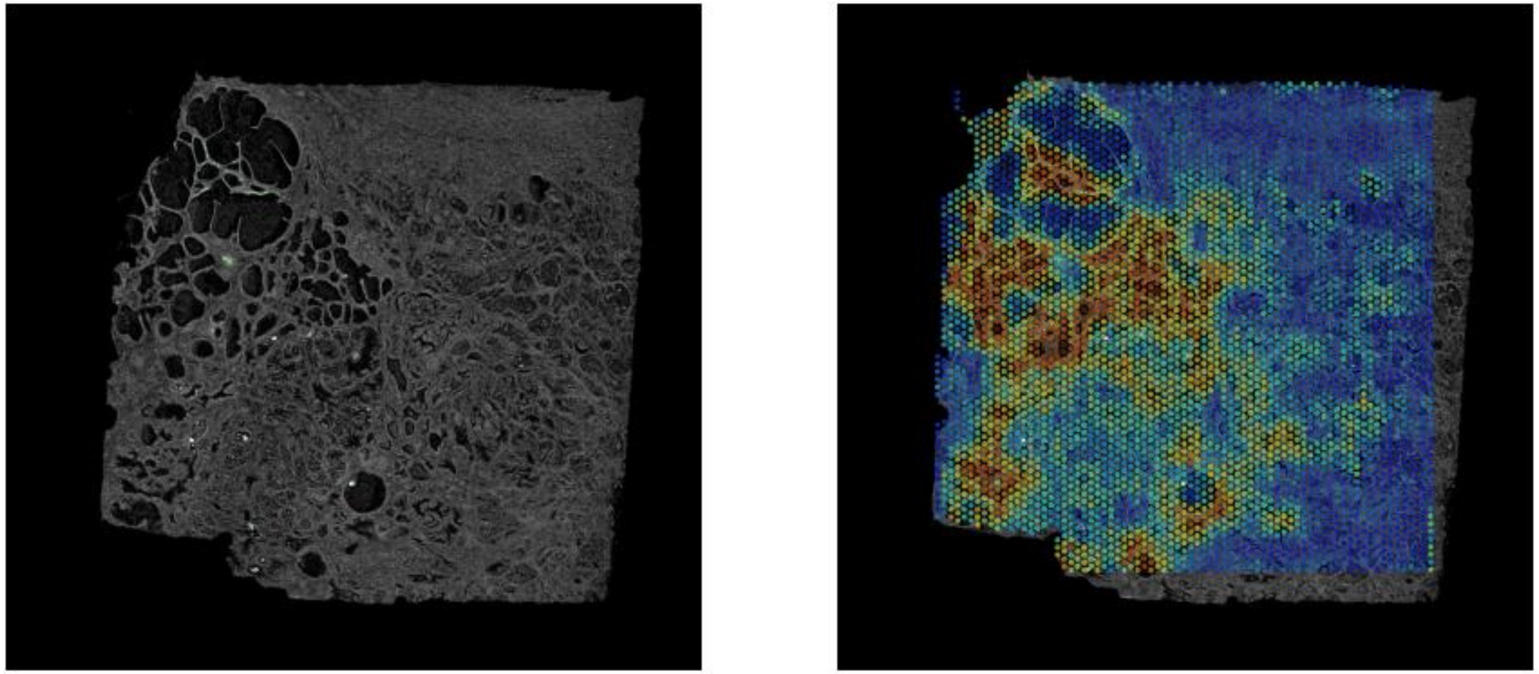
Mahmood Lab
AI for Pathology

AI-driven 3D Spatial Transcriptomics (?)



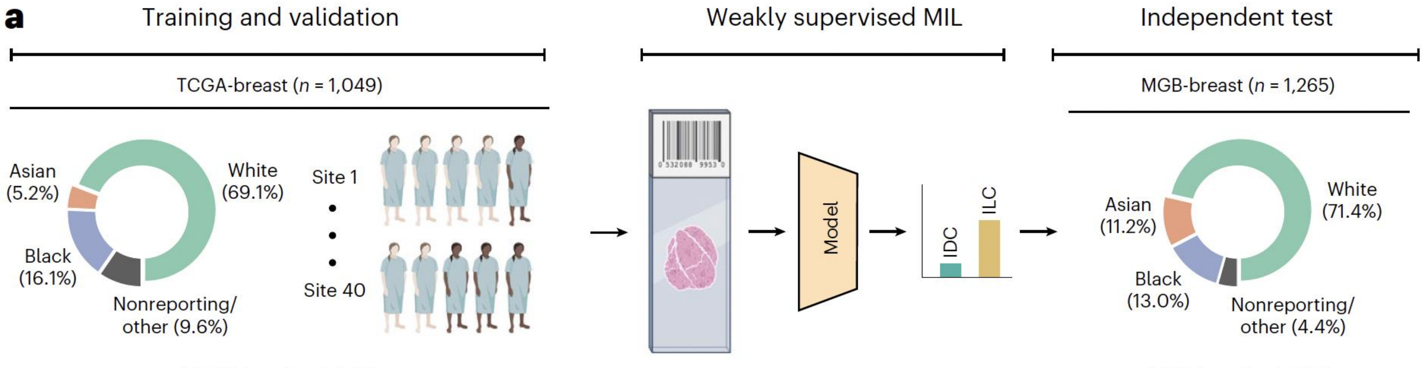
MSMB gene, a prostate cancer marker, is known to be downregulated in cancerous cells compared with benign prostate epithelium.

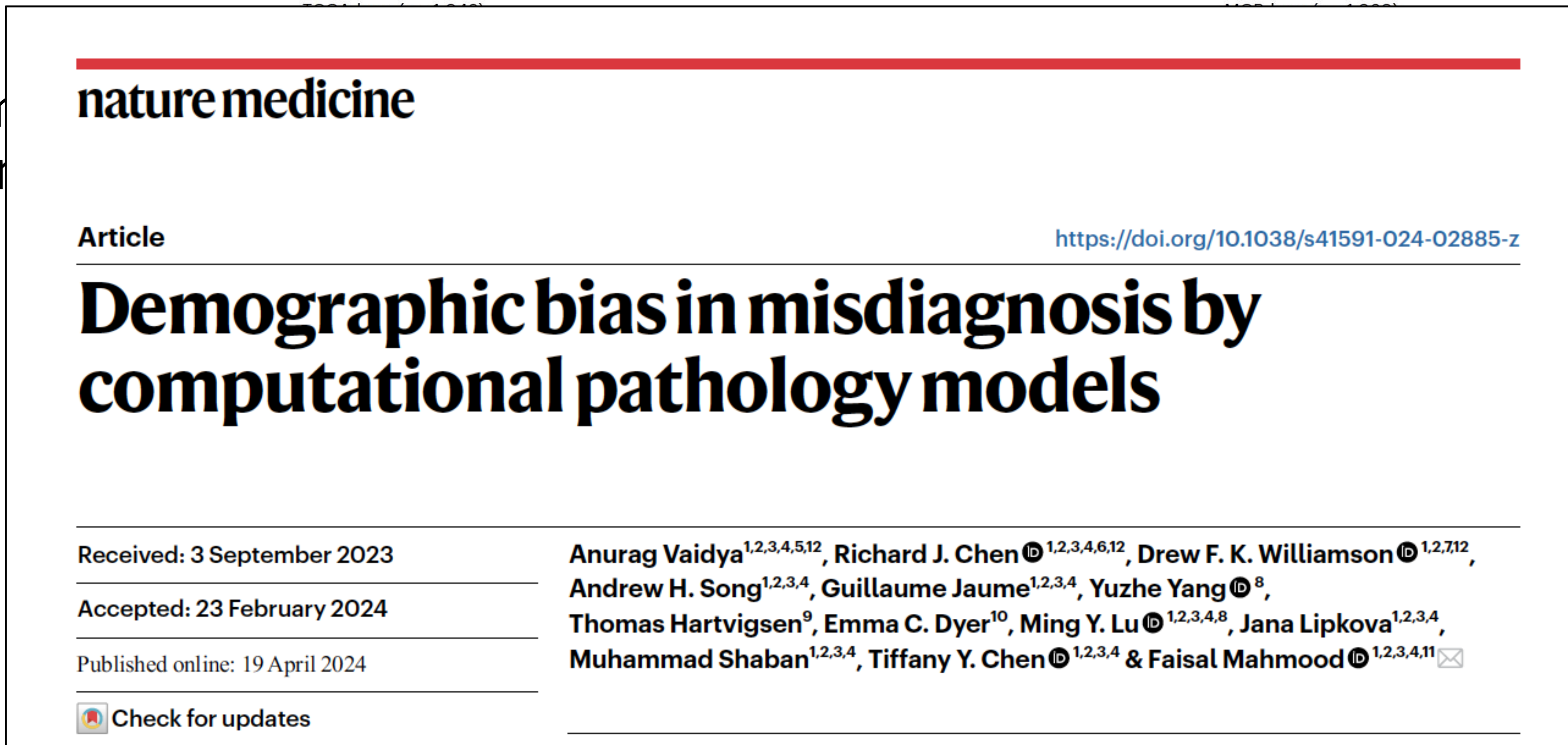
Scrolling up in the Z dimension …



Bias is computational pathology datasets

- ▶ Common datasets over-represent patients from certain demographics
- ▶ Real world populations are diverse
- ▶ Are there biases in algorithm trained for cancer subtyping ar mutation prediction tasks?





nature medicine

Article

https://doi.org/10.1038/s41591-024-02885-z

Demographic bias in misdiagnosis by computational pathology models

Received: 3 September 2023

Accepted: 23 February 2024

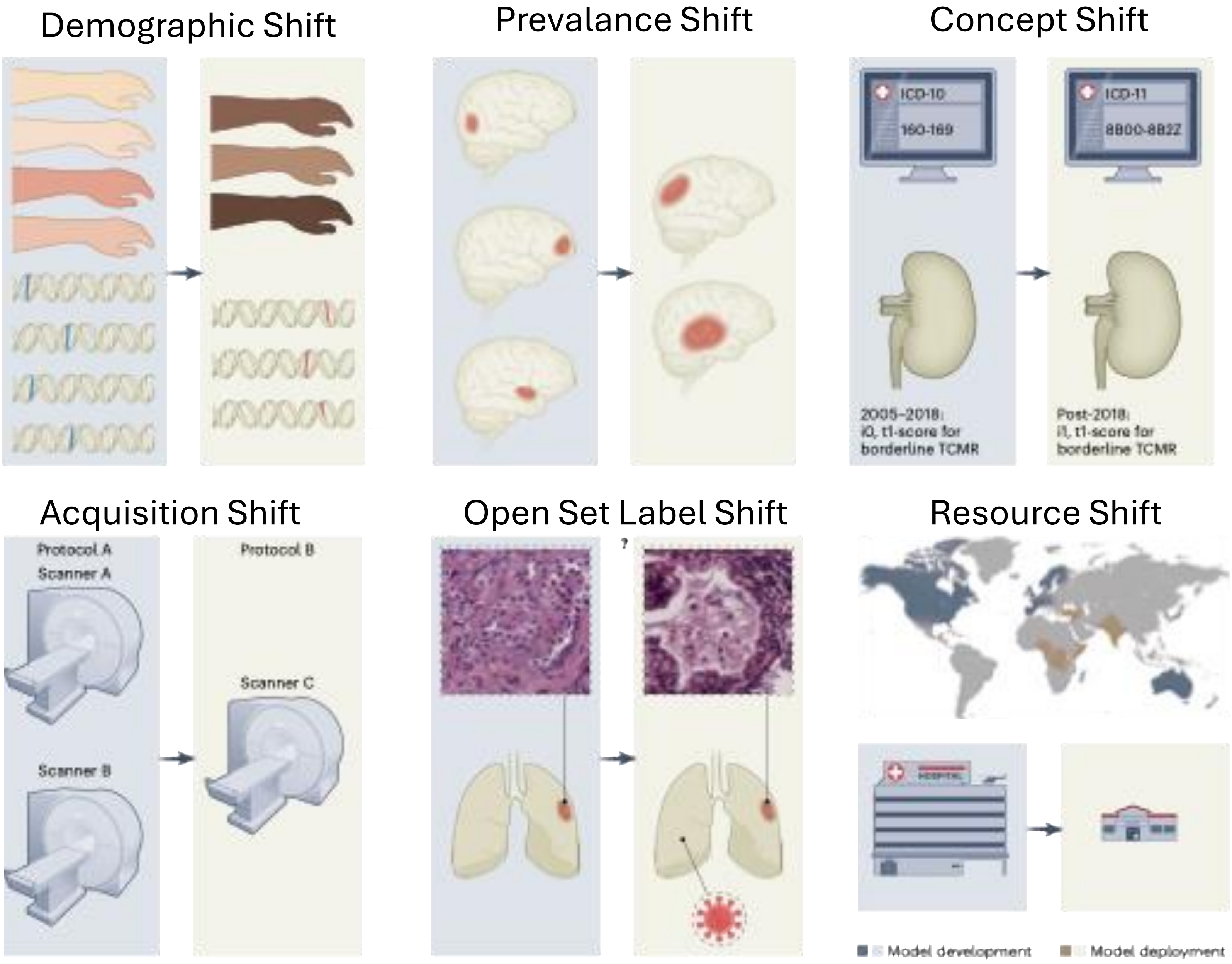
Published online: 19 April 2024

Check for updates

Anurag Vaidya[1,2,3,4,5,12], Richard J. Chen[1,2,3,4,6,12], Drew F. K. Williamson[1,2,7,12], Andrew H. Song[1,2,3,4], Guillaume Jaume[1,2,3,4], Yuzhe Yang[8], Thomas Hartvigsen[9], Emma C. Dyer[10], Ming Y. Lu[1,2,3,4,8], Jana Lipkova[1,2,3,4], Muhammad Shaban[1,2,3,4], Tiffany Y. Chen[1,2,3,4] & Faisal Mahmood[1,2,3,4,11]

(*Nature Medicine*, 2024)

Algorithm Fairness in Healthcare and Medicine



Demographic Shift

Prevalance Shift

Concept Shift

Acquisition Shift

Open Set Label Shift

Resource Shift

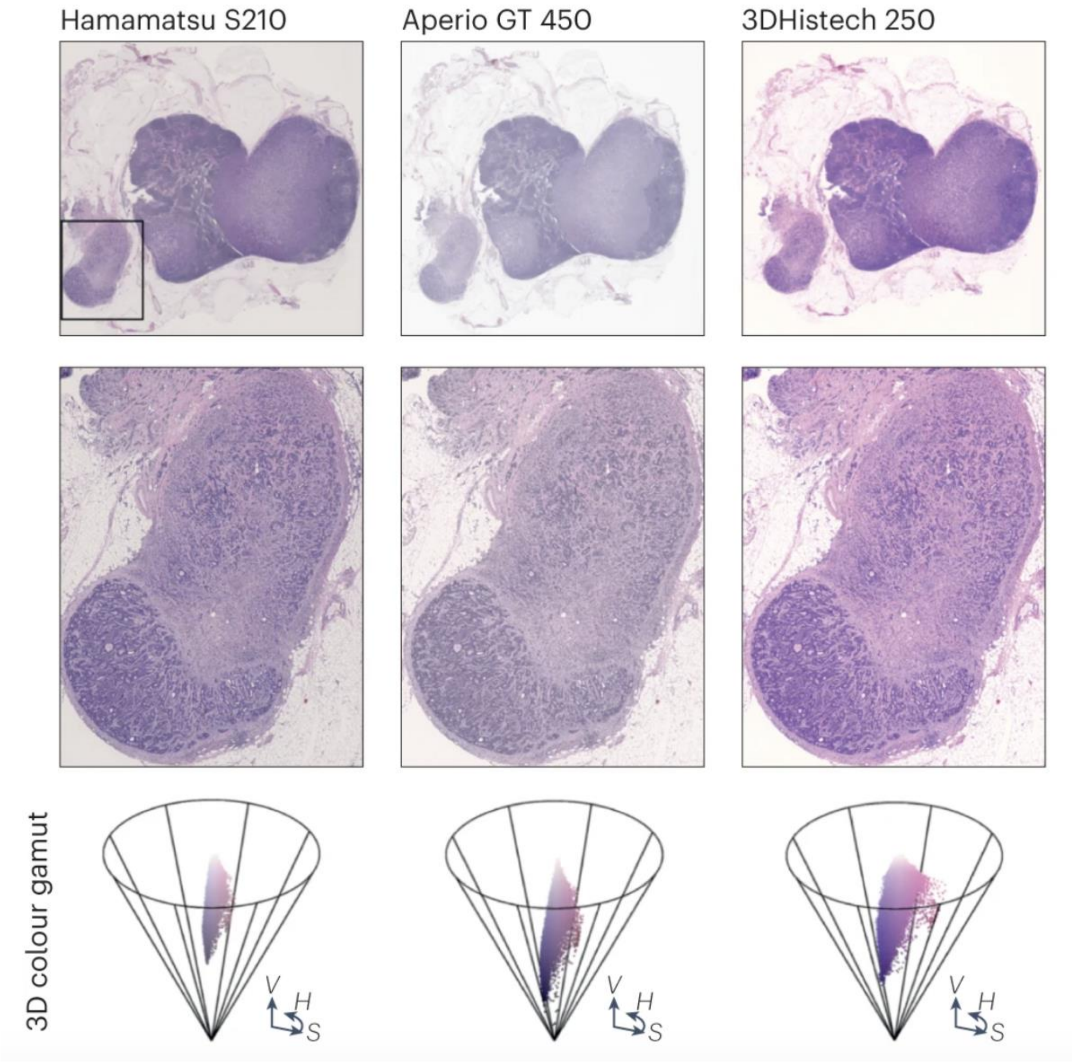
- Many healthcare disparities in medical AI can be understood as arising from dataset shift, e.g.

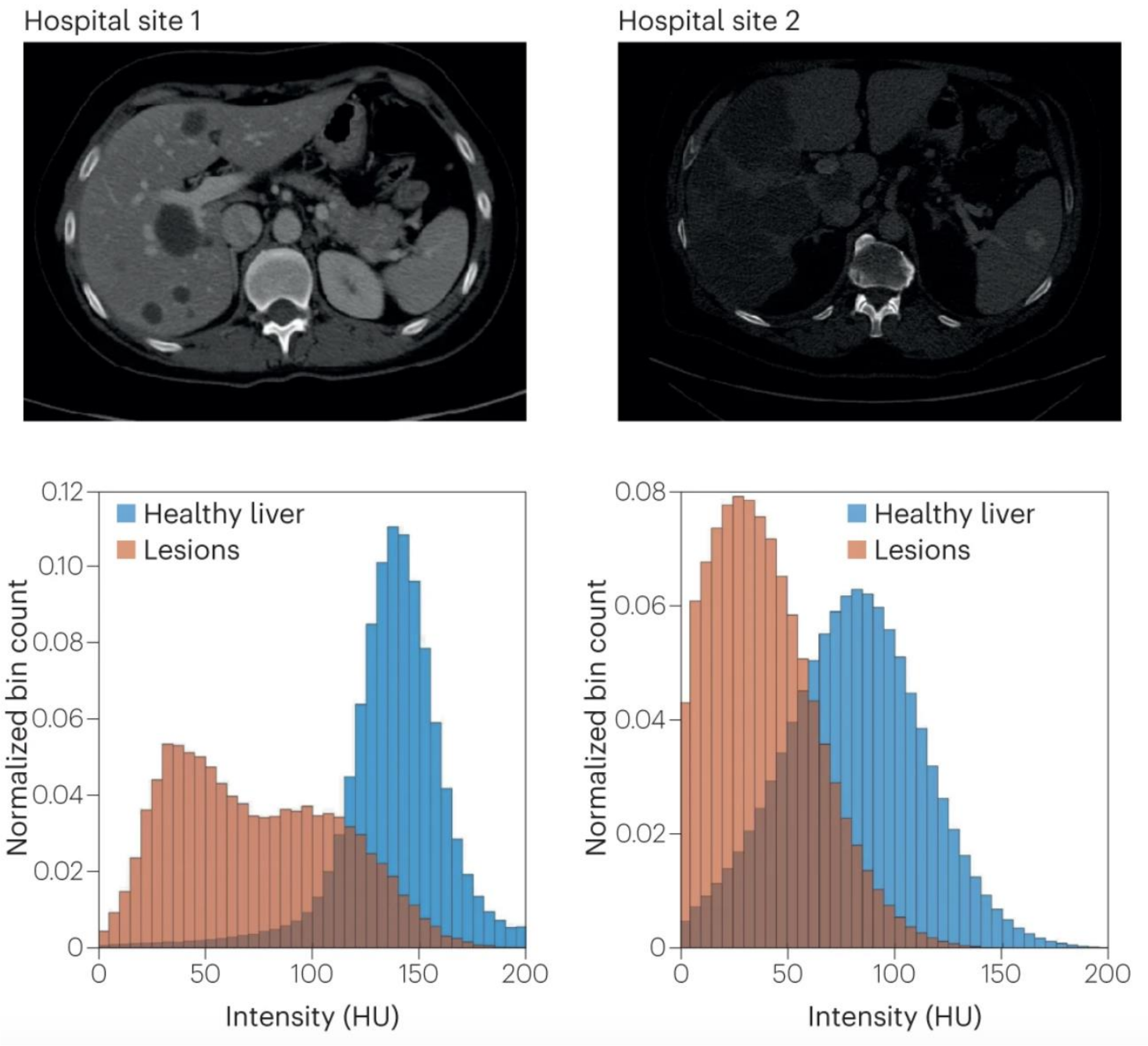
$$P_{\text{train}}(X) \neq P_{\text{test}}(X)$$

$$P_{\text{train}}(Y) \neq P_{\text{test}}(Y)$$

Chen et al., Algorithmic fairness in artificial intelligence for medicine and healthcare. Nature BME, 2024

Algorithm Fairness in Healthcare and Medicine

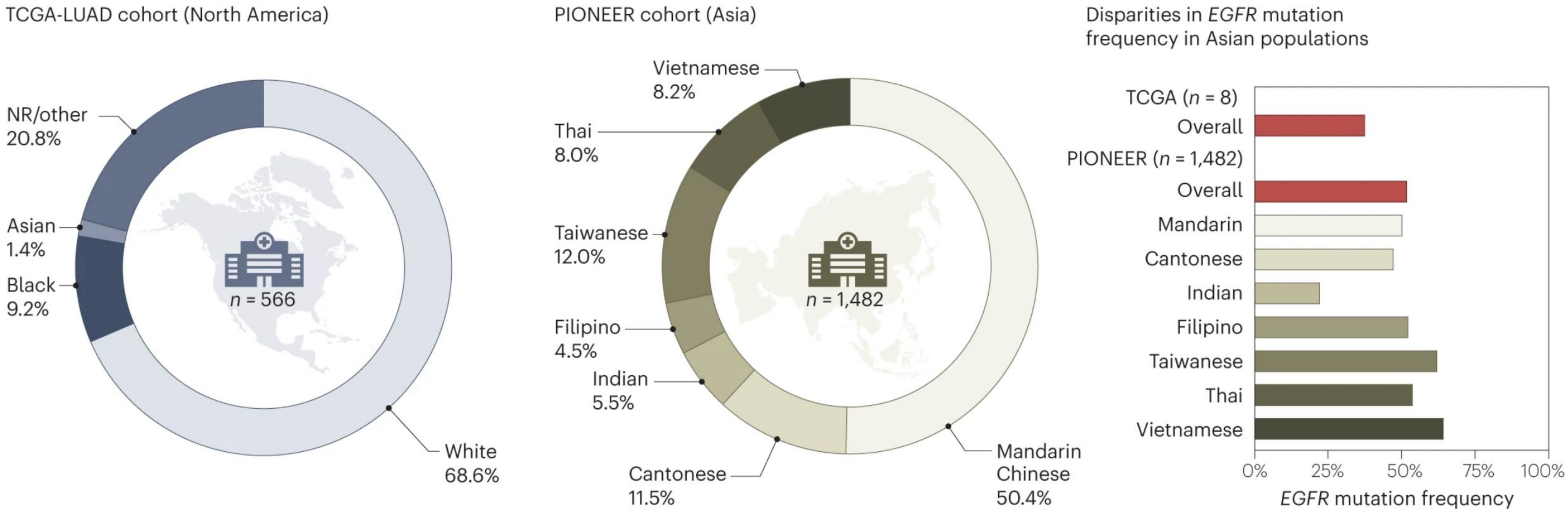




- Image acquisition shift in H&E pathology images (stain variability) and CT (radiointensity variability)

Chen et al., Algorithmic fairness in artificial intelligence for medicine and healthcare. Nature BME, 2024

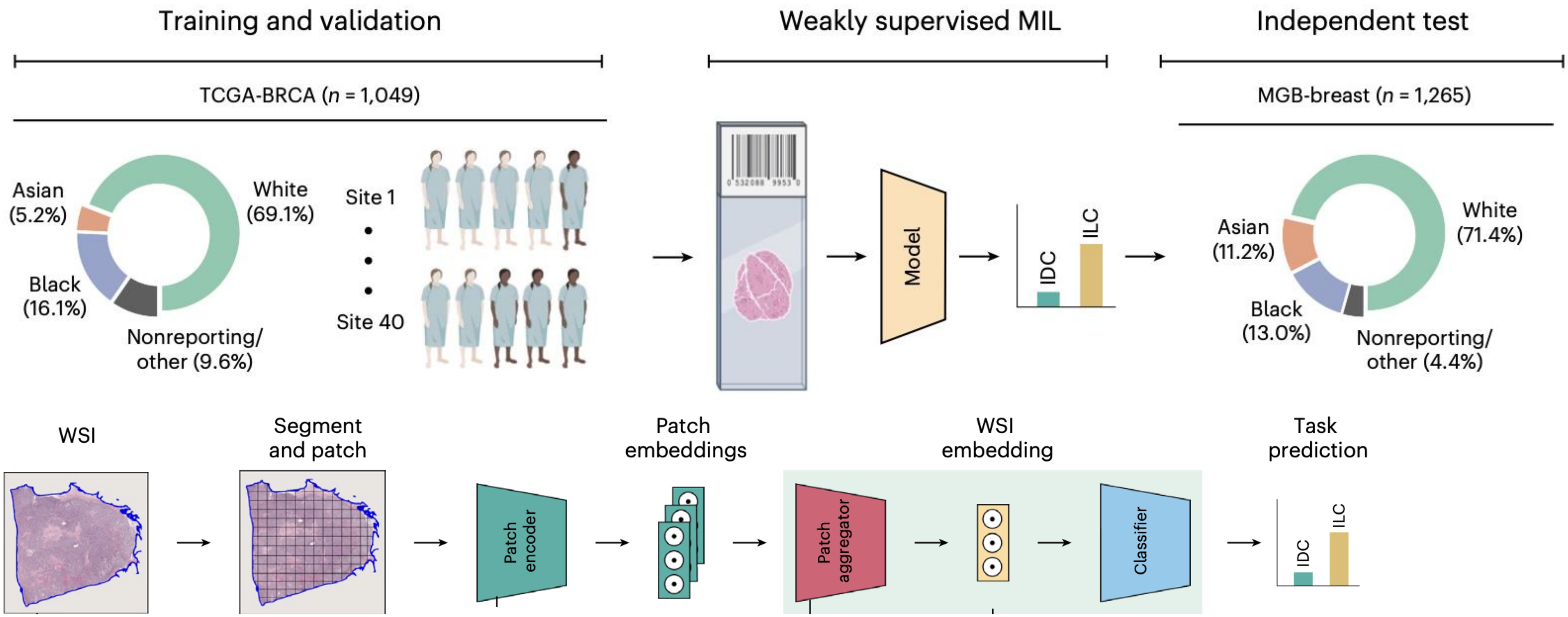
Algorithm Fairness in Healthcare and Medicine



- **The majority of models are trained on datasets that over-represent individuals of European ancestry**, often without the consideration of algorithm fairness
- 82.0% of all cases in the TCGA are from patients with European ancestry – how do AI models behave when trained on predominantly White patients and tested on under-represented minorities?

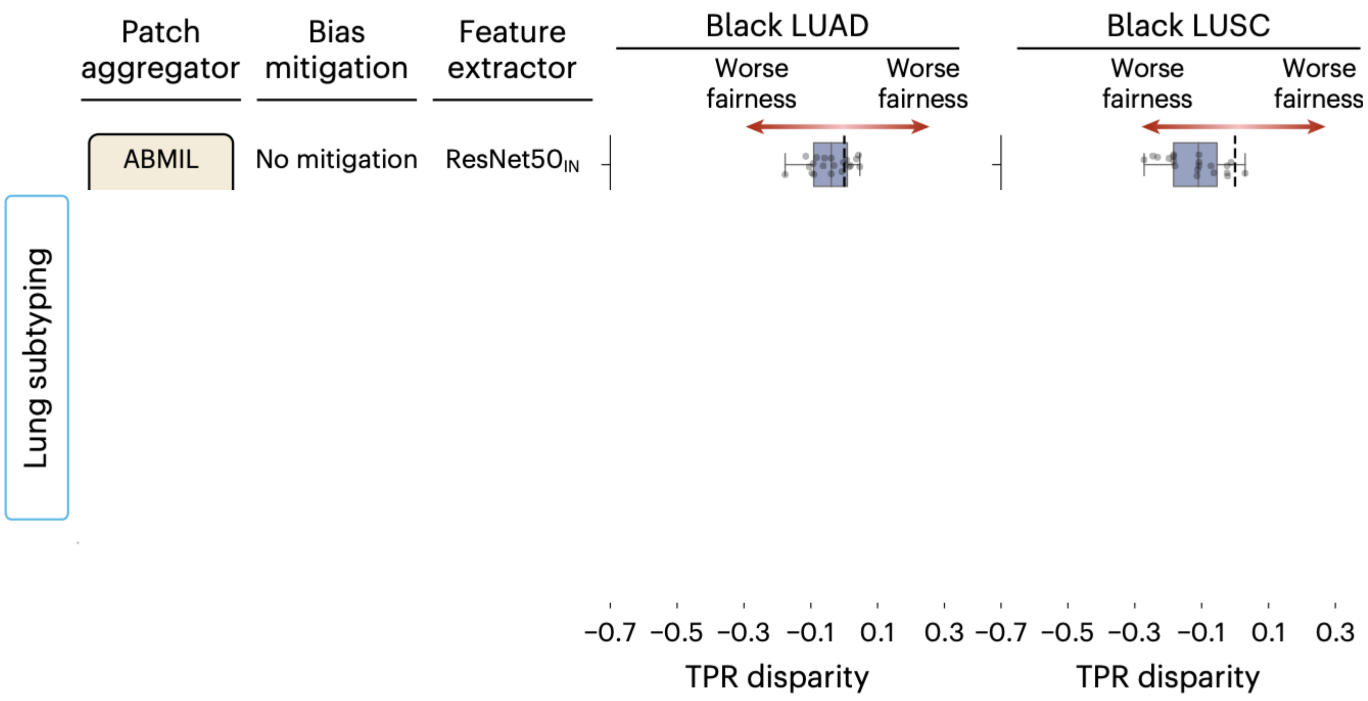
Chen et al., Algorithmic fairness in artificial intelligence for medicine and healthcare. Nature BME, 2024

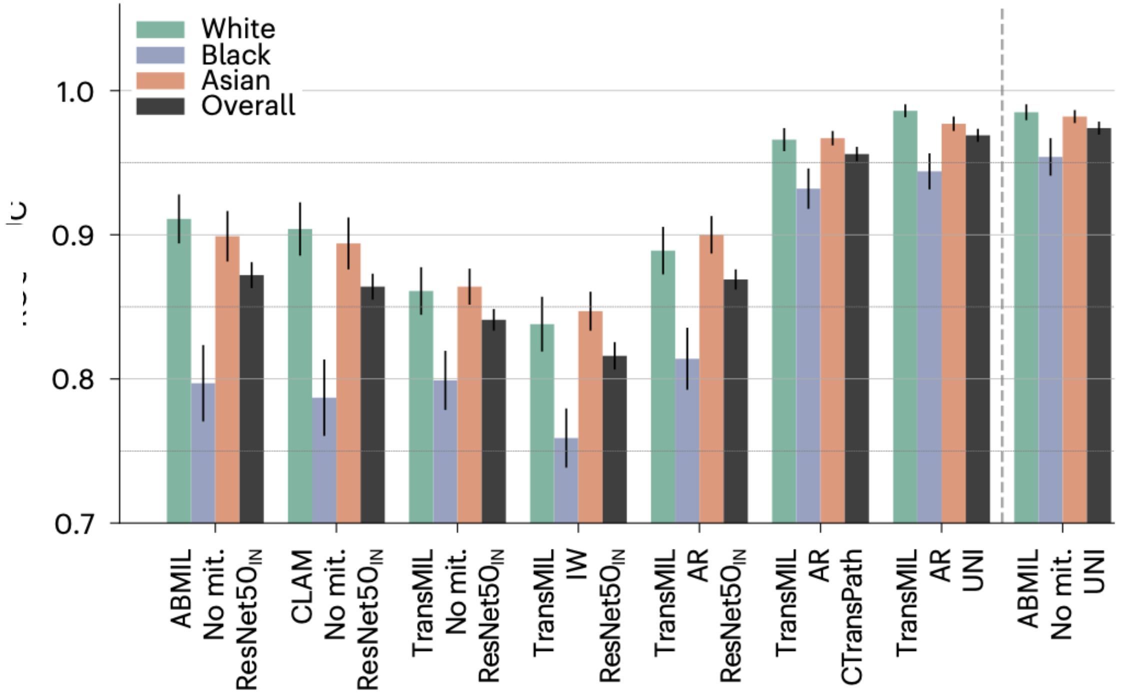
Demographic Bias in Computational Pathology AI models



Vaidya et al., Demographic bias in misdiagnosis by computational pathology models. Nature Medicine, 2024

Demographic Bias in Computational Pathology AI models

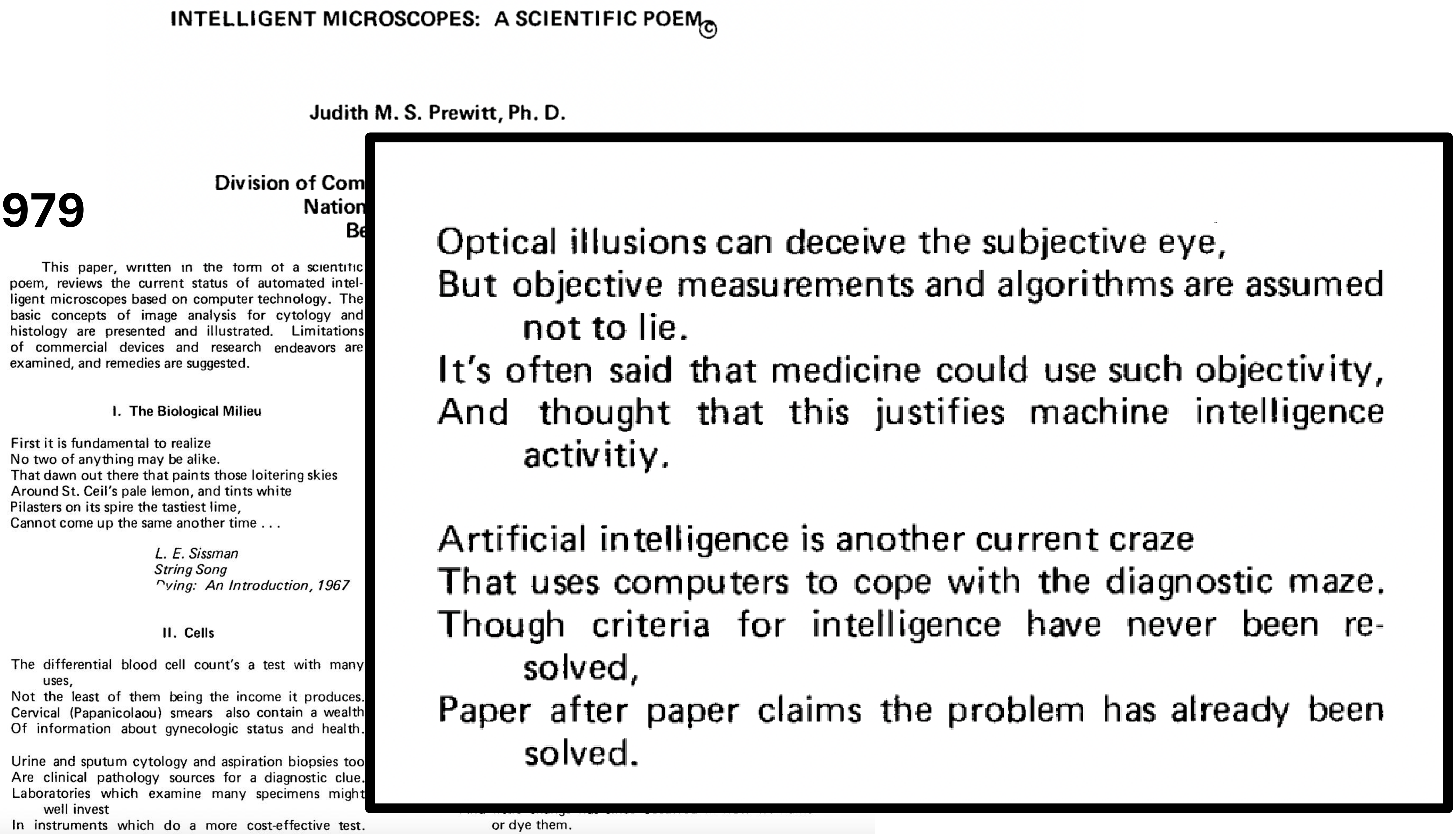




- Self-supervised pathology encoders (UNI) help mitigate performance disparities in cancer subtyping and biomarker prediction

Vaidya et al., Demographic bias in misdiagnosis by computational pathology models. Nature Medicine, 2024

1979



INTELLIGENT MICROSCOPES: A SCIENTIFIC POEM©

Judith M. S. Prewitt, Ph. D.

Division of Com
Nation
Be

This paper, written in the form of a scientific poem, reviews the current status of automated intelligent microscopes based on computer technology. The basic concepts of image analysis for cytology and histology are presented and illustrated. Limitations of commercial devices and research endeavors are examined, and remedies are suggested.

I. The Biological Milieu

First it is fundamental to realize
No two of anything may be alike.
That dawn out there that paints those loitering skies
Around St. Ceil's pale lemon, and tints white
Pilasters on its spire the tastiest lime,
Cannot come up the same another time . . .

L. E. Sissman
String Song
Dying: An Introduction, 1967

II. Cells

The differential blood cell count's a test with many
uses,
Not the least of them being the income it produces.
Cervical (Papanicolaou) smears also contain a wealth
Of information about gynecologic status and health.

Urine and sputum cytology and aspiration biopsies too
Are clinical pathology sources for a diagnostic clue.
Laboratories which examine many specimens might
well invest
In instruments which do a more cost-effective test.

Optical illusions can deceive the subjective eye,
But objective measurements and algorithms are assumed
not to lie.
It's often said that medicine could use such objectivity,
And thought that this justifies machine intelligence
activitiy.

Artificial intelligence is another current craze
That uses computers to cope with the diagnostic maze.
Though criteria for intelligence have never been re-
solved,
Paper after paper claims the problem has already been
solved.

or dye them.

@AI4Pathology

The Mahmood Lab

NIH NIGMS

Frederick National Laboratory for Cancer Research
sponsored by the National Cancer Institute

leidos
Leidos Biomedical Research, Inc.

NIH NATIONAL CANCER INSTITUTE

AACR
American Association for Cancer Research®
FINDING CURES TOGETHER®

NVIDIA®

American Cancer Society®

NIH NHGRI

NSF

HARVARD MEDICAL SCHOOL

BROAD INSTITUTE

BWH BRIGHAM AND WOMEN'S HOSPITAL

Dana-Farber Cancer Institute