Leveraging Foundation Models for H&E and Beyond

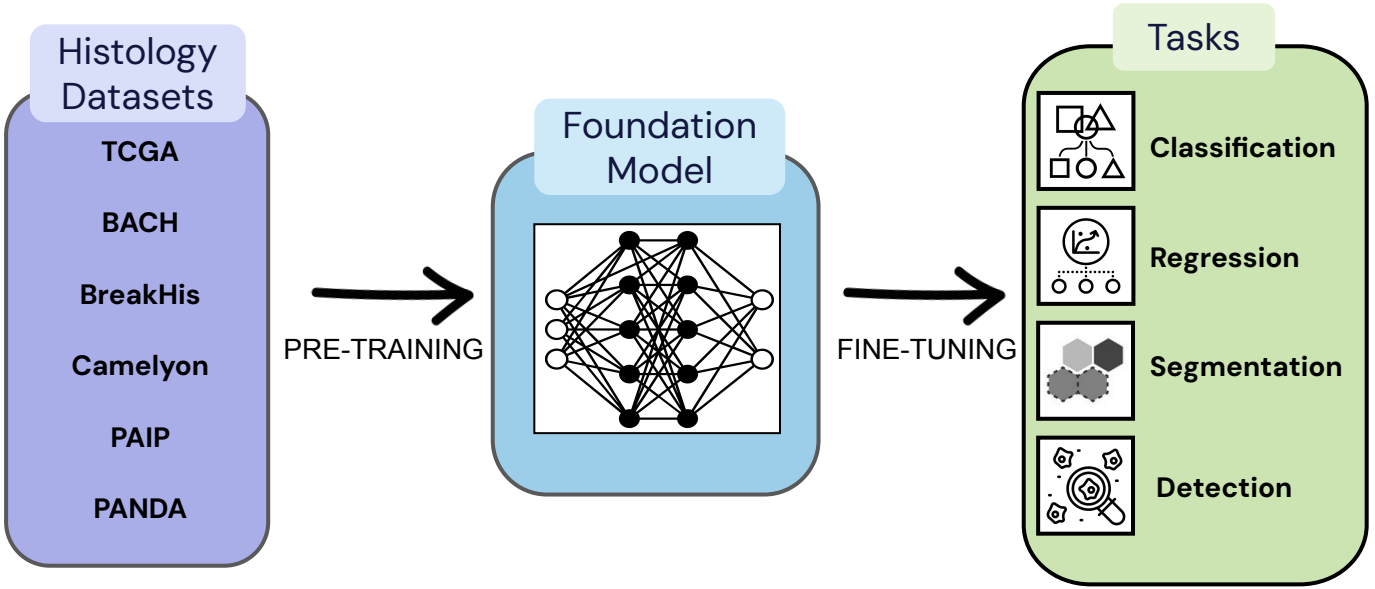
Heather Couture

MICCAI
COMPAYL Workshop

October 6, 2024

Pixel Scientia LABS

Foundation Models for Histology



First publicly available model ~2022
“Self-supervised learning”

How did we get here?

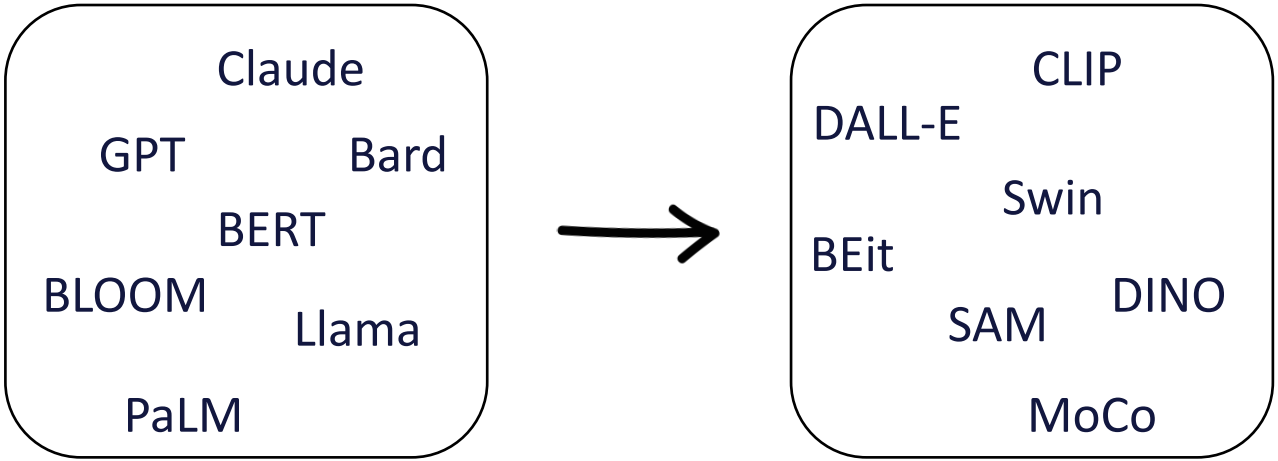
“Foundation model” coined by Stanford Institute for Human-Centered Artificial Intelligence's Center for Research on Foundation Models

“A foundation model is any model that is trained on **broad data** (generally using **self-supervision** at scale) that can be **adapted** (e.g., fine-tuned) to a wide range of downstream tasks.”

Source: Bommasani, On the Opportunities and Risks of Foundation Models, 2021

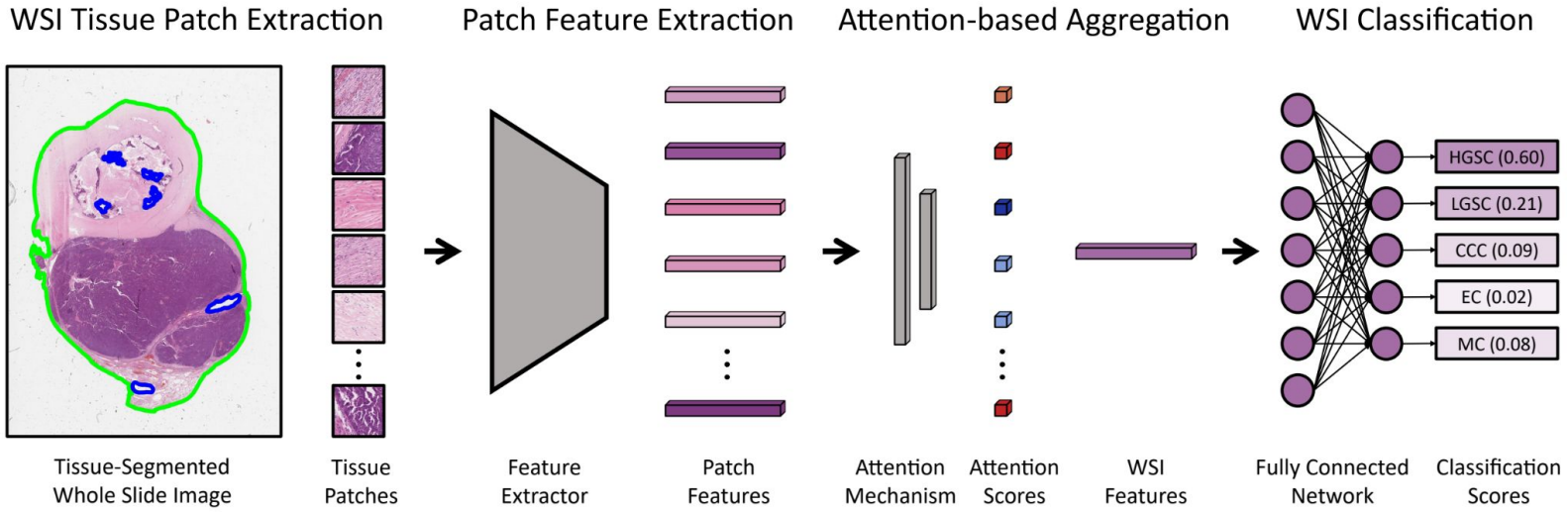
Self-supervision: learn features without labels by solving a pretext task

From Large Language Models to Large Vision Models

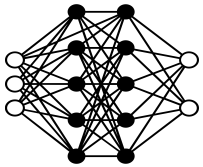


Easy to adapt to new tasks

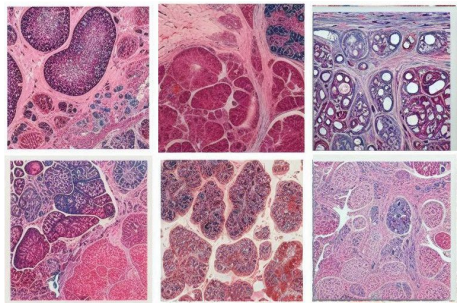
From ImageNet Pretraining to Self-Supervised Learning



Source: Rymarczyk, Kernel Self-Attention in Deep Multiple Instance Learning, 2020

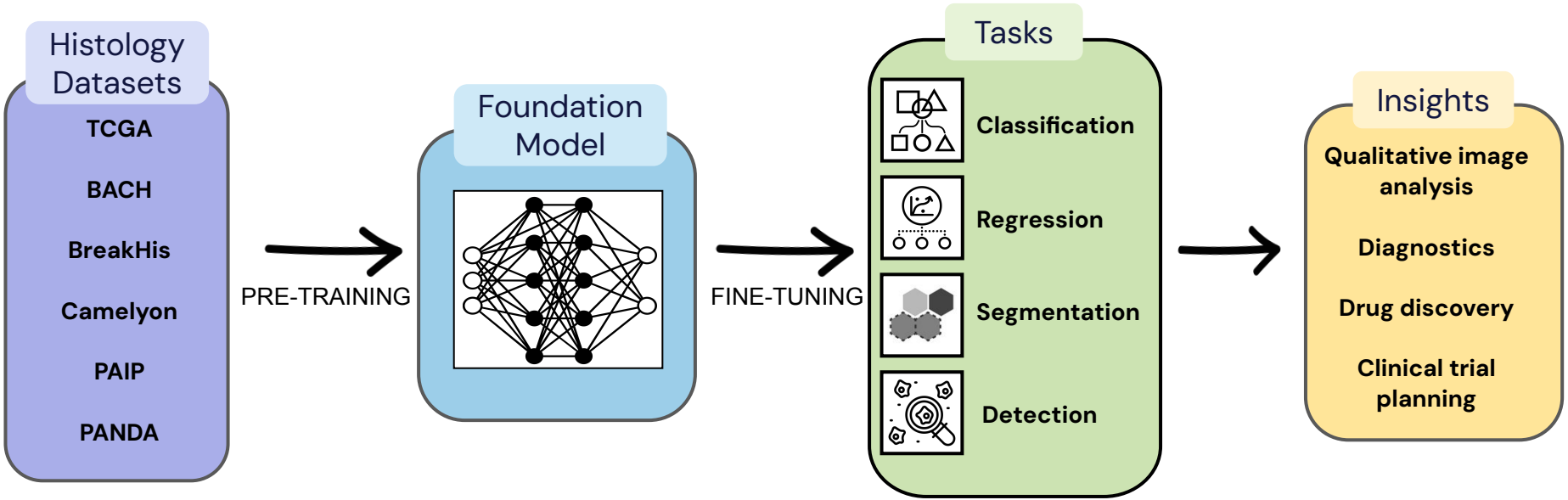


ImageNet vs. Histology



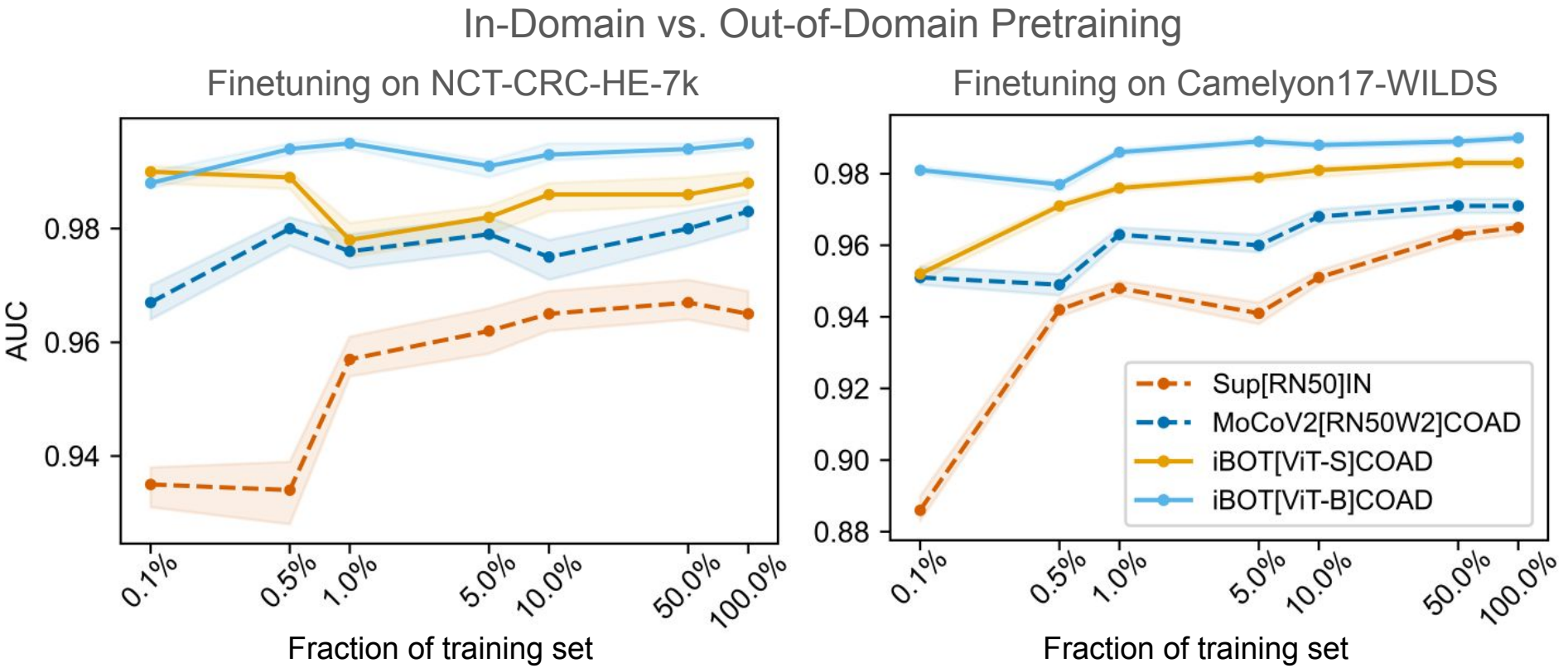
A more generalizable tile embedding

Foundation Models for Histology: Adaptable & Generalizable



Why use foundation models?

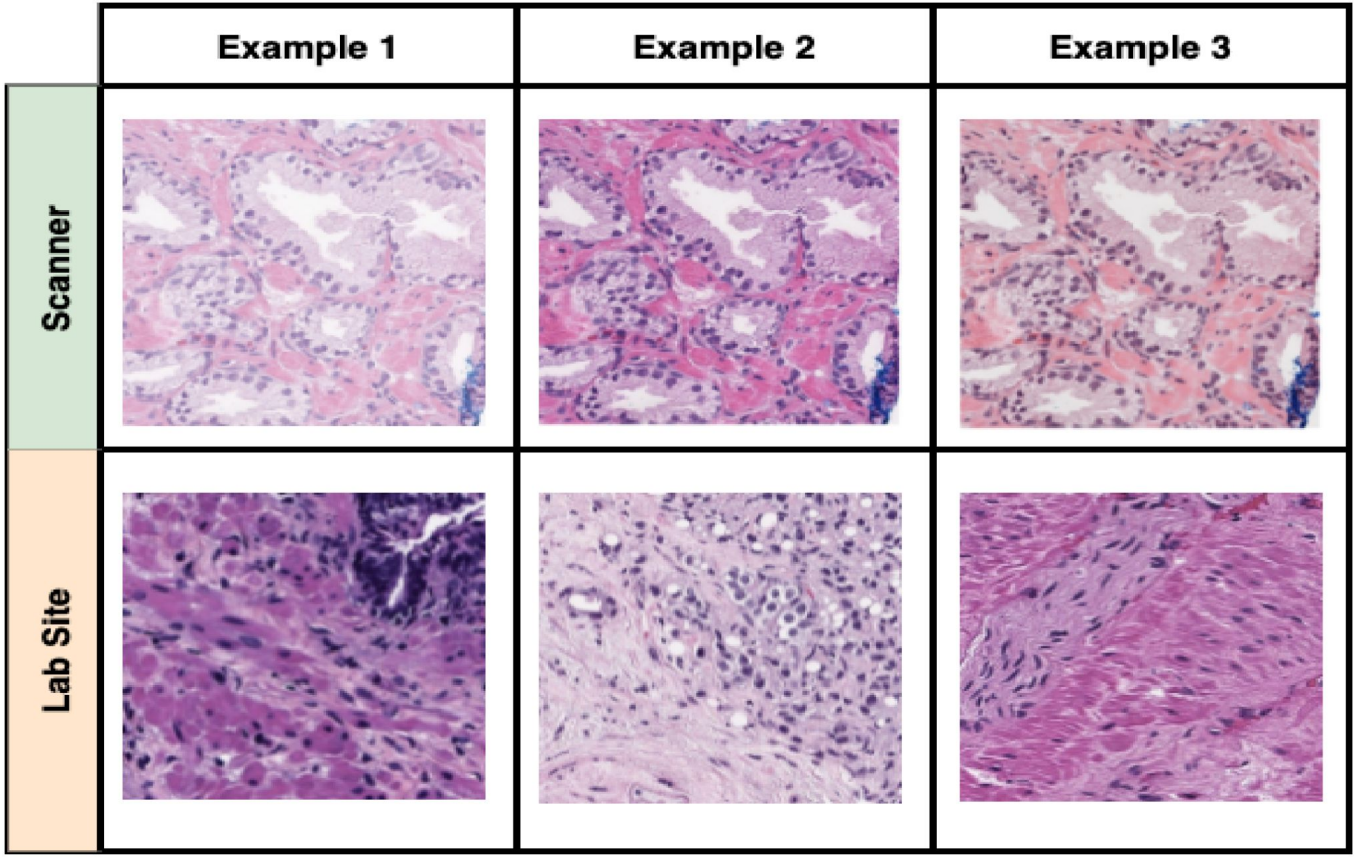
Scarce Labeled Data



Improved accuracy with fewer labels

Source: Filiot, Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling, 2023 (Owkin)

Robustness to Distribution Shifts

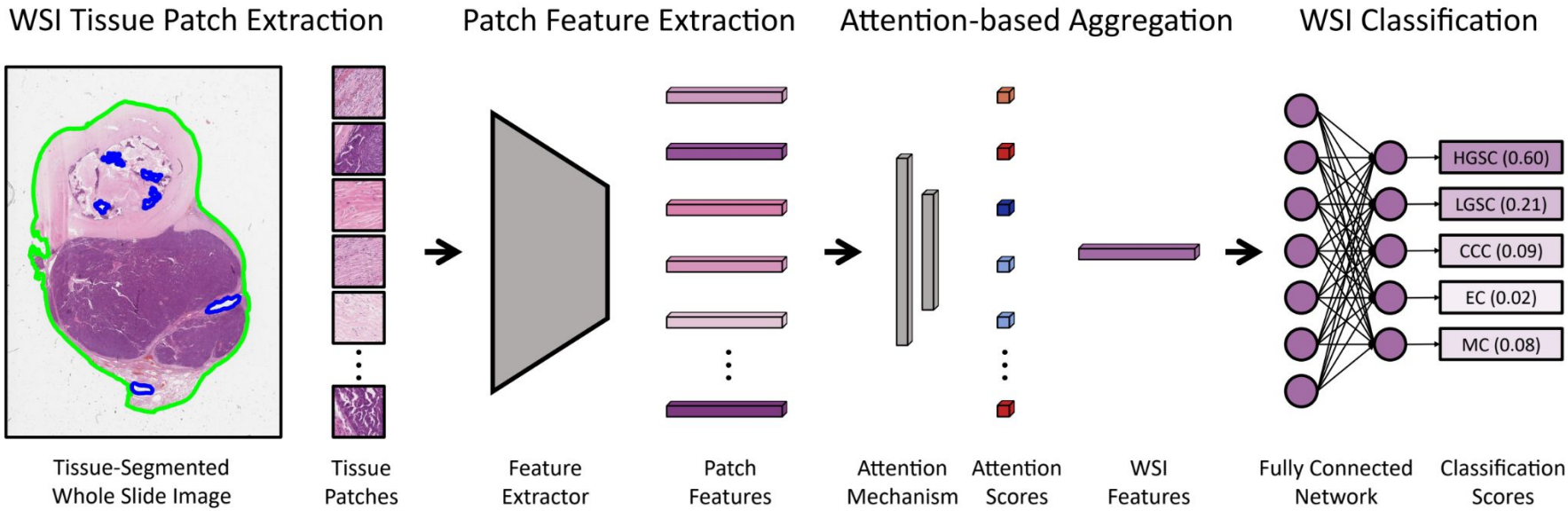


- Inconsistent tissue preparation
- Differences in imaging equipment
- Artifacts
- Batch effects

Improved generalizability

Source: Javed, Rethinking Machine Learning Model Evaluation in Pathology, 2022

Gigapixel Images and Weak Supervision



Slide- or patient-level characteristics:

- Mutations
- Genomic subtype
- Hormone receptor status
- Patient outcome
- Treatment response

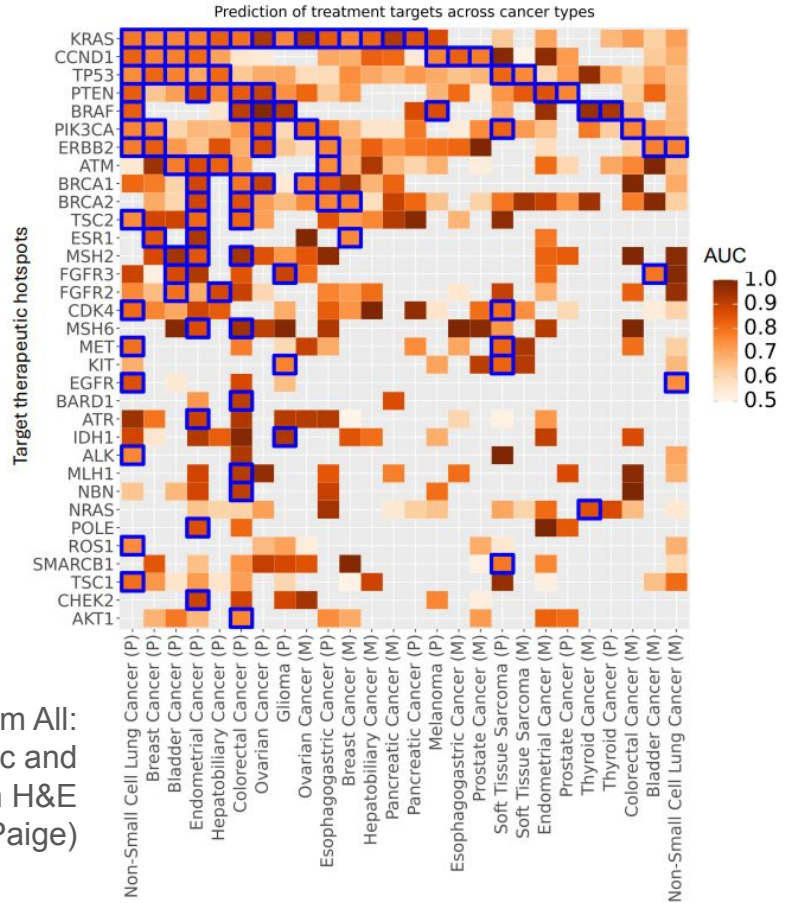
Source: Breen, A Comprehensive Evaluation of Histopathology Foundation Models for Ovarian Cancer Subtype Classification, 2024

Increased accuracy and generalizability

Faster Experimentation

- 48k slides from 39k patients
- Image tiles encoded with Virchow2
- Unified model to predict 1,228 different biomarkers and cancer types simultaneously

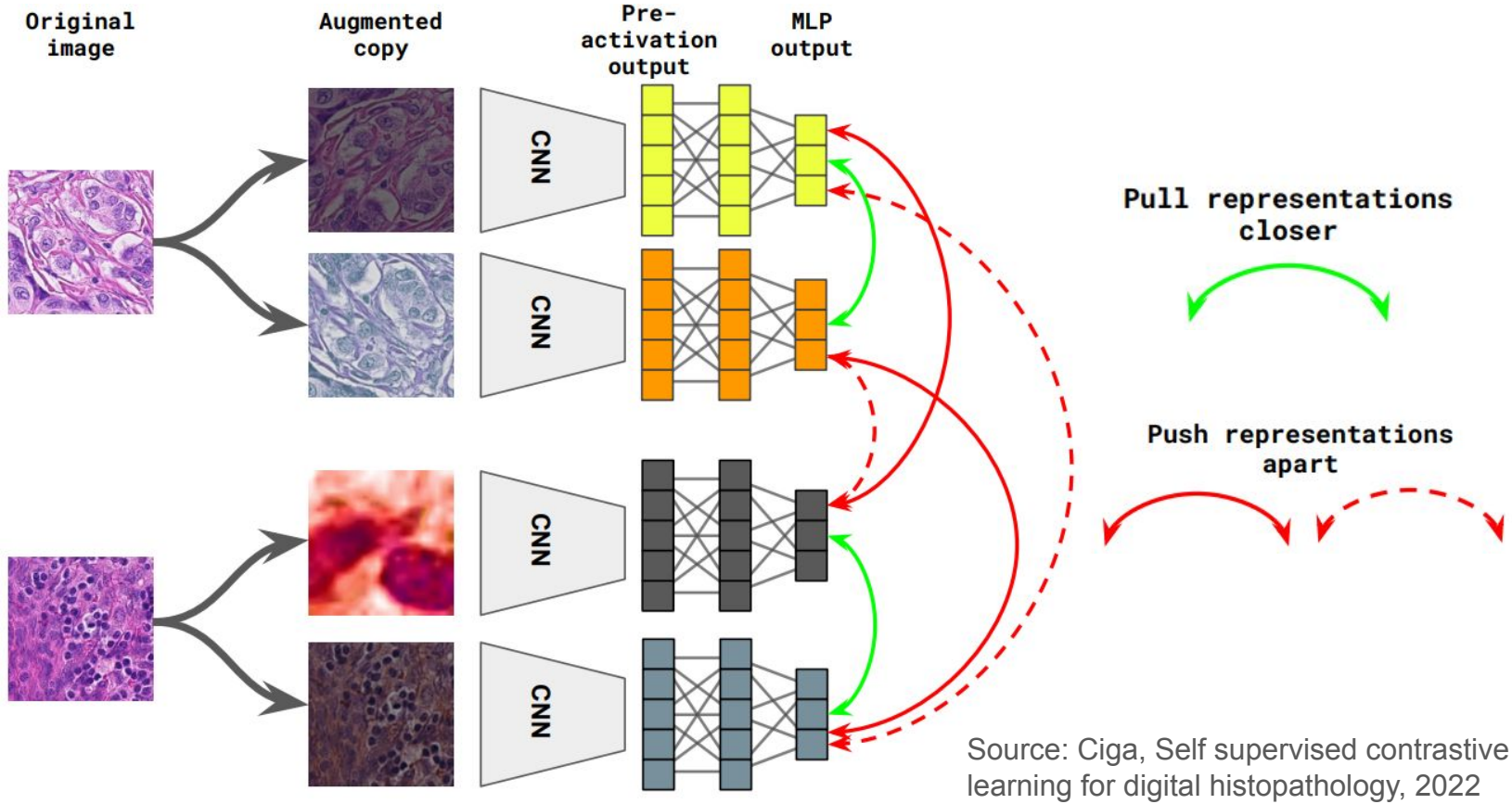
Many cancer types and biomarkers



Source: Wang, Screen Them All: High-Throughput Pan-Cancer Genetic and Phenotypic Biomarker Screening from H&E Whole Slide Images, 2024 (Paige)

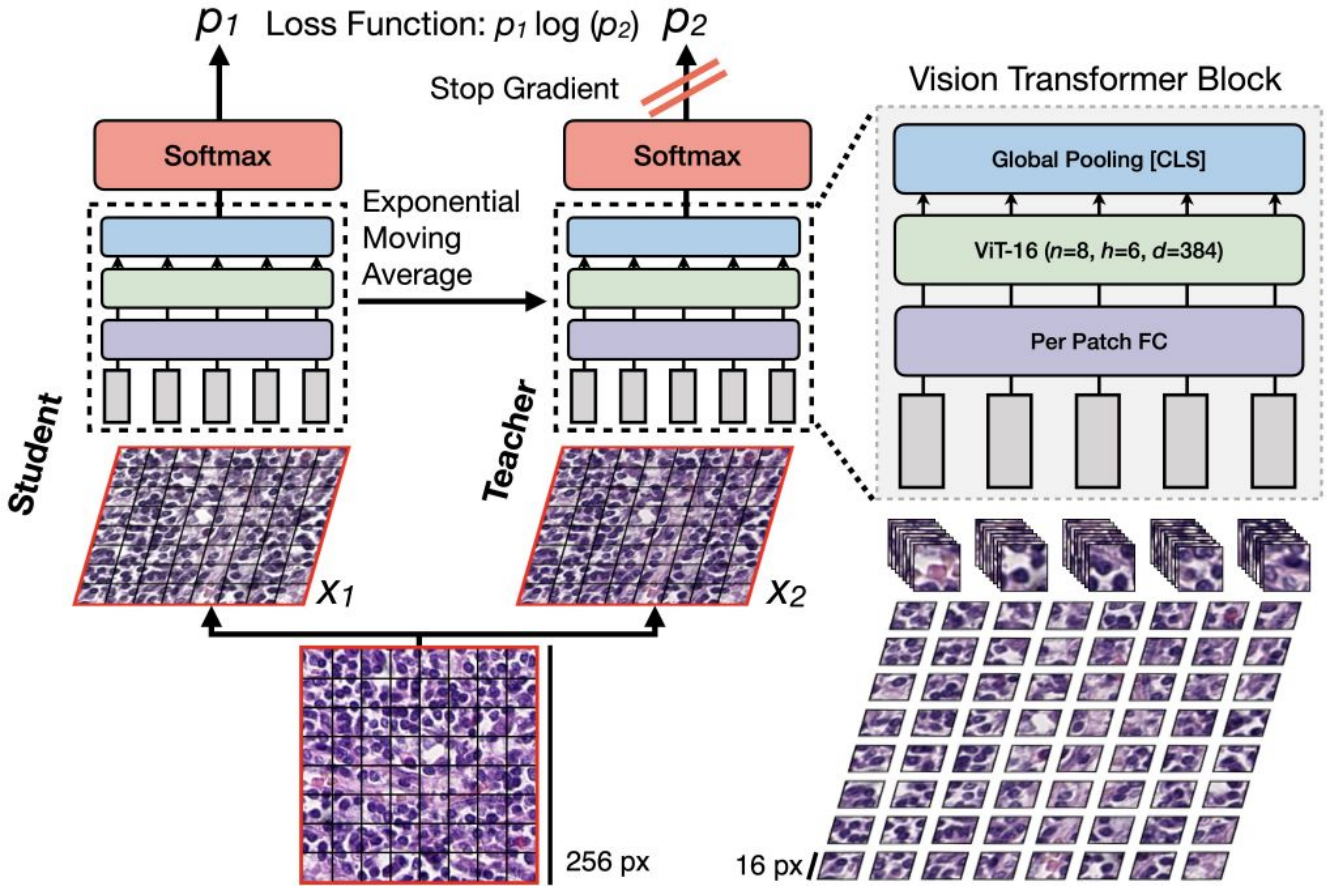
How to train a foundation model?

Self-Supervised Learning: Contrastive



Source: Ciga, Self supervised contrastive learning for digital histopathology, 2022

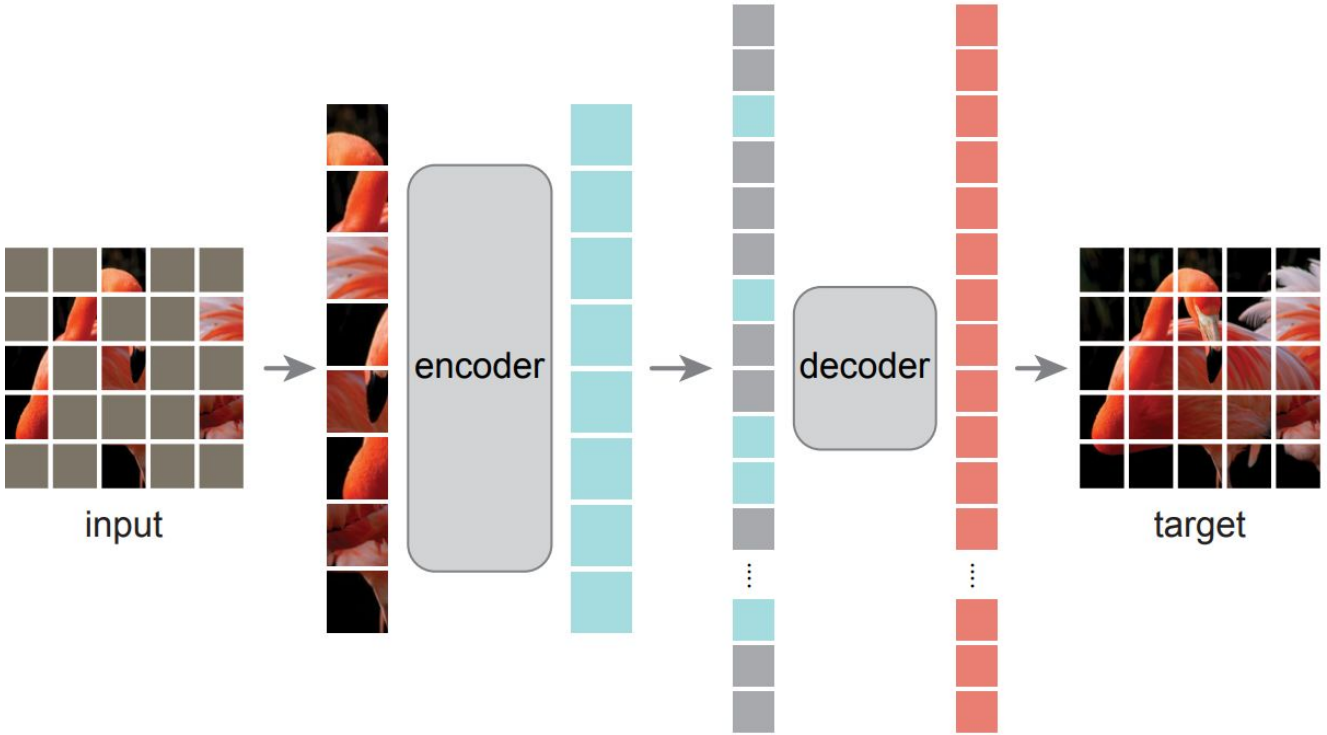
Self-Supervised Learning: Self-Distillation



DINO:
distillation with
no labels

Source: Chen, Self-Supervised Vision Transformers Learn Visual Concepts in Histopathology, 2022

Self-Supervised Learning: Reconstruction



Source: He, Masked Autoencoders Are Scalable Vision Learners, 2021

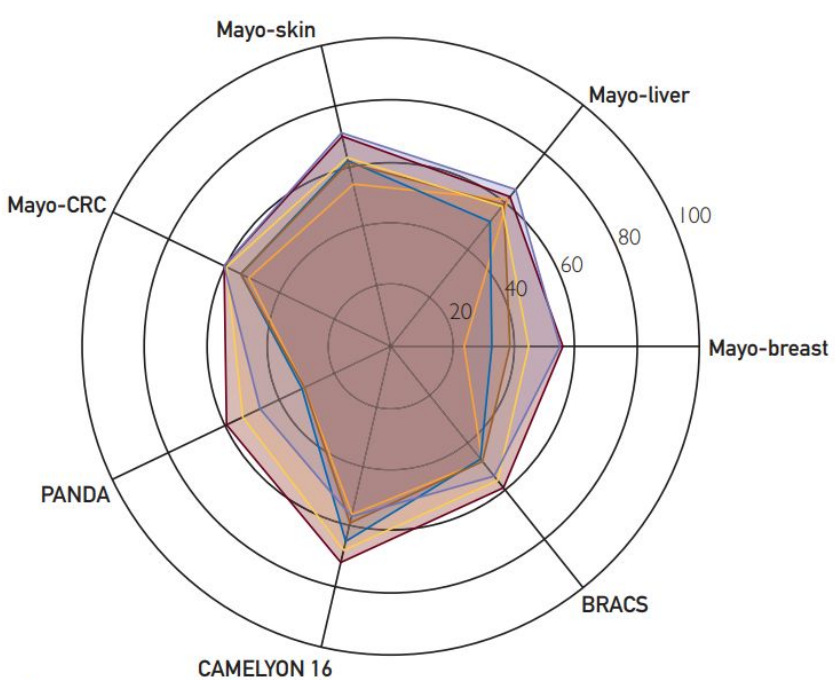
Published Foundation Models

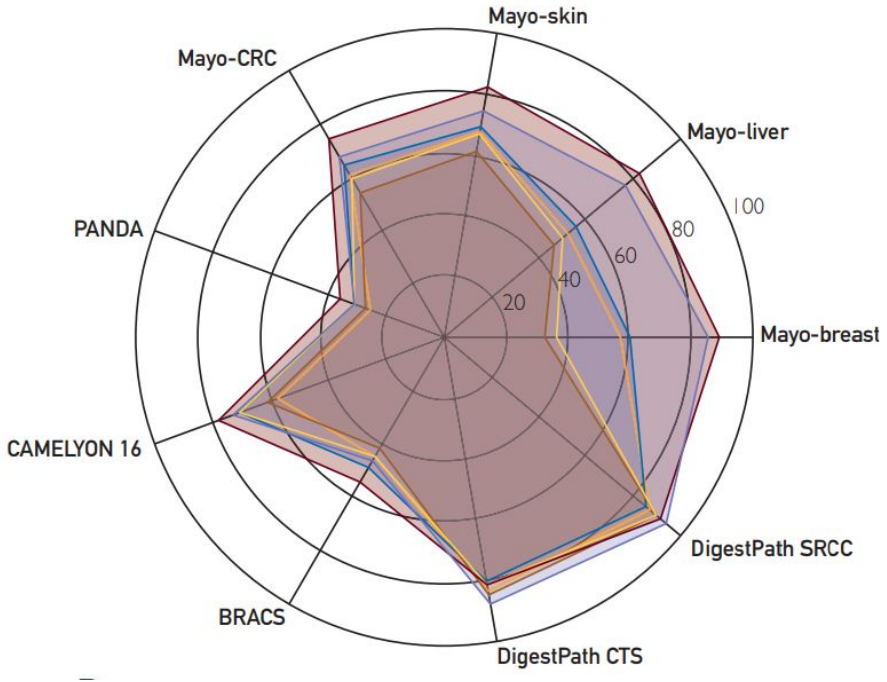
Foundation Model	Backbone	Parameters	Pretraining Algorithm	Pretraining WSIs
Lunit	ViT-S	22 million	DINO	37k
Path Foundation	ViT-S	22 million	SimCLR+MSN	6k
PLUTO	ViT-S	22 million	DINO+iBOT+MAE+Fourier	158k
CTransPath	CNN + SwinT	28 million	Novel SSL	32k
Hibou-B	ViT-B	86 million	DINOv2	1.1m
Phikon	ViT-B	86 million	iBOT	6k
Kaiko-L14	ViT-L	303 million	DINOv2	29k
GPFM	ViT-L	303 million	Novel Distillation	72k
UNI	ViT-L	303 million	DINOv2	100k
RudolfV	ViT-L	303 million	DINOv2	134k
Hibou-L	ViT-L	307 million	DINOv2	1.1m
Phikon-v2	ViT-L	307 million	DINOv2	55k
Virchow	ViT-H	631 million	DINOv2	1.5m
Virchow2-CLS	ViT-H	631 million	DINOv2	3.1m
CanvOI	ViT-g	1,134 million	DINOv2	632k
H-optimus-0	ViT-g	1,134 million	DINOv2	>500k
Prov-GigaPath	ViT-g	1,134 million	DINOv2	171k

Consensus: DINOv2, larger ViT backbone, larger pretraining dataset

What makes a good foundation model?

In-Domain Pretraining

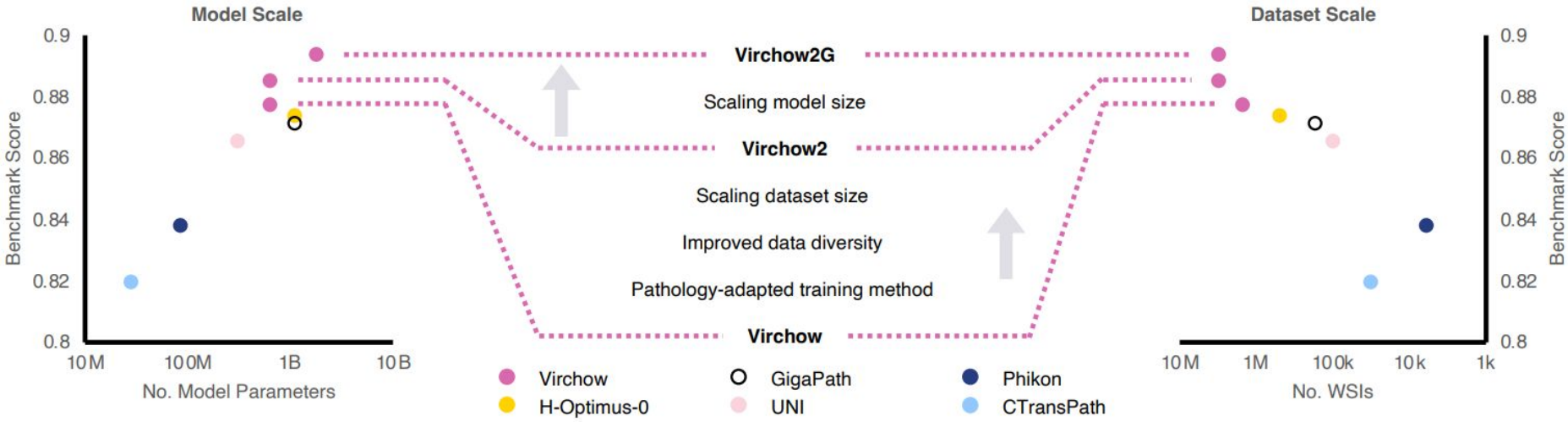




DinoV2 CLIP BiomedCLIP PLIP KimiaNet DinoSSLPath

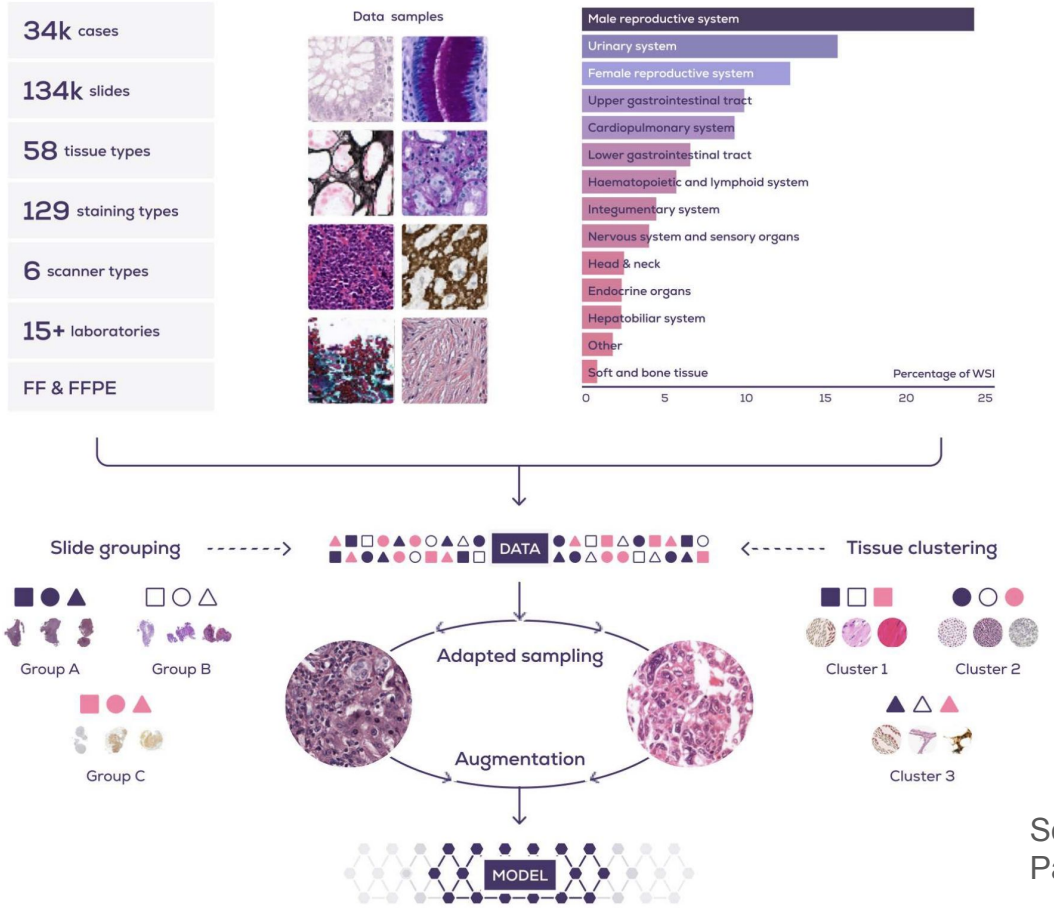
Source: Alfasly, Foundation Models for Histopathology: Fanfare or Flair, 2024 (Mayo Clinic)

Larger Models and Larger Training Sets



Source: Zimmerman, Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology, 2024 (Paige)

Data Diversity



RudolfV outperformed UNI, Phikon, and Virchow on a number of tasks

Hypothesis: pretraining data diversity is critical

Source: Dippel, RudolfV: A Foundation Model by Pathologists for Pathologists, 2024 (Aignostics)

Magnification

Downstream task	40×	20×	10×	5×	{20, 40}×	{5, 10, 20}×	{5, 10, 20, 40}×
BACH (3.47×)*	0.639	0.685	0.659	0.679	0.689	0.683	**0.753**
CRC (20×)	0.935	0.945	0.939	0.927	0.942	0.944	**0.947**
MHIST (5×)	0.744	0.746	0.648	0.710	0.746	0.744	**0.771**
PCam/val (10×)	0.879	**0.898**	0.873	0.859	0.887	0.870	0.887
PCam/test (10×)	0.824	0.874	0.834	0.820	0.874	0.858	**0.876**

Source: Aben, Towards Large-Scale Training of Pathology Foundation Models, 2024 (kaiko.ai)

Multiple magnification levels is best

Self-Supervised Frameworks

DINOV2 for Different Datasets and Architectures

Data	Arch.	Top-1 ACC
Mass-1K	ViT-B	0.453
Mass-1K	ViT-L	0.473
Mass-22K	ViT-B	0.522
Mass-22K	ViT-L	0.508
Mass-100K	ViT-B	0.503
Mass-100K	ViT-L	0.538

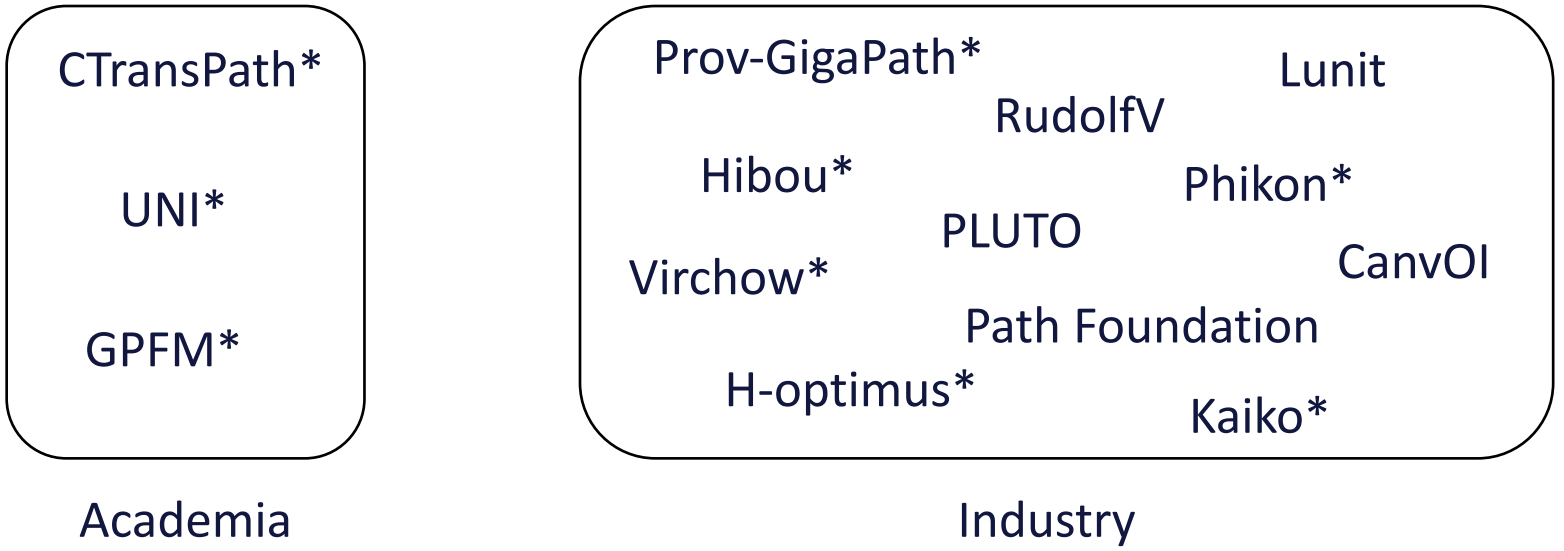
Mass-1K Pretraining for Different SSL Methods

SSL Method	Arch.	Top-1 ACC
MoCoV3	ResNet50	0.399
MoCoV3	ViT-L	0.273
DINOV2	ViT-L	0.473

Source: Chen, Towards a general-purpose foundation model for computational pathology, 2024 (Harvard)

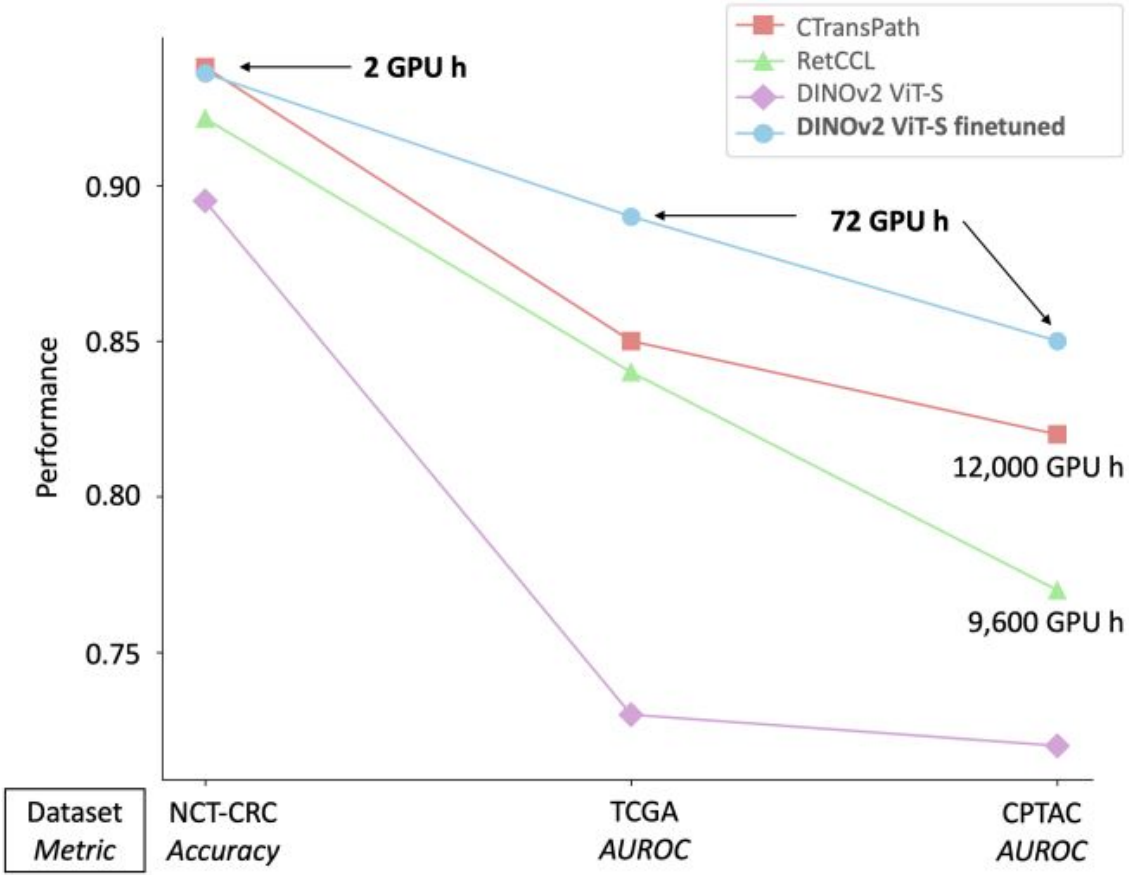
Hypothesis: SSL framework is more important than pretraining dataset and model size

Foundation Models from Academia and Industry



*open source

A Compute-Friendly Path to a Foundation Model

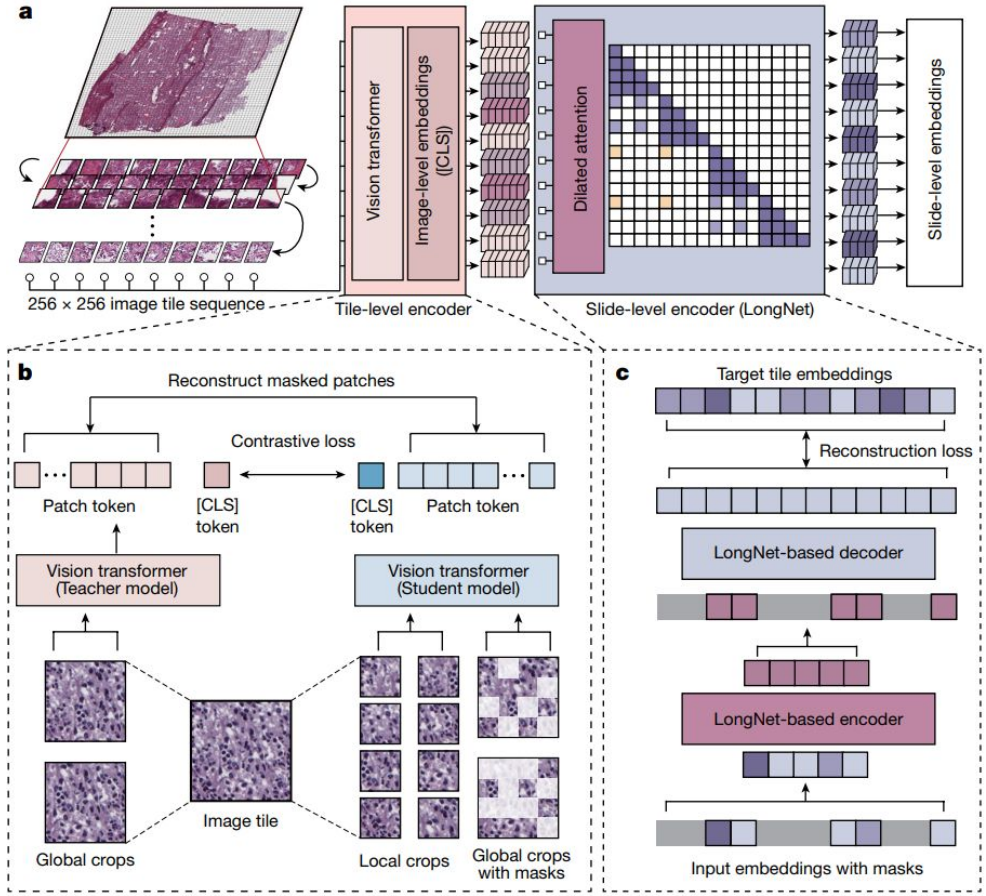


Finetuning publicly-available foundation models uses < 1% of the compute

Source: Roth, Low-Resource Finetuning of Foundation Models Beats State-of-the-Art in Histopathology, 2024 (Technical University of Munich)

Beyond tiles

Slide-Level Foundation Model



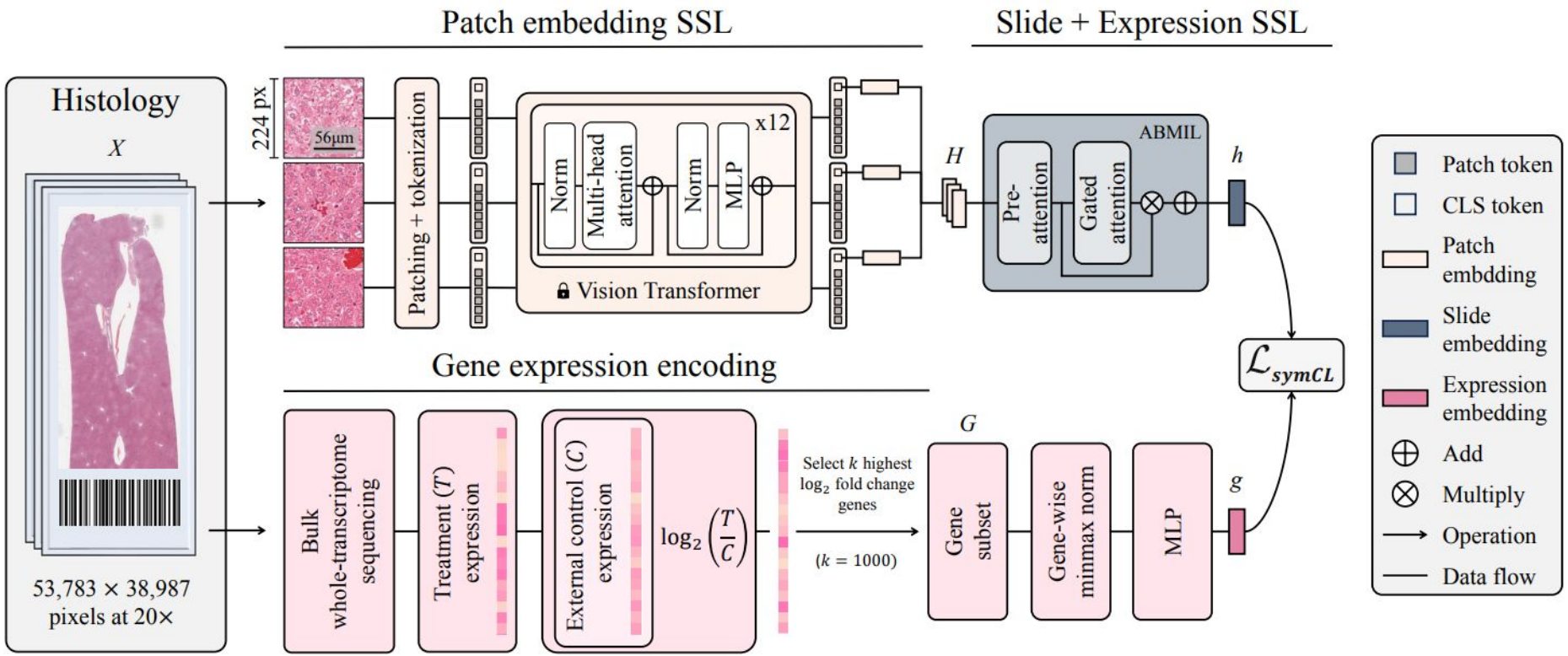
Prov-GigaPath:

- Tile-encoder trained with DINOv2
- Slide-encoder trained with masked autoencoder
- LongNet to handle thousands of tiles per slide
- Trained on 171k WSIs from 30k patients

Source: Xu, A whole-slide foundation model for digital pathology from real-world data, 2024 (Microsoft)

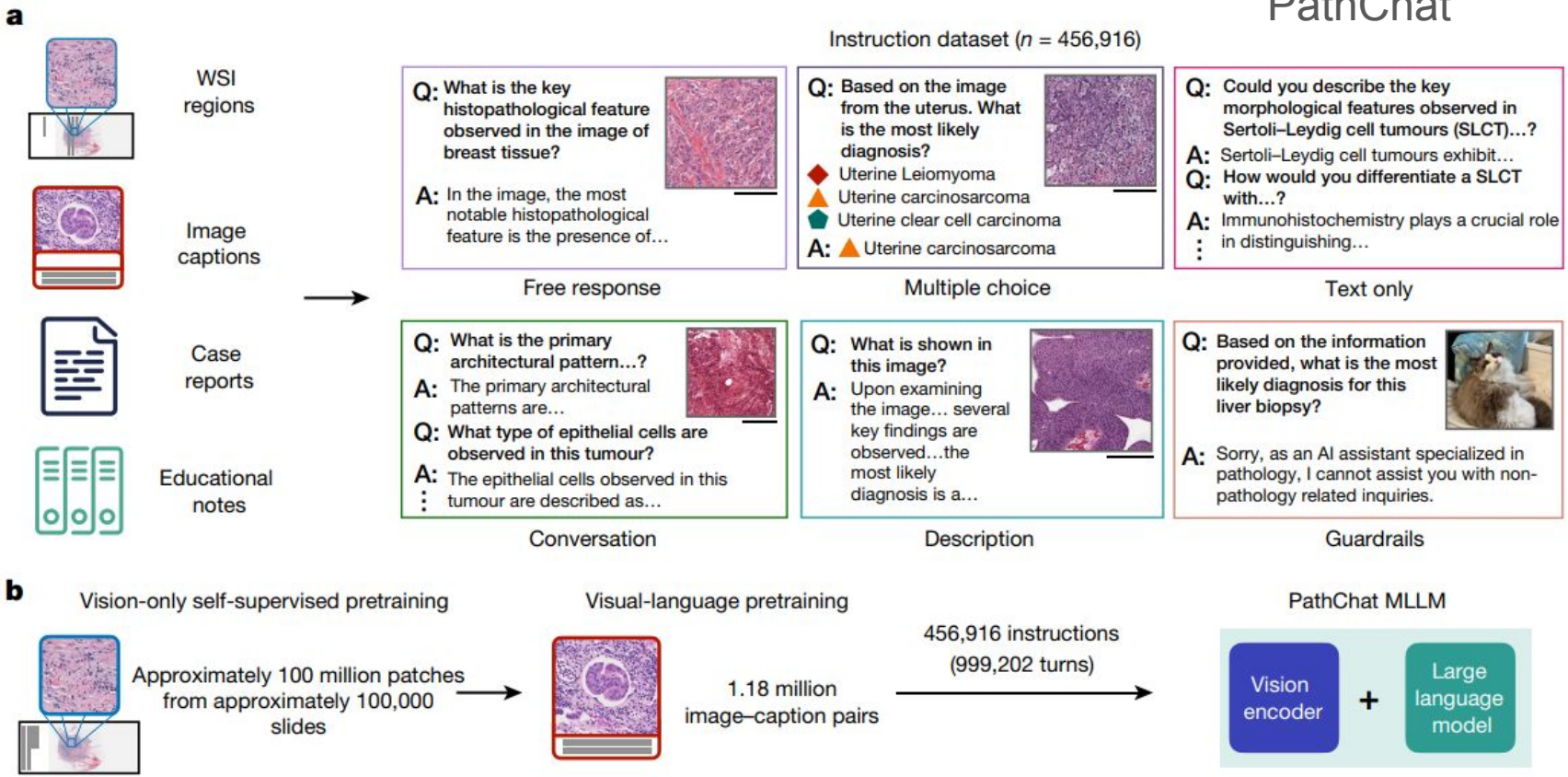
Pretext Task: Gene Expression

TANGLE



Source: Jaume, Transcriptomics-guided Slide Representation Learning in Computational Pathology, 2024 (Harvard)

Multimodal: Pathology + Language



Source: Lu, A multimodal generative AI copilot for human pathology, 2024 (Harvard)

What else do we know about foundation models?

Which foundation model is best for a particular project?

A Clinical Benchmark of Public Self-Supervised Pathology Foundation Models (Campanella, 2024, Mount Sinai/Memorial Sloan Kettering)

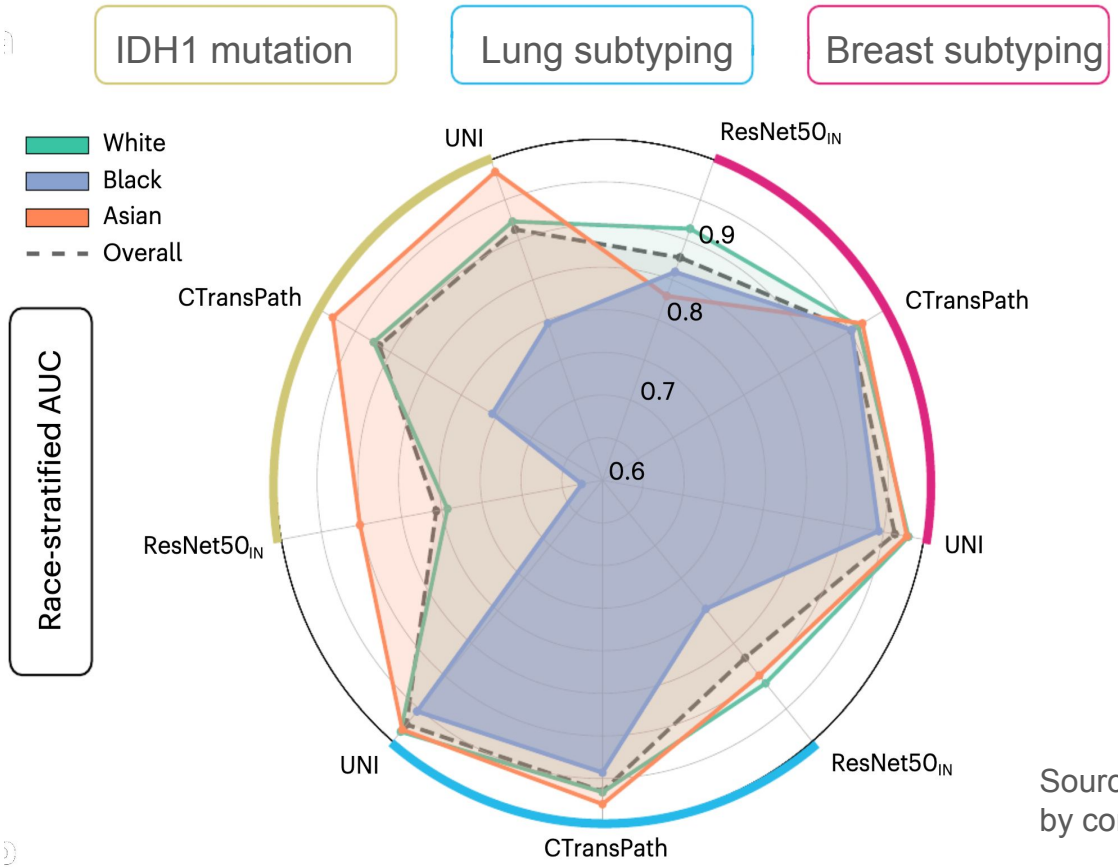
- 9 disease detection and 11 biomarker prediction tasks
- Compared 8 foundation models
- Not released publicly to prevent being scraped into training future foundation models

A Comprehensive Evaluation of Histopathology Foundation Models for Ovarian Cancer Subtype Classification (Breen, 2024, University of Leeds)

- Ovarian cancer morphological subtyping
- Compared 17 feature extraction methods

No single model was best across all tasks

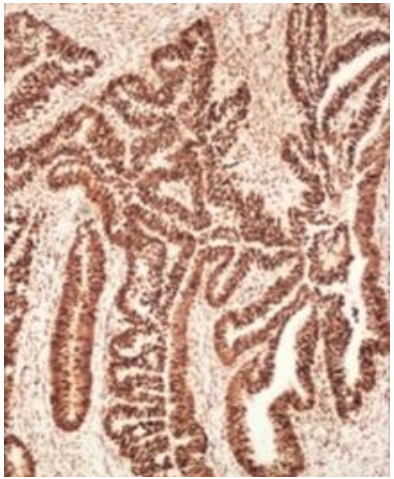
How do foundation models influence reduce bias?



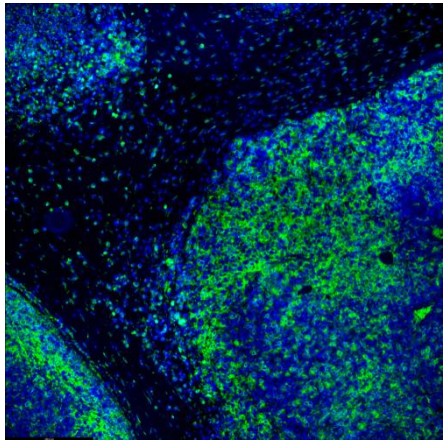
Histology foundation models reduce bias, but don't eliminate it

Source: Vaidya, Demographic bias in misdiagnosis by computational pathology models, 2023 (Harvard)

What about other modalities?

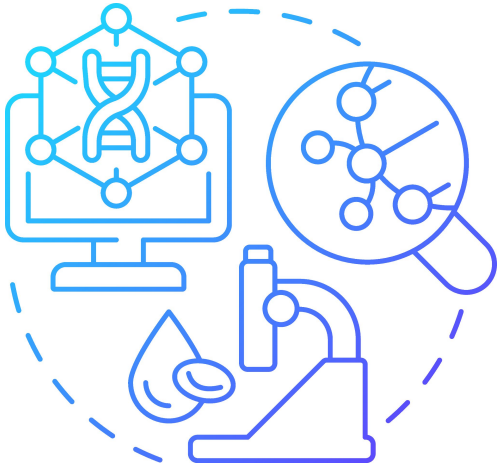


Immunohistochemistry



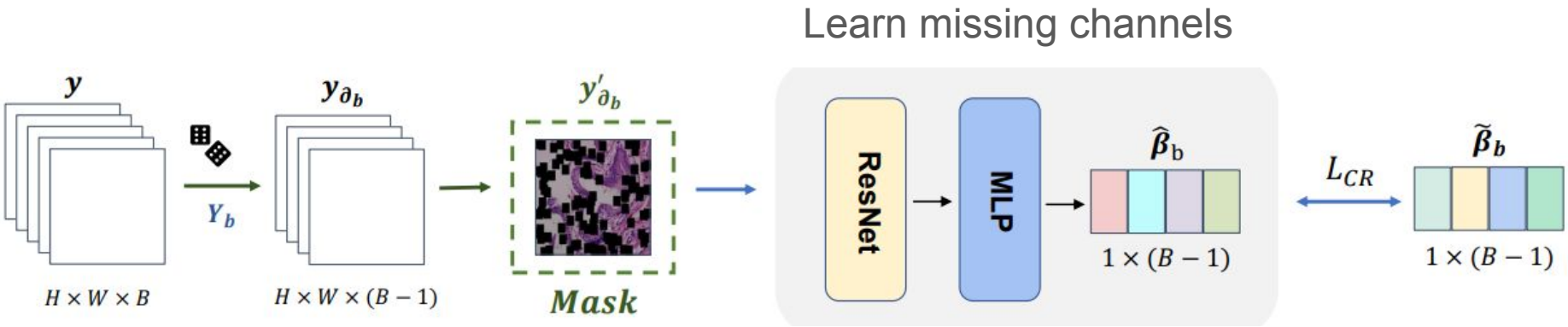
Multiplex immunofluorescence

Multispectral point-of-care histology



‘Omics

Multi/Hyperspectral Imagery



Source: Xie, S3R: Self-supervised Spectral Regression for Hyperspectral Histopathology Image Classification, 2022 (East China Normal University)

Can foundation models can be trained on small datasets?

Method \ Dataset	Breast n=1820	Colorectal n=5000
DCL [4]	0.939 / 0.939	**0.921 / 0.921**
SimCLR [1]	0.925 / 0.924	0.874 / 0.874
SimSiam [2]	0.908 / 0.907	0.898 / 0.898
VICReg [3]	**0.939 / 0.939**	0.915 / 0.915
ImgNet+TL	0.916 / 0.915	0.911 / 0.910

Source: Chattopadhyay, Exploring Self-Supervised Representation Learning for Low-Resource Medical Image Analysis, 2023 (Jadavpur University)

Will this work for even smaller datasets?

What’s next?

- What makes a good foundation model?
  - Look beyond larger datasets and models
  - How does training data selection impact bias?
  - How do SSL frameworks compare?
  - Ablation studies
- How can we train foundation models more efficiently?
- Can we get equivalent performance from a smaller model?
- What techniques are best when pretraining data is limited?
- How to select a foundation model? Can we do a better job of benchmarking?
- Do foundation models solve distribution shift?
- How can this knowledge be applied beyond H&E?

Resources

https://pixelscientia.com/compayl2024/

heather@pixelscientia.com

Links to these slides, articles, podcasts, and other resources.

Computer Vision Insights Newsletter	A biweekly newsletter that often features the latest research in AI for histopathology.
Impact AI Podcast	Learn how to build a mission-driven machine, learning-powered company from the innovators and entrepreneurs who are leading the way.
Consulting	For founders and other leaders who need a practical computer vision strategy to streamline model development and boost investor confidence.

Pixel Scientia LABS