Demystifying Foundation Models for Pathology

How to accelerate model development without sacrificing accuracy

Heather Couture

January 29, 2025
11 am EST

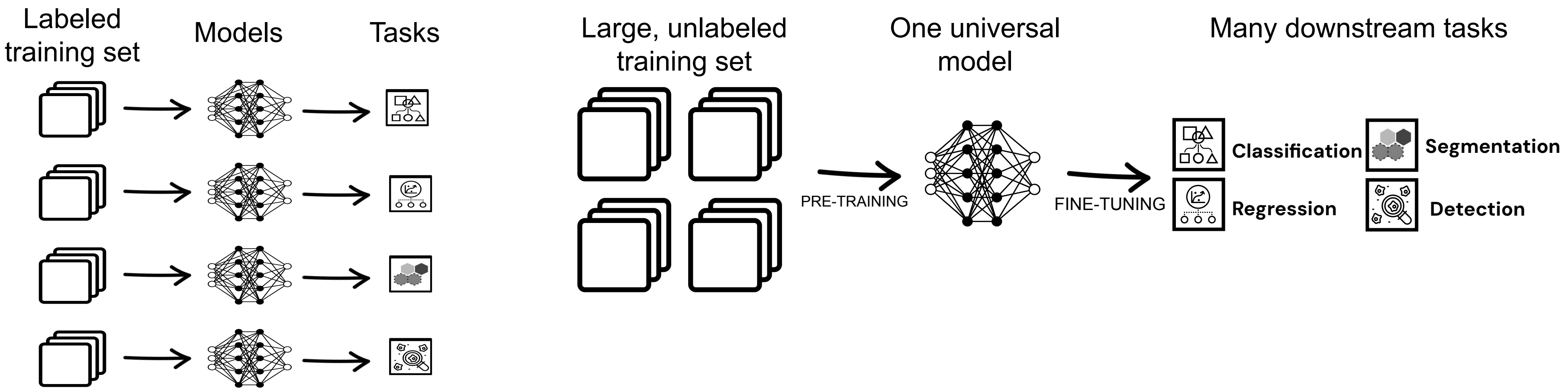
30 minutes + Q&A

Pixel Scientia LABS

The Paradigm Shift

Traditional ML

Foundation Models



How does this approach work for pathology?

- Qualitative image analysis
- Diagnostics
- Drug discovery
- Clinical trial planning

Goals of this webinar

- What a foundation model is
- Why use foundation models
- How to use foundation models
- How to apply foundation models beyond image tiles

Who am I?

- Heather Couture
- Computer vision consultant

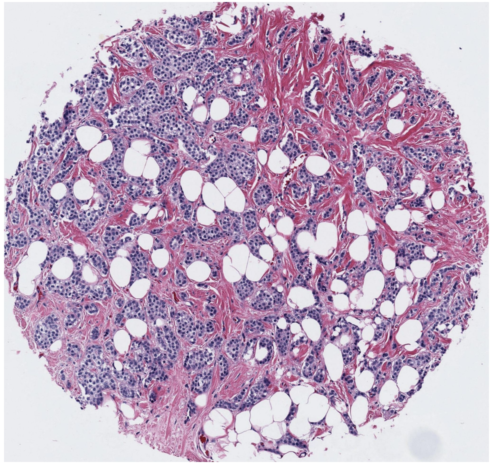
OWKIN DECIPHEX ULTIVUE AGENDIA Leica MICROSYSTEMS QRITIVE

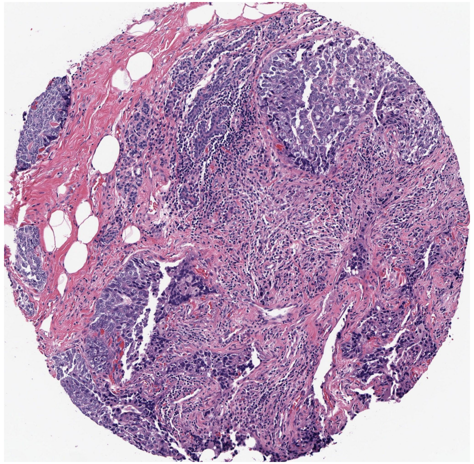
- Keynote speaker at MICCAI workshop on computational pathology
- Contributor to Scientific American, The Pathologist, DPA Blog
- Newsletter and podcast

Computer Vision Insights by Pixel Scientia Labs

- PhD in Computer Science from University of North Carolina

At UNC we created methods to predict tumor-level labels for breast cancer





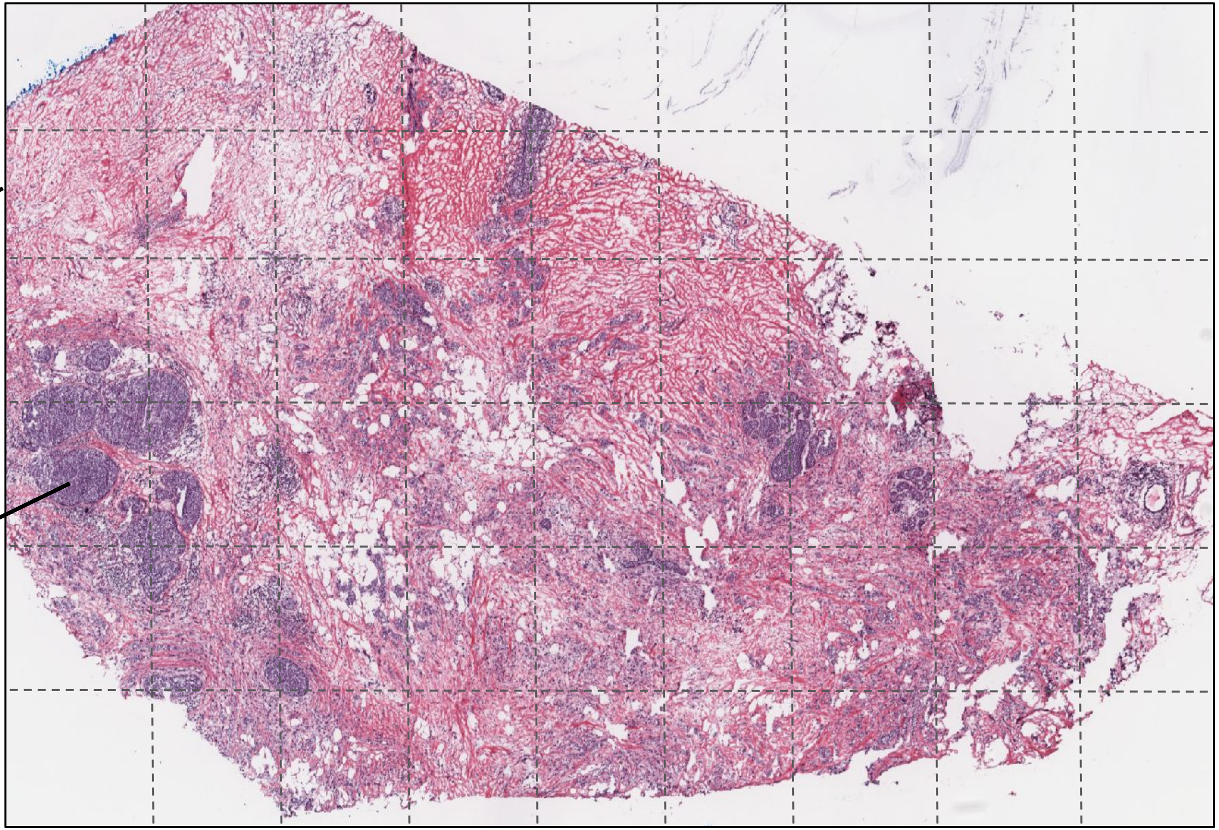
	Class 1	Class 2	Ground Truth
Grade Histologic subtype Estrogen receptor status Genomic subtype	low lobular positive non-Basal	high ductal negative Basal	pathologist pathologist immunohistochemistry gene expression

Couture, et al., **Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype**, 2018
Couture, et al., **Multiple Instance Learning for Heterogeneous Images: Training a CNN for Histopathology**, 2018

Without detailed annotations, this is called weakly supervised learning

patient or slide
label given

tile label unknown

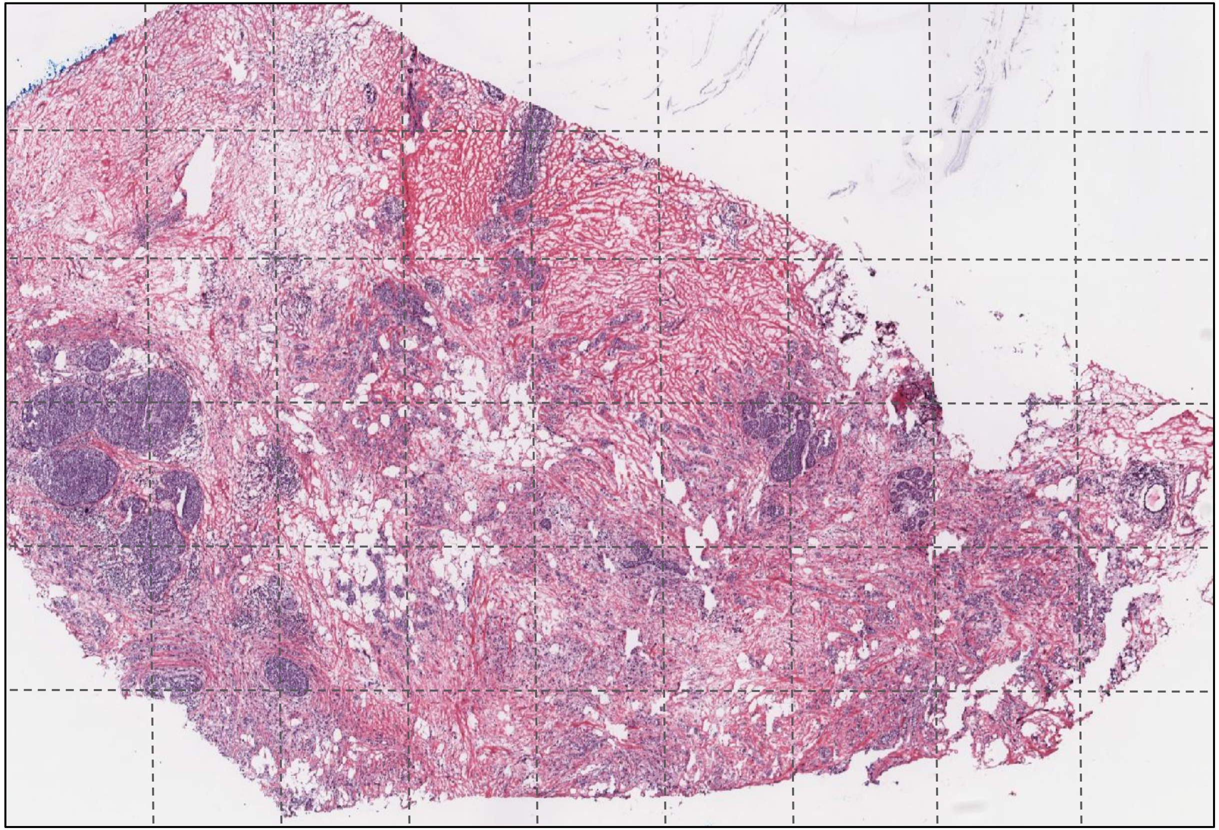


A basic model for predicting biomarkers from whole slide images

1. Extract features from tiles
2. Aggregate tile features
3. Predict class

v1: Extract features with ImageNet model

Can we do better?

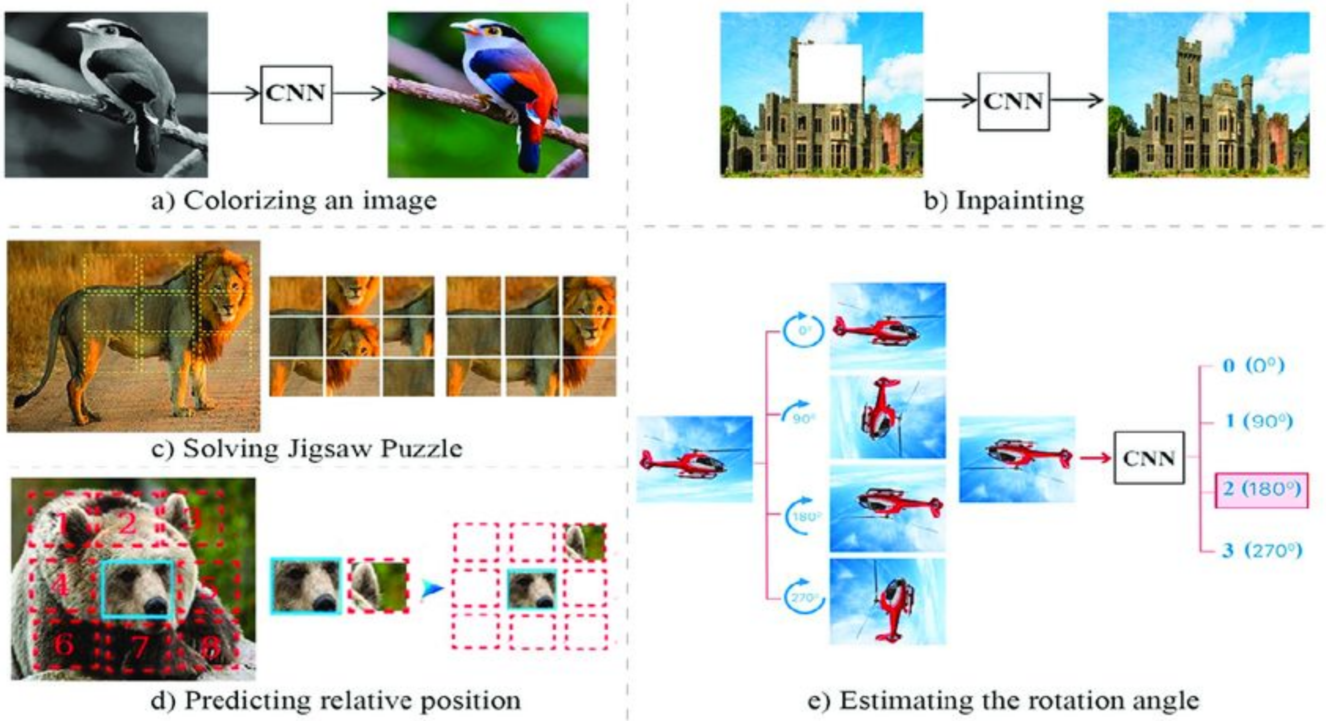


“Foundation model” coined by Stanford Institute for Human-Centered Artificial Intelligence's Center for Research on Foundation Models

“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.”

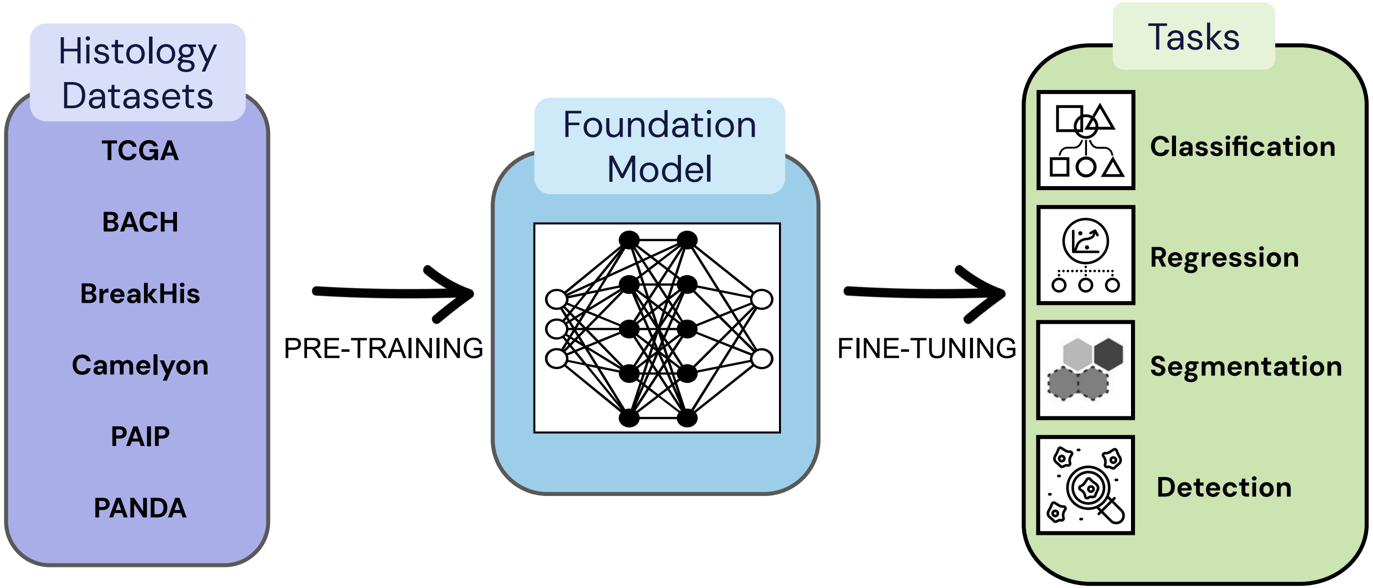
Source: Bommasani, On the Opportunities and Risks of Foundation Models, 2021

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



Source: Albelwi, Survey on Self-Supervised Learning: Auxiliary Pretext Tasks and Contrastive Learning Methods in Imaging, 2022

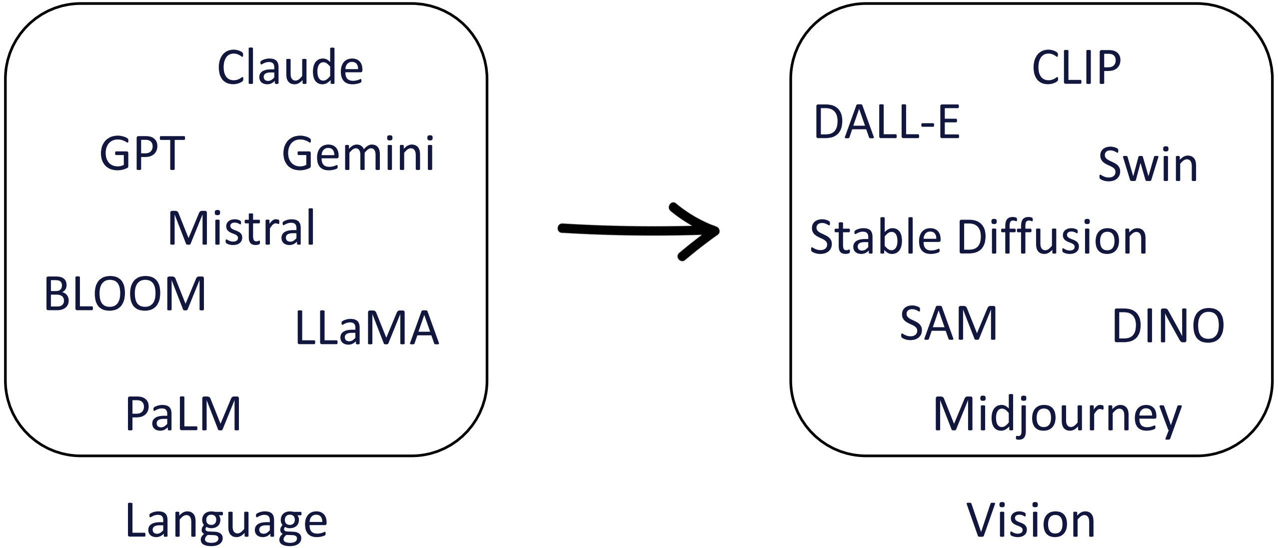
Foundation Models for Histopathology



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

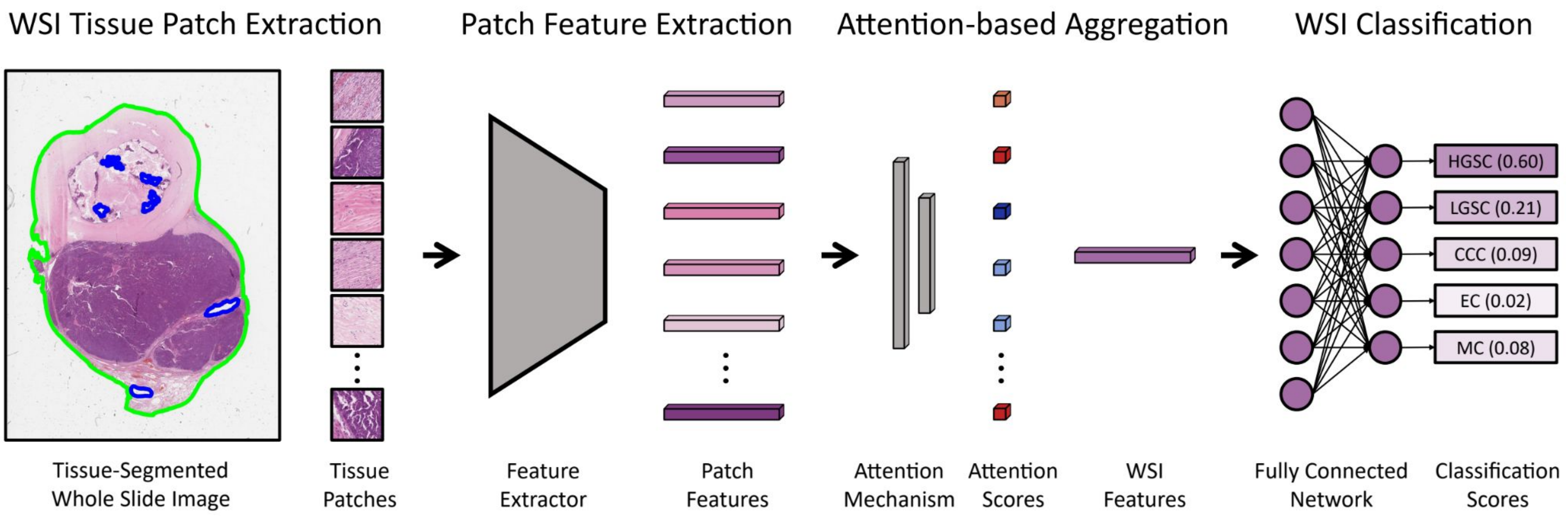
How did we get here?

From Large Language Models to Large Vision Models



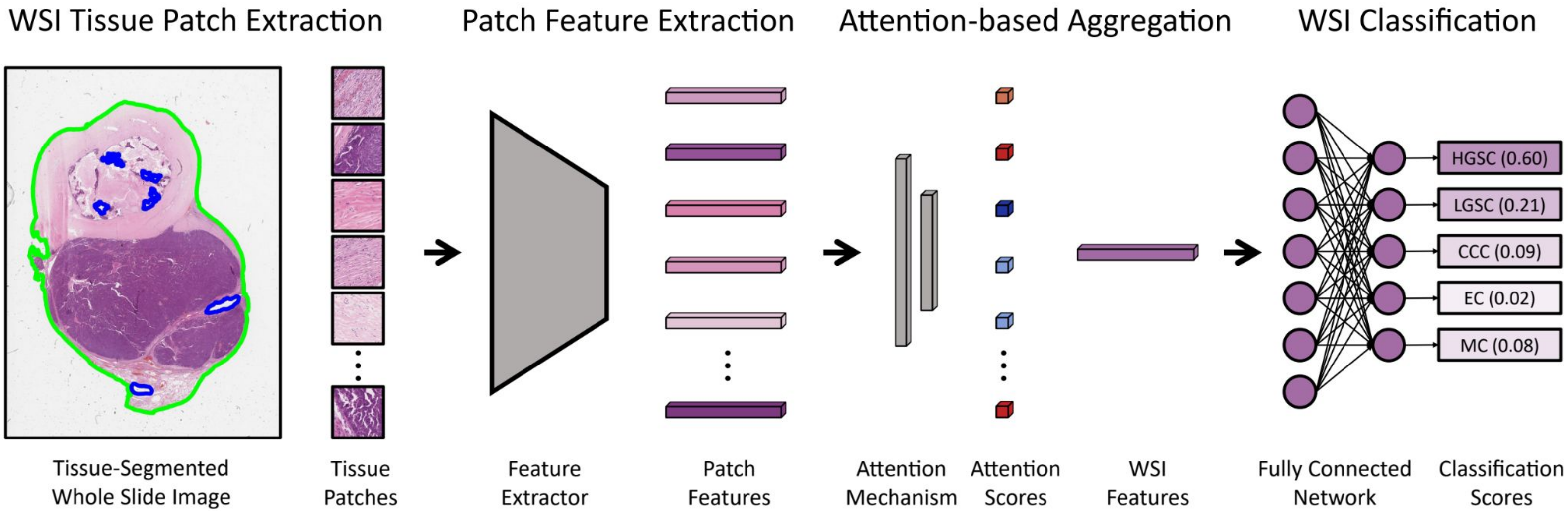
Easy to adapt to new tasks

From ImageNet Pre-Training to Self-Supervised Learning

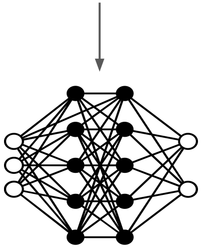


Source: Breen, A Comprehensive Evaluation of Histopathology Foundation Models for Ovarian Cancer Subtype Classification, 2024

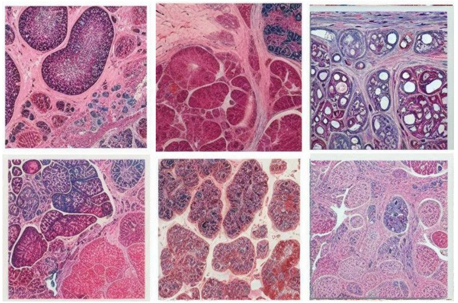
From ImageNet Pre-Training to Self-Supervised Learning



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

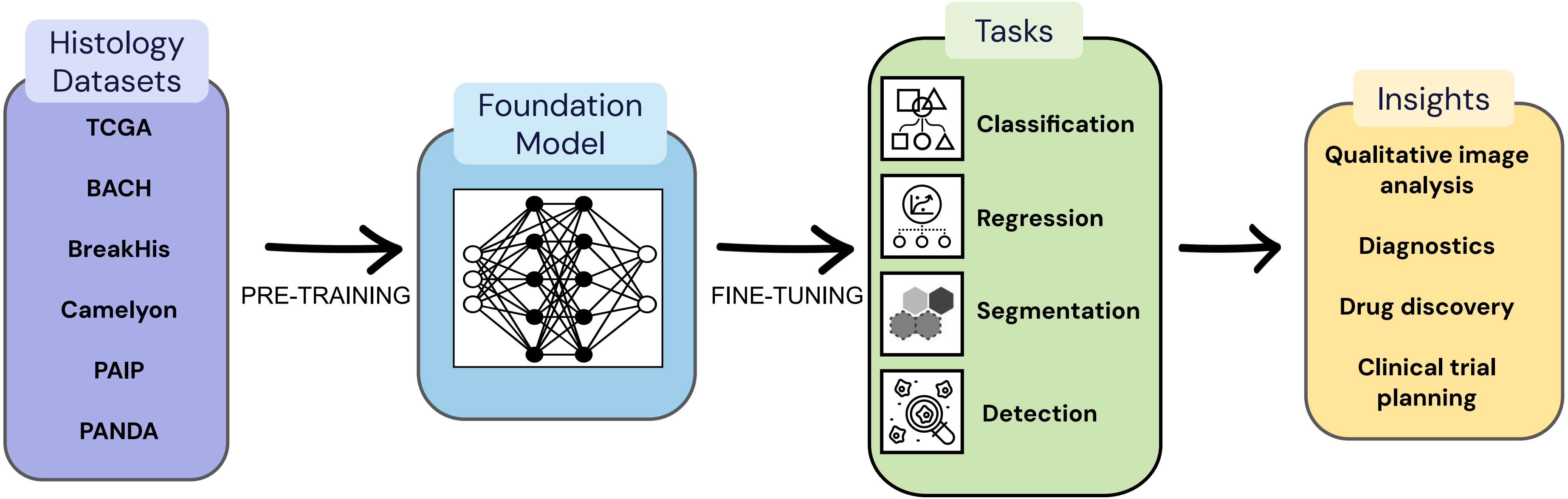


ImageNet vs. Histology



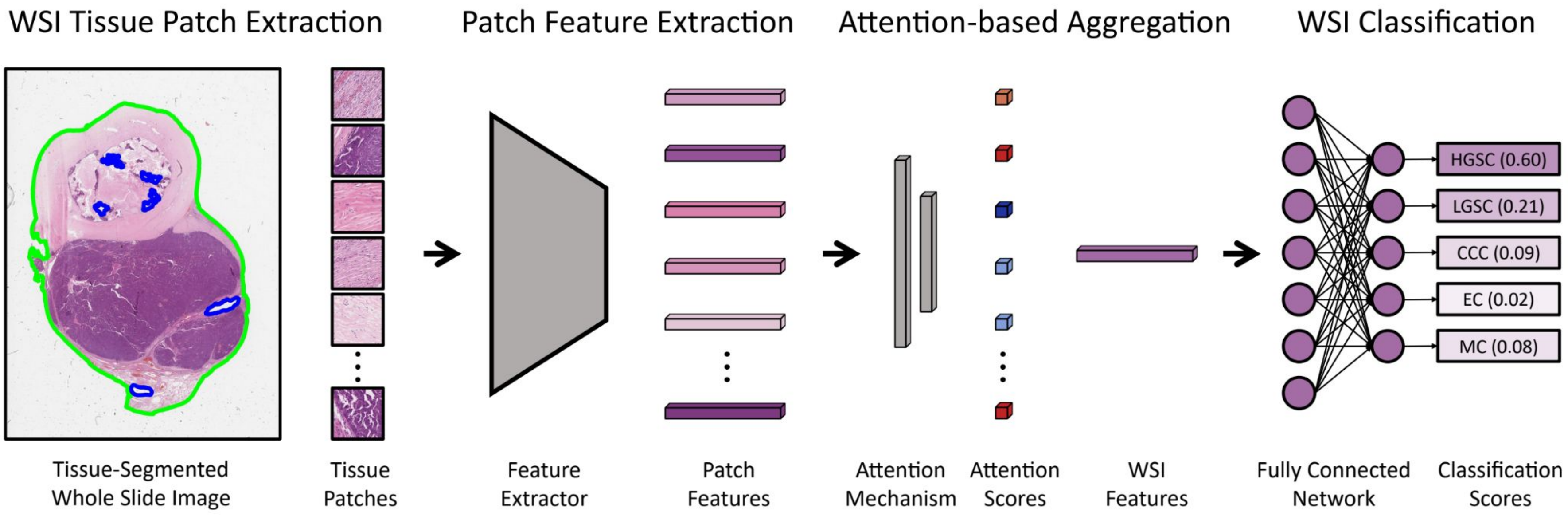
A more generalizable tile embedding

Foundation Models for Histology: Adaptable & Generalizable



Why use foundation models?

Gigapixel Images and Weak Supervision

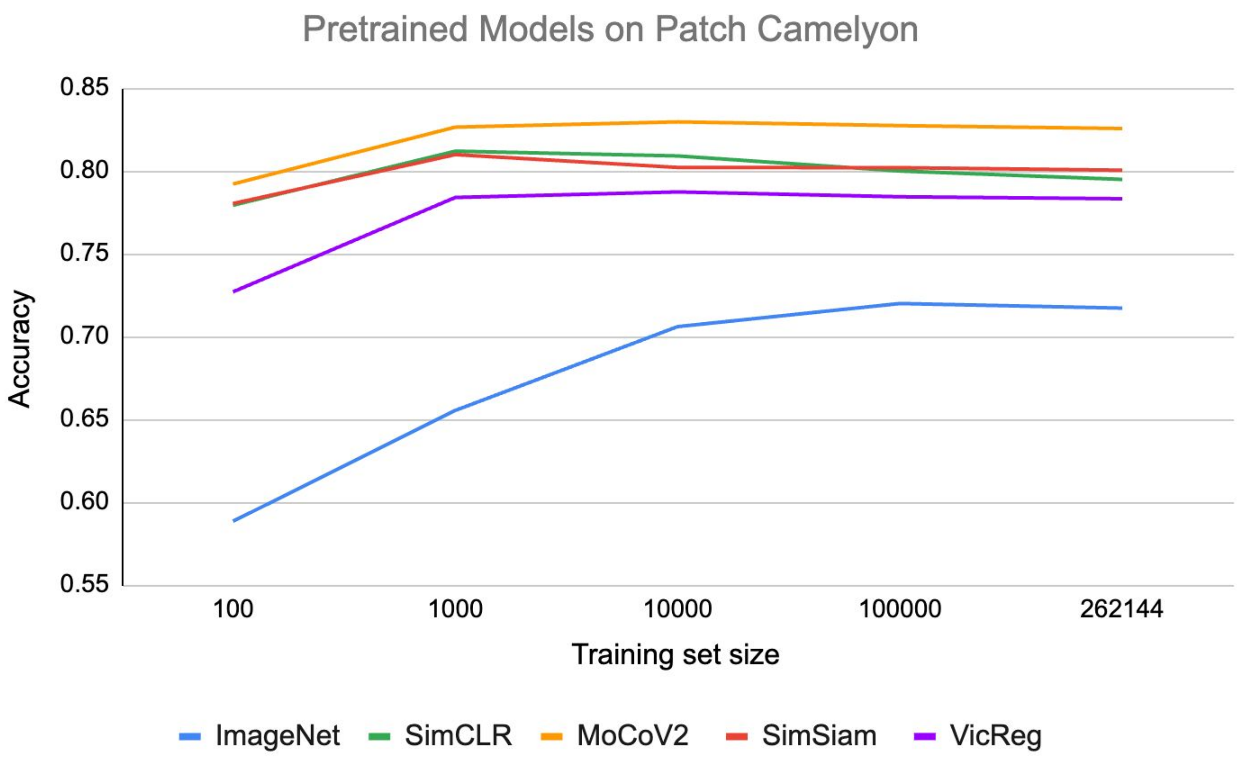


Slide- or patient-level characteristics:

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

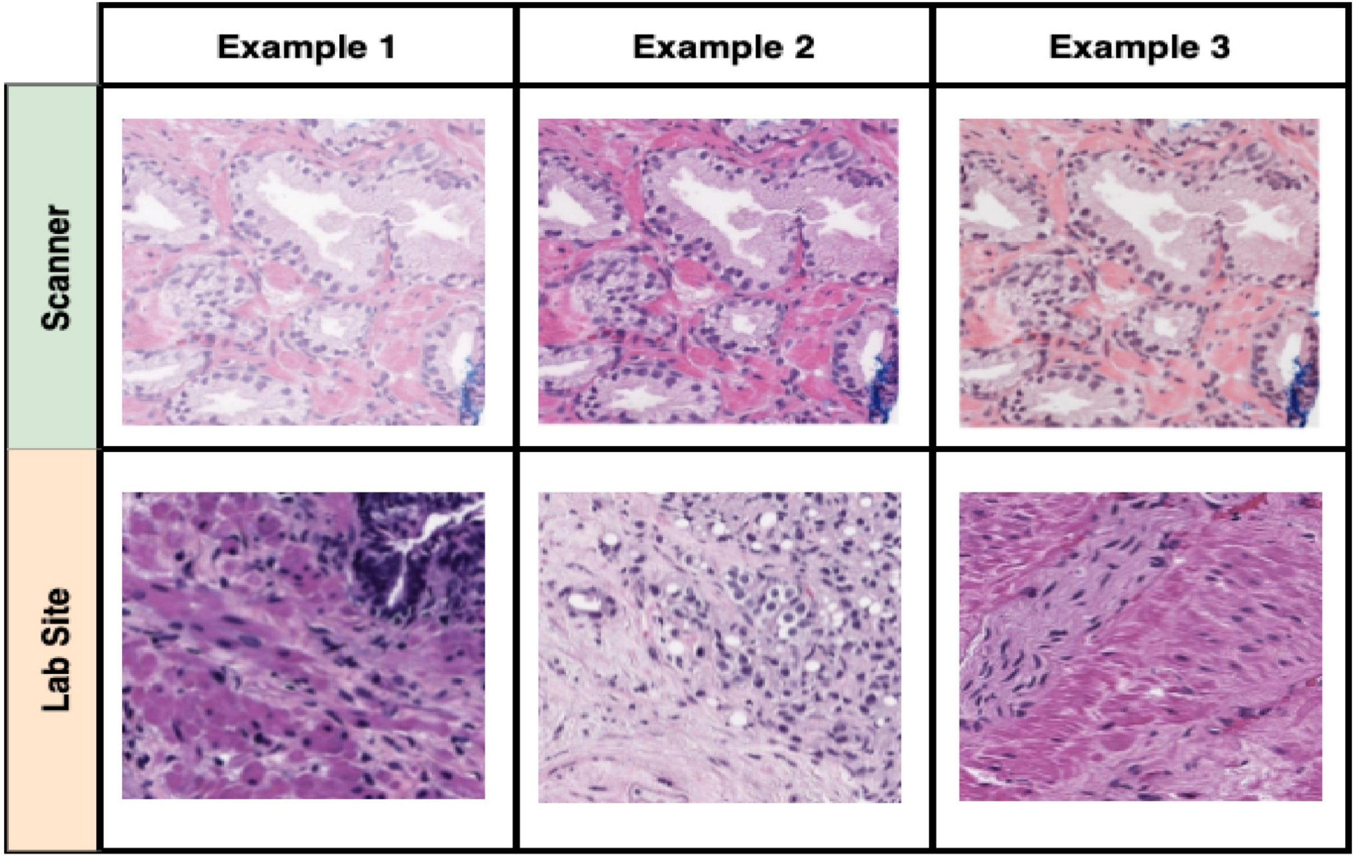
Source: Breen, A Comprehensive Evaluation of Histopathology Foundation Models for Ovarian Cancer Subtype Classification, 2024

Limited Labeled Data



Improved accuracy with fewer labels

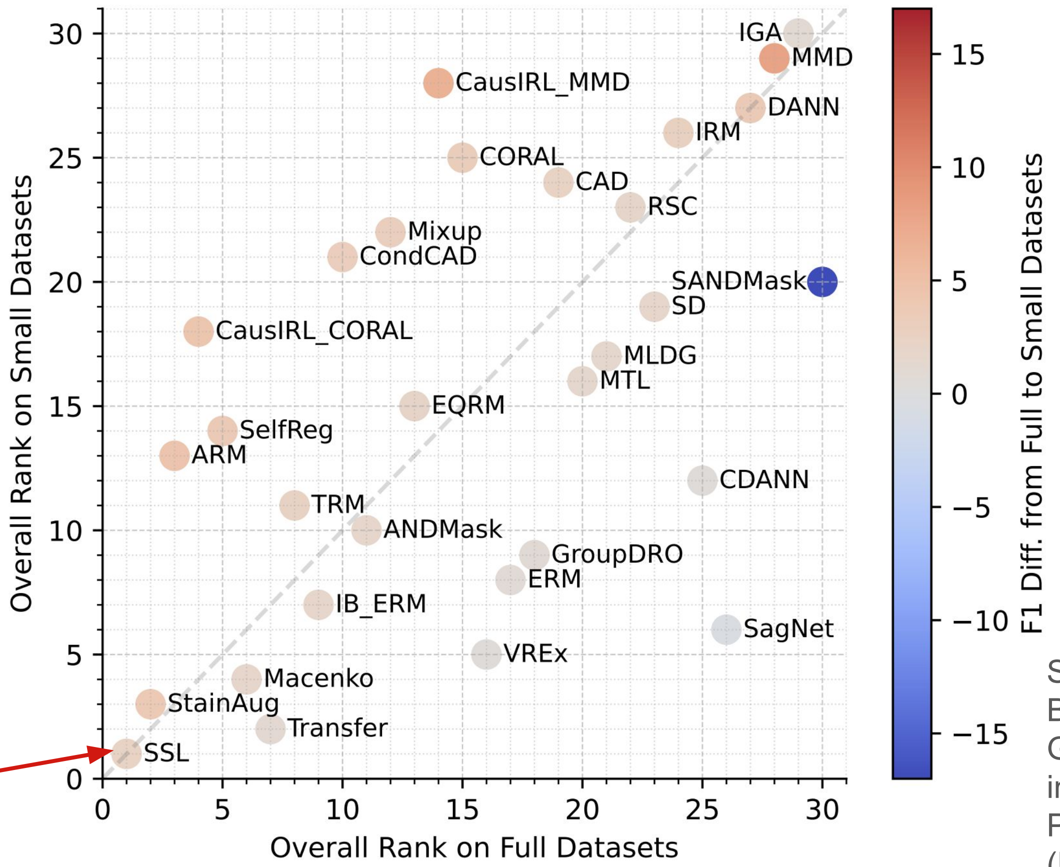
Robustness to Distribution Shifts



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

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

Robustness to Distribution Shifts



SSL better than any other approach

Source: Zamanitajeddina, Benchmarking Domain Generalization Algorithms in Computational Pathology, 2024 (University of Warwick)

Rapid Prototyping

Patch classification:

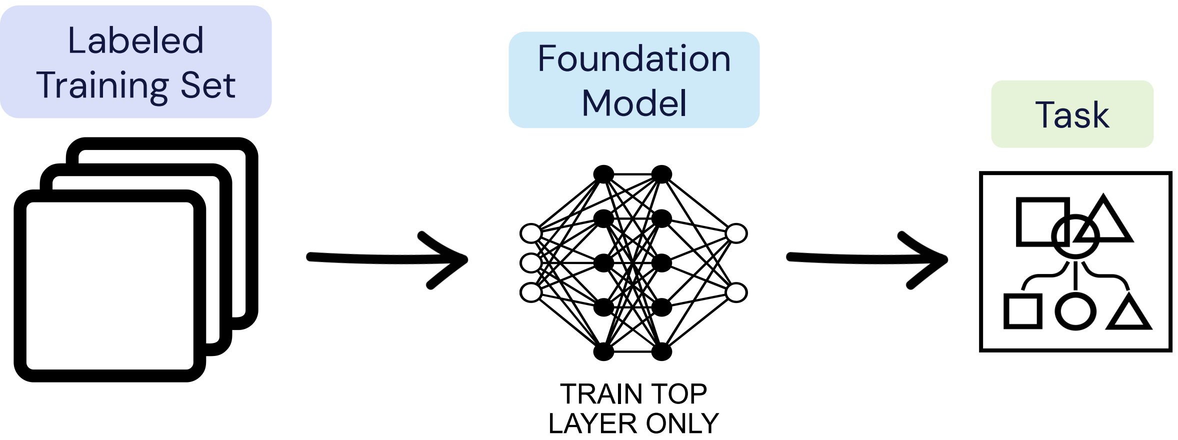
1) Label 100 tiles
2) Extract features from each tile using foundation model
3) Train logistic regression classifier using features
4) Validate and iterate!

Slide classification:

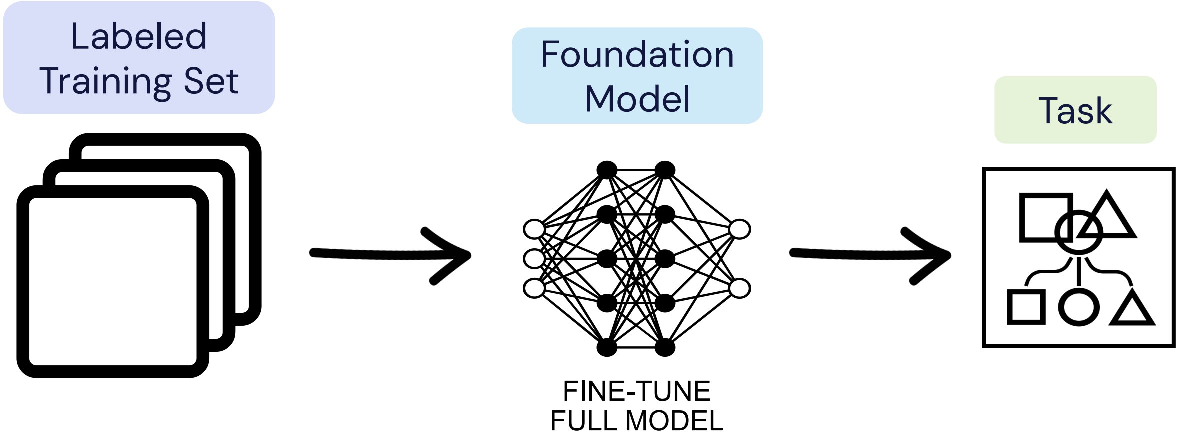
1) Label 100 slides
2) Tile each slide and exclude non-tissue tiles
3) Extract features from tissue tiles using foundation model
4) Compute mean feature vector for each slide
5) Train logistic regression classifier on slide features
6) Validate and iterate!

How to use foundation models?

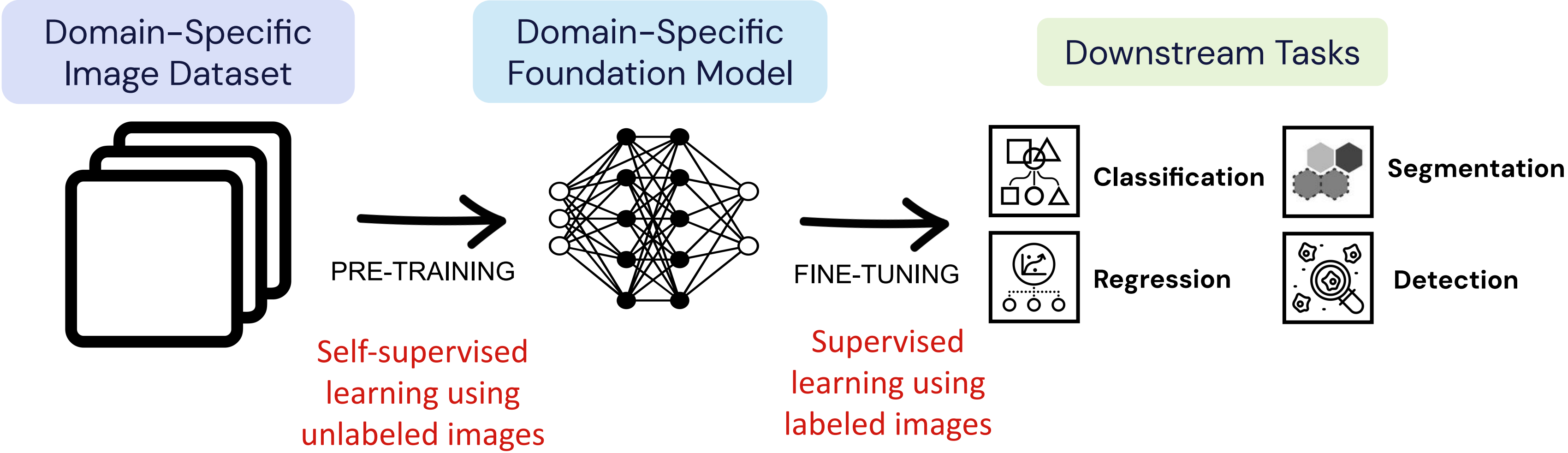
Linear probing



Fine-tune for task



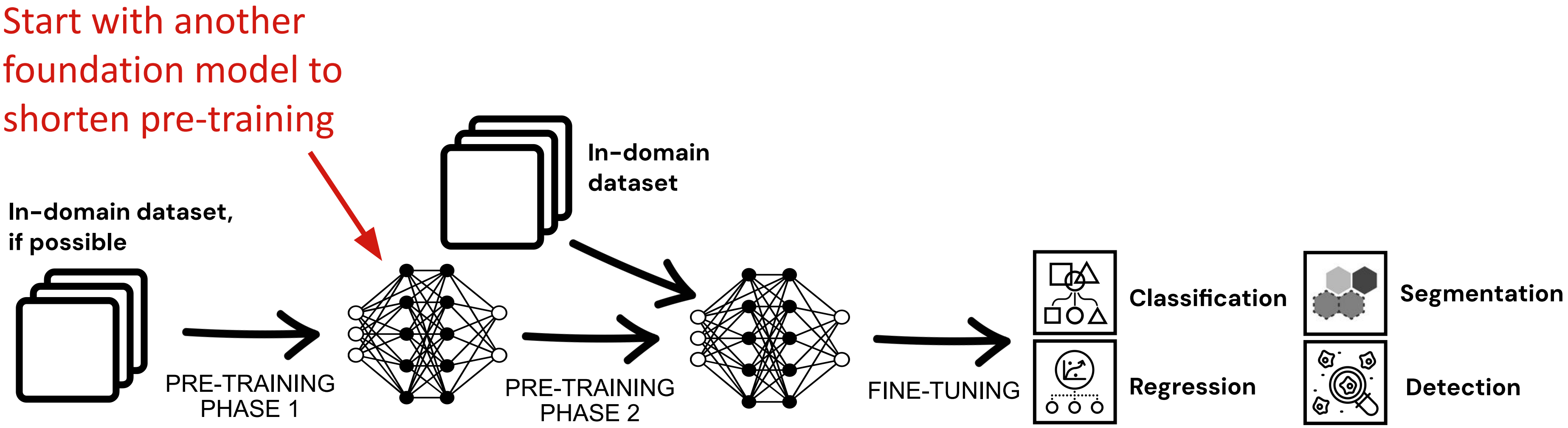
Pre-train from scratch



- Most computationally-intensive step
- Done once
- Uses a pretext task instead of manual labels

- For each downstream task
- Less computation

Additional Pre-Training



Comparison of Uses

	Linear Probing	**Fine-Tuning**	**Additional Pre-Training**	**Pre-Train from Scratch**
Pros	Quick Little computation Low risk of overfitting	Learn fine-grained features	Learn a better representation for your domain	Learn a representation exclusively for your domain
Cons	May miss subtle features	Requires more labeled training data Can overfit	Extra computation and training time	Requires significant computation and training time
When to use	Quick prototype Obvious features	Linear probing insufficient	Your data is more specialized Unlabeled images available	You images are quite different than existing foundation models Lots of unlabeled images available

Published Foundation Models for Histopathology Tiles

Model Name	Backbone	Parameters	SSL Algorithm	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*	ViT-H	631 million	DINOv2	3.1m
Atlas	ViT-H	632 million	DINOv2	1.2 m
UNI2-h*	ViT-H	681 million	DINOv2	350k
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

* open weights

How to select a model?

1) Was it trained on images like yours?
2) Did it perform well on tasks like yours?
3) How important are compute costs?

Which foundation model is best for a particular task?

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

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

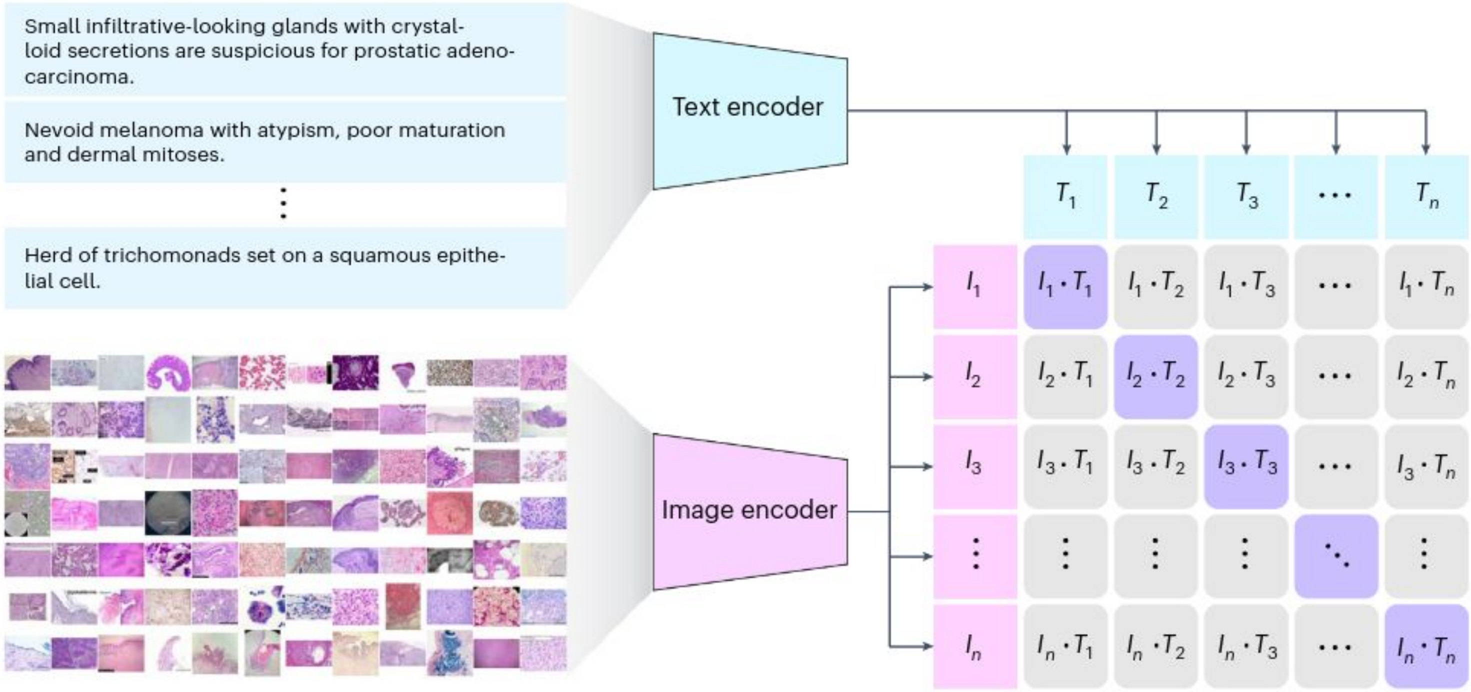
- Ovarian cancer morphological subtyping
- Compared 17 feature extraction methods

No single model was best across all tasks

Beyond Tiles

Multimodal: H&E + Language

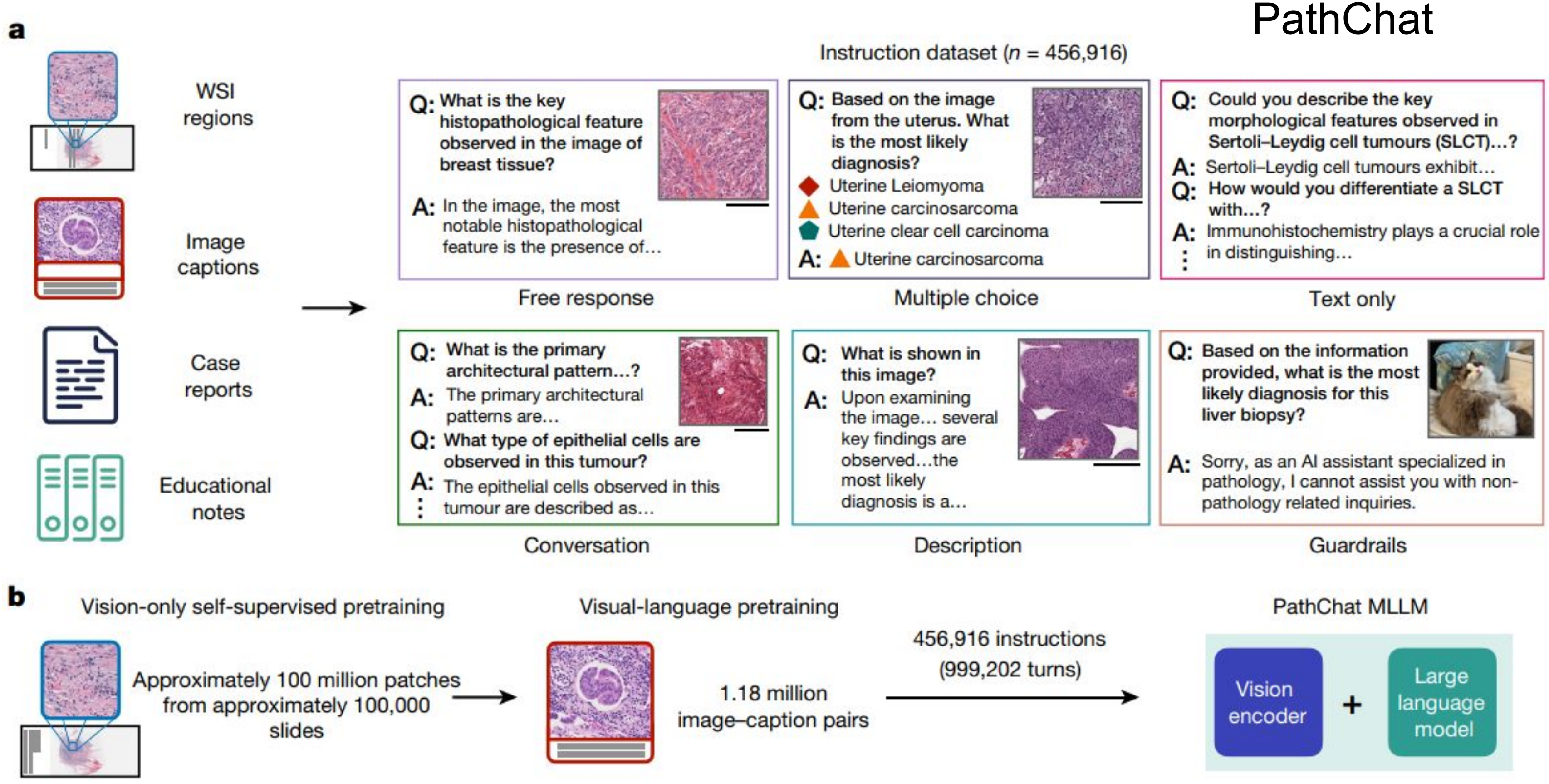
PLIP: Fine-tuned CLIP on 208k pathology images + natural language description pairs



Source: Huang, A visual–language foundation model for pathology image analysis using medical Twitter, 2023 (Stanford)

Multimodal: H&E + Language

PathChat



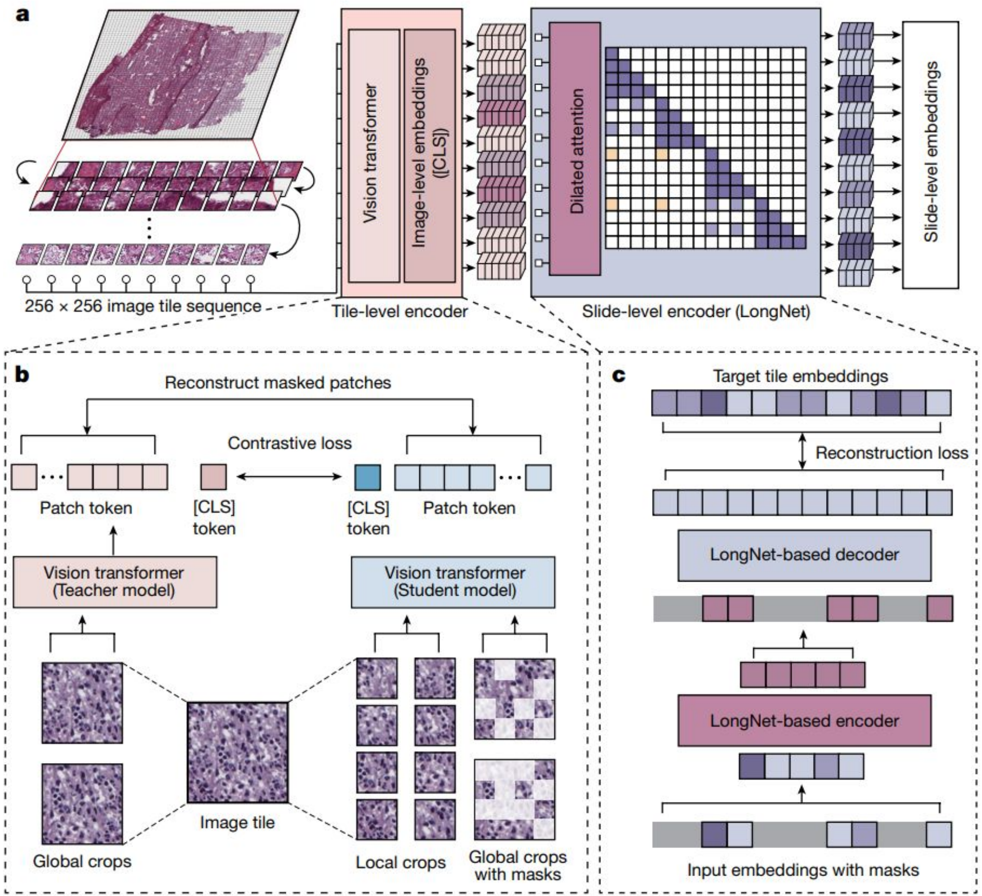
Source: Lu, A multimodal generative AI copilot for human pathology, 2024 (Harvard)

Multimodal Tile Foundation Models

Model Name	Additional Modality	Pairs
PLIP*	Text descriptions	208k
CONCH*	Image captions	1.17m
PathChat	Image captions + Q&A	1.18m
MUSK*	Clinical reports	1m

* open weights

Slide-Level Foundation Models



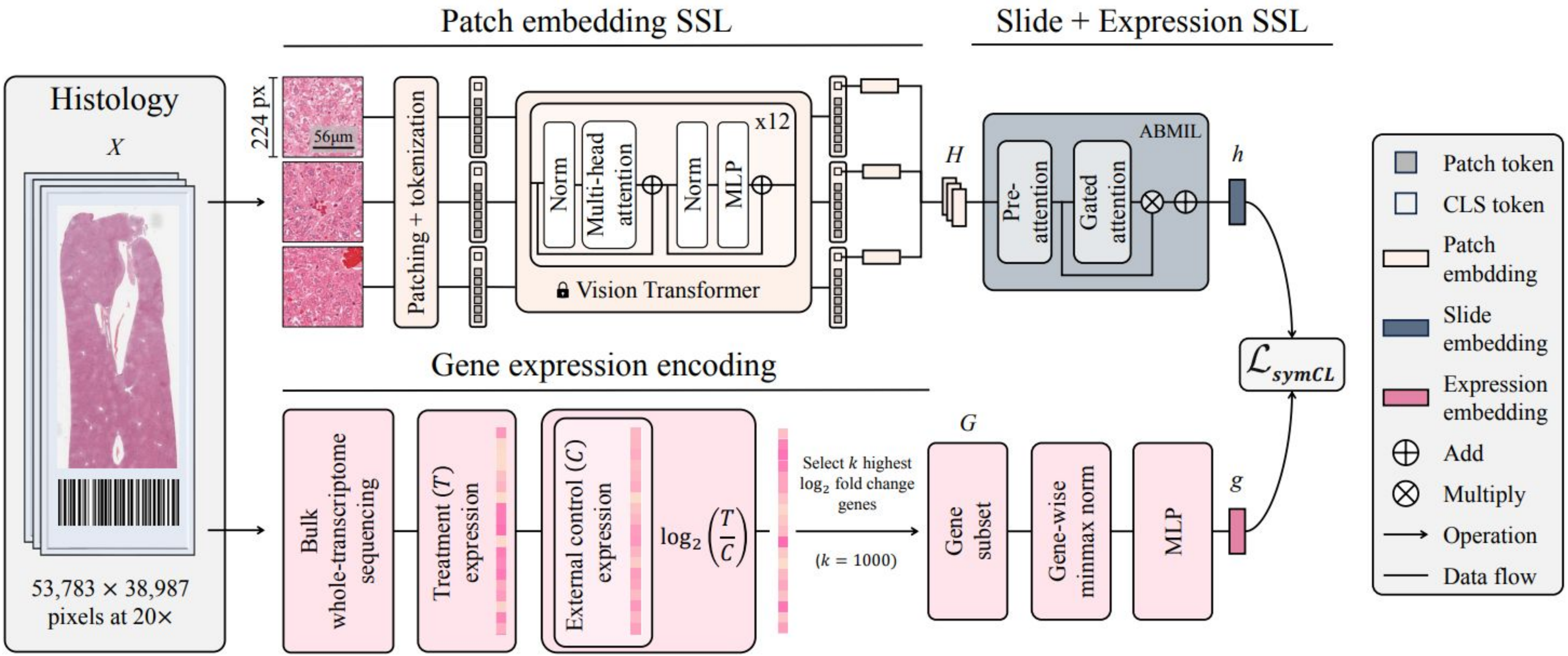
Prov-GigaPath:

- 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)

Slide-Level Multimodal: H&E + Gene Expression

TANGLE



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

Slide-Level Models

Model Name	Additional Modality	SSL Algorithm	WSIs
GigaSSL*		SimCLR	12k
PRISM*	Clinical reports	Contrastive	587k
Prov-GigaPath*		Masked autoencoder	171k
MADELEINE*	IHC	Contrastive	16k
TANGLE*	Gene expression	Contrastive	9k
mSTAR	Path reports, RNASeq	Contrastive	26k
CHIEF*		?	61k
COBRA*	Multiple FM embeddings	MoCo-V3	3k
TITAN*	Clinical reports	iBOT	335k

* open weights

Summary

Why use foundation models for pathology?

- Weakly supervised whole slide images
- Scarce labeled data
- Robustness to distribution shifts
- Rapid prototyping

How?

- Linear probing
- Fine-tune
- Additional pre-training
- Pre-train from scratch

Going beyond tiles:

- Multimodal
- Slide-level

Ready to get started?

How I work with clients

Who I work with:

- Founders and other leaders
- Their technical team

Example results:

- Streamline model development
- Keep up with AI trends and innovations
- Boost investor confidence

Advisory services:

- Monthly strategy call
- Weekly office hours
- Private Slack channel
- And more

Bonus Offer

Team Workshop: Harnessing the Power of Foundation Models for Pathology

In just 90 minutes, you'll gain:

- Deep insights into cutting-edge foundation model architectures
- Practical strategies for using foundation models in your workflow
- A concrete action plan for starting with foundation models

$1,500 (25% discount if booked in the next 48 hours)

Resources

Team workshop bonus offer (25% discount if booked in the next 48 hours):

http://pixelscientia.com/workshops/pathology-foundation-models

Other consulting services: heather@pixelscientia.com

Pixel Scientia LABS