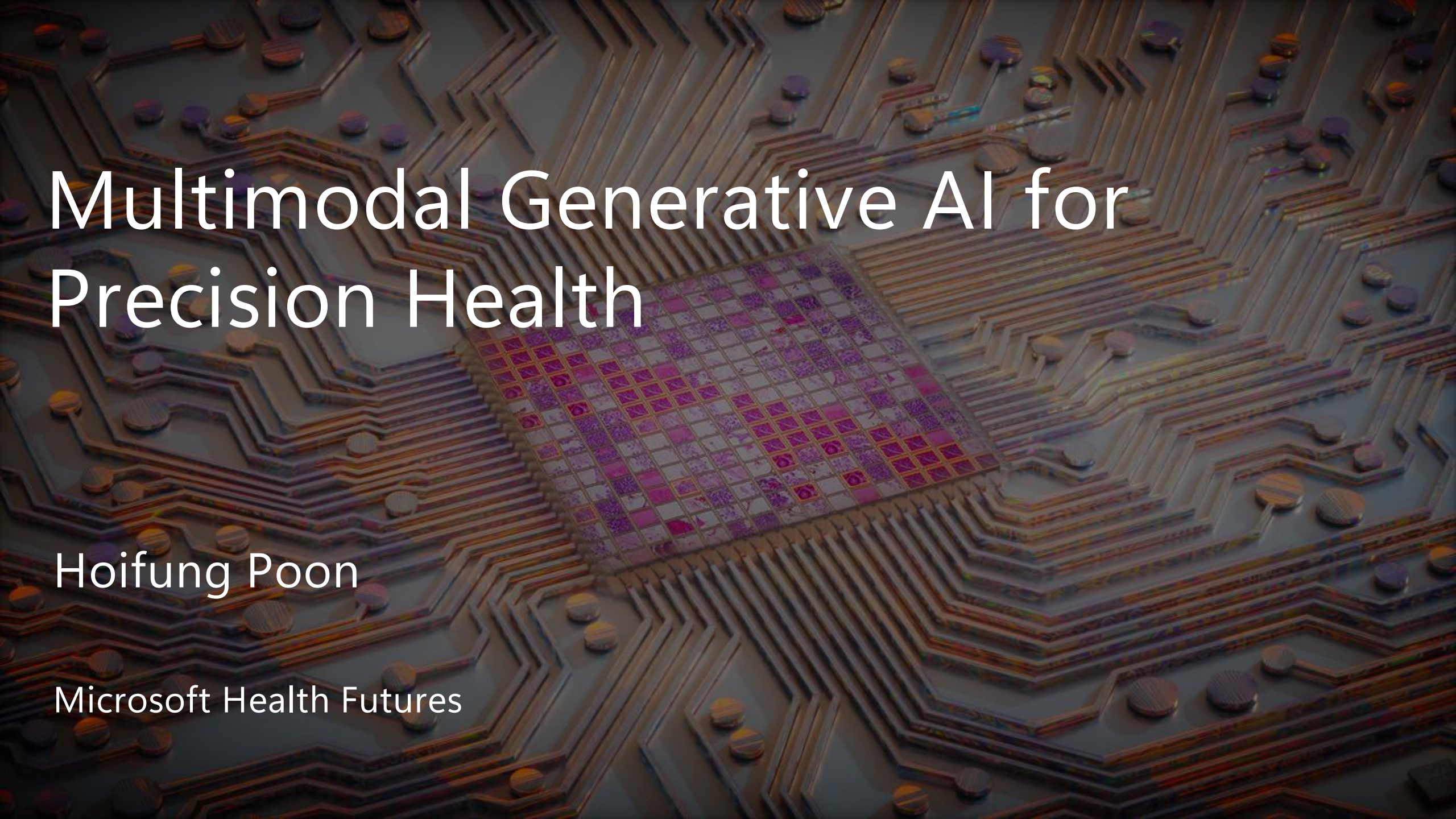
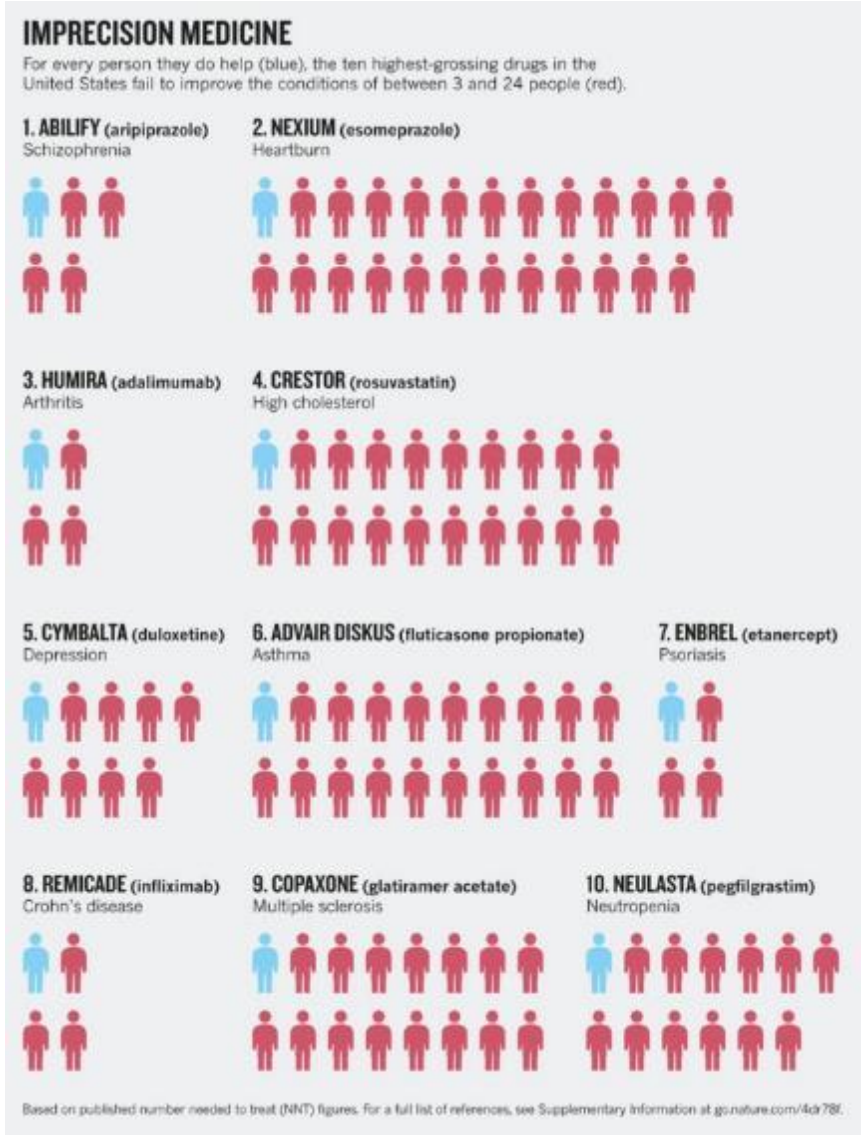


Multimodal Generative AI for Precision Health

Hoifung Poon

Microsoft Health Futures

Medicine Today Is Imprecise



Top 20 drugs
80% non-responders

Wasted
1/3 health spending
$1 Trillion / year

Cancer: Immunotherapy

Keytruda: immunotherapy blockbuster ($25B, 2023)

FDA approved for many cancer indications

But only work for minority of patients

Wanted: Complex biomarker for better stratification

Insight Consumer

Pharma, Payor, Regulator

Data Producer

Provider, EHR Vendor

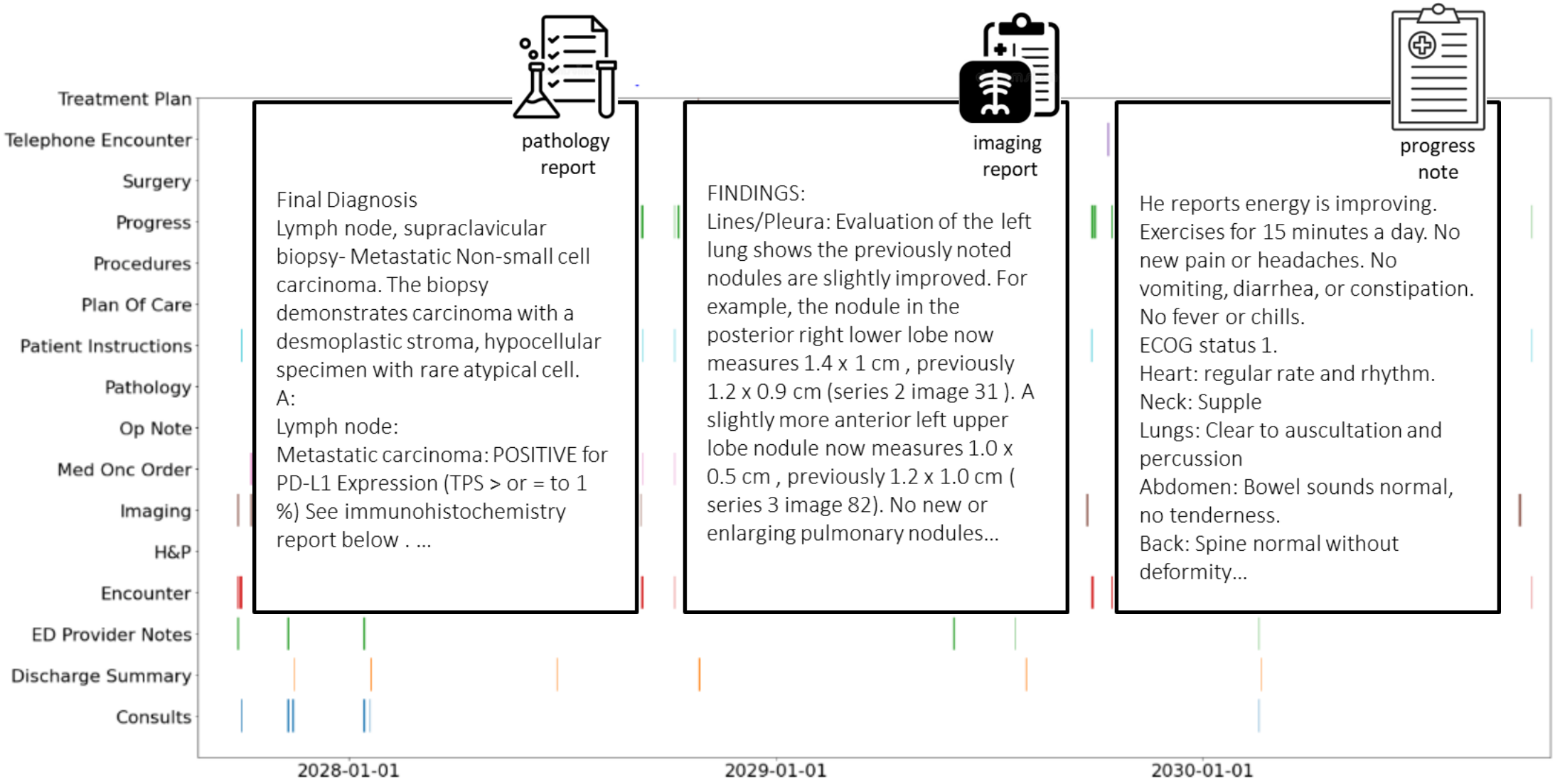
Real-World Evidence (RWE)

Real-world data (RWD) are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.

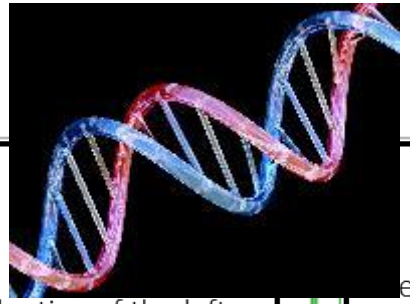
Real-world evidence (RWE) is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD

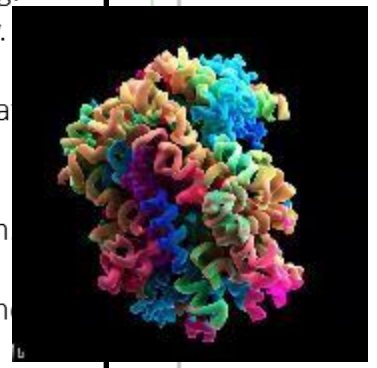
Population-scale free lunch

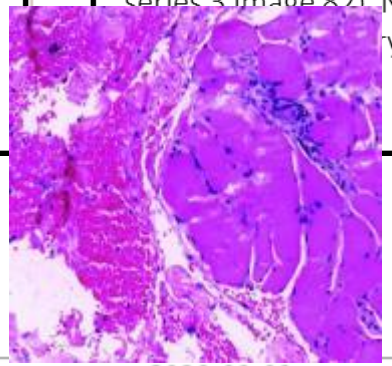
Multimodal Patient Journey

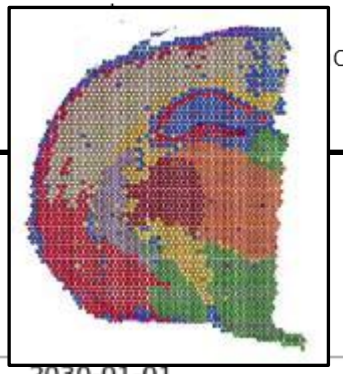


Multimodal Patient Journey









Treatment Plan
Telephone Encounter
Surgery
Progress
Procedures
Plan Of Care
Patient Instructions
Pathology
Op Note
Med Onc Order

Final Diagnosis
Lymph node, s
biopsy- Metastatic Non-small cell carcinoma. The biopsy

Lymph node:
Metastatic carcinoma: POSITIVE for PD-L1 Expression (TPS > or = to 1

FINDINGS:
Lines/Pleura: Evaluation of the left lung shows the previously noted nodules are slightly improved. For

slightly more anterior left upper lobe nodule now measures 1.0 x 0.5 cm , previously 1.2 x 1.0 cm (series 3 image 82). No new or

progress note

e reports energy is improving.
Exercises for 15 minutes a day.
new pain or headaches. No
vomiting, diarrhea, or constipa

Lungs: Clear to auscultation an
percussion
Abdomen: Bowel sounds normal,

2028-01-01
2029-01-01
2030-01-01

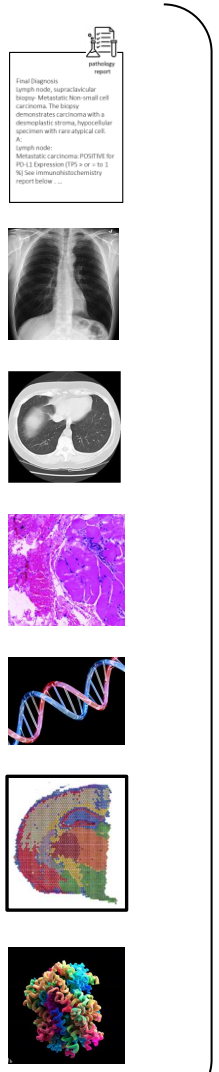
Growing array of info-rich modalities .....

Multimodal Patient Journey

https://www.adventisthealthcare.com/living-well/the-blind-men-and-the-elephant/

Precision Health Is a Multimodal Generative Problem

f



=

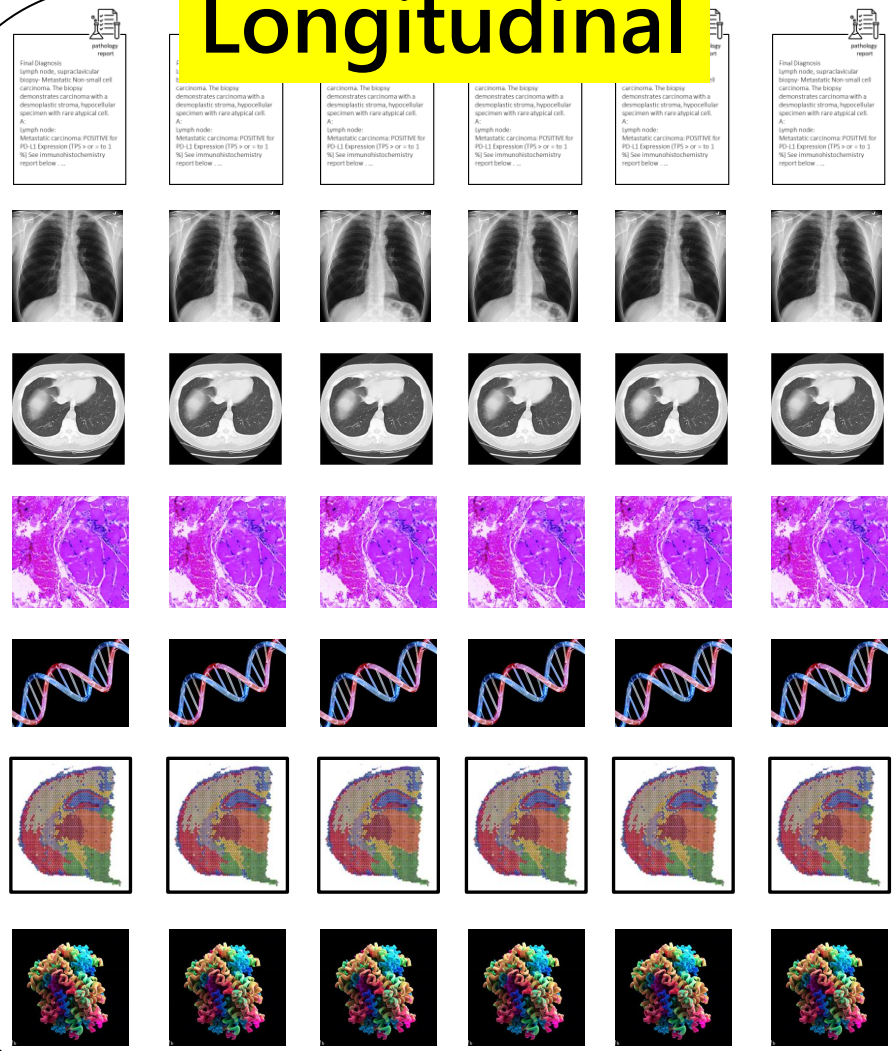
Disease Progression

Treatment Response

Precision Health Is a Multimodal Generative Problem

f

Longitudinal



=

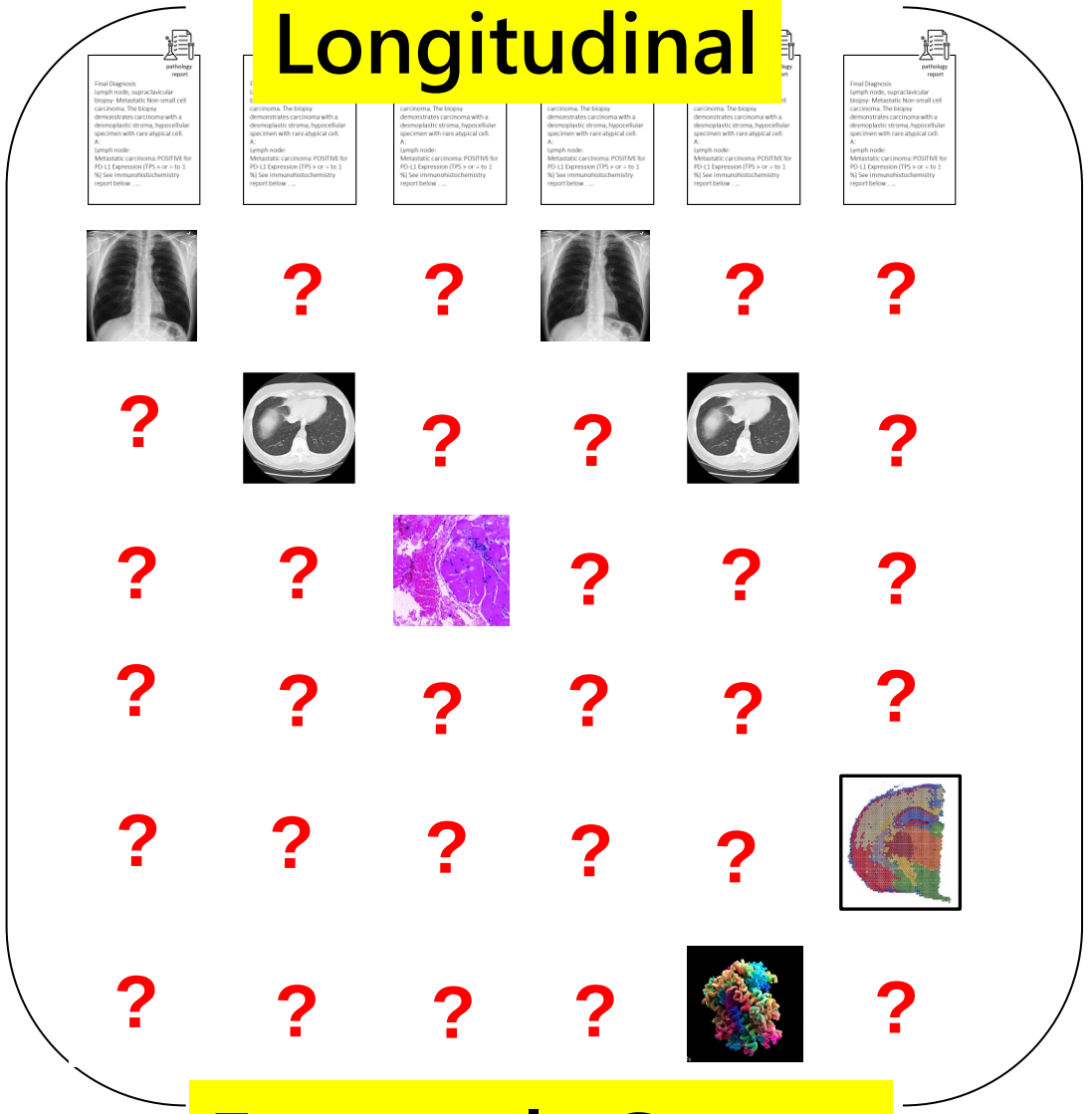
Disease Progression

Treatment Response

Precision Health Is a Multimodal Generative Problem

Longitudinal

f



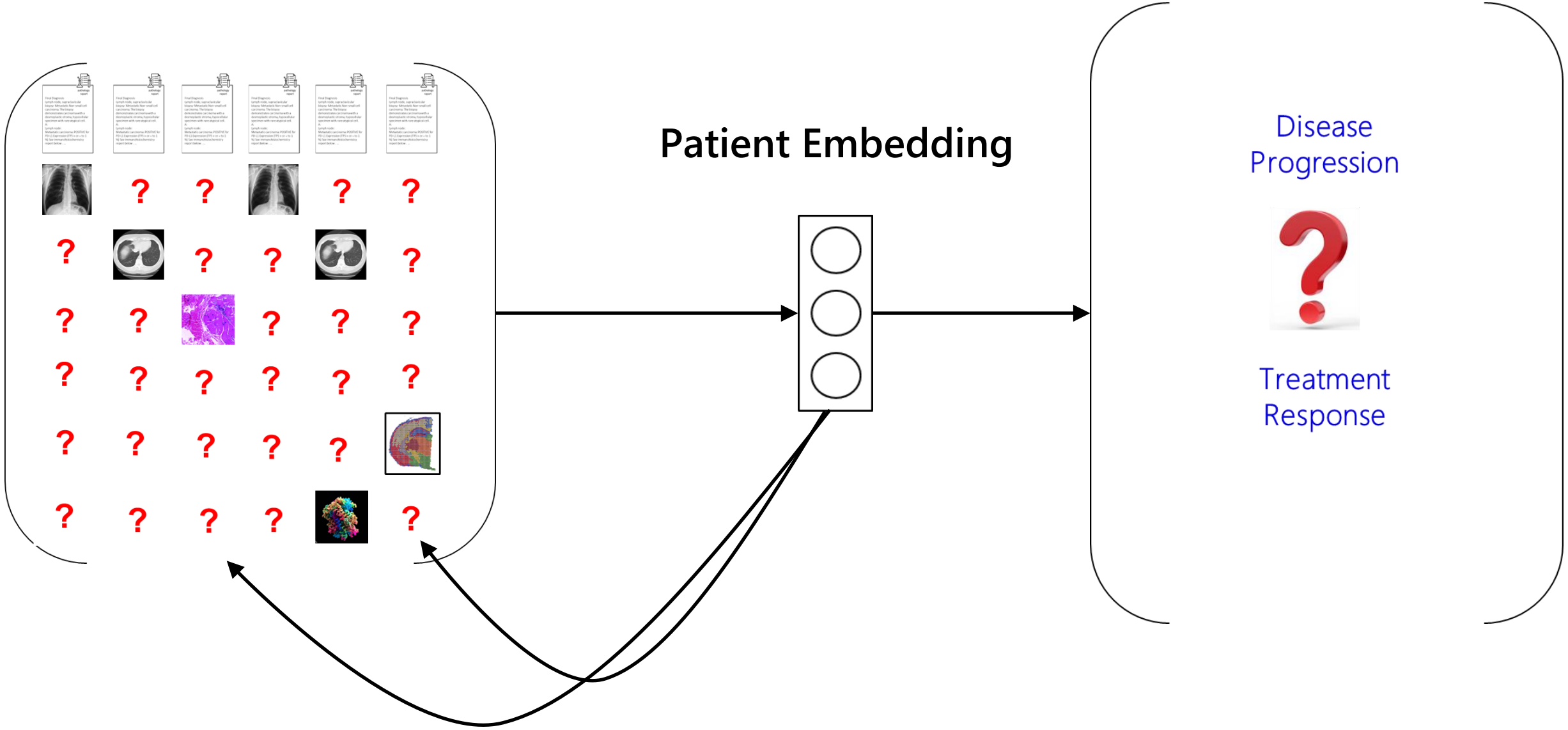
Extremely Sparse

=

Disease Progression

Treatment Response

Multimodal Patient Embedding



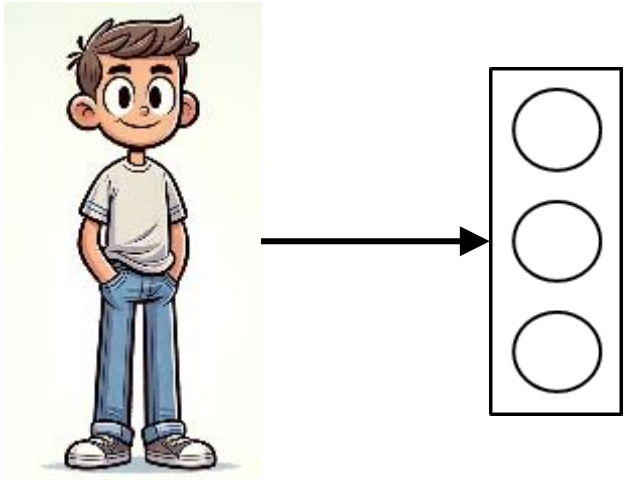
Multimodal Patient Embedding

Pretraining

Real-World Data (RWD)

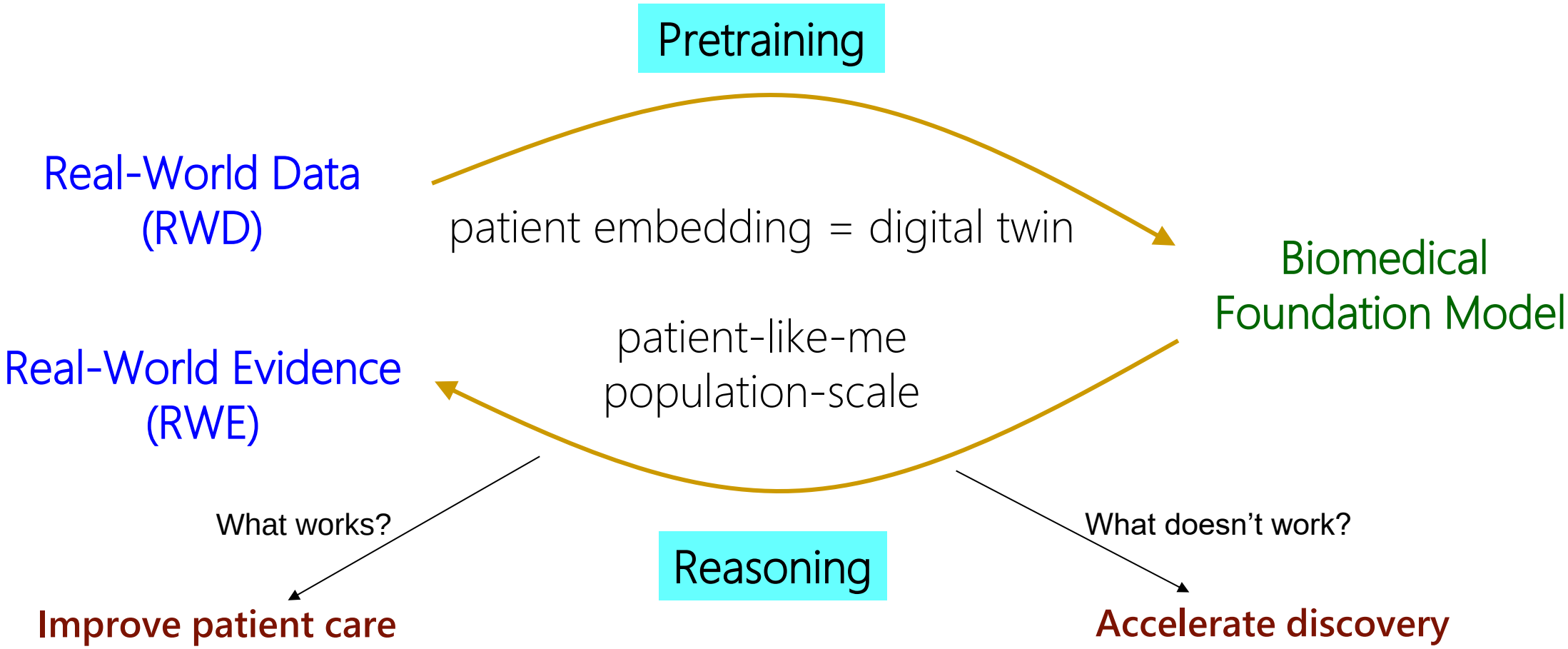
patient embedding = digital twin

Biomedical Foundation Model



GenAI disruption:
High-fidelity patient embedding

Multimodal Patient Embedding



Population-Scale RWE → "Emergent Capabilities"

The Curse of Multimodal Complexities

Structured data: discrete, ordinal, numerical, …

Clinical notes: unstructured, noisy, …

Radiology: 2D X-ray → 3D CT, MRI, …

Digital pathology: Gigapixel images (120K X 120K)

Genomics: 30-gene panel → 600-gene panel → WGS

Spatial transcriptomics: 500 genes → 4000 genes → WES

……

The Curse of Multimodal Complexities

Structured data: discrete, ordinal, numerical, …

Clinical notes: unstructured, noisy, …

Radiology: 2D X-ray → 3D CT, MRI, …

Digital pathology: Gigapixel images (120K X 120K)

Genomics: 30-gene panel → 600-gene panel → WGS

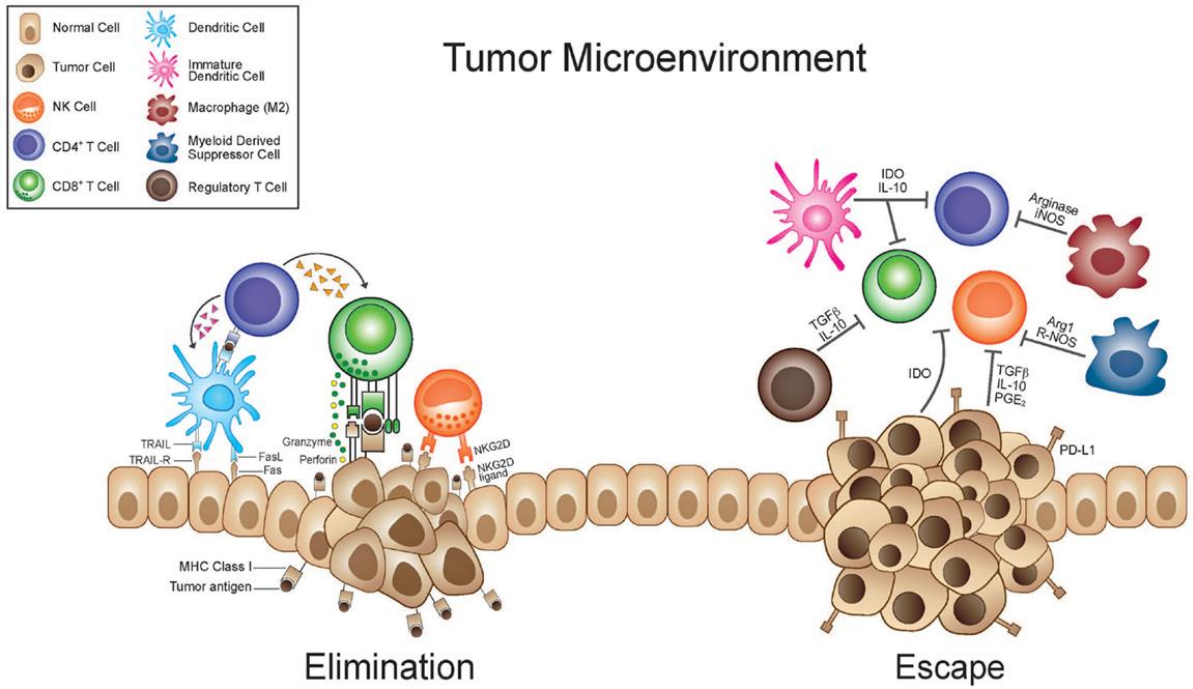
Spatial transcriptomics: 500 genes → 4000 genes → WES

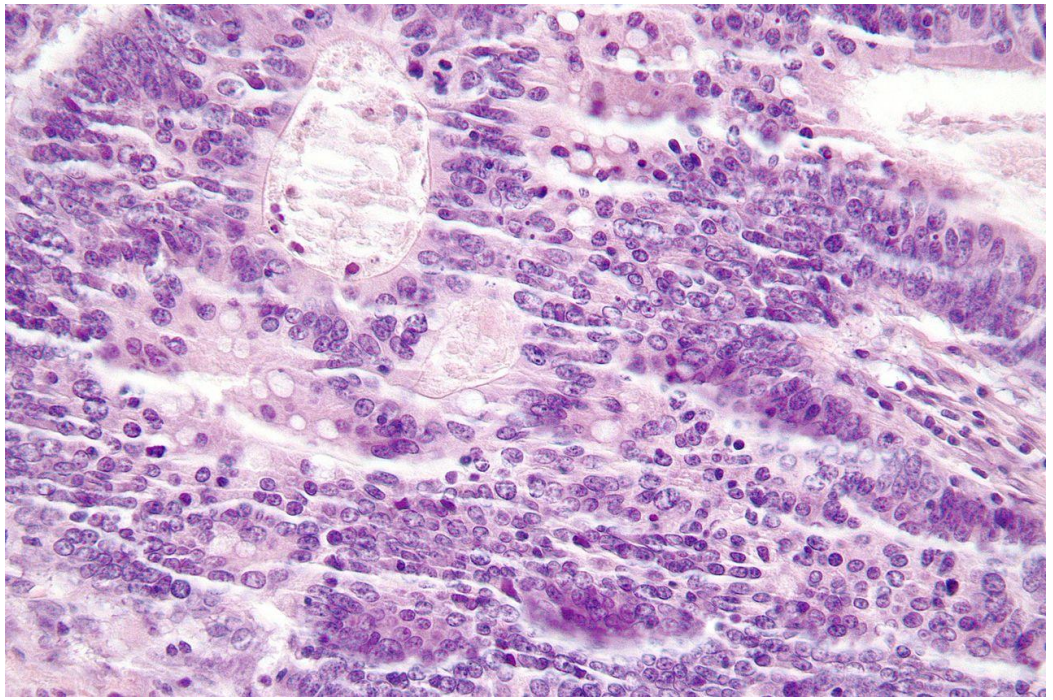
……

Case Study: Immunotherapy

Given Keytruda cohort, find exceptional responder

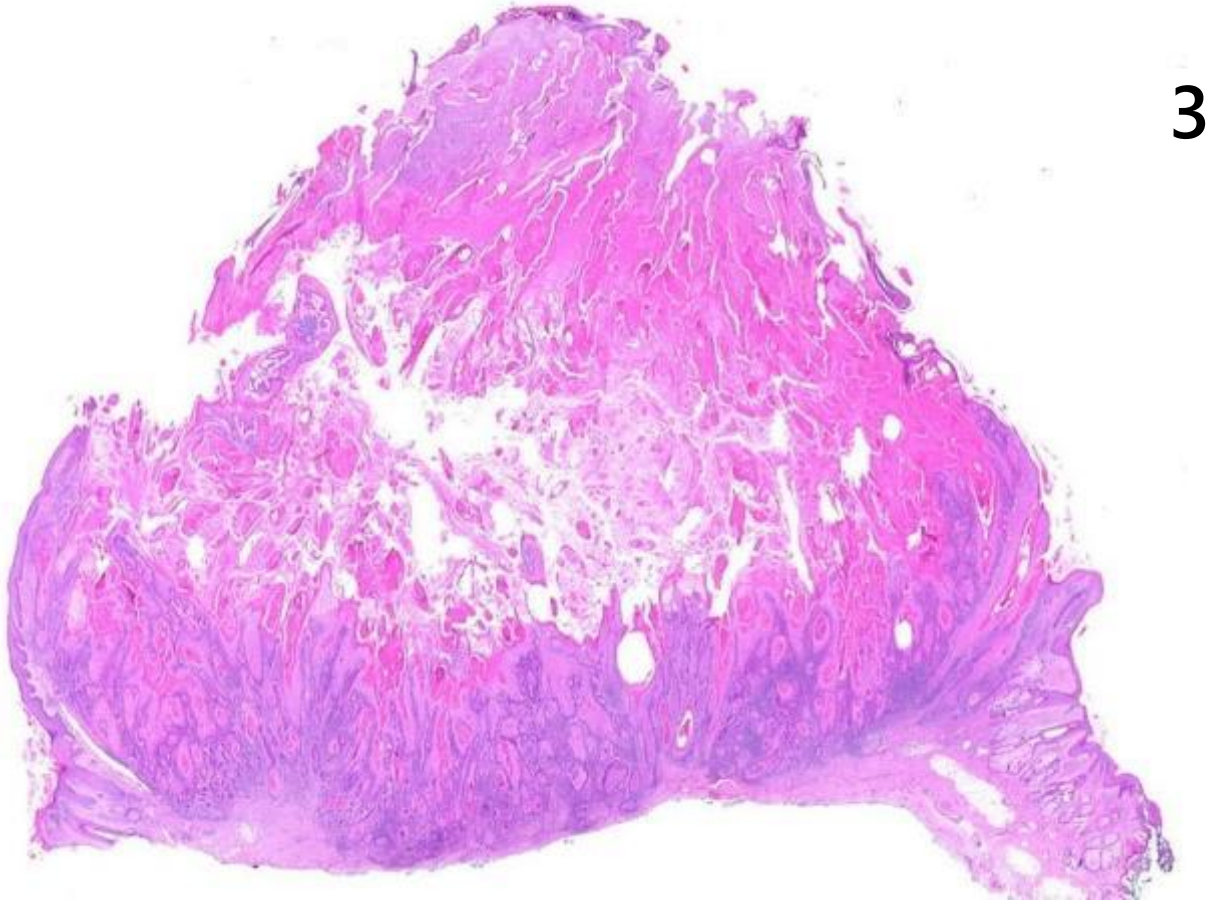
Need to model tumor microenvironment





https://en.wikipedia.org/wiki/Tumor-infiltrating_lymphocytes

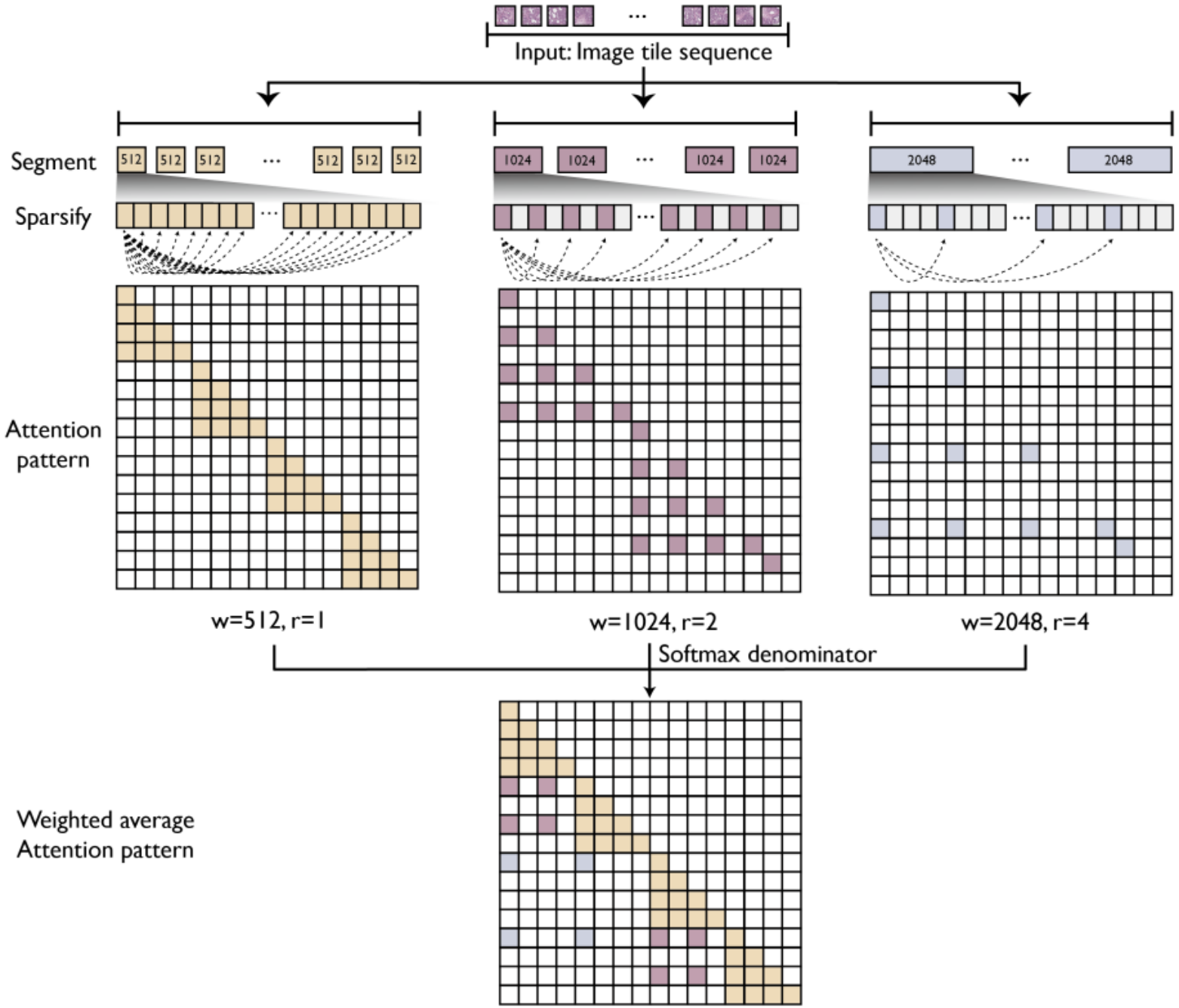
Digital Pathology: Transformer blows up…



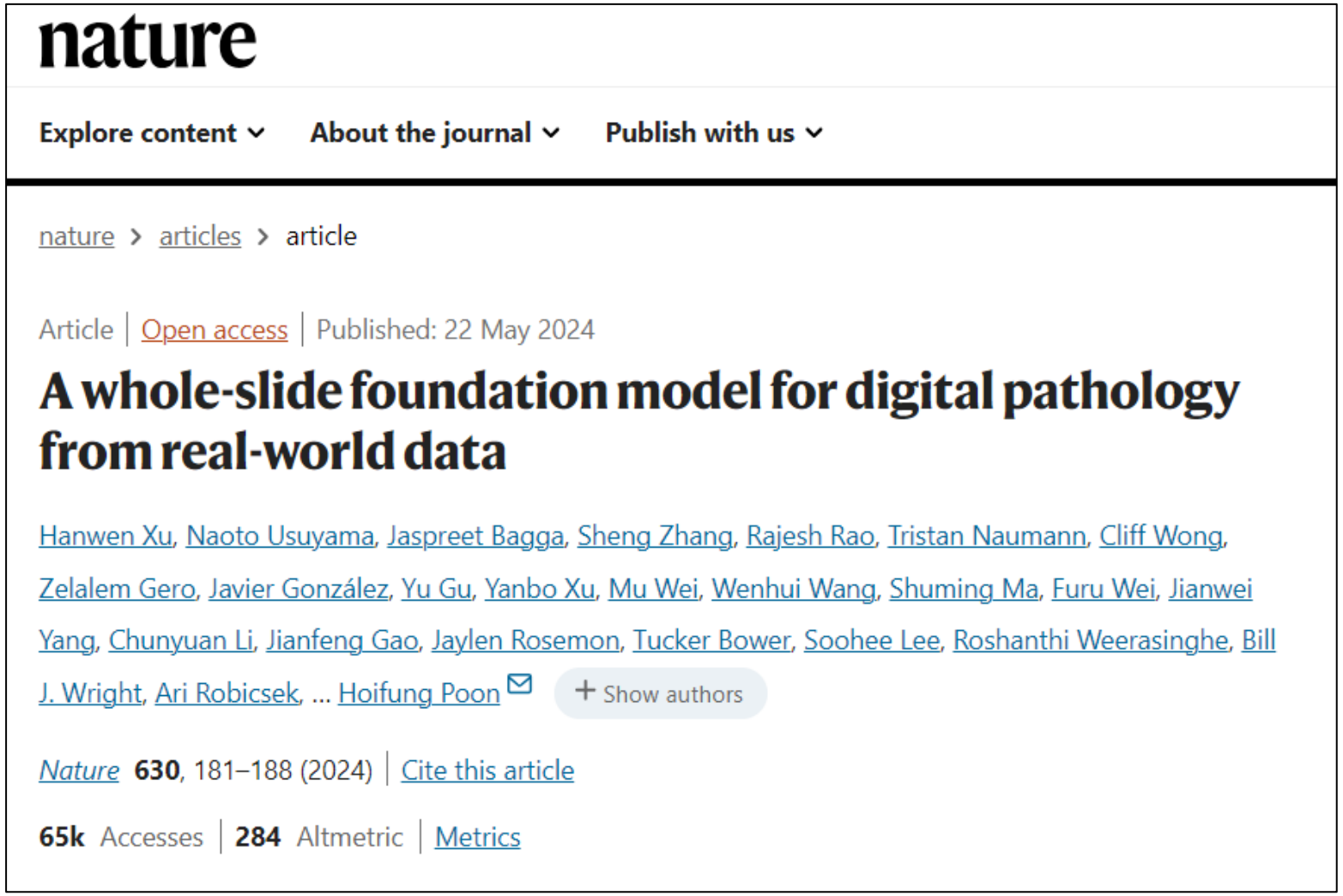
3cm x 3cm, 40x → 120K x 120K pixels

16 x 16 patch: 56 million tokens

Compute: Billions of times more than web images



Dilated attention



nature

Explore content ∨ About the journal ∨ Publish with us ∨

nature › articles › article

Article | Open access | Published: 22 May 2024

A whole-slide foundation model for digital pathology from real-world data

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, Yanbo Xu, Mu Wei, Wenhui Wang, Shuming Ma, Furu Wei, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Jaylen Rosemon, Tucker Bower, Soohee Lee, Roshanthi Weerasinghe, Bill J. Wright, Ari Robicsek, ... Hoifung Poon + Show authors

Nature **630**, 181–188 (2024) | Cite this article

65k Accesses | **284** Altmetric | Metrics

Microsoft

Providence

UNIVERSITY OF WASHINGTON 1861 LVX SIT

GigaPath: First Whole-Slide Digital Pathology Foundation Model
170K whole slides (1.3B image tiles)

https://aka.ms/gigapath

GigaPath: Monthly Downloads > 90K

Hugging Face | Search models, datasets, users... | Models | Datasets | Spaces | Posts | Docs | Solutions | Pricing

prov-gigapath / **prov-gigapath** like 85

Image Feature Extraction | timm | PyTorch | vision | medical | License: prov-gigapath-license (other)

Model card | Files and versions | Community 11 | Settings | Use this model

Edit model card

Gated model You have been granted access to this model

Prov-GigaPath

A whole-slide foundation model for digital pathology from real-world data

[Code] [Model] [Paper] [BibTeX]

Hanwen Xu*, Naoto Usuyama*, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, Yanbo Xu, Mu Wei, Wenhui Wang, Shuming Ma, Furu Wei, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Jaylen Rosemon, Tucker Bower, Soohee Lee, Roshanthi Weerasinghe, Bill J. Wright, Ari Robicsek, Brian Piening, Carlo Bifulco, Sheng Wang, Hoifung Poon (*Equal Contribution)

Downloads last month
90,166

Inference API

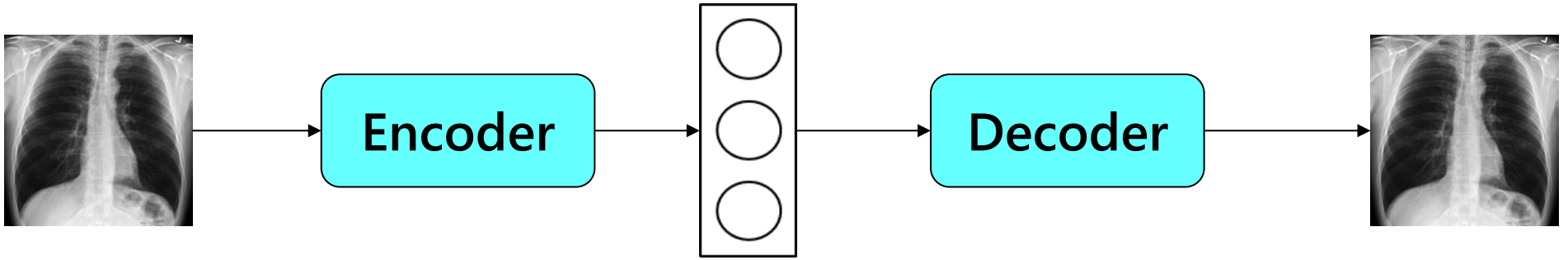
Image Feature Extraction

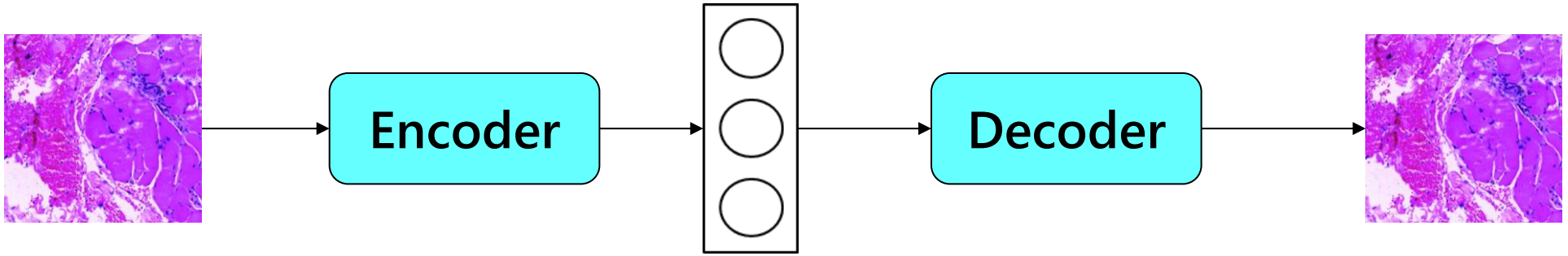
Inference API (serverless) does not yet support timm models for this pipeline type.


NORTH
AMERICA
SOUTH
AMERICA
Atlantic
Ocean
Pacific
Ocean
AFRICA
ASIA
Indian
Ocean
Southern
Ocean

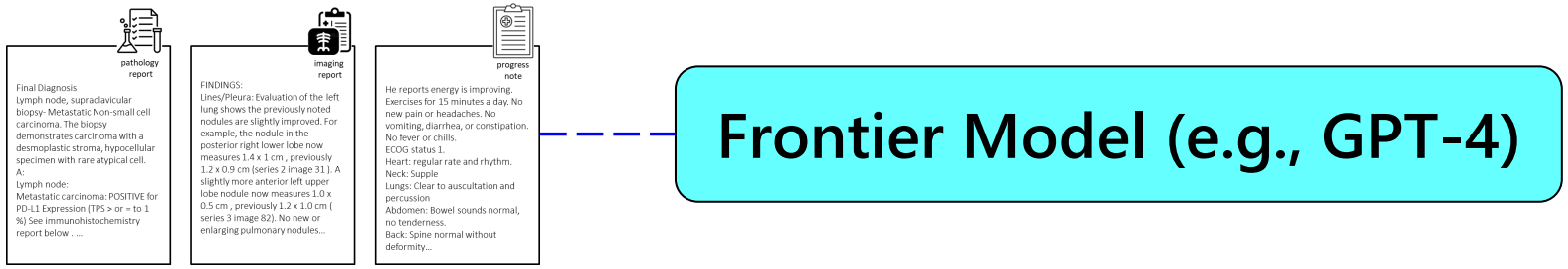
Unimodal: Encoder / Decoder

Modality-Specific Self-Supervision

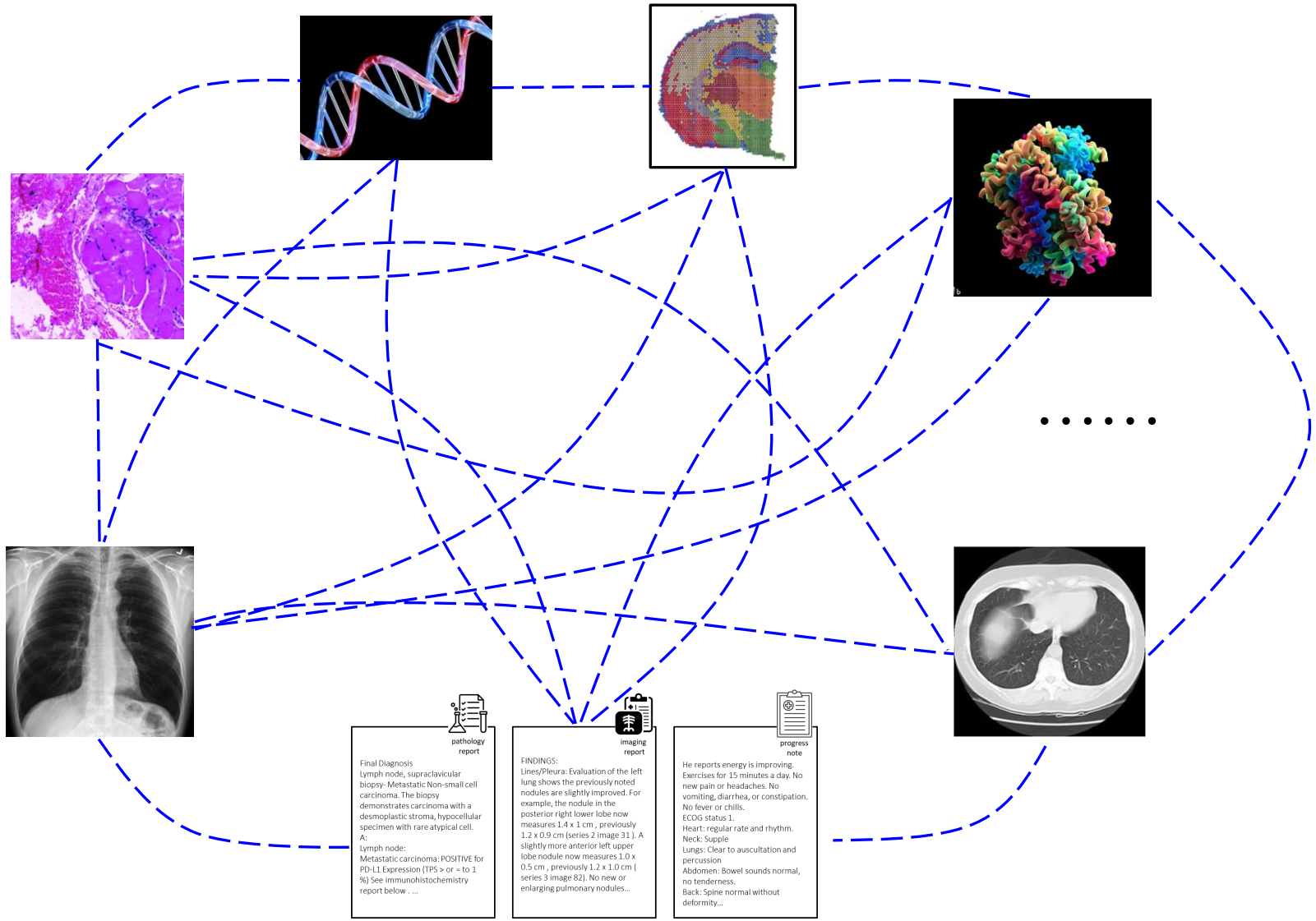




. . . .



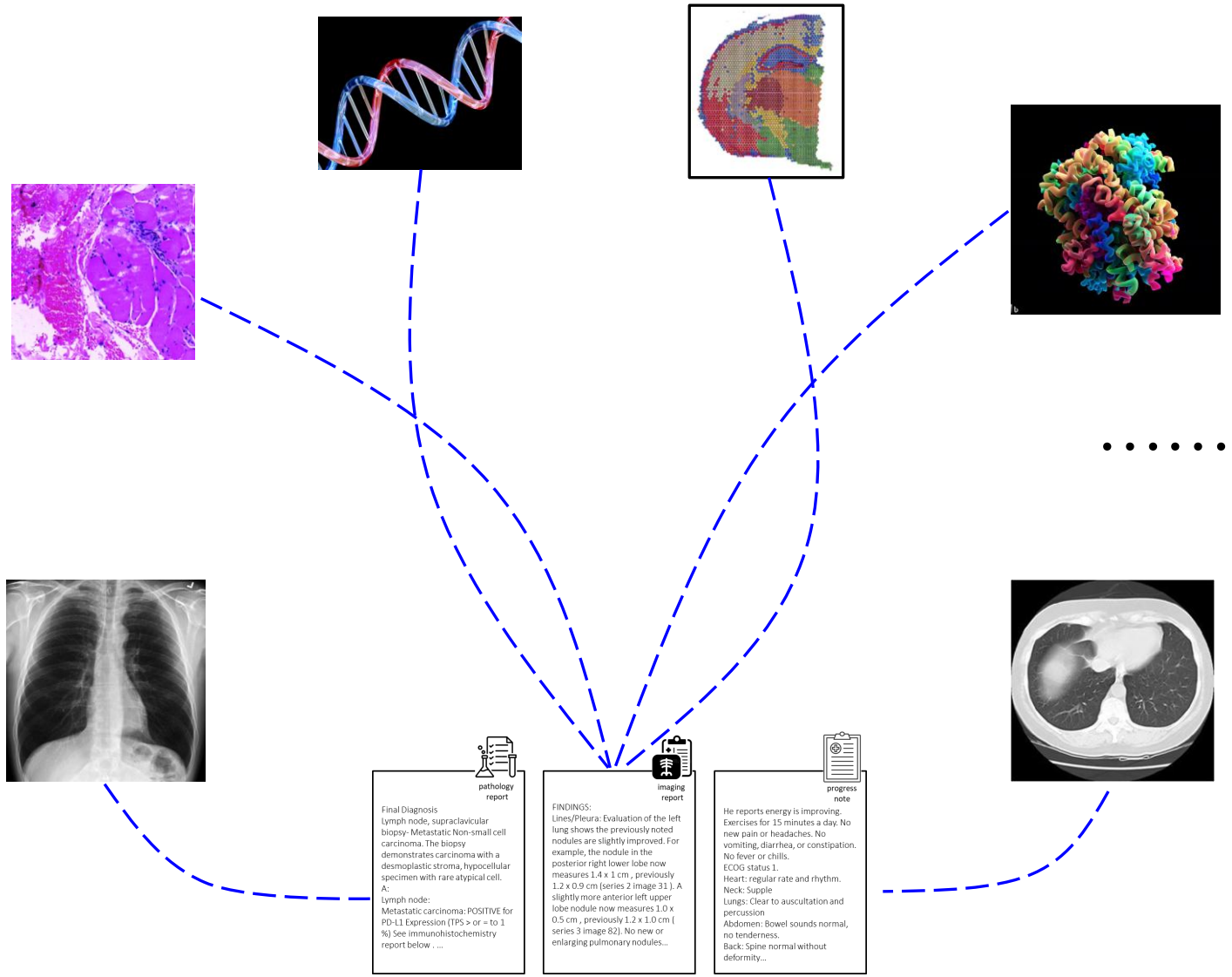
The Curse of Multimodal Complexities



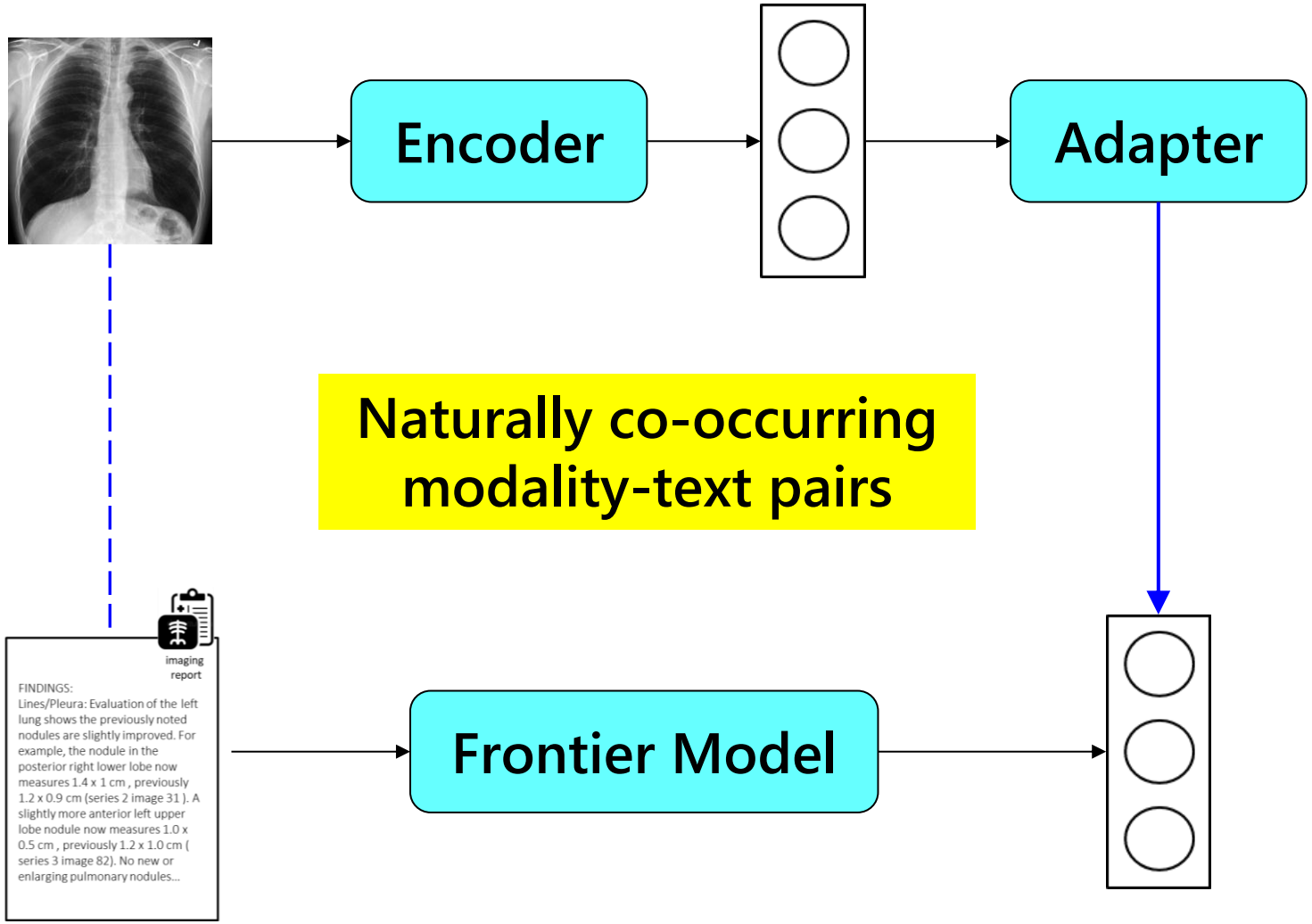
Combinatorial Explosion

Multimodal Babel Tower

Text = Interlingua



Cross-Modal: Adapters



LLaVA-Med: General Recipe

Li*, Wong*, Zhang*, et al. LLaVA-Med: Training a Large Language-and-Vision Assistant in One Day. *NeurIPS 2023, Spotlight.*

Image Encoder

Adapter

Text Decoder

<Answer Text>

Instruction

E.g.:
"Describe this image."
"Where is the lesion located?"
"Is it extended into the chest wall?"

Instruction-following data

GPT-4 → Multimodal Synthetic Data

BiomedJourney: In Silico Imaging

Gu*, Yang*, et al. BiomedJourney: Counterfactual Biomedical Image Generation by Instruction-Learning from Multimodal Patient Journeys, *arxiv 2310.10765*.

Image Encoder

Adapter

Image Decoder

Text Encoder

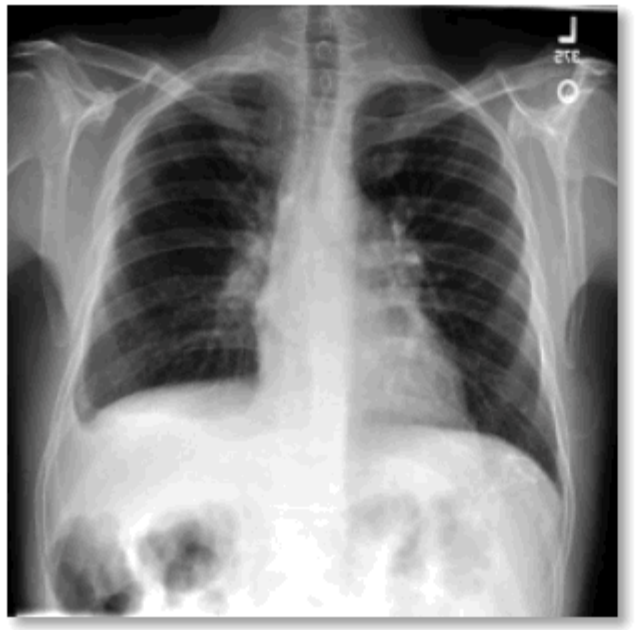
Progression

E.g.:
"Increasing right pleural effusion and prominence of cardiac silhouette"

Instruction-following data

GPT-4 → Multimodal Synthetic Data

Towards Generative Diagnostics



Patient Image

BiomedParse: Image Parsing

Zhao*, Yu*, et al. BiomedParse: a biomedical foundation model for image parsing of everything everywhere all at once, *arxiv 2405.12971*.

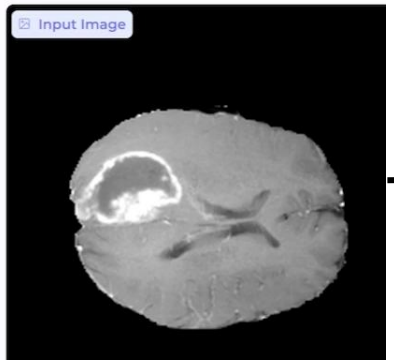
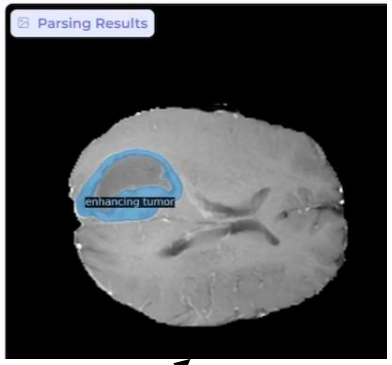


Image Encoder

Adapter

Mask Decoder



Text Encoder

Text Prompt

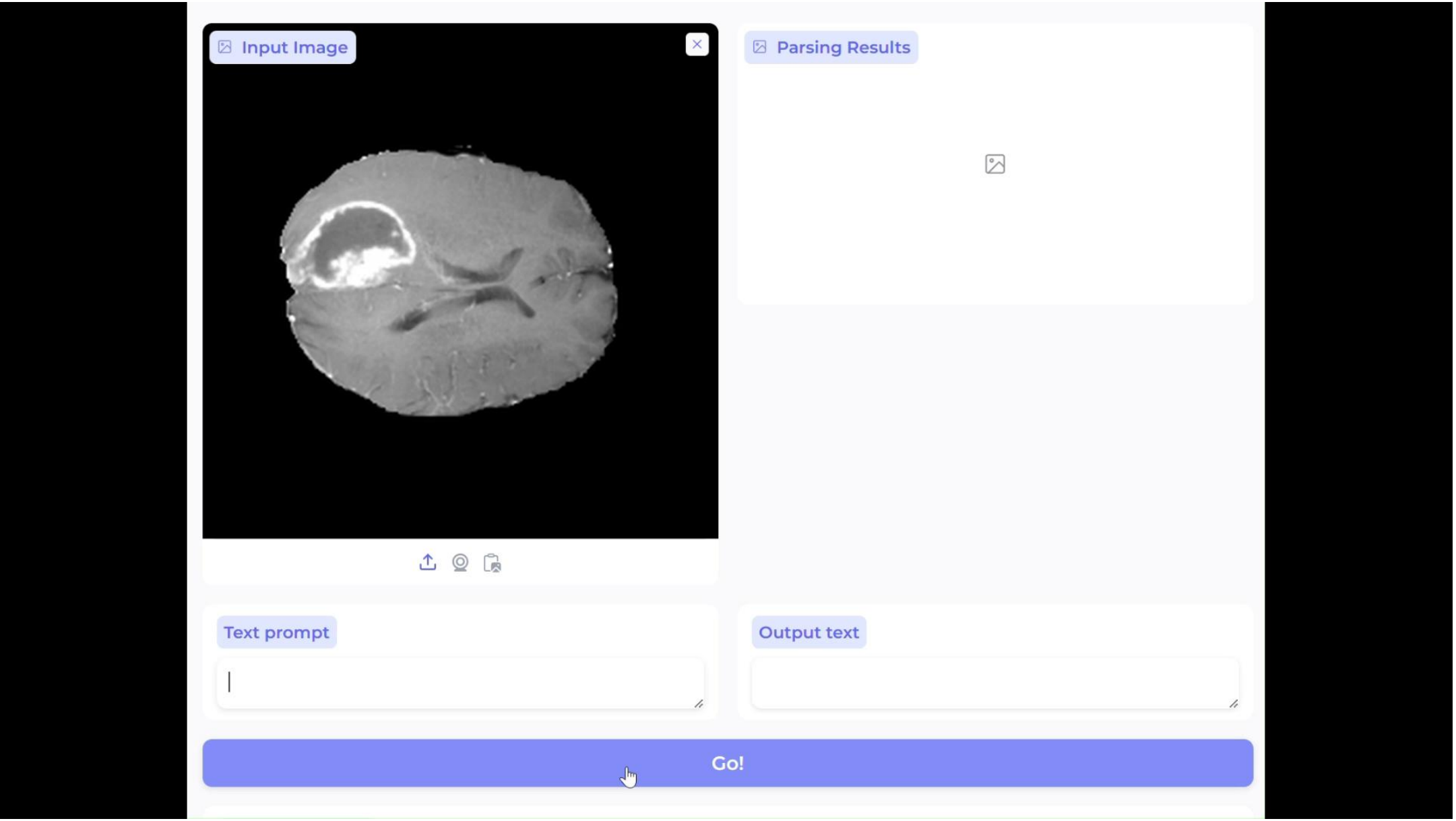
E.g.:
"Are there any immune cells?"
"Lung nodule in this chest CT"
"Find me all organs"

Instruction-following data

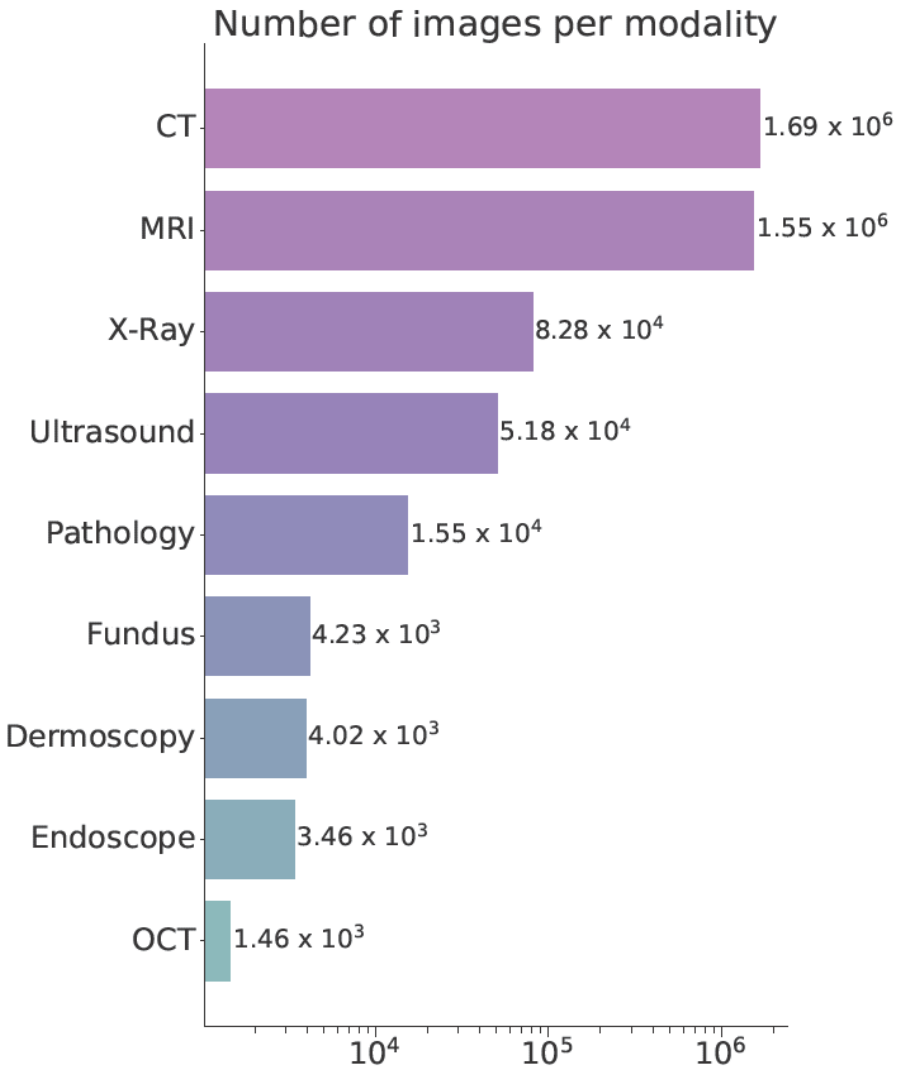
GPT-4 → Multimodal Synthetic Data

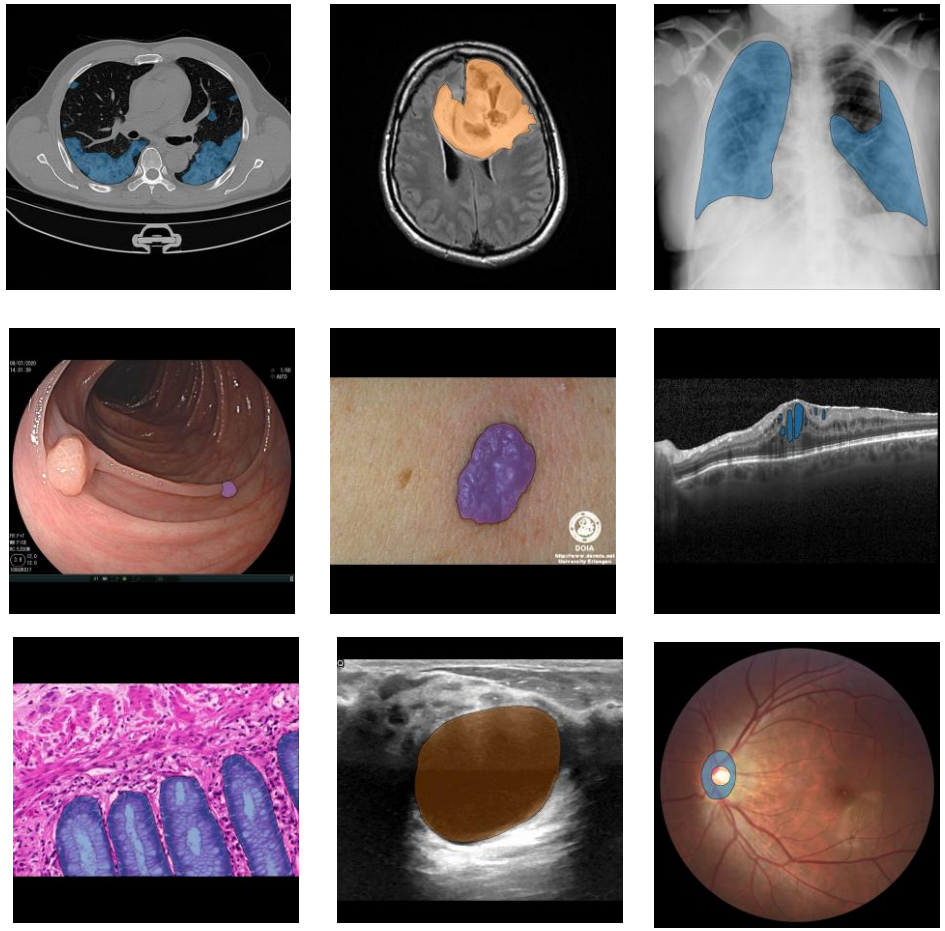
Image Parsing: Segmentation, Detection, Recognition

https://microsoft.github.io/BiomedParse/



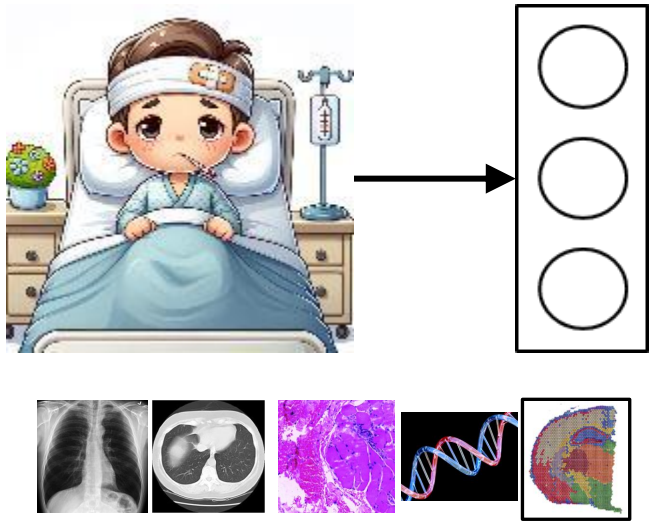
Foundation Model: Nine Modalities

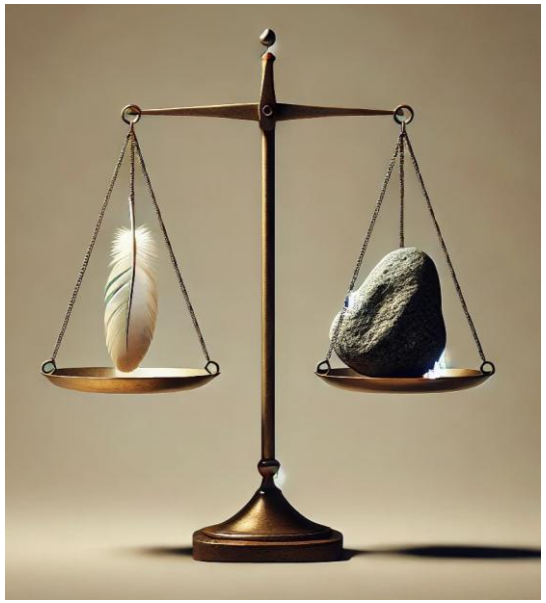




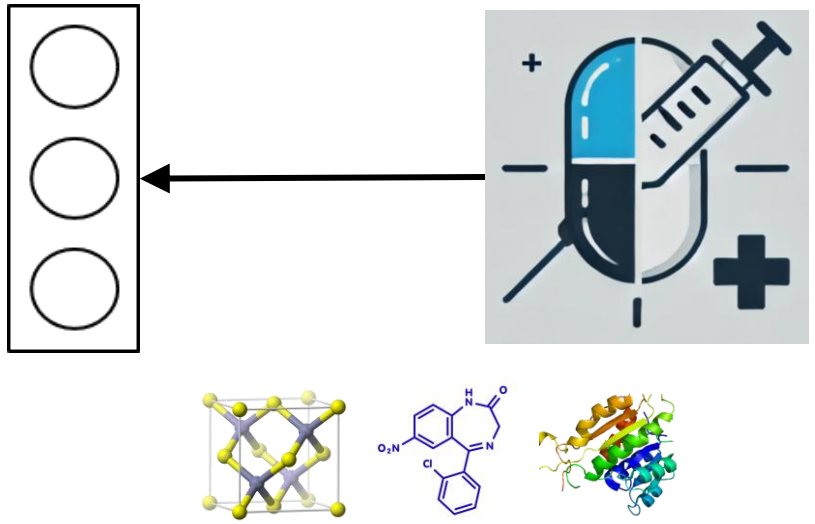
Universal Embedding

Patient





Intervention



Optimize clinical care

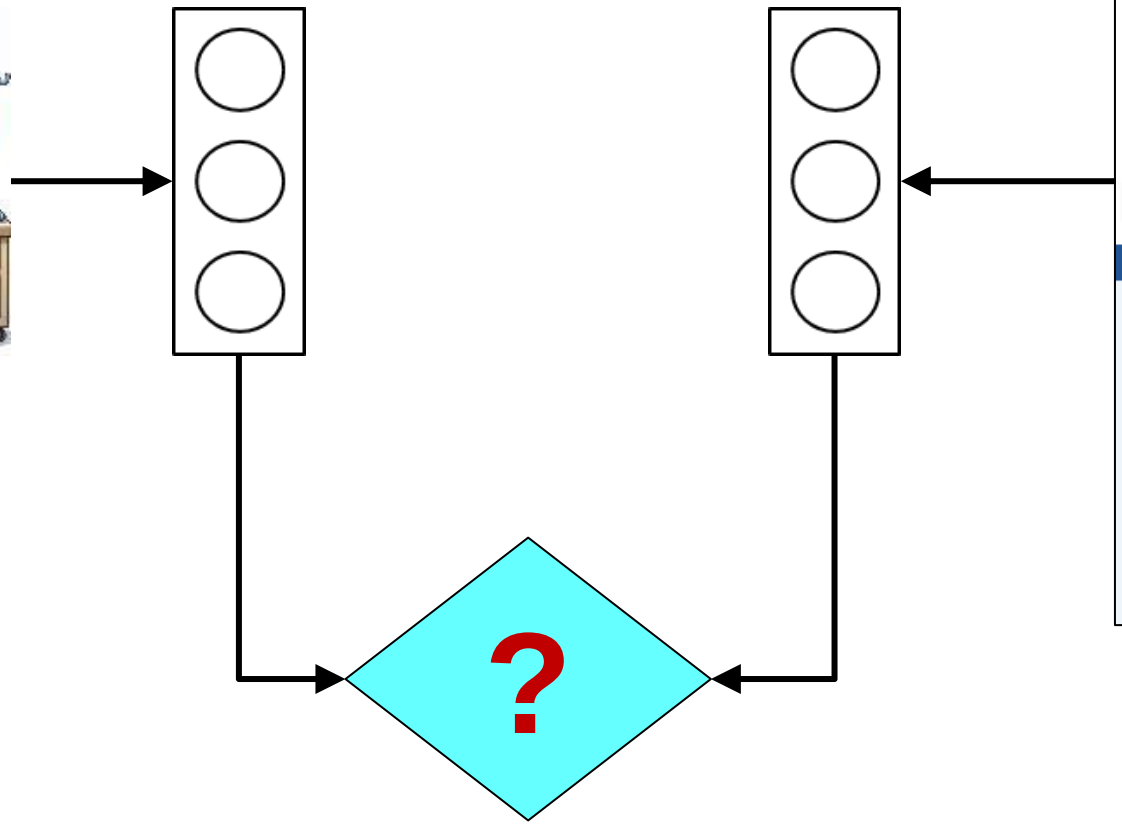
Accelerate biomed discovery

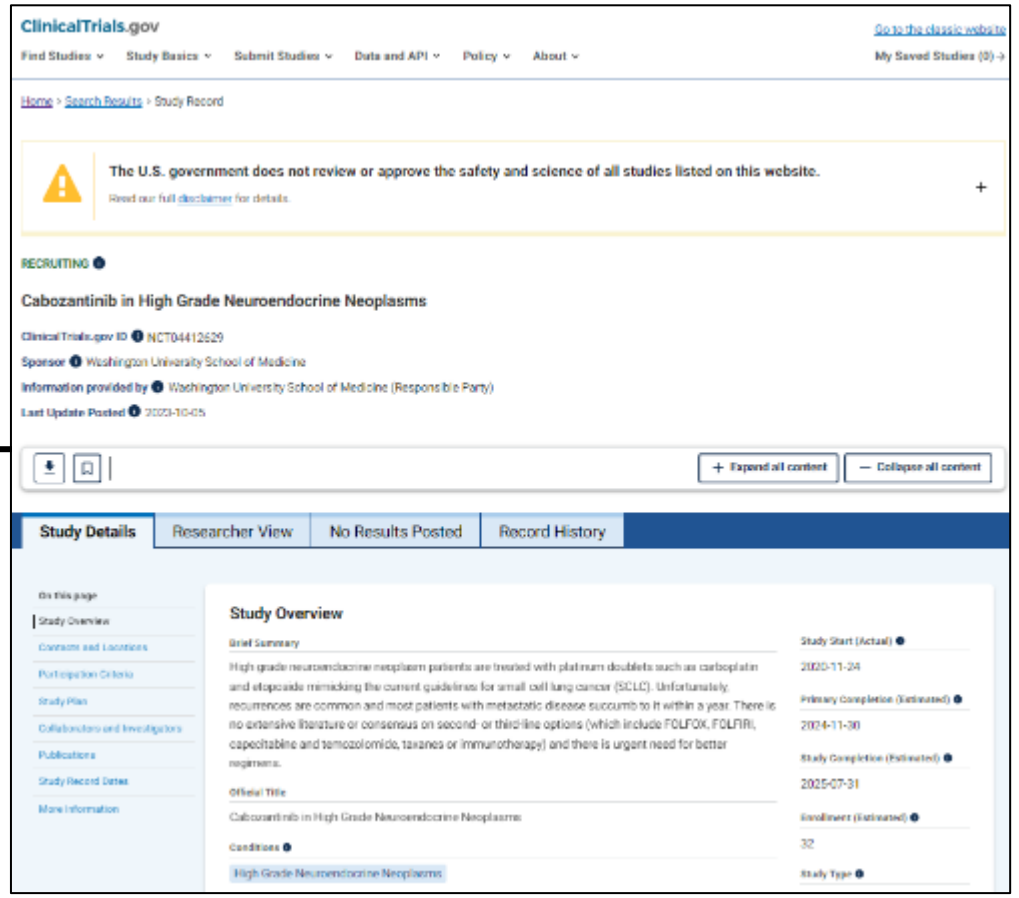
Information Access Can Be Life or Death

Marty Tenenbaum
Late-stage melanoma (late 1990s)
Initial prognosis: 6 months
Saved by Phase III trial of Canvaxin

US: Less than 3% cancer patients enroll in trials
40% cancer trial failures due to insufficient patients
New drug costs $2-10 billion and takes 10+ years

US alone, 2M new cancer patients every year





400K clinical trials at ct.gov
Thousands of active cancer trials

Just-in-time clinical trial matching

ClinicalTrials.gov BETA

Resources About

Home > Search Results > Study Record

ClinicalTrials.gov Identifier: NCT04520711

RECRUITING

Hotspot TCR-T: A Phase I/Ib Study of Adoptively Transferred T-cell Receptor Gene-engineered T Cells (TCR-T)

Information provided by Providence Health & Services (Responsible Party)

Last Updated: May 6, 2022

Dr. Rom Leidner

Providence

Pilot: 3 test sites, 30 clinicians, 5,000 patients/year

Next: 51 hospitals, 950 clinics, 500 oncologists, 50,000 patients/year

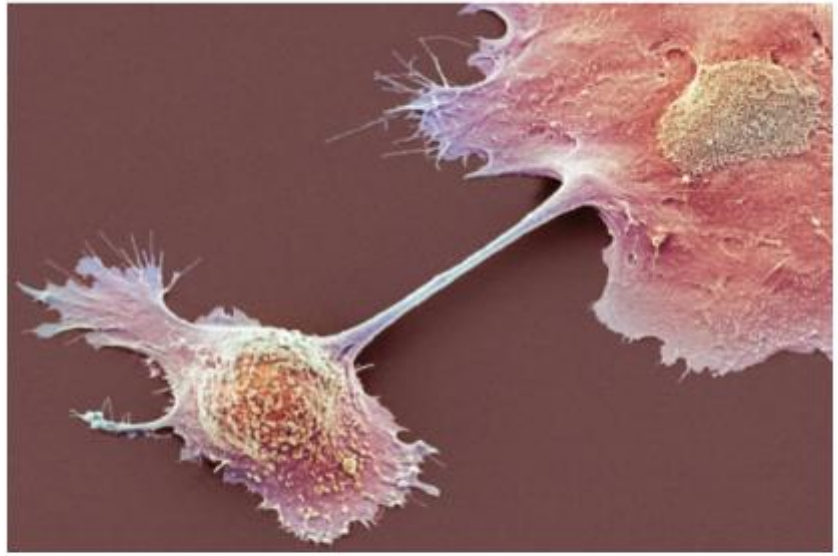
Wong et al. "Scaling Clinical Trial Matching Using Large Language Model: A Case Study in Oncology", *MLHC, 2023*.

The New York Times

Reprogrammed Cells Attack and Tame Deadly Cancer in One Woman

Another patient who had the same treatment did not survive. But the demonstration of the technique could help with other cancers.

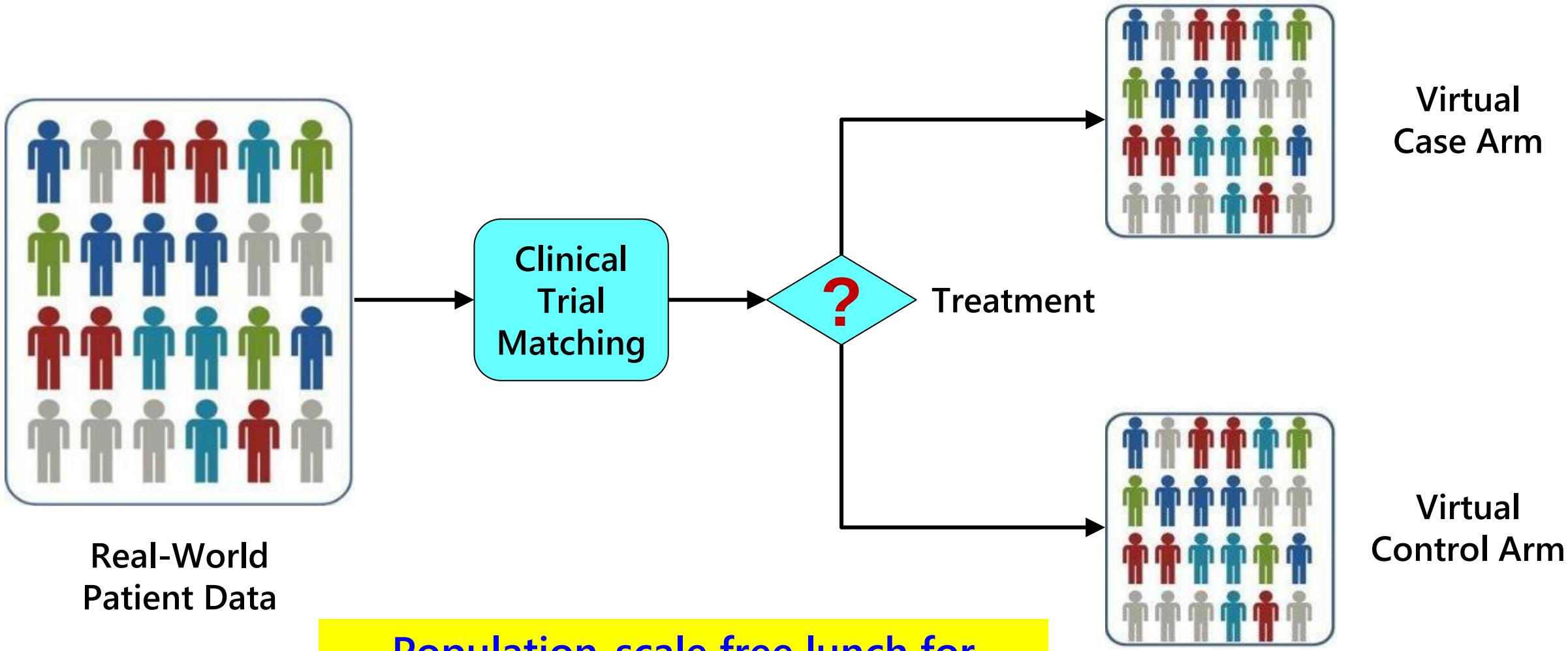
Give this article



A colored scanning electron micrograph of pancreatic cancer cells. Steve Gschmeissner/Science Source

By Gina Kolata

Real-World Evidence: In Silico Clinical Trial Simulation



Population-scale free lunch for hypothesis generation and testing

TrialScope: In Silico Clinical Trial Simulation

Trial	RCT		Simulation HR				HR
	HR	95%CI	HR	95%CI	C	T	match?
FLAURA	0.63	(0.45, 0.88)	0.57	(0.43, 0.77)	255	169	✓
			0.76	(0.61, 0.95)	458	347	
CHECKMATE057	0.73	(0.59, 0.89)	0.63	(0.46, 0.86)	109	136	✓
			0.77	(0.64, 0.93)	304	413	
CHECKMATE078	0.68	(0.52, 0.9)	0.79	(0.60, 1.03)	140	198	✓
			0.79	(0.65, 0.97)	305	415	
KEYNOTE010	0.71	(0.58, 0.88)	0.70	(0.56, 0.87)	187	539	✓
			0.74	(0.62, 0.88)	332	1044	
OAK	0.73	(0.62, 0.87)	0.63	(0.33, 1.19)	129	33	✓
			0.47	(0.32, 0.69)	345	88	
KEYNOTE024	0.63	(0.47, 0.86)	0.68	(0.5, 0.93)	104	524	✓
			0.79	(0.61, 1.02)	250	1178	
STELLA	1.108	(0.27, 1.48)	1.10	(0.75, 1.61)	1507	54	✓
			1.31	(1.00, 1.73)	4004	100	
NCT00130728	0.78	(0.79, 1.17)	0.87	(0.67, 1.14)	264	91	✓
			1.12	(0.91, 1.37)	517	173	
CHECKMATE017	0.59	(0.44, 0.79)	0.87	(0.53, 1.43)	36	66	×
			0.77	(0.64, 0.93)	304	413	
EMPHASIS	?	?	0.76	(0.59, 0.97)	192	322	?
			0.77	(0.62, 0.94)	330	485	
NCT02604342	?	?	0.42	(0.23, 0.78)	1001	29	?
			0.52	(0.34, 0.82)	1742	53	

Gonzalez et al. "TrialScope: A Unifying Causal Framework for Scaling Real-World Evidence Generation with Biomedical Language Models", *arxiv 2311.01301*.

Providence

Team


Hoifung Poon


Tristan Naumann


Cliff Wong


Naoto Usuyama


Sheng Zhang


Zelalem Gero


Jass Bagga

Collaborators

Providence: Carlo Bifulco, Brian Piening, Roshanthi Weerasinghe, Rom Leidner

JnJ: Xiaoying Wu, Tommaso Mansi

U. Wash: Sheng Wang

USC: Muhao Chen

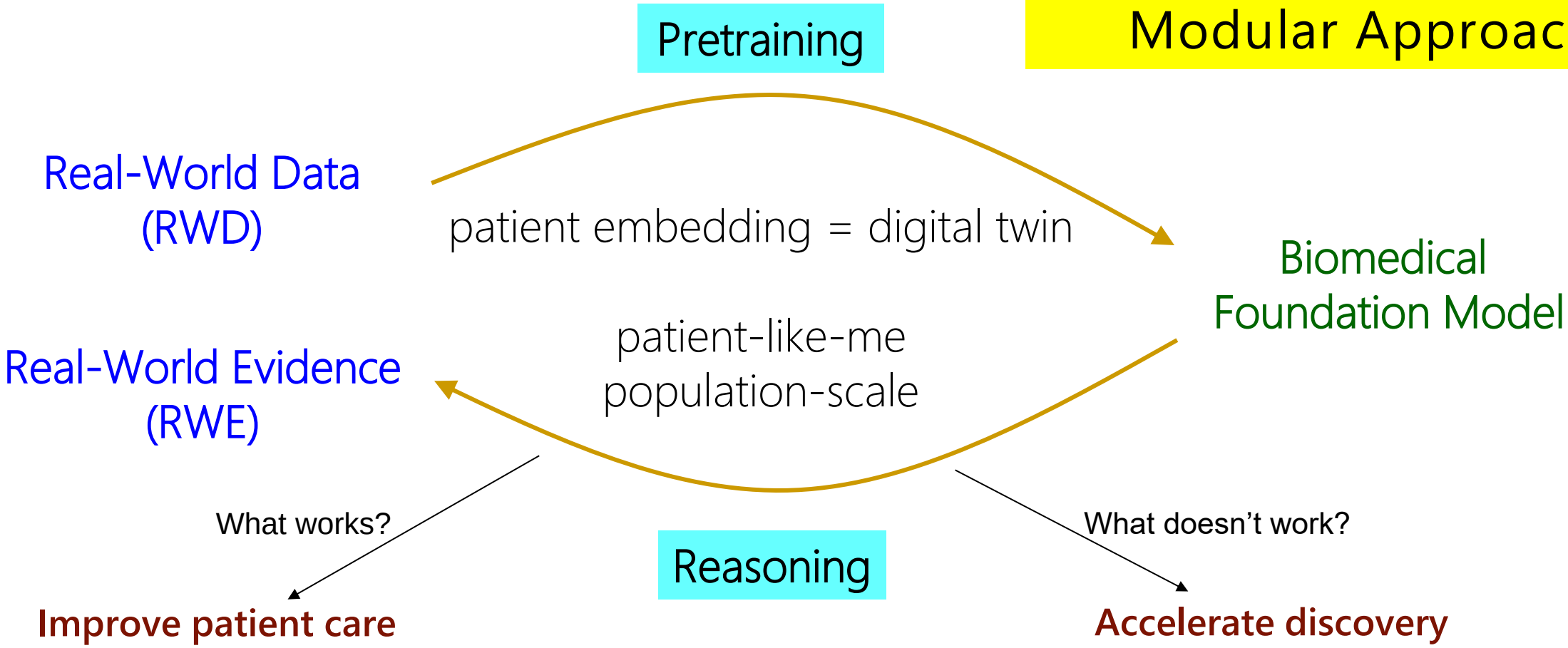
Stanford: Akshay Chaudhari

Microsoft: Jianfeng Gao, Javier Alvarez-Valle, Javier Gonzalez, Tao Qin, Tie-Yan Liu, Furu Wei, Mu Wei, Sam Preston, Emre Kiciman, Naveen Valluri, Paul Viozila, Matt Lungren, Harjinder Sandhu

Interns: Maxim Grechkin, Ankur Parikh, Victoria Lin, Sheng Wang, Stephen Mayhew, Daniel Fried, Violet Peng, Hai Wang, Robin Jia, Matthew McDermott, Alexis Ross, Zelalem Gero, Sarthak Jain, Jenny Chen, Hunter Lang, Benedikt Boecking, Varsha Kishore, Jinfeng Xiao, Michelle Li, Wenxuan Zhou, Neha Hulkund, Risa Ueno, Peniel Argaw, Hanwen Xu, Juan Manuel Zambrano Chaves, Mars Huang, Yiqing Xie, Isabel Chien, Eduard Orakvin, Alicia Curth

Multimodal Patient Embedding

Multimodal Complexities: Modular Approach



Population-Scale RWE → "Emergent Capabilities"