CPH 100c / Data c146:
Foundations for Computational Precision Health

Instructors:

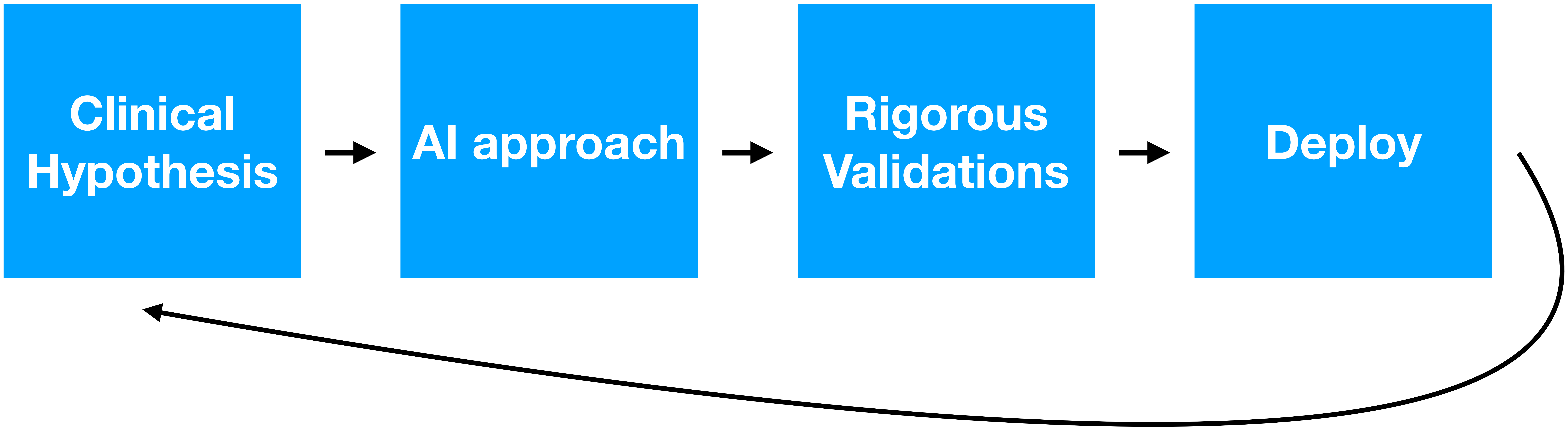
Adam Yala, PhD (yala@berkeley.edu)

Irene Chen, PhD (iychen@berkeley.edu)

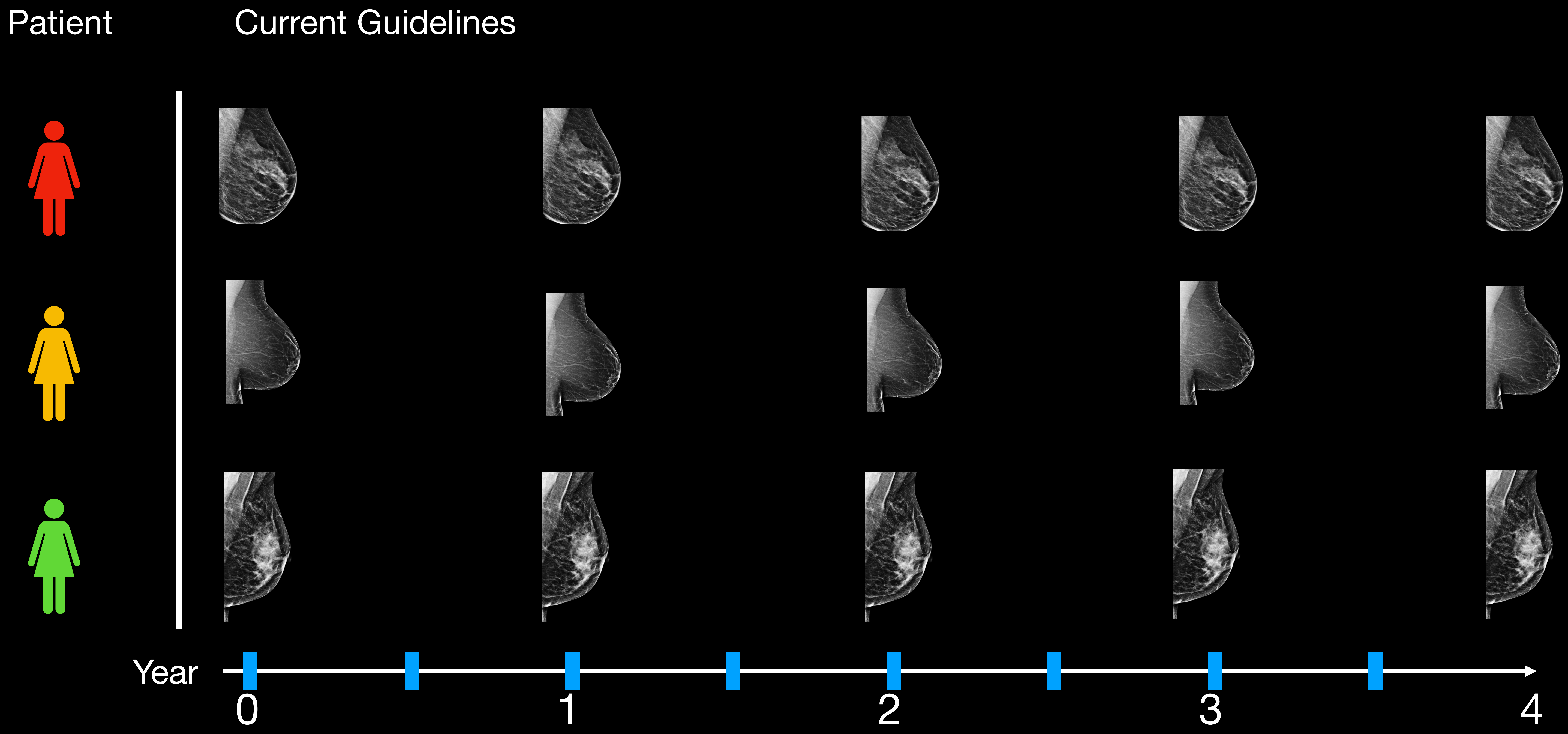
Welcome to CPH!

- What is Computational Precision Health?
  - *Applying computation to real-world settings to improve the quality, efficiency, and equity of medicine and public health*
- Spans many computational and medical domains
- Key unifying features:
  - *Precision* in problem formulation, solutions and deployment

A practical instantiation: *My own research journey*

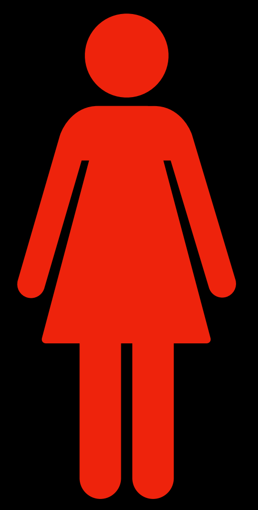


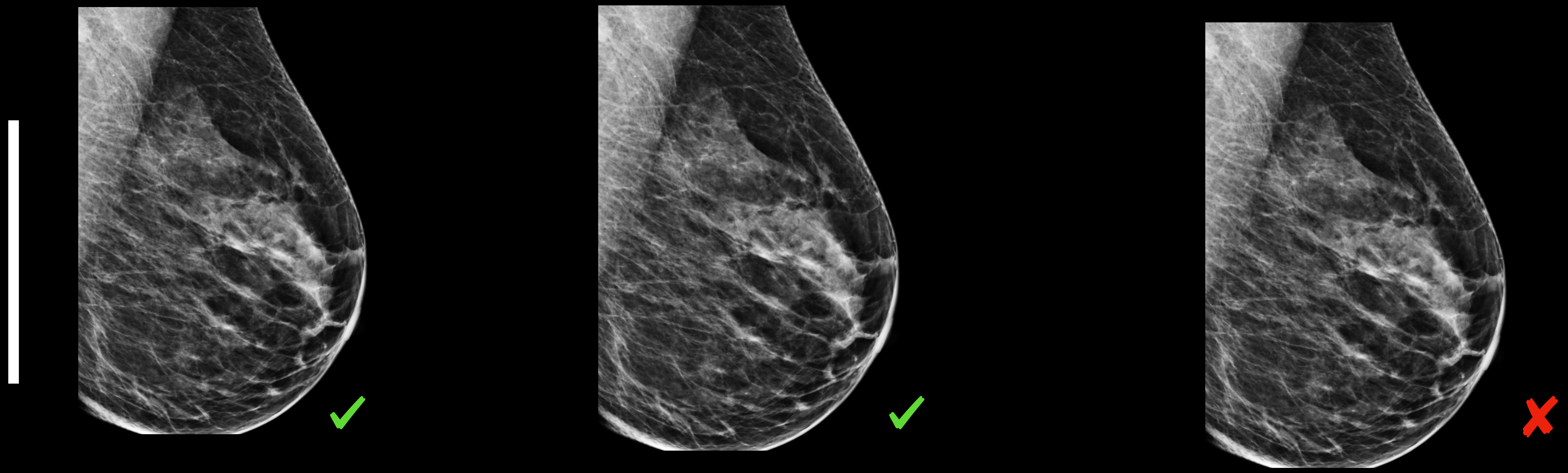
Motivating example: Screening today - one size fits all



The harms of late diagnosis

Patient



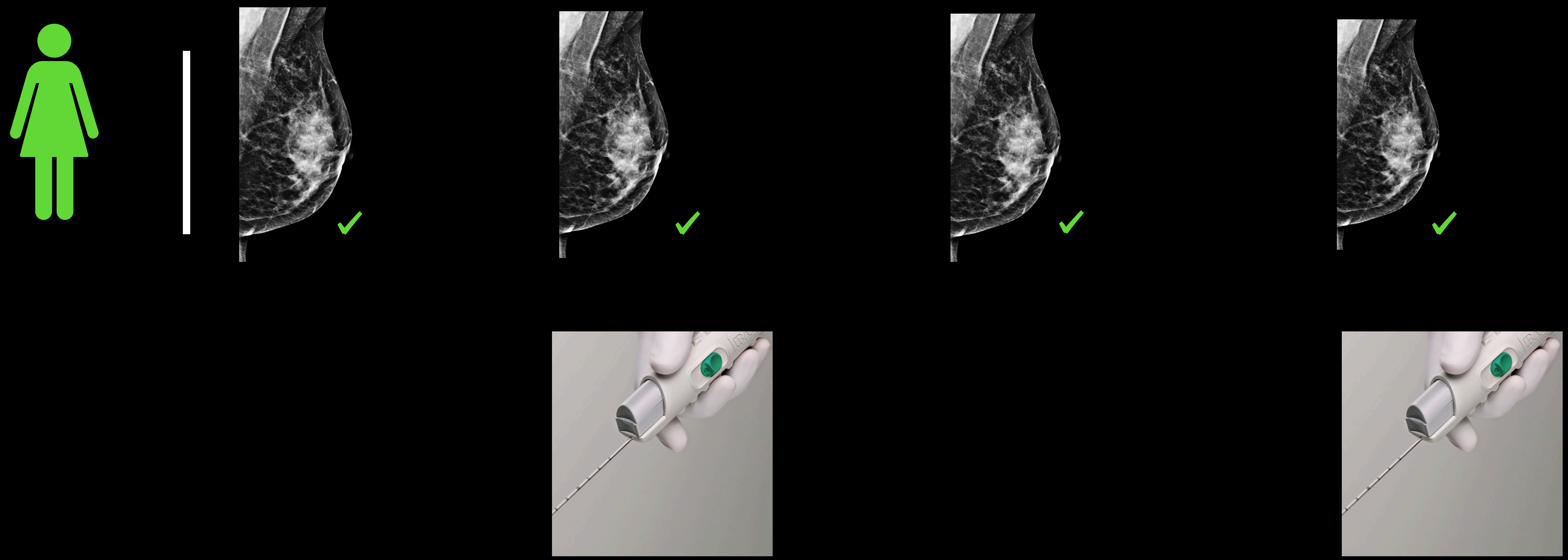


Morbid treatment options, poor chances of survival

We should have done more

The harms of over screening

Patient

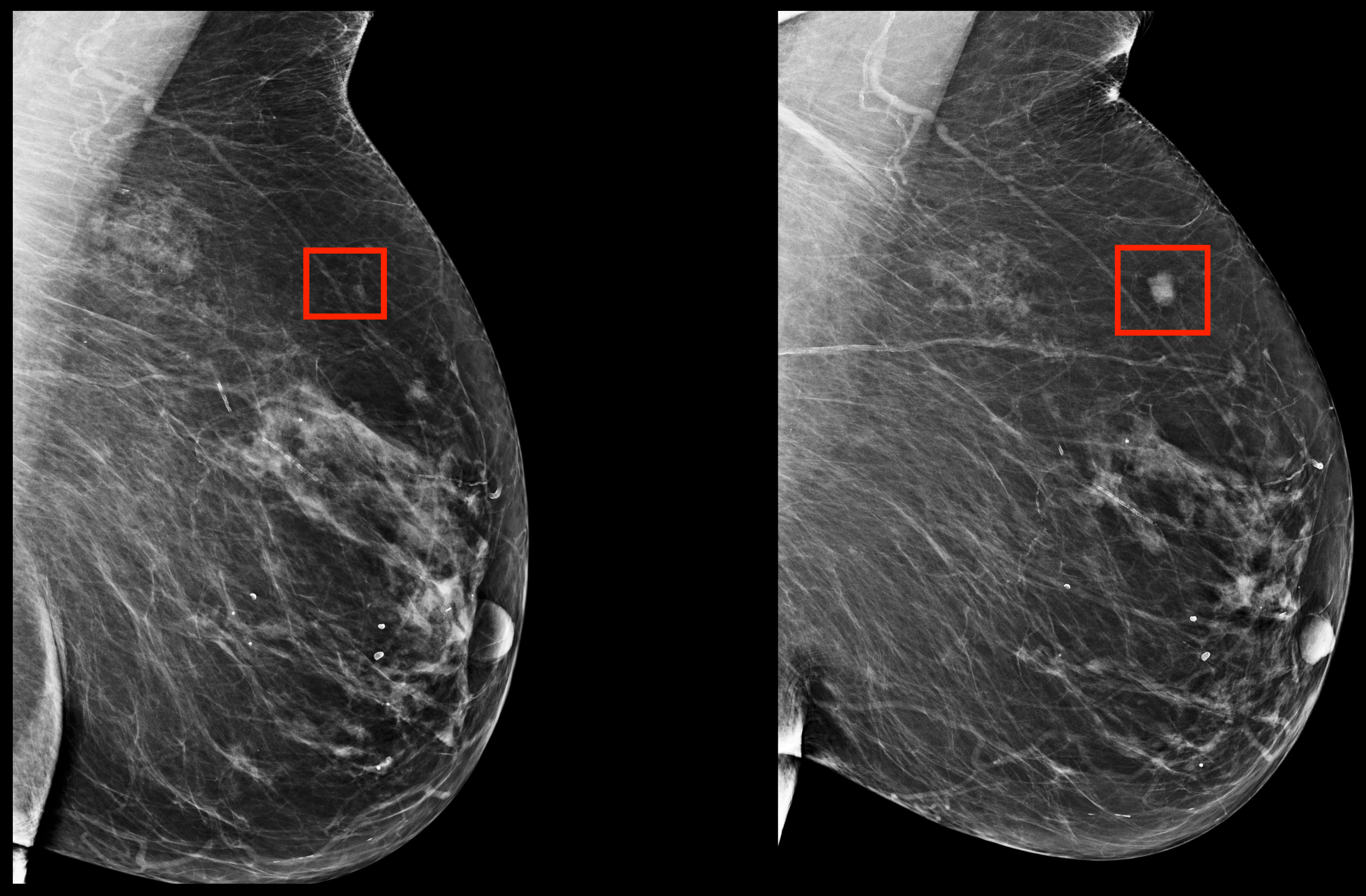


Unnecessary biopsies, terrible anxiety

We should have done less

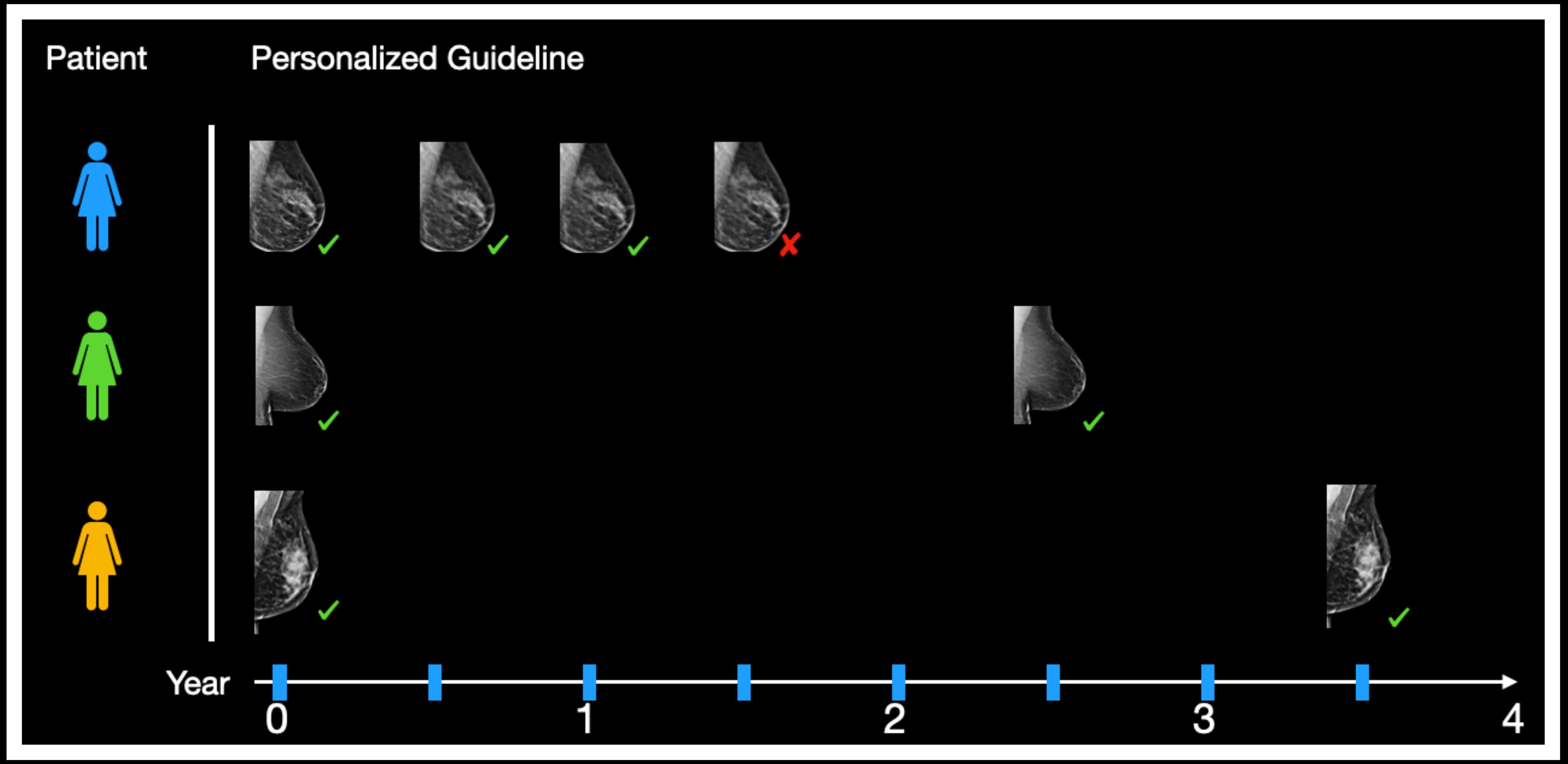
How to catch cancer earlier

Predict Cancer Risk



Year 0 Year 5

Create personalized screening policy



Obstacle: Current clinical tools are insufficient

Questionnaire based (<< 1KB of data per patient)

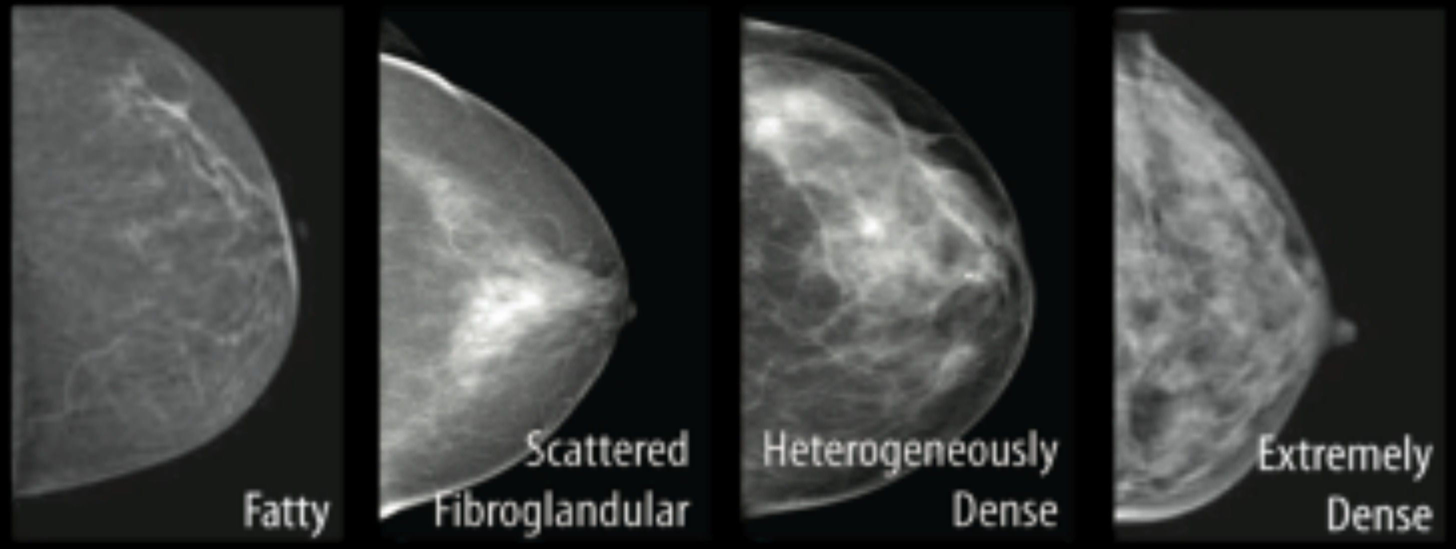
Age
Family History
Prior Breast Procedure
Breast Density

→ Model → **Risk**

Identify **<25%** of future cancers as "high risk"

>95% of "high risk" patients won't get cancer

Hypothesis 1: Maybe density is too variable?

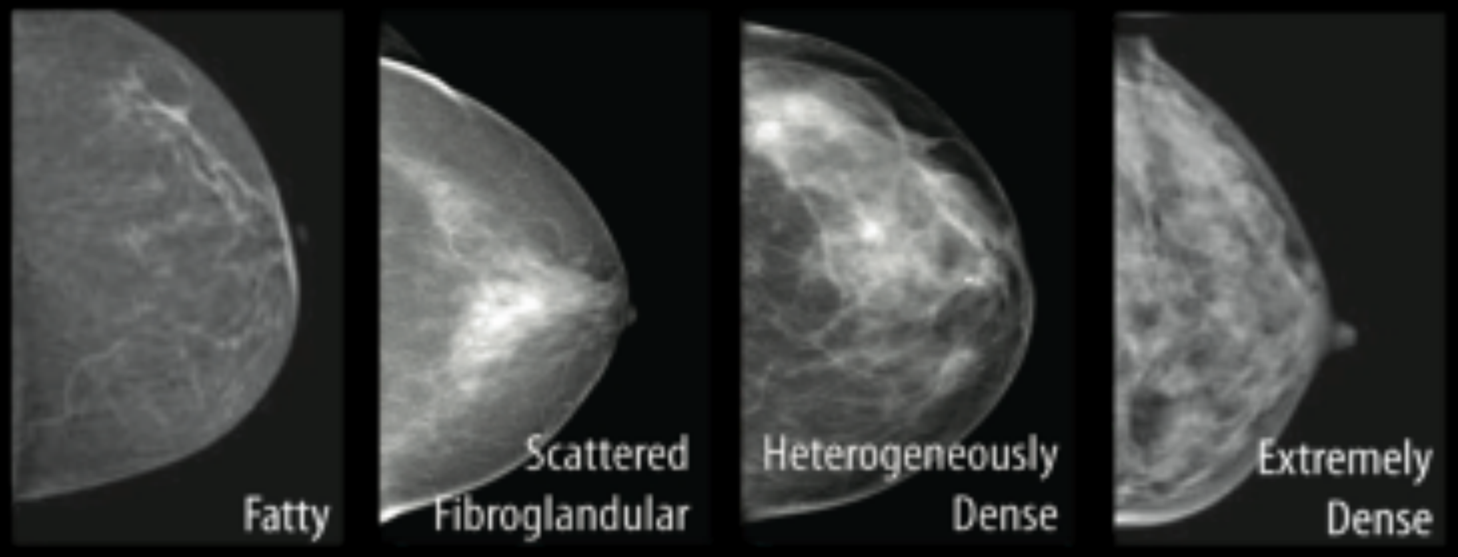


Density

Results: Overall, 36.9% of mammograms were rated as showing dense breasts. Across radiologists, this percentage ranged from $\boxed{6.3\% \text{ to } 84.5\%}$ (median, 38.7% [interquartile range, 28.9% to 50.9%]), with multivariable adjustment for patient characteristics having little effect (interquartile range, 29.9% to 50.8%).

*Brian L. Sprague, PhD; Emily F. Conant, MD; Tracy Onega, PhD; Michael P. Garcia, MS; Elisabeth F. Beaber, PhD; Sally D. Herschorn, MD; Constance D. Lehman, MD, PhD; Anna N.A. Tosteson, ScD; Ronilda Lacson, MD, PhD; Mitchell D. Schnall, MD, PhD; Despina Kontos, PhD; Jennifer S. Haas, MD, MSc; Donald L. Weaver, MD; William E. Barlow, PhD; on behalf of the PROSPR Consortium **

Hypothesis 1: Maybe density is too variable?



Density

Results: Overall, 36.9% of mammograms were rated as showing dense breasts. Across radiologists, this percentage ranged from 6.3% to 84.5% (median, 38.7% [interquartile range, 28.9% to 50.9%]), with multivariable adjustment for patient characteristics having little effect (interquartile range, 29.9% to 50.8%).

*Brian L. Sprague, PhD; Emily F. Conant, MD; Tracy Onega, PhD; Michael P. Garcia, MS; Elisabeth F. Beaber, PhD; Sally D. Herschorn, MD; Constance D. Lehman, MD, PhD; Anna N.A. Tosteson, ScD; Ronilda Lacson, MD, PhD; Mitchell D. Schnall, MD, PhD; Despina Kontos, PhD; Jennifer S. Haas, MD, MSc; Donald L. Weaver, MD; William E. Barlow, PhD; on behalf of the PROSPR Consortium **

Radiology

Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation

Constance D. Lehman, Adam Yala, Tal Schuster, Brian Dontchos, Manisha Bahl, Kyle Swanson, Regina Barzilay

88% binary accuracy on previous logs
97% agreement with an expert radiologist

But we actually solve the problem?

Questionnaire based (<< 1KB of data per patient)

Age

Family History

Prior Breast Procedure

Breast Density

→ Model → **Risk**

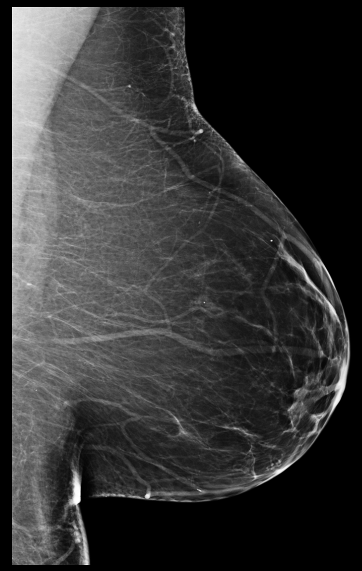
Identify **<25%** of future cancers as "high risk"

>95% of "high risk" patients won't get cancer

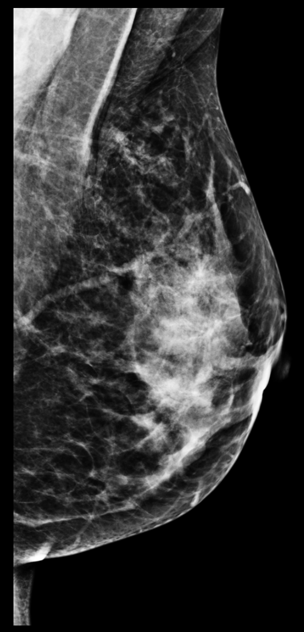
Hypothesis 2: We need to rethink risk

Patient **Data** **Future Outcome**

→ 3 year cancer



→ No cancer



→ 5 year cancer

MB of data per patient

Journal of Clinical Oncology®
An American Society of Clinical Oncology Journal

Check for updates

Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model

Adam Yala, MEng[1,2]; Peter G. Mikhael, BS[1,2]; Fredrik Strand, MD, PhD[3,4]; Gigin Lin, MD, PhD[5]; Siddharth Satuluru, BS[6];

SCIENCE TRANSLATIONAL MEDICINE

Toward robust mammography-based models for breast cancer risk

Adam Yala[1,2]*, Peter G. Mikhael[1,2], Fredrik Strand[3,4], Gigin Lin[5], Kevin Smith[6,7], Yung-Liang Leslie Lamb[8], Kevin Hughes[9], Constance Lehman[8†], Regina Barzilay[1,2†]

Radiology

A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction

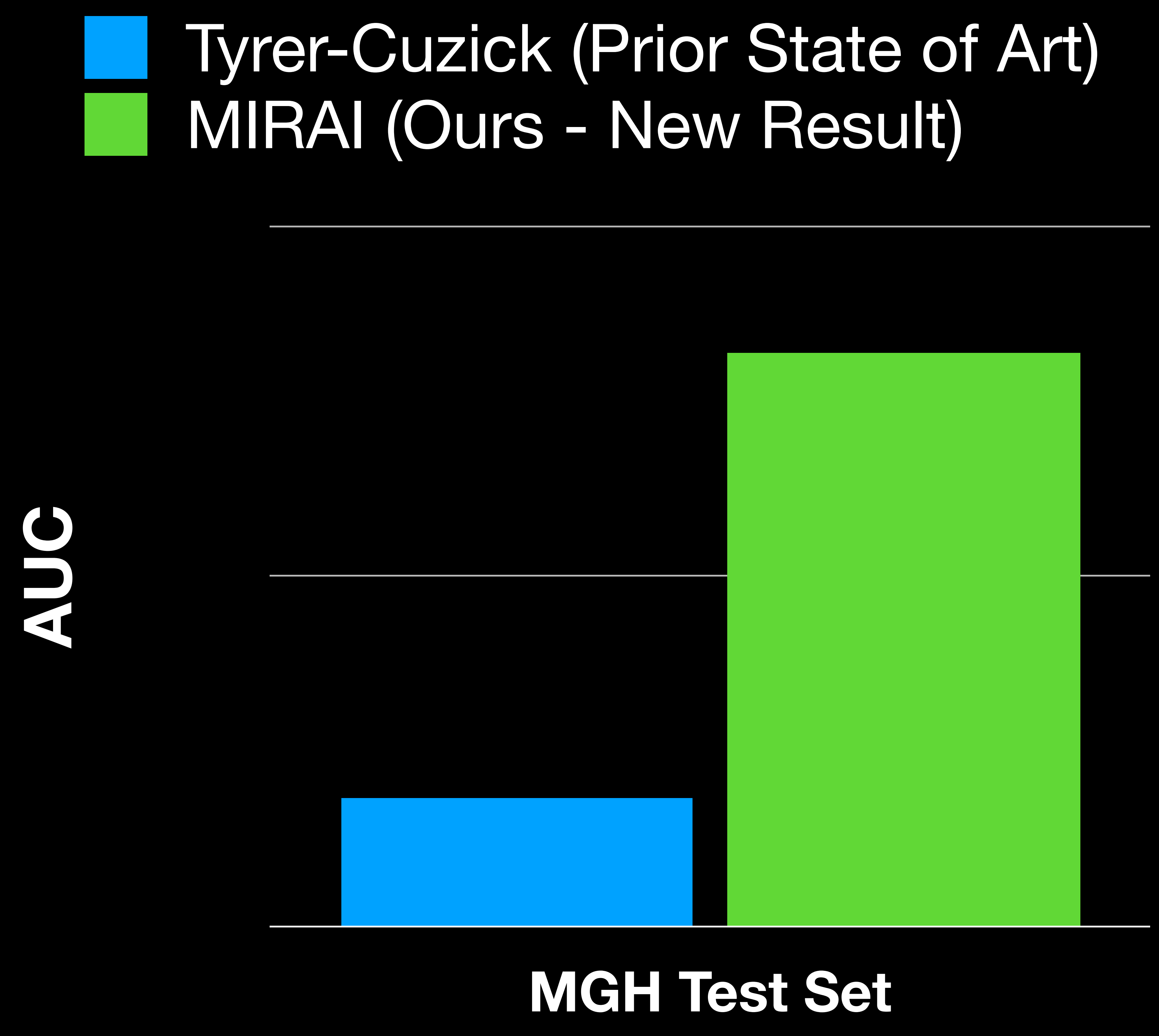
Adam Yala, MEng • Constance Lehman, MD, PhD • Tal Schuster, MS • Tally Portnoi, BS • Regina Barzilay, PhD

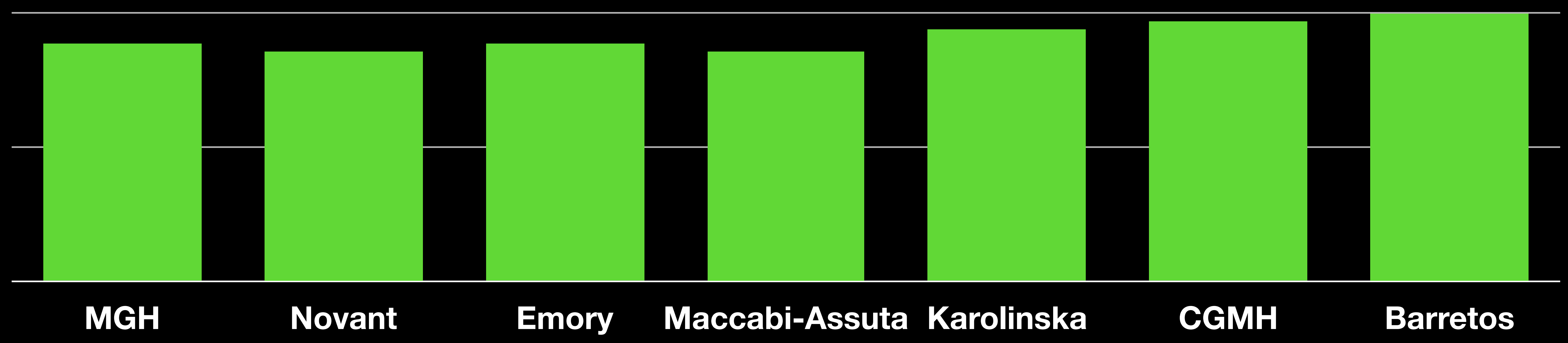
Journal of Clinical Oncology®
An American Society of Clinical Oncology Journal

Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography

Peter G. Mikhael, BSc[1,2]; Jeremy Wohlwend, ME[1,2]; Adam Yala, PhD[1,2]; Ludvig Karstens, MSc[1,2]; Justin Xiang, ME[1,2]; Angelo K. Takigami, MD[3,4]; Patrick P. Bourgouin, MD[3,4]; PuiYee Chan, PhD[5]; Sofiane Mrah, MSc[4]; Wael Amayri, BSc[4]; Yu-Hsiang Juan, MD[6,7]; Cheng-Ta Yang, MD[6,8]; Yung-Liang Wan, MD[6,7]; Gigin Lin, MD, PhD[6,7]; Lecia V. Sequist, MD, MPH[3,5];

Maintains accuracy across diverse populations





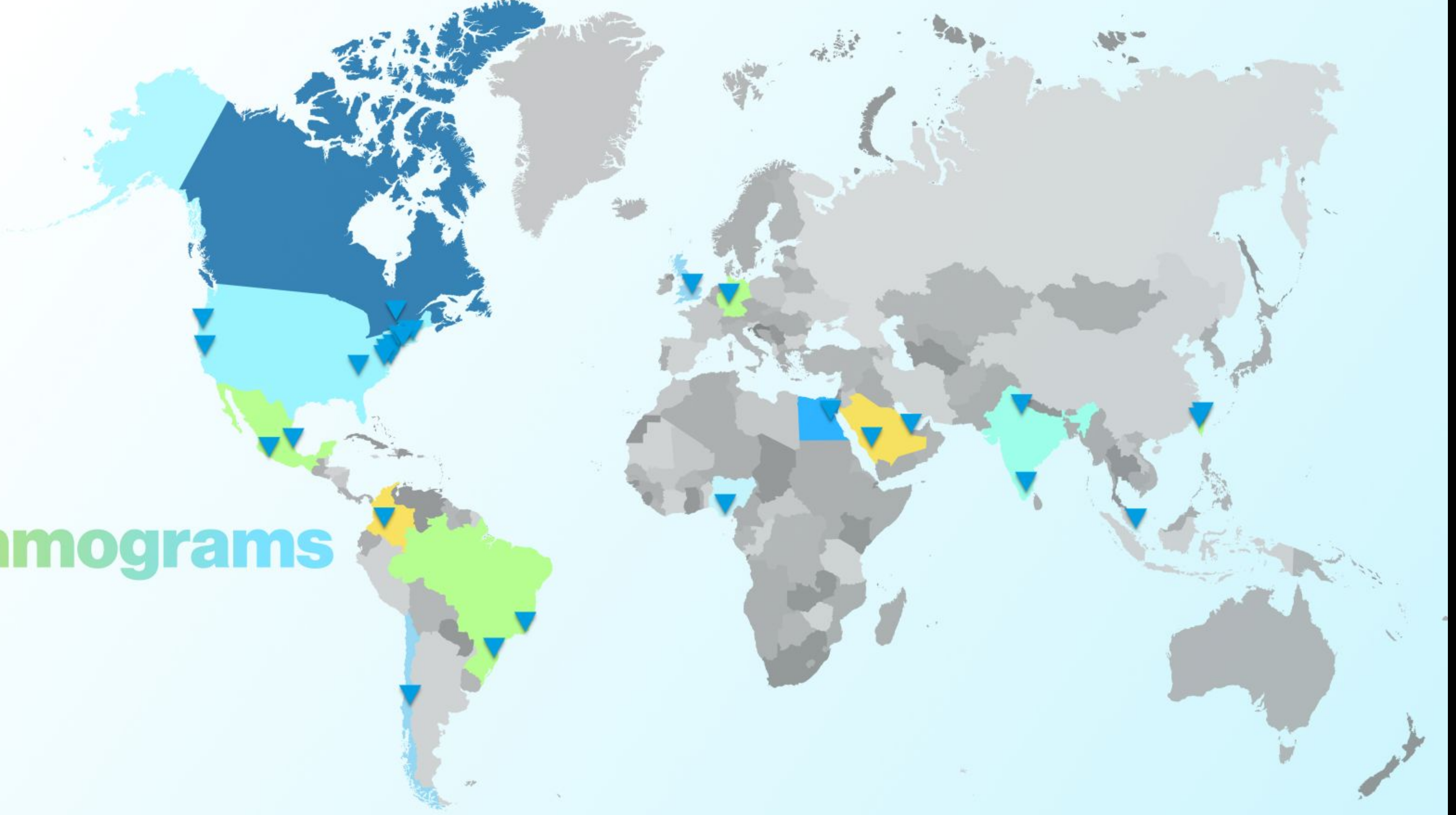
Jameel Clinic

MIRAI

1.5M+ mammograms

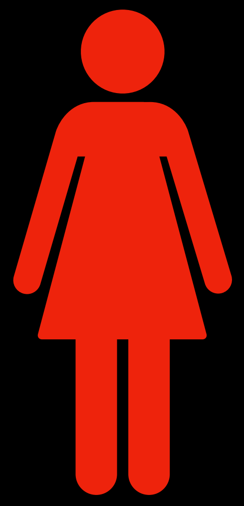
43 hospitals

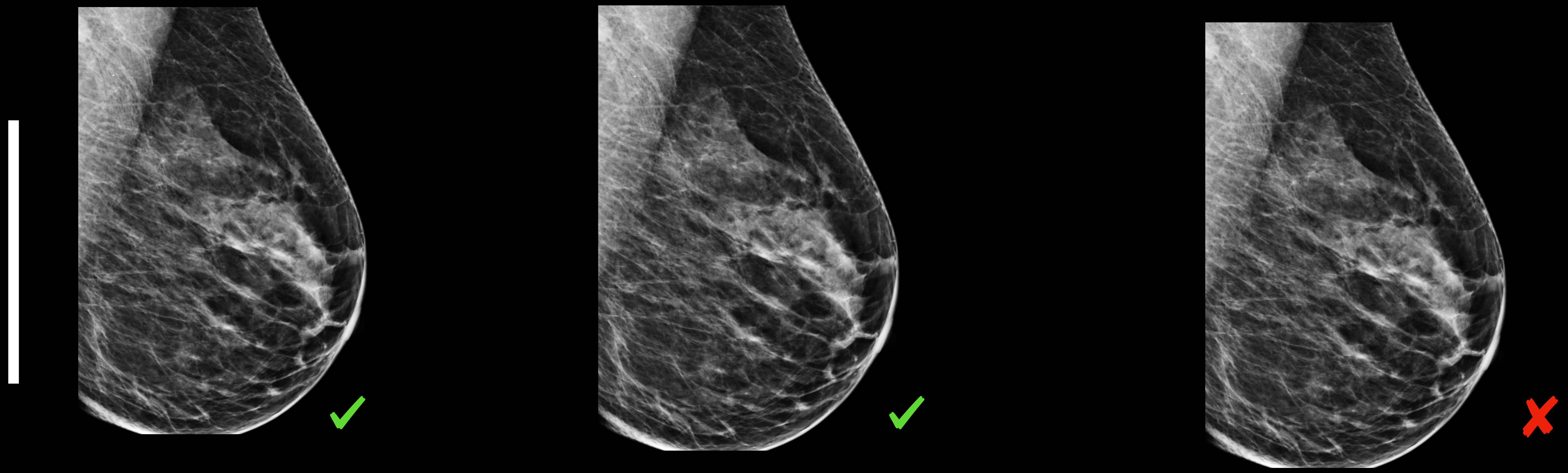
14 countries



The harms of late diagnosis

Patient

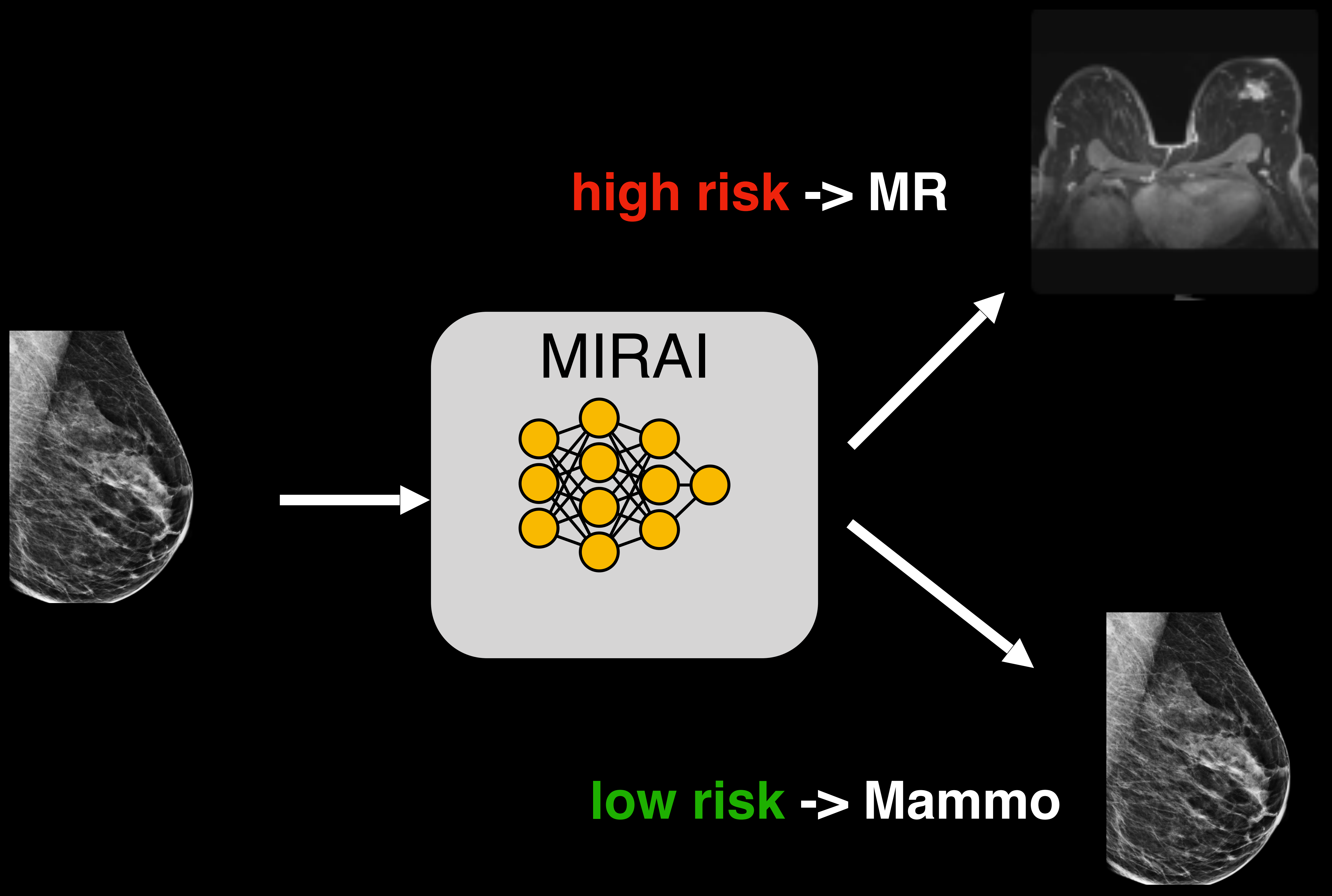




Morbid treatment options, poor chances of survival

We should have done more

Ongoing Prospective Trials: Mirai-MRI

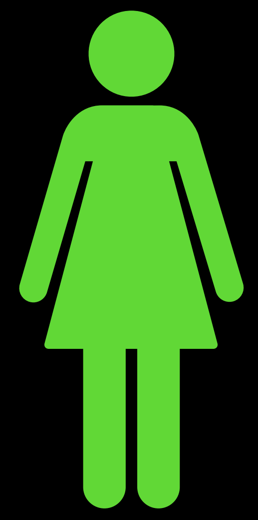


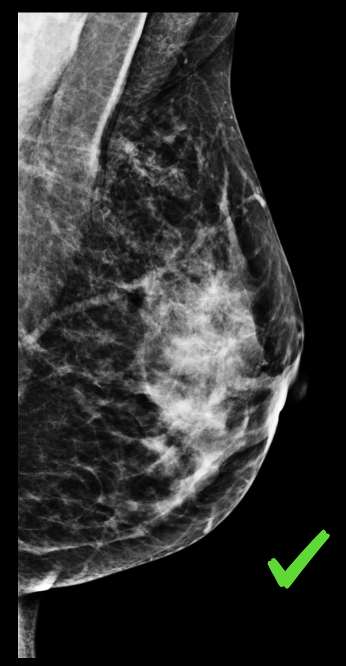
Mirai-based Supplemental Imaging

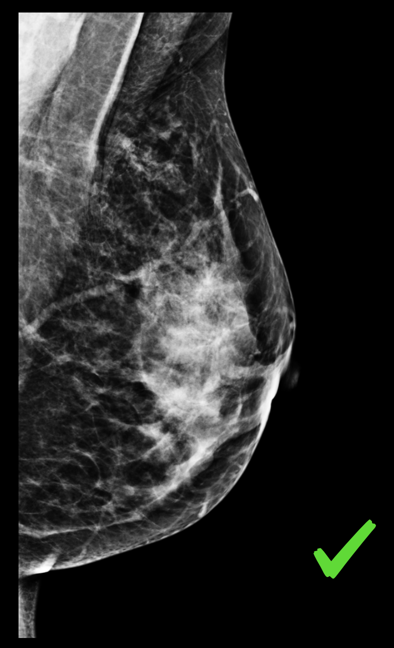
NCT 05968157

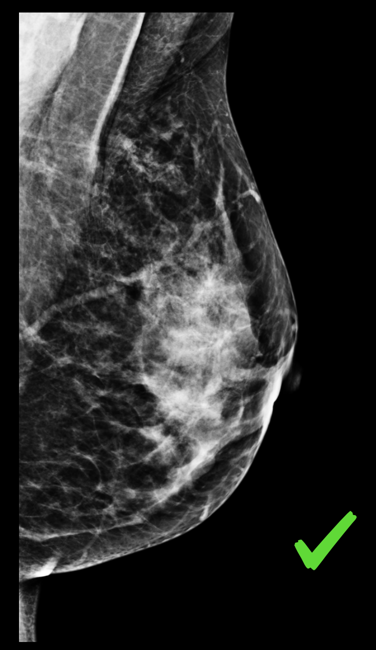
The harms of over screening

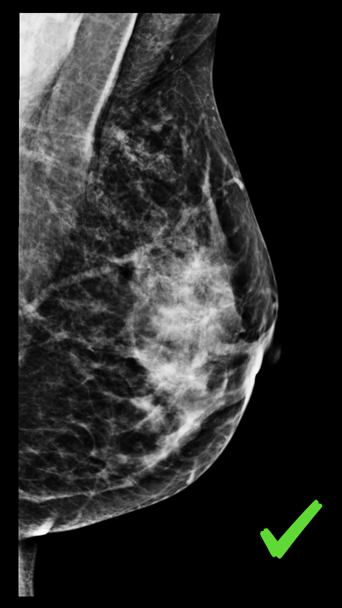
Patient











Unnecessary biopsies, terrible anxiety

We should have done less

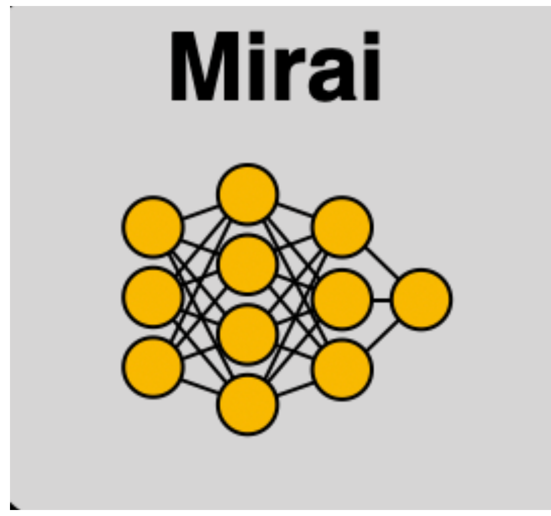
Ongoing Prospective Trials: Mirai-SDA

SDA Workflow:

- **Realtime** AI-based cancer risk assessment
- Invite high risk patient for **same-day** diagnostic exam
- Expectation: >50% of cancer cases will receive same-day diagnostic
- **100%** enrollment so far!

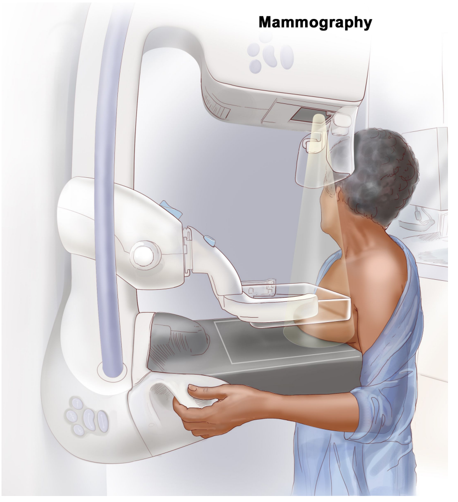
Screen

<5 seconds



AI Risk

Top 10% Risk



Same Day Diagnostic

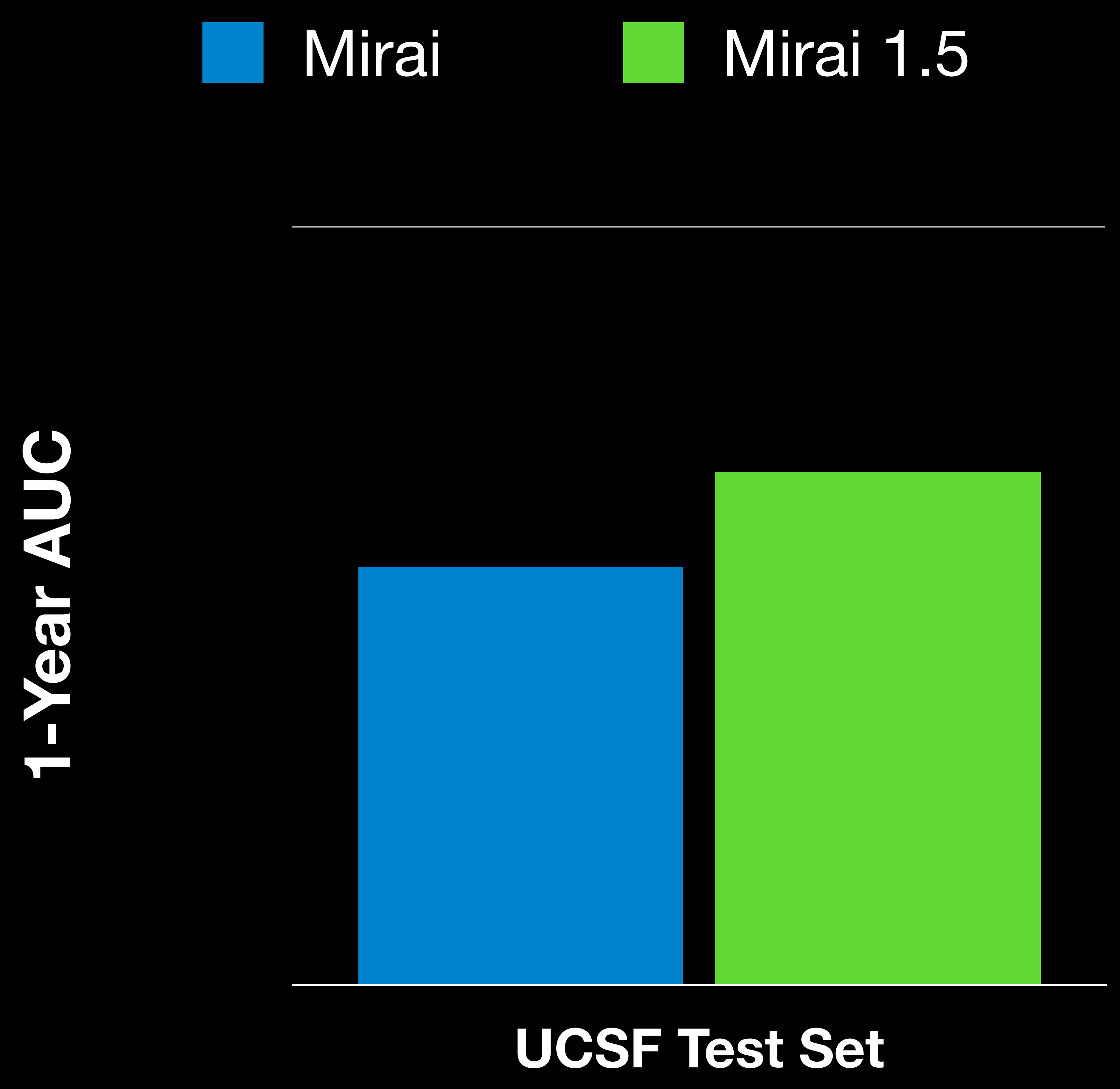
A practical instantiation: *My own research journey*

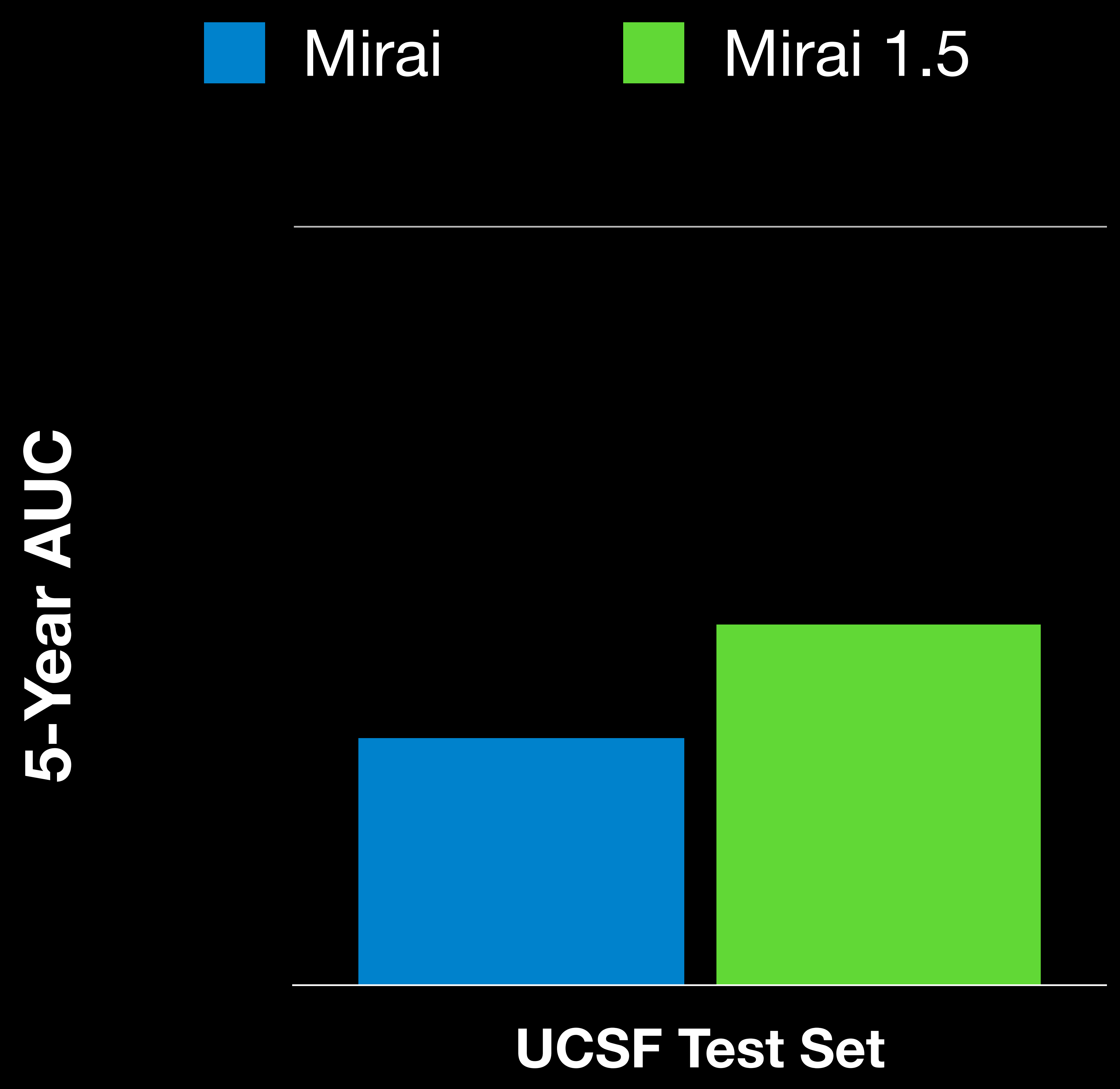
Clinical Hypothesis → **AI approach** → **Rigorous Validations** → **Deploy**

Computational
PRECISION HEALTH

Berkeley | UCSF

Mirai 1.5: Improving mammography risk models



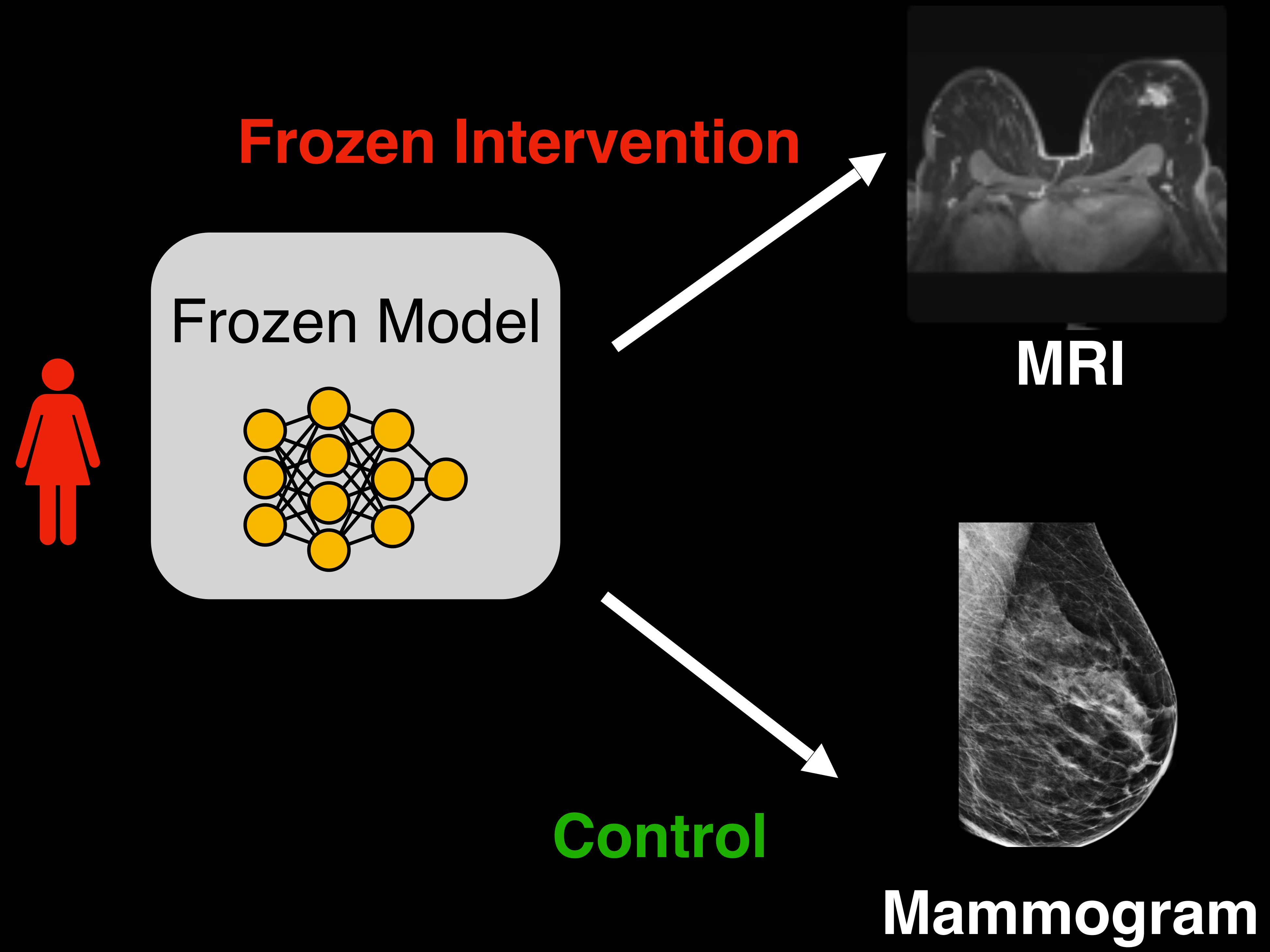


** New unpublished model*

How do we evaluate constant evolving AI tools?

Evaluation

Traditional Randomized Trials

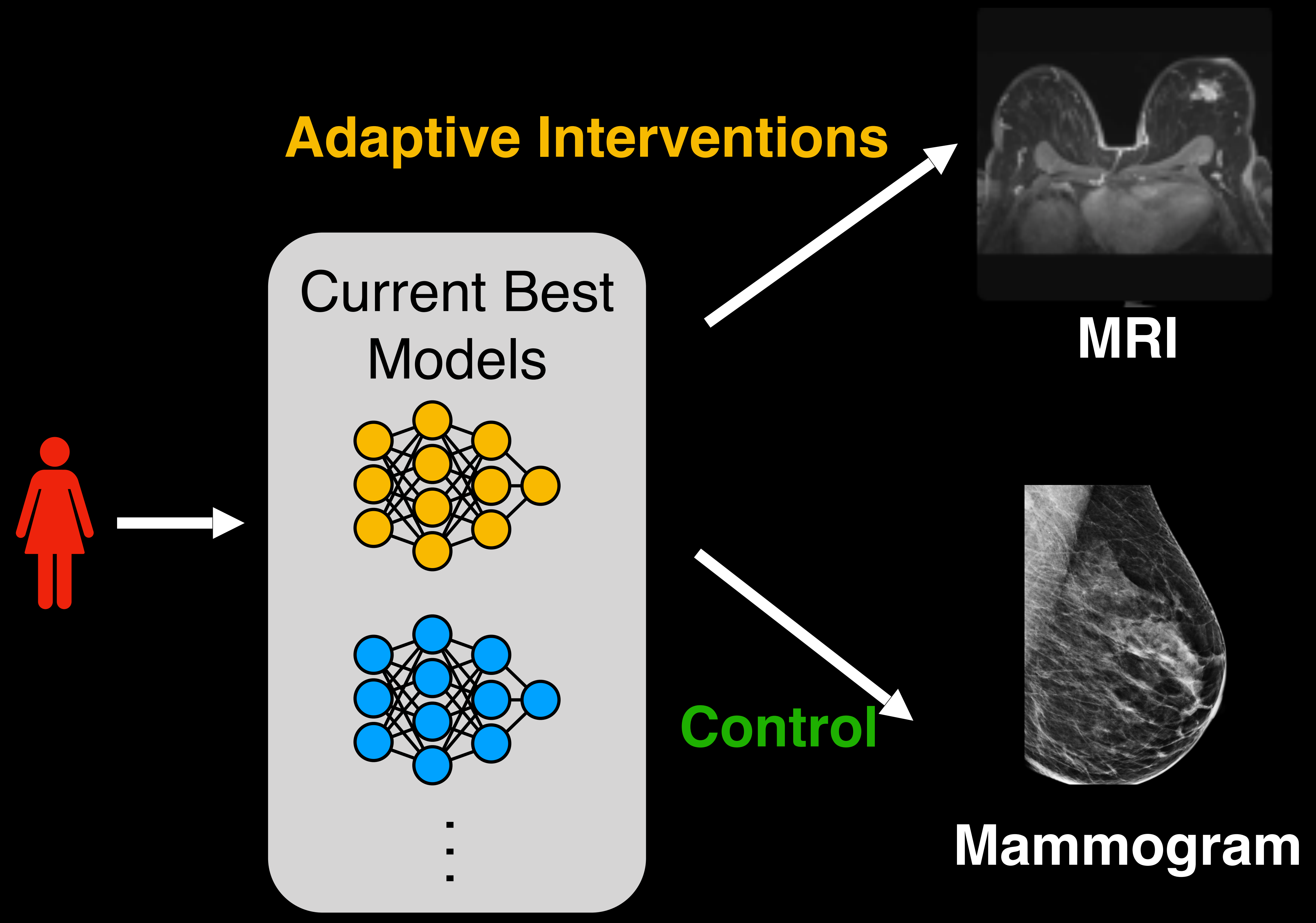


AI **obsolete** by end of trial..
Incompatible with rapid model innovation

Led by: Wenxin Zhang

Simulation: Mirai-SDA using BRIDGE Trials

BRIDGE Trials: Data Reuse for AI Trials



Continuous **platform trials** for improving AI
Combine evidence across all models + historical data (RWE). **Fast**

Mirai-1 Trial:

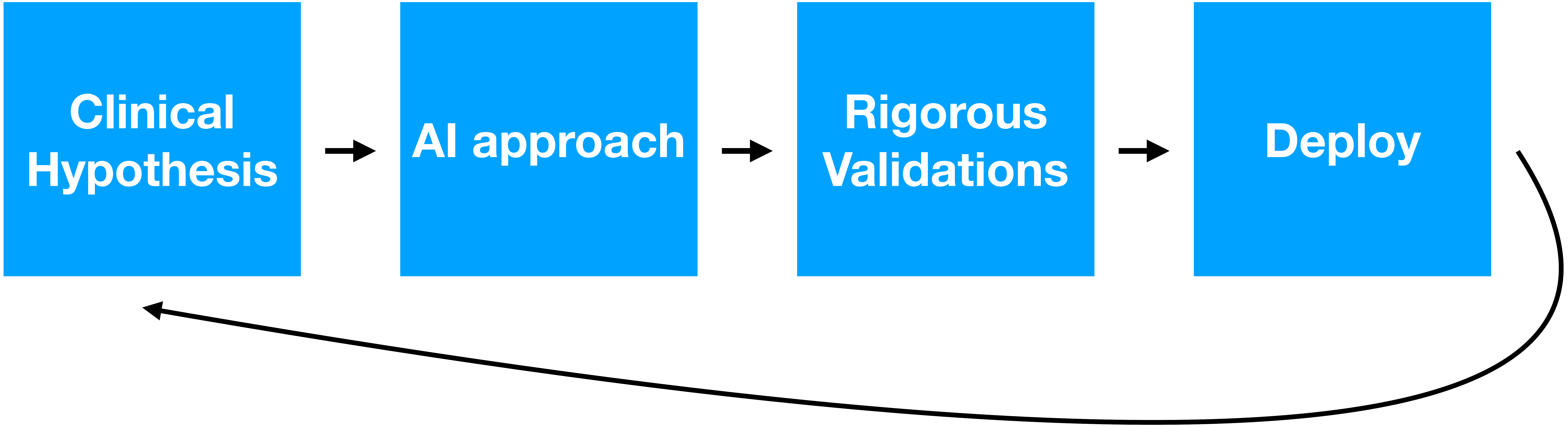
- Enroll 100 patients

Mirai-1.5 Trial

- Naive trial: 100 new patients
- **70 %** patients high risk by M1.5 and M1.0
- With data-reuse: 30 new patients

Trials will only get faster and easier

This class: How do empower students to do CPH work?



Course overview

- What is this class?
    - Foundations for CPH research
    - Broad exposure to:
        - diverse health care areas
        - core machine learning methods
- 14 weeks of class:
    - Clinical and Method lectures
    - Precision Health Proposals
    - Hands on projects with real data. Build models, analyze clinical impact.

CPH100 Roadmap

Clinical Survey

Primary Care

Emergency Medicine

Cardiology

Cancer

Method Arc

ML Foundations

NN Foundations

Evaluation + Deployment

LLMs for Clinicians and Patients

Final Project Presentations

Computational
PRECISION HEALTH

Berkeley | UCSF

CPH 100: Projects and Assignments

- **Computational Health Proposals**
    - Pitch computational projects + evaluation plans based on guest lectures
- **Hands on projects (Project 1 and Project 2):**
    - Implementing components of SOTA models
    - Evaluating and clinical impact

Computational Health Proposals

CPH100: Computational Health Proposal

Due Dates: See Course Schedule

Length: ≤ 1 page (single-spaced, 12 pt font)

Overview

This recurring assignment challenges you to identify a meaningful healthcare problem based on one of the recent clinical guest lectures and propose a corresponding computational approach to address it. Choose a healthcare problem inspired by a clinical guest lecture delivered since your last proposal. Your proposal should articulate:

1) A specific clinical challenge
2) A plausible computational strategy
3) A clinically grounded evaluation framework

Guided Class Projects

- Hands-on projects: **github.com/yala/CPH100_25_release**
  - **Project 1:** ML tools to improve lung cancer screening
    - Due: **Sept 23rd**
  - **Project 2:** Deep learning tools for medical imaging
    - Due: **Oct 16th**
  - **Mix of Individual Code + Team (5 student) reports**

Final Projects: Choose your own CPH Adventure

- Details shared in 2nd half of course with Prof Irene!

Resources Status Check

- **Computational Resources:**
  - Local laptops
  - Savio Cluster
- **How to get help?**

- **Class philosophy:** we are building this together

Expectations and grading

- Class is letter graded
- Grade breakdown:
  - Project 1: 20%
  - Project 2: 20%
  - Computational Health Proposals: 20%
  - Final Project: 40%
- Class recordings posted when possible

How to reach me

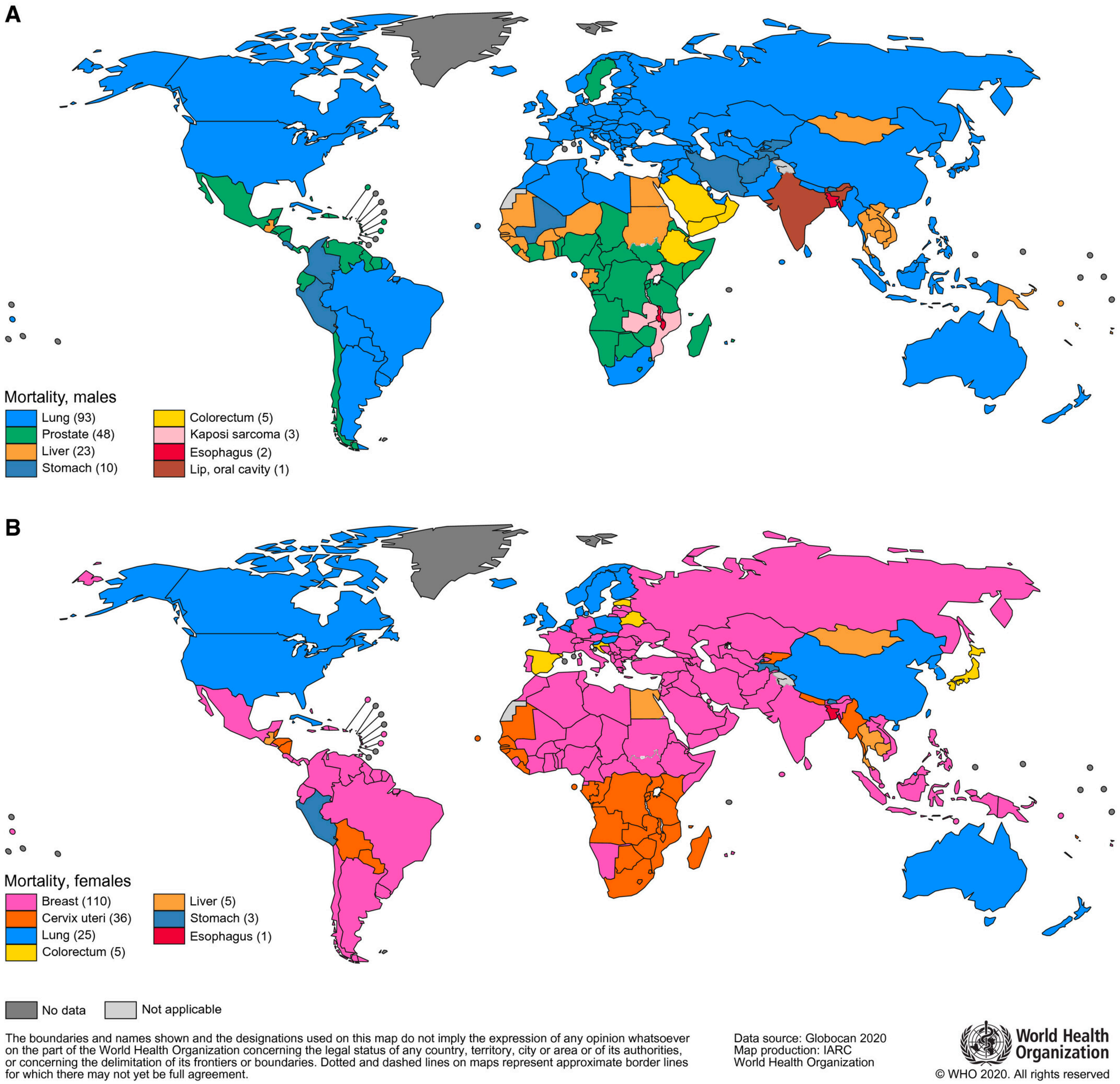
- Email: yala@berkeley.edu
- Office locations:
  - CPH Suite 120AB, Earl Warren Hall, UC Berkeley
    - Office Hours: Weekly Weds
- **I am here to support you!**
- **Email me or come to OH if you want to chat about the class, research or anything else.**
- **We are building this class together!** Will regularly check-in on pacing, projects, more

Course Overview Questions?

What can CPH do for cancer care?

Global cancer burden

- **2nd leading cause of death**
- 19 million diagnoses in 2020
- 10 million deaths in 2020
- Cancer incidence varies globally
- Evolving epidemiology
- Diversity in genetics, environments and lifestyles

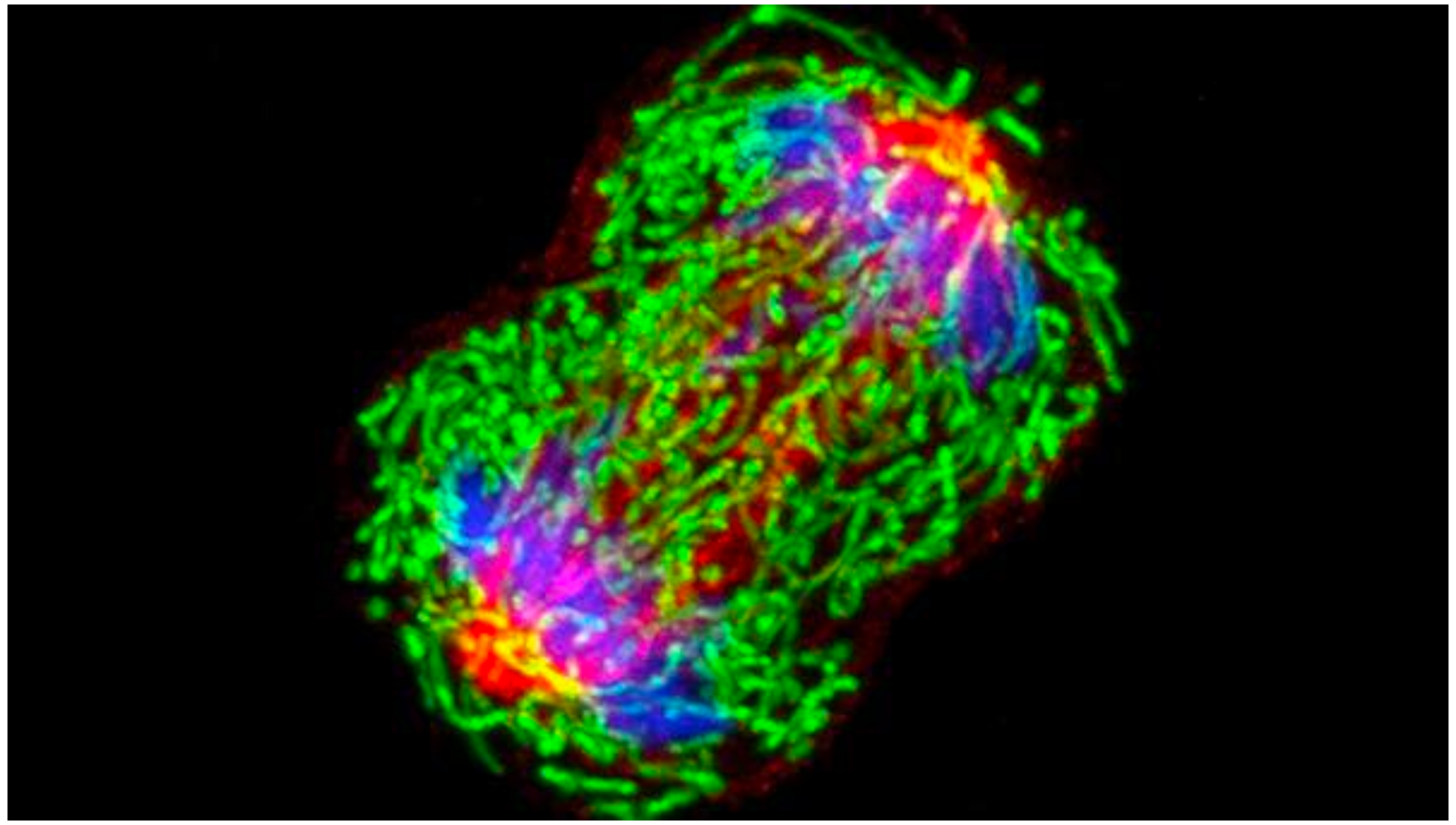


Computational
PRECISION HEALTH

Berkeley | UCSF

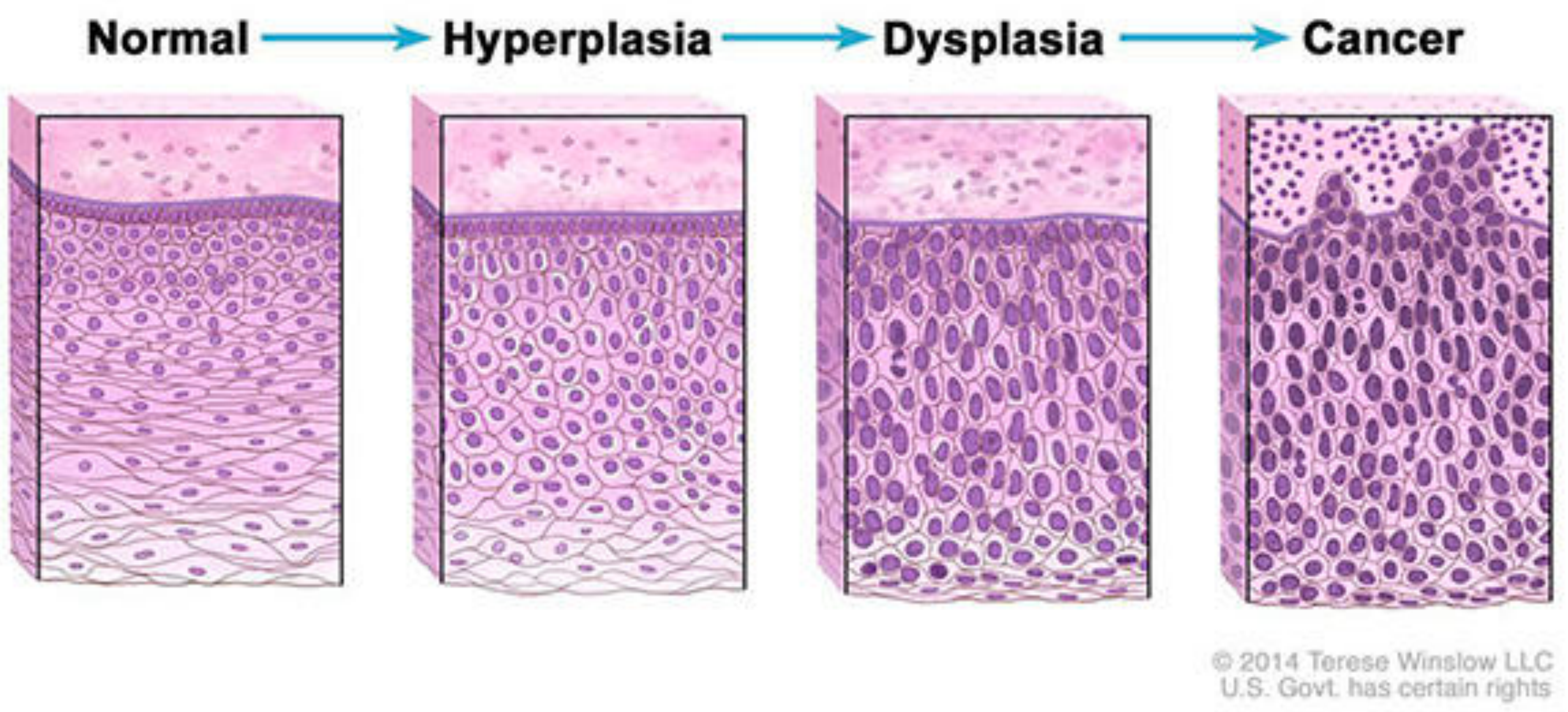
What is cancer?

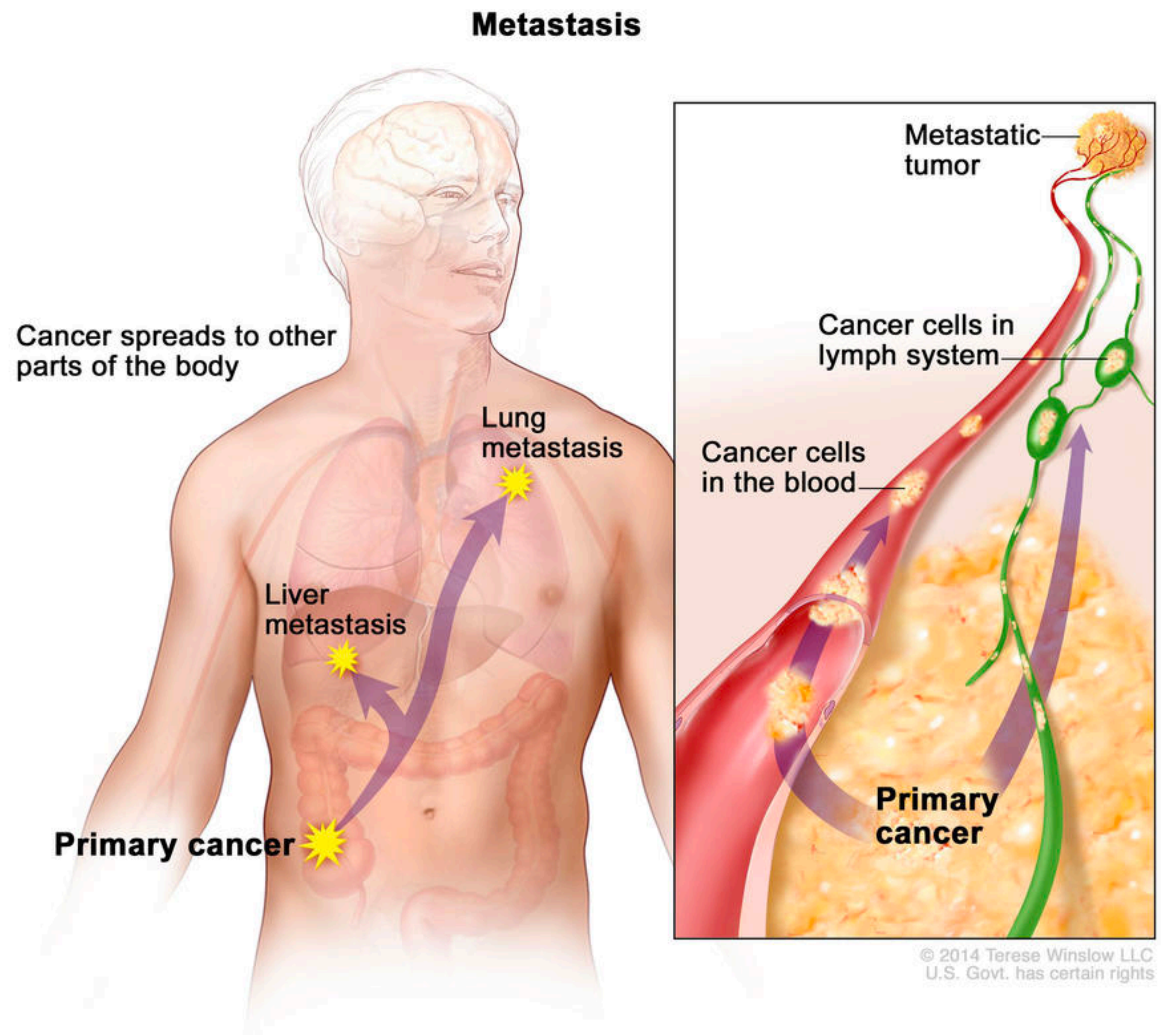
- Diseases of **uncontrollable cell growth**
- Absence of normal regulatory circuits:
  - Grow without appropriate growth signals
  - Ignore cell-death signals
  - Hide from or suppress immune system
- Caused by genetic defects in underlying cells



What is cancer?

- Not all abnormal cell growth is cancer
- **Cancer invades** other tissue (metastasis)
- Biology of distant cancer driven by primary
- >100 types of cancer, **diverse biology**

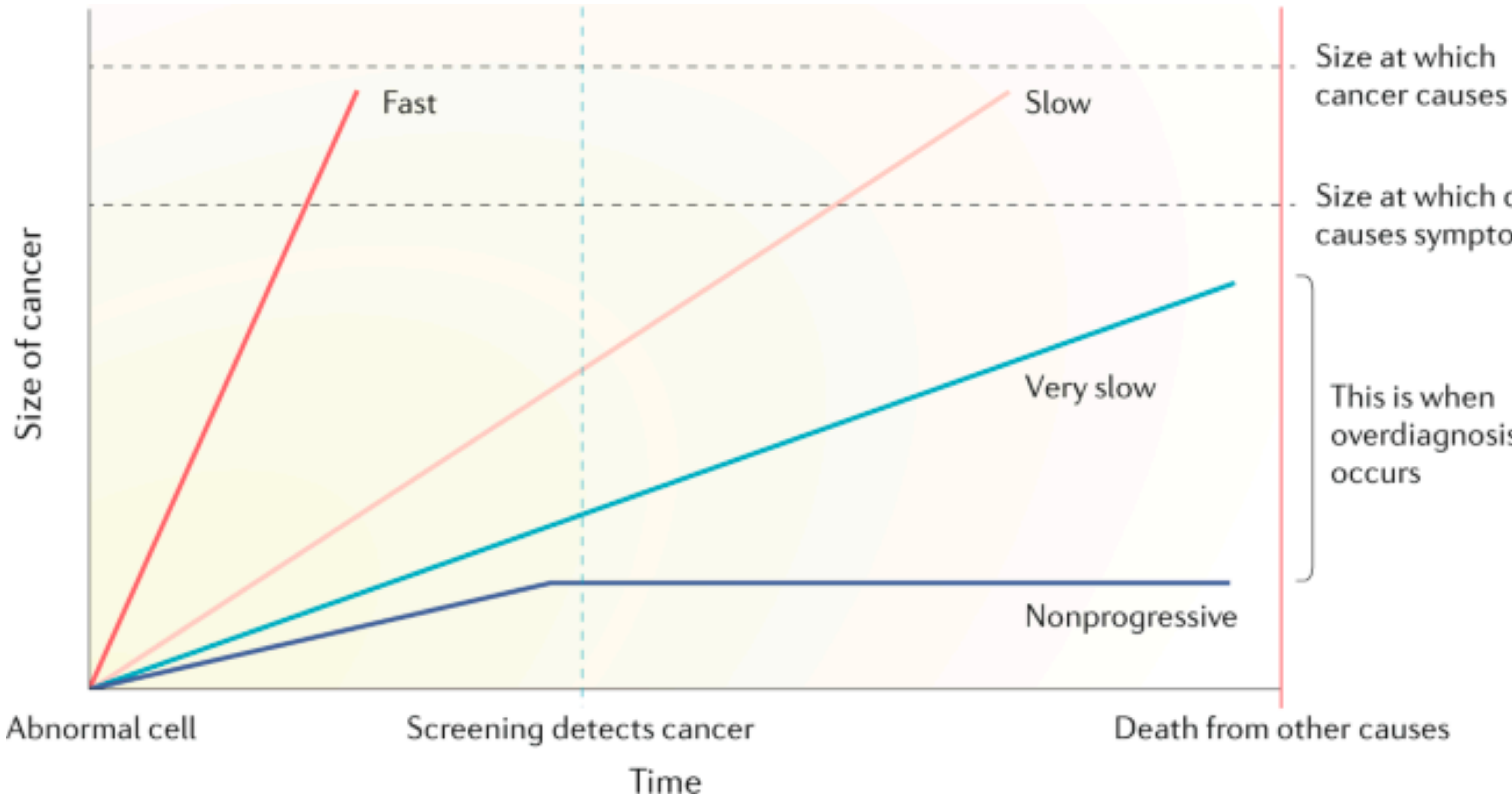




Computational
PRECISION HEALTH

What is cancer?

- **diverse biology:** snails to hares
- Accumulates mutations as it grows
- Generally harder to treat as it grows
  - More heterogeneous
  - Adaptive adversary



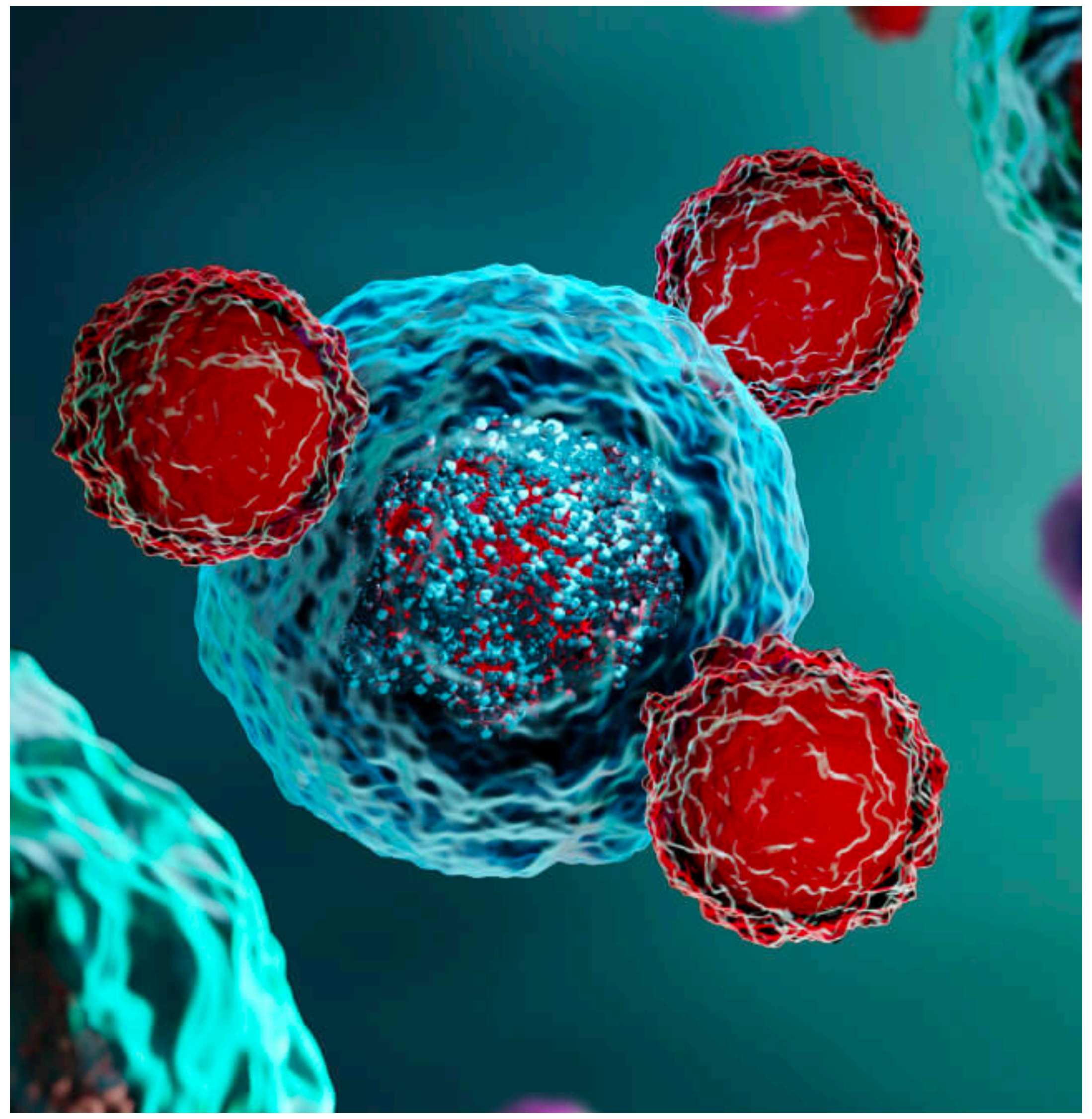
Srivastava, Sudhir, et al. "Cancer overdiagnosis: a biological challenge and clinical dilemma." *Nature Revie*
Cancer 19.6 (2019): 349-358.

Computational
PRECISION HEALTH

Core challenges in cancer care

Screening

Diagnosis



Treatment

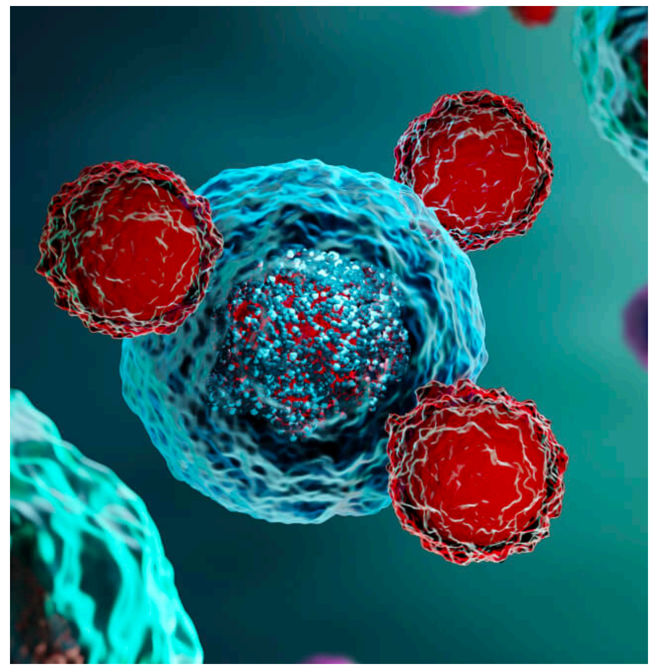
Computational
PRECISION HEALTH

Berkeley | UCSF

Core challenges in cancer care

Screening

Diagnosis

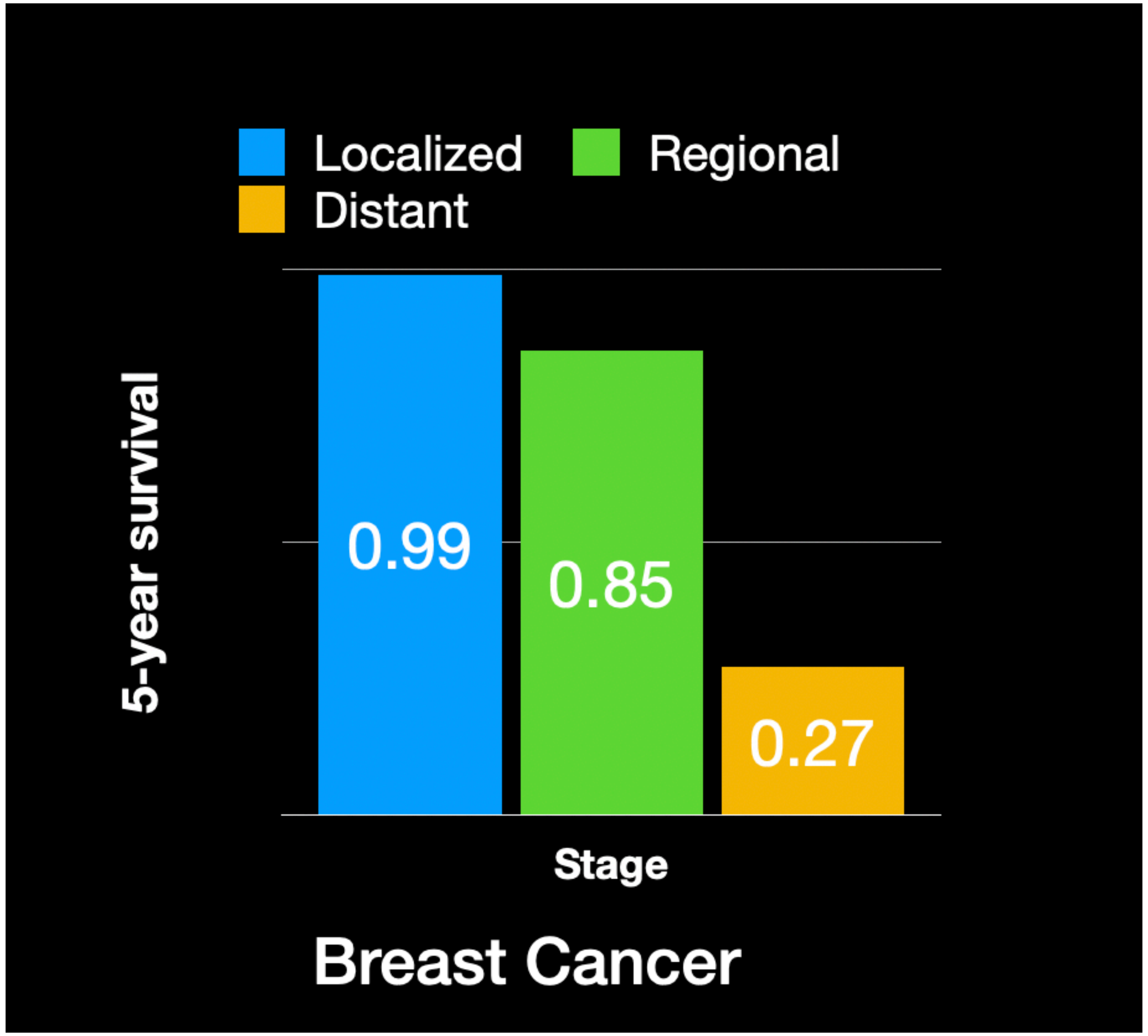


Treatment

Identifying the right intervention at the right time

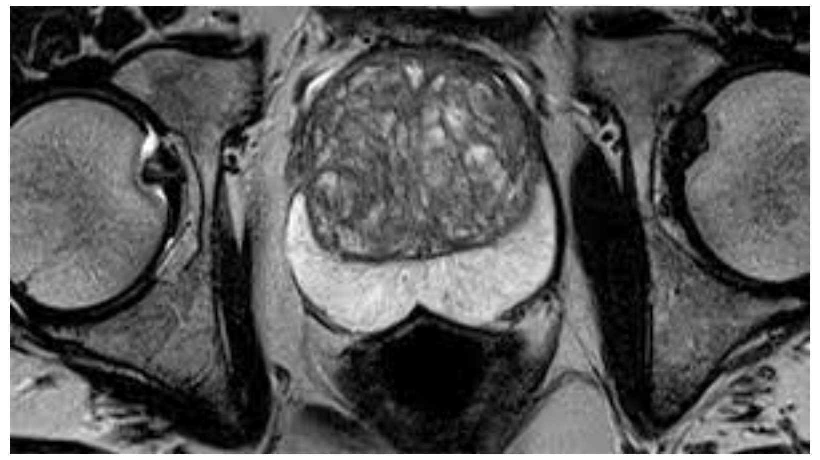
How to screen for cancer?

- Every screening has harms
- Tiny portion of population has cancer
    - 0.5% of mammograms find cancer
- **Questions:**
    - Who should get screened?
    - When?
    - And with what?



How to diagnose cancer?

Imaging



Biopsy

Surgery

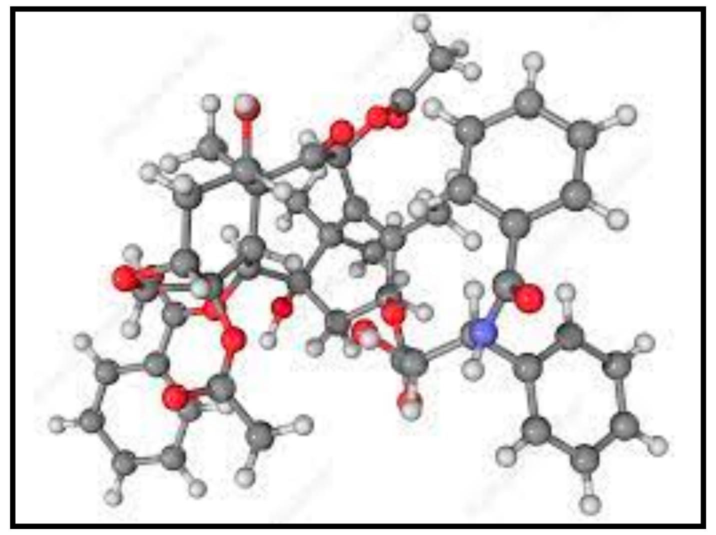
- Wide arsenal of tests
- Balance sensitivity, specificity and harms
- **Question:**
  - Who has cancer?
  - How do we design the diagnostic workflow?
  - Which cancers are worth treating?

How to treat cancer?

- Wide array of treatment pathways
    - Radiation, Surgery, Chemo, etc.
- **Question:**
    - Who will respond to a particular treatment pathway?
    - How do you can adapt treatment as you see new information?
    - What do clinical outcomes teach us about drug design?

Surgery

Radiation

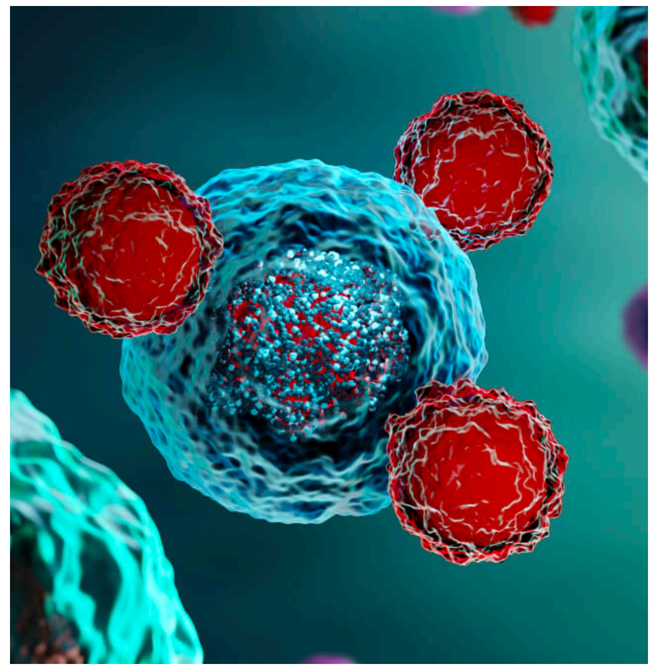


Drugs

Core challenges in cancer care

Screening

Diagnosis



Treatment

Identifying the right intervention at the right time

Core computational challenges in cancer care

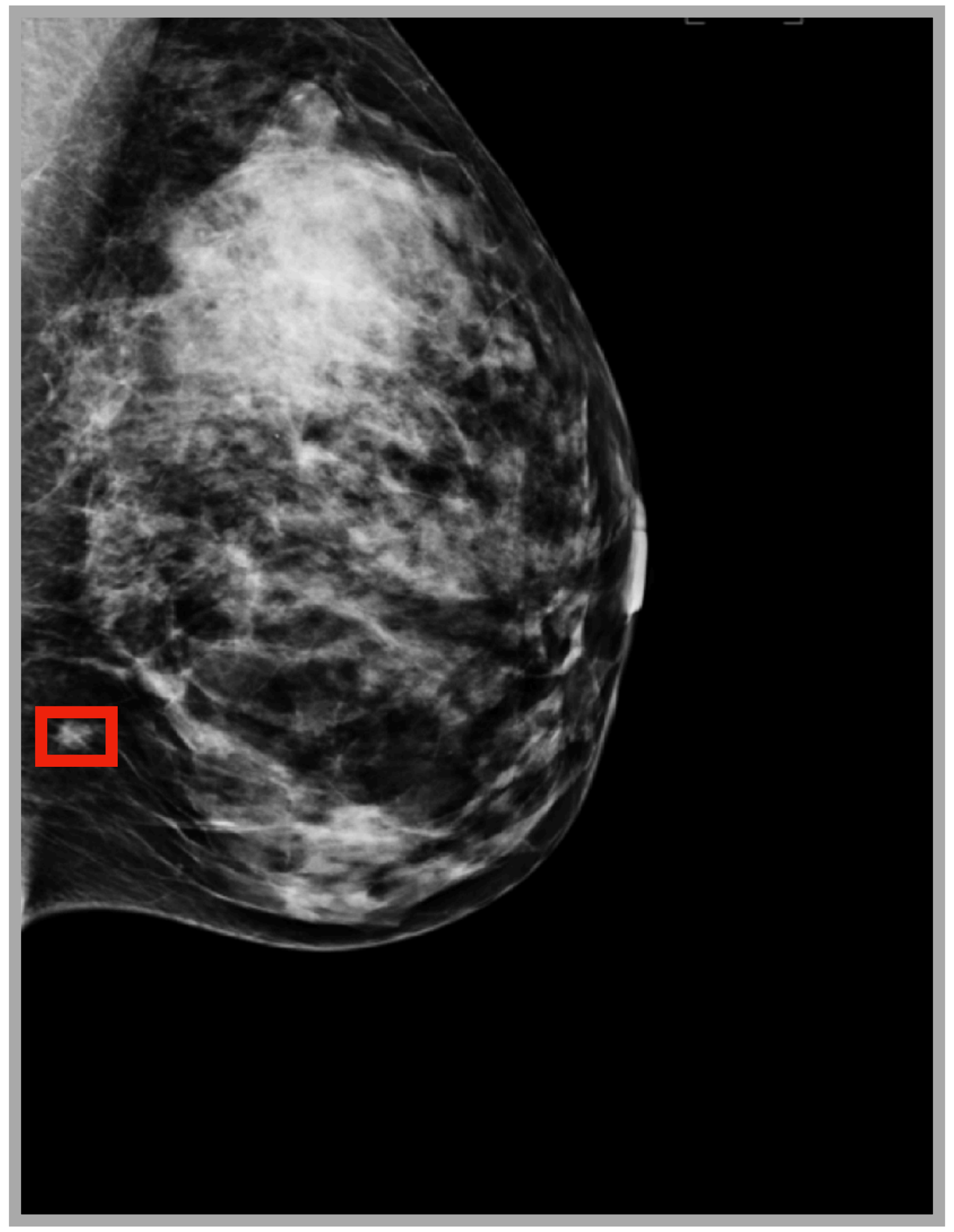
- Cancer is one of the most data-rich areas of medicine
    - e.g. EMR, radiology, pathology, single-cell gene expression
- Each cancer patient has many **GB of data**
    - Context length of GPT-5 in ~<MB range
    - Standard of care generally relies on a few Bits (e.g. age, family history)

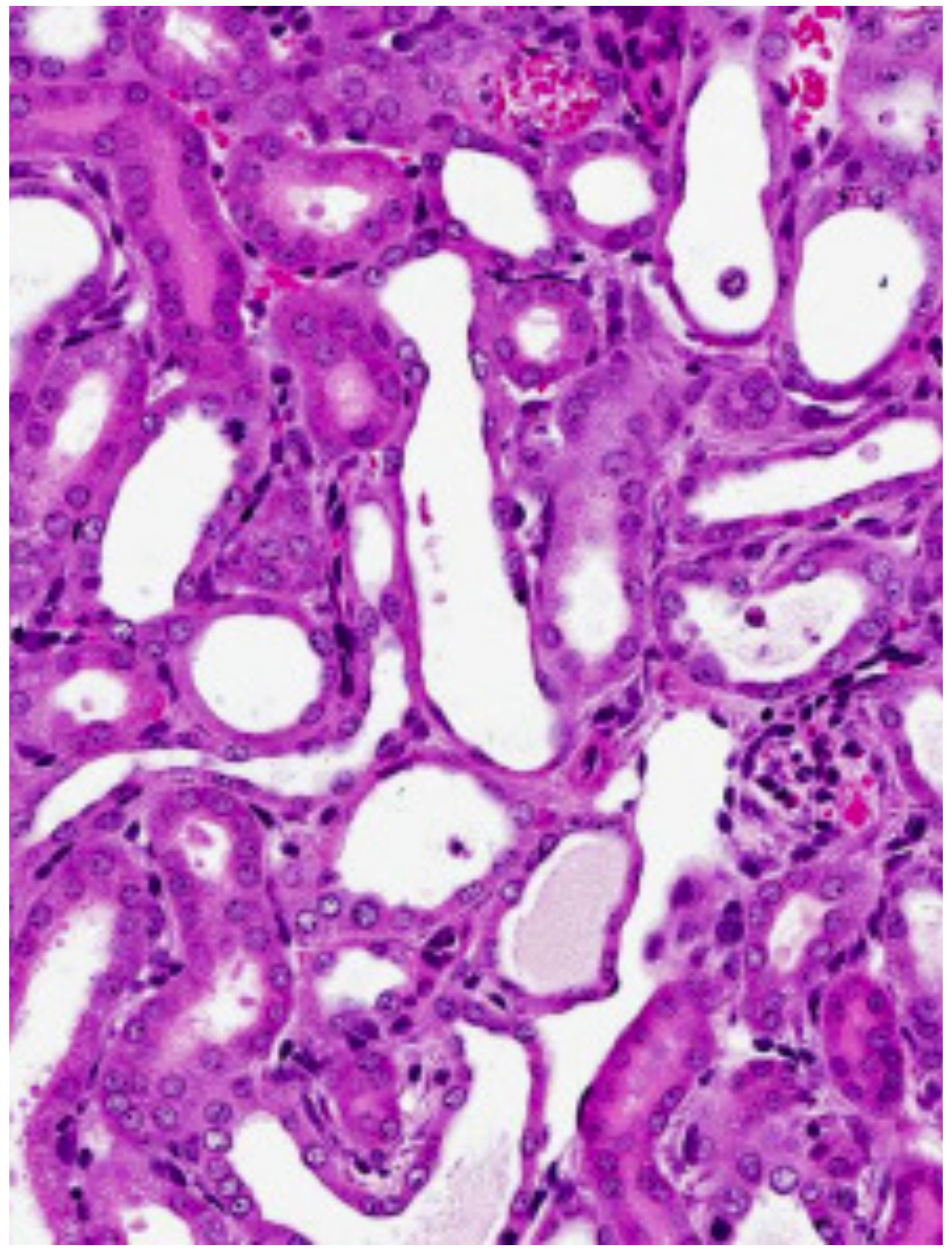
Berkeley | UCSF

Dealing with scale: The data are too large

200 x 220

50 x 50px





ImageNet Scale: 224 x 224

Mammogram Scale: 3k x 2k

Pathology Scale: 100k x 100k

Core computational challenges in cancer care

- Cancer is one of the most data-rich areas of medicine
  - e.g. EMR, radiology, pathology, single-cell gene expression
- Each cancer patient has many **GB of data**
  - Context length of GPT-5 in ~<MB range
  - Standard of care generally relies on a few Bits (e.g. age, family history)
- Key challenge:
  - **Prediction:** How leverage all we have to better understand the disease?
  - **Control:** How can we use that capacity to improve care?
  - **Translation:** How do build evidence for use?

This class: How do empower students to do CPH work?

Clinical Hypothesis → AI approach → Rigorous Validations → Deploy

Questions?

Computational
PRECISION HEALTH

Berkeley | UCSF