De-identified Digital Pathology Training Session

March 21, 2025

Basic pathology asset structure

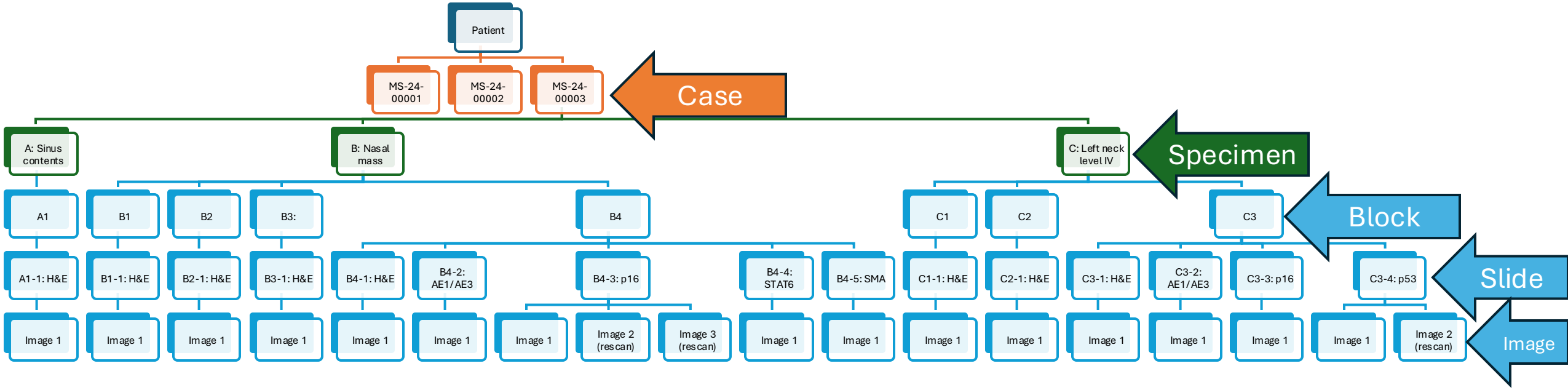
One to many to many to many to many to many

- A patient can have many cases
- A case can have many specimens
- A specimen can have many blocks
- A block can have many slides (which can be different stains/orders)
- A slide can have many images (i.e., rescans)

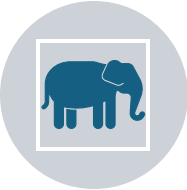
- No designated "best image" for a patient

Basic pathology asset structure

One to many to many to many to many to many



Overview of pathology image data



Large (easily > 2 GB each)

40x resolution (~0.25 microns/pixel)

Contains all tissue detected by the scanner (usually reliable)

Most images rescanned if QC issues present (e.g., out of focus)

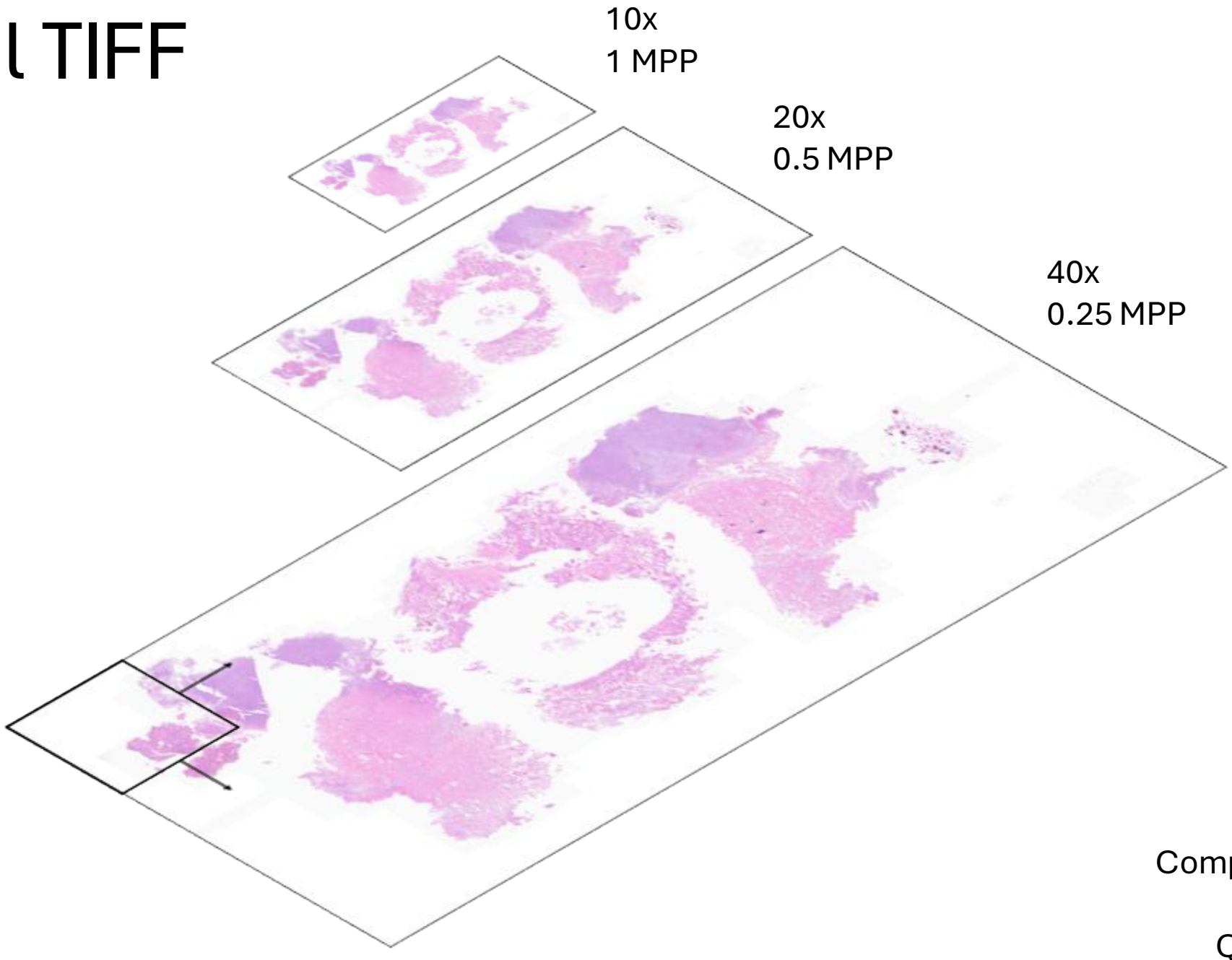
TIFF format

Open-source tools available

186 000 pixel

>x300k

Pyramidal TIFF Files



Compression:
JPEG
Quality 90

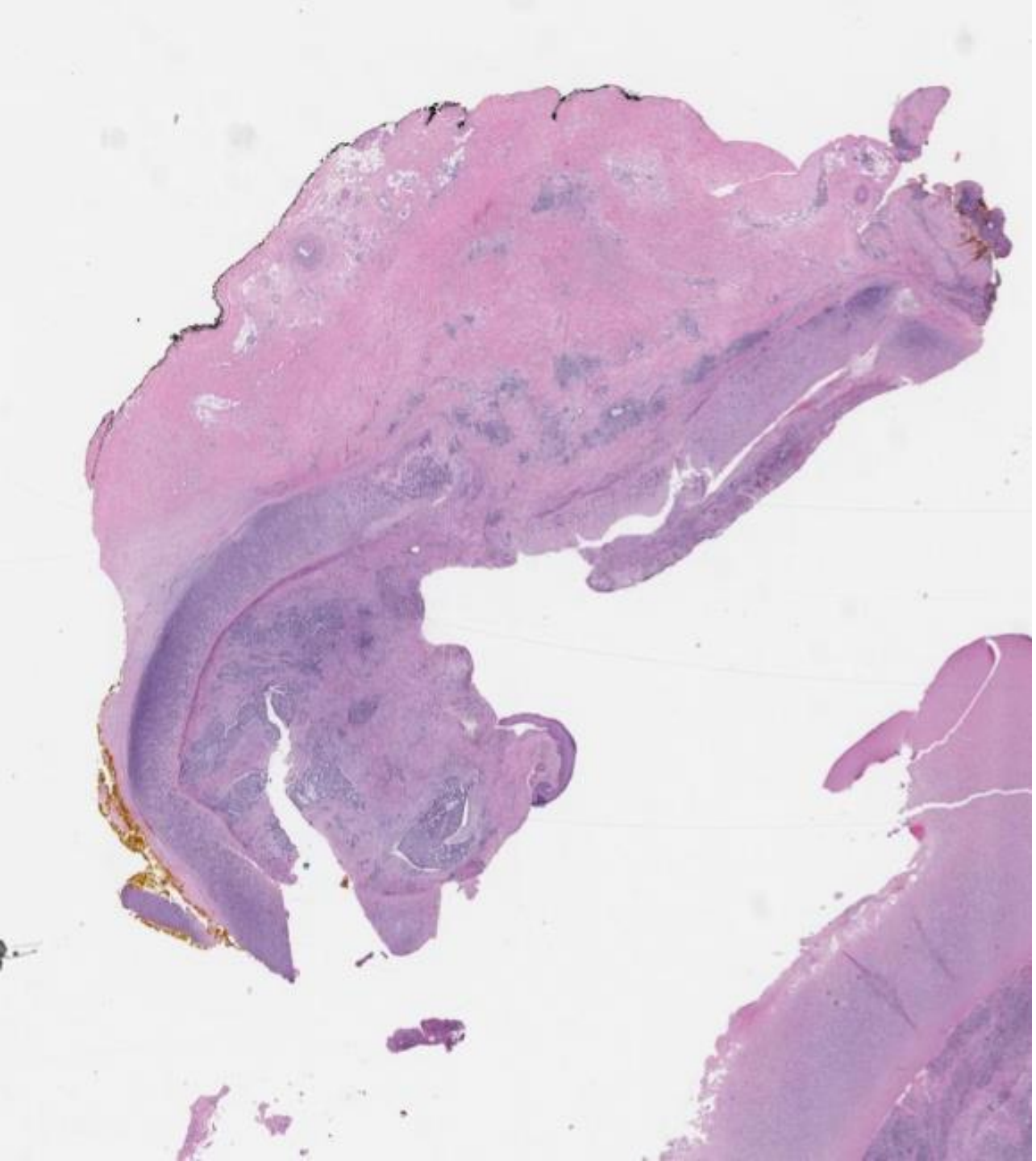
How to interact with images

- Files can’t be loaded into memory all at once
- Image has to be tiled or downsampled for analysis
- Need specialized packages to interact with the files (python)
  - `openslide`
  - `cucim`
- Needs specialized viewers to view the files
  - QuPath (local)
  - Digital Slide Archive (browser app with backend)

How to interact with images

```
import openslide
slide = openslide.OpenSlide('path/to/slide.tiff')
base_mpp = float(slide.properties[openslide.PROPERTY_NAME_MPP_X])
thumb = slide.get_thumbnail((1000,1000))
downsamples = slide.level_downsamples
tile40x = slide.read_region((14356, 24674), 0, (256,256)).convert('RGB')
tile20x = slide.read_region((14356, 24674), 1, (256,256)).convert('RGB')
```

https://openslide.org/api/python/



Quality Control

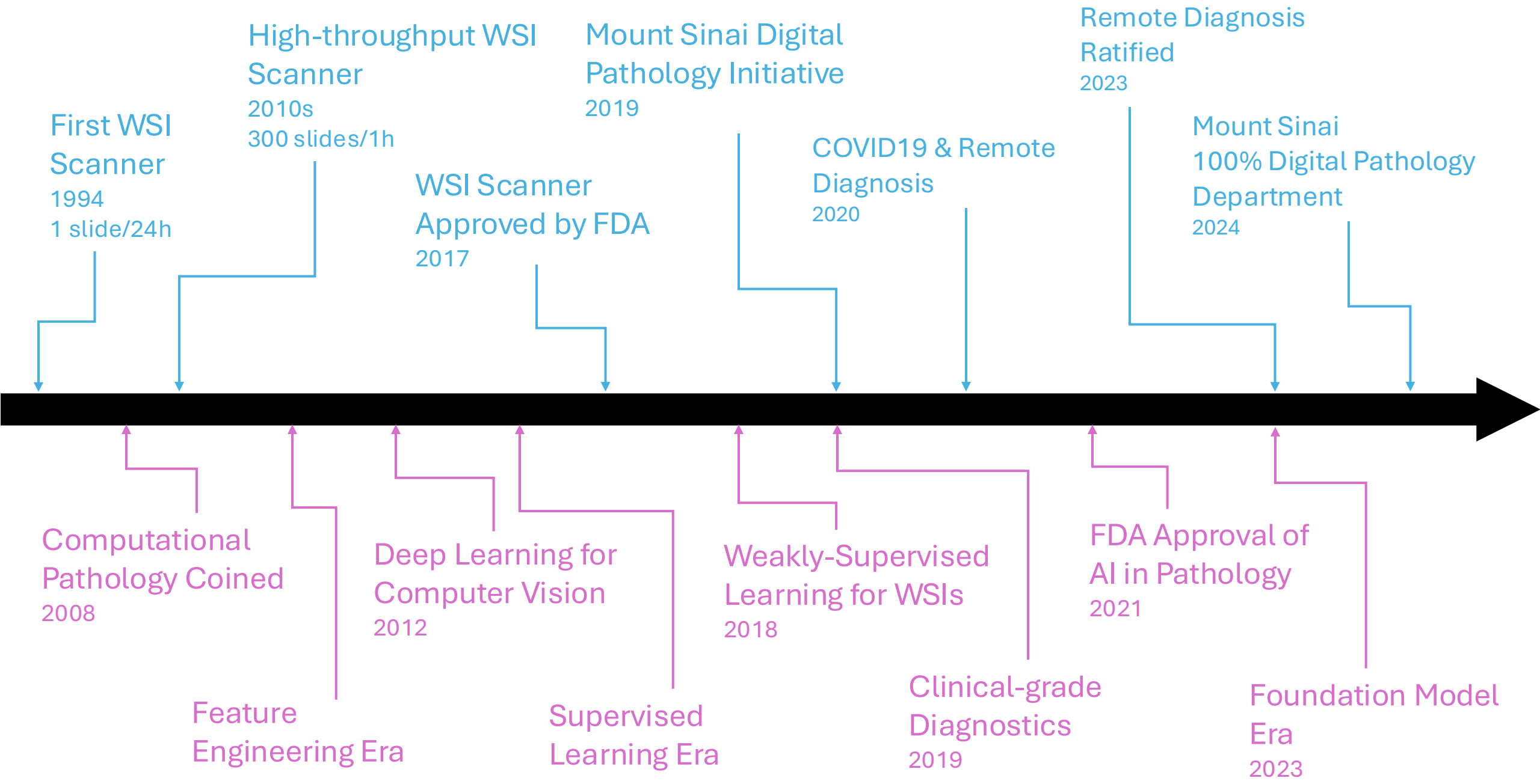
- Detect blurred regions
- Detect tissue folds
- Stain normalization

Diagnostic Reporting

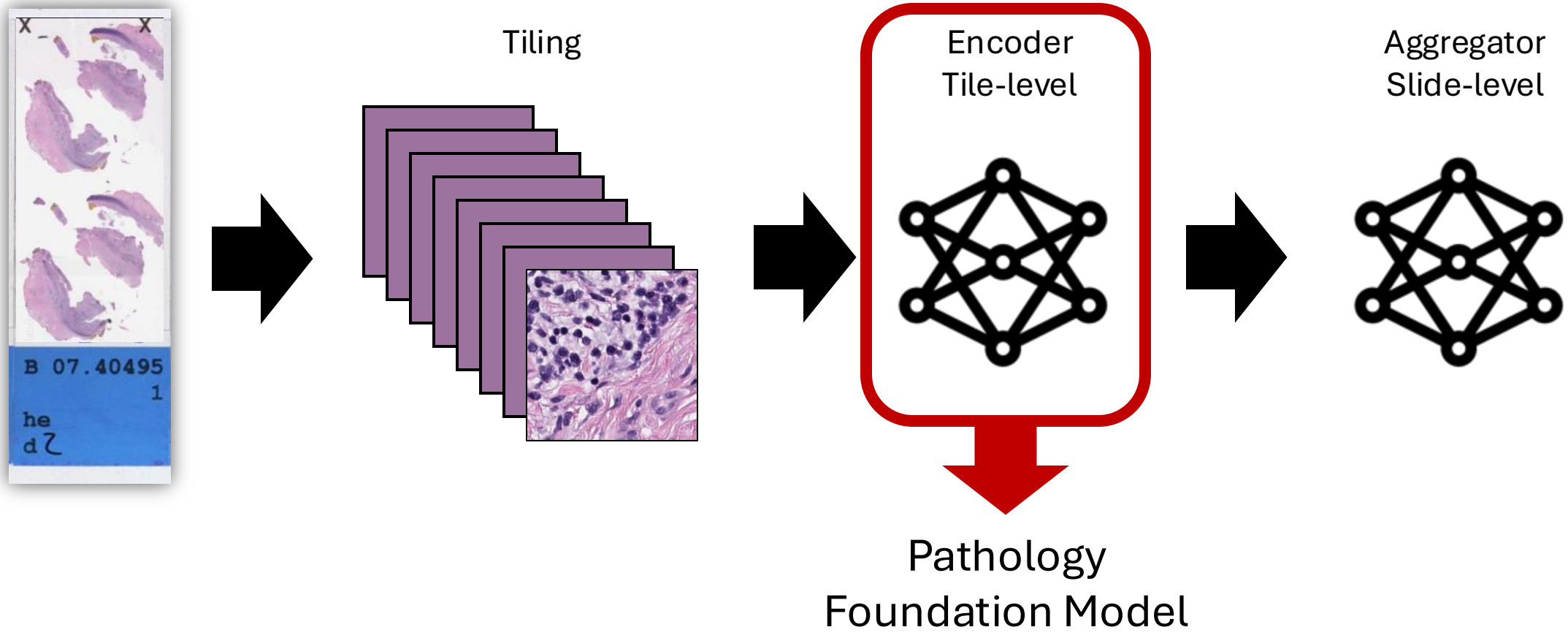
- Tumor detection
- Tumor segmentation
- Tumor staging
- Cell detection
- Slide retrieval

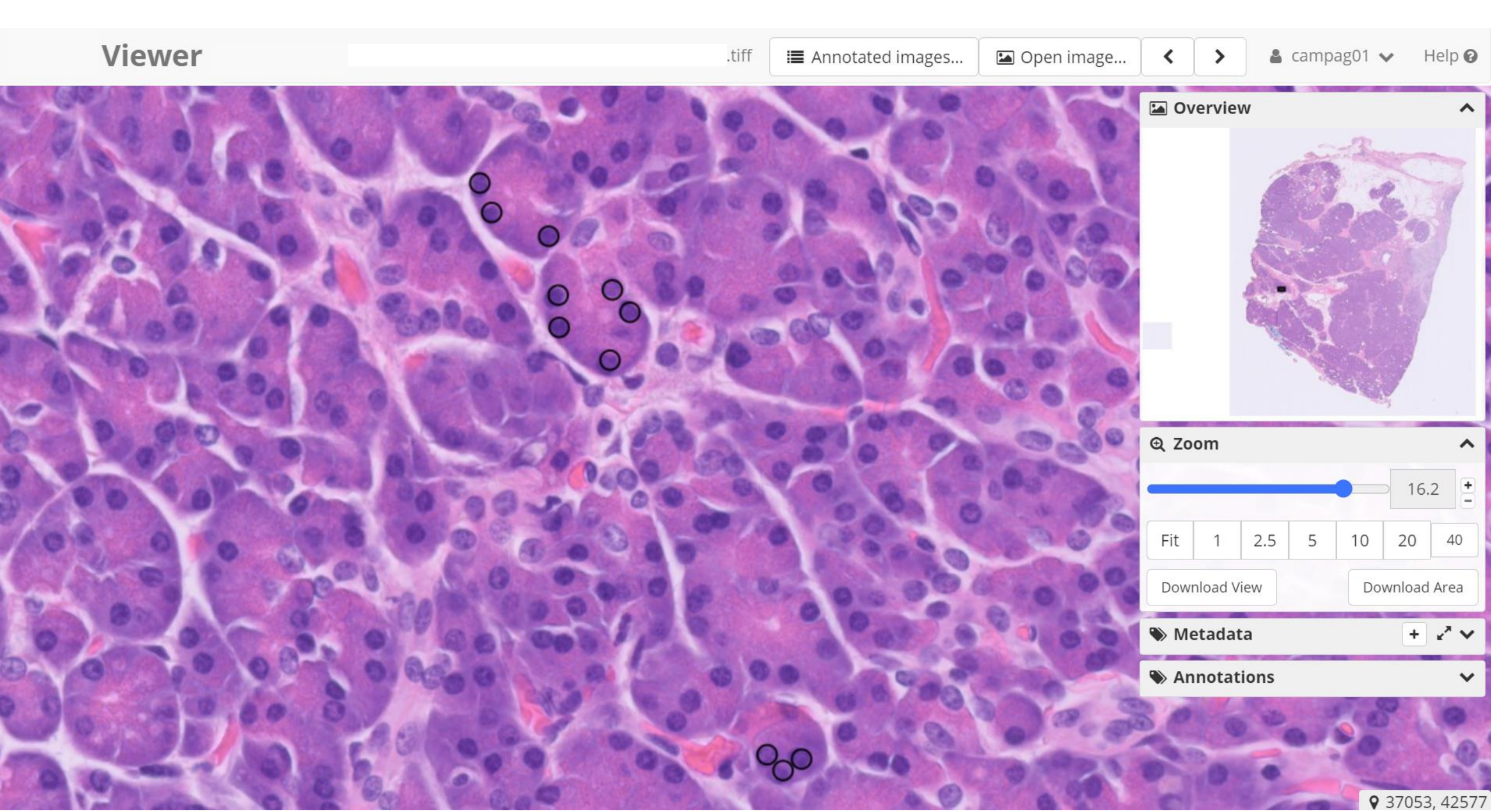
Prognostication

- Biomarker prediction
- Response prediction
- Survival analysis

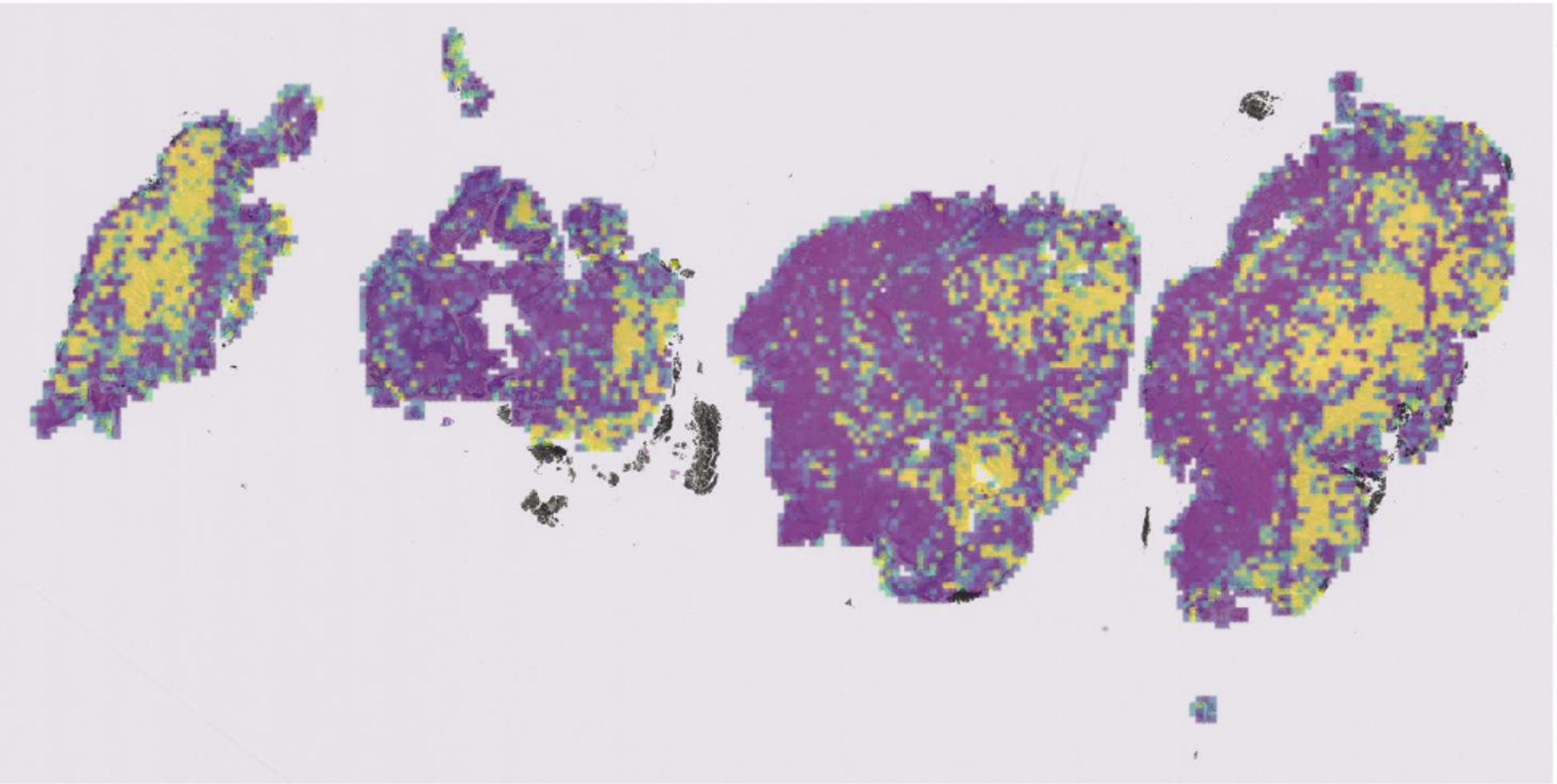

First WSI Scanner
1994
1 slide/24h
High-throughput WSI Scanner
2010s
300 slides/1h
WSI Scanner Approved by FDA
2017
Mount Sinai Digital Pathology Initiative
2019
COVID19 & Remote Diagnosis
2020
Remote Diagnosis Ratified
2023
Mount Sinai 100% Digital Pathology Department
2024
Computational Pathology Coined
2008
Feature Engineering Era
Deep Learning for Computer Vision
2012
Supervised Learning Era
Weakly-Supervised Learning for WSIs
2018
Clinical-grade Diagnostics
2019
FDA Approval of AI in Pathology
2021
Foundation Model Era
2023

Computational Pathology

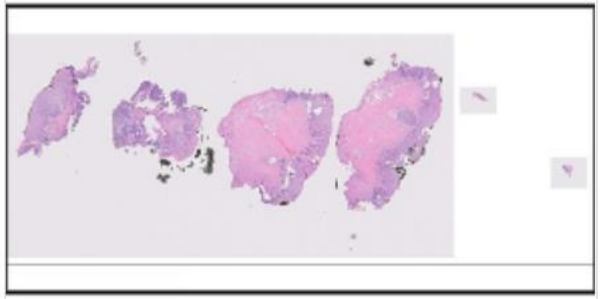



Viewer
.tiff
Annotated images...
Open image...
campag01
Help
Overview
Zoom
16.2
Fit
1
2.5
5
10
20
40
Download View
Download Area
Metadata
Annotations
37053, 42577

Viewer .tiff Annotated images... Open image... campag01 Help



Overview



Zoom

0.2

Fit 1 2.5 5 10 20 40

Download View Download Area

Metadata

Annotations

89341, -3065

Case Study

Goal: Obtain a diagnostic biopsy slide cohort of breast cancer patients with invasive ductal carcinoma and their receptor status.

Tasks:

- Identify the right cases
- Identify representative slide
- *Obtain diagnosis, filter by IDC*
- *Obtain hormone receptor status (ER/PR/HER2)*

Identify Case

- Manual review
- Rely on pathology Sign-out knowledge
  - Search breast cases where report contains “biopsy” & “ductal carcinoma”
  - Cons
    - Will capture DCIS
    - May lose cases where biopsy is not explicitly mentioned
    - Typos
    - Free text fallacy

Identify Case

- Rely on non-pathology data sources
  - EPIC, cancer registry, etc.
  - Match with collection dates of specimens
  - Cons
    - Will not capture biopsies done in outpatient clinic or consults
    - Potential inconsistent MRNs

Identify Slide

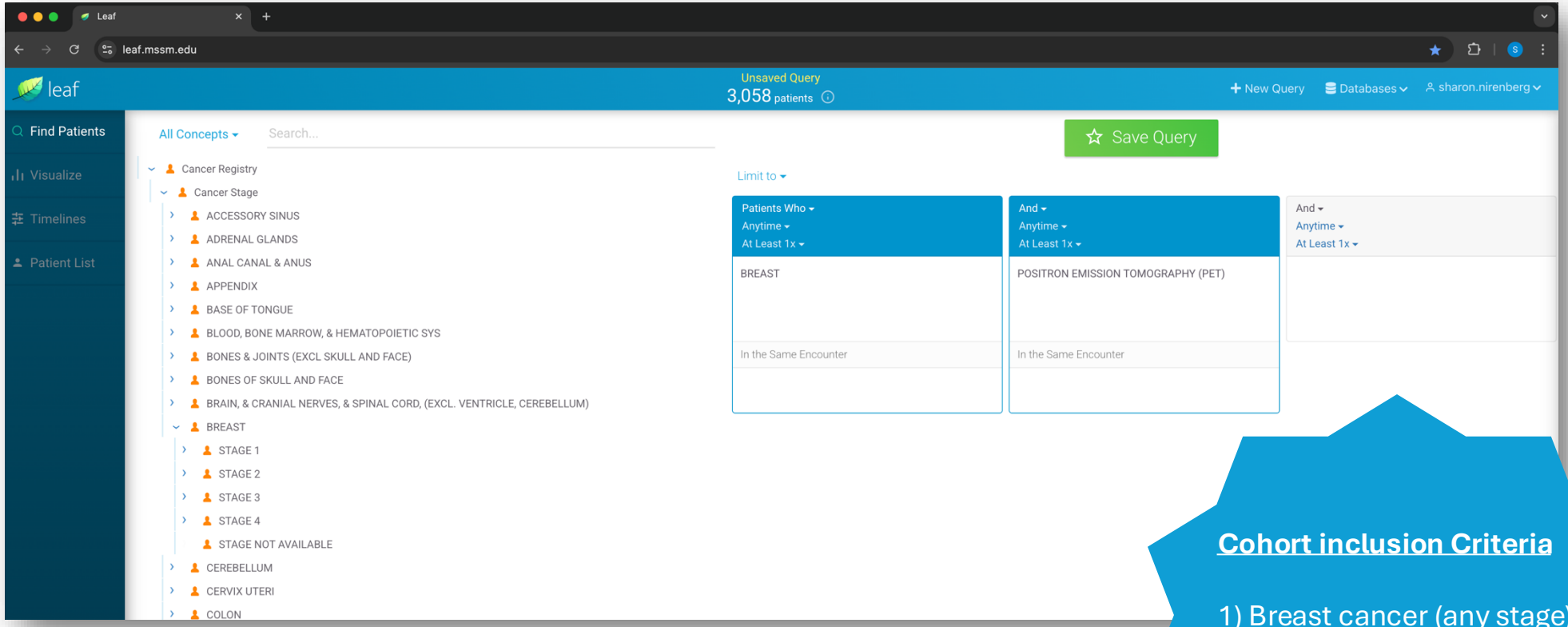
Preface: pathology report gives specimen level information, not slide level

- Manual slide review
- Rely on specialty specific knowledge
  - For breast cancer, only tumor slides receive ER/PR/HER2 testing
- Rely on synoptic reporting
  - Newer synoptic reports include “best” tumor block
  - Cons
    - Implemented only recently
    - No synoptic for biopsies
- AI – vision model

TLDR

Pathologist collaboration is essential!

Identify Cohorts with Digital Pathology Slides on Minerva Using Leaf

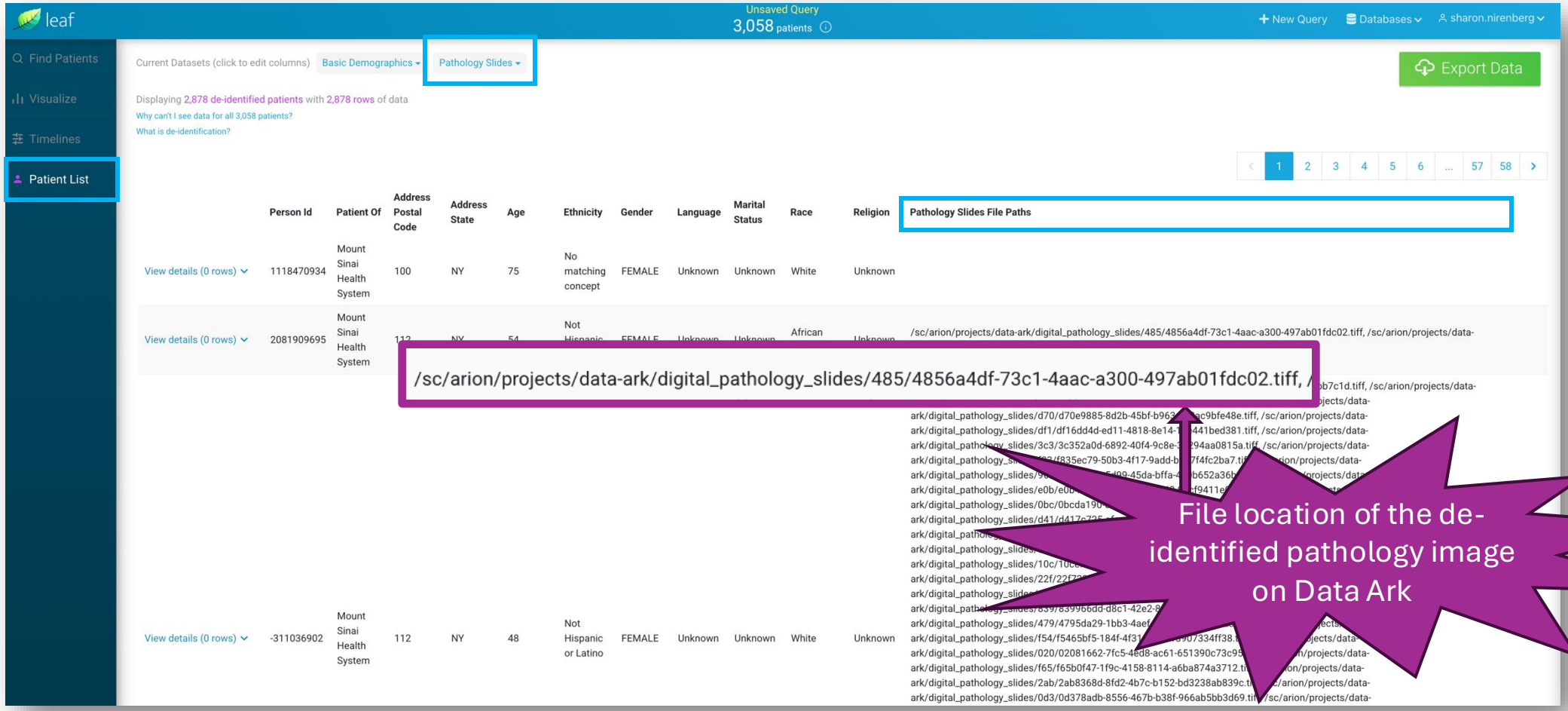


Cohort inclusion Criteria

1) Breast cancer (any stage)

2) PET Scan in IRW 2.0

Using Leaf to Locate Digital Pathology Images on Data Ark



Learn how to access the Digital Pathology Images on **Minerva**

Data Ark Training Session

Thursday, March 27, 1-2PM

Icahn School of Medicine building (1425 Madison Ave) Room L3-36 and Zoom

Pathologist collaboration is essential!

- Choosing a single representative image is challenging
  - Generally requires subspecialty expertise by a pathologist
  - Context-dependent: a selected image for one project is not necessarily the best image for another project
- They can also make sure that the findings reported in the study make sense in the context of pathology

- For assistance in finding a pathologist interested in research, email the director of the Biorepository and Pathology CORE, Rachel Brody, PhD (rachel.brody@mountsinai.org)