NIH NATIONAL CANCER INSTITUTE

FDA U.S. FOOD & DRUG ADMINISTRATION

ANALYSIS OF A PILOT STUDY COLLECTING PATHOLOGIST ANNOTATIONS FOR VALIDATING MACHINE LEARNING ALGORITHMS

Katherine Elfer, PhD, MPH

NCI-FDA Interagency Oncology Task Force Cancer Prevention Fellow

Division of Imaging, Diagnostics, Software Reliability

Office of Science and Engineering Laboratories

Center for Devices and Radiological Health

U.S. Food and Drug Administration

22 February 2023

Disclosures

In the past 12 months, I have not had any significant financial interest or other relationship with the manufacturers of the products or providers of the services that will be discussed in my presentation.

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

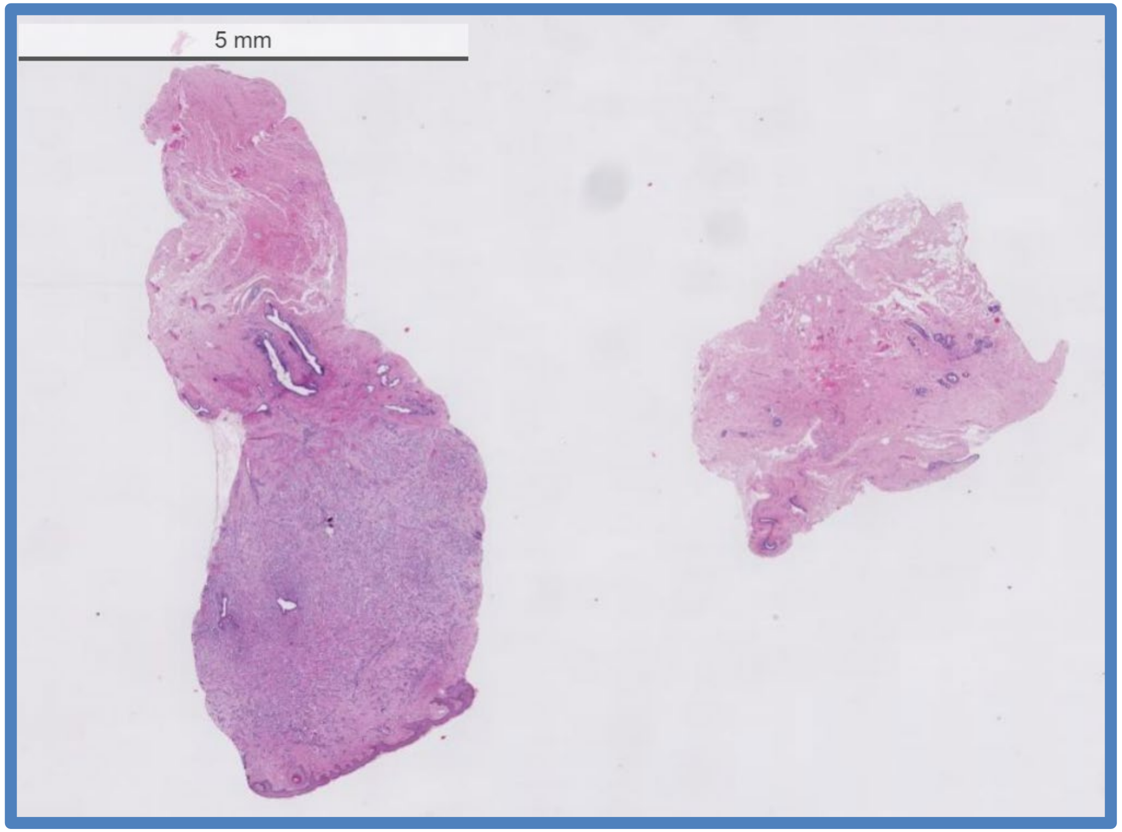
Outline

- Overview:
  - High Throughput Truthing (HTT) Project
- Study 1:
  - HTT Pilot Study
- Study 2:
  - eeDAP Registration Accuracy
- Future Work
  - HTT Pivotal Study

Pain Point:
Validation of Digital Pathology Technology

Recent surge in Digital Pathology technologies:
- Whole Slide Image Scanners can create images on the order of Gpx and up to 120GB in size.
- Algorithms detect and diagnose disease

Evaluation by Algorithm:
- Reproducible
- Reduce burden on pathologists
- Increase Speed & Accuracy



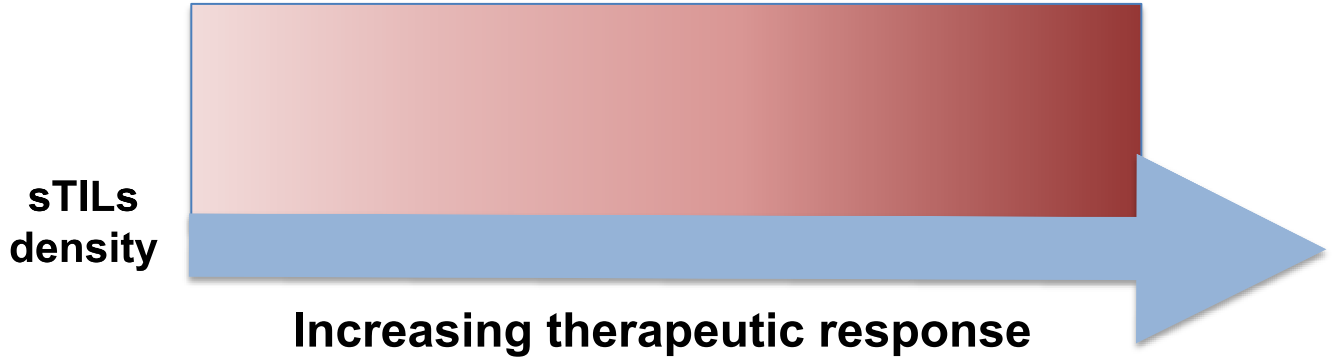
H&E Breast Cancer Whole Slide Image (WSI)

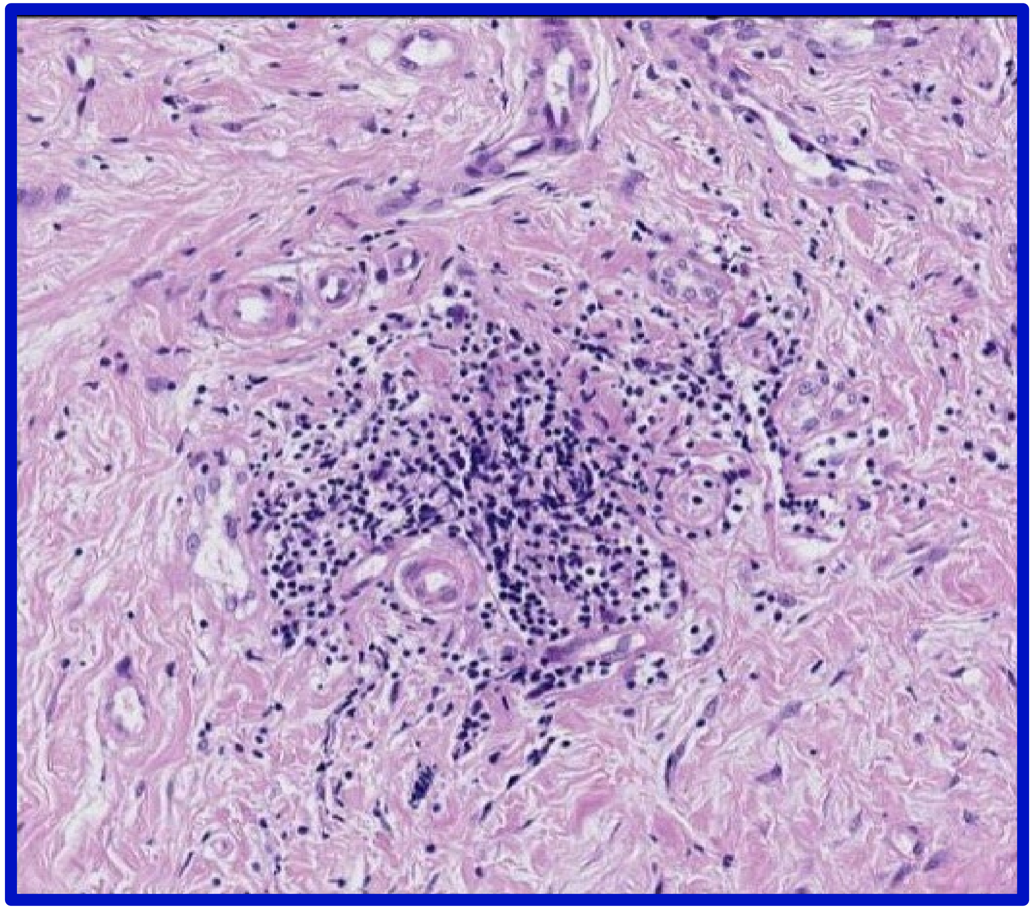
Use Case:

Stromal Tumor Infiltrating Lymphocytes (sTILs)

Clinical application:

stromal Tumor Infiltrating Lymphocytes (TILs) density are a **quantitative**, prognostic biomarker.

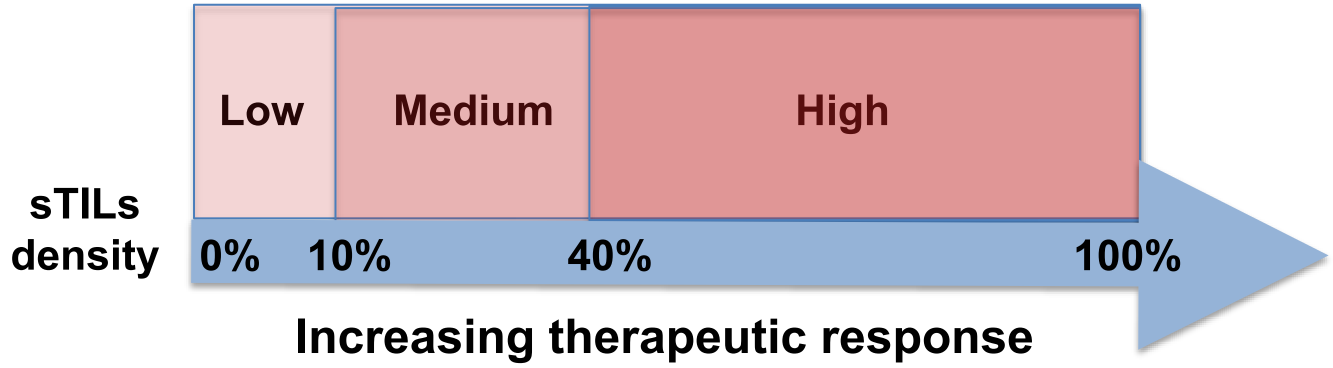


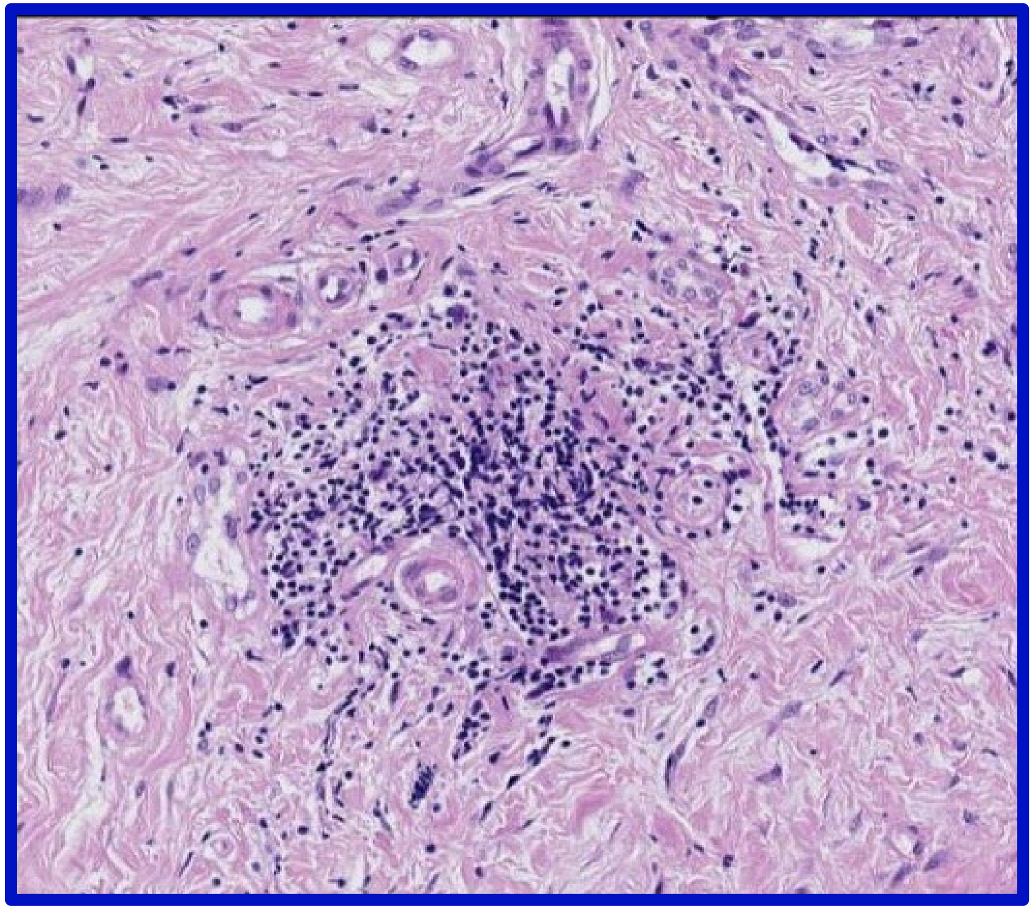


sTILs in Breast Cancer

Use Case:
Stromal Tumor Infiltrating Lymphocytes (sTILs)

Clinical application:
stromal Tumor Infiltrating Lymphocytes (TILs) density are a **quantitative**, prognostic biomarker.

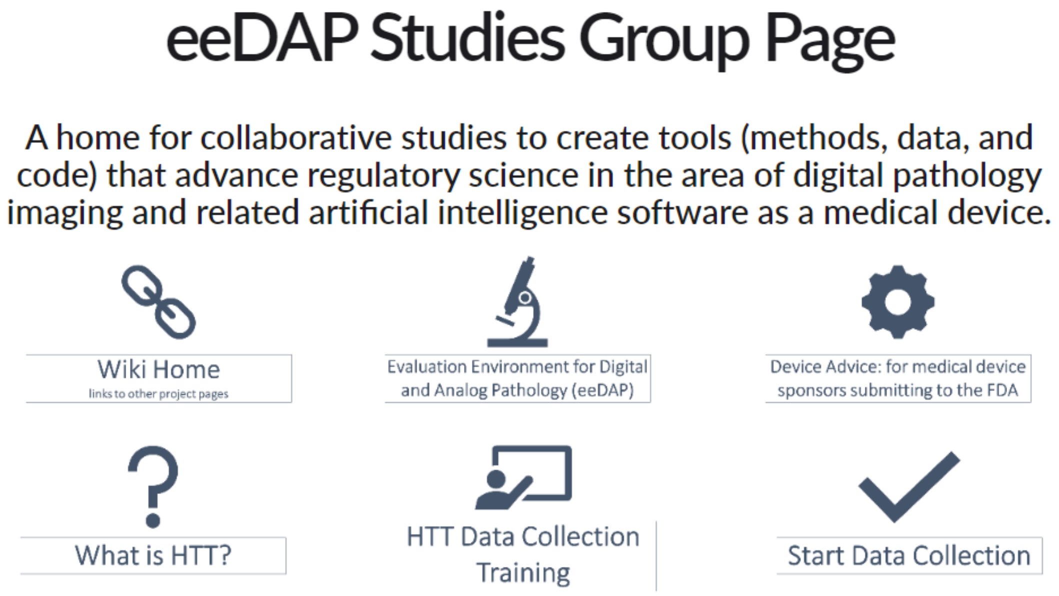




sTILs in Breast Cancer

High-Throughput Truthing (HTT) Project

- An international volunteer collaboration
- **Goal:** Create a dataset of Triple Negative Breast Cancer (TNBC) slides & images with pathologist annotations of a quantitative biomarker (sTILs)
  - To be used for the validation of AI/ML and computational pathology models
  - Pursuing qualification as a medical device development tool (MDDT)



https://ncihub.org/groups/eedapstudies

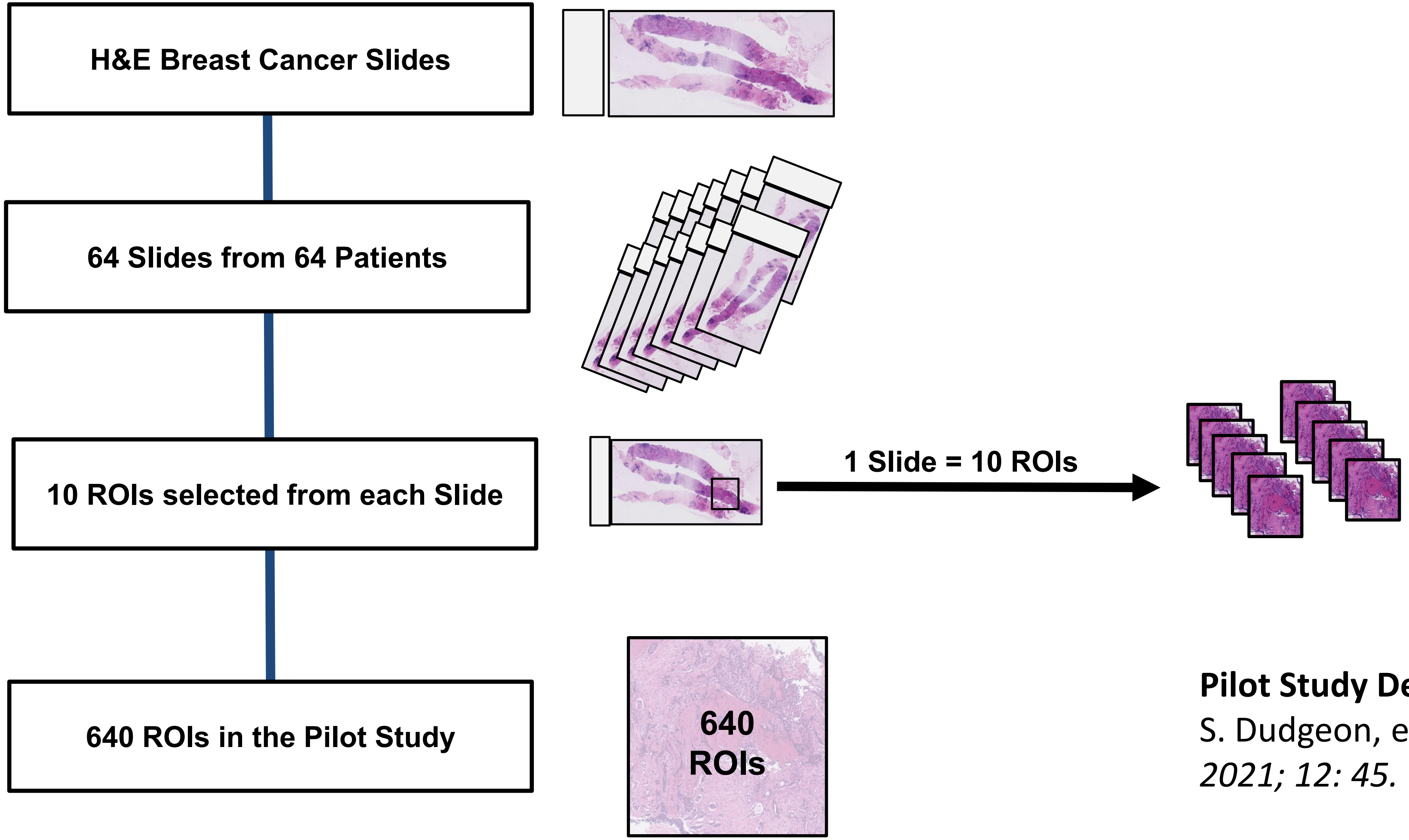
https://www.zotero.org/groups/4384613/eedap_studies_presentations_publications_and_studies/collections/9ABM9D8M

Study 1

STUDY GOAL:
Exploratory Analysis of HTT Pilot Data

Publication: K. Elfer, et. Al/ J. Med. Imag. 9(4) 047501, 2022.

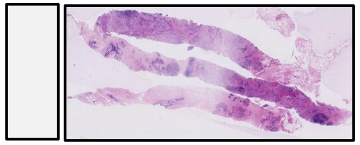
HTT Pilot Study Materials: Feb 2020 – May 2021



Pilot Study Description:
S. Dudgeon, et. al., *J Pathol Inform. 2021; 12: 45.*

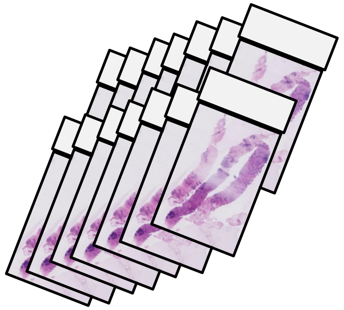
HTT Pilot Study Materials: Feb 2020 – May 2021

H&E Breast Cancer Slides

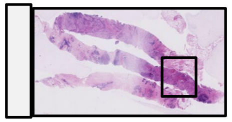


3 Annotations Platforms

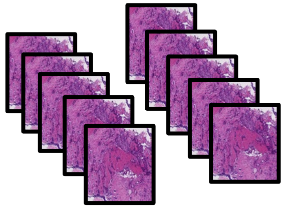
64 Slides from 64 Patients



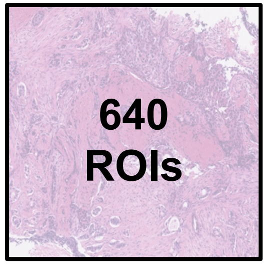
10 ROIs selected from each Slide



1 Slide = 10 ROIs



640 ROIs in the Pilot Study

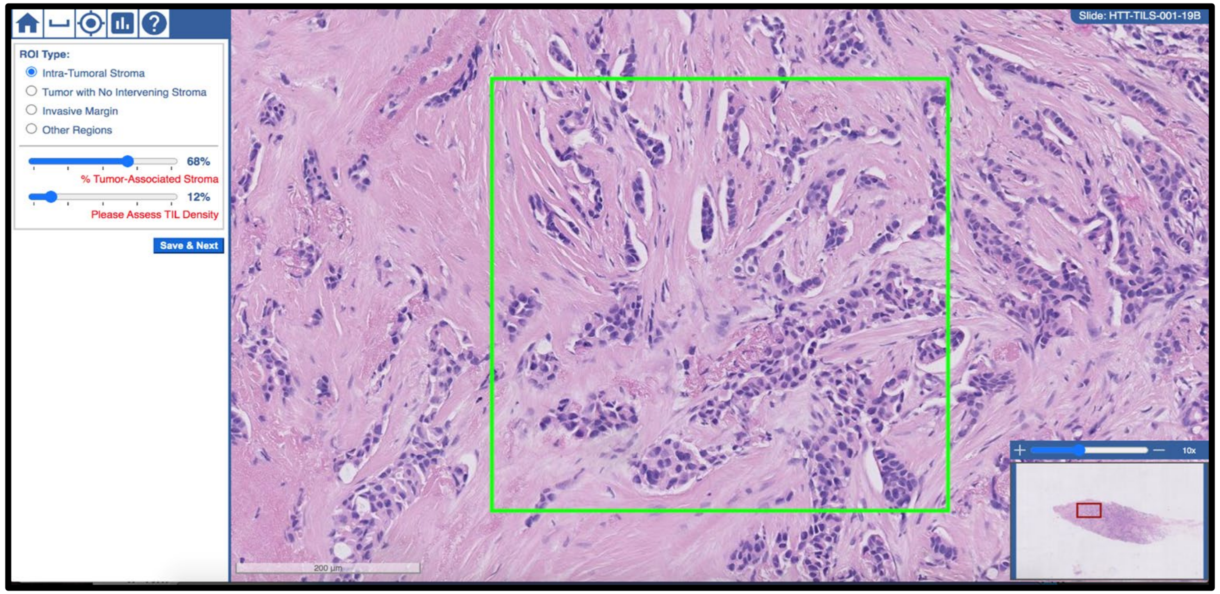


Pilot Study Description:
S. Dudgeon, et. al., *J Pathol Inform. 2021; 12: 45.*

HTT Data Collection Platforms: Digital Modalities

FDA

Platform 1: caMicroscope

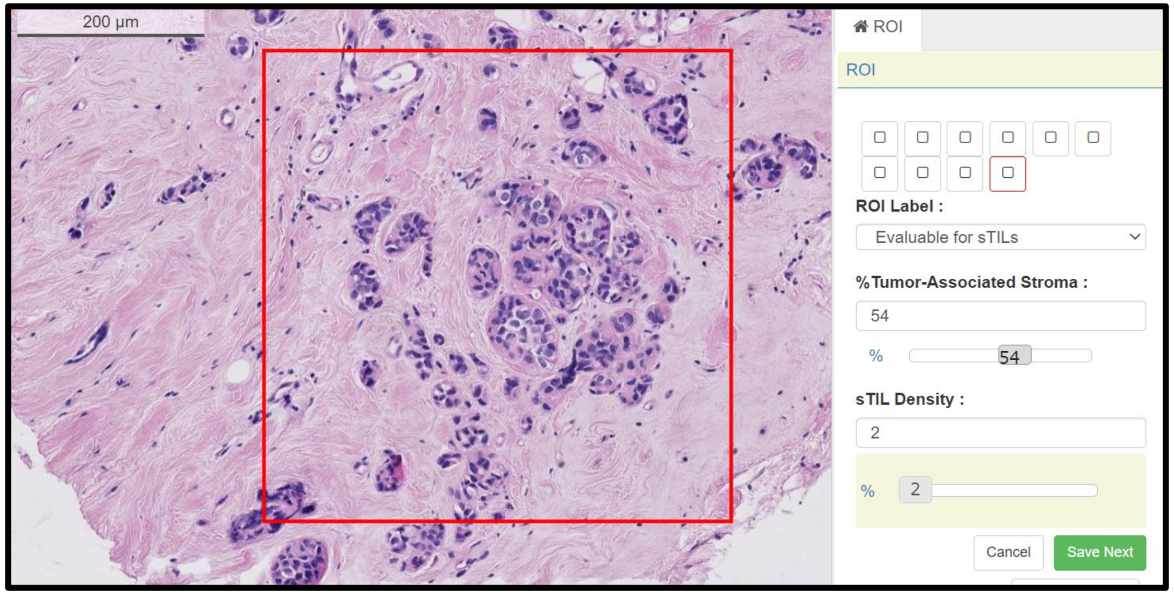


caMicroscope is an open-source platform hosted on precisionFDA

precisionFDA

https://github.com/camicroscope/caMicroscope

Platform 2: PathPresenter



PathPresenter is a licensed commercial entity volunteering server space and platform customization

PathPresenter®

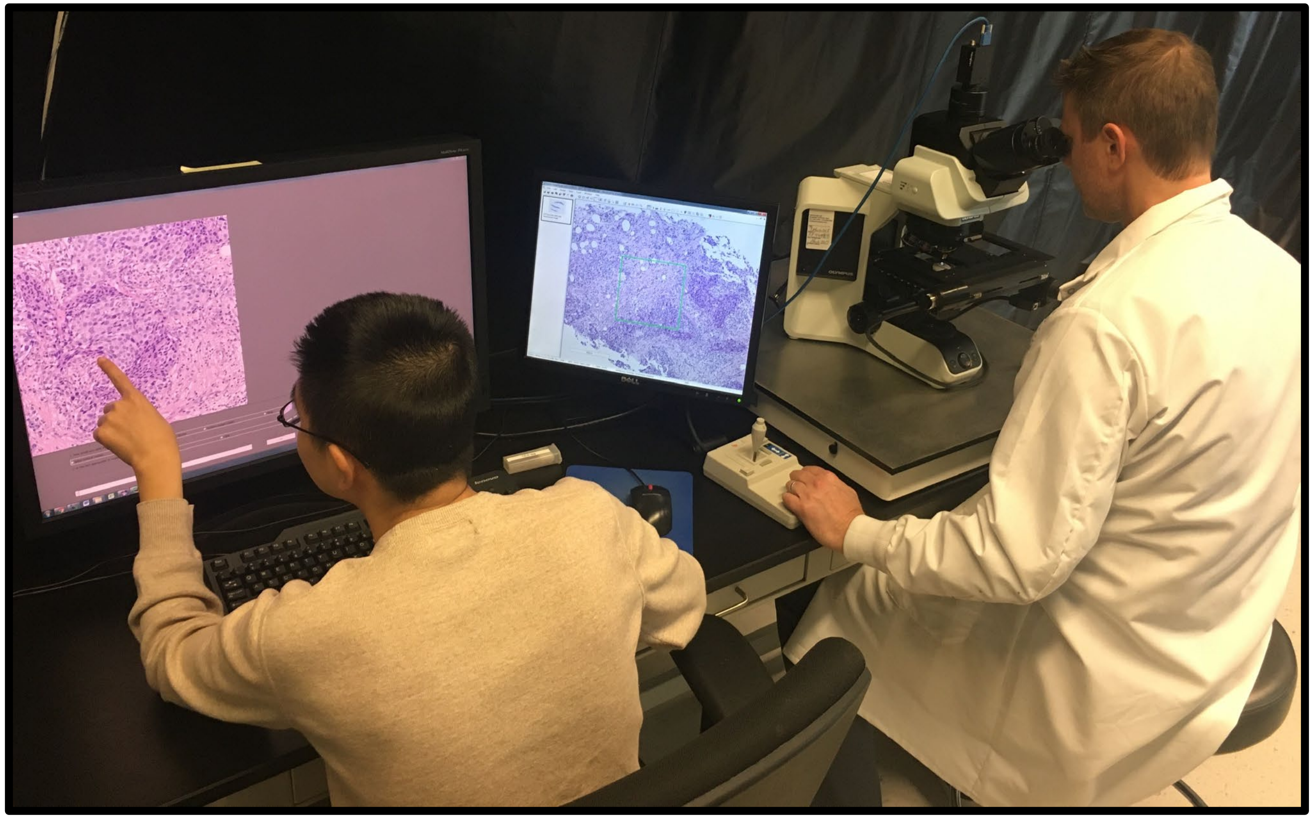
https://htt.pathpresenter.net

HTT Data Collection Platforms: Microscope Modality

Microscope: eeDAP

eeDAP evaluation environment for Digital and Analogue Pathology

previously presented: Qi Gong, et. al. 2018, B.D. Gallas, et. al., 2014



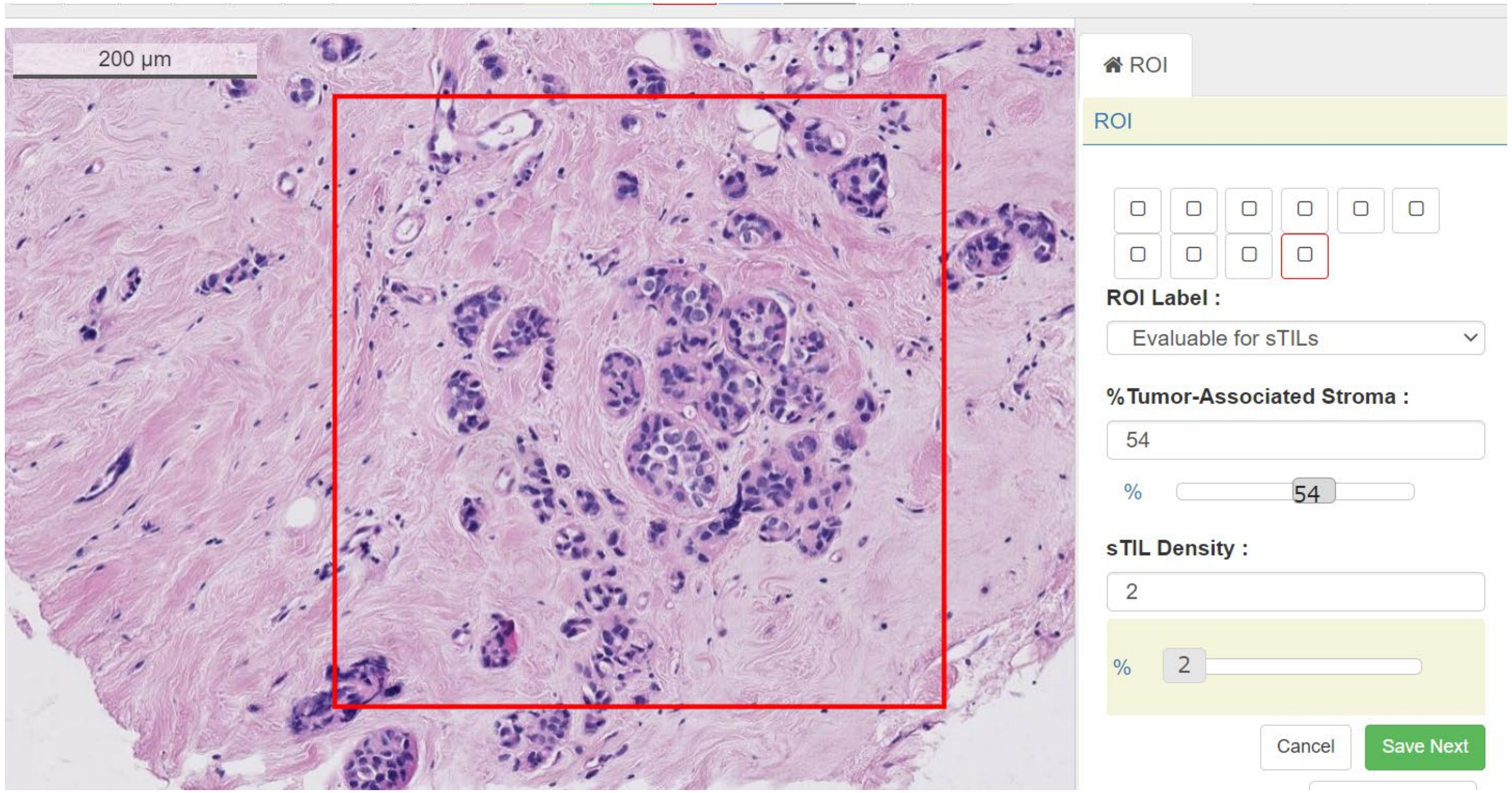
Reference Standard Technology

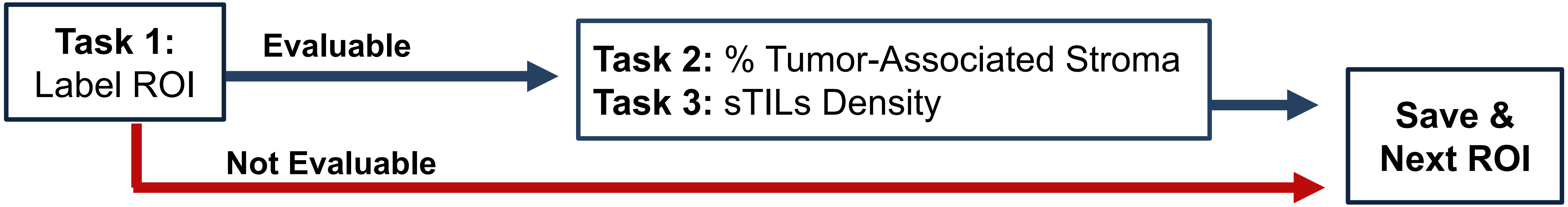
Enables exact studies on glass slides & WSIs

https://github.com/DIDSR/eeDAP

HTT Annotations

htt.pathpresenter.net





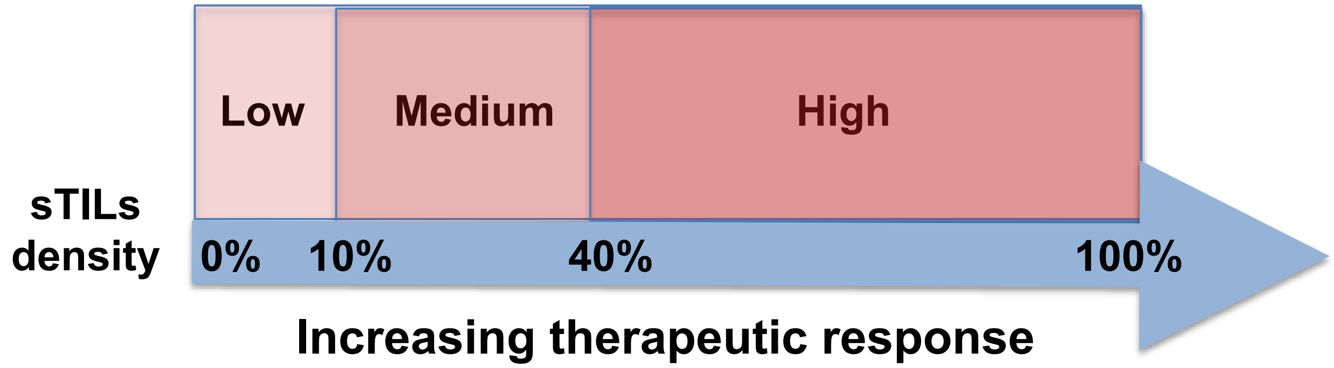
HTT Pilot Study Results: All Platforms

Platform	Readers	Observations
eeDAP	7	440
PathPresenter	10	1833
caMicroscope	20	5100
All	37	7373

Publication:
K. Elfer, et. al., J. Med. Imag. 9(4) 047501, 2022.

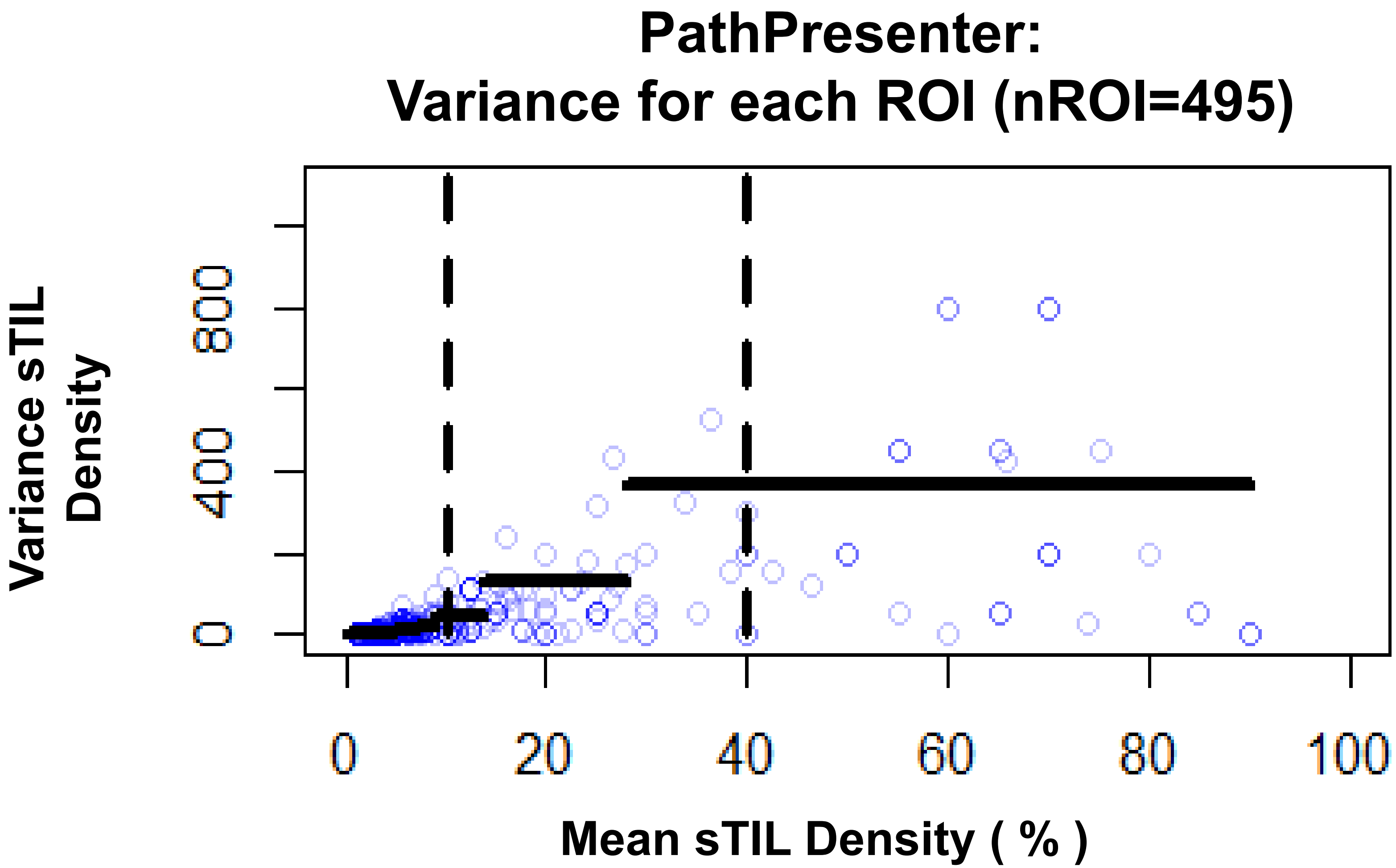
HTT Pilot Study Results: All Platforms

Platform	Readers	Observations	Pre-Selected Density Bins		
			Low: 0-10%	Med: 11-40%	High: 41-100%
eeDAP	7	440	323	21	10
PathPresenter	10	1833	1,127	211	79
caMicroscope	20	5100	3,042	912	233
All	37	7373	4,492	1,144	322



Publication:
K. Elfer, et. al., J. Med. Imag. 9(4) 047501, 2022.

HTT Pilot Study Results



Each blue circle represents one ROI (n=495) with at least two sTIL density estimates.

Pathologist variance depends on the ROI and increases with the mean.

Publication:

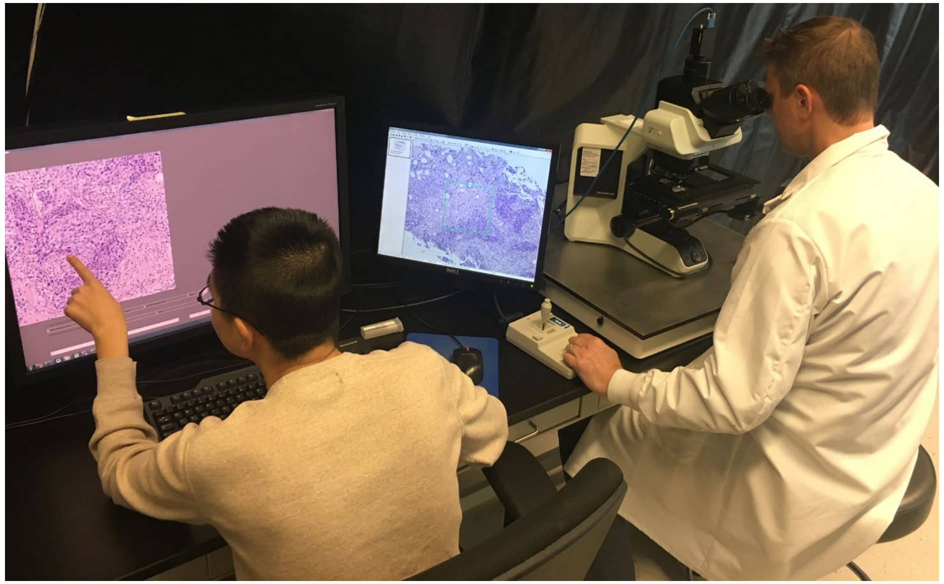
K. Elfer, et. al., J. Med. Imag. 9(4) 047501, 2022.

Take-Aways from the HTT Pilot Study

Accomplishments:

- Training methods to reduce variability:
  - Expert Panel: Pearls and Pitfalls of sTILs Assessment
    - V. Garcia, et. al., Cancers (Basel). 2022 May 17;14(10):2467.
  - Development of a medical training course in sTILs Assessment
  - Development of interactive training materials

- Improvement of data-collection platforms:
  - Improvement of digital platforms
  - Improved hardware for eeDAP



Study 2

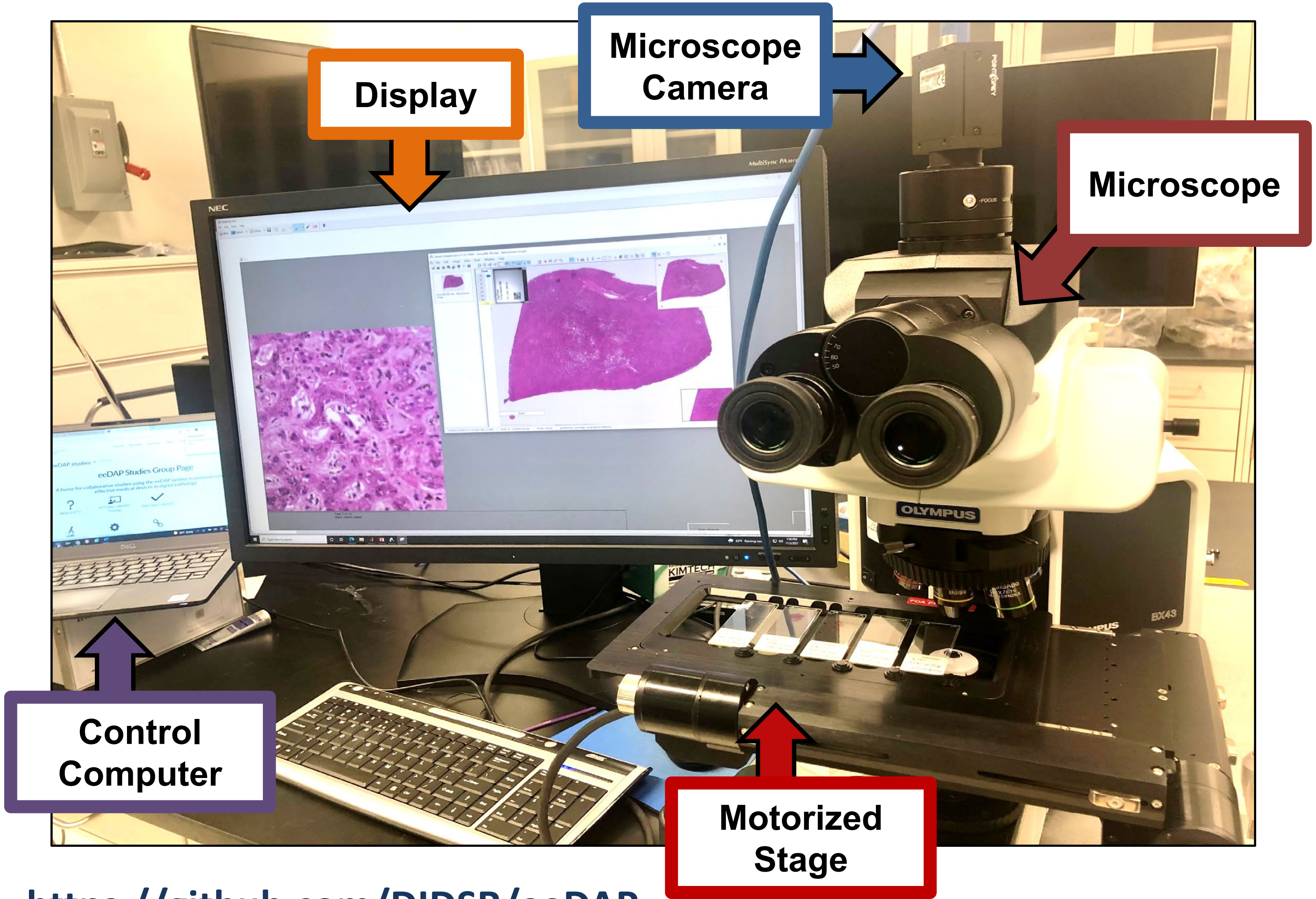
STUDY GOAL:

Registration Accuracy of eeDAP

Continuation of :

Qi Gong, et. al.,SPIE Med Imag, Proceedings, 2018

eeDAP: evaluation environment for Digital and Analogue Pathology



https://github.com/DIDSR/eeDAP

eeDAP drives the microscope stage to an x-y target location corresponding to the WSI's spatial coordinates

Uses a set of custom input files to define tasks for annotators.

B.D. Gallas, et. al., Proc. SPIE 9037, Medical Imaging 2014.

B. D. Gallas, et. Al,J Med Imaging 2014 Oct;1(3):037501.

eeDAP Registration Methods

FDA

4 Registration Methods:

1) **Global Registration**
   – defines relationship between WSI and glass slide
2) **Local: Automatic**
3) **Local: Fast** - utilizes padding to extend boundary of image
4) **Local: Best** - best fit; no-padding

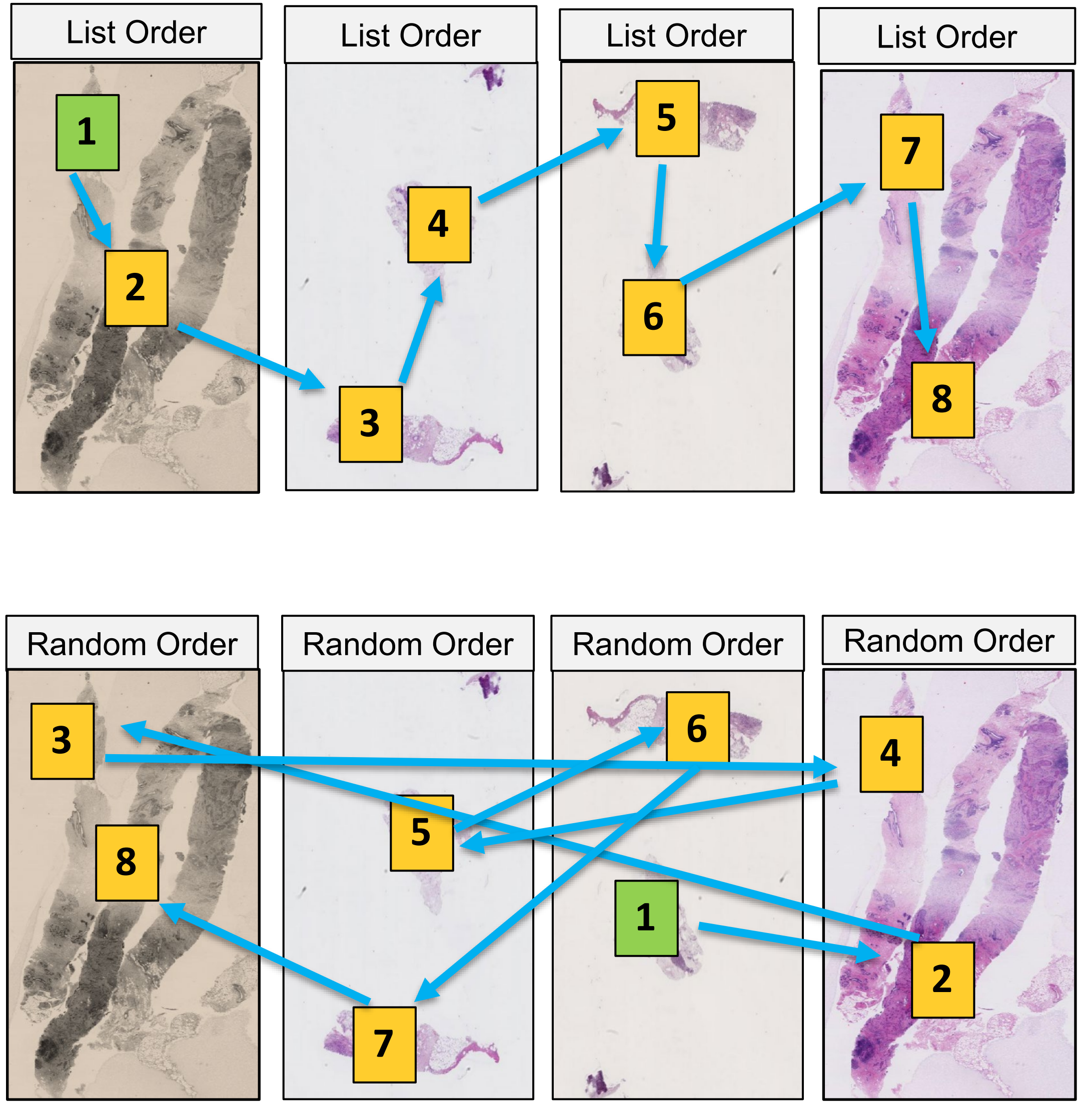
eeDAP Registration Methods

4 Registration Methods:

1) **Global Registration**
   – defines relationship between WSI and glass slide
2) **Local: Automatic**
3) **Local: Fast** - utilizes padding to extend boundary of image
4) **Local: Best** - best fit; no-padding

2 Orders of Operation:

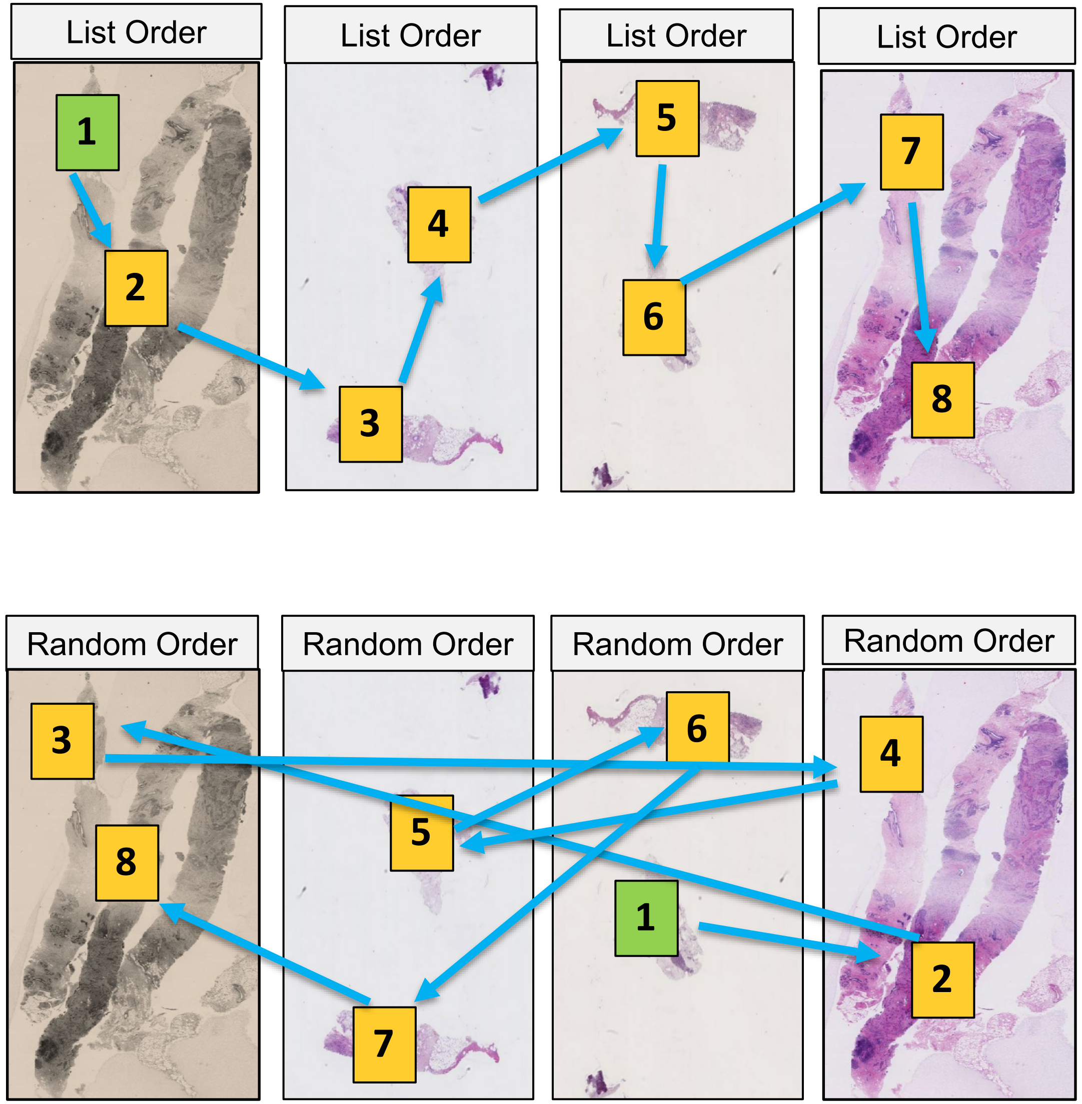
1) List Order
2) Random Order



eeDAP Registration Methods

4 Registration Methods:

1) **Global Registration**
   – defines relationship between WSI and glass slide
2) **Local: Automatic**
3) **Local: Fast** - utilizes padding to extend boundary of image
4) **Local: Best** - best fit; no-padding



2 Orders of Operation:

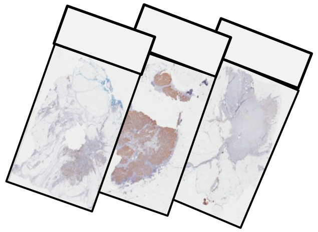
1) List Order
2) Random Order

8 Study Conditions

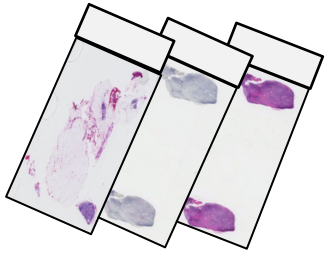
eeDAP Registration Accuracy Study Materials

6 Slides:
Human and Canine
H&E, PHH3, ER, HER2

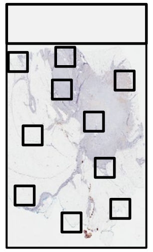
3 Slides: 20X



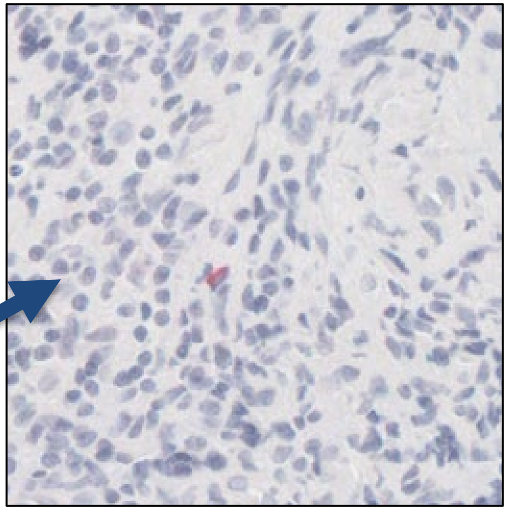
3 Slides: 40X



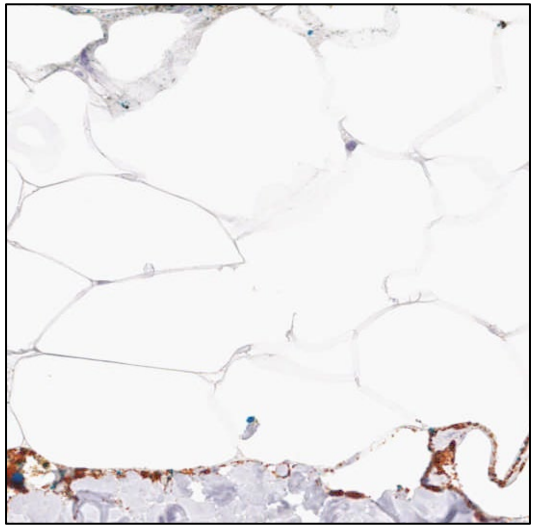
10 ROIs from each Slide
(500 μm x 500 μm)



1 Slide = 10 ROIs

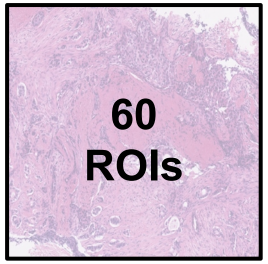


Canine Oral Melanoma (PHH3)



Human Breast (HER2)

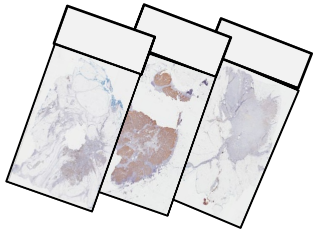
60 ROIs in the Study



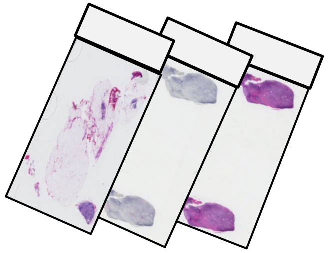
eeDAP Registration Accuracy Study Materials

6 Slides:
Human and Canine
H&E, PHH3, ER, HER2

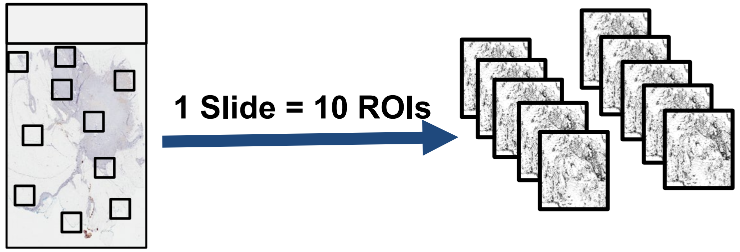
3 Slides: 20X



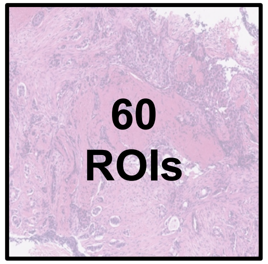
3 Slides: 40X



10 ROIs from each Slide
(500 μm x 500 μm)



60 ROIs in the Study



2018: 2 Operators

Camera: PointGrey Grasshopper 3

0.3 MP ; pixel size = 7.5 μm

Stage: Ludl BioPrecision 2

Velocity = 33 mm/s ; accuracy = 6 μm

Qi Gong, et. al.,SPIE Med Imag, Proceedings, 2018

2022: 1 Operator

Camera: PointGrey Grasshopper 3

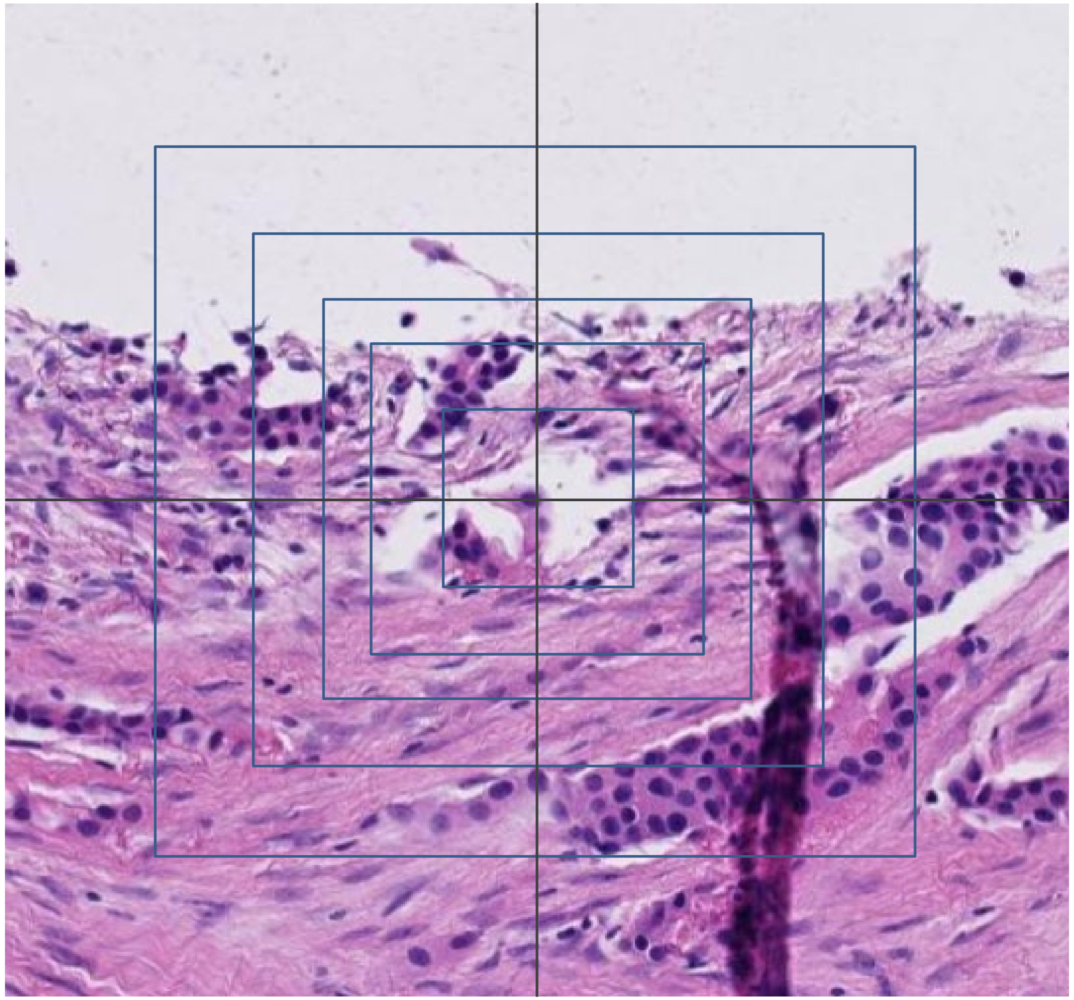
2.3 MP ; pixel size = 5.36 μm

Stage: Thorlabs High Speed X-Y

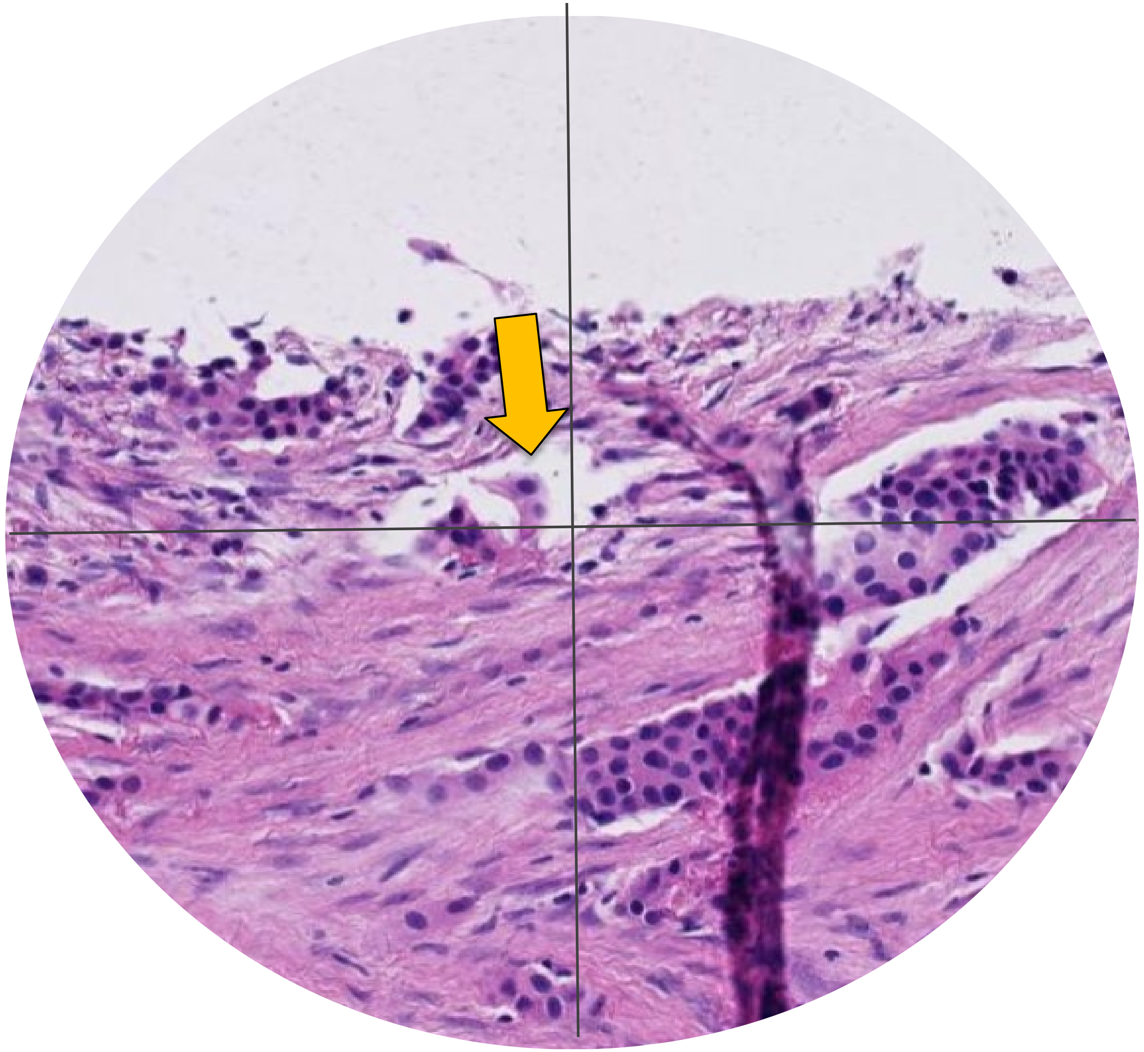
Velocity = 250 mm/s ; accuracy = 0.25 μm

Elfer, et. al.,SPIE Med Imag, Proceedings, 2023

Measuring eeDAP Registration Accuracy



A. ROI with virtual bounding box

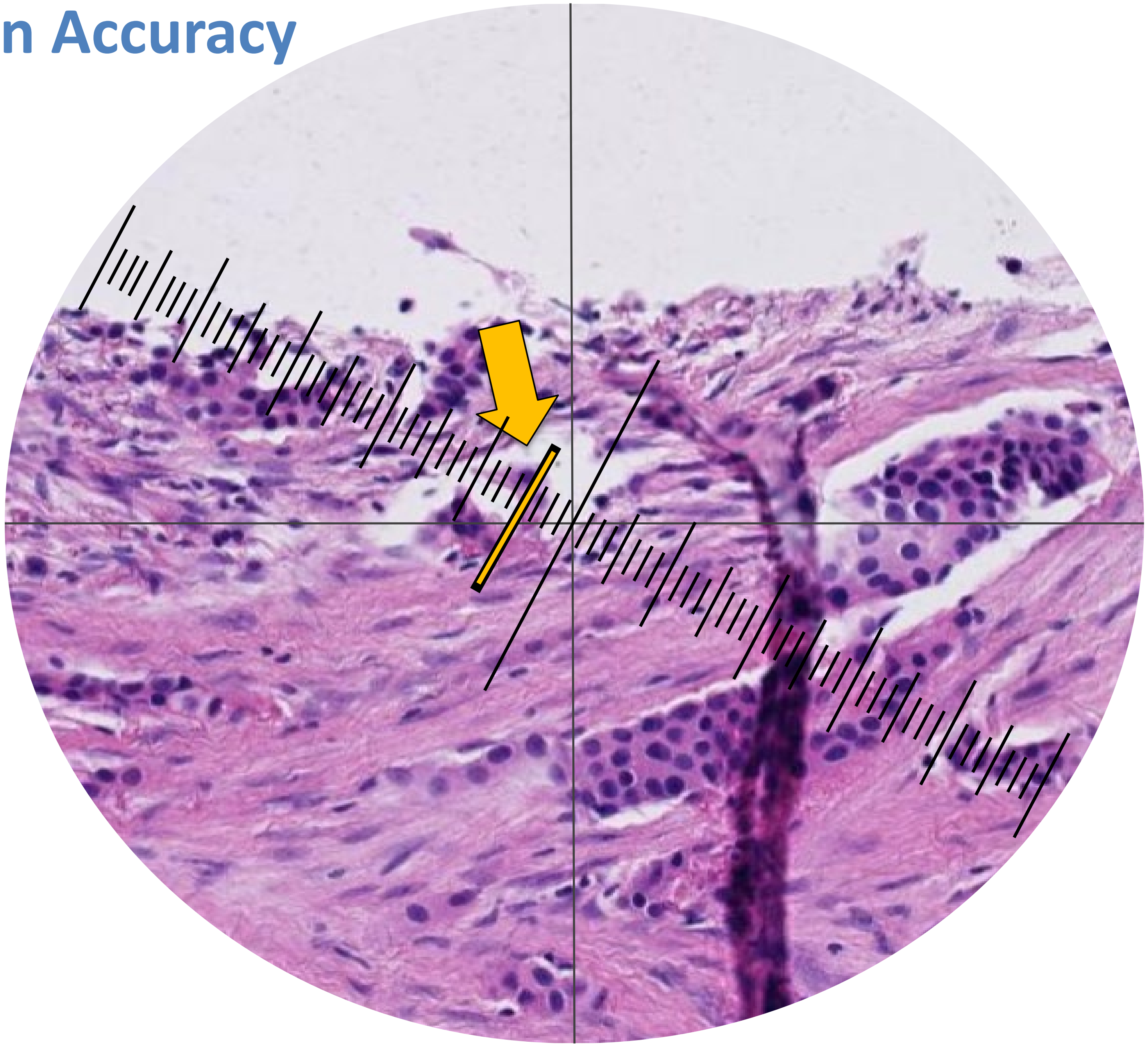


B. FOV with reticle ruler

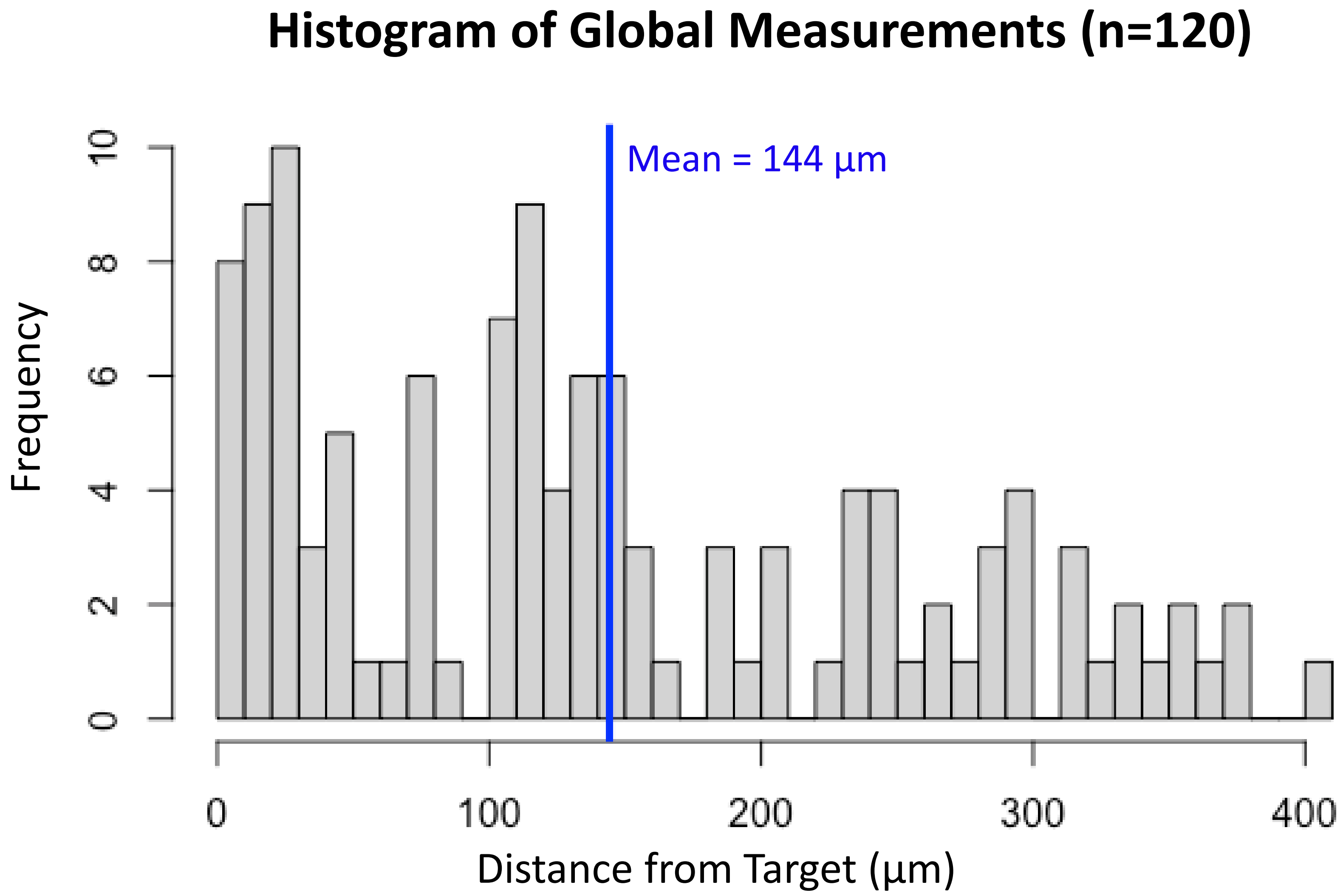
Measuring eeDAP Registration Accuracy

Reticle Ruler: 10 mm: 100 divisions

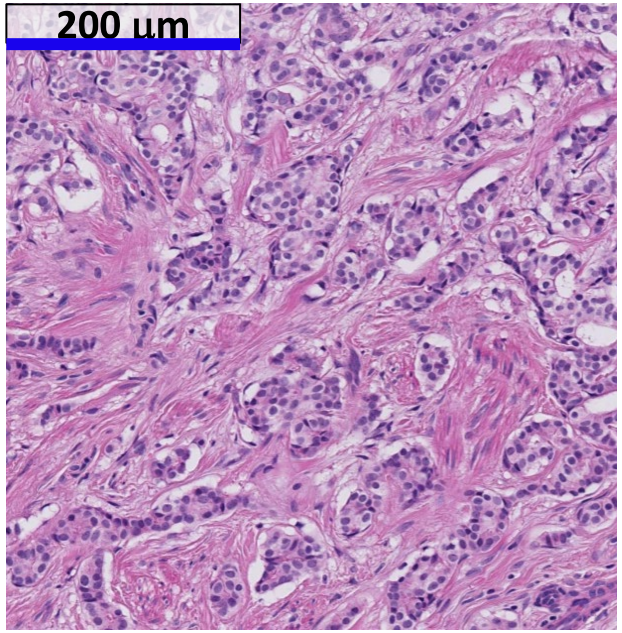
We measure radial distance from center of cross-hairs to target

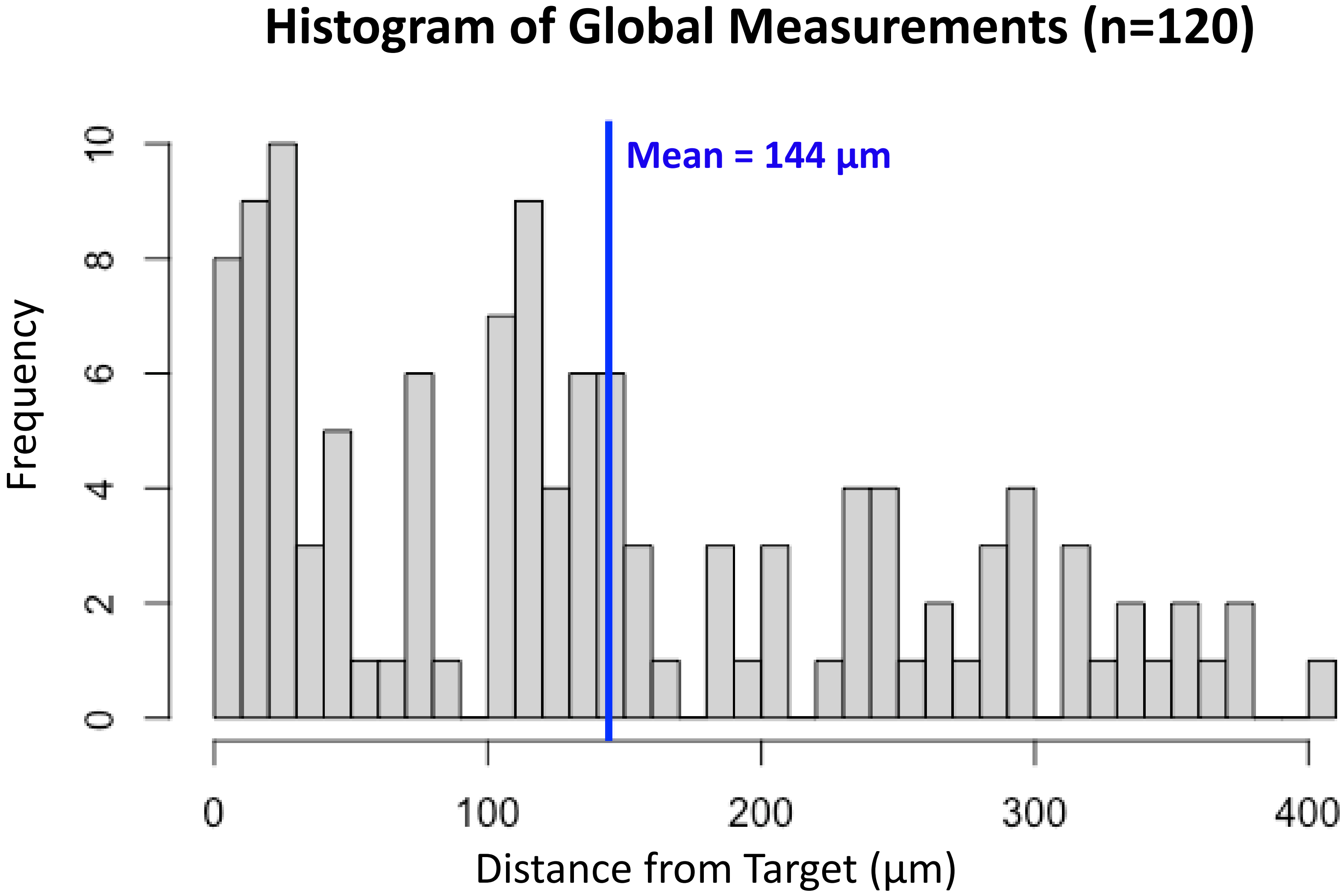


Global Registration Measurements



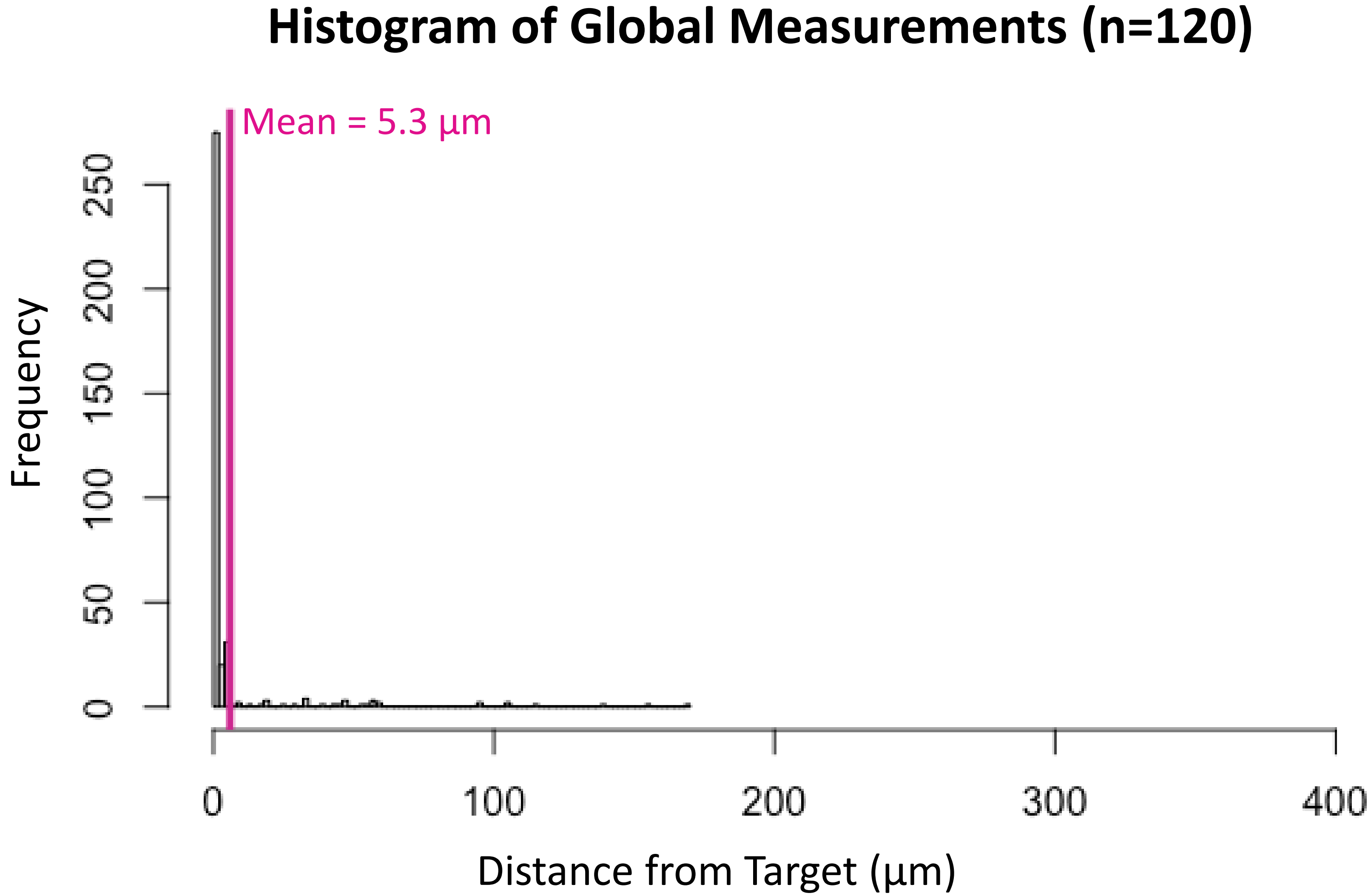
Global Registration Measurements





Local Registration Measurements

With Local:Fast and Local:Best, 97% of measurements were within 5 µm



Multi-Reader Multi-Case Analysis for sizing a validation study

Variance of Percent Correct (PC) where success = 1; failure = 0

Binary Data: Success ≤ 5 µm : Failure > 5 µm

3 Variance Components:

$$var\left(\widehat{PC}\right) = \frac{\sigma_C^2}{N_C} + \frac{\sigma_R^2}{N_R} + \frac{\sigma_{RC}^2}{N_R N_C},$$

Reader Terms

Case Terms

Interaction Terms between Cases & Readers

iMRMC Github:
https://github.com/DIDSR/iMRMC

Multi-Reader Multi-Case Analysis for sizing a validation study

12 Readers pooled across 2018 & 2022 readers and methods

3 Readers X (Local:Fast + Local:Best) X (List + Random) = 12 Readers

3 Variance Components:

$$var\left(\widehat{PC}\right) = \frac{\sigma_C^2}{N_C} + \frac{\sigma_R^2}{N_R} + \frac{\sigma_{RC}^2}{N_R N_C},$$

Reader Terms

Case Terms

Interaction Terms between Cases & Readers

iMRMC Github:
https://github.com/DIDSR/iMRMC

Multi-Reader Multi-Case Analysis for sizing a validation study

Using these variability results, we can estimate the N_C,N_R of future validation studies, like one for the HTT project

Component	MRMC Result
Average registration success rate	0.97
MRMC standard error of the average registration success rate	0.013
σ_C^2 = variability from cases	6.8 x10^{-3}
σ_R^2 = variability from readers	2.7 x10^{-4}
σ_{RC}^2 = variability from interaction between readers and cases.	2.3 x10^{-2}

$$var\left(\widehat{PC}\right) = \frac{\sigma_C^2}{N_C} + \frac{\sigma_R^2}{N_R} + \frac{\sigma_{RC}^2}{N_R N_C},$$

iMRMC Github:
https://github.com/DIDSR/iMRMC

Current and Future Work

Launch of the HTT Pivotal Study

Launch of training course and modules for pathologist sTILs Assessment

Recent Publications:

S. Wen & B.D. Gallas, Three-Way Mixed Effect ANOVA to Estimate MRMC Limits of Agreement. *Statistics in Biopharmaceutical Research*, 2022.

H. Du, et. al., Single reader between-cases AUC estimator with nested data. *Statistics in Biopharmaceutical Research* 31(11), 2022.

eeDAP Studies Group Page

A home for collaborative studies to create tools (methods, data, and code) that advance regulatory science in the area of digital pathology imaging and related artificial intelligence software as a medical device.

https://ncihub.org/groups/eedapstudies

https://www.zotero.org/groups/4384613/eedap_studies_presentations_publications_and_studies/collections/9ABM9D8M

Acknowledgements

FDA

Salgado et al: Slides and Images: We would like to acknowledge the Pathology Department of Institut Jules Bordet in Brussels, Belgium for providing the pathology slides for this work.

Kate Elfer would like to acknowledge the funding and support of the National Cancer Institute: Division of Cancer Prevention

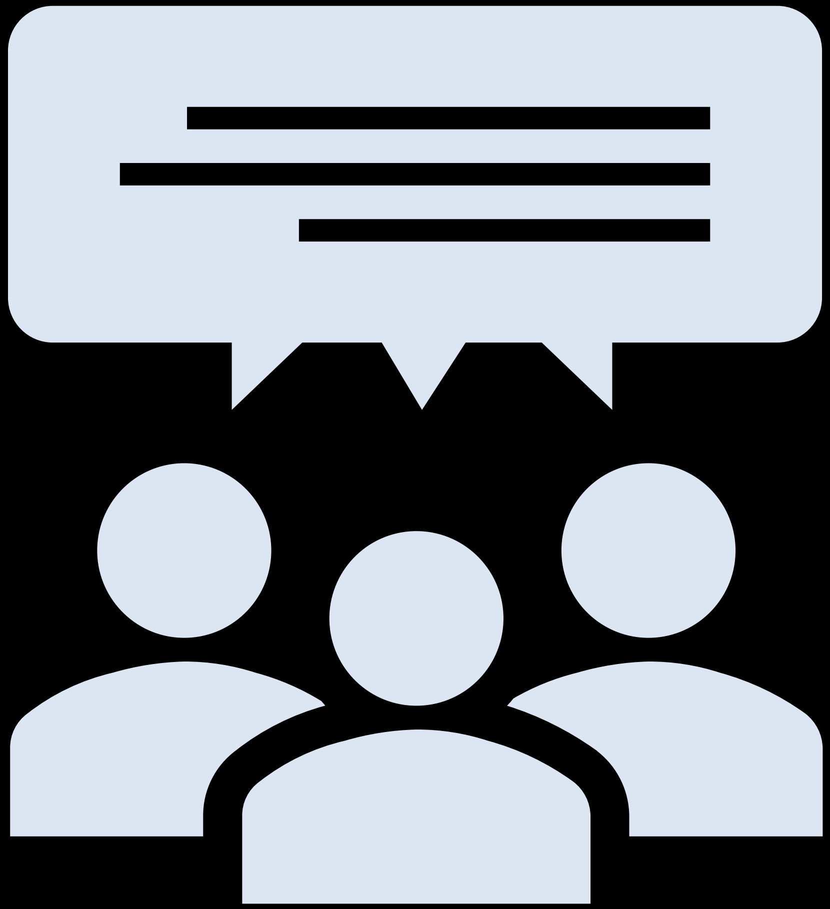
Cancer Prevention Fellowship Program https://cpfp.cancer.gov/

NIH NATIONAL CANCER INSTITUTE
Cancer Prevention Fellowship Program

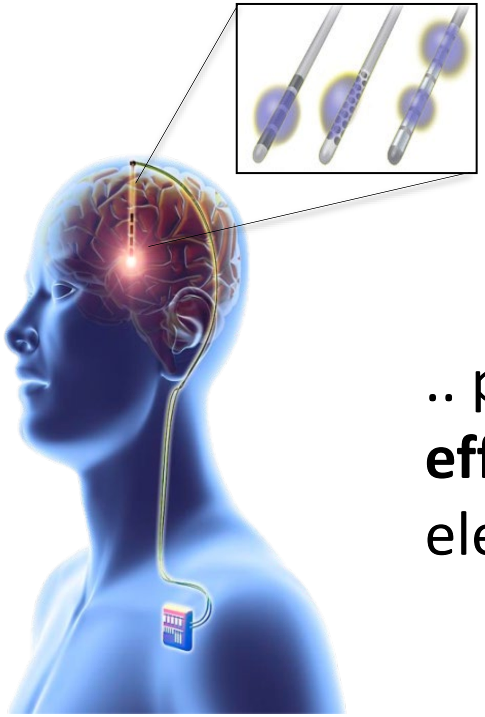
Collaborators

- **Mohamed Amgad, MD**
  - Department of Pathology, Northwestern University
- **Kim Blenman, PhD**
  - Yale School of Medicine
- **Weijie Chen, PhD**
  - FDA/CDRH/OSEL/DIDSR
- **Sarah Dudgeon, MPH**
  - CORE Center for Computational Health Yale-New Haven Hospital
- **Brandon Gallas, Phd**
  - FDA/CDRH/OSEL/DIDSR
- **Victor Garcia, MD**
  - FDA/CDRH/OSEL/DIDSR
- **Ryan Birmingham**
  - FDA/CDRH/OSEL/DIDSR
- **Emma Gardecki,**
  - FDA/CDRH/OSEL/DIDSR
- **Rajarsi Gupta, MD/PhD**
  - Stony Brook Medicine Dept of Biomedical Informatics
- **Matthew Hanna, MD**
  - Memorial Sloan Kettering Cancer Center
- **Steven Hart, PhD**
  - Department of Health Sciences Research, Mayo Clinic
- **Evangelos Hytopoulos, PhD**
  - iRhythm Technologies Inc
- **Joe Lennerz, MD**
  - Pathology Innovation Collaborative Community
- **Xiaoxian "Bill" Li, MD/PhD**
  - Emory University School of Medicine
- **Anant Madabhushi, PhD**
  - Case Western Reserve University
- **Roberto Salgado, PhD**
  - Division of Research, Peter Mac Callum Cancer Centre, Melbourne, Australia; Department of Pathology, GZA-ZNA Hospitals
- **Tony Pan, PhD**
  - Emory University, caMicroscope
- **Joel Saltz, MD/PhD**
  - Stony Brook Medicine Dept of Biomedical Informatics
- **Manasi Sheth, PhD**
  - FDA/CDRH/OPQE/Division of Biostatistics
- **Rajendra Singh, MD**
  - Northwell health and Zucker School of Medicine
- **Evan Szu, PhD**
  - Arrive Bio
- **Darick Tong, MS**
  - Arrive Bio
- **Si Wen, PhD**
  - FDA/CDRH/OSEL/DIDSR
- **Bruce Werness, MD**
  - Arrive Bio

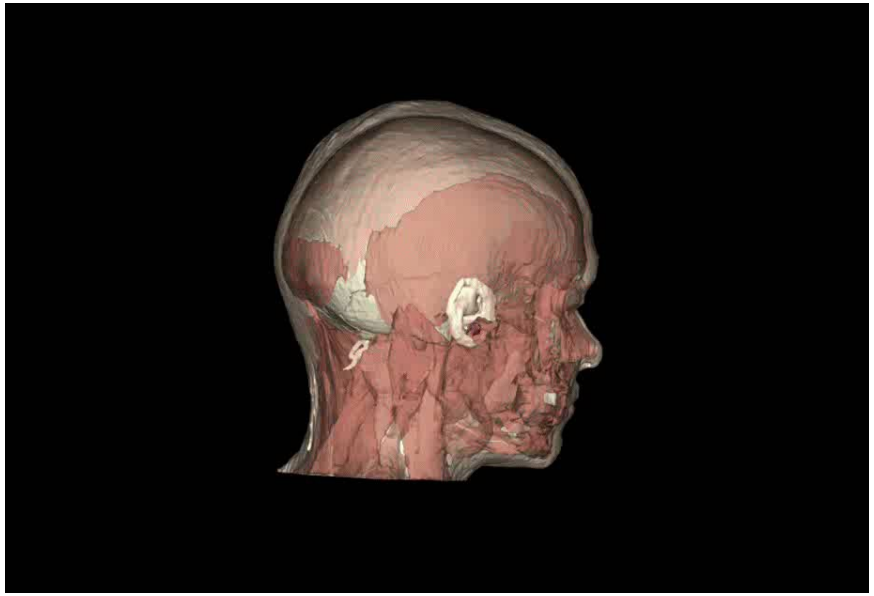
Open for Questions



CDRH Mission



.. protect and promote the health of the public by ensuring the **safety** and **effectiveness** of **medical devices** and the safety of radiation-emitting electronic products…



We facilitate medical device innovation by advancing regulatory science, providing industry with predictable, consistent, transparent, and efficient regulatory pathways, and assuring consumer confidence in devices marketed in the U.S.

CDRH Snapshot

1900
EMPLOYEES

18k
Medical Device Manufacturers

183k
Medical Devices
On the U.S. Market

22k/year
Premarket Submissions
including supplements and amendments

570k
Proprietary Brands

25k
Medical Device Facilities Worldwide

1.4 MILLION/year
Reports on medical device adverse events and malfunctions

Office of Science and Engineering Laboratories (OSEL)

- Conduct laboratory-based regulatory research to facilitate development and innovation of safe and effective medical devices and radiation emitting products
- Provide scientific and engineering expertise, data, and analyses to support regulatory processes
- Collaborate with colleagues in academia, industry, government, and standards development organizations to develop, translate, and disseminate science and engineering-based information regarding regulated products
- https://www.fda.gov/about-fda/cdrh-offices/office-science-and-engineering-laboratories

OSEL Snapshot

183
FEDERAL EMPLOYEES
Up to 180 visiting scientists

140 Projects
In 20 Program Areas

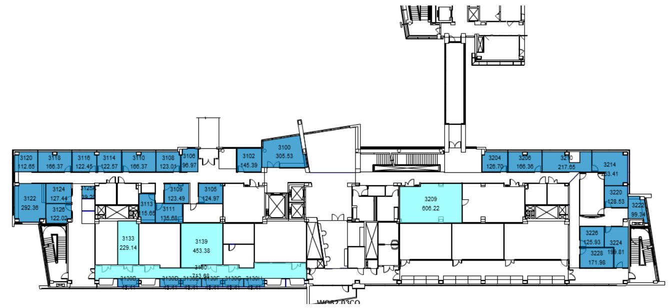
400/year
Peer reviewed presentations, articles, and other public disclosures

2,500/year
Premarket Regulatory consults

75
Standards and conformity assessment committees

70%
Staff with post graduate degree

55,000 ft^2
Lab facilities



Division of Imaging, Diagnostics and Software Reliability (DIDSR)

- Develop least burdensome approaches for regulatory evaluation of imaging and big-data devices
  - Efficient clinical trials accounting for reader variability, simulation tools, in silico phantoms and imaging trials, addressing issues related to imperfect / missing reference standards, and limited data for training/testing of machine classifiers
- Develop measures of technical effectiveness of imaging and big-data technologies
  - Phantoms, laboratory measurements, computational models

DIDSR Snapshot

35
FEDERAL EMPLOYEES
14 Fellows/Students
3 Open Staff Positions

145/year
Peer reviewed articles, code and presentations

4 Program Areas
- AI/ML
- Medical Imaging and Diagnostics
- Digital Pathology
- Mixed Reality (AR/VR/XR)

550/year

Premarket Regulatory consults

~15,000 ft^2
DIDSR Lab and facilities