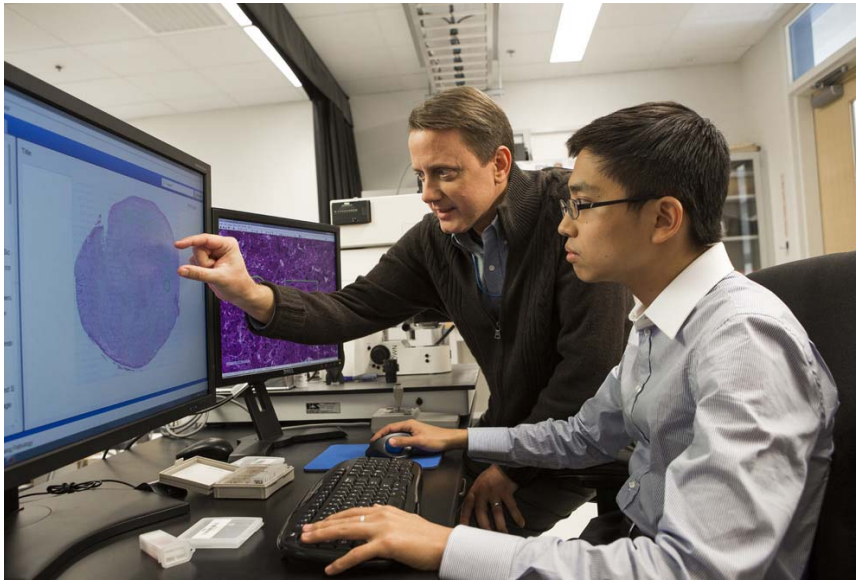
Reader Studies/Feature Studies, statistics

Dr. Brandon Gallas
CDRH/OSEL/DIDSR

Outline

- Clinical Performance Assessment
  - Technical
  - Analytical=Features=Reproducibility
  - Clinical
- My niche: Design, execution, and analysis of studies involving the reader
- Mitotic counting study
- Evaluation environement for digital and analog pathology
- Medical Device Development Tool (MDDT)
- WSI Working Group

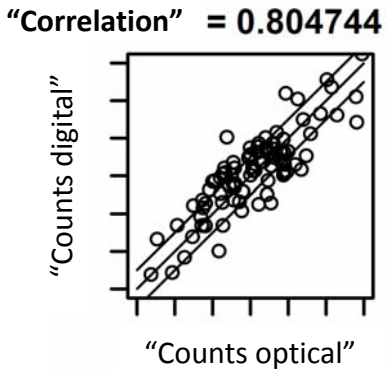


Clinical Performance Assessment

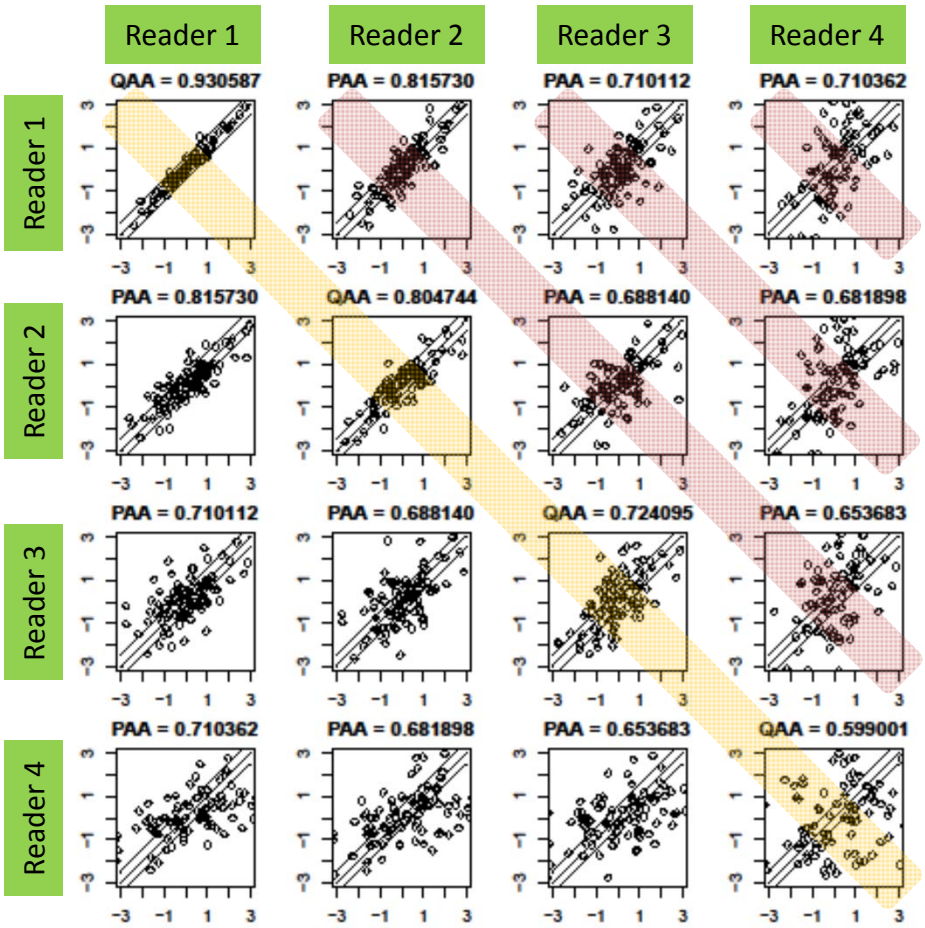
- Current research: MRMC analysis of
  - Within-reader agreement
  - Between-reader agreement

- Evaluate digital pathology (or any new imaging technology)
  - Truth not known
  - Reference for comparison determined by readers
  - ***"Reproducibility" Study***
  - Example tasks
    - Impact of viewing mode on mitotic count (digital vs. optical)
    - Impact of stain on mitotic count (H&E vs. PHH3)

- Study designs and MRMC analysis methods are lacking



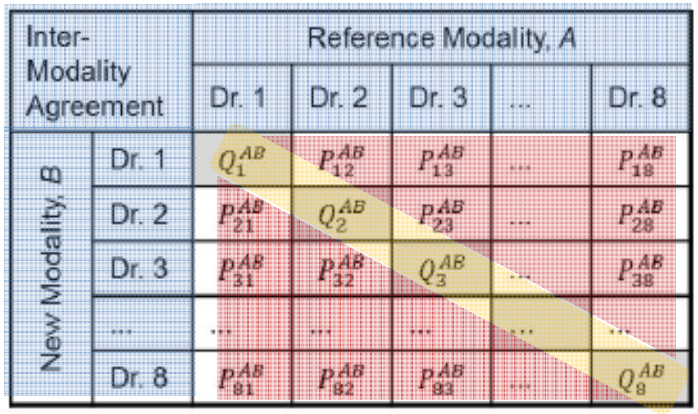
Clinical Performance Assessment

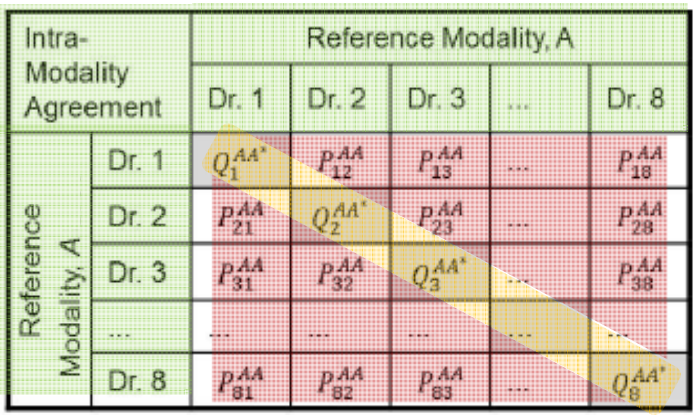


- Multiple readers evaluate Multiple cases
- Variability from cases
- Variability from readers

Clinical Performance Assessment

- Is between-modality agreement (New Device Performance) worse than within-modality agreement (Baseline Performance)?
  - Within-reader?
  - Between-reader?





requires replicated readings

Study Design & Analysis

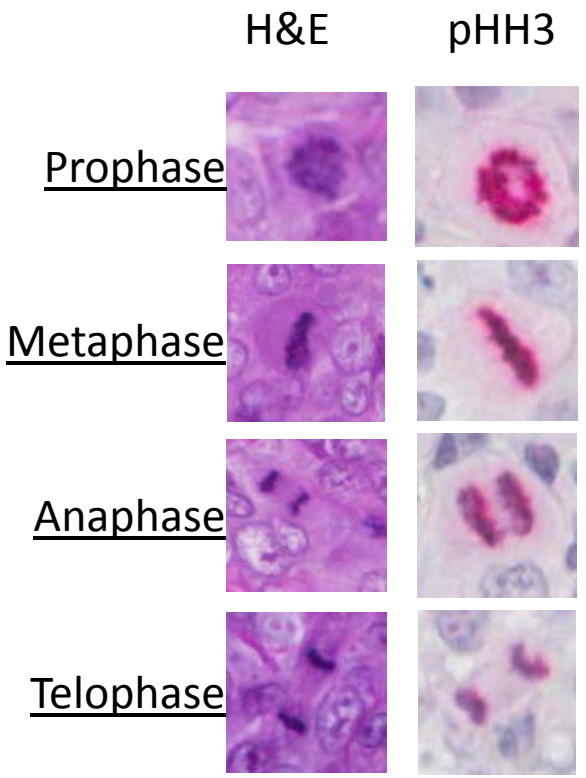
- What about error bars for study endpoints?

- **MRMC** analysis: account for variability from
  - **M**ultiple **R**eaders
  - **M**ultiple **C**ases

- Gallas, Anam, Chen, Wunderlich, Zhang (2016), "MRMC analysis of agreement studies." *In Proc. SPIE,* ***9787, pp. 97870F-97870F-12.***
  - *U-statistics*
  - *Agreement by concordance*
  - *Novel, builds on methods for area under the ROC curve*
  - *Peer reviewed manuscript in development*

Clinical Performance Assessment

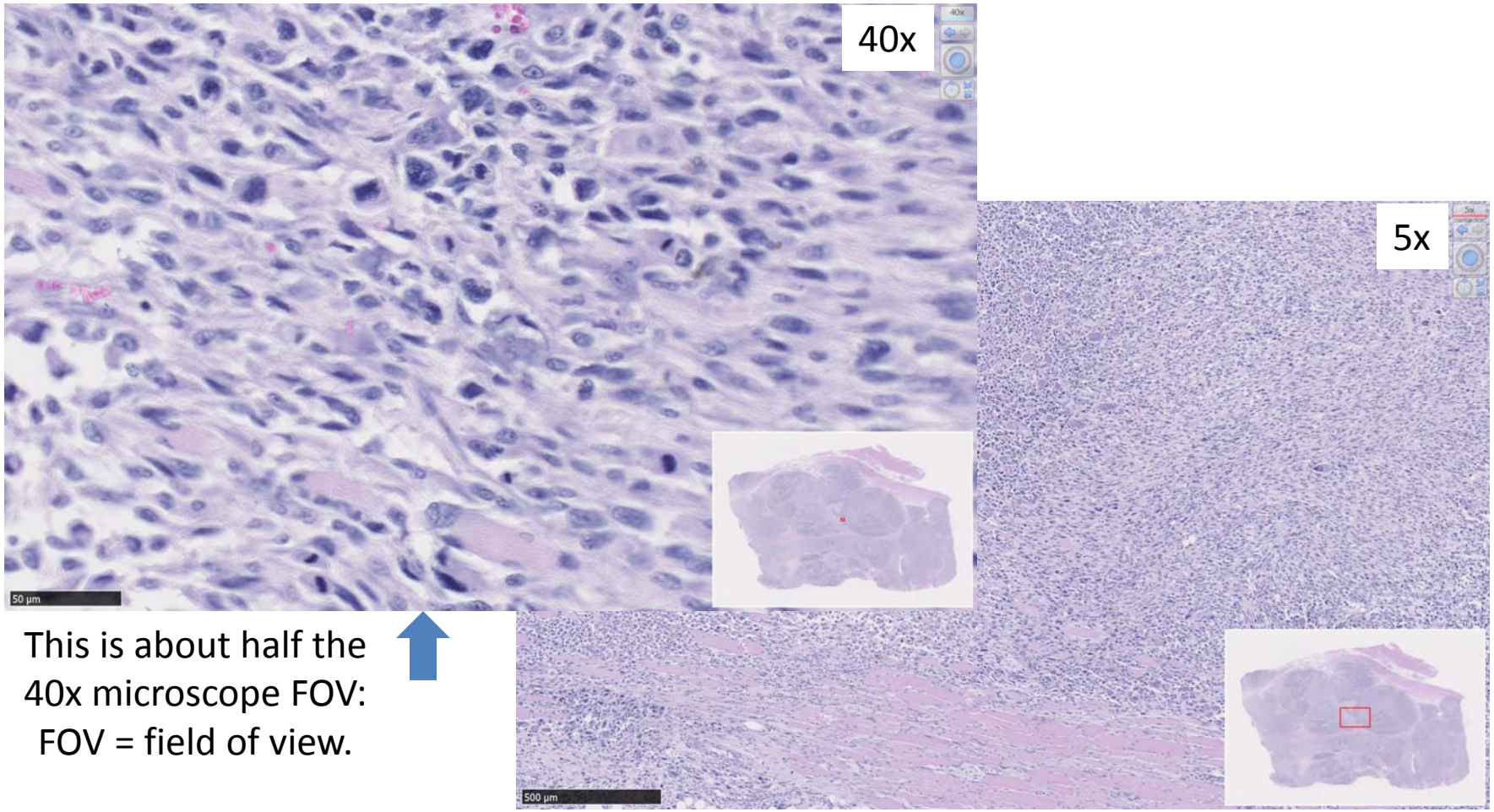
- Reproducibility study = feature study
  - Surrogate for clinical task
- Mitotic Counting
  - NIH Study, Mark Simpson (DVM, PhD) et al.
  - Quantitative prognostic feature for survival
  - Compare stains: H&E vs. pHH3
  - Compare microscope to WSI
- 12 readers:
  - veterinary pathologists at 4 sites
  - NCI, 2 academic centers, one regional reference lab
- 113 cases:
  - 1 case = 1 slide
  - specimens from canines with oral melanoma



Looks easy, but …

Clinical Performance Assessment

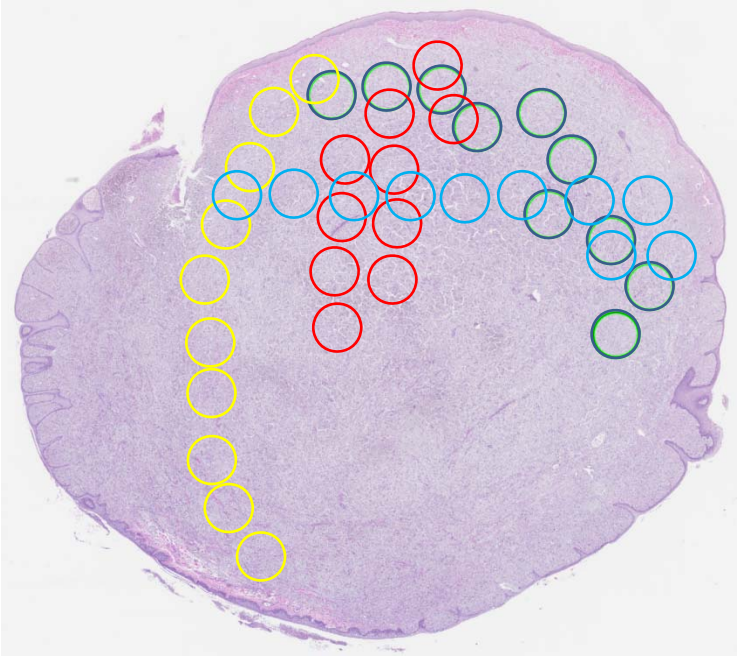
- It's hard to find mitotic figures because there are so many cells!



This is about half the 40x microscope FOV: FOV = field of view.

Clinical Performance Assessment

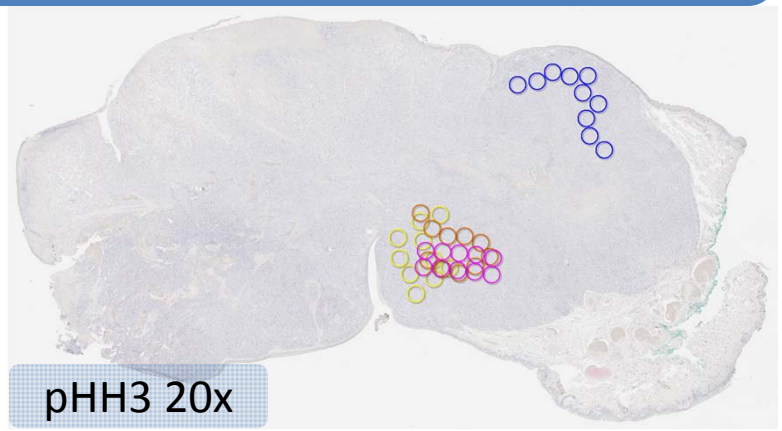
- How do pathologists count mitotic figures in practice?
  - Identify region of high mitotic activity
  - Count mitotic activity in 10 sequential 40x FOVs

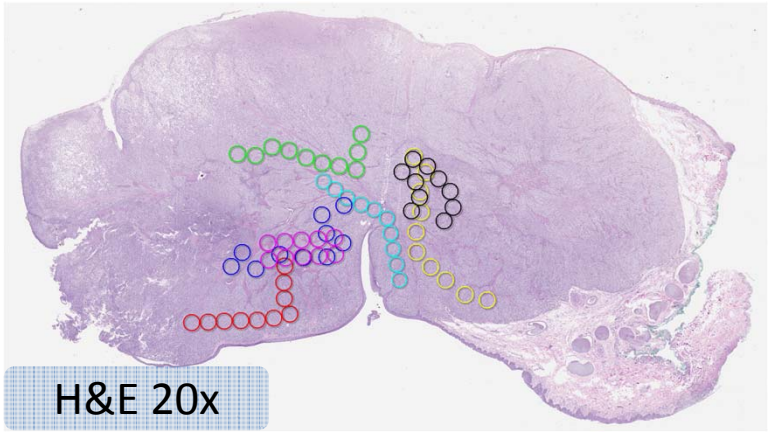


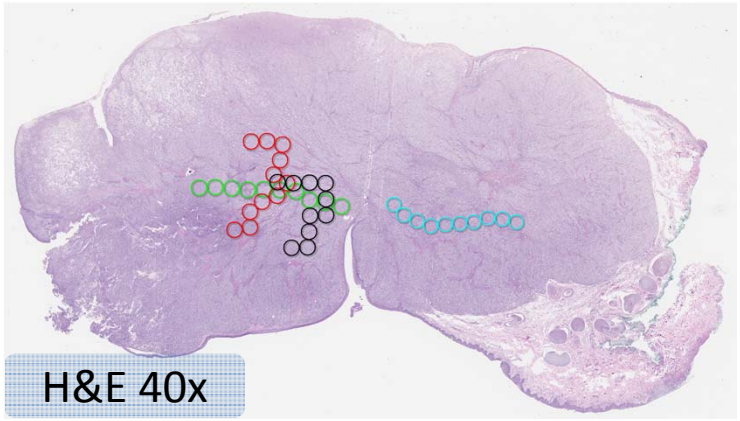
Clinical Performance Assessment

- NIH Study data
  - FOV locations saved for each pathologist in digital mode

Counts come from different tissue!
Clinical practice vs. technology evaluation



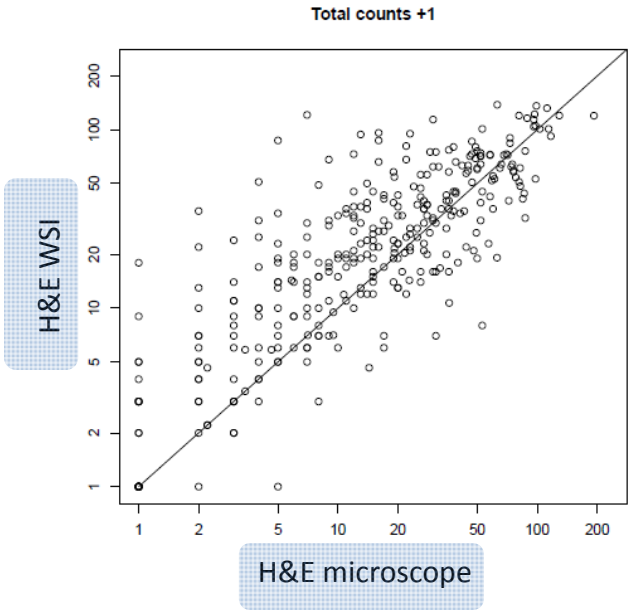


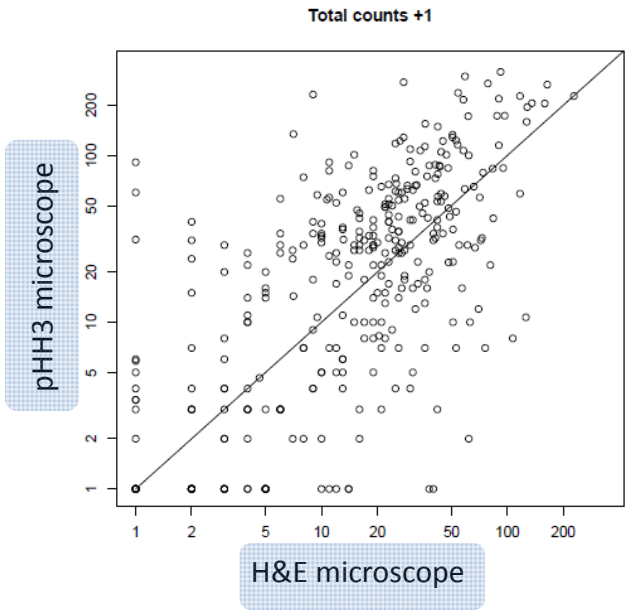


Clinical Performance Assessment

- NIH Study data

Each point shows the counts from two readers evaluating the same case. Each count is the sum of 10 FOVs.



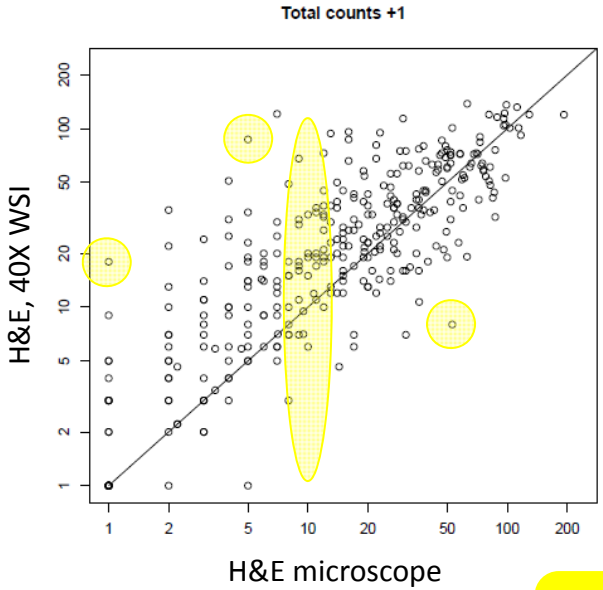


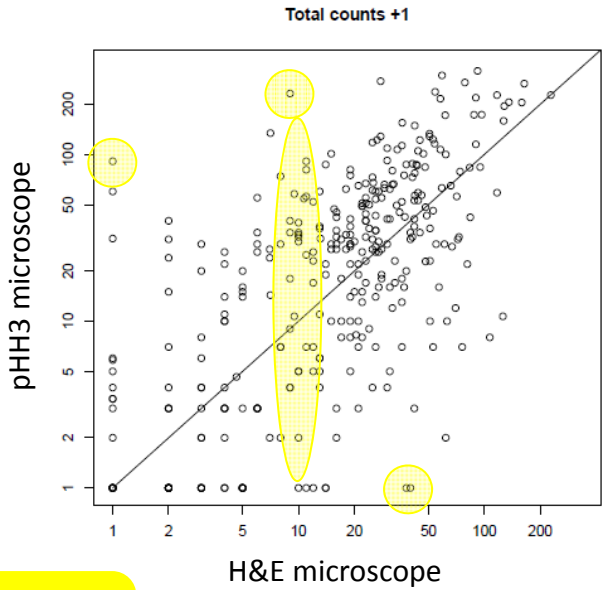
Clinical Performance Assessment

- NIH Study data

Each point shows the counts from two readers evaluating the same case. Each count is the sum of 10 FOVs.

Very large coefficient of variation.



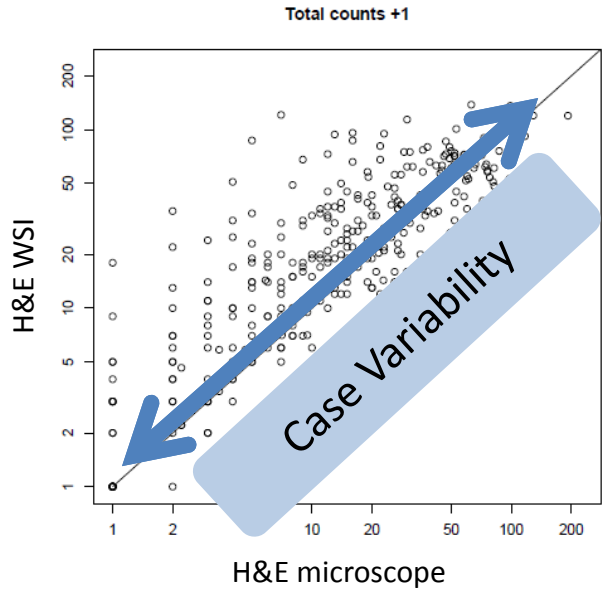


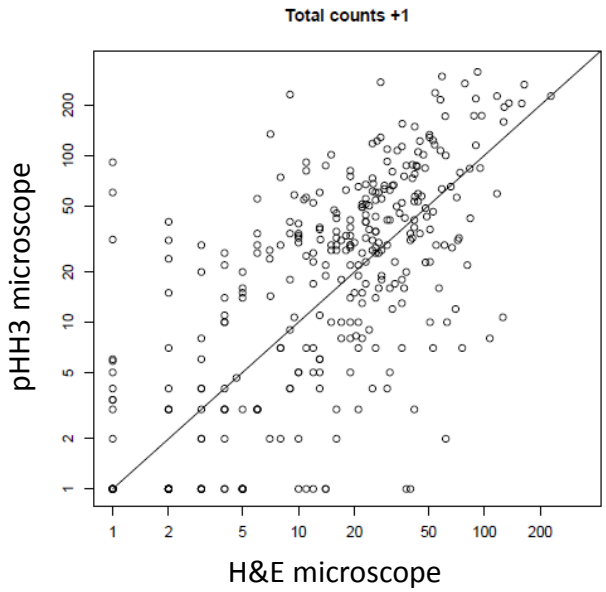
Warning: log-log plot

Clinical Performance Assessment

- NIH Study data

Each point shows the counts from two readers evaluating the same case. Each count is the sum of 10 FOVs.



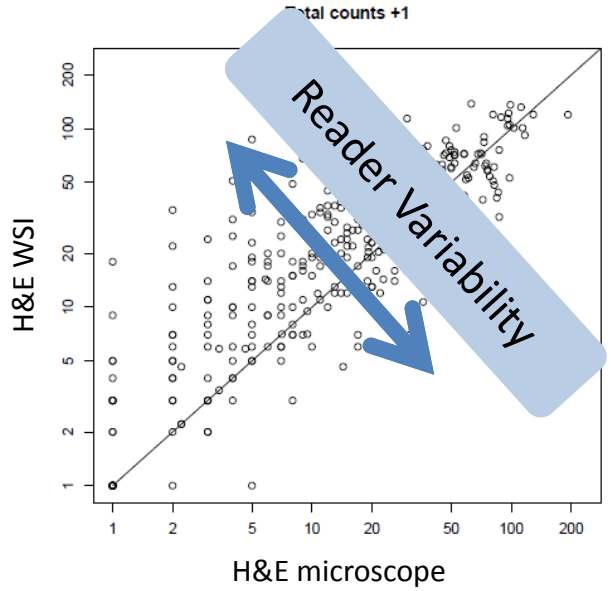


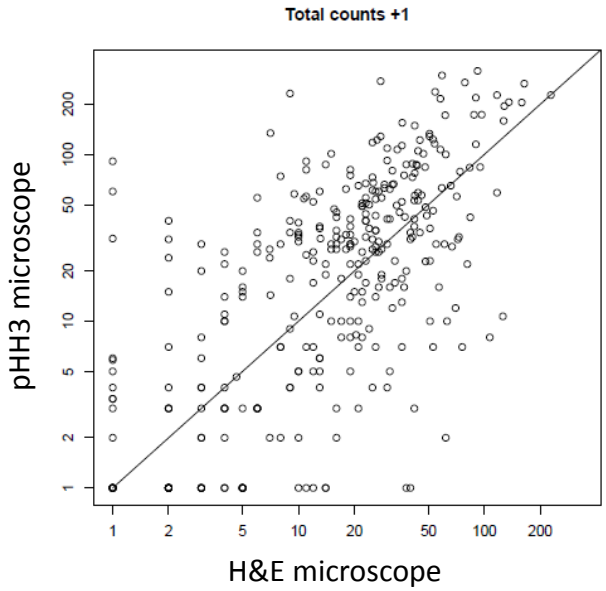
Clinical Performance Assessment

- NIH Study data

Each point shows the counts from two readers evaluating the same case. Each count is the sum of 10 FOVs.

Variability from readers? FOVs? technologies?



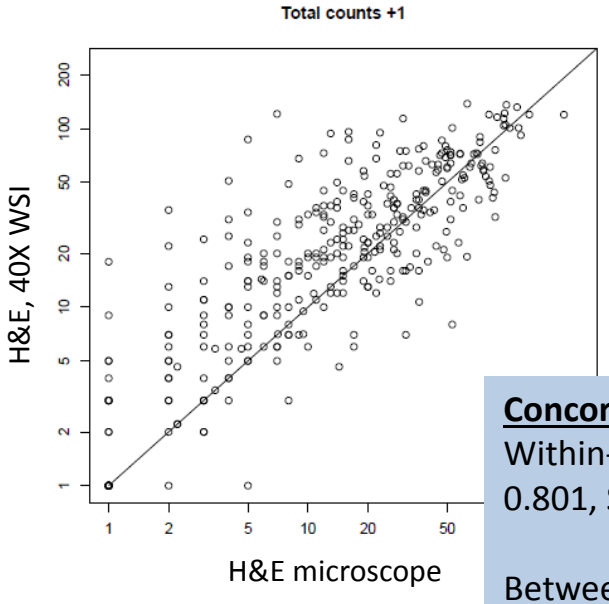


Clinical Performance Assessment

- NIH Study data

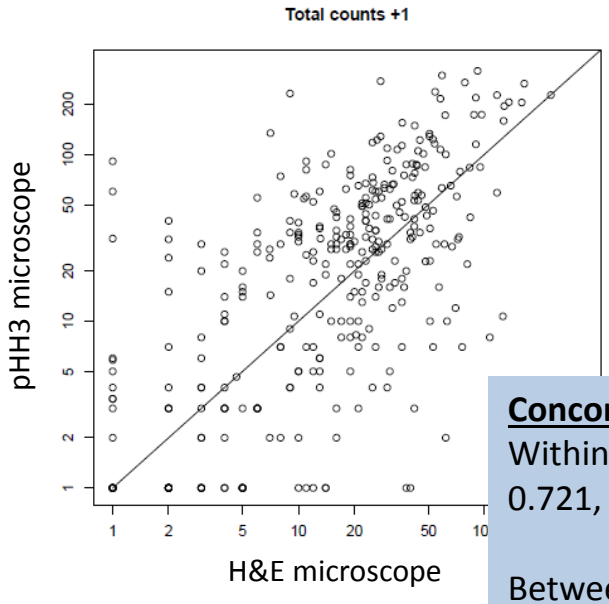
Each point shows the counts from two readers evaluating the same case. Each count is the sum of 10 FOVs.

Current research:
SE accounts for reader variability



Concordance Rate:
Within-reader
0.801, SE = 0.018

Between-reader
0.807, SE = 0.059



Concordance Rate:
Within-reader
0.721, SE = 0.027

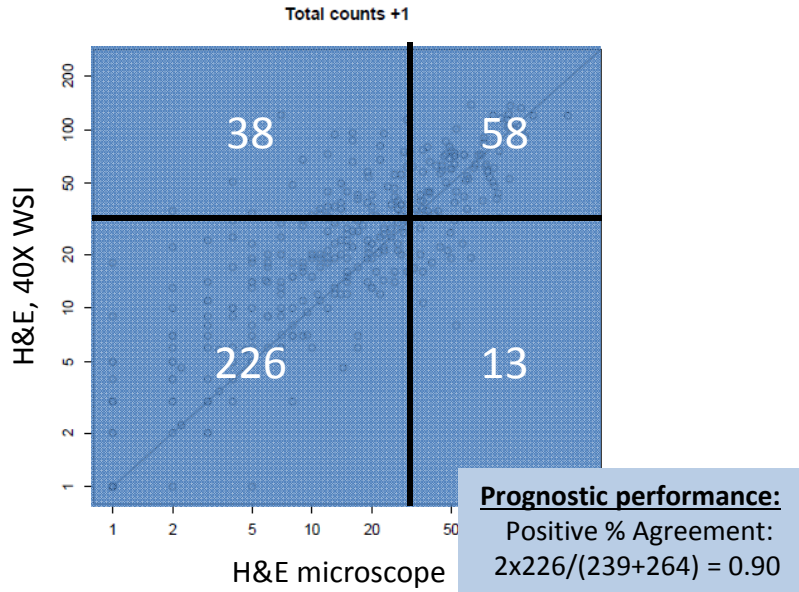
Between-reader
0.721, SE = 0.020

Clinical Performance Assessment

- NIH Study data

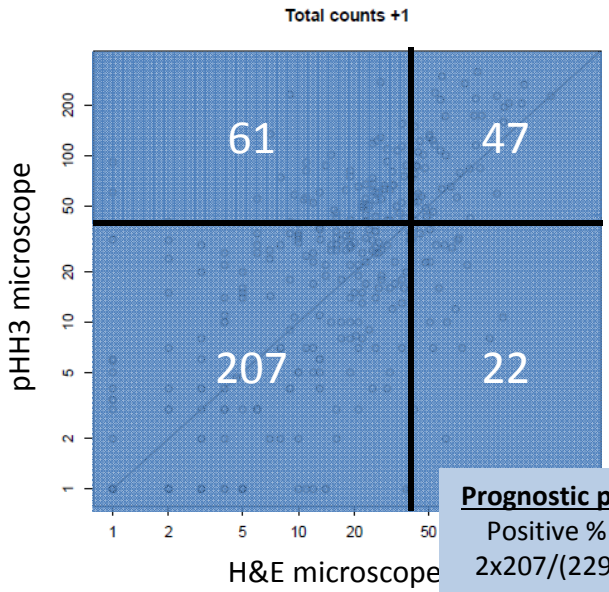
40 total counts is a
clinical prognostic threshold
for death or euthanasia within 1 year

Sensitivity = 90%, Specificity = 84%
Bergin2011_Vet-Pathol_v48p41



Prognostic performance:
Positive % Agreement:
2x226/(239+264) = 0.90

Negative % Agreement:
2x58/(96+71) = 0.69

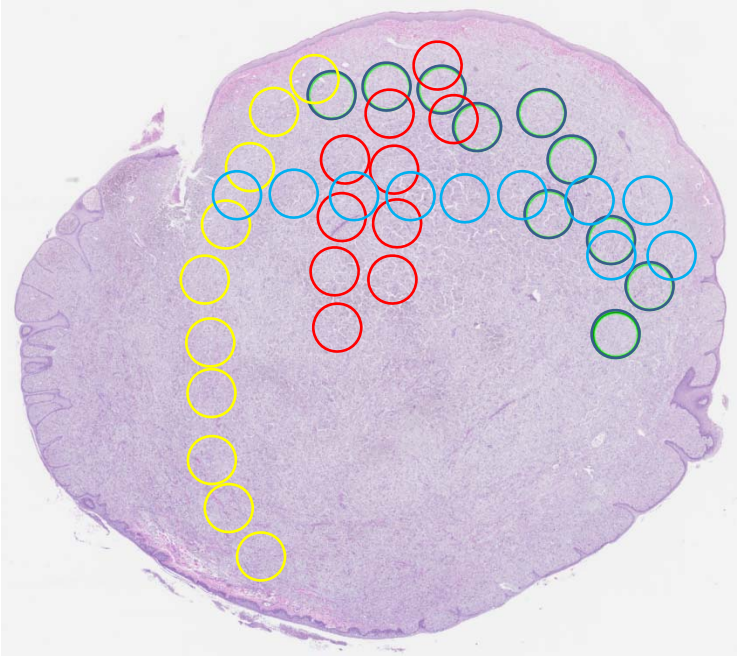


Prognostic performance:
Positive % Agreement:
2x207/(229+268) = 0.83

Negative % Agreement:
2x47/(108+69) = 0.53

Technology Performance Assessment

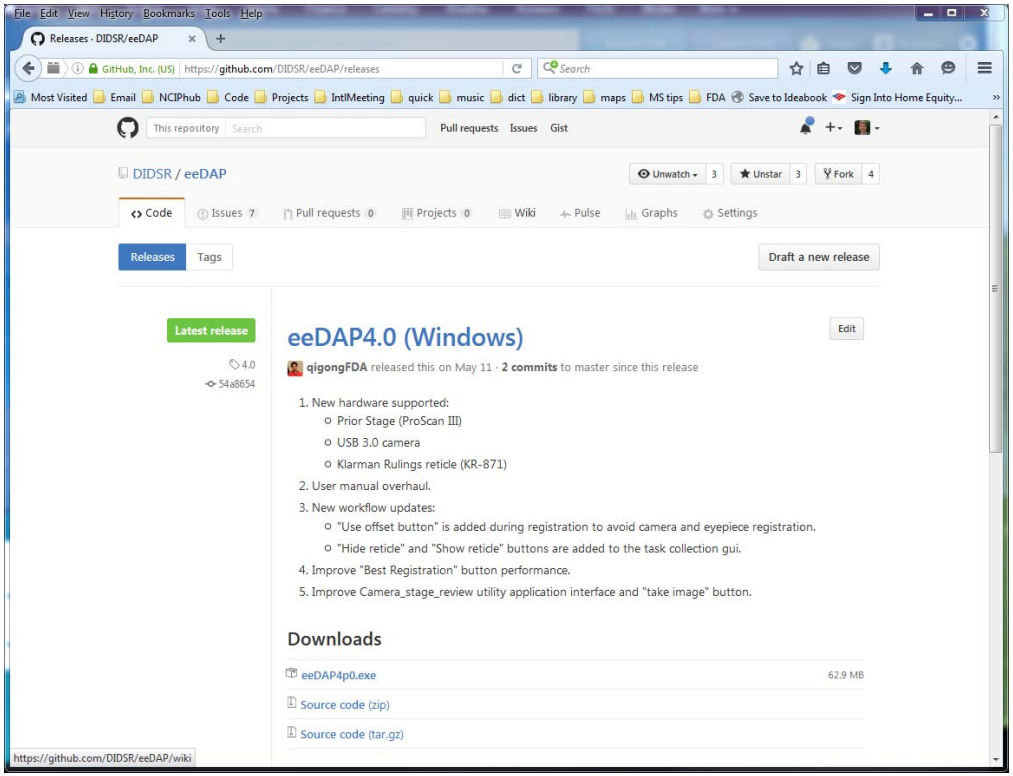
- Can we eliminate or reduce reader and location variability for technology assessment?



eeDAP

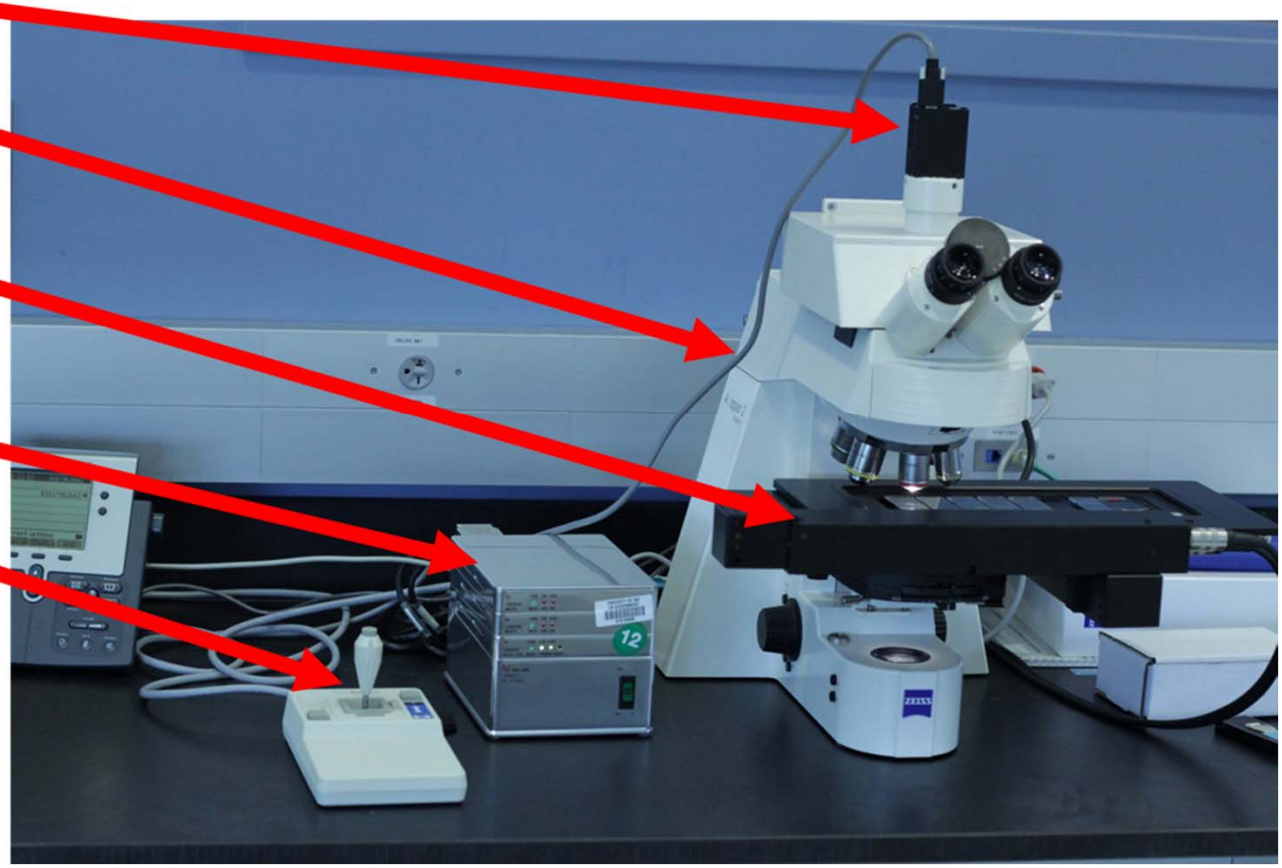
- Evaluation environment for digital and analog pathology
  - View and evaluate same fields of view in digital mode and on the microscope.

- License-free app
- Source code
- Documented



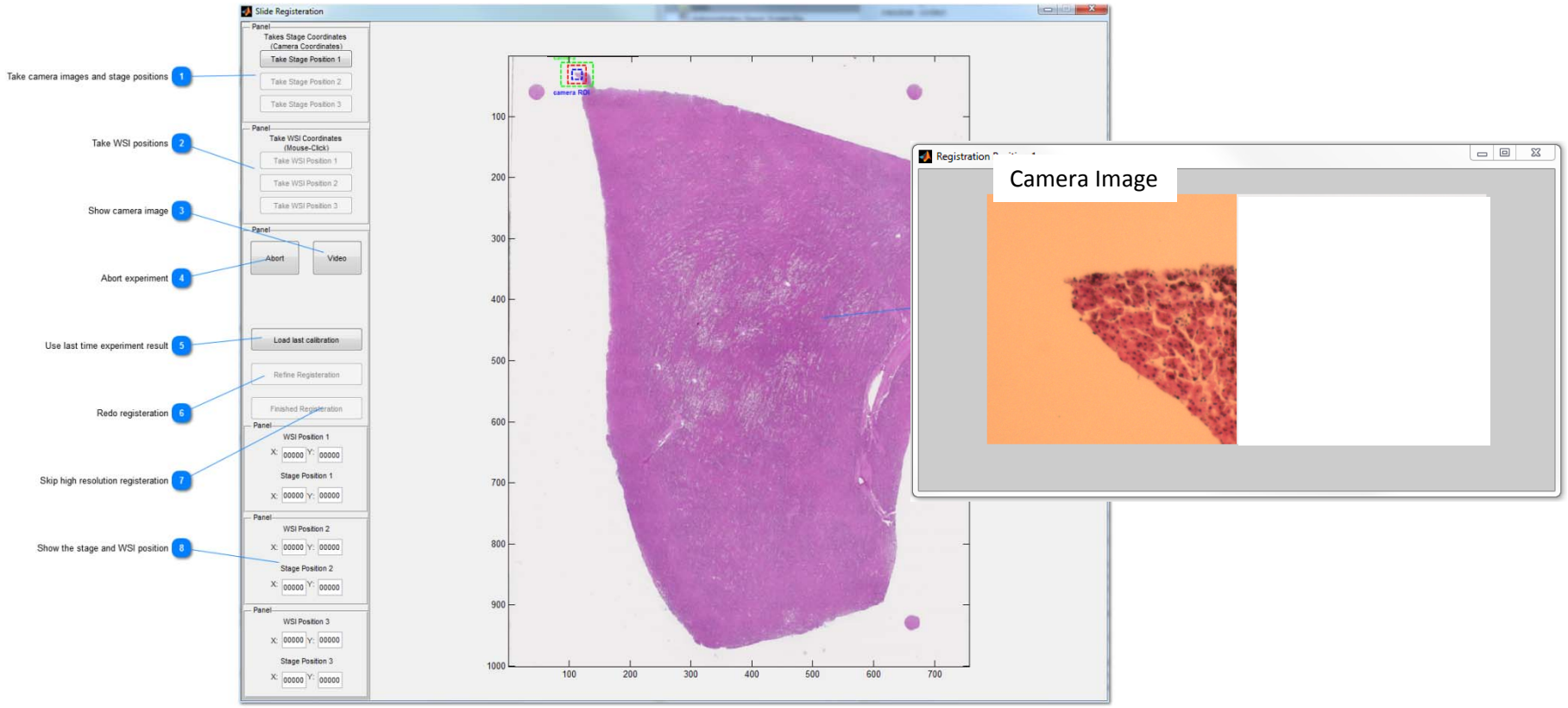
eeDAP

- Camera
- Microscope
- Moving stage with multiple slides
- Stage controller
- Joystick for stage control
- Computer and monitor not shown



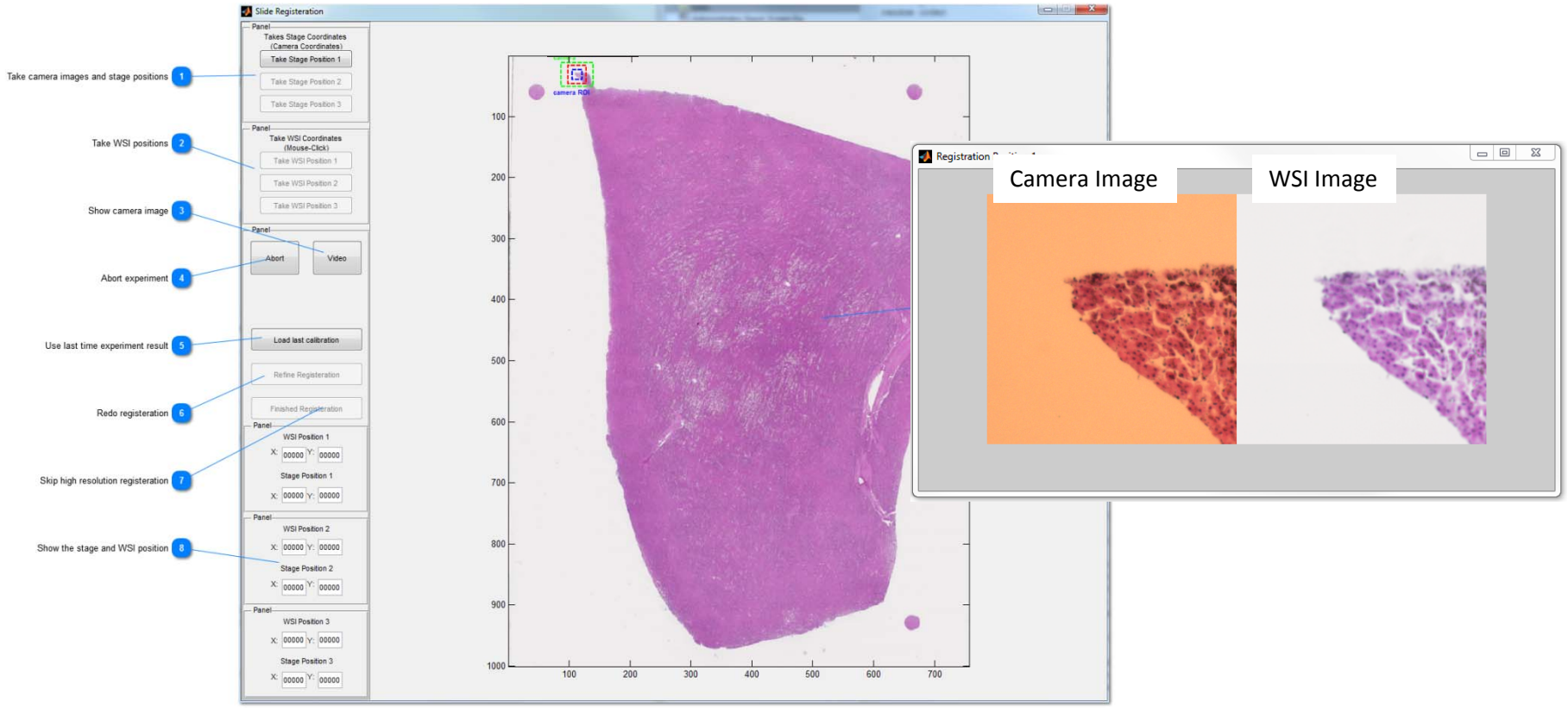
eeDAP

- Register WSI to glass slide on the microscope



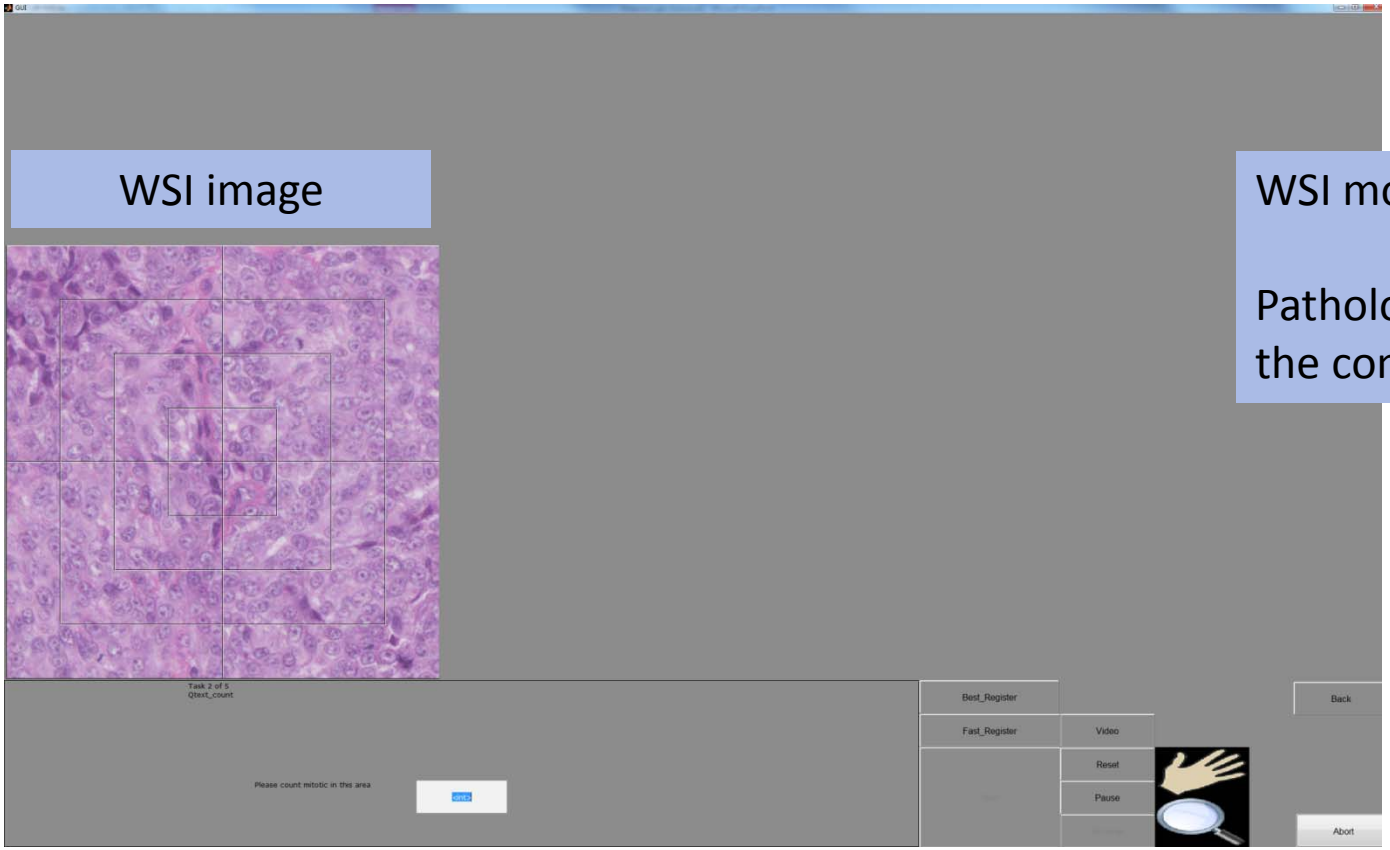
eeDAP

- Register WSI to glass slide on the microscope



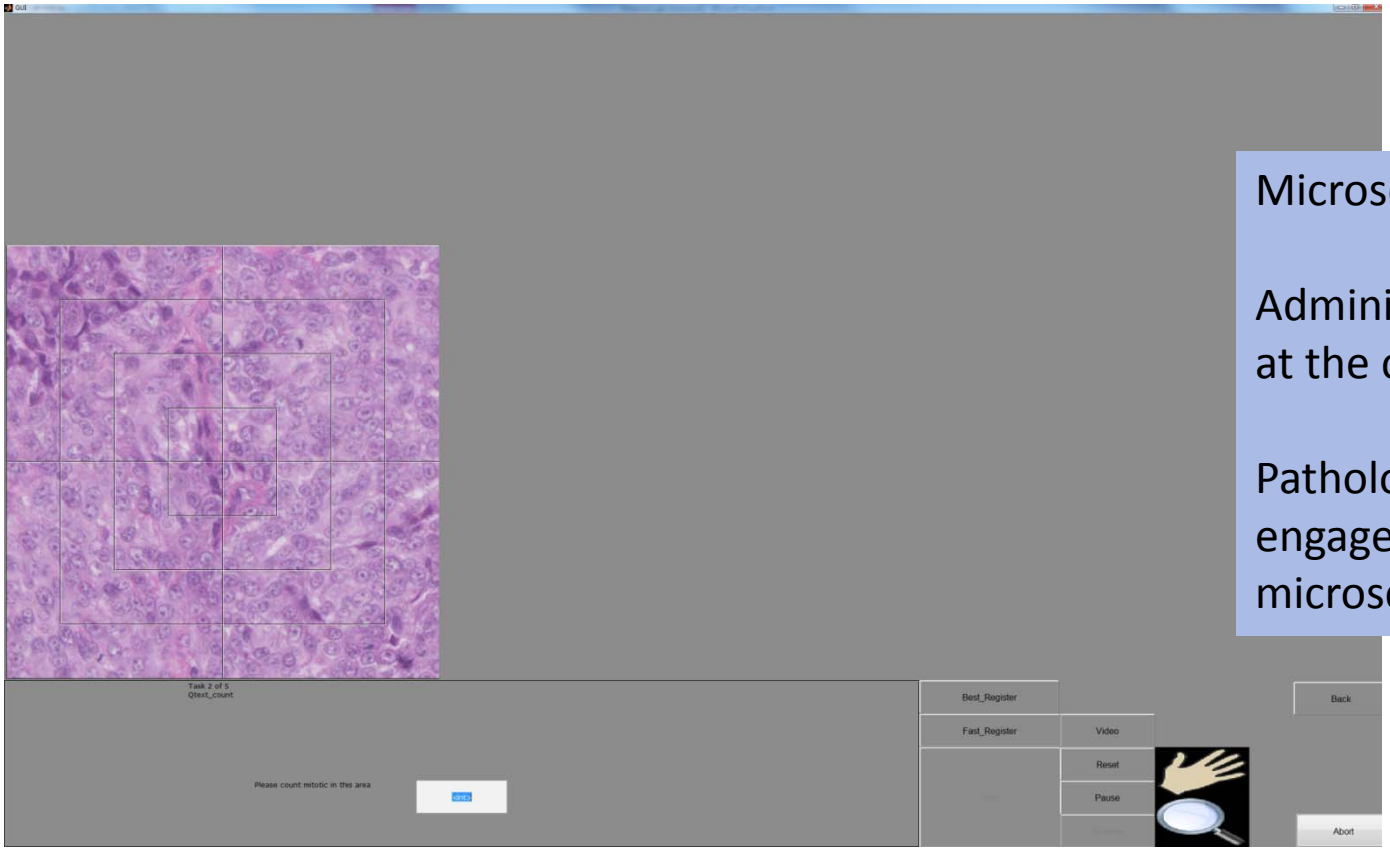
eeDAP

- Collect Data in WSI mode or Microscope mode



eeDAP

- Collect Data in WSI mode or Microscope mode



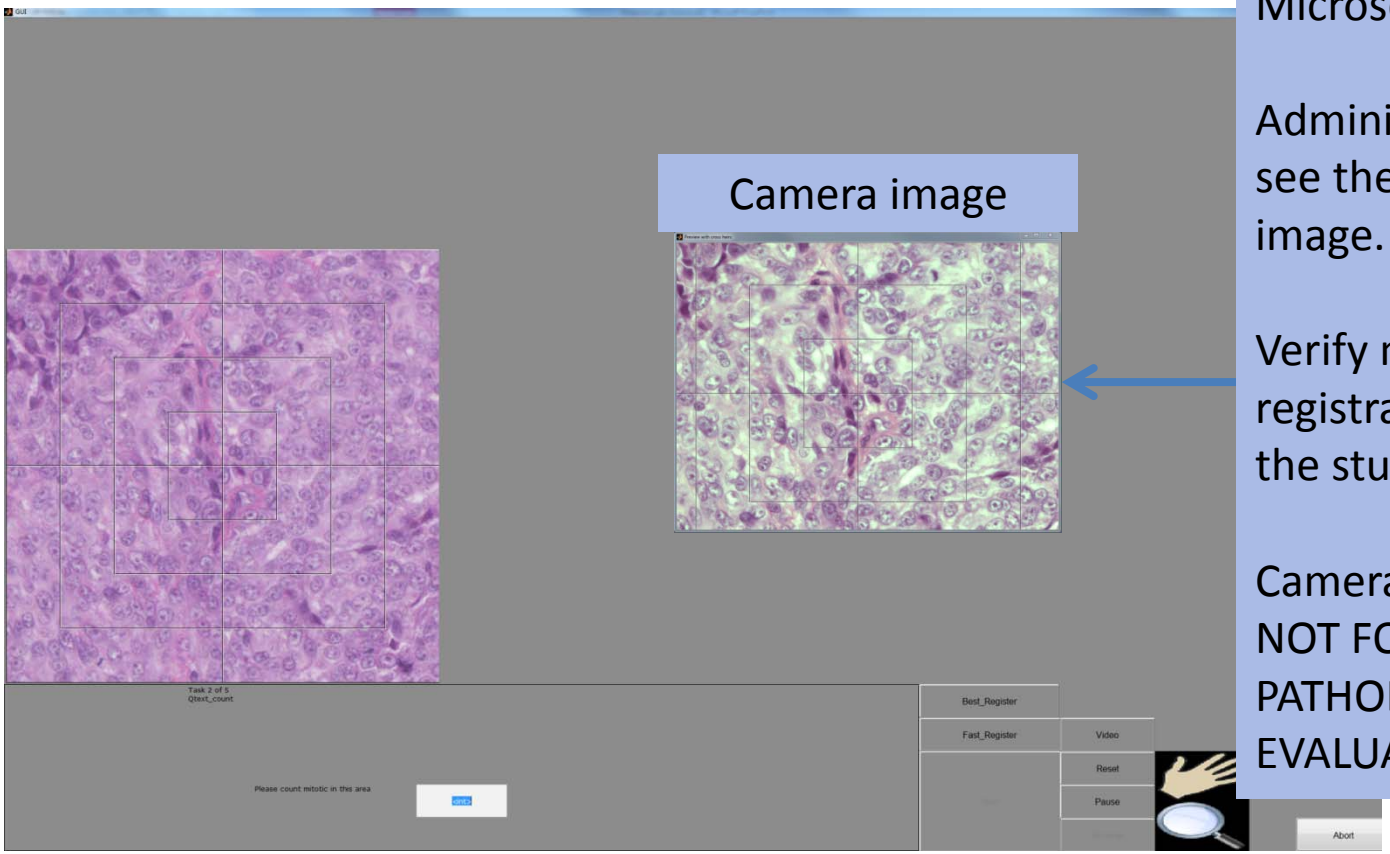
Microscope mode:

Administrator sits at the computer.

Pathologist is engaged with microscope.

eeDAP

- Collect Data in WSI mode or Microscope mode



Microscope Mode:

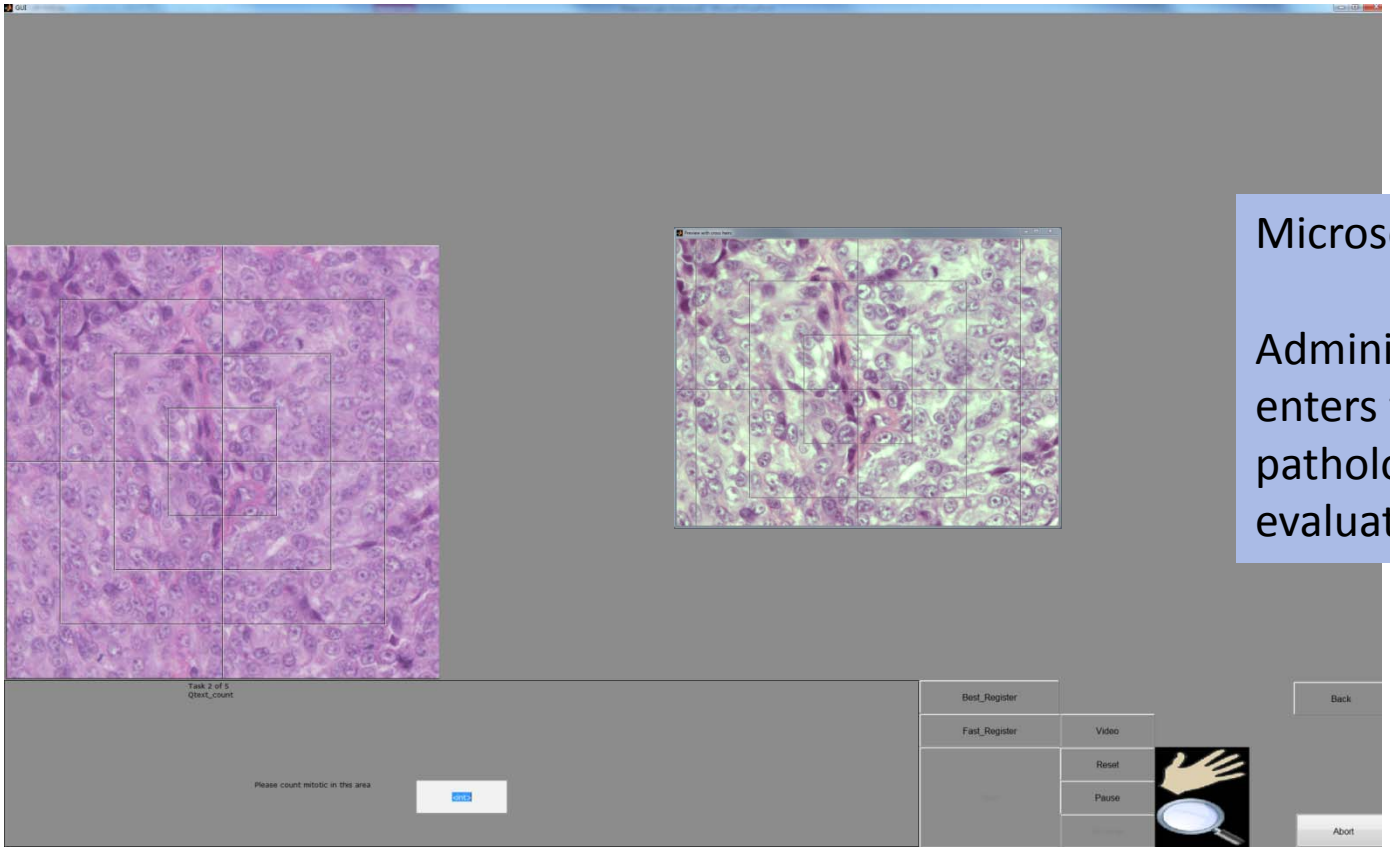
Administrator can see the camera image.

Verify microscope registration during the study.

Camera image is NOT FOR PATHOLOGIST EVALUATION

eeDAP

- Collect Data in WSI mode or Microscope mode



Microscope Mode:

Administrator enters the pathologist's evaluation.

eeDAP and MDDT Program

- "An MDDT ifs a scientifically validated tool … that aids device development and regulatory evaluation. Qualification reflects CDRH's expectation that within a specified *context of use*[1], the results of an assessment that uses an MDDT can be relied upon to support device development and regulatory decision-making."

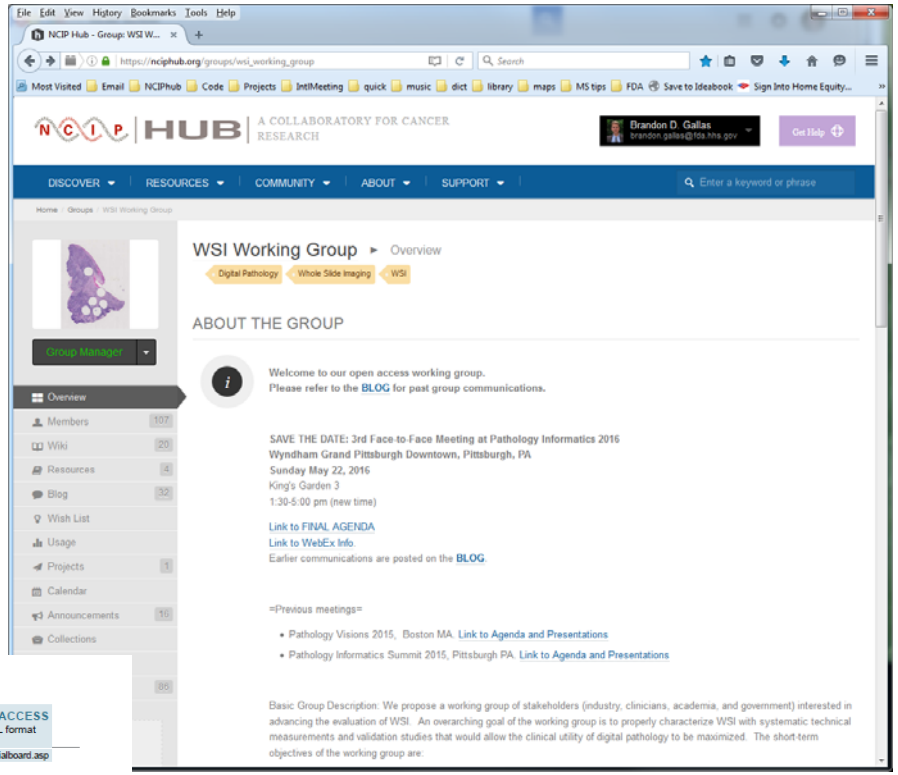
- An MDDT submission and qualification are not drastically different from a medical device submission, except the purpose of the MDDT is to ***support*** regulatory decision-making (an MDDT is not a medical device). An MDDT gets reviewed by FDA/CDRH pretty much like a medical device would get reviewed.

- "Once an MDDT is qualified for a specific context of use, FDA's expectation is that it can be used by any medical device developer for that context of use. CDRH reviewers should accept the MDDT for the qualified context of use without the need to reconfirm the suitability of the MDDT."

- Share the MDDT submission documents and process with the digital pathology community
  - WSI Working Group

WSI Working Group

- 107 members!
  - Broad representation of stakeholders.
  - Proposed group and collected emails
    - 5/2014: Pathology Informatics Summit
  - All work is in the public space
    - 8/2014: Established group at NCIPhub
  - Published Mission 2015




J Pathol Inform

Commentary

Evaluating whole slide imaging: A working group opportunity

Darren Treanor[1], Brandon D. Gallas[2], Marios A. Gavrielides[2], Stephen M. Hewitt[3]

[1]Leeds Teaching Hospitals NHS Trust and University of Leeds, Leeds, England, [2]Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, USA, [3]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, USA

E-mail: *Stephen M. Hewitt - genejock@helix.nih.gov
*Corresponding author

WSI Working Group

- 3 Face-to-face meetings
  - Pathology Informatics 2015, 2016
    - Association for Pathology Informatics
  - Pathology Visions 2015
    - Digital Pathology Association (DPA) conference
    - “Industry” group
  - Special Format:
    Presentations followed by critical feedback
  - Presenters have all been from industry (+ FDA)
    - Study designs
    - Methods
    - Tools

WSI Working Group

- Building trust and credibility
  - Desire exists
  - There are active members, but …
  - Momentum is hard to maintain
  - Challenges to momentum
    - Protecting IP
    - Business case for investing time and effort

- Next step: Submit eeDAP to MDDT
  - Peek inside the business and culture of FDA
  - Learn how to draft a context of use (== intended use)
  - Doesn't require WSI WG member participation (members are invited)
  - Might spawn participation (other MDDT efforts)

eeDAP and MDDT

- Q: Why eeDAP?
- *A: Eliminate location variability for faster and more precise results*
- Planning eeDAP study
  - Complement NIH study
  - Mitotic counting
  - H&E vs. PHH3
  - Optical vs. Digital

Summary

- FDA is developing and disseminating study design and analysis methods
  - Research programs: modeling, computer simulation, phantoms (VICTRE)

- Mitotic counting studies
  - Compare eeDAP study to clinical study
  - Hypothesis: eeDAP will produce faster and more precise results

- Medical Device Development Tool (MDDT)
  chance to engage
- WSI Working Group
  and provide a window onto FDA culture

- Looking for partners for eeDAP study


FDA
U.S. FOOD & DRUG
ADMINISTRATION