Pilot reader studies to compare digital microscopic images versus the microscope

Division of Imaging and Applied Mathematics, CDRH/FDA:

Brandon D. Gallas
Wei-Chung Cheng
Marios A. Gavrielides
Adam Ivansky*
Tyler Keay*
Adam Wunderlich

Laboratory of Pathology, Center for Cancer Research, NCI/NIH:

Stephen Hewitt
Jason Hipp*

*No longer with the gov.

Motivation

- Pathology Going Digital
  - Whole Slide Imaging
  - Virtual or Digital Microscopy

- Facilitated Consultations
- Reach underserved populations
- Quantitation
- Image Analysis
- Archiving, Search, Data Mining
- Integration with Patient Record

- There is a need for

Task-Based Evaluation of Whole Slide Imaging

- How well can humans perform pathology tasks given WSI?
- How good is one WSI scanner compared to another?
- Reference:

  glass slide and microscope

Goals:

- Compare Digital and Optical Pathology
- Test the limits of the technology
  - Develop data collection methods
  - Develop data analysis methods
  - Execute reader studies

Big Picture Project Goals

Current Work Being Presented

Under development

Planned

- Identify sources of variability
- Eliminate or reduce variability
- Account for sources of variability
  - Readers = Pathologists
  - Cases = Patients, Slides, Regions of Interest, Cells

Study Design

Common Study

- Reflect Clinical Case Load
  - Multiple tasks
  - Multiple tissue types
- Entire case
  - Patient Info
  - multiple slides
  - different stains
- Primary Diagnosis
- Reference: Expert panel consensus
- Free text reports

Limitations

- Results depend on case mix
- Decisions based on different
  - information
  - slides
  - areas
- May not strongly depend on image quality
- Experts:
  qualifications, recruitment, consensus
- Adjudication Panel

Study Design

Common Study

- Reflect Clinical Case Load
  - Multiple tasks
  - Multiple tissue types
- Entire case
  - Patient Info
  - multiple slides
  - different stains
- Primary Diagnosis
- Reference: Expert panel consensus
- Free text reports

Target Study

- Single narrow task
  - Find and count target cells
  - Classify individual cells
- Reduce information to evaluate
  - No patient info
  - Small Region of Interest (ROI)
  - Single Cell
- Task chosen to stress imaging
- Agreement with qualified pathologists on reference modality
- Scores directly and objectively amenable to analysis

Outline

- **eeDAP**
  - *evaluation environment for Digital and Analog Pathology*
  - Software and Hardware

  - **KEY**

  Registration between digital image and microscope slide

  - **Registration Reduces or Eliminates Variability**

  Search Training, Search Ability, Evaluation Area

- Feasibility studies
  - design, results, problems, solutions

Tasks for Feasibility Studies

Classification

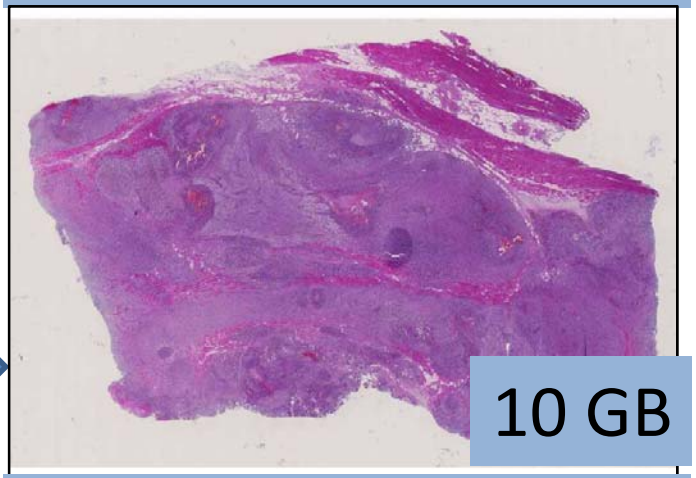
- Plasma cell or not
- Source: Colon

2 GB

9 mm x 11.7 mm
51,200 x 39,680 pixels

scanner “resolution”
pixel size: 0.23 um at 40x

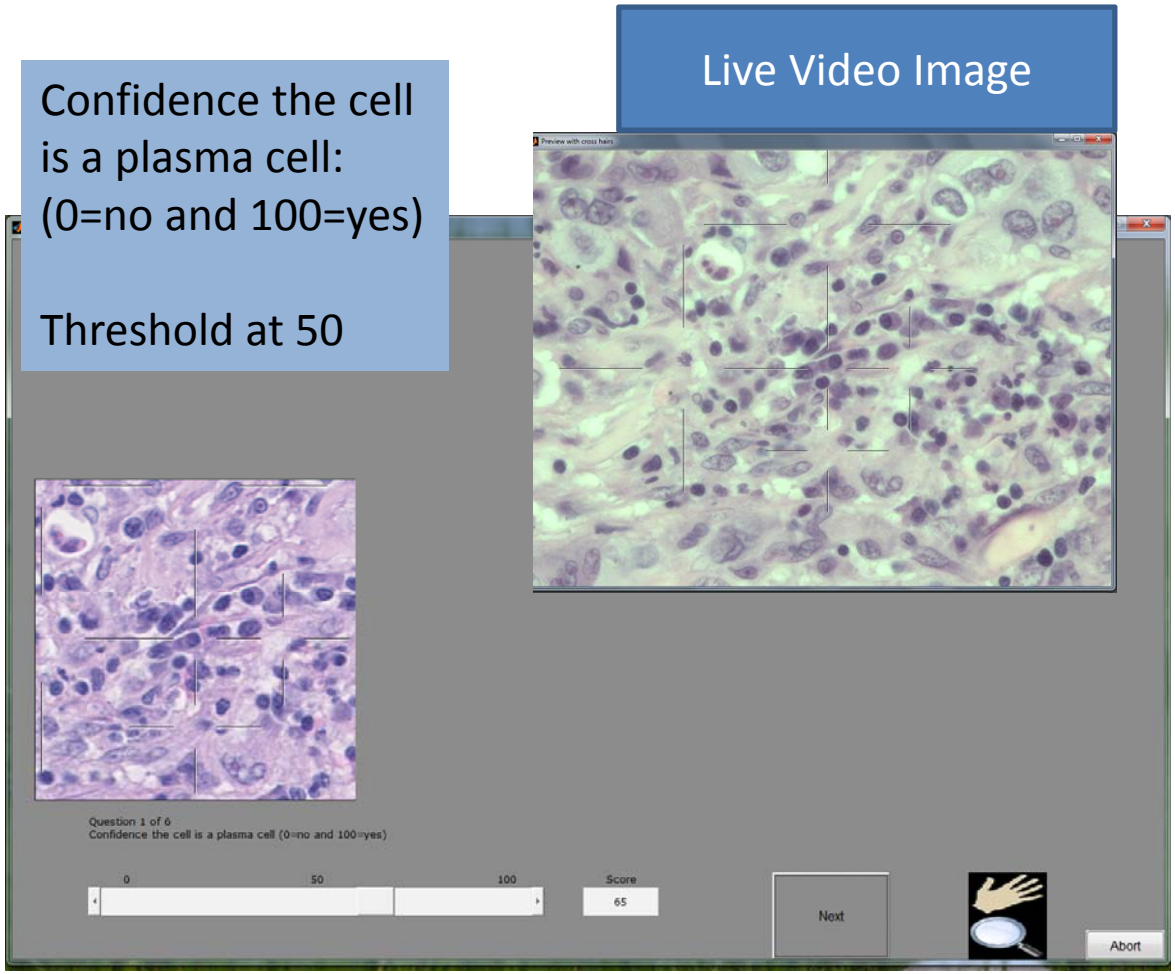
- Mitotic figure or not
- Source: Sarcoma



10 GB

28.0 mm x 18.8 mm
123,008 x 82,688 pixels

eeDAP Data Collection GUI



- *Digital Mode*
- *MicroRT Mode*

Pathologist
- Engaged with microscope

Administrator
- Operates GUI
- Enters Data
- Checks and Maintains Good Registration
- Video Image of Microscope View
- Button to register

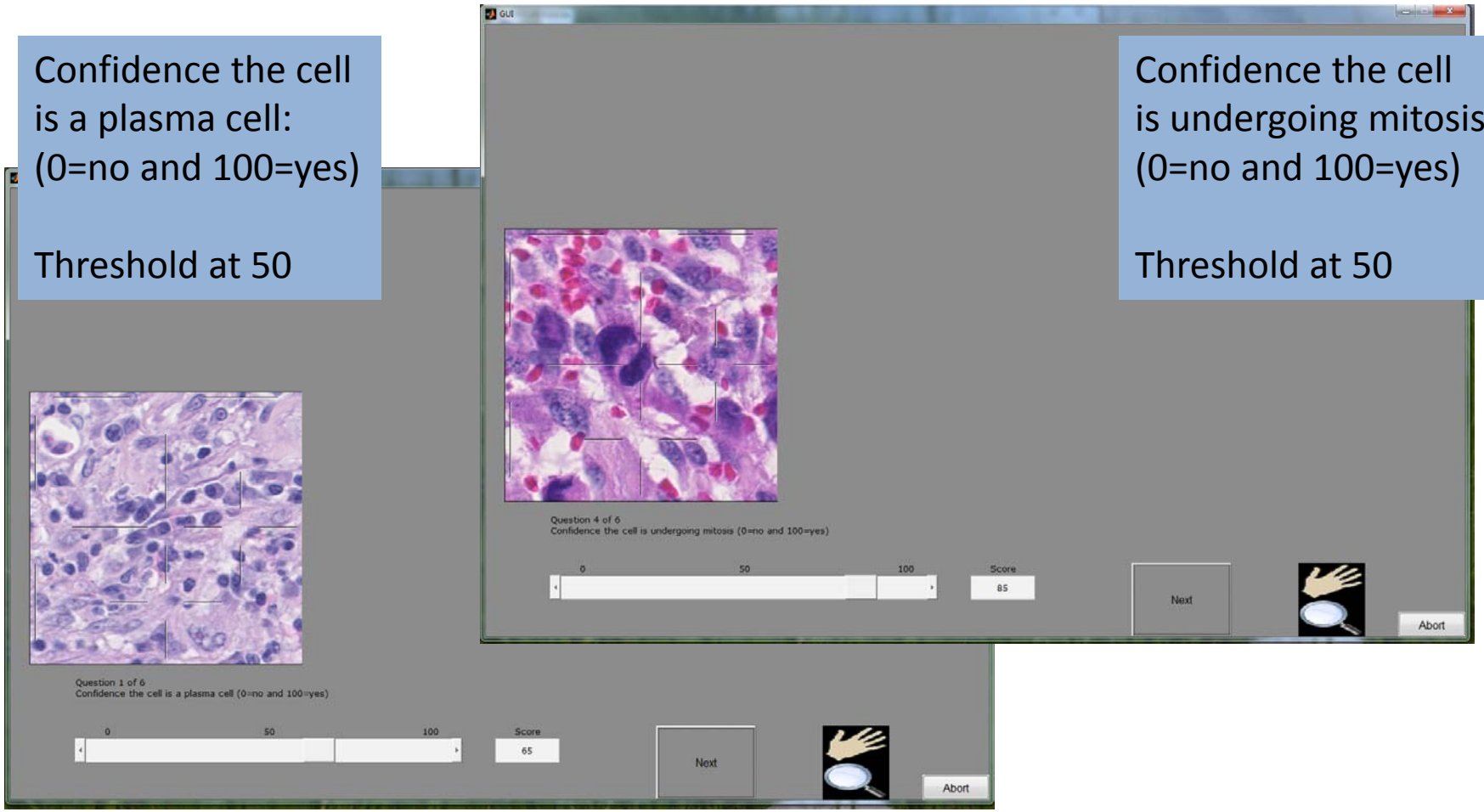
eeDAP Data Collection GUI

Confidence the cell is a plasma cell: (0=no and 100=yes)

Threshold at 50

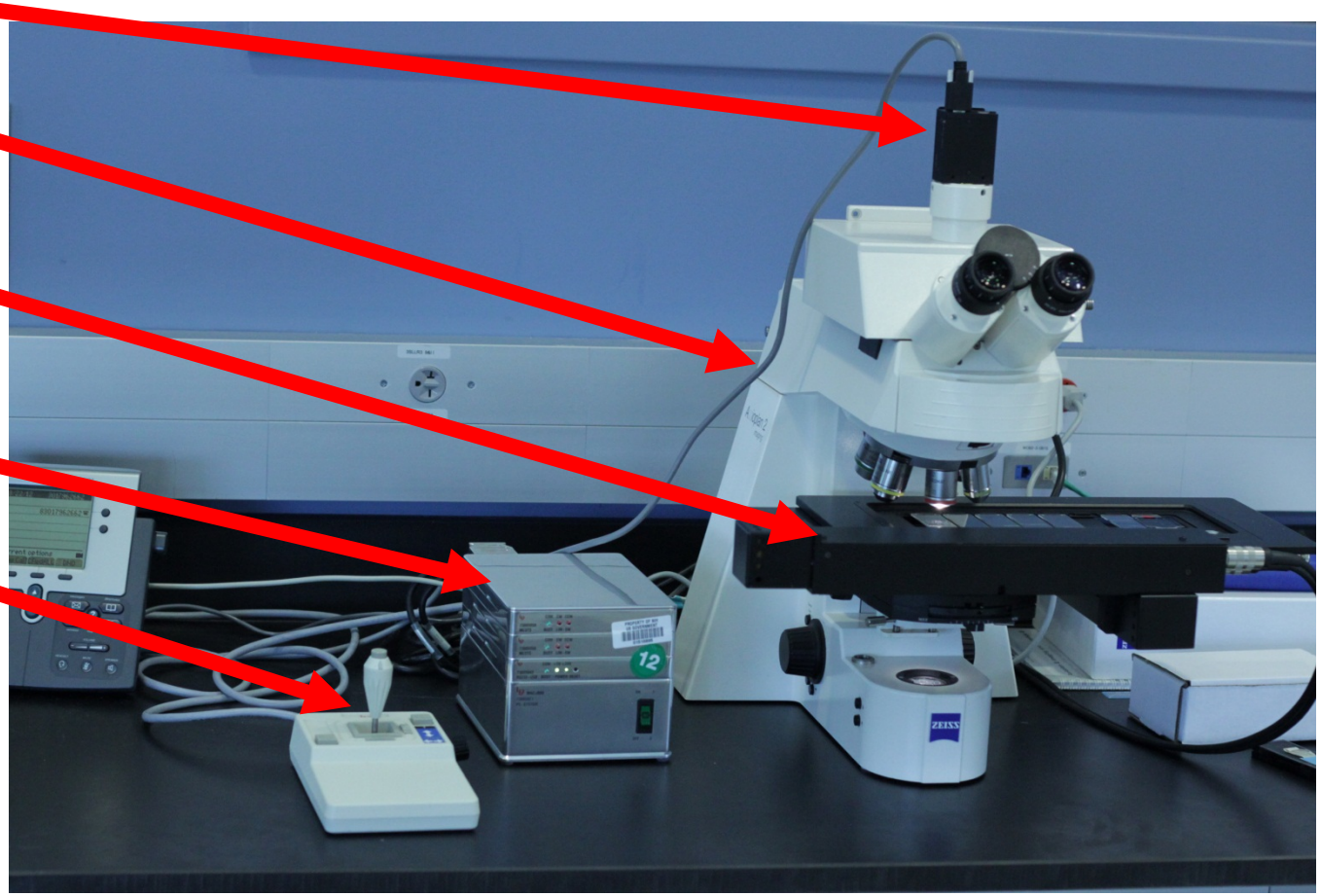
Confidence the cell is undergoing mitosis: (0=no and 100=yes)

Threshold at 50



eeDAP Hardware

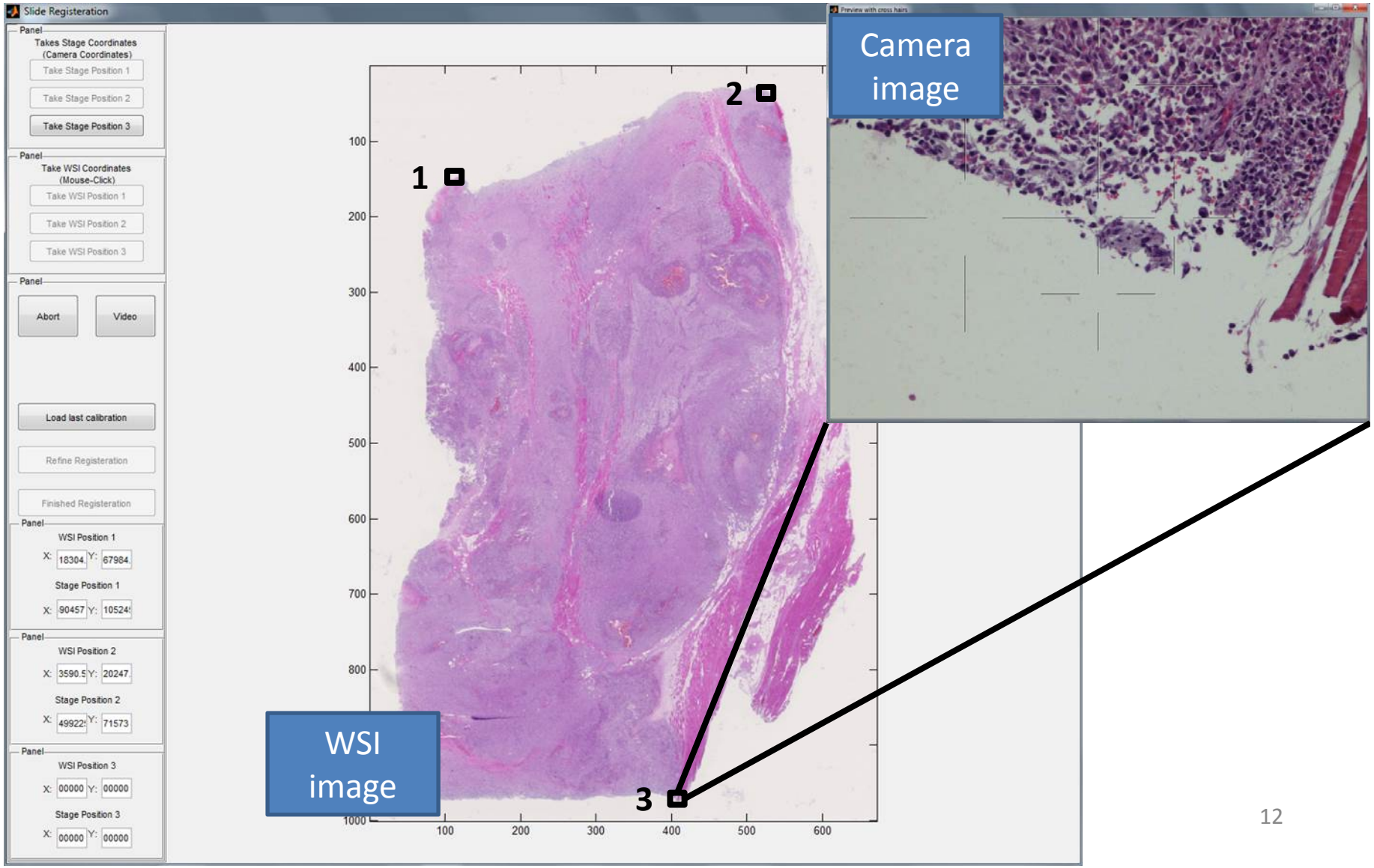
- Camera
- Microscope
- Moving stage with multiple slides
- Stage controller
- Joystick for stage control
- Computer and monitor not shown



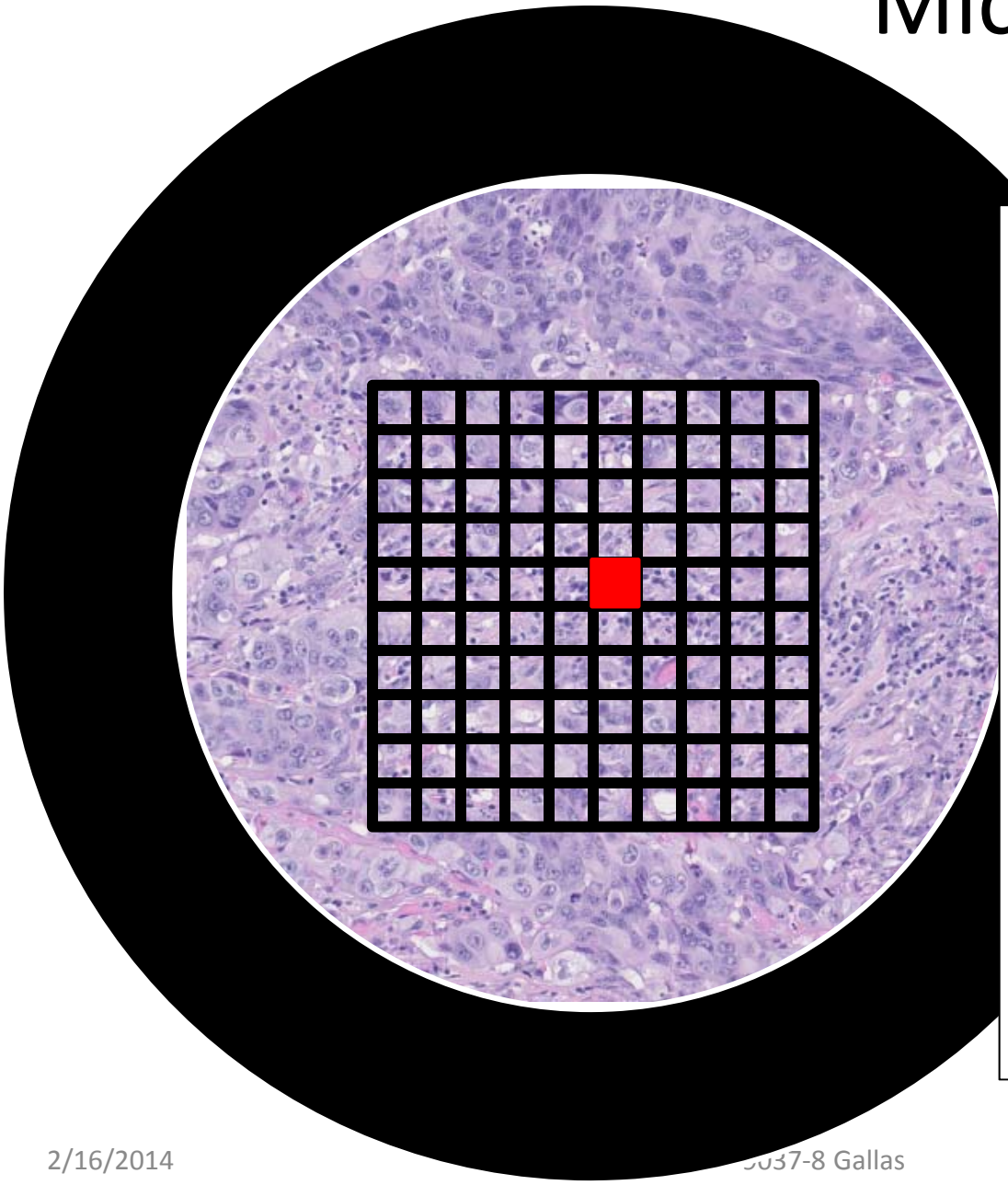
eeDAP Registration
Global

- Global registration = mapping between
  - WSI coordinates
  - Glass Slide coordinates = Stage Coordinates

eeDAP Registration GUI
3 Local → Global

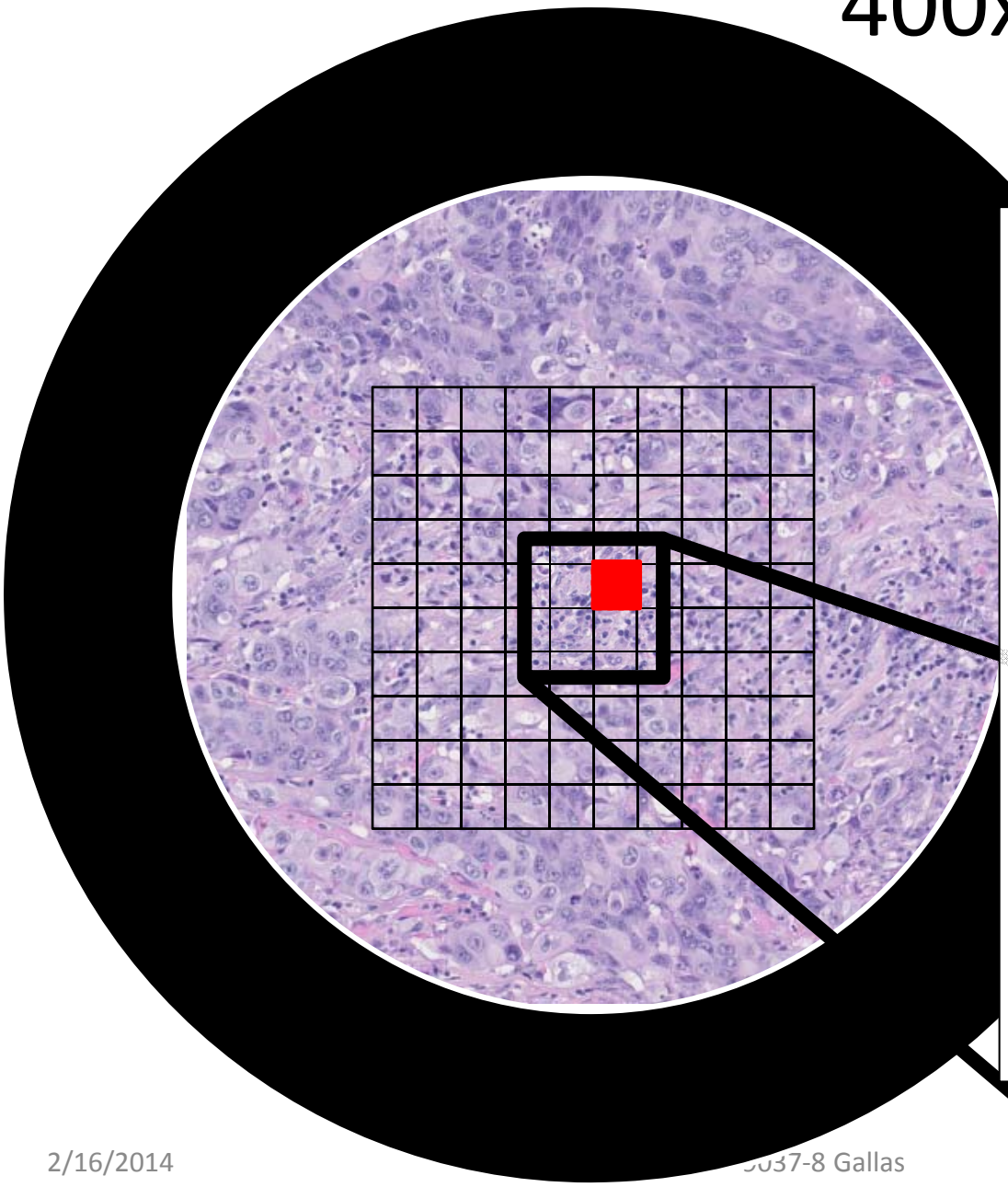


Microscope Image with Reticle

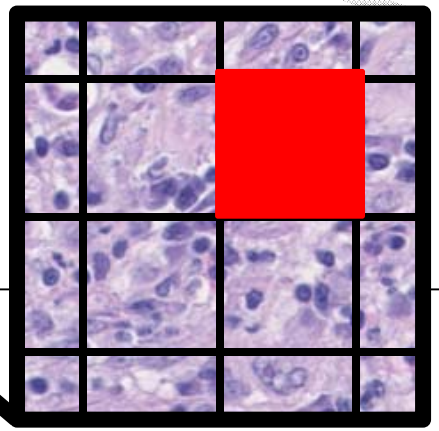


- Allows evaluation of small ROIs (regions of interest)
- Lines are exaggerated
- Lives in the Eyepiece

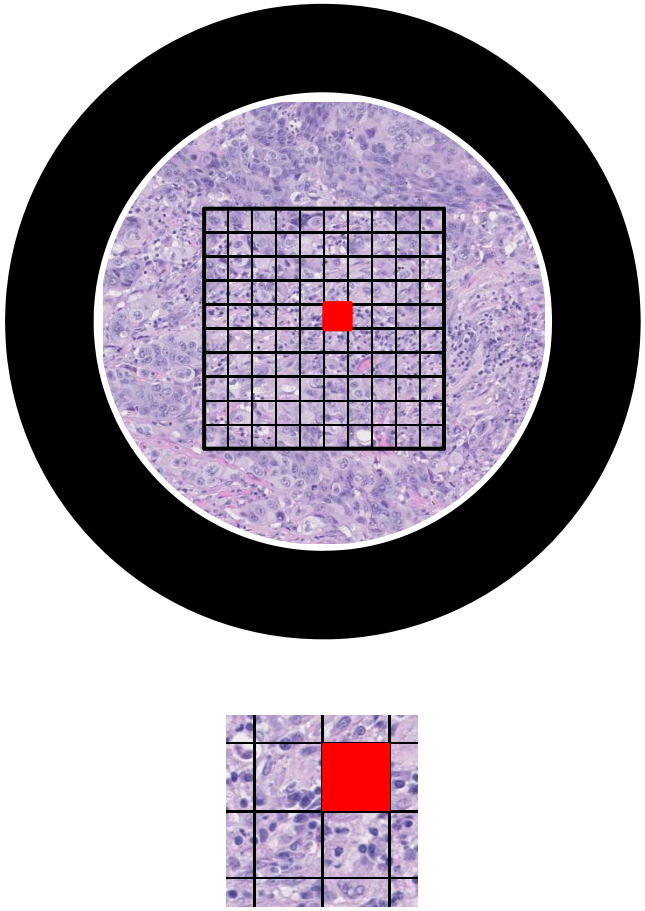
400x400 WSI Patch with Reticle



- Allows evaluation of small ROIs (regions of interest)
- Lines are exaggerated
- Digital Overlay



Feasibility Studies, Plasma Cells



- Purpose: Identify problems
  - study design
  - reading protocol
  - imaging protocol
  - training
  - software bugs
- Score: Yes/No and 101 point scale
  - center ROI
  - cells with more than 50% nucleus inside ROI
  - score most likely candidate
- 50 cases
  - 25 expert identified plasma cells
  - 25 selected random ROIs

Agreement by Concordance

Experiment: Two pathologists score two cases

Possible Results:

1. Concordance

Cases ranked in same order

2. Discordance

Cases ranked in opposite order

3. Tie by first doctor
4. Tie by second doctor
5. Tie by both doctors

Summarize:

- Rate of getting any kind of tie
- Rate of concordance conditional on no ties
  - Random = 0.5
  - Perfect = 1.0

- Work in progress
  - Investigating summary measures
  - MRMC variance analysis

Feasibility Studies, Plasma Cells

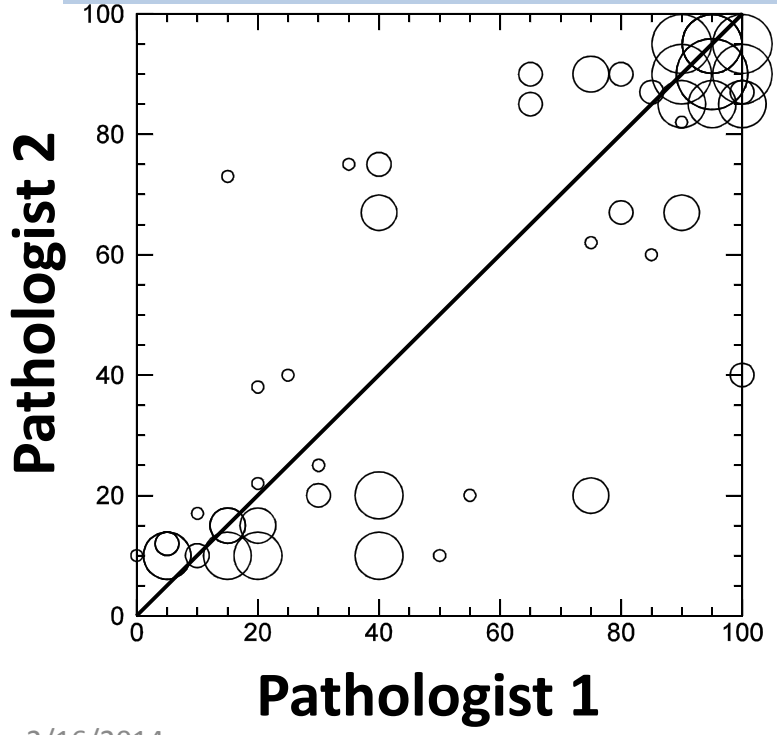
Inter-Reader, Intra-Modality

Joint distribution of scores

- 50 cases

Size of circle

- proportional to frequency



Microscope Mode

- Rate of Any Tie = 0.12
- Rate of Concordance = 0.84

- Good concordance

Feasibility Studies, Plasma Cells

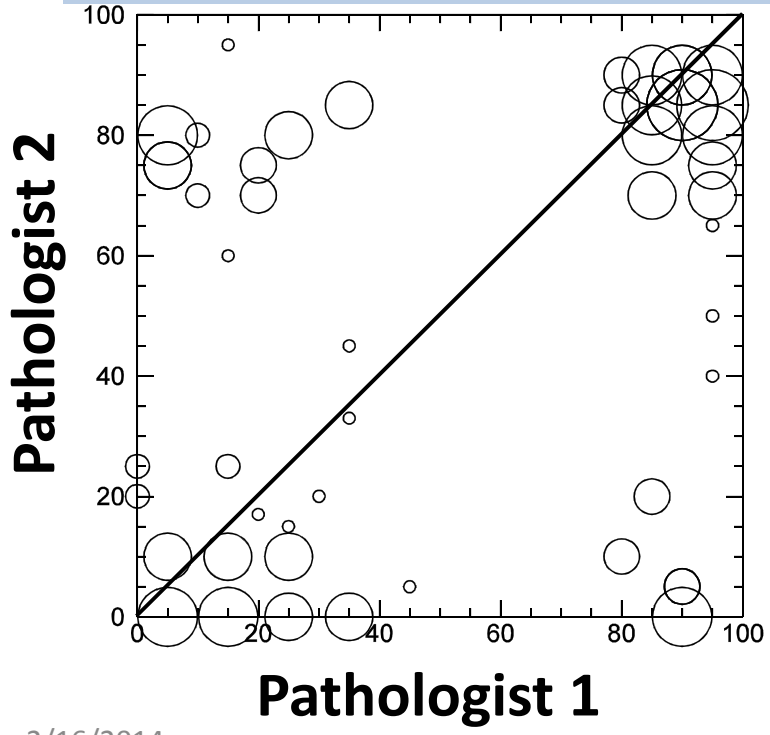
Inter-Reader, Intra-Modality

Joint distribution of scores

- 50 cases

Size of circle

- proportional to frequency



Digital Mode

- Rate of Any Tie = 0.13
- Rate of Concordance = 0.59

- Very poor concordance

Feasibility Studies, Plasma Cells

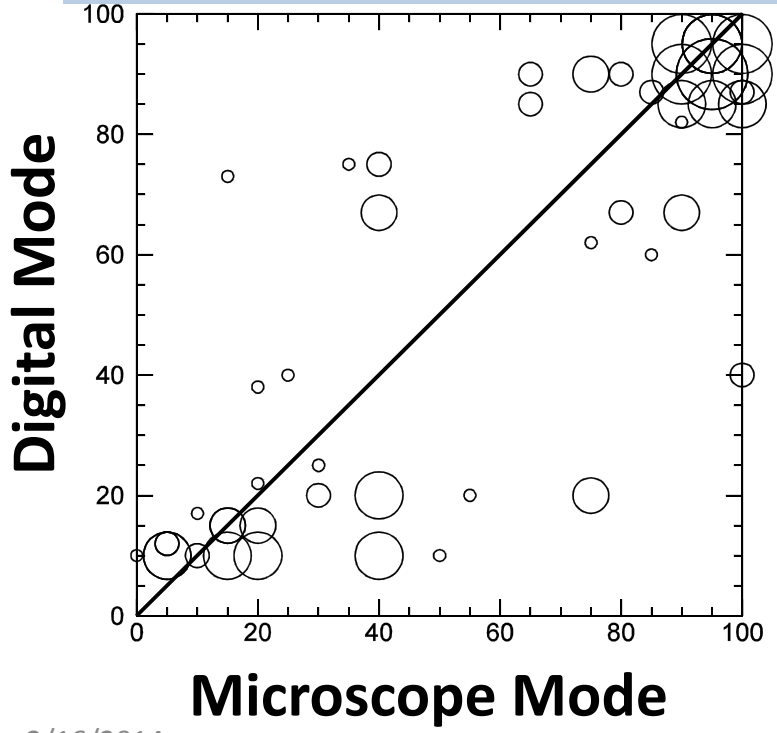
Intra-Reader, Inter-Modality

Joint distribution of scores

- 50 cases

Size of circle

- proportional to frequency



Pathologist 1

- Rate of Any Tie = 0.12
- Rate of Concordance = 0.67

- Poor concordance

Feasibility Studies, Plasma Cells

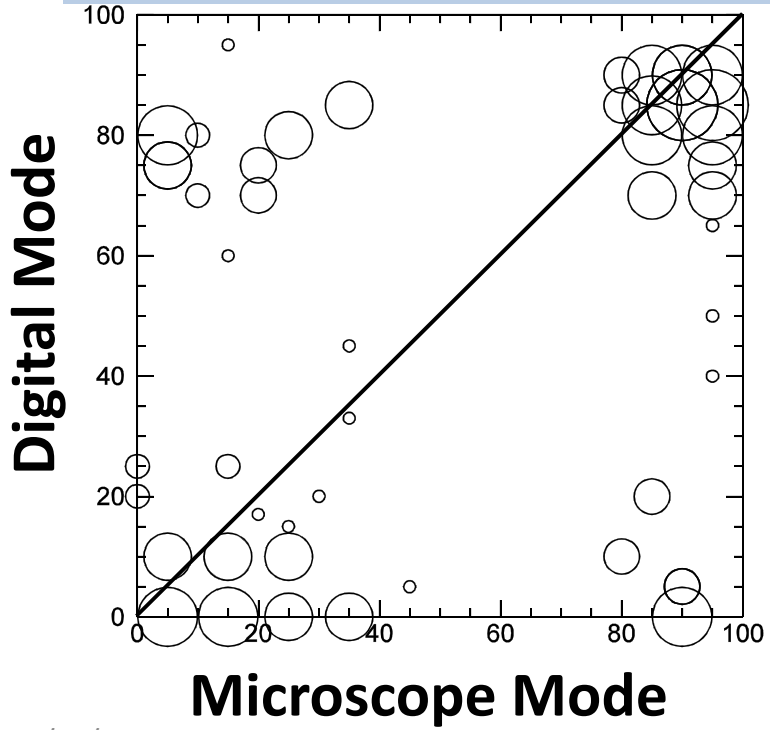
Intra-Reader, Inter-Modality

Joint distribution of scores

- 50 cases

Size of circle

- proportional to frequency



Pathologist 2

- Rate of Any Tie = 0.13
- Rate of Concordance = 0.58

- Very poor concordance

Feasibility Studies
Summary of Results

- Concordance Rates
  - Plasma Cell Detection: Not very good
  - Mitotic Figure Detection: Worse (not shown)

- Investigated Disagreements Case-by-Case
  - Feedback from pathologists helpful
  - Very few fundamental disagreements
  - Data collection method caused disagreement

Data Collection Problems and Fixes

Goal: Eliminate Disagreements not related to the task and image quality

- *Problem*: Scoring ROI adds variability
- *Solution*: Better reticle, point at single cells

- *Problem*: Registration precision degrades as stage moves
- *Solution*: Register at each ROI

- *Problem*: Pathologists disagree
- *Solution*: Better training and better instructions

- *Problem*: Scores bunch at extremes
- *Solution*: Training on scoring

Conclusions & Status Update

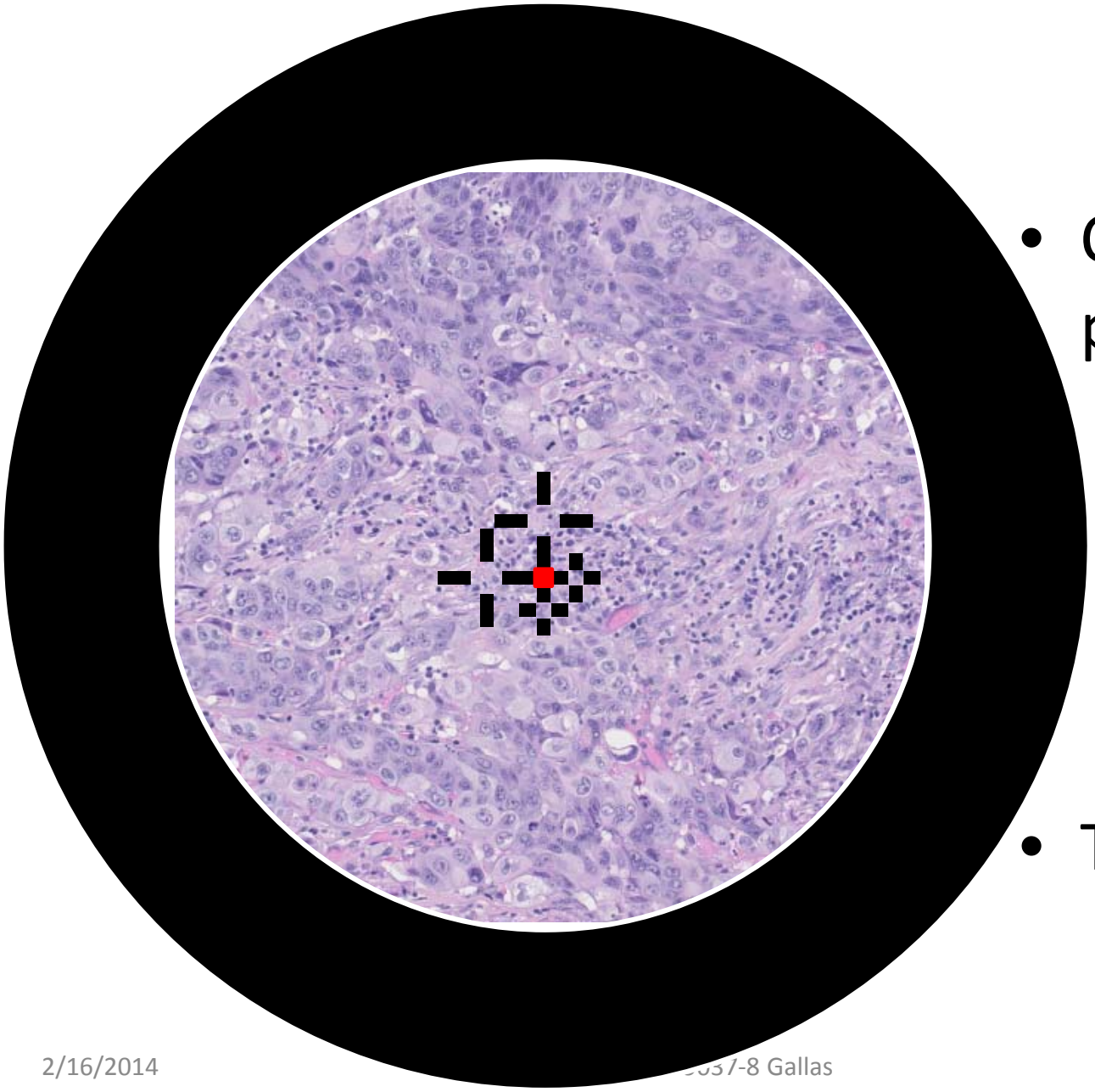
- Early feasibility studies completed
  - Problems with data collection
- Improvements have been made
  - Reticle with cross-hairs
  - Register every ROI
  - Better training and instructions: objects and scoring
- Improvements being tested
  - Additional feasibility studies ongoing
  - Improved inter-reader agreement in digital mode
  - Successfully training novices

More Details

- Proceedings paper provides details
  - Software/Hardware Info
  - Registration Algorithms
  - Imaging specifications
  - Color: software dependent tone reproduction curves

- code.google.com
  - project “eeDAP”
  - Matlab source code
  - Precompiled, stand-alone, license free application

New Reticle



- Cross-hairs point at gaps

- Thank You!

Thank You

Feasibility Studies A, Plasma Cells

- Yes/No Threshold
  - at 50
- Marginals very close!
  - *IF* you don't collect paired observations at the level of an ROI
  - *THEN* you don't see disagreement.

Inter-reader

Micro		Obs 1		
		neg	pos	
Obs 2	neg	20	3	23
	pos	4	23	27
		24	26	

Inter-reader

Digital		Obs 1		
		neg	pos	
Obs 2	neg	16	7	23
	pos	11	16	27
		27	23	

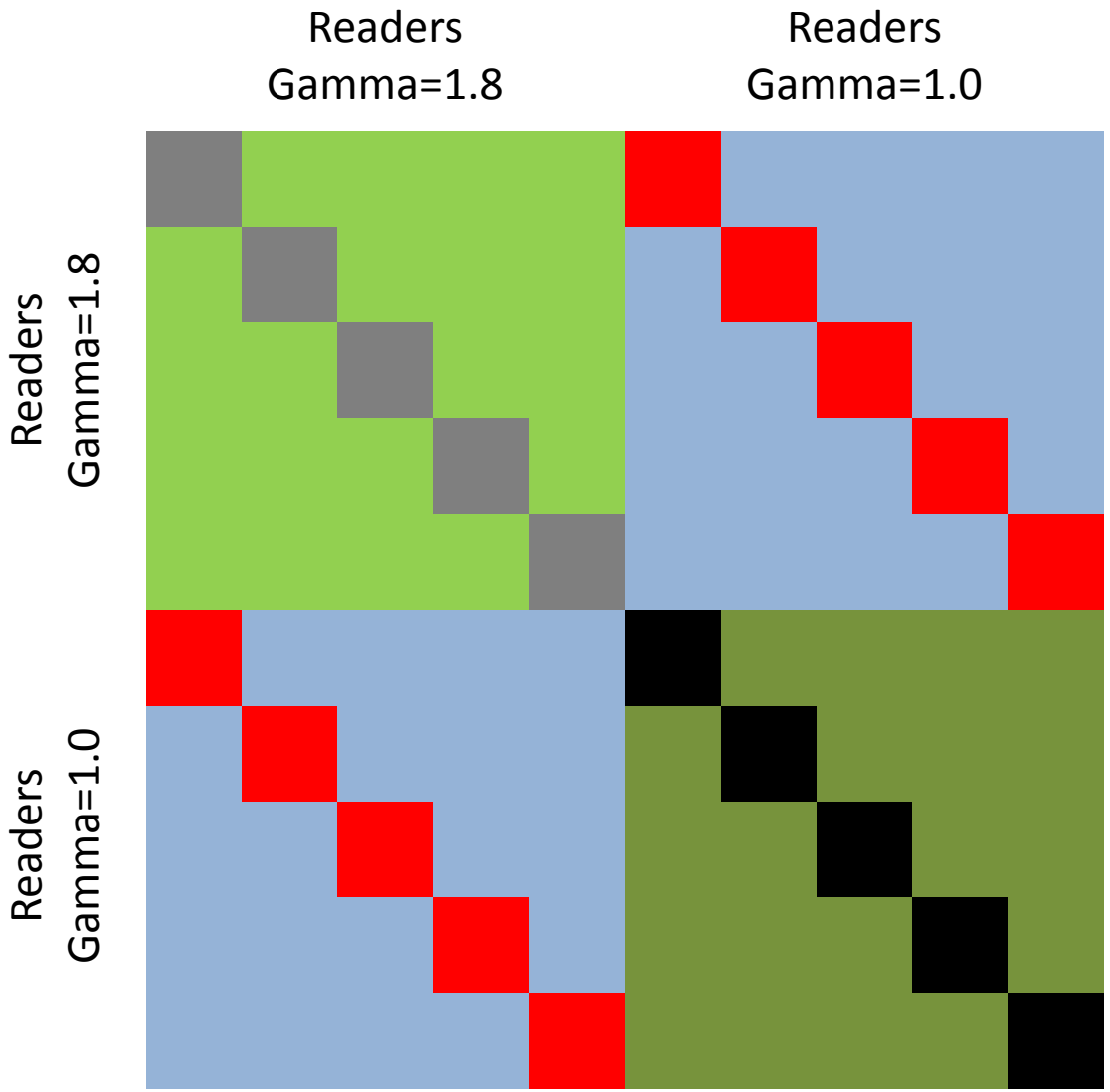
Inter-Modality

Obs 1		Digital		
		neg	pos	
Micro	neg	15	9	24
	pos	12	14	26
		27	23	

Inter-Modality

Obs 2		Digital		
		neg	pos	
Micro	neg	14	9	23
	pos	9	18	27
		23	27	

Comparing Scores



Intra-Reader
Gamma 1.8
Need Replicates

Intra-Reader
Gamma 1.0
Need Replicates

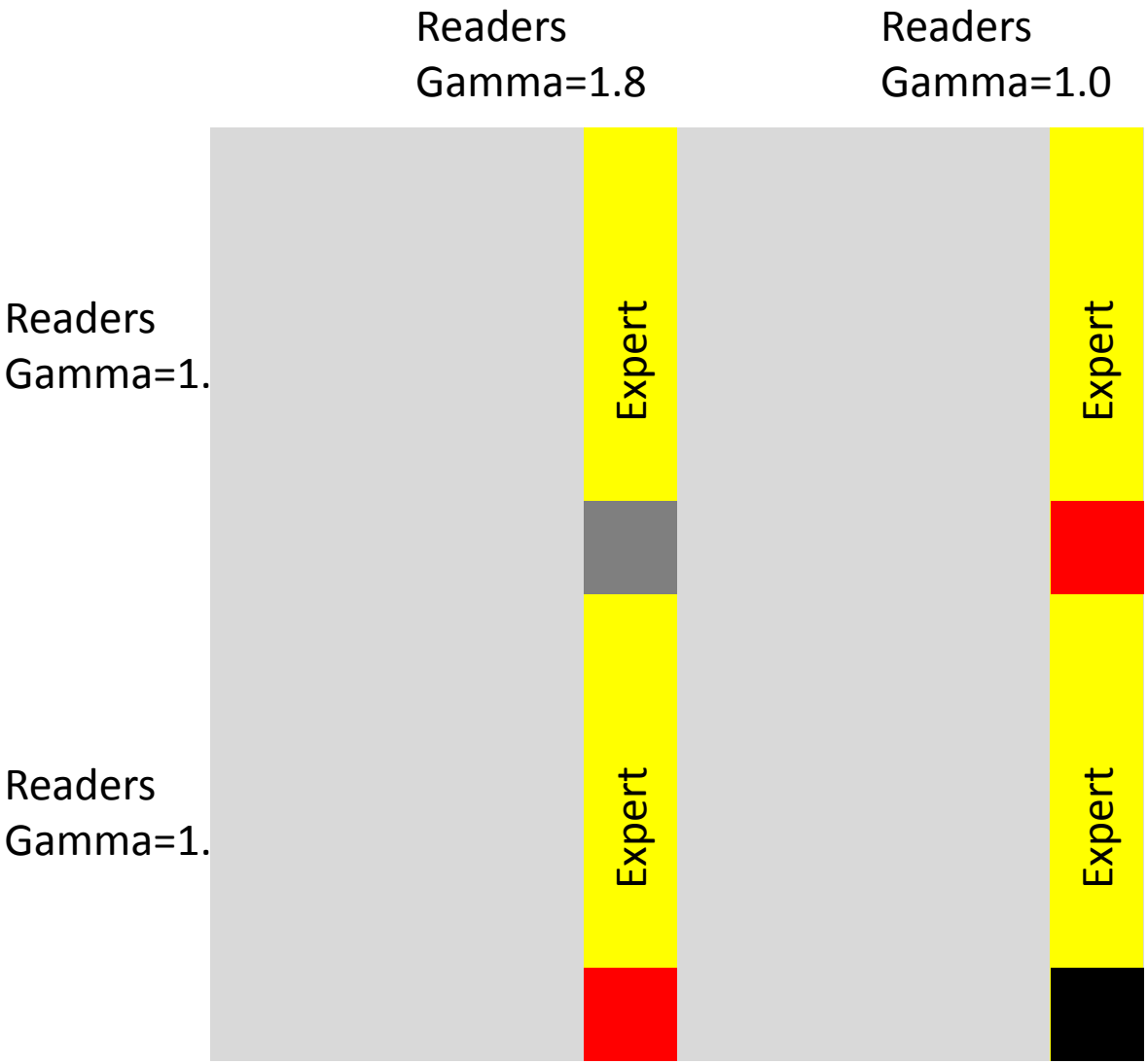
Intra-Reader
Inter-Modality

Inter-Reader
Gamma = 1.8

Inter-Reader
Gamma 1.0

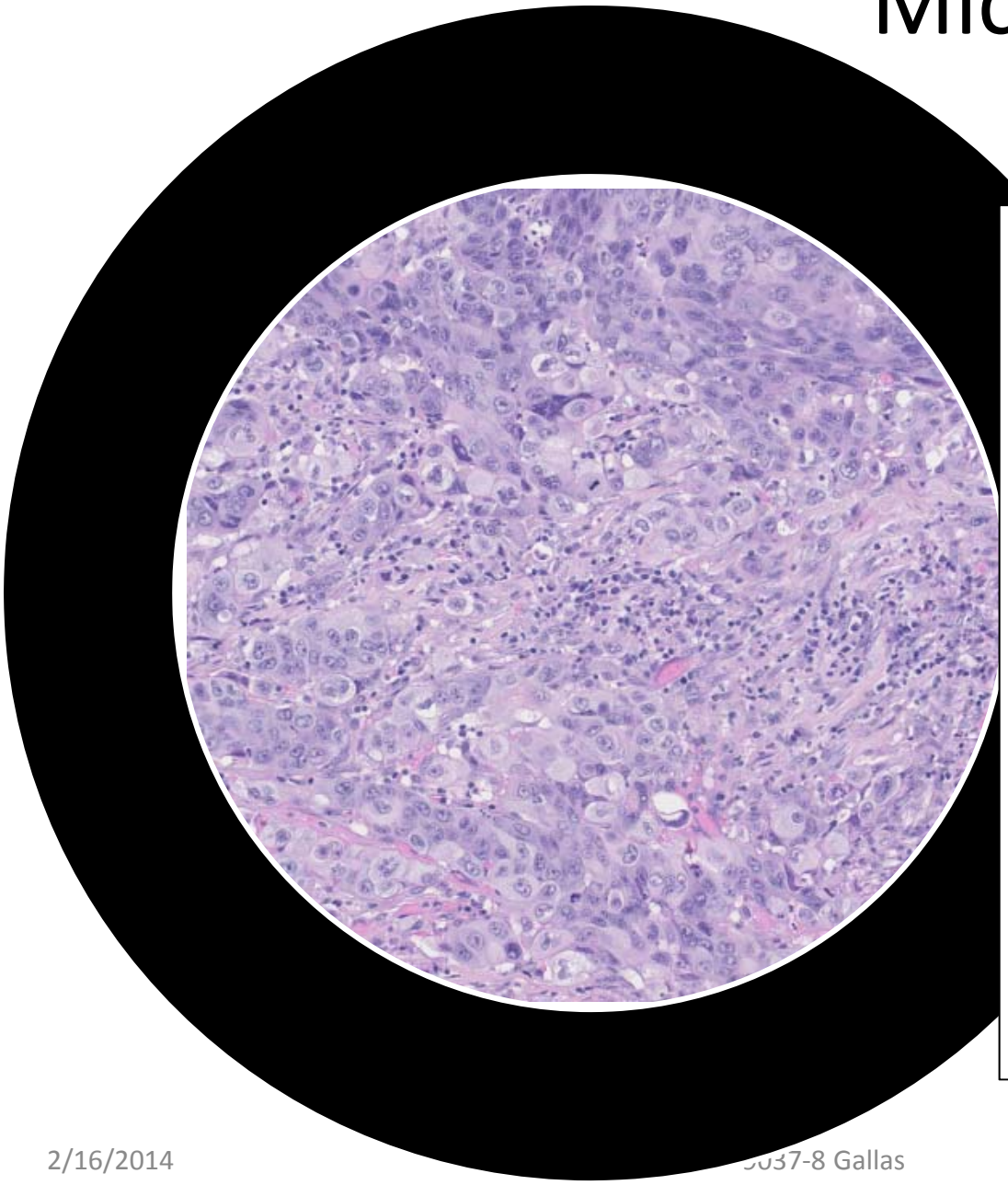
Inter-Reader
Inter-modality

Classify Plasma Cells



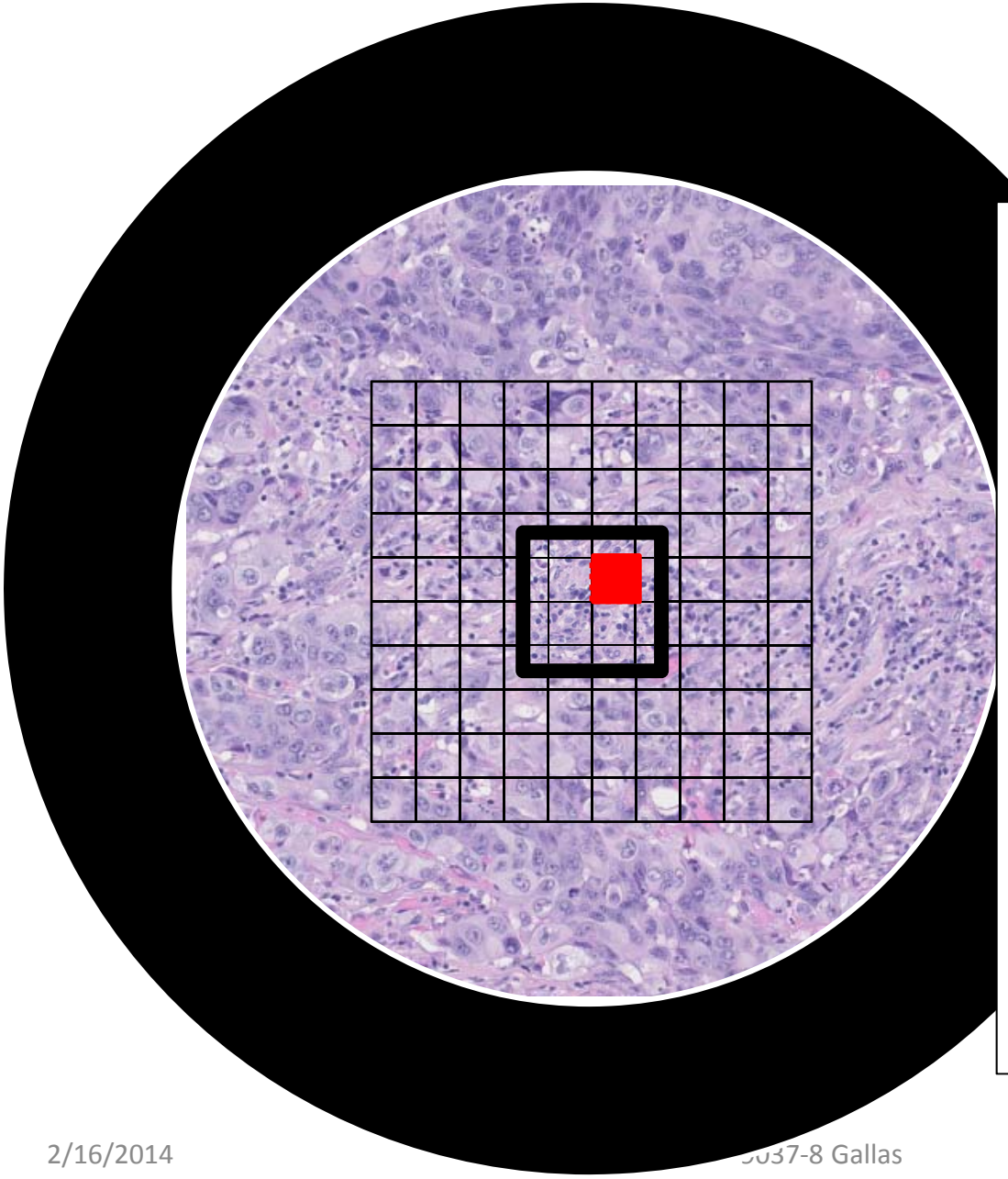
Agreement with expert

Microscope Image



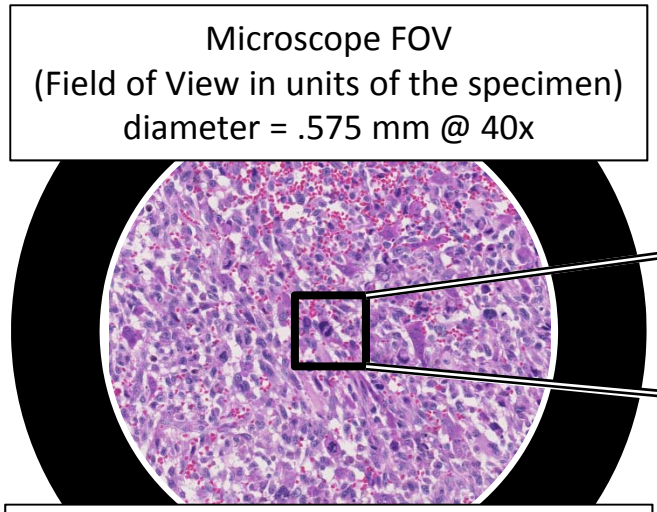
- Microscope FOV
  - Field of View
  - Units of the specimen
  - Diameter = .575 mm @ 40x
- Effective Image Size
  - As perceived by eye
  - 23 cm diameter
  - viewed at 25 cm

400x400 WSI Patch

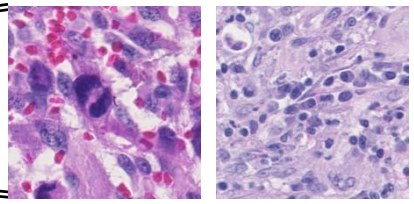


- Image Size (FOV)
  - Units of the specimen
  - width = 0.092 mm
- Perceived Image Size
  - computer monitor: 258 um pixels
  - width = 10.32 cm

Image Sizes
Microscope and Computer Monitor

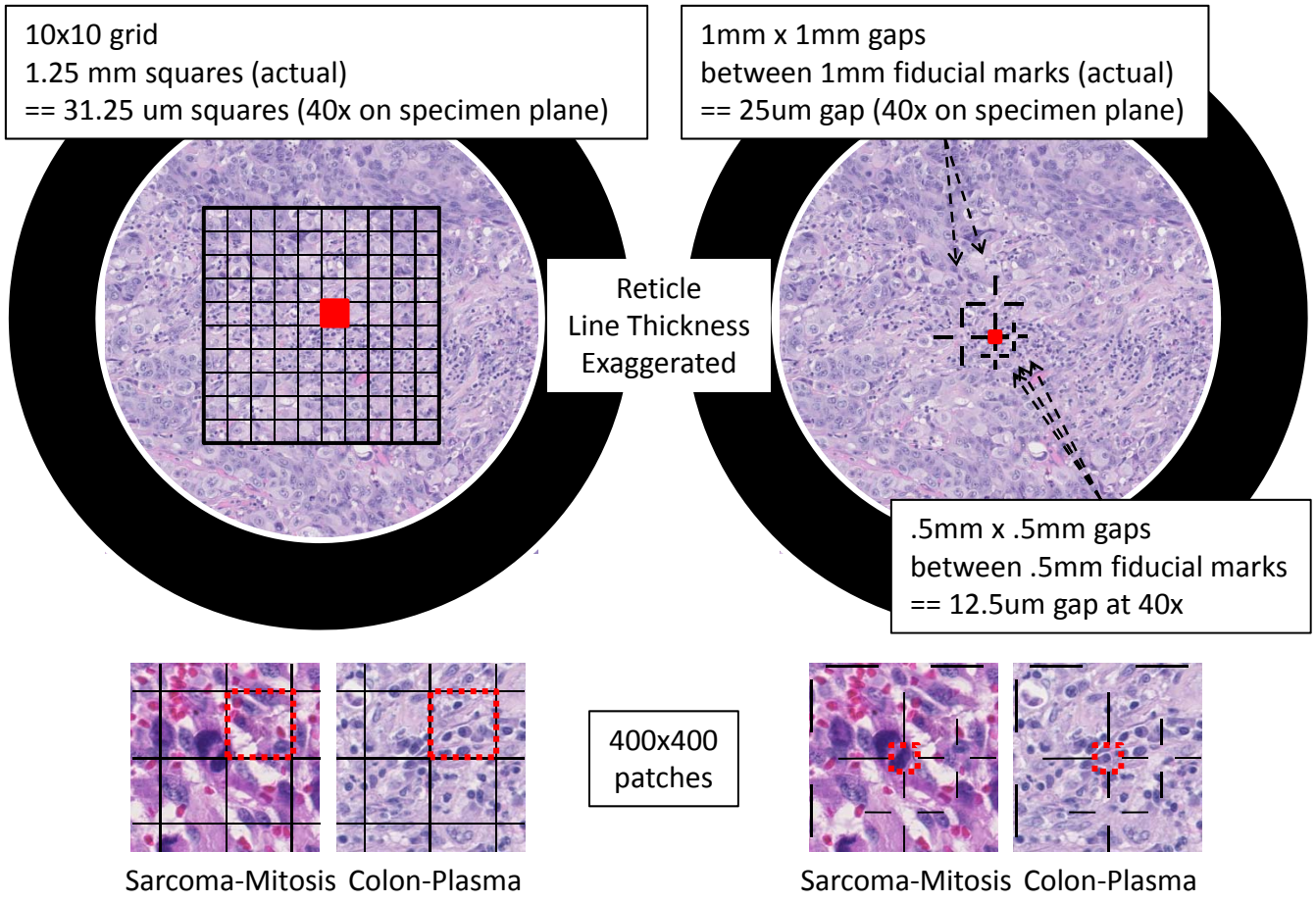


400x400 ROI patch of WSI
in units of the specimen:
width = 0.092 mm



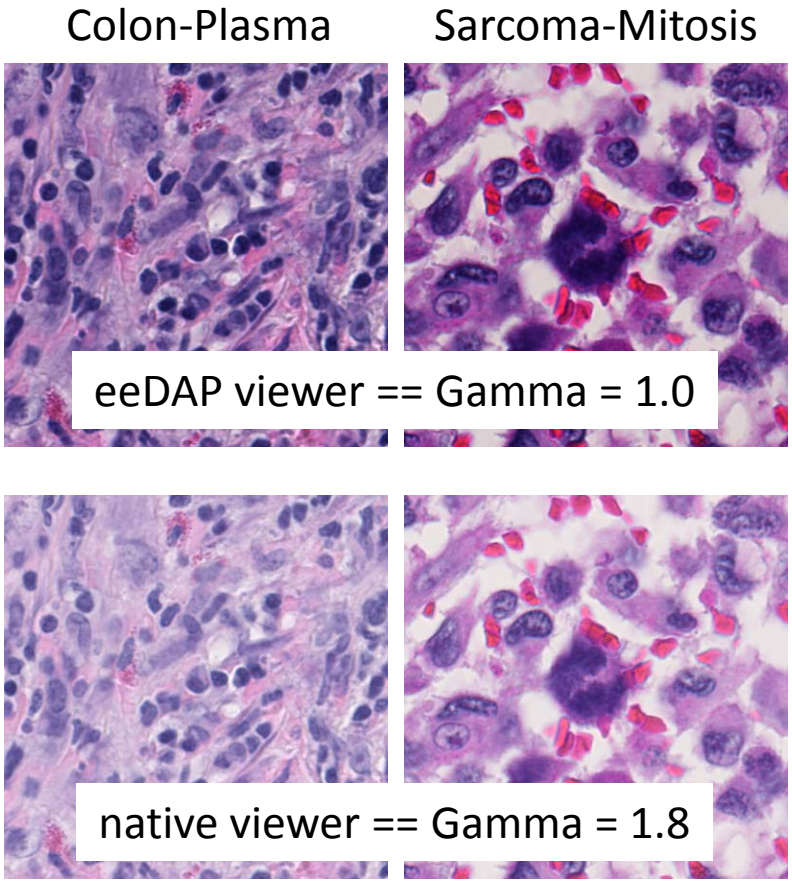
Sarcoma-Mitosis Colon-Plasma

400x400 ROI patch of WSI
computer monitor: 258 um pixels
ROI Size: width = 10.32 cm


10x10 grid
1.25 mm squares (actual)
== 31.25 um squares (40x on specimen plane)
1mm x 1mm gaps
between 1mm fiducial marks (actual)
== 25um gap (40x on specimen plane)
Reticle
Line Thickness
Exaggerated
.5mm x .5mm gaps
between .5mm fiducial marks
== 12.5um gap at 40x
400x400
patches
Sarcoma-Mitosis
Colon-Plasma
Sarcoma-Mitosis
Colon-Plasma

Color Differences Observed and Measured

- eeDAP viewer
  - generally dark
  - little contrast in nucleus
- native viewer
  (developed by scanner mfr)
  - brighter
  - more contrast in nucleus



Feasibility Studies B

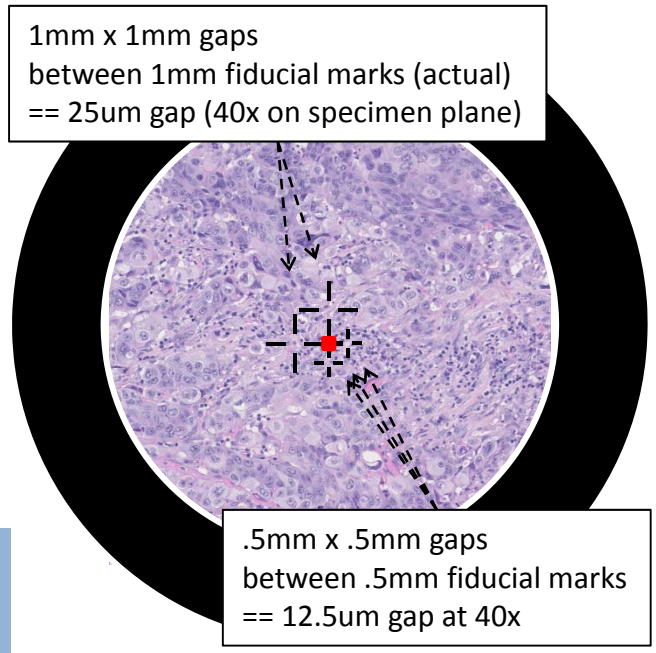
- Purpose: Identify problems
  - study design
  - reading protocol
  - imaging protocol
  - training
  - software bugs

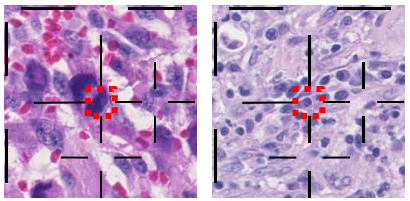
AND
begin to investigate color differences

- Score: Yes/No and 101 point scale

-- score cell who's nucleus is at center of cross hairs

- 50 cases
  - 25 selected as targets
  - 25 selected at random*





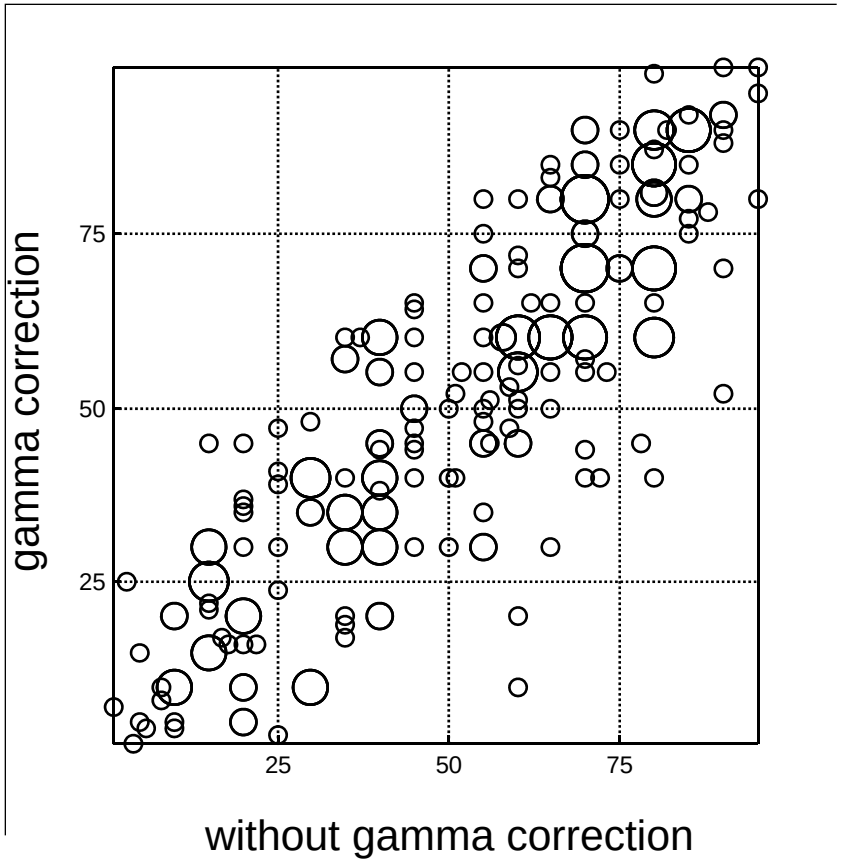
Sarcoma-Mitosis Colon-Plasma

Feasibility Studies B, Plasma Cells Intra-reader, Inter-modality

Scatter plots:
joint distribution of scores
Size of circle
proportional to frequency

- 250 pairs of data
  - 50 cases
  - 5 readers
- Good Agreement
- Nice distribution of scores

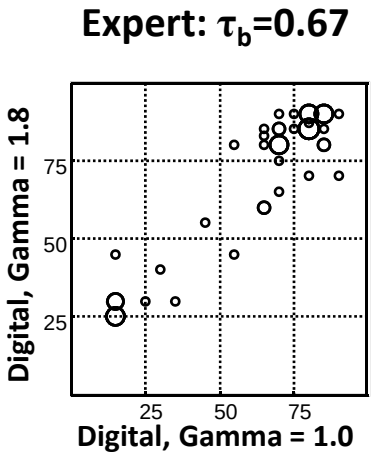
Average τ_b = 0.67

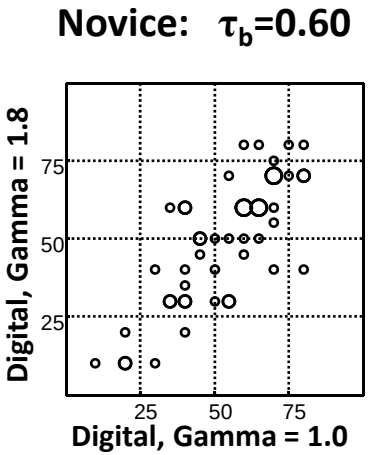


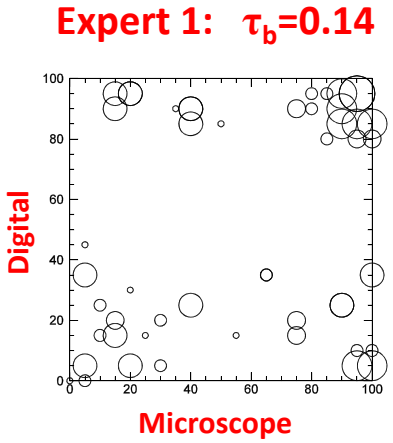
Feasibility Studies B, Plasma Cells Intra Reader (Expert and Novice)

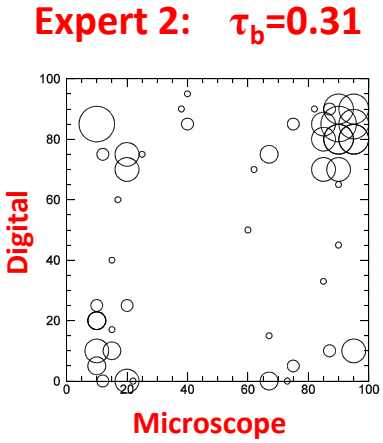
- Gamma = 1.0 vs. Gamma = 1.8
- Agreement by novice good
- Expert still bimodal

- Compare to Feasibility Studies A:
  - Digital only experiment







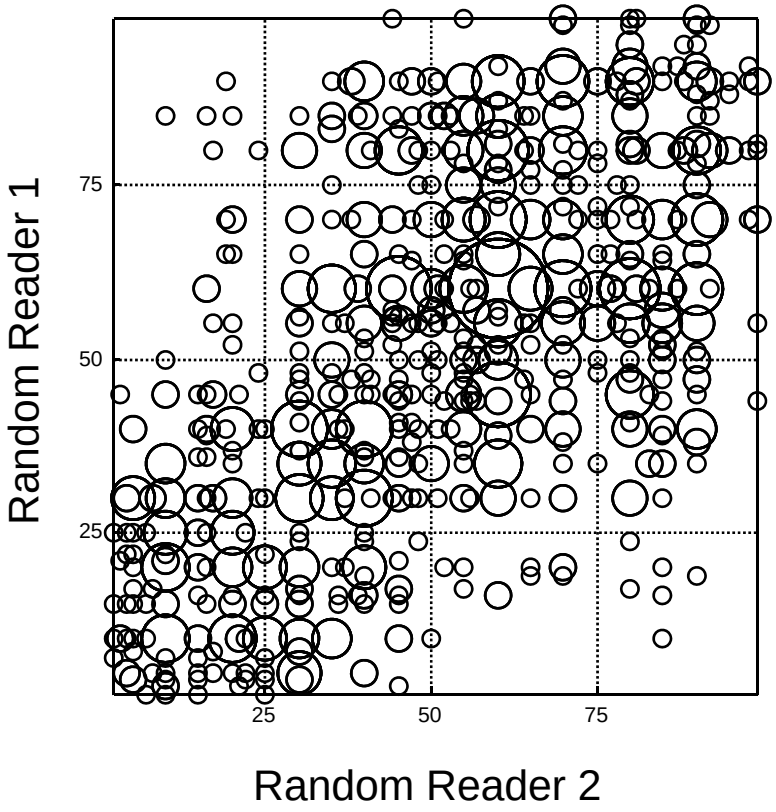


Feasibility Studies B, Plasma Cells
Inter-reader, Gamma = 1.8

Scatter plots:
joint distribution of scores
Size of circle
proportional to frequency

- 1000 observations
  - 50 cases
  - 20 pairs of readers
- Good Agreement
- Nice distribution of scores

Average τ_b = 0.53

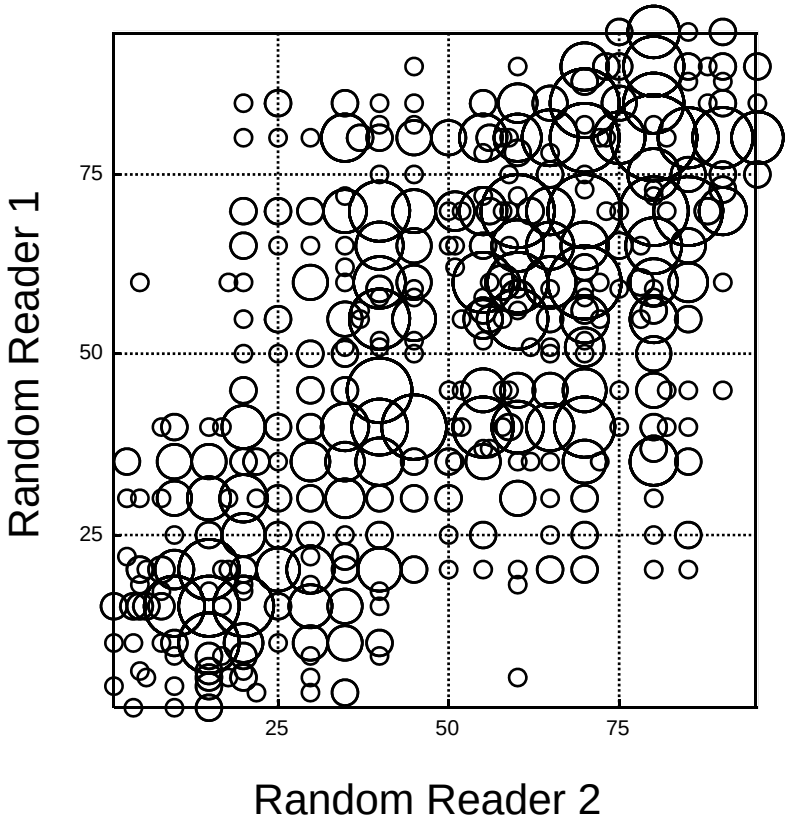


Feasibility Studies B, Plasma Cells
Inter-reader, Gamma = 1.0

Scatter plots:
joint distribution of scores
Size of circle
proportional to frequency

- 5 readers x 50 cases
- 1000 pairs of data
- Good Agreement
- Nice distribution of scores

Average τ_b = 0.58

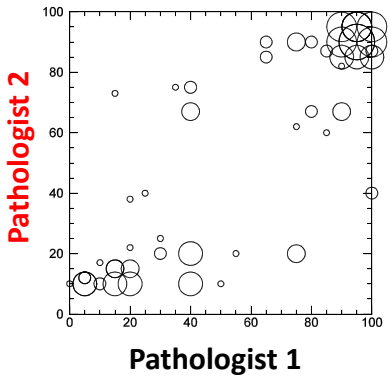


Feasibility Studies B, Plasma Cells
Inter-Reader (Expert vs. Novice)

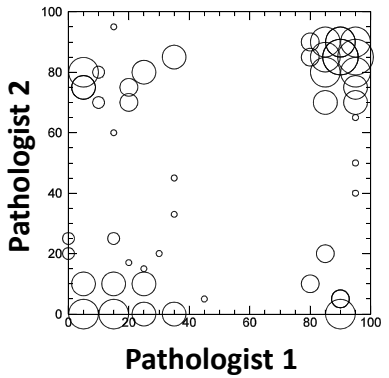
- Compare to Feasibility Studies A:
  - Digital only experiment
  - Pathologist vs. Pathologist

- Expert vs. Novice
- Agreement good

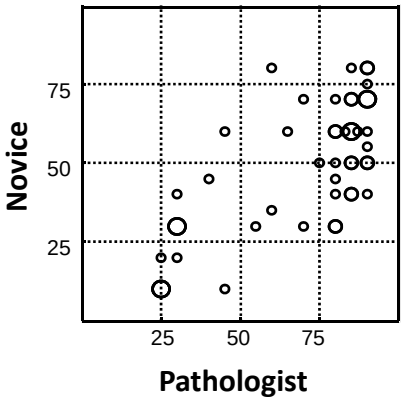
Microscope: τ_b=0.63



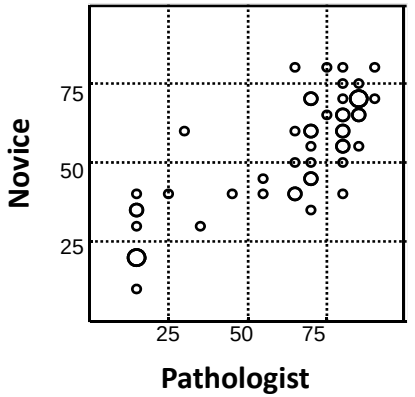
Digital: τ_b=0.17



Gamma = 1.8: τ_b=0.52



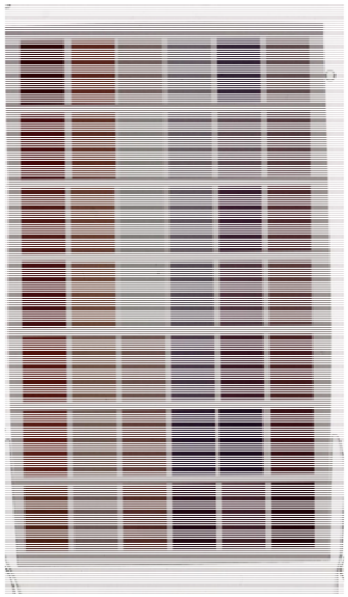
Gamma = 1.0: τ_b=0.61

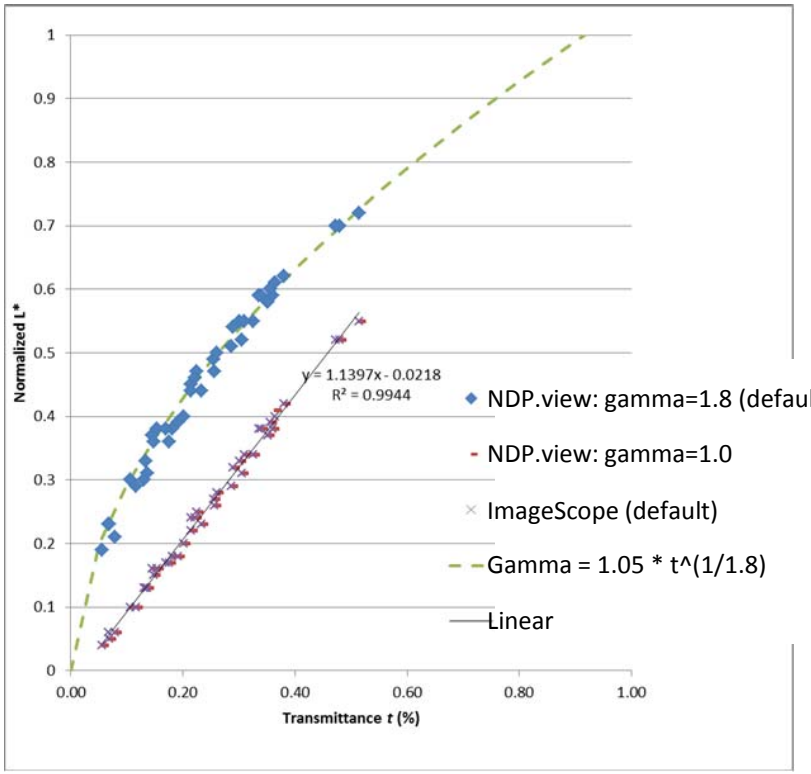


I measured the transmittance of the 42 color patches on the TY-2 color phantom as the truth. Then I retrieved the RGB values of the scanned image from 3 viewers/modes: NDP Standard, NDP Analysis, and Aperio. Then the RGB values were converted into CIELAB L* to generate the following plot. Two additional curves were added to fit the data.

The transmittance range of the 42 color patches is limited to [6%, 51%], so the data may not be very representative. However within the range, the NDP Analysis (linear) mode and the Aperio viewer are almost identical and very linear. The linear regression result is 1.14x – 0.02 with R=0.99, which means that the contrast is enhanced a little bit.

The NDP Standard mode can be fitted by a gamma curve of 1.05*t^(1/1.8), the same as what you predicted with again a little contrast increase.





Future Work

- Training
  - Instructions for 100 point scale
  - Need examples/descriptions of False Positives
  - Need feedback: training module
- Study Limitations
  - Single pair of readers
  - Cell selection biased
  - QA/QC for computer monitors (color calibration)