A Reader Study on a 14-head Microscope

Brandon D. Gallas, Qi Gong
FDA/CDRH/OSEL/DIDSR, Silver Spring, MD, US

Jamal Benhamida, Matthew G. Hanna, S. Joseph Sirintrapun, Kazuhiro Tabata, Yukako Yagi
Memorial Sloan Kettering Cancer Center (MSKCC), Pathology Informatics, New York, NY, US

Partha P. Mitra
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, US

A Reader Study on a 14-head Microscope

Purpose

- Purpose of this work
  - Demonstration … proof of concept … technology demonstration … method development

- Technology evaluation, not clinical performance

  Task-based evaluation of image quality
  - Task: Detection and classification of mitotic figures (MFs)
  - Images: Glass slides and WSI
  - Readers: Pathologists
  - Performance: Within- and Between-Reader Agreement

Clinically relevant task
Part of every pathologist's training
Challenging task
(substantial reader variability)
Convenient samples

Agreement … No ground truth

Count differences (calibration)
Pairwise Concordance (correlation)

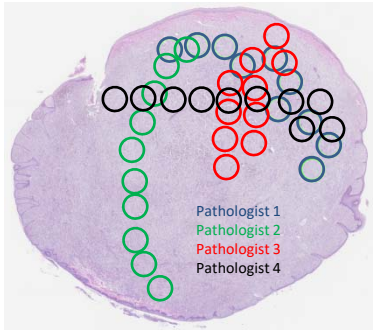
"MRMC" analyses
account for variability from
Multiple **R**eaders and **M**ultiple **C**ases

Microscope still the gold standard

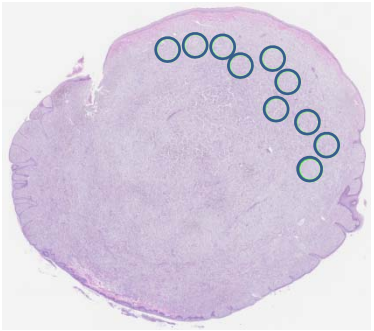
Remove search from technology evaluation

- Eliminate location variability for faster and more precise results.

<u>Clinical practice</u>
Pathologists choose
Fields of View to evaluate

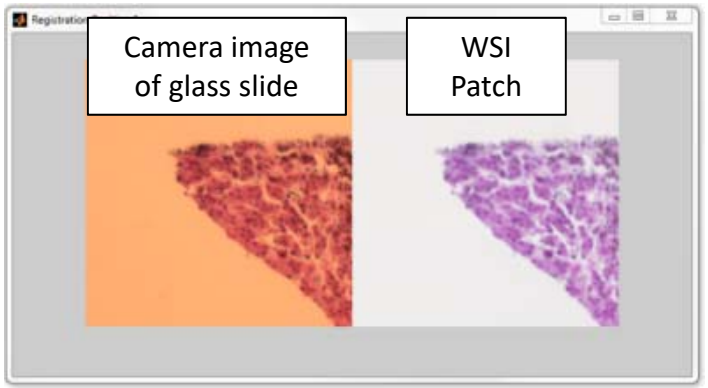


<u>Technology Evaluation</u>
All pathologists evaluate
same Fields of View



eeDAP: Evaluation Environment for Digital and Analog Pathology

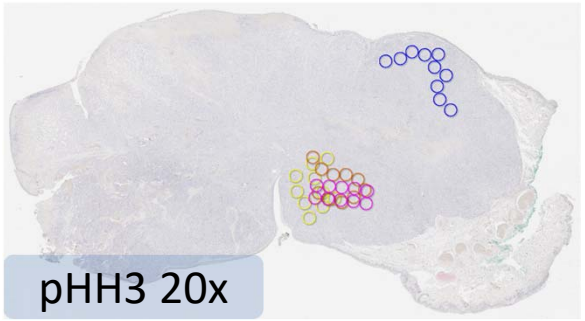
- eeDAP: Evaluation Environment for Digital and Analog Pathology
- Registration allows pathologists to evaluate the same fields of view

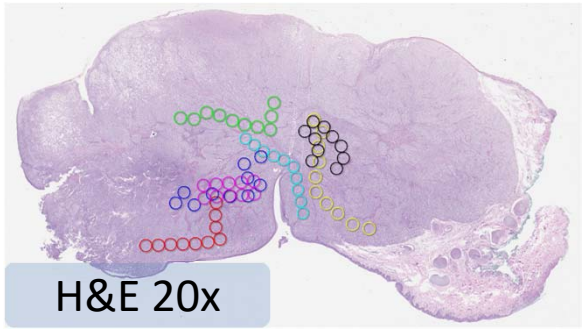


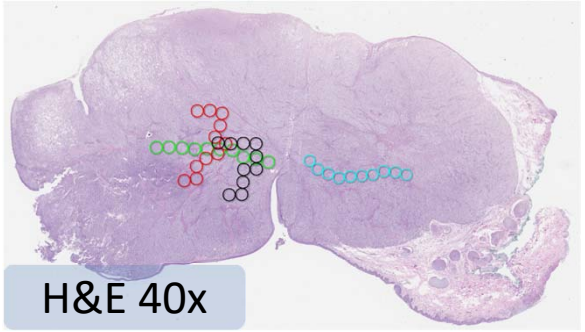
NIH Mitotic Counting Study

- NIH Study data (Mark Simpson)
  - FOV locations saved for each pathologist in digital mode
  - Preliminary agreement results given during WSIWG meeting

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

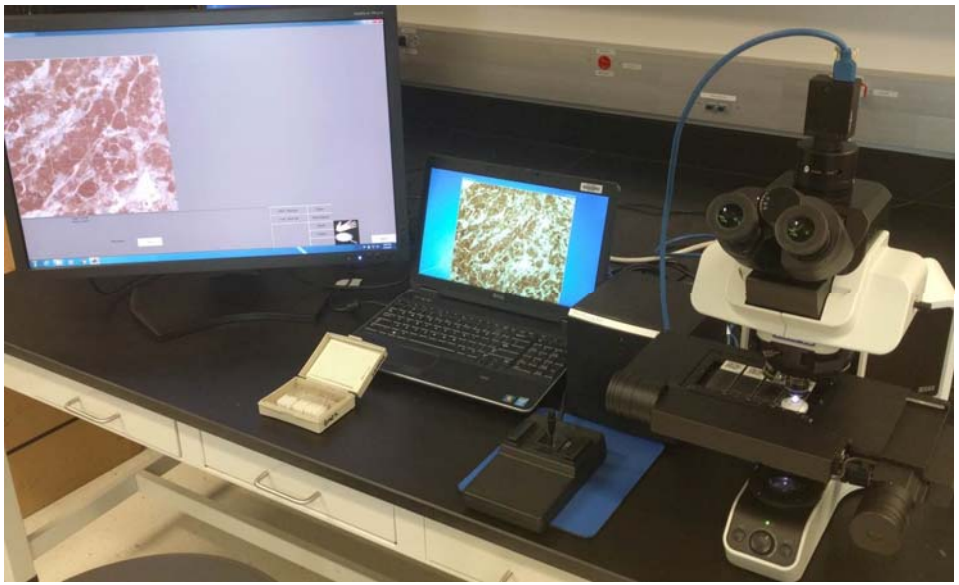


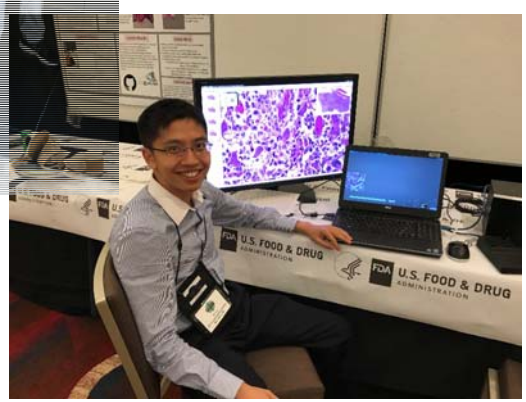




eeDAP on the road last year …

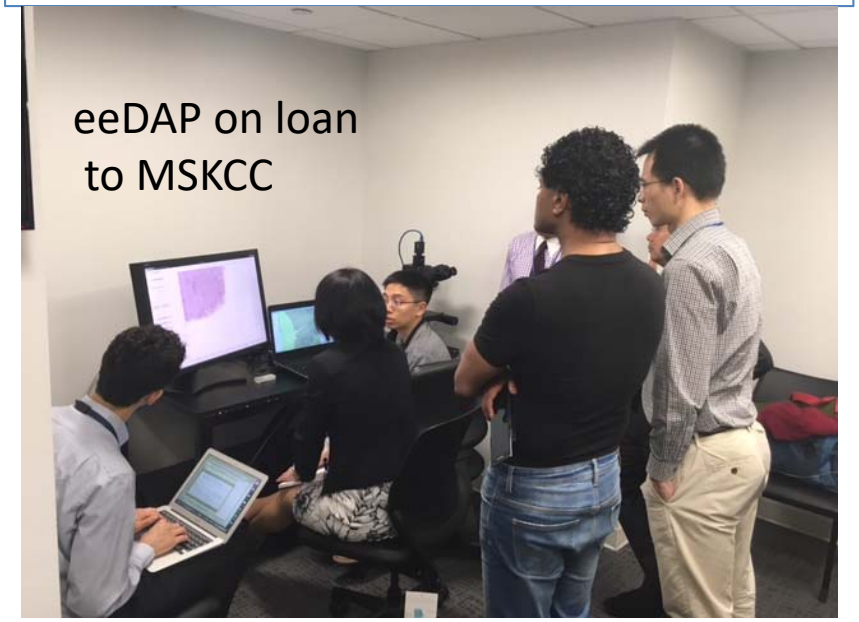
Monitor, Computer, motorized stage with joystick, microscope with mounted camera, reticle in eyepiece





Mitotic Counting and Classification

Install, Demo, Train at MSKCC



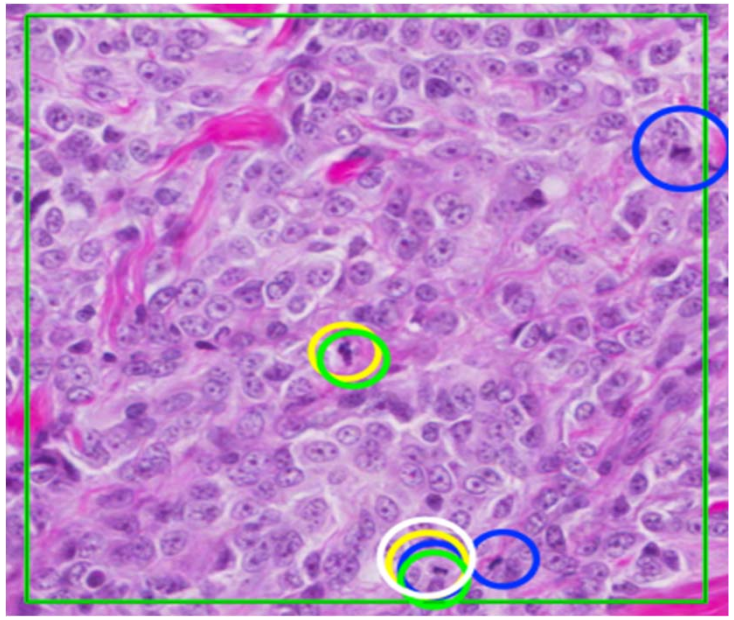
Study Design

- 4 slides from Mark Simpson at NCI
  - HE: canine oral melanoma
- 10 ROIs per slide from tumor
  - ROI
  - 800 x 800 pixels @ 0.25um/pixel
    200um x 200um
    17% of the entire FOV (0.24 mm^2)
- 4 pathologists from MSKCC

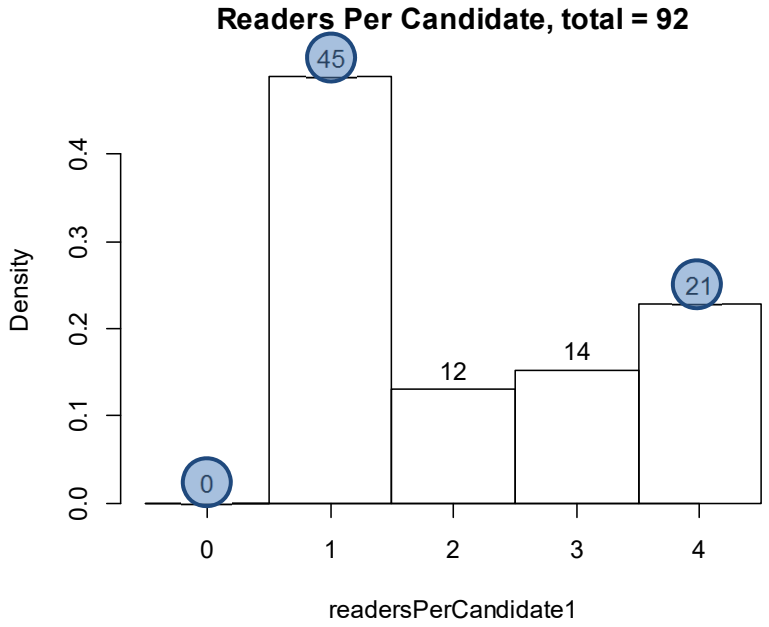
Quick look at first study

- Circles: mitotic figures identified by pathologists.
- "Candidate MFs" = marked cells
- Each color corresponds to a different pathologist.

WSI image



Readers per Candidate MF



- 45/92 = 49% marked by only one
- 21/92 = 23% unanimously marked
- Build these candidate MFs into next study: Classification task
- Need some low-probability candidates from ROIs with zero or one candidates -> yield 34

Can we use eeDAP on this multi-headed microscope?

- Same microscope frame … 14 heads!
- Stage mounts fine
- Camera mounts fine
- Let’s do it.

Mitotic counting and Classification: Multi-head microscope

High-throughput reader study

Study Design

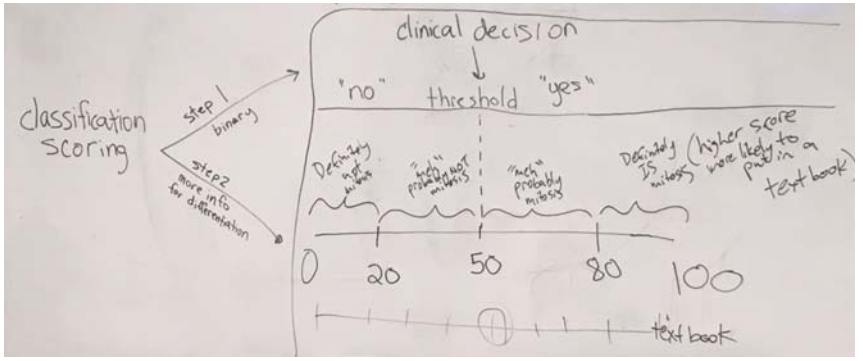
- 4 slides from Mark Simpson at NCI
  - HE: canine oral melanoma
- 10 ROIs per slide from tumor
  - ROI = 800 x 800 pixels @ 0.25um/pixel
    = 200um x 200um
    = 17% of the entire FOV (0.24 mm^2)
- 126 (=92+34) Candidate MFs
- 10 pathologists*
- Collect data on paper
  - ~1 hour training
  - ~2 hours for data collection

Mitotic counting and Classification: Multi-head microscope

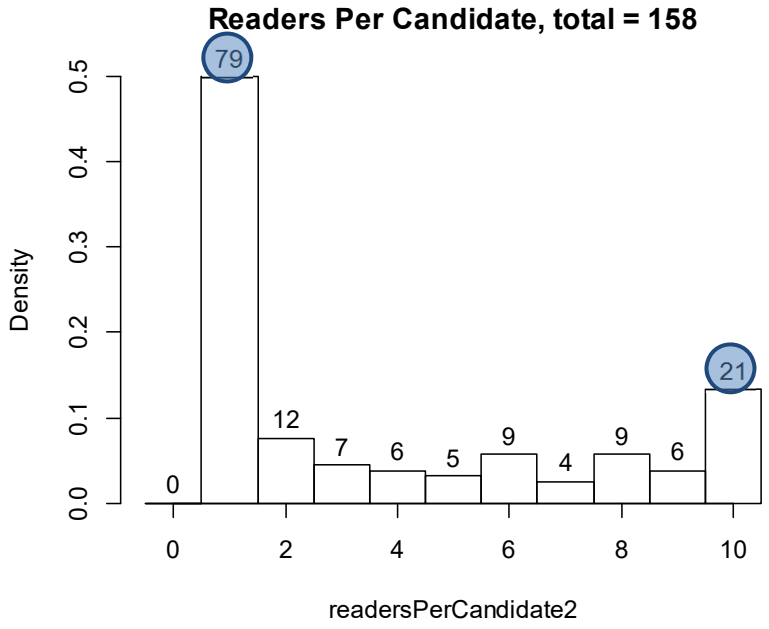
High-throughput reader study

Workflow

- Mark and count in ROI
- Classify candidates in same ROI

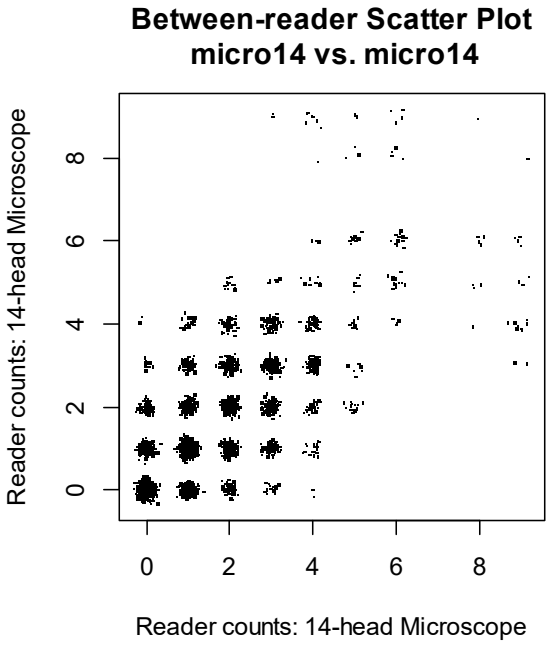


Readers per Candidate: Multi-head study



- Similar characteristics as before
- 79/158 = 49% marked by only one
- 21/158 = 13% unanimously marked
  - 13 agree with previous, 8 new ones

Counting Results



- Each point =
  - One ROI and a pair of readers
  - Appears twice (transpose x,y)
  - Has noise added for visualization

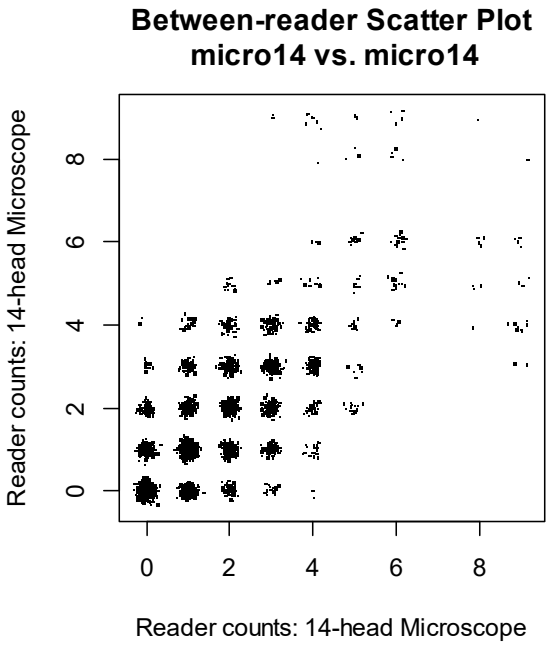
- How do we summarize this?

Agreement … No ground truth

Count differences (calibration)
Pairwise Concordance (correlation)

"MRMC" analyses
account for variability from
Multiple **R**eaders and **M**ultiple **C**ases

Results: Count Differences

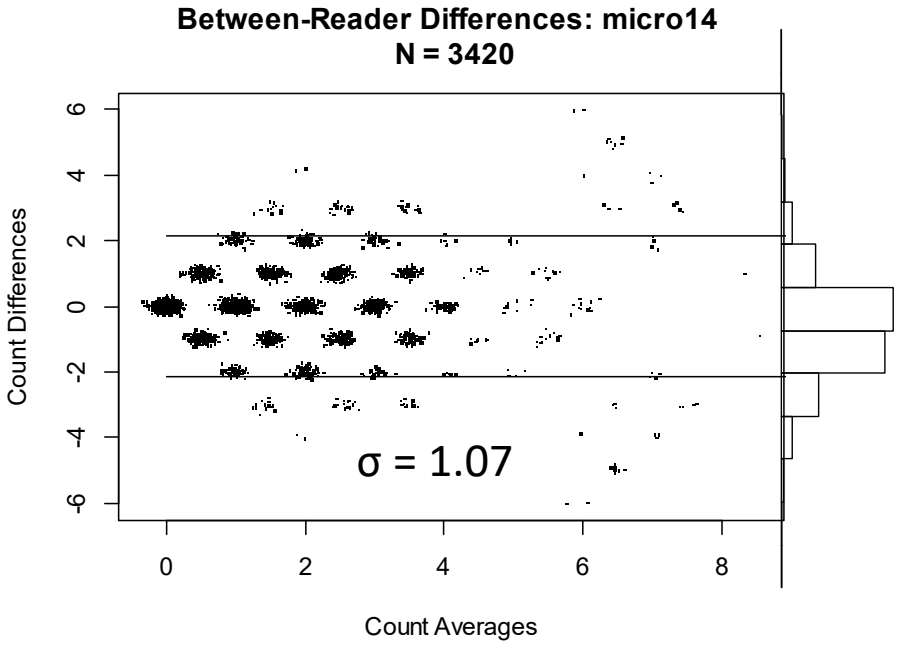


- Rotate 45 and

  rescale x-axis

-> Bland-Altman plot

Results: Count differences



- Rotate 45 and
  rescale x-axis

-> Bland-Altman plot

- Limits of agreement
  - Characterize spread of differences
  - σ = 1.07

  "MRMC" analyses:
  account for variability from
  Multiple **R**eaders and **M**ultiple **C**ases

- Not the standard error
  - SE characterizes the spread of the mean difference

Results: Count Differences

Study 1:	Average Counts	SE Average Counts	Std of Between-Reader Count Differences
Digital	1.22	0.23	1.29
Microscope	1.48	0.27	1.12
Microscope - Digital	0.26	0.12	1.20

- Study 1:
  - More MFs with microscope
  - Count differences were larger with digital
- Study 2:
  - Microscope results consistent with Study 1

Results: Count Differences

Study 1:	Average Counts	SE Average Counts	Std of Between-Reader Count Differences
Digital	1.22	0.23	1.29
Microscope	1.48	0.27	1.12
Microscope - Digital	0.26	0.12	1.20

- Study 1:
  - More MFs with microscope
  - Count differences were larger with digital
- Study 2:
  - Microscope results consistent with Study 1

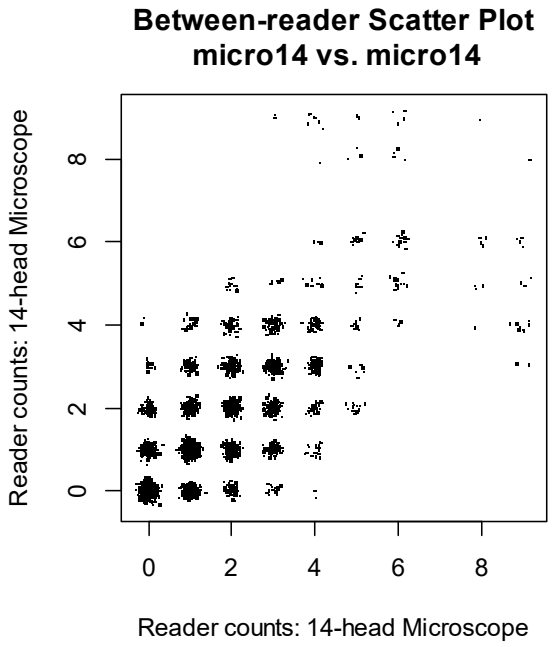
Results: Count Differences

Study 1:	Average Counts	SE Average Counts	Std of Between-Reader Count Differences
Digital	1.22	0.23	1.29
Microscope	1.48	0.27	1.12
Microscope – Digital	0.26	0.12	1.20
Study 2:			
14-head Microscope	1.54	0.25	1.07

- Study 1:
  - More MFs with microscope
  - Count differences were larger with digital
- Study 2:
  - Microscope results consistent with Study 1

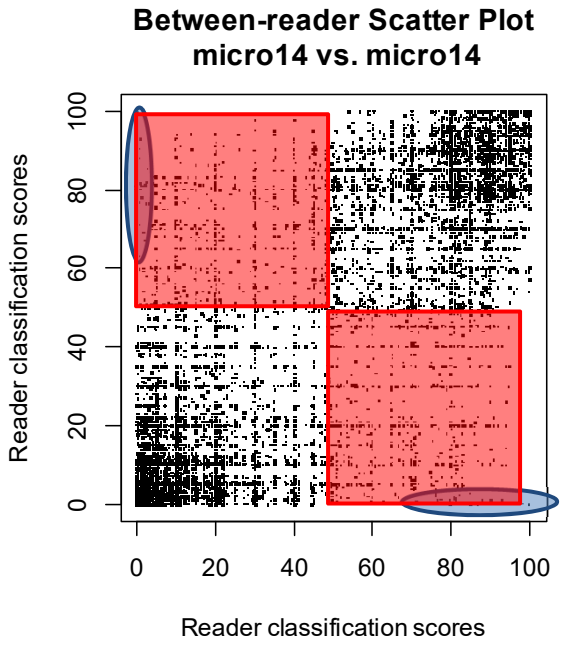
Pairwise Concordance

A probability that tracks with correlation



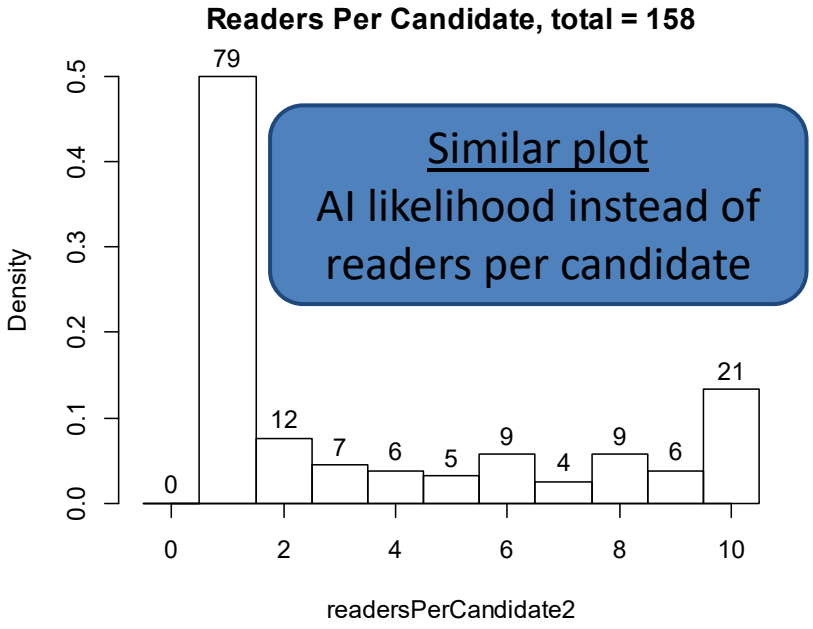
No time for concordance results

Classification scores



No time for concordance results

Generalize to evaluating computational pathology



Similar plot
AI likelihood instead of readers per candidate

- FDA qualification of images with annotations
  - MDDT: Medical Device Development Tools
  - Support FDA submissions of computational pathology
- Generate candidates from
  - Pathologists

  AND
  - Algorithm(s)
- Candidates cover range in likelihood the candidate is a MF
- Use same agreement measures

Reduces bias in the comparison

Summary

- Collected and analyzing:
  - MF counts, locations, and classifications
- Agreement analyses
  - MRMC analysis
  - Calibration
  - Correlation
  - Unit of analysis:
    cells > ROIs > slides
- Limitations
  - Anecdotal feedback
    - Pathologists felt rushed
    - Focus handling not perfect
    - No reticles in eyepieces
  - No Ground Truth
- Future work
  - Generalize to other ROIs?
  - Generalize to other specimens (organs)?
- Evaluate AI algorithms
  - Use similar study design
  - Use similar analysis tools
  - Need “candidates” from algorithms and pathologists for unbiased evaluation
- FDA qualification of images with annotations
  - MDDT: Medical Device Development Tools
  - Test sets for FDA submissions of computational pathology