

High-Throughput Truthing (HTT): Pathologist Agreement from a Pilot Study

Brandon D. Gallas

Division of Imaging, Diagnostics, Software Reliability

Office of Science and Engineering Laboratories

Center for Devices and Radiological Health

U.S. Food and Drug Administration

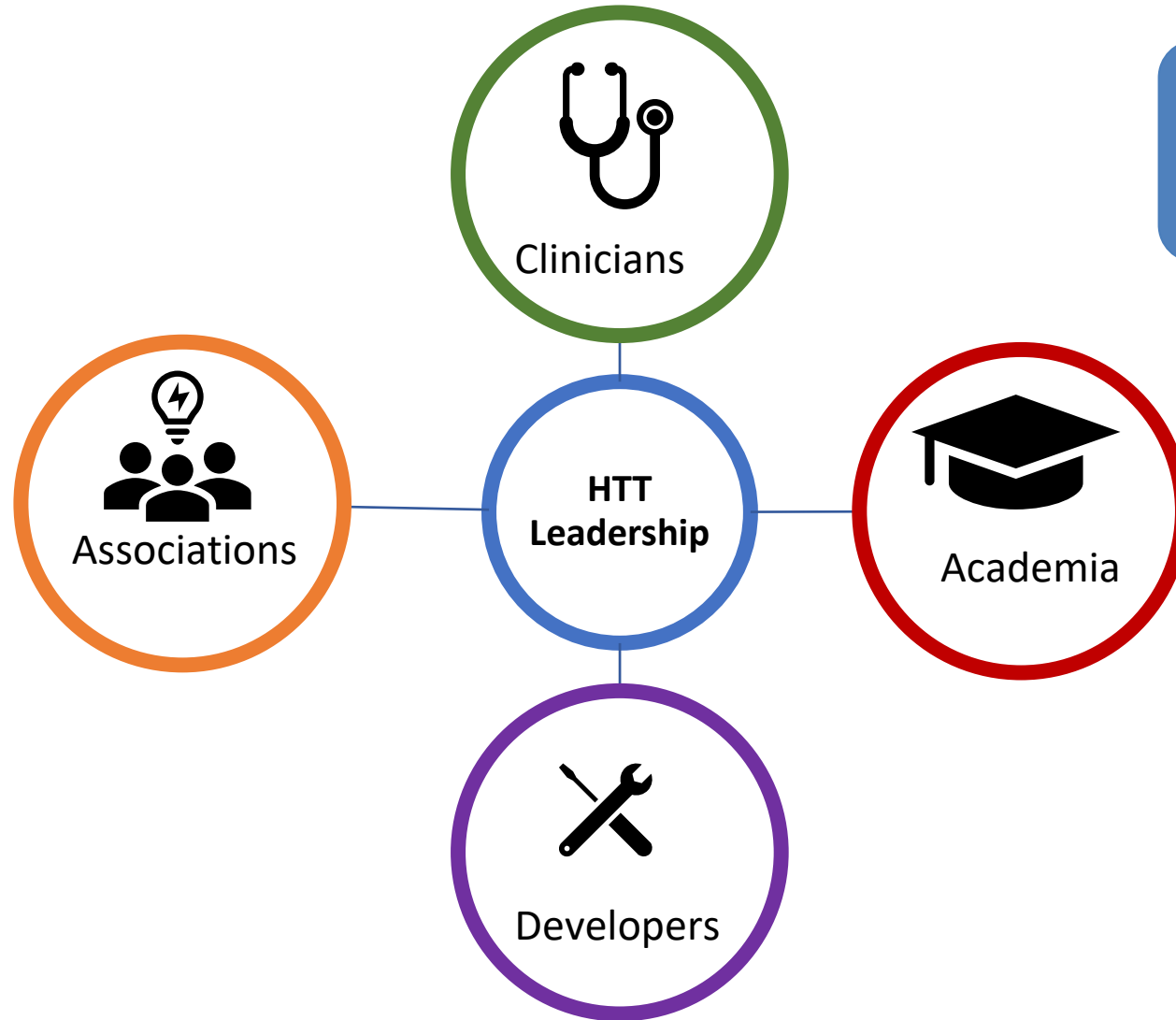


Co-Authors

- **Katherine Elfer, PhD, MPH**
 - FDA/CDRH/OSEL/DIDSR
- **Mohamed Amgad, MD**
 - Department of Pathology, Northwestern University
- **Weijie Chen, PhD**
 - FDA/CDRH/OSEL/DIDSR
- **Sarah Dudgeon, MPH**
 - CORE Center for Computational Health Yale-New Haven Hospital
- **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
- **Richard Huang, MD**
 - Massachusetts General Hospital/Harvard Medical School
- **Evangelos Hytopoulos, PhD**
 - iRhythm Technologies Inc
- **Denis Larsimont, MD**
 - Department of Pathology, Institut Jules Bordet
- **Xiaoxian Li, MD/PhD**
 - Emory University School of Medicine
- **Anant Madabhushi, PhD**
 - Case Western Reserve University
- **Hetal Marble, PhD**
 - Massachusetts General Hospital/Harvard Medical School
- **Roberto Salgado, PhD**
 - Division of Research, Peter Mac Callum Cancer Centre, Melbourne, Australia; Department of Pathology, GZA-ZNA Hospitals
- **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

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Collaboration of Volunteers



There is room for you!

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

DISCLOSURE

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 any commercial products herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Outline

- Overview of the HTT project
 - High-Throughput Truthing
- Explore the Data
- Questions and Current Work
- Next steps
- Conclusions

Work in
progress

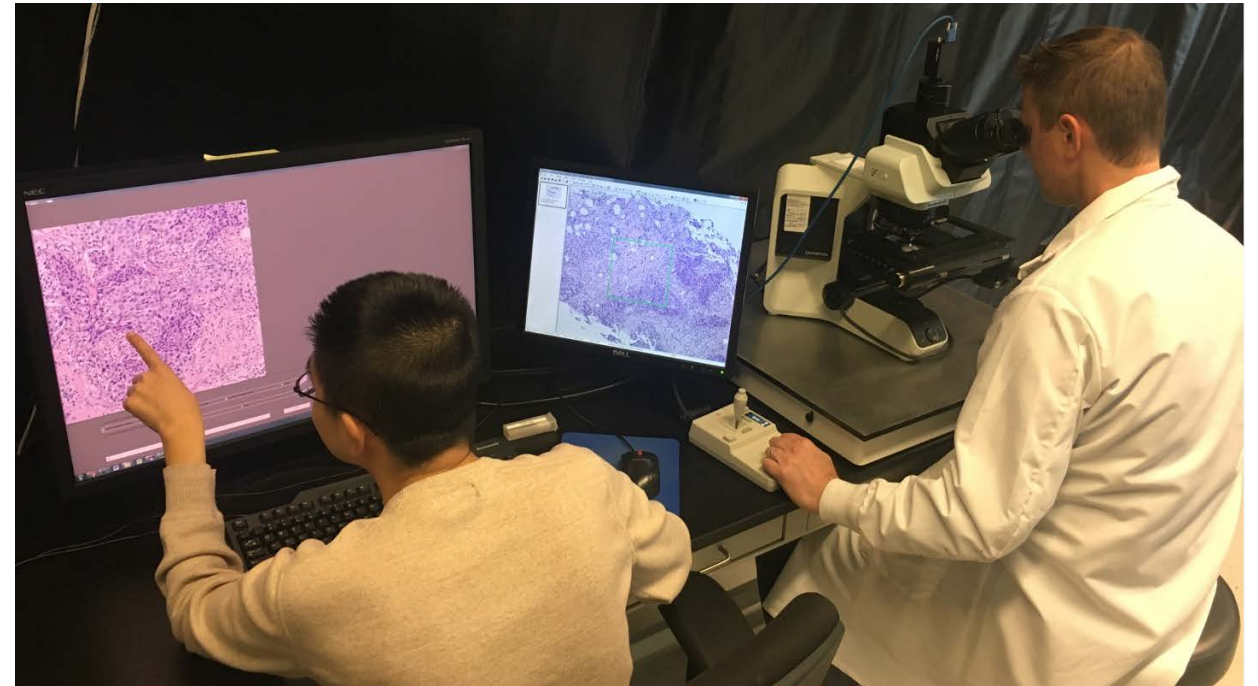
Overview of the HTT project

- Clinical Application and Relevance
- Regulatory Deliverable
- Validation Data and Methods
- Standardized Evaluations of a Quantitative Biomarker

- Project description accepted for publication at the Journal of Pathology Informatics
 - S. N. Dudgeon *et al.*, “A Pathologist-Annotated Dataset for Validating Artificial Intelligence: A Project Description and Pilot Study,” *arXiv:2010.06995 [eess, q-bio]*, vol. Accepted for publication by the Journal of Pathology Informatics, Oct. 2020, Accessed: Oct. 29, 2020. [Online]. Available: <http://arxiv.org/abs/2010.06995>

Clinical Application and Relevance

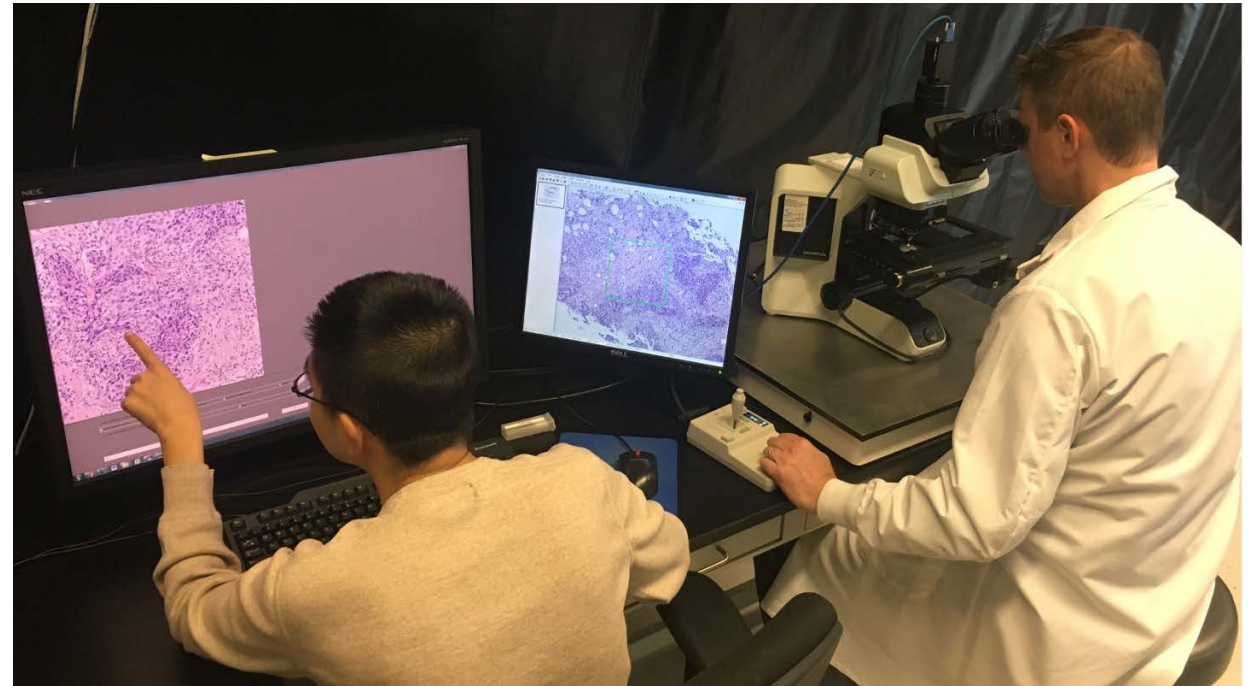
- Clinical application:
 - Stromal Tumor Infiltrating Lymphocytes (sTILs) in breast cancer
- Clinical relevance of sTILs:
 - Prognostic for survival
 - Expected to inform patient management
 - Expected to reduce use of toxic chemotherapies
- Software as a medical device (SAMd)
 - Reduce burden on pathologist
 - Reproducible
 - Quantitative



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Regulatory Deliverable

- **Regulatory Science Question**
 - How can we use pathologist annotations to support SaMD validation?
- **Deliverables**
 - Validation data
 - Methods
- **Pursue Regulatory Deliverable:**
 - Medical Device Development Tool (MDDT)



<https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt>

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Validation Data and Methods

Data

- Multiple sites
- Represents defined population
- Reproducible protocol
- Proficient Pathologists

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Validation Data and Methods

Data

- Multiple sites
- Represents defined population
- Reproducible protocol
- Proficient Pathologists

Methods

- Interchangeability
- Quantitative biomarker
 - Quantitative agreement
- Human evaluation of a quantitative biomarker
 - Rank-based agreement
 - Qualitative agreement

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Validation Data and Methods

Quantitative Agreement Endpoint:

MSD = Mean-Squared Deviation

- Algorithm-pathologist agreement

$$\text{MSD} = E \left[(Y_{kl} - X_{jkl})^2 \right]$$

Score from SaMD

Score from pathologist j

Same case k and location l

- Pathologist-pathologist agreement

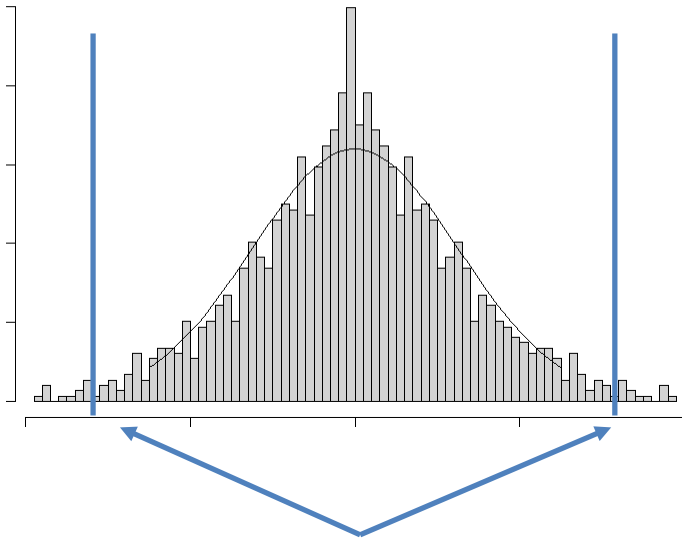
$$\text{MSD} = E \left[(X_{j'kl} - X_{jkl})^2 \right]$$

Score from pathologists j and j'

Same case k and location l

Validation Data and Methods

Distribution of differences between pathologists



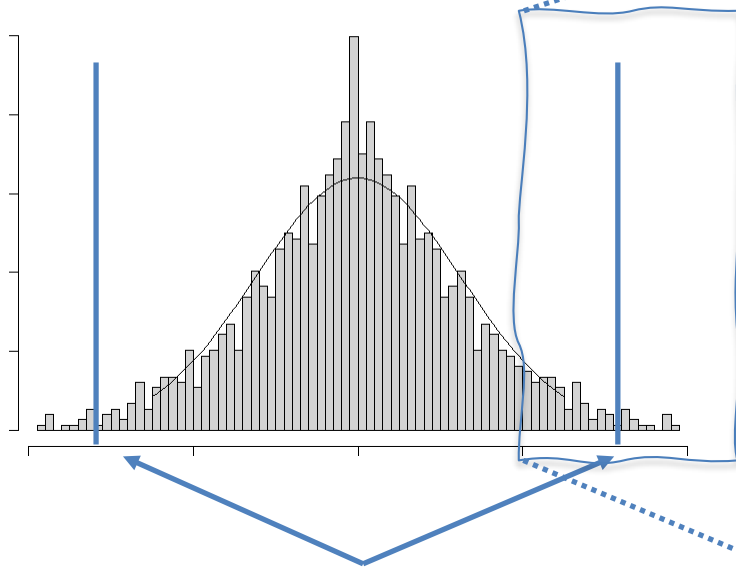
Limits of agreement

Observed differences will be within the LOA
~95% of the time

- Limits of agreement are proportional to
 - Standard deviation
 - Square root of MSD

Validation data and methods

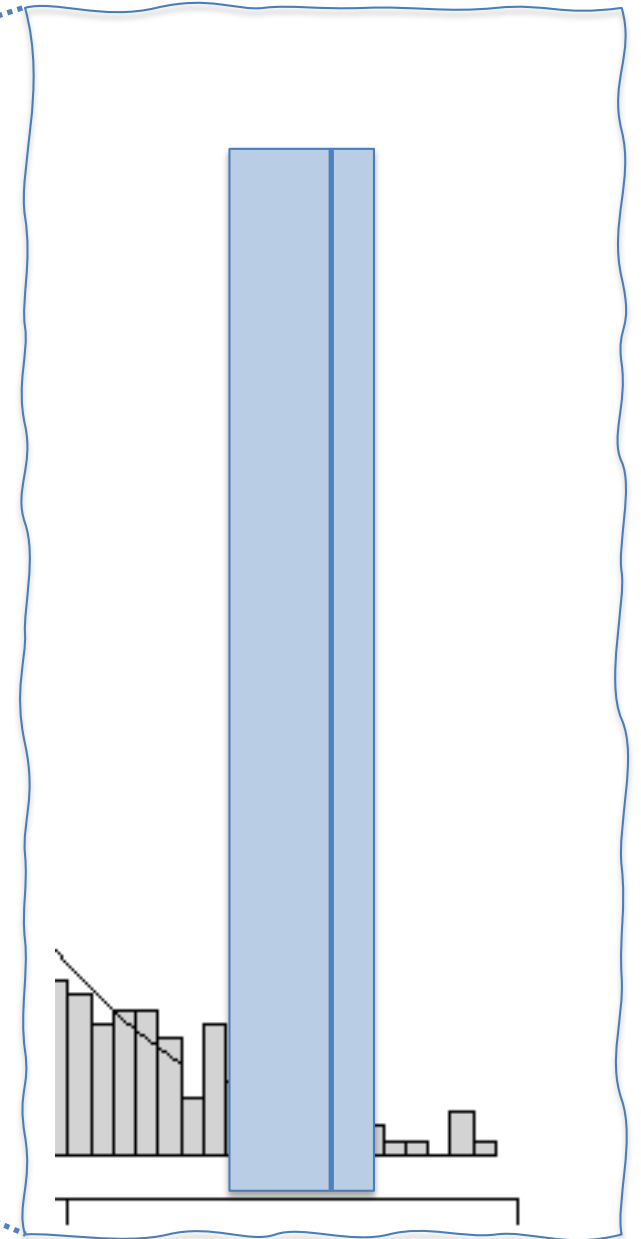
Distribution of differences between pathologists



- Confidence interval for reference panel LOA (not symmetric)

Limits of agreement

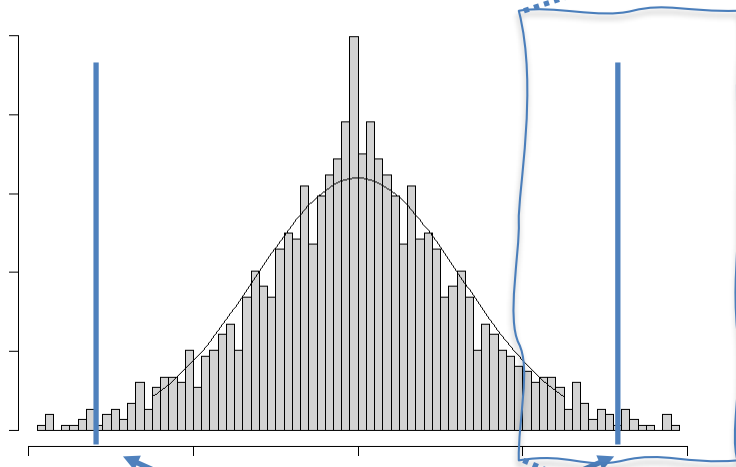
Observed differences will be within the LOA ~95% of the time



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Validation Data and Methods

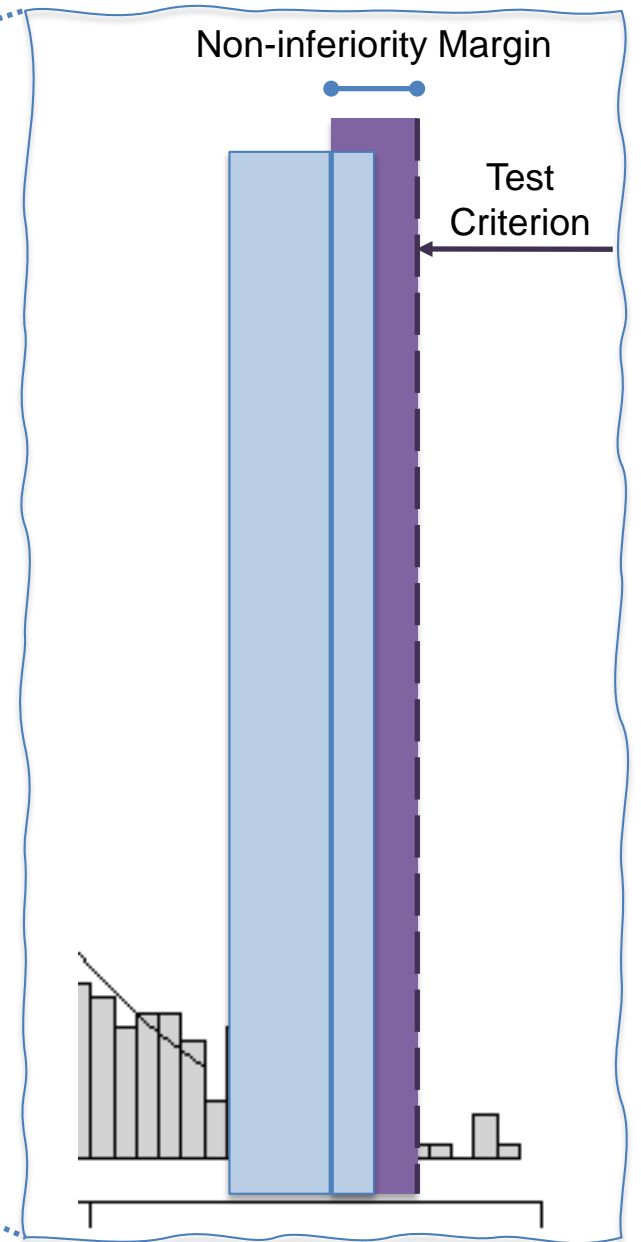
Distribution of differences between pathologists



- Clinically tolerable non-inferiority margin – 5% ?

Limits of agreement

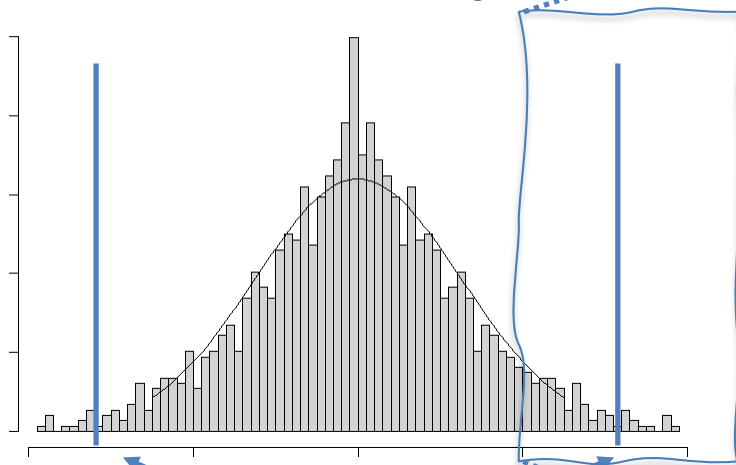
Observed differences will be within the LOA ~95% of the time



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Validation Data and Methods

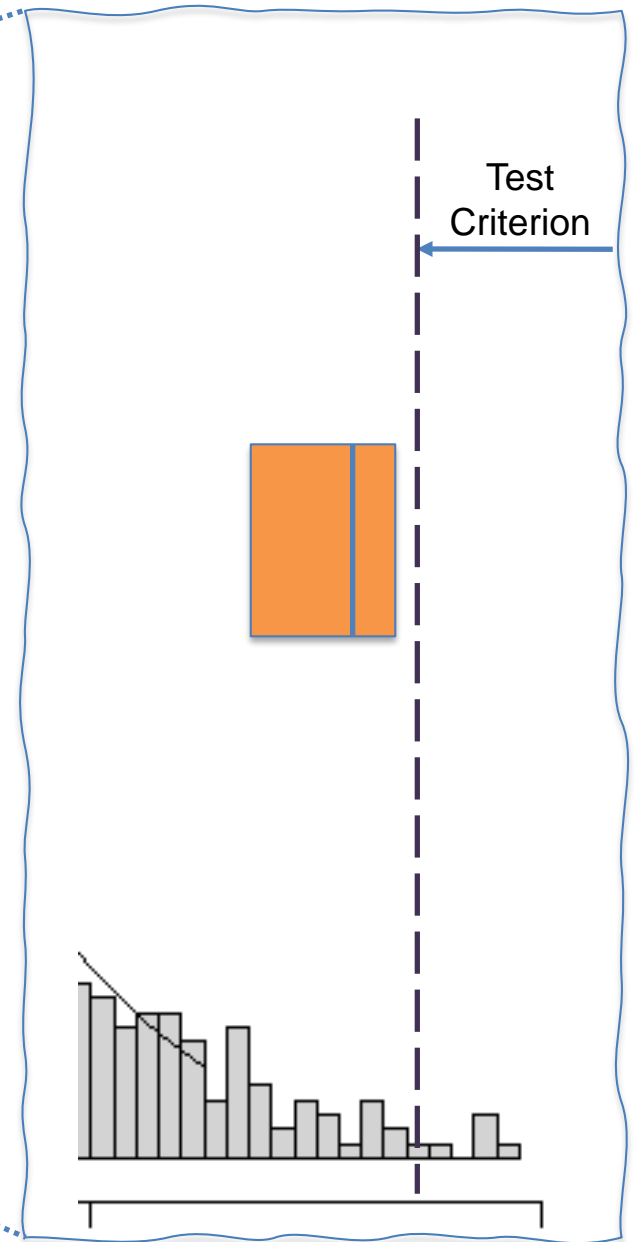
Distribution of differences between pathologists



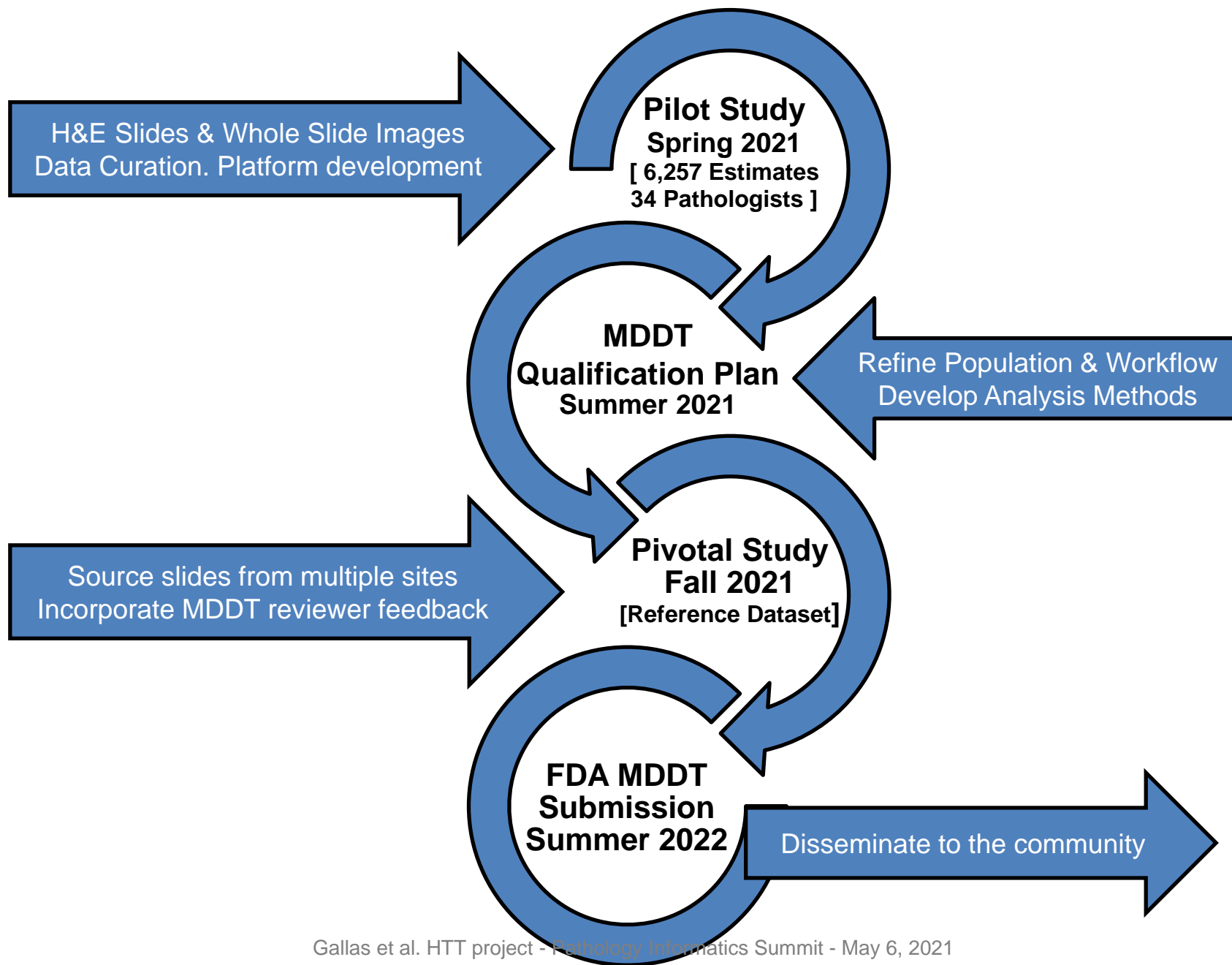
- Confidence interval of the algorithm-pathologist LOA
 - Compare to test criterion

Limits of agreement

Observed differences will be within the LOA ~95% of the time



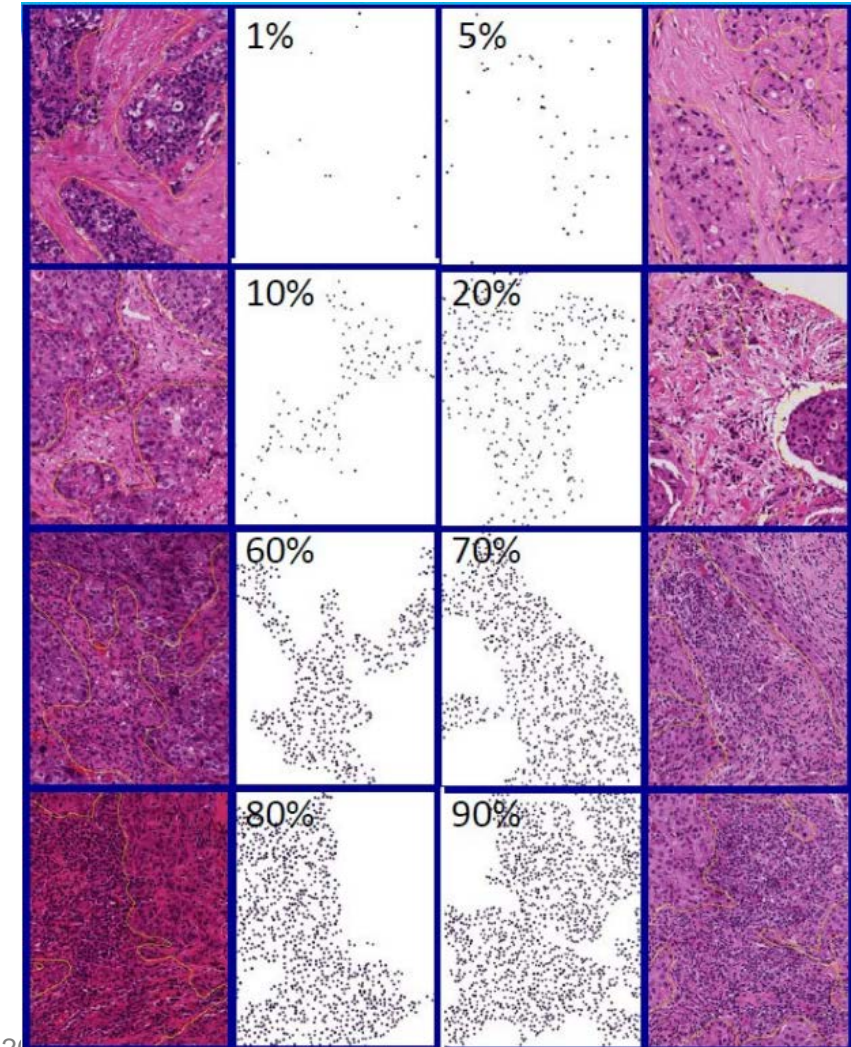
Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Standardized Evaluations of a Quantitative Biomarker

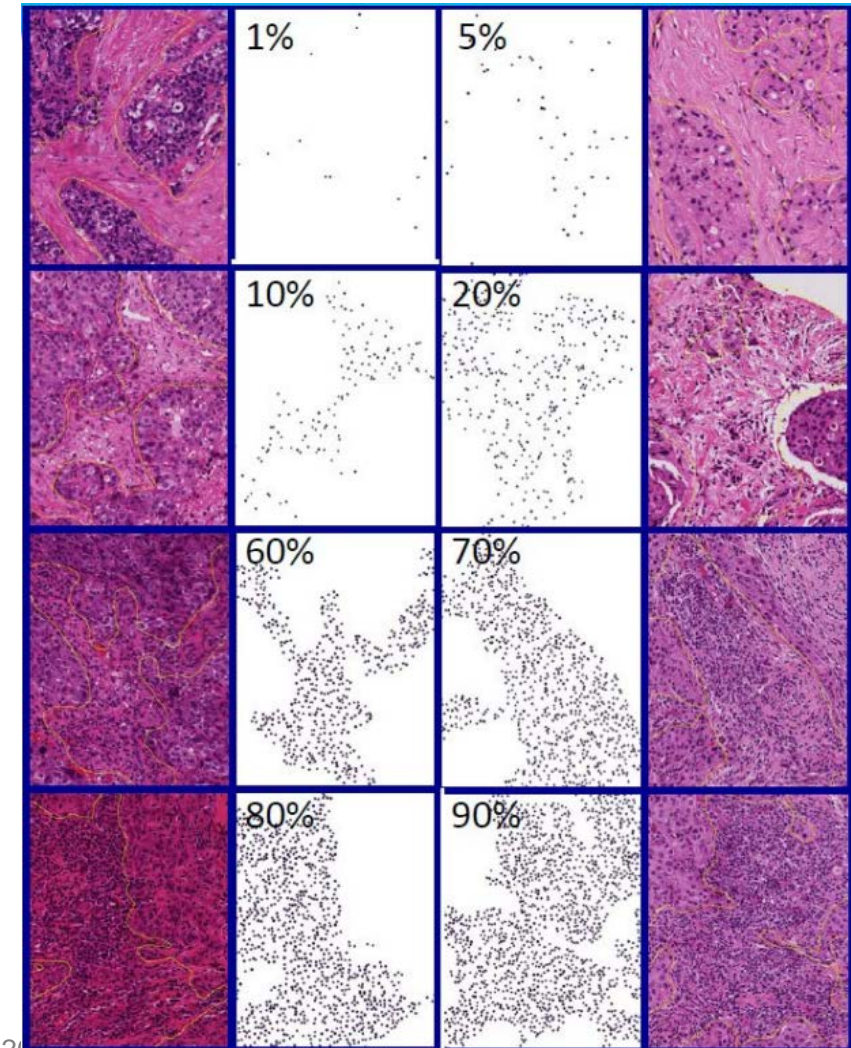
- Pathologist Evaluation
 - Density Estimates percent stromal Tumor Infiltrating Lymphocytes (sTILs) 0%-100%
 - Density Estimates percent stroma 0%-100%
- R. Salgado *et al.*, “The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: **recommendations** by an International TILs Working Group 2014,” *Ann. Oncol.*, vol. 26, no. 2, pp. 259–271, Feb. 2015, doi: [10.1093/annonc/mdu450](https://doi.org/10.1093/annonc/mdu450).



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2015

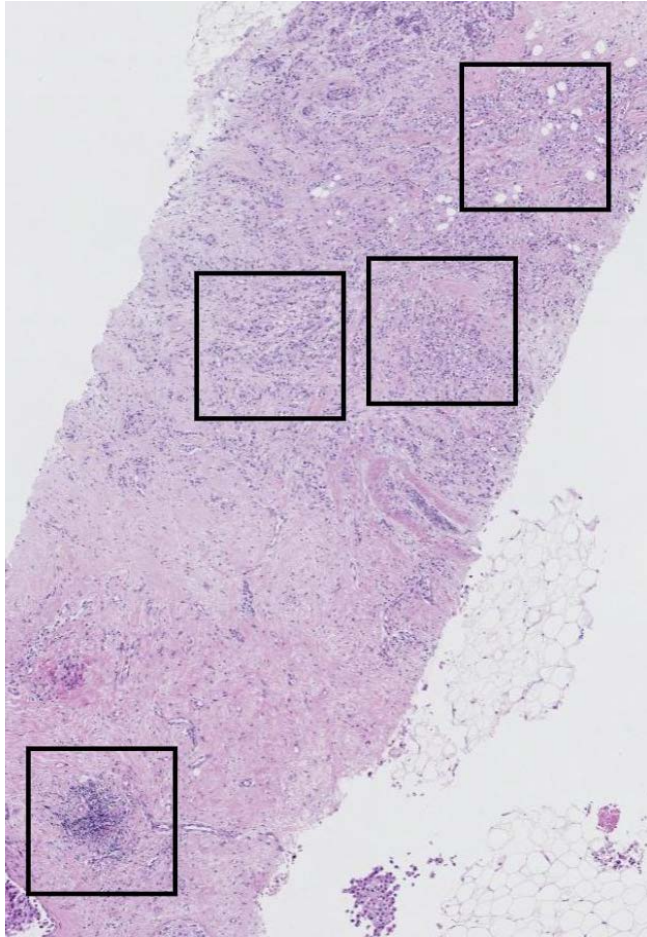
Standardized Evaluations of a Quantitative Biomarker

- “Required” training included a 13-minute video
 - Training not monitored
- “Optional” training included
 - Link to the recommendations
 - Project overview (video)
 - Platform operation overview (video)



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2011

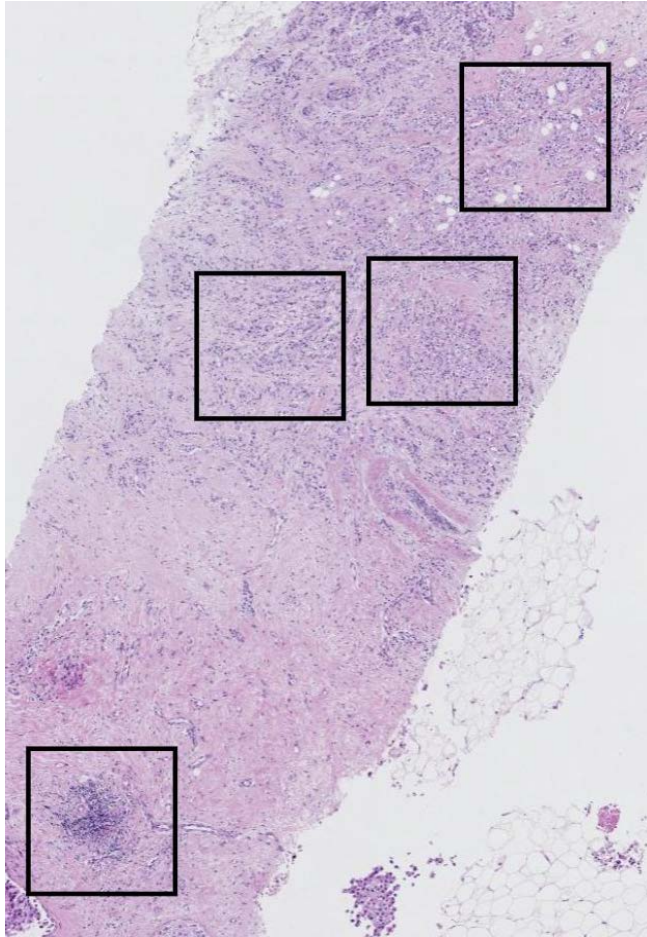
Pre-select Regions of Interest (ROIs)



- **Intra-tumoral stroma**
(Tumor-associated stroma)
 - Select ~3 ROIs
- **Invasive margin**
(Tumor-stroma transition)
 - Select ~2 ROIs
- **Tumor with no intervening stroma**
 - Select ~2 ROIs, if possible
- **Other regions**
 - Select ~3-4 ROIs

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Pilot Study Materials



- **64** Hematoxylin & Eosin Slides
 - “40X” Imaging (0.23 um/pixel)
- **10** ROIs per Slide
- **640** ROIs Total
 - 8 batches of 8 slides
- **500 um x 500 um** squares

No patient
information or
meta-data

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Evaluation Platforms

- Digital
 - caMicroscope
 - PathPresenter

- Microscope
 - eeDAP

We did not specify the display

We did collect the display size in pixels

caMicroscope (Digital)

Slide: HTT-TILS-001-19B

Home, Back, Rotate, Zoom, Help icons

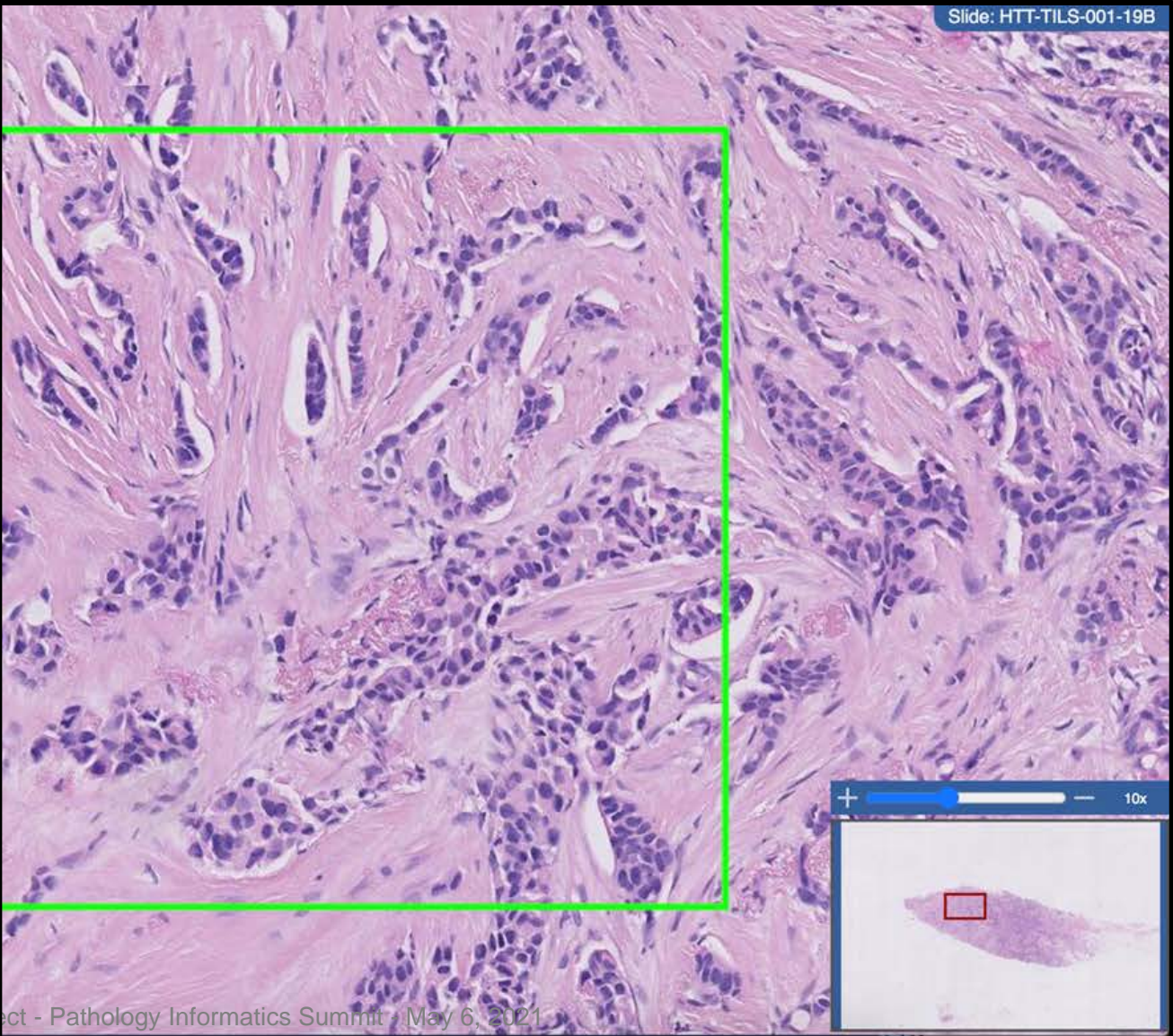
ROI Type:

- Intra-Tumoral Stroma
- Tumor with No Intervening Stroma
- Invasive Margin
- Other Regions

68%
% Tumor-Associated Stroma

12%
Please Assess TIL Density

Save & Next



The main image is a histology slide showing tumor tissue. A green box highlights a region of interest (ROI) within the tumor. A 10x inset in the bottom right corner shows a magnified view of the ROI, with a red box indicating the area of interest. The inset includes a zoom control with a plus sign, a slider, and a minus sign, and the text '10x'.

ect - Pathology Informatics Summit - May 6, 2021

PathPresenter (Digital)

The screenshot displays the PathPresenter software interface. On the left, a histology image is shown at 40x magnification, with a 100 µm scale bar. The top toolbar includes navigation icons, magnification options (0.25x, 2x, 5x, 10x, 20x, 40x, 60x, 100x), a search icon, and a zoom level of 20. The area of interest is 0.25 mm. On the right, a data entry panel is visible with the following fields:

- ROI Label :** Intra tumoral stroma
- Description :** Test Description
- %Tumor-Associated Stroma :** 23
- TILs :** 21

Each percentage field includes a slider control. The TILs field is highlighted in yellow. At the bottom right of the panel are 'Cancel' and 'Save' buttons.

Gallas et al. HTI project - Pathology Informatics Summit

eeDAP (Microscope)



Registers stage coordinates with whole slide image via camera

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

PathPresenter (Digital)



Task 1 of 40
HTT study

1. How would you label this ROI?

Switch to WSI thumbnail

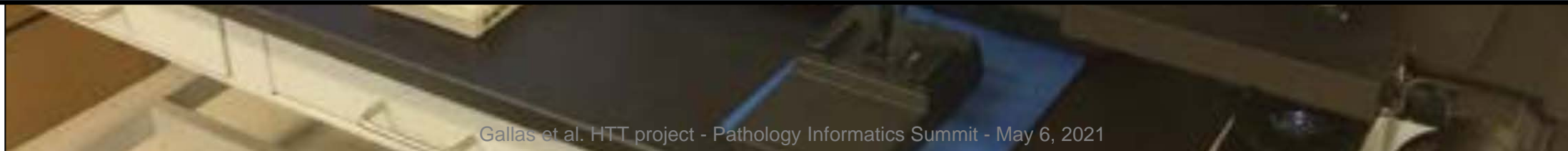
Intra-tumoral stroma
 Tumor with NO intervening stroma
 Invasive margin (1mm)
 Other

2: What is the percent (%) Tumor-Associated Stroma?

0	50	100	Score
<input type="range" value="66"/>			66

3: What is the intra-tumoral stromal TIL density?

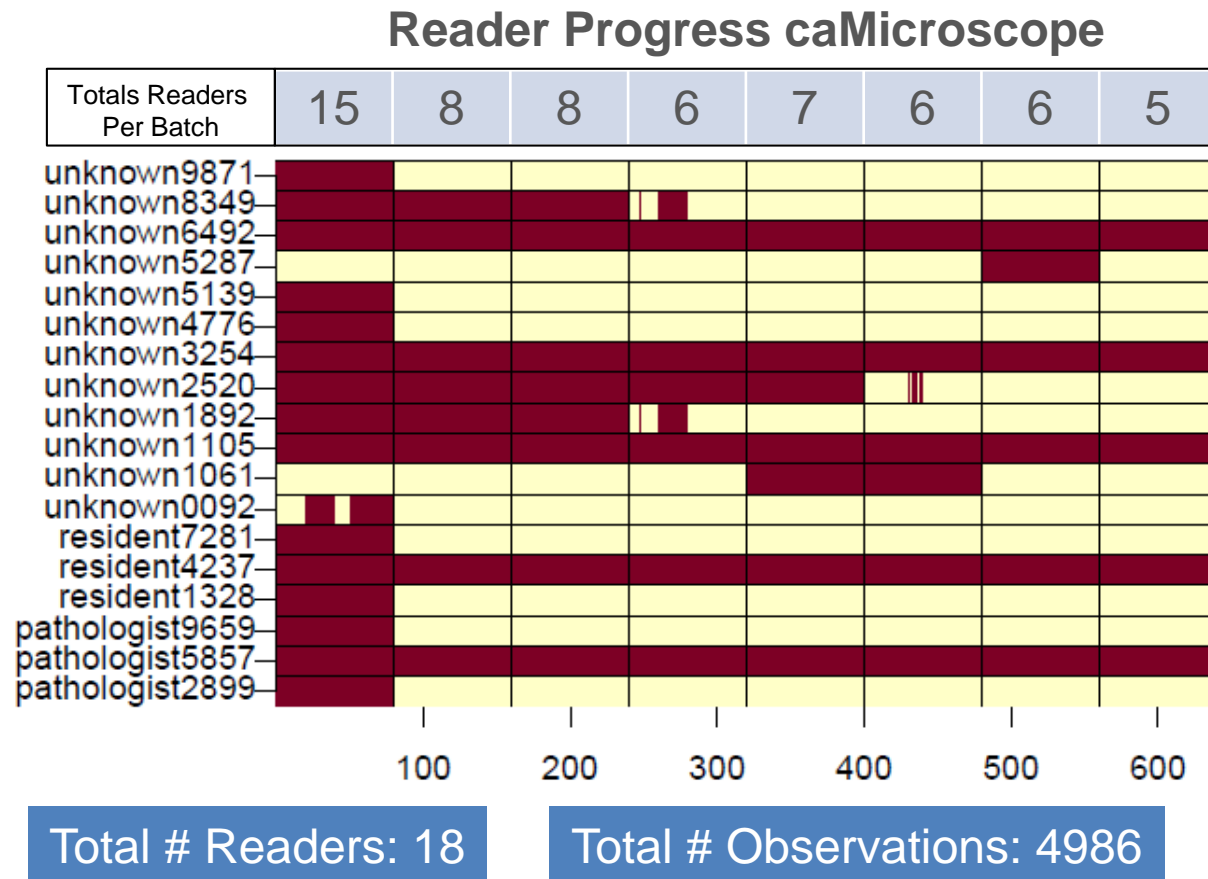
0	50	100	Score
<input type="range" value="66"/>			66



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Data Collected

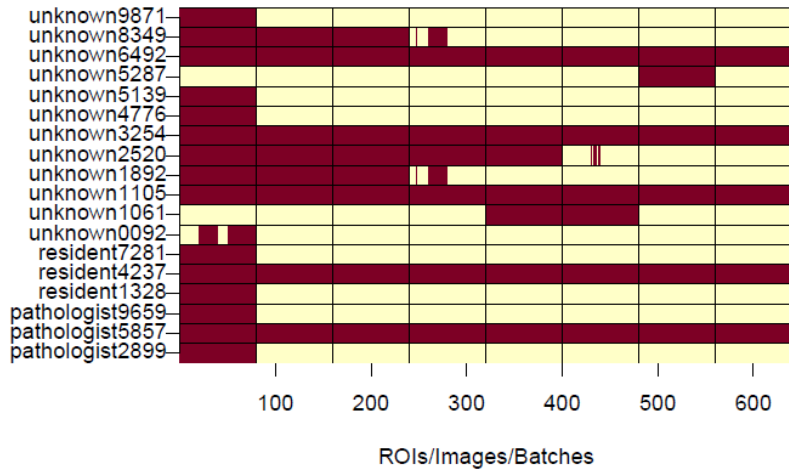
- Each tick mark is an observation
- Vertical lines partition the data by batch
- We are following up with the “unknowns” to get their experience



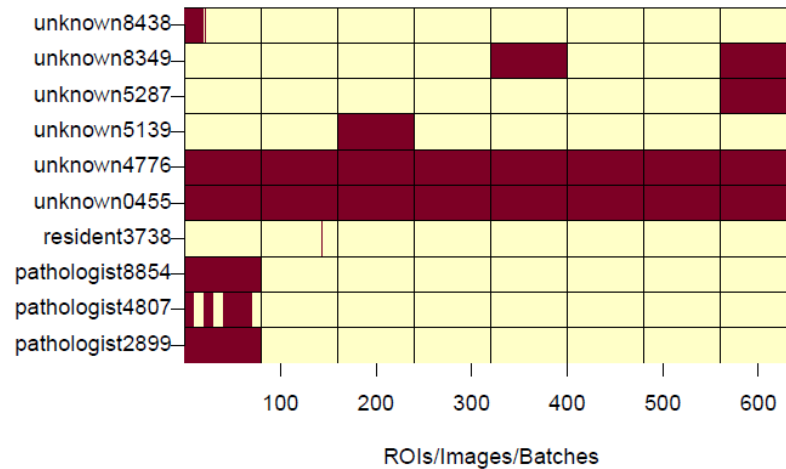
Data Collected

- Hit target:
 - 5 readers per ROI
 - Total observations: 7,259
 - Readers: 35
- Data-collection portals still open
- Two sites planned for eeDAP data collection this summer

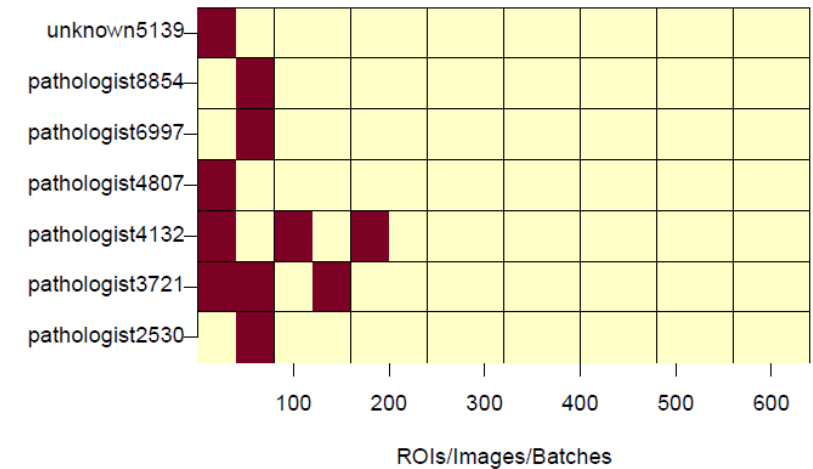
Reader Progress caMicroscope



Reader Progress PathPresenter



Reader Progress eeDAP



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Explore the data

- R Data Package for Sharing
- **CV**: Coefficient of Variation = $STD/Mean$
- Mean-variance relationship
- Scatter plots
- **LOA**: Limits Of Agreement

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

R Data Package

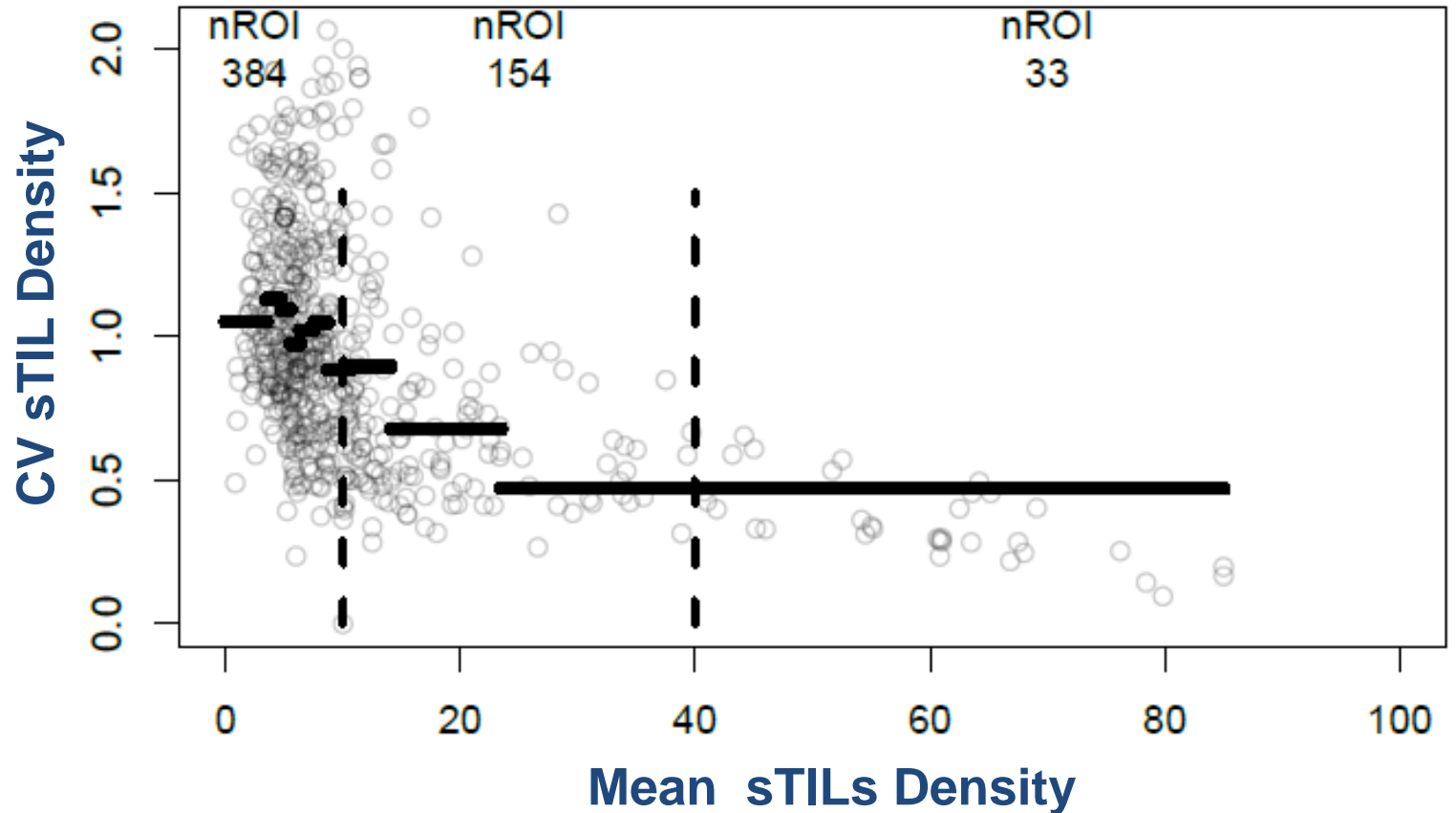
- Plan to share evaluation data summer 2021
- Use API's to pull data from platforms
- Use scripts to convert data into a standardized data frame
- Key variables:

caseID == ROI	readerID	modalityID	labelROI	percentStroma	densityTILs
HTT-TILS-001-73B.ndpi_x34892.2190_y45830.2190	unknown5287	camic	Intra-Tumoral Stroma	NA	20
HTT-TILS-001-73B.ndpi_x34892.2190_y45830.2190	pathologist5857	camic	Intra-Tumoral Stroma	39	45
HTT-TILS-001-73B.ndpi_x34892.2190_y45830.2190	resident4237	camic	Intra-Tumoral Stroma	15	20
HTT-TILS-001-73B.ndpi_x34892.2190_y45830.2190	unknown1105	camic	Intra-Tumoral Stroma	3	5
HTT-TILS-001-73B.ndpi_x34892.2190_y45830.2190	unknown6492	camic	Intra-Tumoral Stroma	20	10
HTT-TILS-001-73B.ndpi_x34892.2190_y45830.2190	unknown3254	camic	Intra-Tumoral Stroma	30	40

CV: Coefficient of Variation = STD/Mean

- Each circle is one ROI
- Mean and CV are averages over all readers
- Horizontal lines:
 - Average CV in 10% bins of the data (57 ROIs)
- Vertical dashed lines:
 - “Clinical” bins
 - low ($\leq 10\%$)
 - medium ($>10\% \ \& \ \leq 40\%$)
 - high ($>40\%$)

Coefficient of Variation (n=571, caMic)

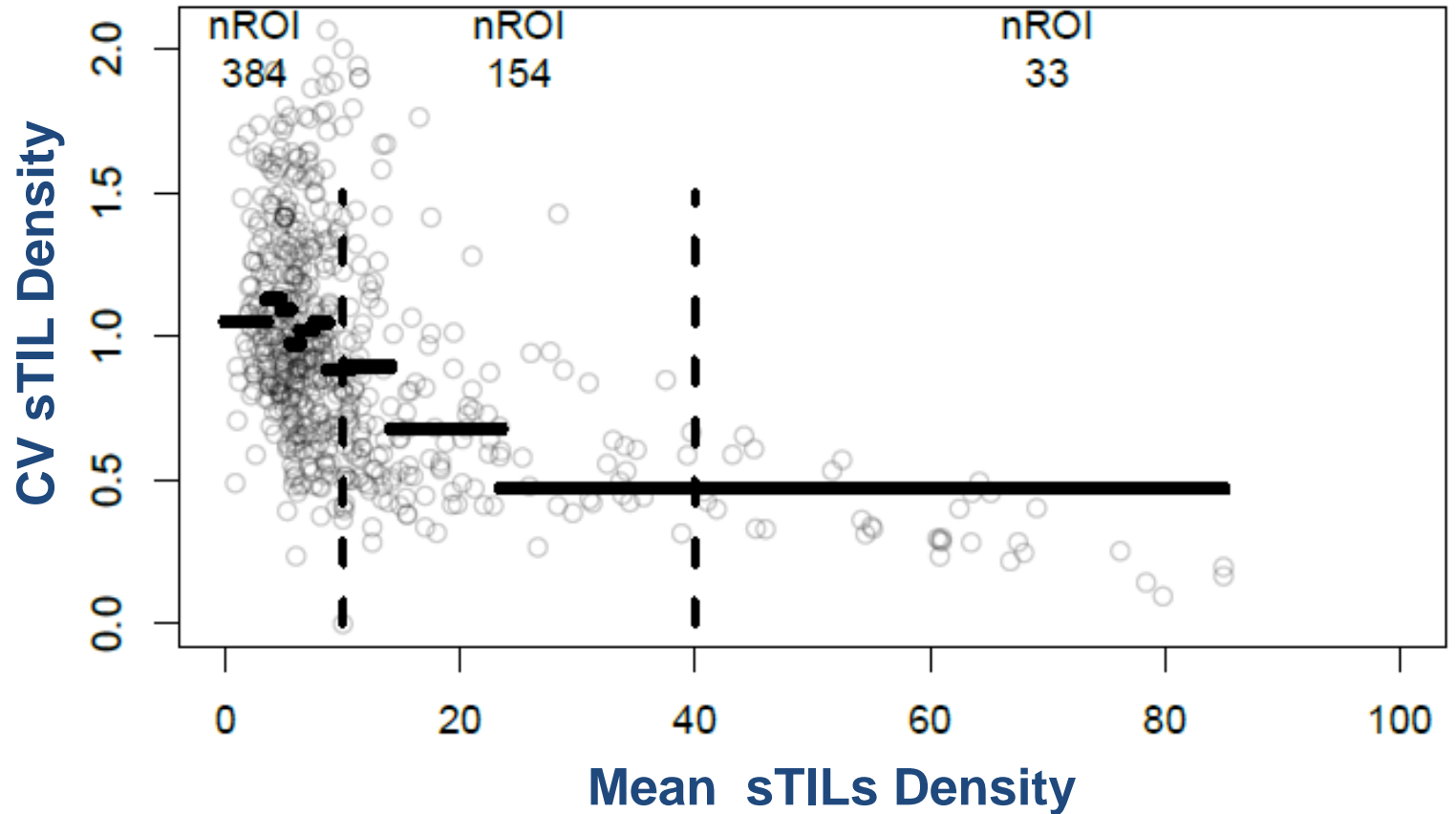


Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

CV: Coefficient of Variation = STD/Mean

- Clinical Interpretation:
 - Difficult for pathologists to quantitate scores, especially below 10
- Statistical Interpretation:
 - Standard deviation is not proportional to the mean
 - What is the mean-variance relationship?

Coefficient of Variation (n=571, caMic)

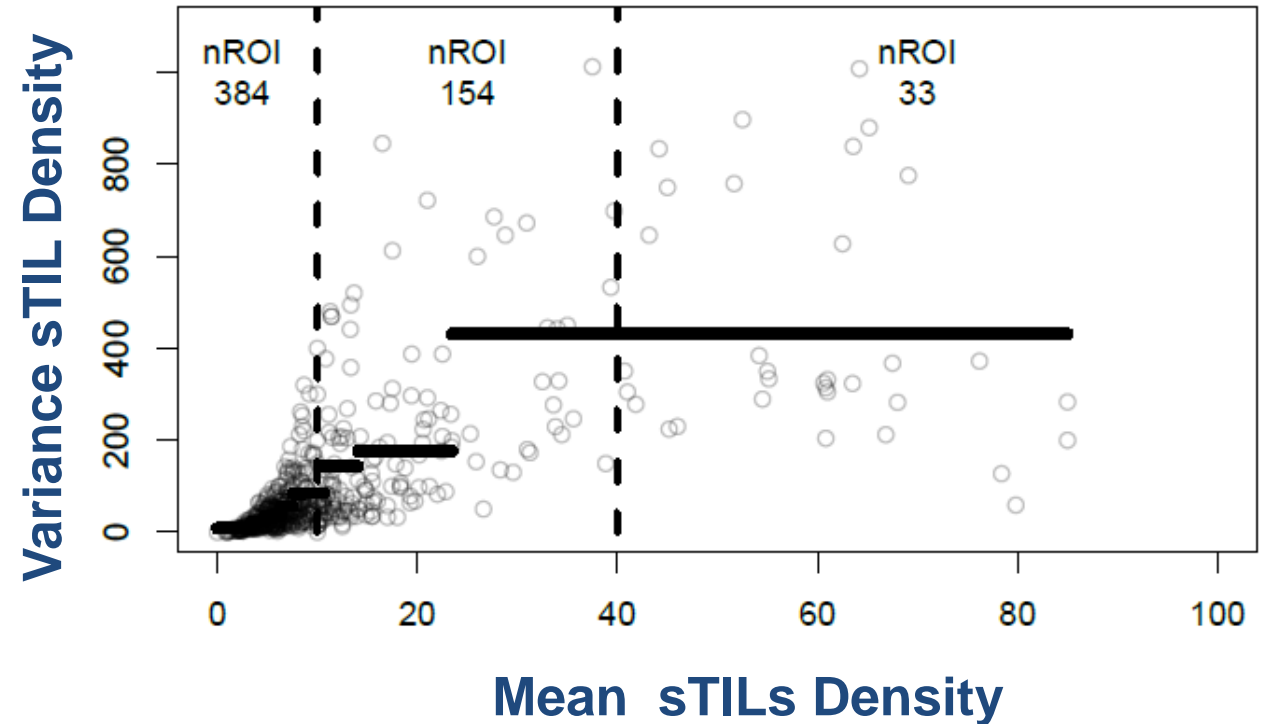


Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Mean-variance relationship

- Statistical Interpretation:
 - Variance increases with the mean
- Can't pool the data
 - Pick best readers
 - Transform the data
 - Log
 - Square-root
 - Bin the data
 - Average over ROIs per WSI
 - Ranks-based correlation

Mean-Variance (n=571, caMic)



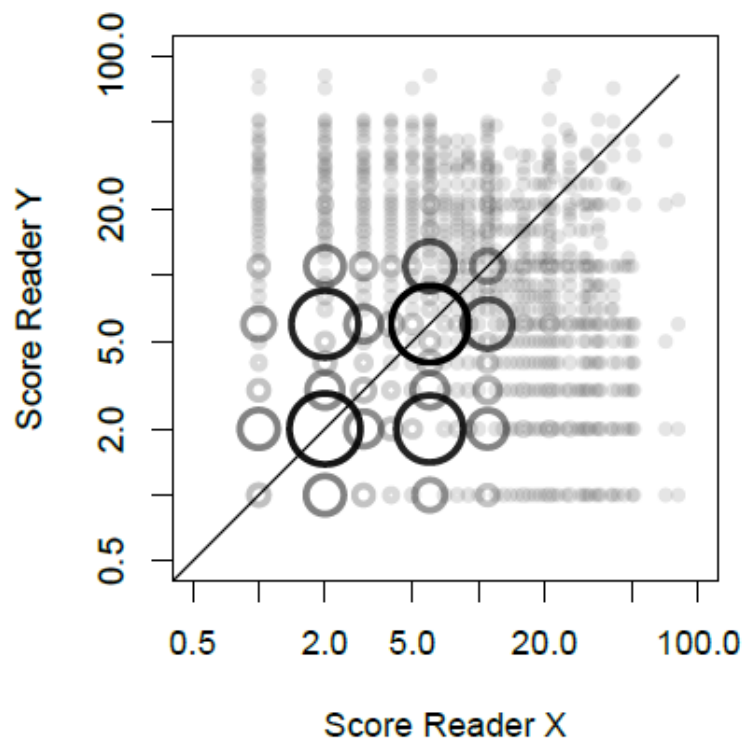
Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Scatter Plots

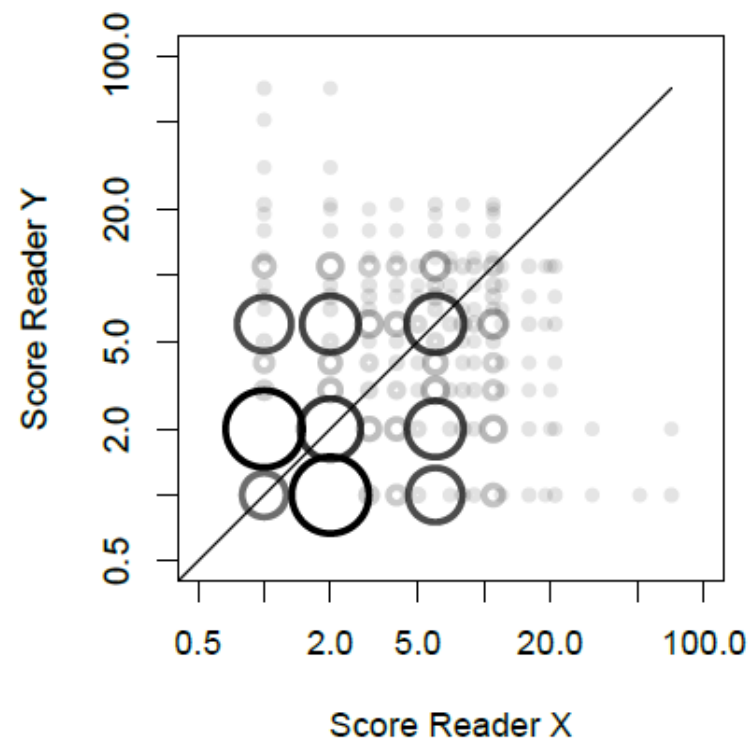
$0 \leq \text{scores} \leq 10$

Symmetrized: We plot (x,y) and (y,x) since we are pooling over readers and none is the reference.

Size of symbol and transparency are scaled with number of paired observations



n = 44784 , Largest symbol == 2602 observations



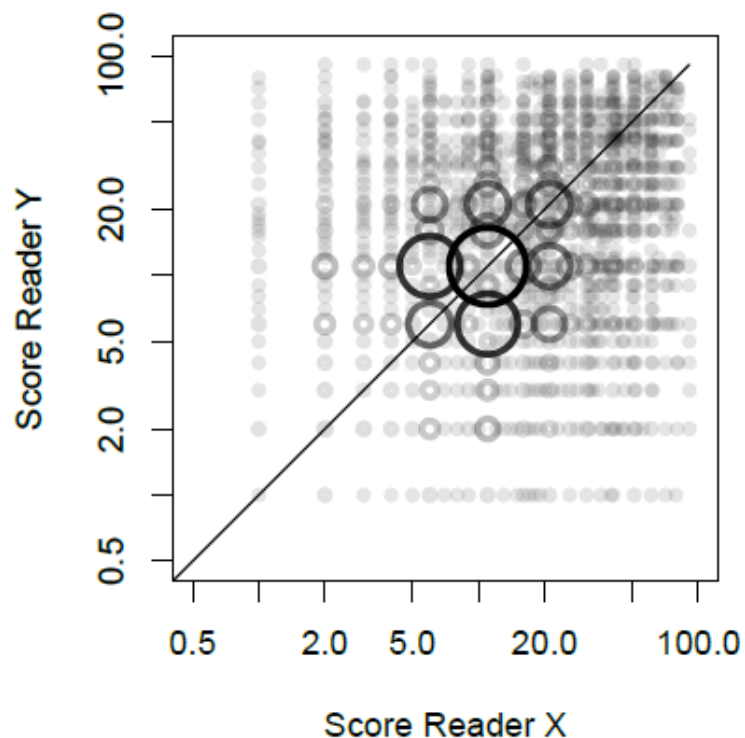
n = 4274 , Largest symbol == 343 observations

Scatter Plots

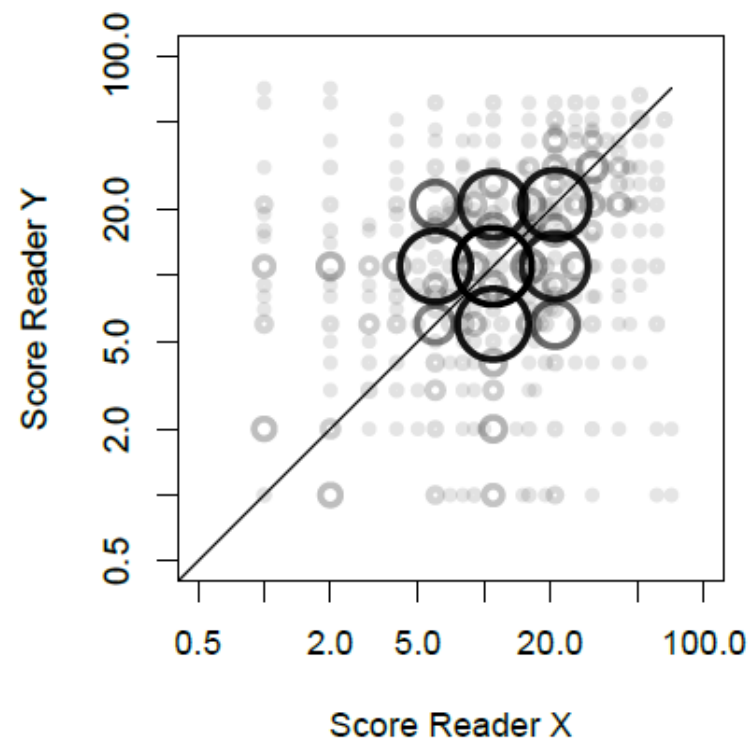
$10 < \text{scores} \leq 40$

Symmetrized: We plot (x,y) and (y,x) since we are pooling over readers and none is the reference.

Size of symbol and transparency are scaled with number of paired observations



n = 20080 , Largest symbol == 982 observations



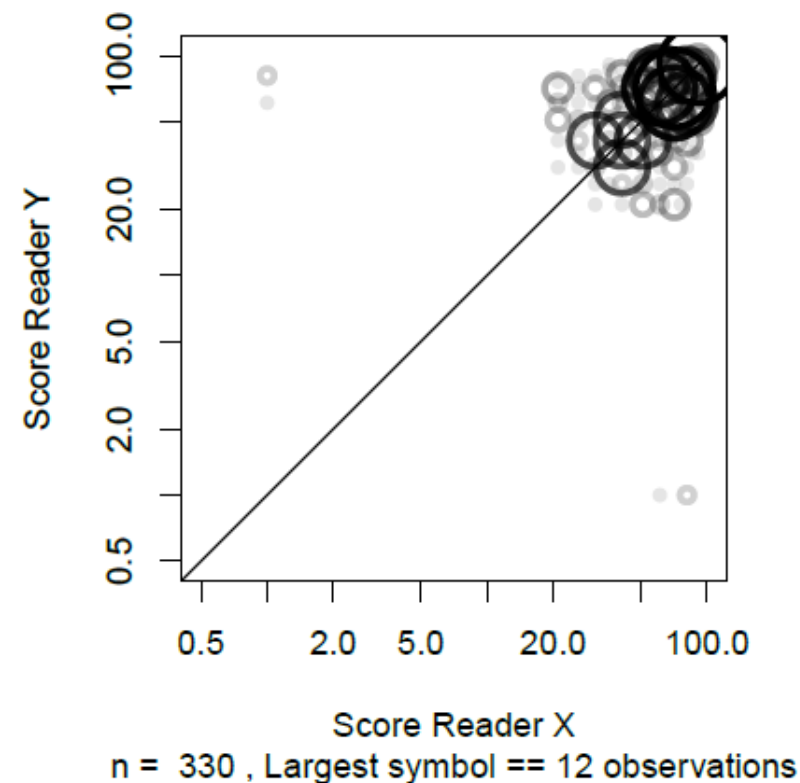
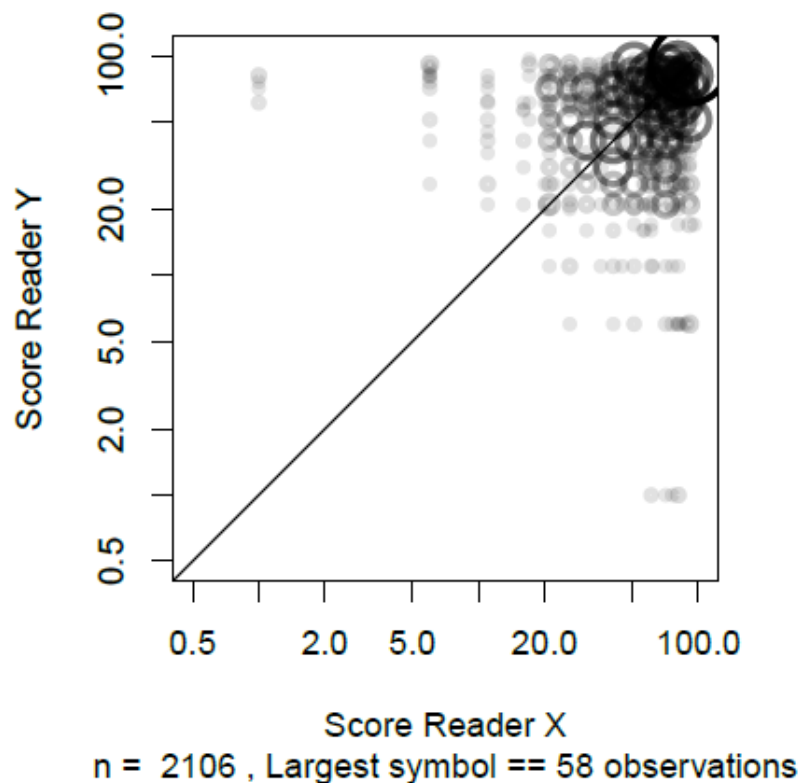
n = 1554 , Largest symbol == 70 observations

Scatter Plots

$40 < \text{scores} \leq 100$

Symmetrized: We plot (x,y) and (y,x) since we are pooling over readers and none is the reference.

Size of symbol and transparency are scaled with number of paired observations



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Limits Of Agreement

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

- LOA accounts for reader and case variability
- LOA reduced by 30% with panel (except for high scores)
- What's the precision of these estimates?
- Still need to account for correlations between ROIs in an image

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

Q: Are these limits clinically acceptable?

- A: Discuss with partners and community



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

Q: Are these limits clinically acceptable?

- Compare to other studies
 - Denkert et al, **Ring Study**, **Modern Pathology**, 2016.

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
0 ≤ scores ≤ 10	18.3	12.4
10 < scores ≤ 40	38.0	27.0
40 < scores ≤ 100	66.2	62.6

- Align with previous work



PROGNOSIS TOOL for Triple Negative Breast Cancer (TNBC)

Welcome to the online TIL and Prognosis tool for TNBC.

tilsinbreastcancer.org

Q: Are these limits clinically acceptable?

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

A: Find Breast Cancer experts

A: Find TIL evaluation experts

Q: How can we tighten LOAs?



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

Q: How can we tighten LOAs?

A: Improve training

- Emphasize calibration cheat sheet
- Test with feedback.
- Proficiency test

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

Q: How can we tighten LOAs?

A: Improve training



Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

- Which pathologists are interchangeable with the panel?
- Which algorithm is interchangeable with the panel?

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Questions and Current Work

	Limits of Agreement (Point Estimates)	
	All Readers	Panel of Four
	Score differences	Score differences
$0 \leq \text{scores} \leq 10$	18.3	12.4
$10 < \text{scores} \leq 40$	38.0	27.0
$40 < \text{scores} \leq 100$	66.2	62.6

- To Investigate:
 - Image-based assessment
 - Average ROIs per image
- To Investigate:
 - Rank-based correlation agreement metrics
 - Smaller evaluation intervals
 - Agreement Rates per Interval

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021

Next Steps

- Update Pathologist Training
 - Immediate
 - Emphasize the calibration cheat sheet
 - For pivotal study
 - Test with feedback
 - Proficiency test
- Continue with pilot study
 - Collect more PathPresenter data
 - Collect microscope-mode data
 - Road trip!
 - Looking for sites and pathologists to help with data collection
- Finalize pivotal study statistical analysis plan
 - Determine study size and power
 - Simulation methods
- Get feedback from the community (including MDDT)
- Source and curate pivotal study slides
 - Looking for one or two more sites
- Plan and execute data-collection

Conclusions

- Continue to make progress on this challenging project
 - Many thanks to all the collaborators
 - Are you interested in getting involved?
- We have collected 7,259 pathologist evaluations (and counting)
 - Building platforms and pipelines
 - Learning about pathologist agreement
 - Developing methods
- We plan to leverage the platforms, pipelines, methods, experience, **and relationships**
 - Other quantitative biomarkers
 - Other pathologist evaluations (qualitative biomarkers, marks, segmentations)

Gallas et al. HTT project - Pathology Informatics Summit - May 6, 2021