The Many Efforts to Create a Dataset to Validate AI/ML Models in Digital Pathology

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

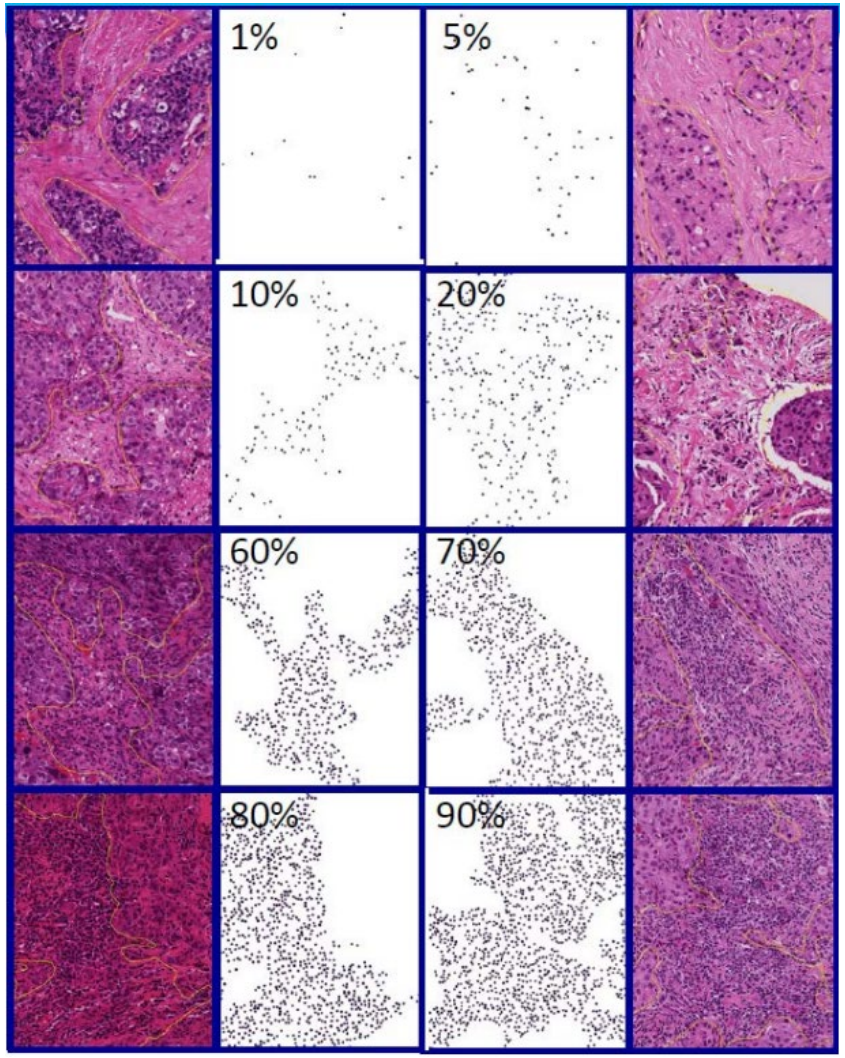
Disclaimers

- 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. This is a contribution of the U.S. Food and Drug Administration and is not subject to copyright.
- This is a contribution of the U.S. Food and Drug Administration and is not subject to copyright.

Outline

- HTT: High-Throughput Truthing project
  - Overview
  - Pilot Study: Pathologist Variability
  - Deep Dive Expert Panel Sessions
- Training Materials
- Pivotal Study Data
- Data Curation
- Project Resources
- Summary

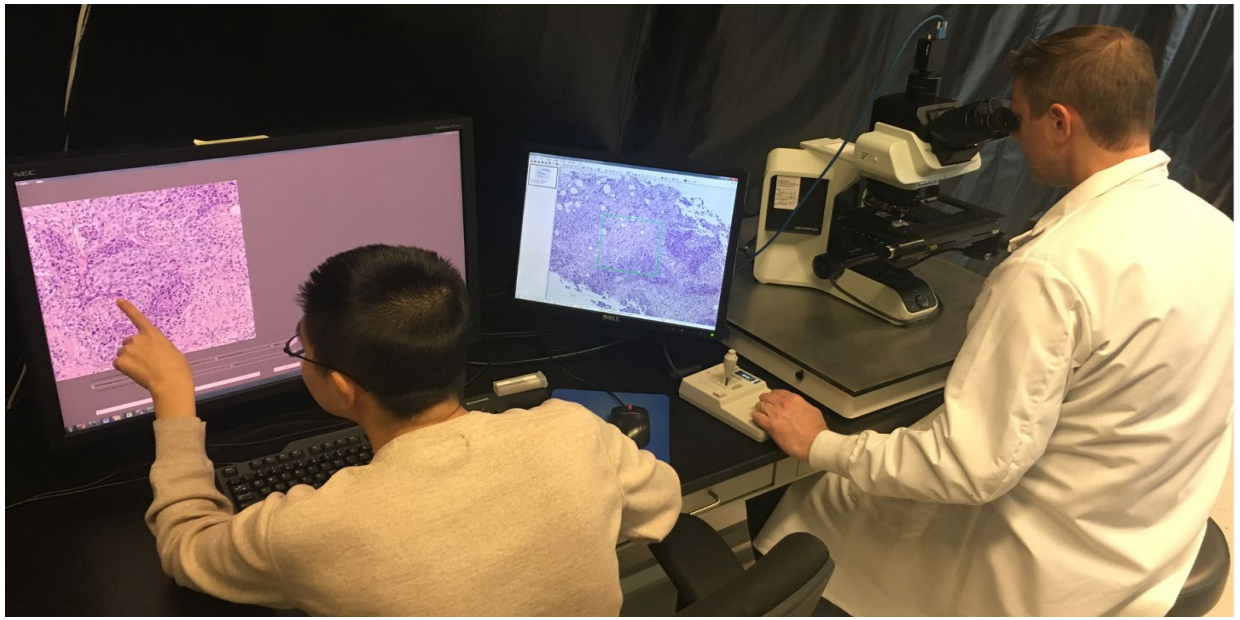
Quantitative Biomarker
TILS: Tumor Infiltrating Lymphocytes



OSEL Accelerating patient access to innovative, safe, and effective medical devices through best-in-the-world regulatory science

High-Throughput Truthing (HTT) Project

- Clinical context:
  - Breast cancer
  - Quantitative Pathology Biomarker: Stromal Tumor Infiltrating Lymphocytes (sTILs)
- Clinical relevance of sTILs:
  - Prognostic for survival
  - Expected to inform patient management
  - Expected to reduce use of toxic chemotherapies
- Biomarker Evaluation by an Algorithm
  - Reduce burden on pathologist
  - Reproducible
  - Quantitative



- Deliverables/Tools
  - Reference standard data set from pathologists
  - Data-collection methods and platforms
  - Methods to validate a quantitative algorithm

References for Clinical Context of TILs in Breast Cancer

1. Mao, Y.; Qu, Q.; Chen, X.; Huang, O.; Wu, J.; Shen, K. The Prognostic Value of Tumor-Infiltrating Lymphocytes in Breast Cancer: A Systematic Review and Meta-Analysis. PLoS ONE **2016**, 11, e0152500. [CrossRef] [PubMed]
2. Loi, S.; Drubay, D.; Adams, S.; Pruneri, G.; Francis, P.A.; Lacroix-Triki, M.; Joensuu, H.; Dieci, M.V.; Badve, S.; Demaria, S.; et al. Tumor-Infiltrating Lymphocytes and Prognosis: A Pooled Individual Patient Analysis of Early-Stage Triple-Negative Breast Cancers. J. Clin. Oncol. **2019**, 37, 559–569. [CrossRef] [PubMed]
3. Savas, P.; Salgado, R.; Denkert, C.; Sotiriou, C.; Darcy, P.K.; Smyth, M.J.; Loi, S. Clinical Relevance of Host Immunity in Breast Cancer: From TILs to the Clinic. Nat. Rev. Clin. Oncol. **2016**, 13, 228–241. [CrossRef] [PubMed]
4. Hendry, S.; Salgado, R.; Gevaert, T.; Russell, P.A.; John, T.; Thapa, B.; Christie, M.; Estrada, M.; Gonzalez-Ericsson, P.; Sanders, M.; et al. Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immunooncology Biomarkers Working Group: Part 1: Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma in Situ, Metastatic Tumor Deposits and Areas for Further Research. Adv. Anat. Pathol. **2017**, 24, 235–251. [CrossRef]
5. Stanton, S.E.; Disis, M.L. Clinical Significance of Tumor-Infiltrating Lymphocytes in Breast Cancer. J. Immunother. Cancer **2016**, 4, 59. [CrossRef]
6. Lotfinejad, P.; Asghari Jafarabadi, M.; Abdoli Shadbad, M.; Kazemi, T.; Pashazadeh, F.; Sandoghchian Shotorbani, S.; Jadidi Niaragh, F.; Baghbanzadeh, A.; Vahed, N.; Silvestris, N.; et al. Prognostic Role and Clinical Significance of Tumor-Infiltrating Lymphocyte (TIL) and Programmed Death Ligand 1 (PD-L1) Expression in Triple-Negative Breast Cancer (TNBC): A Systematic Review and Meta-Analysis Study. Diagnostics **2020**, 10, 704. [CrossRef]
7. Denkert, C.; von Minckwitz, G.; Darb-Esfahani, S.; Lederer, B.; Heppner, B.I.;Weber, K.E.; Budczies, J.; Huober, J.; Klauschen, F.; Furlanetto, J.; et al. Tumour-Infiltrating Lymphocytes and Prognosis in Different Subtypes of Breast Cancer: A Pooled Analysis of 3771 Patients Treated with Neoadjuvant Therapy. Lancet Oncol. **2018**, 19, 40–50. [CrossRef]
8. Wein, L.; Savas, P.; Luen, S.J.; Virassamy, B.; Salgado, R.; Loi, S. Clinical Validity and Utility of Tumor-Infiltrating Lymphocytes in Routine Clinical Practice for Breast Cancer Patients: Current and Future Directions. Front. Oncol. **2017**, 7, 156. [CrossRef]
9. Park, J.H.; Jonas, S.F.; Bataillon, G.; Criscitiello, C.; Salgado, R.; Loi, S.; Viale, G.; Lee, H.J.; Dieci, M.V.; Kim, S.-B.; et al. Prognostic Value of Tumor-Infiltrating Lymphocytes in Patients with Early-Stage Triple-Negative Breast Cancers (TNBC) Who Did Not Receive Adjuvant Chemotherapy. Ann. Oncol. **2019**, 30, 1941–1949. [CrossRef]
10. Luen, S.J.; Salgado, R.; Dieci, M.V.; Vingiani, A.; Curigliano, G.; Gould, R.E.; Castaneda, C.; D'Alfonso, T.; Sanchez, J.; Cheng, E.; et al. Prognostic Implications of Residual Disease Tumor-Infiltrating Lymphocytes and Residual Cancer Burden in Triple-Negative Breast Cancer Patients after Neoadjuvant Chemotherapy. Ann. Oncol. **2019**, 30, 236–242. [CrossRef]
11. Denkert, C.; von Minckwitz, G.; Brase, J.C.; Sinn, B.V.; Gade, S.; Kronenwett, R.; Pfitzner, B.M.; Salat, C.; Loi, S.; Schmitt,W.D.; et al. Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant ChemotherapyWith orWithout Carboplatin in Human Epidermal Growth Factor Receptor 2–Positive and Triple-Negative Primary Breast Cancers. JCO **2015**, 33, 983–991. [CrossRef]
12. Balic, M.; Thomssen, C.; Würstlein, R.; Gnant, M.; Harbeck, N. St. Gallen/Vienna 2019: A Brief Summary of the Consensus Discussion on the Optimal Primary Breast Cancer Treatment. Breast Care **2019**, 14, 103–110. [CrossRef]
13. Cardoso, F.; Kyriakides, S.; Ohno, S.; Penault-Llorca, F.; Poortmans, P.; Rubio, I.T.; Zackrisson, S.; Senkus, E. Early Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. Ann. Oncol. **2019**, 30, 1194–1220. [CrossRef]
14. Morigi, C. Highlights of the 16th St Gallen International Breast Cancer Conference, Vienna, Austria, 20–23 March 2019: Personalised Treatments for Patients with Early Breast Cancer. Ecancermedicalscience **2019**, 13, 924. [CrossRef]
15. Salgado, R.; Denkert, C.; Demaria, S.; Sirtaine, N.; Klauschen, F.; Pruneri, G.; Wienert, S.; Van den Eynden, G.; Baehner, F.L.; Penault-Llorca, F.; et al. The Evaluation of Tumor-Infiltrating Lymphocytes (TILs) in Breast Cancer: Recommendations by an International TILsWorking Group 2014. Ann. Oncol. **2015**, 26, 259–271. [CrossRef]

Collaborators – Current and Past

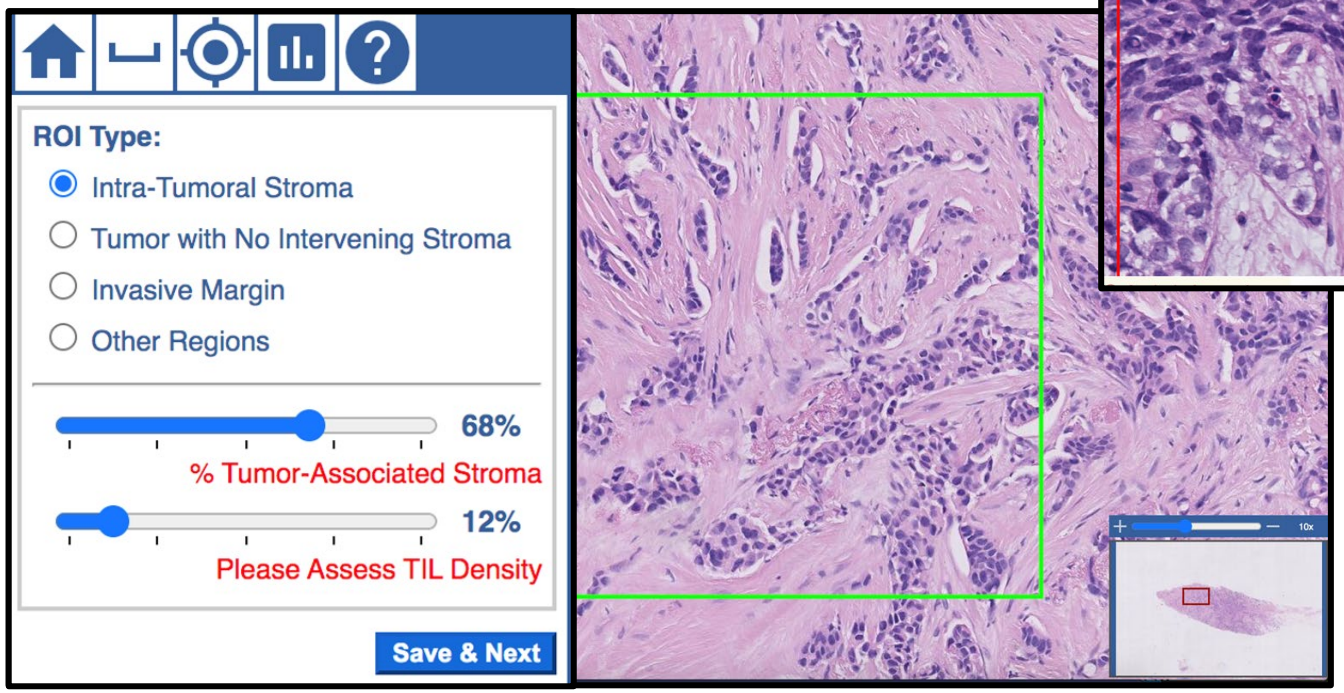
Pathologists, Academics, Industry, International

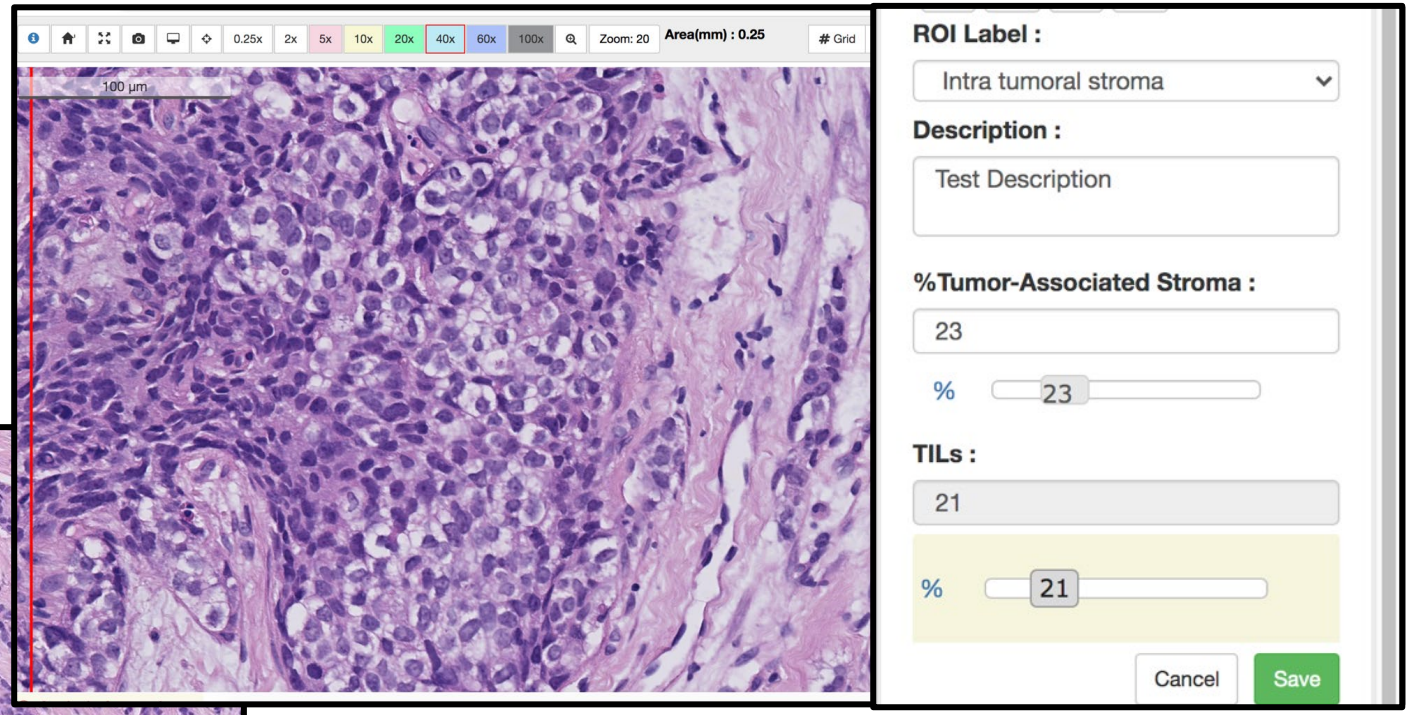
Volunteers

- **Mohamed Amgad, MD, PhD**
  - Northwestern University - The Feinberg School of Medicine
- **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
- **Kate Elfer, MPH**
  - FDA/CDRH/OSEL/DIDSR
- **Anna Ehinger**
  - Lund University
- **Emma Gardecki, BS**
  - FDA/CDRH/OSEL/DIDSR
- **Victor Garcia, MD**
  - 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
- **Denis Larsimont, MD**
  - Department of Pathology, Institut Jules Bordet
- **Xiaoxian Li, MD/PhD**
  - Emory University School of Medicine
- **Amy Ly, MD**
  - Massachusetts General Hospital
- **Anant Madabhushi, PhD**
  - Case Western Reserve University
- **Hetal Marble, PhD**
  - Immuto Scientific
- **Dieter Pieters**
  - Sint-Maarten Hospital; University of Antwerp; CellCarta
- **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**
  - PathPresenter Corporation
- **Evan Szu, PhD**
  - Arrive Bio
- **Darick Tong, MS**
  - Arrive Bio
- **Si Wen, PhD**
  - FDA/CDRH/OSEL/DIDSR
- **Bruce Werness, MD**
  - Arrive Bio

Data-Collection Platforms: Digital

caMicroscope: Open Source
https://github.com/camicroscope/caMicroscope

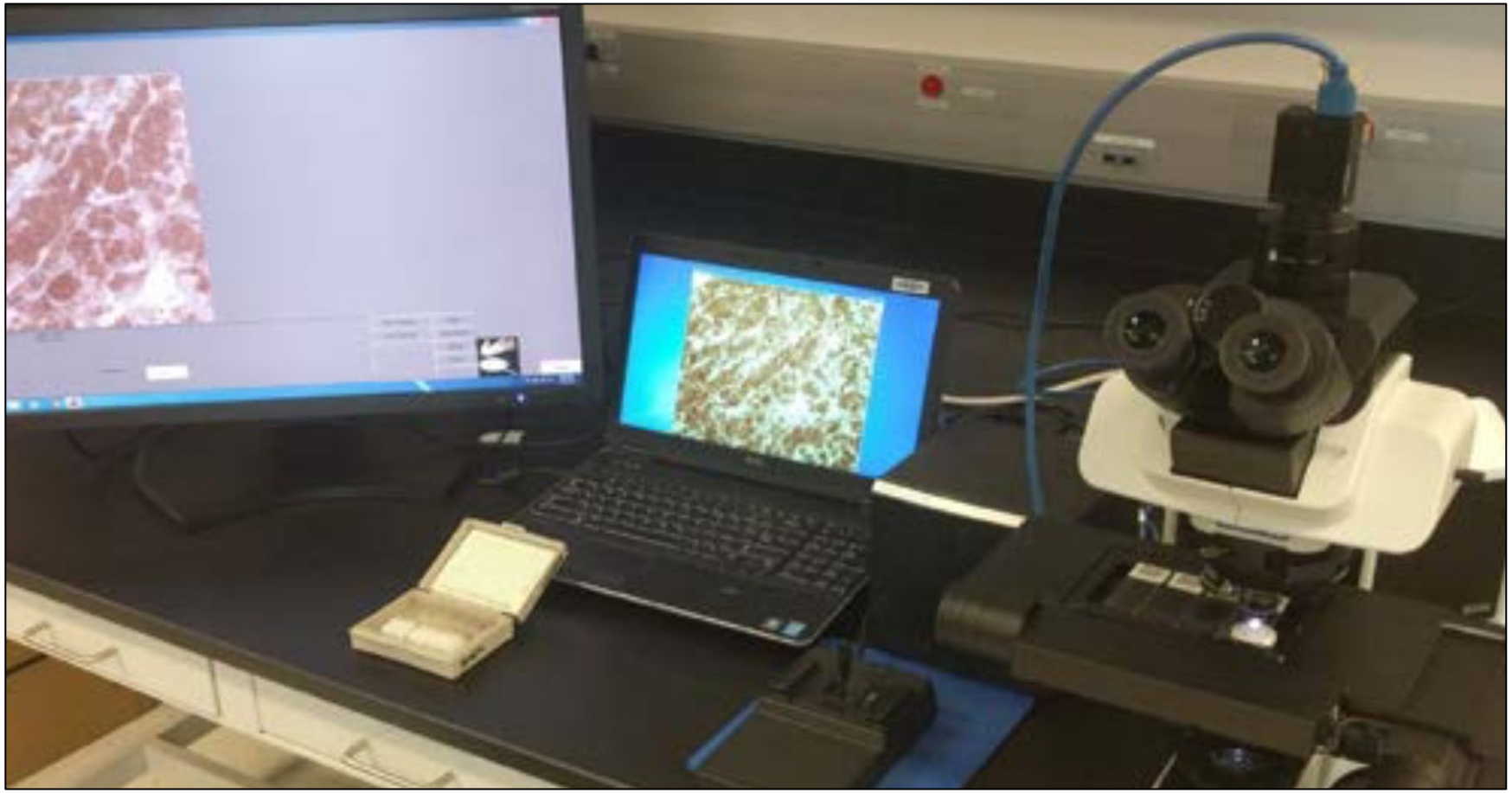




PathPresenter:
https://pathpresenter.net/about

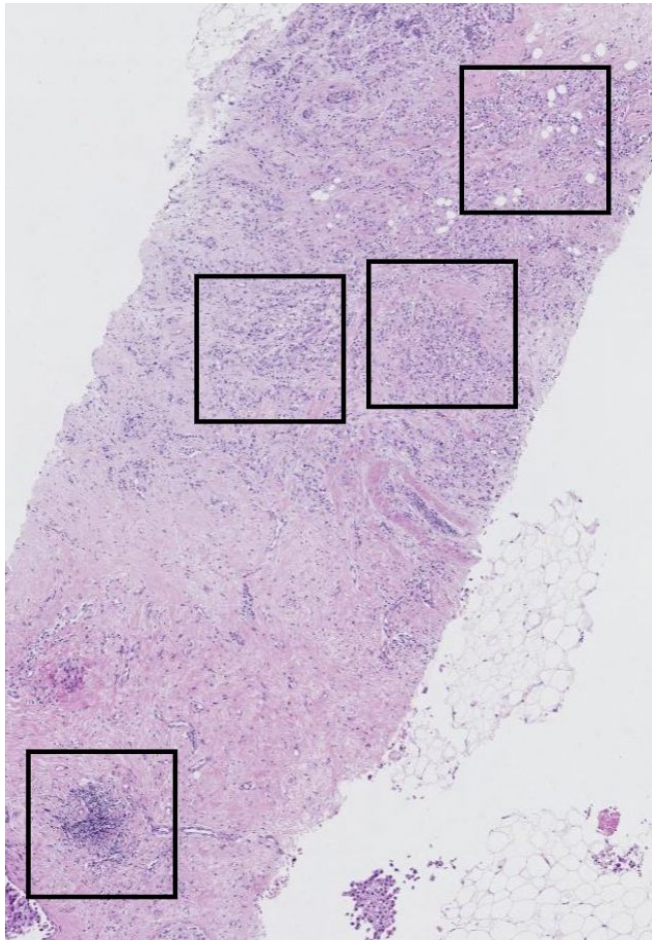
Data-Collection Platforms: Microscope

- Registers stage coordinates with whole slide image via camera
- Allows replication of the digital-mode study design
- Computer drives the stage from ROI to ROI
- Annotations are independent of the scanner and viewer



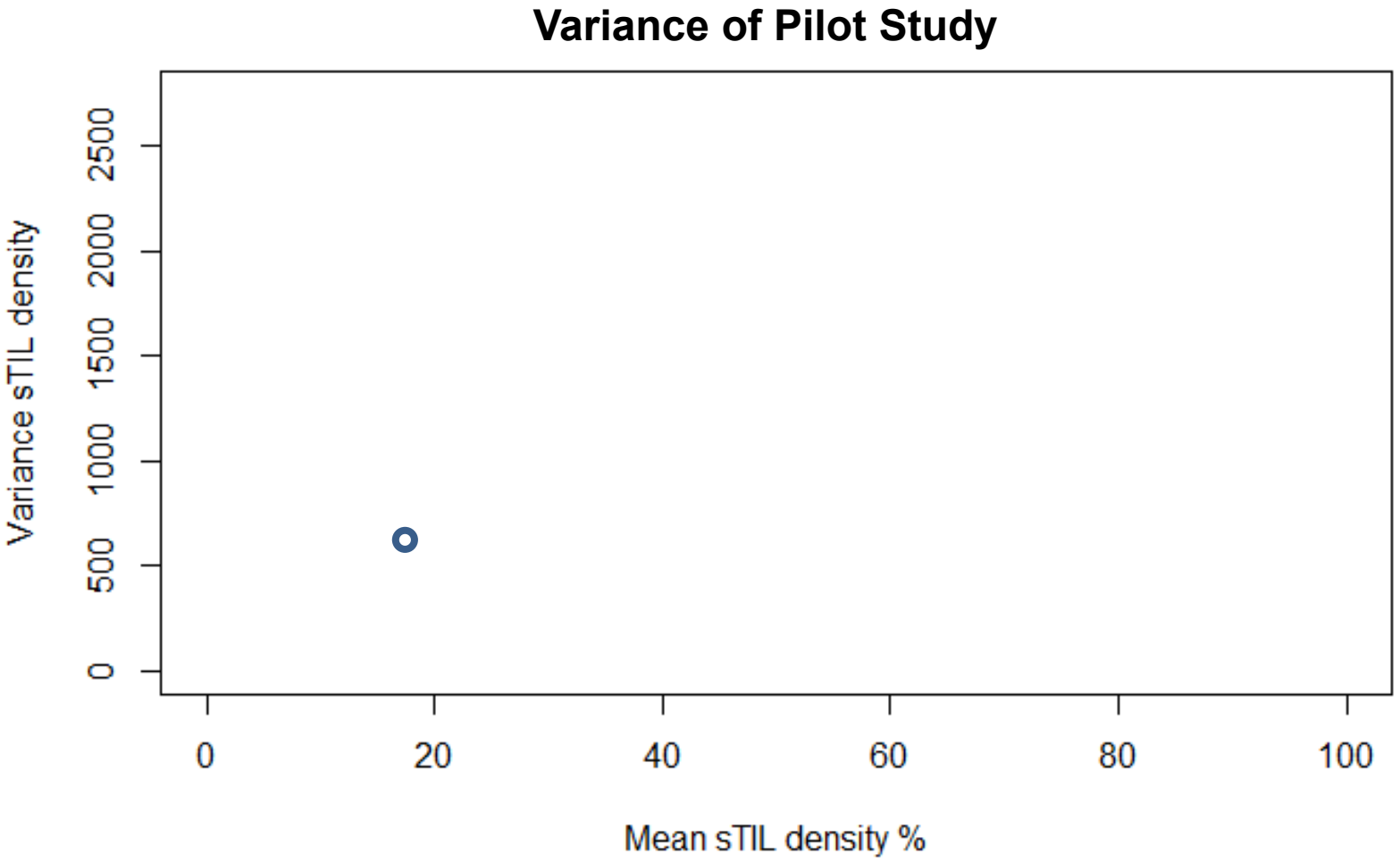
Pilot Study

- Cases:
  - 64 H&E Slides
  - 10 Regions of Interest (ROIs) per Slide
  - Some ROIs are not appropriate for sTIL evaluation
- Evaluation Platforms:
  - 2 digital and 1 microscope
- Readers:
  - 37 readers
  - 7 crowd readers with complete data
  - 7 expert readers are on the collaboration team
- 7,898 Observations



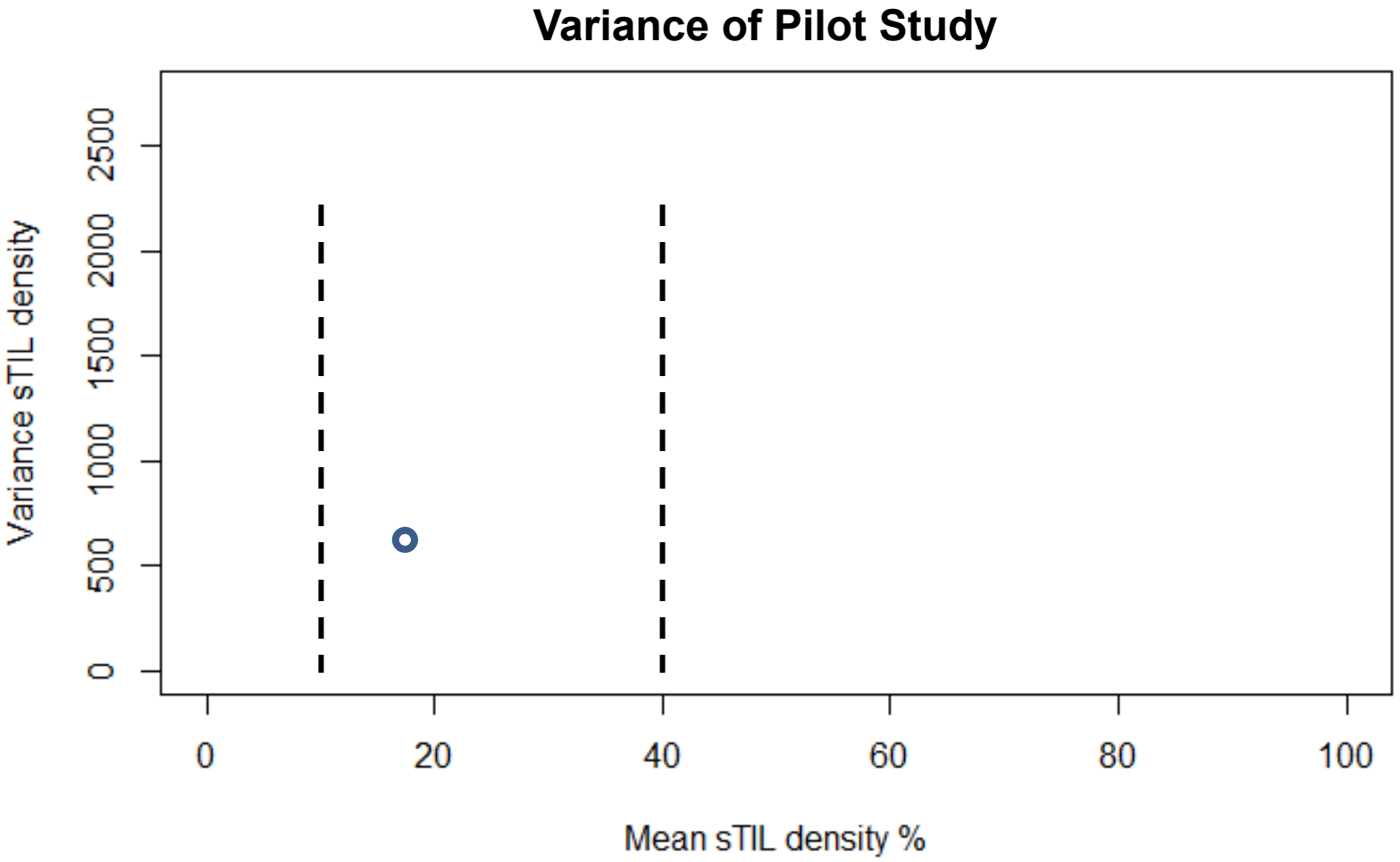
R Data Package
https://github.com/DIDSR/HTT

Initial Analysis of Pilot Study



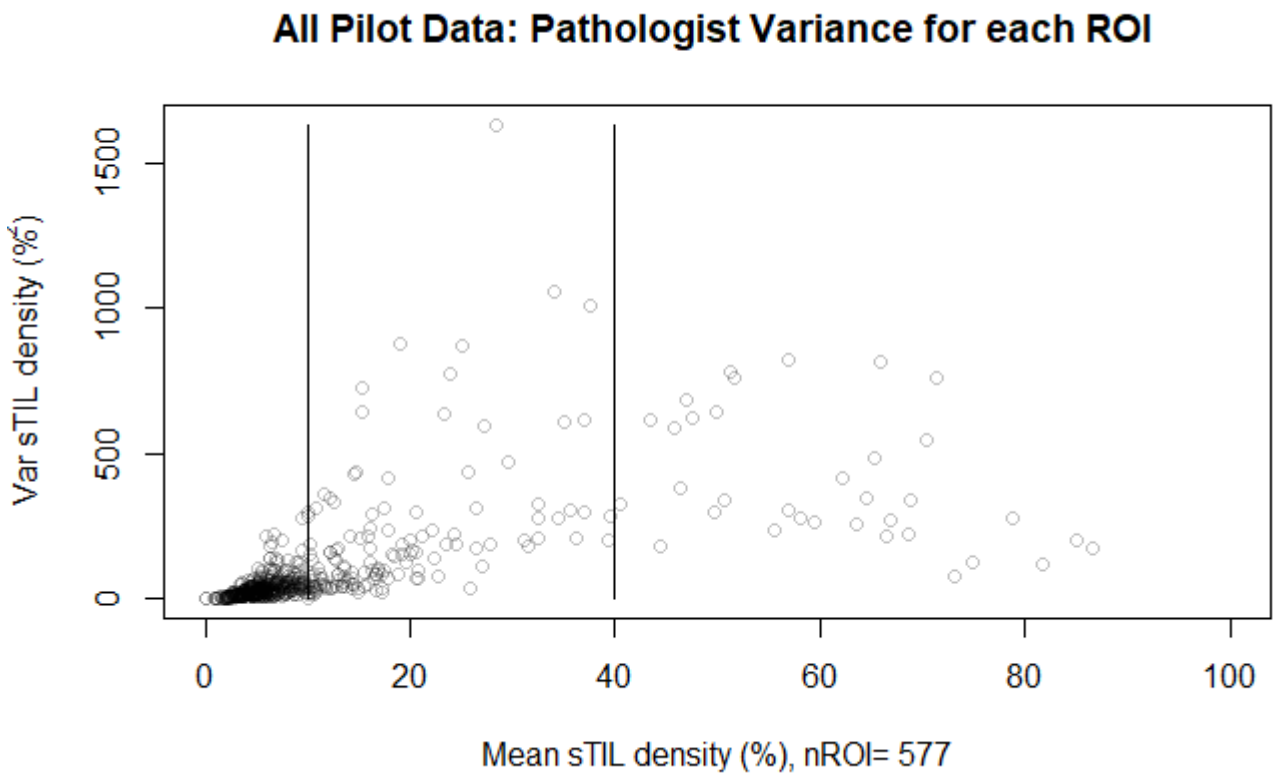
- Mean and Variance are averages over all readers

Initial Analysis of Pilot Study



- Mean and Variance are averages over all readers
- Vertical dashed lines represent clinical bins
  - low (≤ 10%)
  - medium (>10% & ≤ 40%)
  - high (>40%)Horizontal

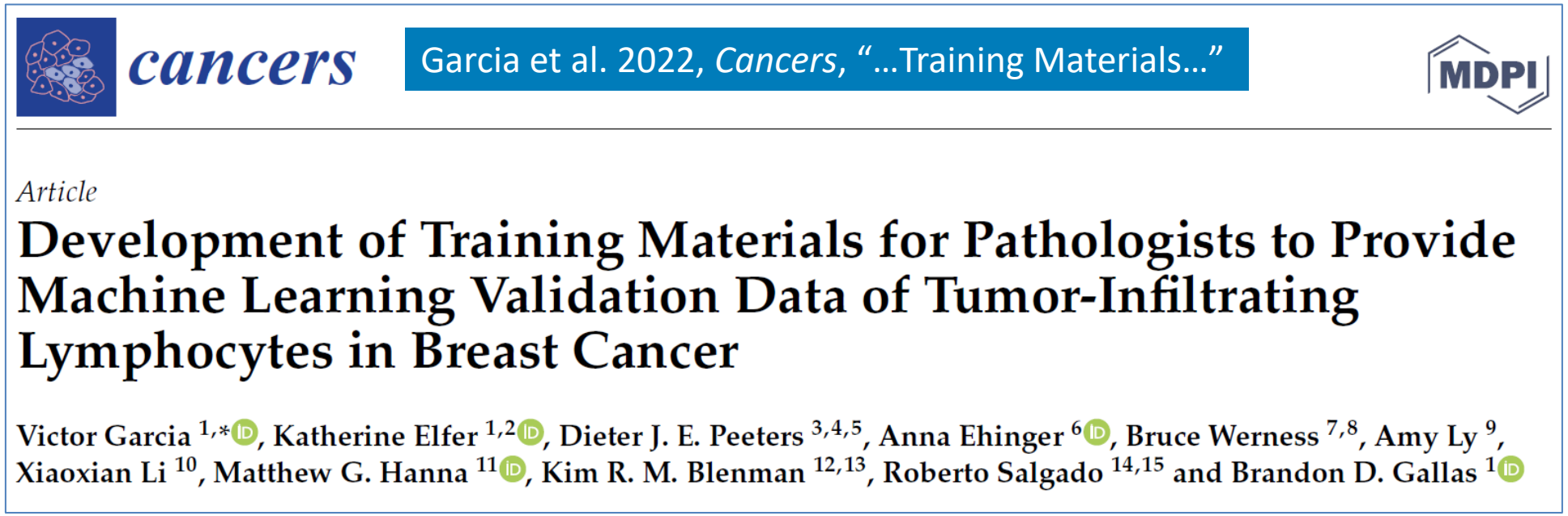
Initial Analysis of Pilot Study



- Means and Variances are averages over all readers
- Vertical lines represent clinical bins
  - low (≤ 10%)
  - medium (>10% & ≤ 40%)
  - high (>40%)
- Variance is increasing with the mean

Pilot Study Deep Dive: Expert Panel Sessions

- Primary purpose
  - Understand pathologist variability
  - Improve instructions to reduce variability
- Subsequent Opportunities
  - Clinical practice training materials
  - Reference standard for pilot study
  - Explore analysis methods



cancers

Garcia et al. 2022, *Cancers*, "…Training Materials…"

MDPI

Article

Development of Training Materials for Pathologists to Provide Machine Learning Validation Data of Tumor-Infiltrating Lymphocytes in Breast Cancer

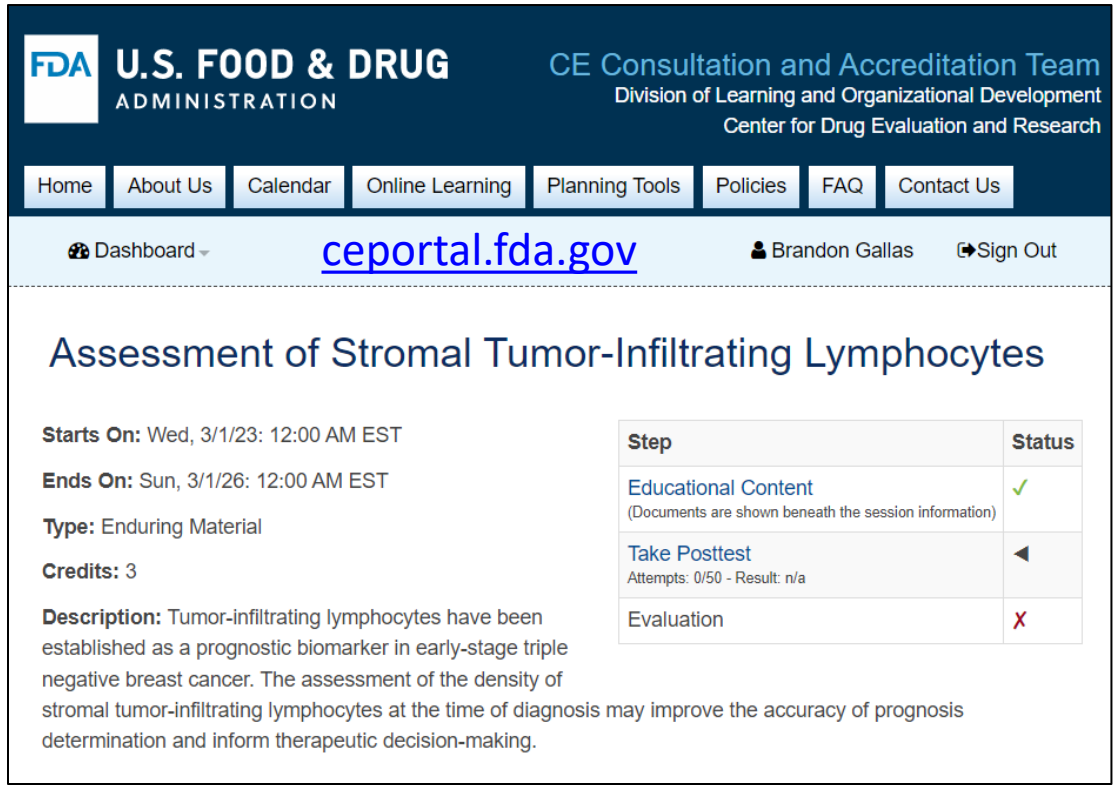
Victor Garcia [1,*], Katherine Elfer [1,2], Dieter J. E. Peeters [3,4,5], Anna Ehinger [6], Bruce Werness [7,8], Amy Ly [9], Xiaoxian Li [10], Matthew G. Hanna [11], Kim R. M. Blenman [12,13], Roberto Salgado [14,15] and Brandon D. Gallas [1]

CME Course:

Assessment of Stromal Tumor-Infiltrating Lymphocytes

Objectives

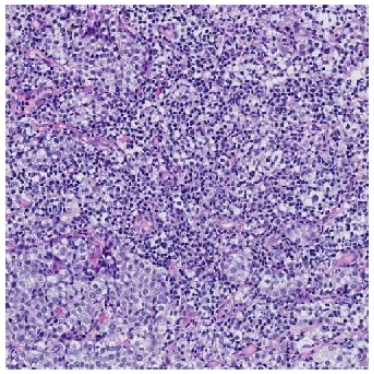
- Describe the **significance** of stromal tumor-infiltrating lymphocytes in triple negative breast cancer.
- Demonstrate knowledge of the **approach** to determining the density of stromal tumor-infiltrating lymphocytes.



Faculty

- Victor Garcia, MD
- Amy Ly, MD
- Matthew Hanna, MD
- Dieter Peeters, MD, PhD
- Roberto Salgado, MD, PhD
- Xiaoxian Li, MD, PhD
- Kim Blenman, PhD, MS
- Katherine Elfer, PhD, MPH
- Bruce Werness, MD
- Anna Ehinger, MD
- Brandon Gallas, PhD

sTILs Reference Document and Pitfalls



caseID: HTT-TILS-001-04B.ndpi_x24343.2190_y11775.2190

Expert Panel Annotations

ROI Type	Percent Tumor-Associated Stroma	sTILs Density
Evaluable	30	90
Evaluable	60	95
Evaluable	50	92
Evaluable	50	75
Evaluable	60	90
Evaluable	60	90

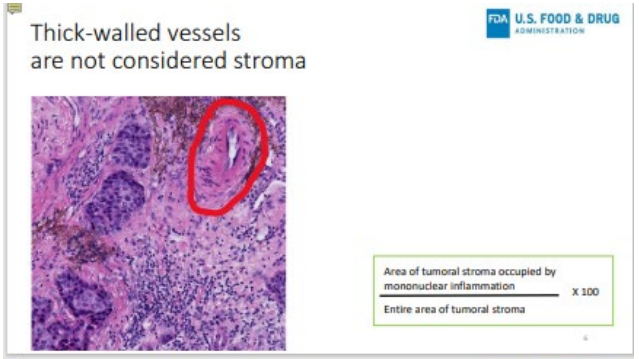
Mean Percent Tumor-Associated Stroma: 51.7

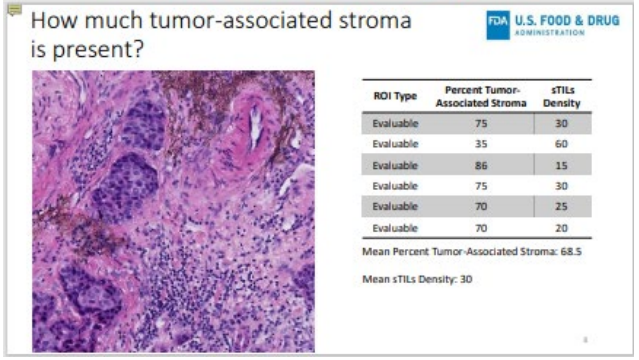
Mean sTILs Density: 88.7

Comments: A challenging case. The high density of lymphocytes results in difficulty determining whether the lymphocytes are located in stroma, or whether they infiltrate tumor cell nests. The presence of small blood vessels and small gaps between lymphocytes suggest the lymphocytes reside within stroma. Occasional tumor cells with small nuclei (possibly degenerating) may be confused for lymphocytes.

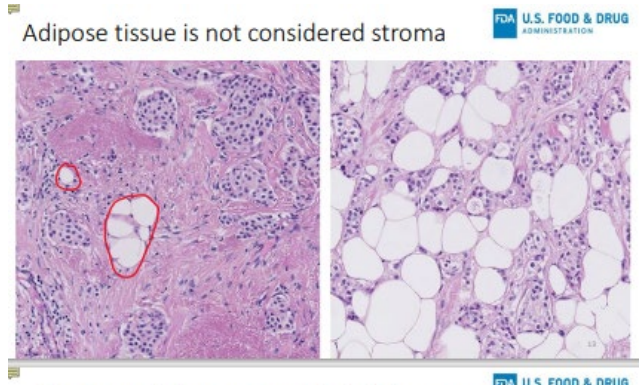
Pitfalls: In regions where the sTILs density is very high, the underlying stroma may be obscured. Non-lymphocytes with small nuclei may be confused for lymphocytes.

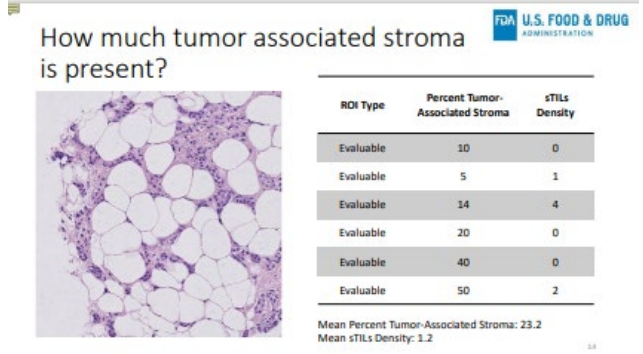
2



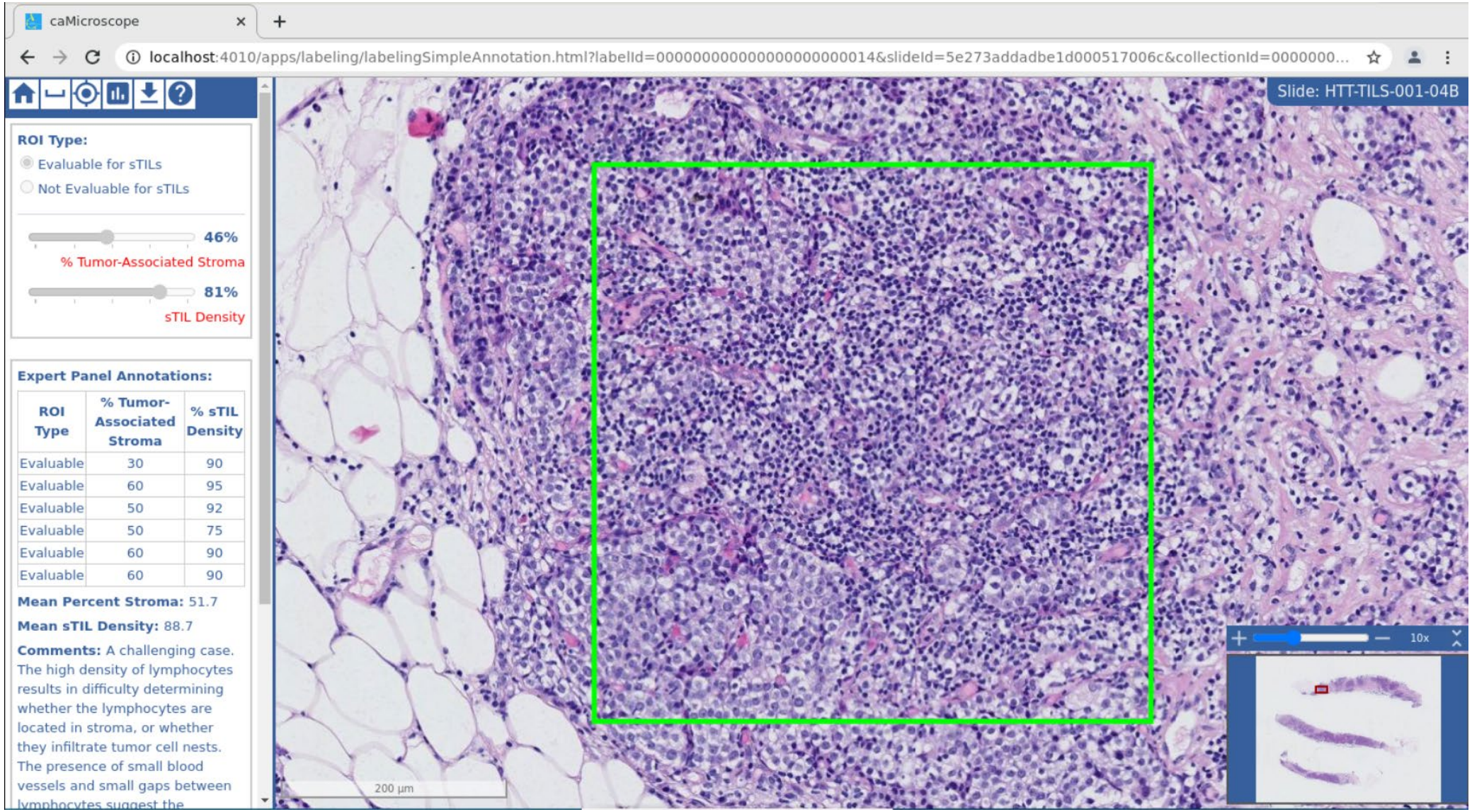


Example Pitfalls





Interactive Test with Feedback and Proficiency Test



Pathologist-Specific Test Reports

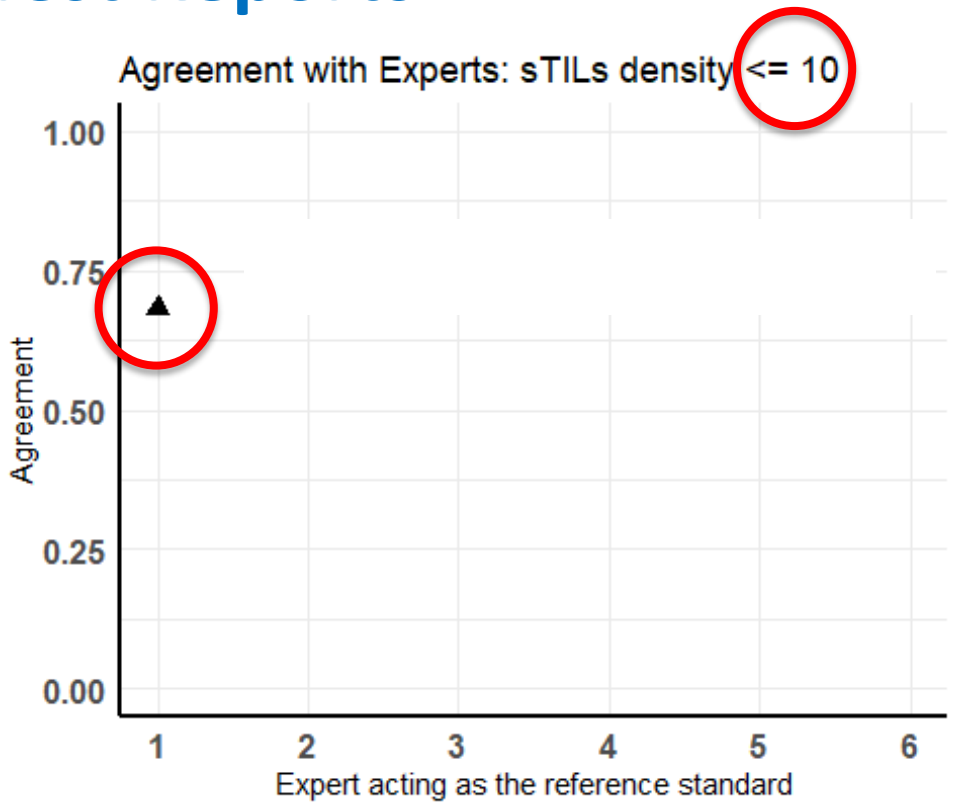
Threshold = 10	reader. NotEvaluable	reader. LE	reader. GT	Fraction Agree	Rate Agree
expert. GT	0	2	11	11/13	0.846
expert. LE	3	15	4	15/22	0.682
expert. NotEvaluable	0	0	0	4/4	NA

Primary performance metrics

- Apply threshold
- Create 3x3 table
- Determine rates of agreement
  - LE threshold
  - GT threshold
  - Not Evaluable
- Repeat for all experts
- Consider other thresholds
- Consider multiple thresholds

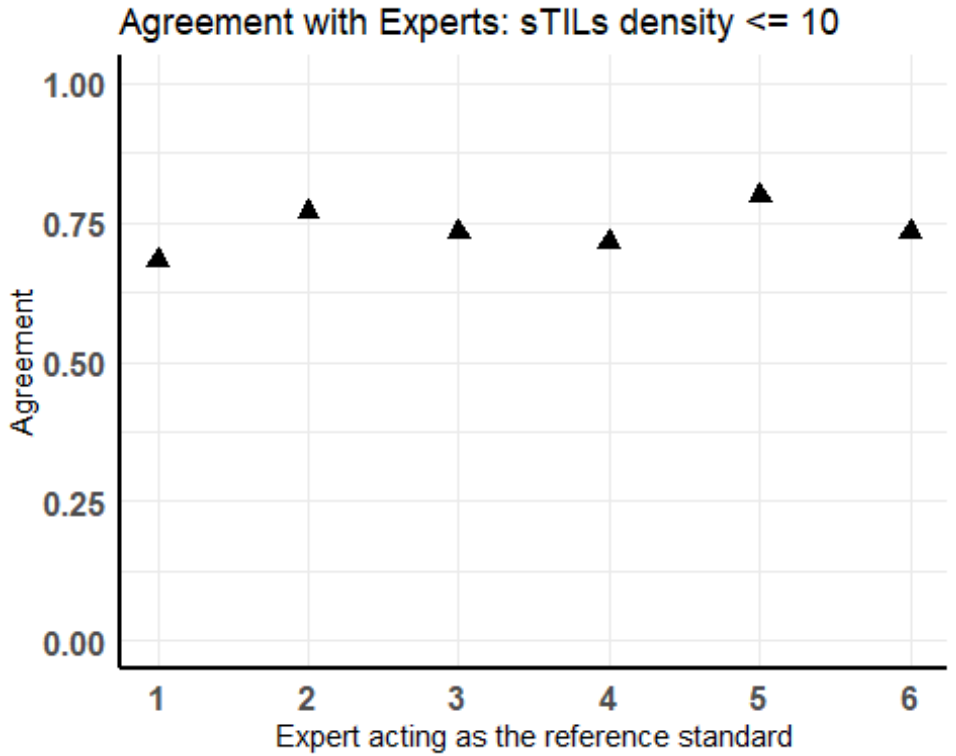
Pathologist-Specific Test Reports

Threshold = 10	reader. NotEvaluable	reader. LE	reader. GT	Fraction Agree	Rate Agree
expert. GT	0	2	11	11/13	0.846
expert. LE	3	15	4	15/22	0.682
expert. NotEvaluable	0	0	0	4/4	NA



Pathologist-Specific Test Reports

- Black Triangles
  - Reader vs. Experts Agreement



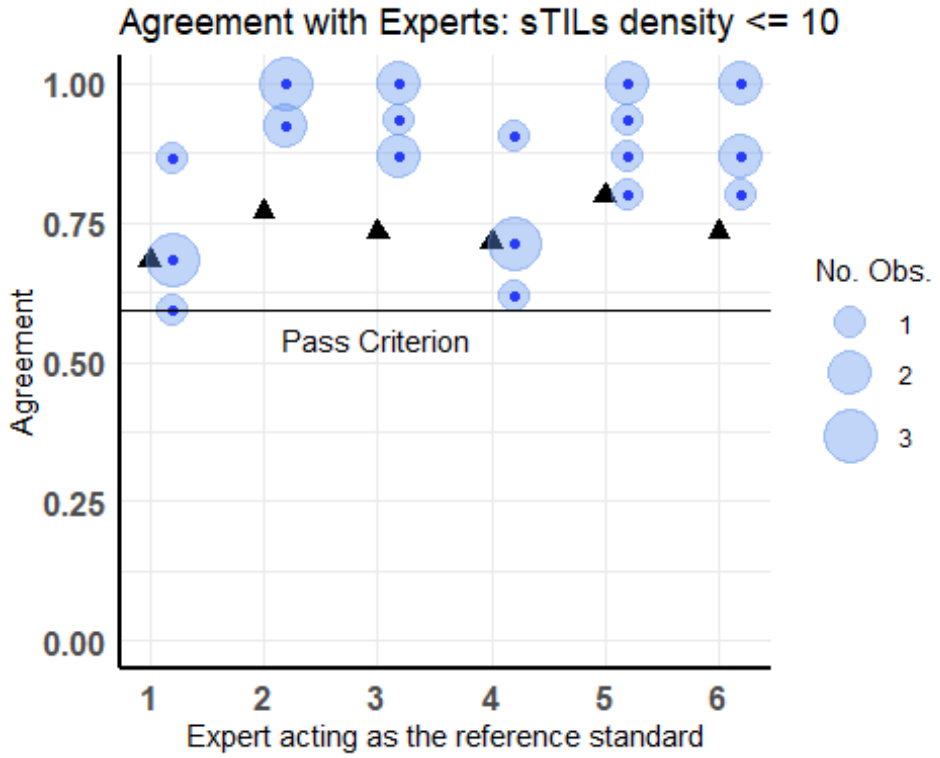
Pathologist-Specific Test Reports

- Black Triangles
  - Reader vs. Experts Agreement
- Blue Circles:
  - Experts vs. Experts Agreement

- Four criteria
  - sTILs density ≤ 10
  - sTILs density > 10
  - sTILs density ≤ 40
  - sTILs density > 40

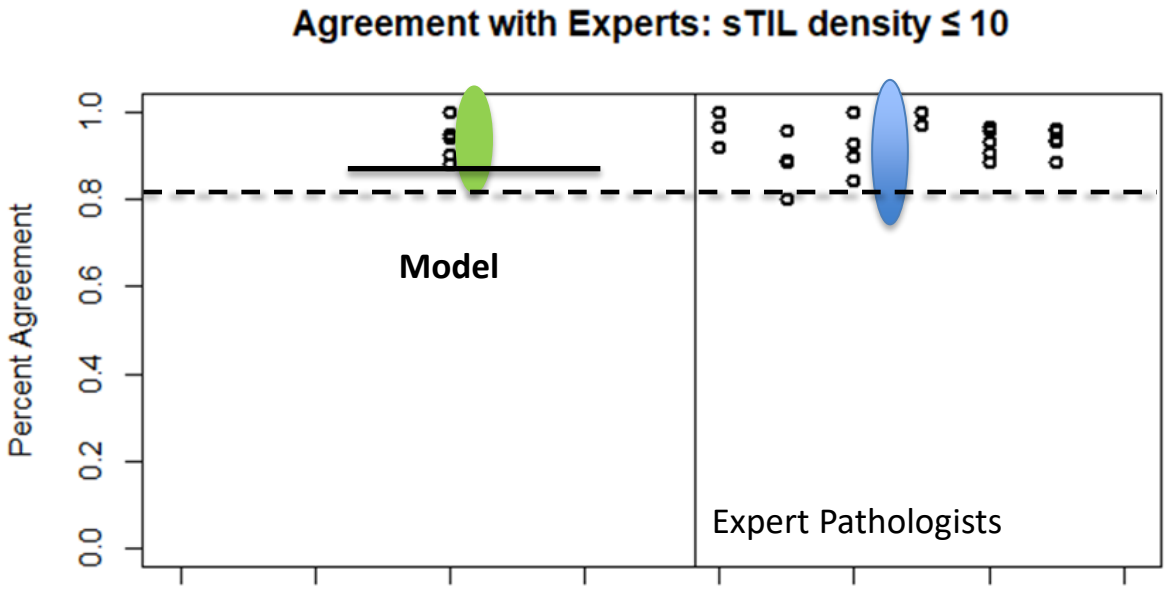
- Pathologist-specific performance reports
  - Feedback test includes reader and expert data

- Requirements for HTT participants
  - Take and pass CME course
  - Take and pass proficiency test



AI/ML model assessment

- *Under development*
  - *Model Distributions*
- Multi-Expert Multi-Case (MEMC) analysis method
  - Account for expert and case variability and correlations
  - For each threshold and agreement above and below (multiple hypotheses) …
  - Study result is the lower 97.5 percentile of model-to-expert agreement
  - Comparator is the lower 97.5 percentile of expert-expert agreement



OSEL Accelerating patient access to innovative, safe, and effective medical devices through best-in-the-world regulatory science

Pivotal Study Data

- Inclusion Criteria
  - Core biopsies of triple negative breast cancer (TNBC: ER/PR/HER2 negative)
  - Slides that have been stained with hematoxylin and eosin within the last 7 years
- Exclusion Criteria
  - Tissue collected after administration of any therapy (e.g., neoadjuvant, chemotherapy, radiation therapy).

Feature	Description	Possible Values
Age	Age of patient at time of sample acquisition. If patient >89 years of age, Age reported as 90.	Continuous whole numbers
Sex	Patient's sex as defined in medical records. There was no differentiation for gender.	Female Male Intersex
Race	Patient's race. More than one response allowed.	American Indian or Alaska Native Asian Black or African American Native Hawaiian or Other Pacific Islander White
Ethnicity	Patient's ethnicity	Hispanic or Latino Not Hispanic or Latino
Breast Cancer Stage	Denotes breast cancer stage (tumor, lymph node, metastasis) at time of biopsy	I II III IV

- Clinical Metadata
  - Need to demonstrate that dataset includes spectrum of clinical population
- ROI Selection
  - Protocol gives instructions to target diverse morphology
  - Collect study annotations
  - Identify pitfalls

Pivotal Study Data

- Determining reference standard is hard and takes resources
- Statistical precision determines study size

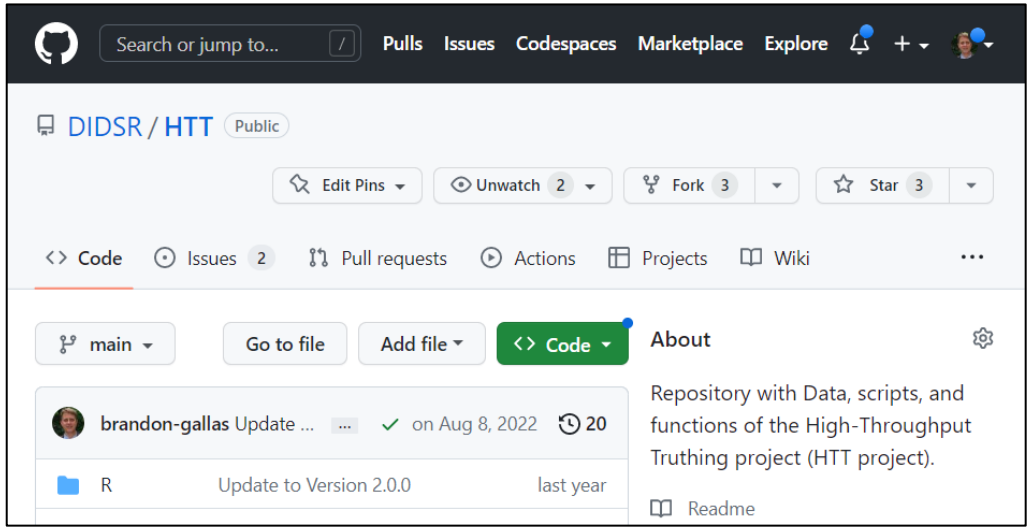
- Given a statistically determined study size, preferentially sample underrepresented cases
  - Cases with ROIs with medium and high sTILs densities
  - Cases from underrepresented groups
  - Cases with rare pitfalls

Enrichment

- Data Curation – Batch Selection Protocol
  - Priority Scoring of Cases
  - Hierarchical Sort

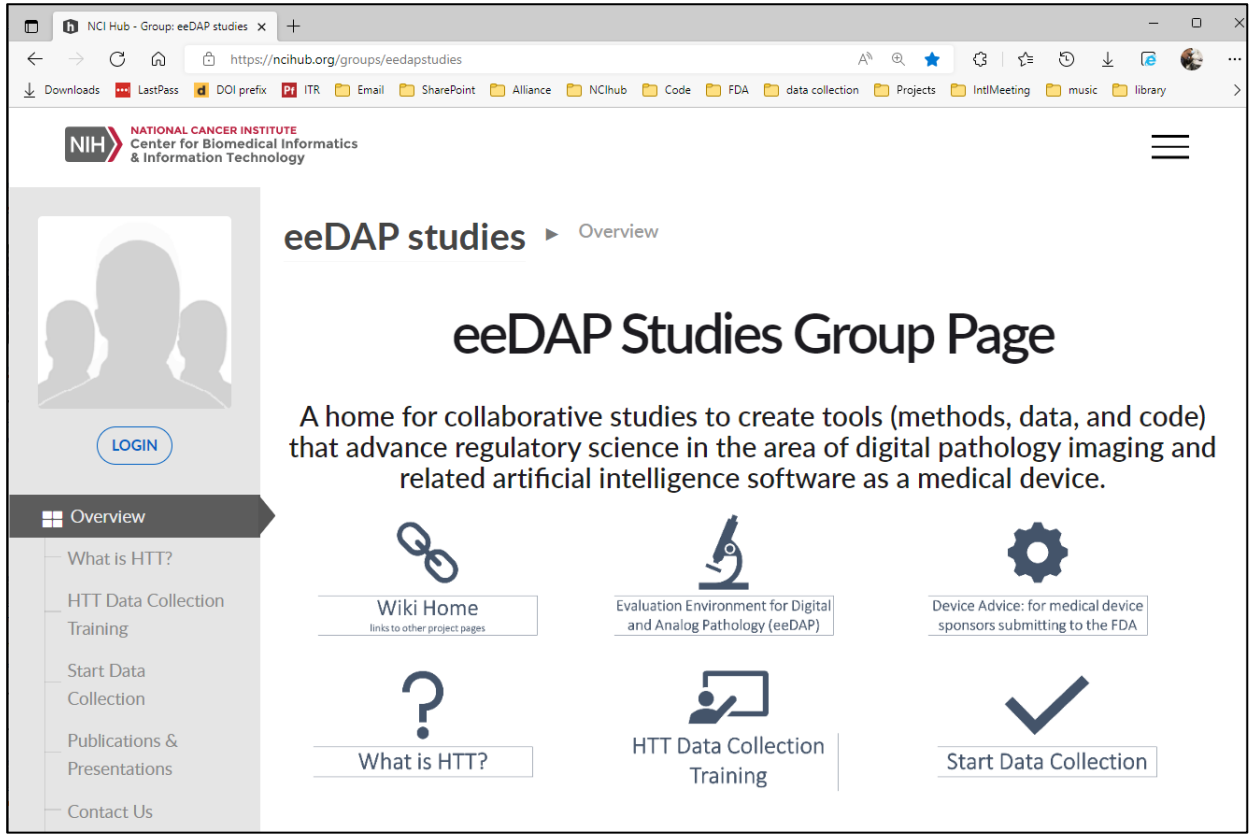
HTT Resources

Pilot Study Data and Analysis Methods
Publicly Shared



caseID	readerID	modalityID	labelROI	percentStroma	densityTILs
HTT-TILS-001-21B.ndpi_x27235.2190_y10576.2190	pathologist5857	camic	Intra-Tumoral Stroma	40	21
HTT-TILS-001-19B.ndpi_x17766.2190_y11985.2190	pathologist5857	camic	Intra-Tumoral Stroma	30	0
HTT-TILS-001-19B.ndpi_x19294.2190_y9536.2190	pathologist5857	camic	Intra-Tumoral Stroma	40	4
HTT-TILS-001-26B.ndpi_x4855.2190_y34952.2190	pathologist5857	camic	Intra-Tumoral Stroma	31	0
HTT-TILS-001-26B.ndpi_x5333.2190_y19777.2190	pathologist5857	camic	Intra-Tumoral Stroma	40	3
HTT-TILS-001-26B.ndpi_x5333.2190_y25656.2190	pathologist5857	camic	Intra-Tumoral Stroma	36	4
HTT-TILS-001-26B.ndpi_x22046.2190_y16263.2190	pathologist5857	camic	Intra-Tumoral Stroma	30	2
HTT-TILS-001-32B.ndpi_x24721.2190_y9052.2190	pathologist5857	camic	Intra-Tumoral Stroma	33	2
HTT-TILS-001-32B.ndpi_x21440.2190_y9947.2190	pathologist5857	camic	Intra-Tumoral Stroma	31	1

Project presentations and publications
Pathologist training materials
Access to data-collection Platforms



Summary and Thoughts

- A lot has been done. A lot still to do.
- Lessons learned from pilot study (and deep dive)
  - Pathologist variability can be significant
  - Pathologist variability can be reduced
  - Pathologist variability is not well behaved
  - Need tools to account for pathologist variability
- Tools (deliverables) from pilot study (and deep dive)
  - Pathologist training materials
  - Data to explore and model
  - Data-collection tools
- Lessons and tools broadly support
  - Biomarker validation
  - AI/ML model validation
  - Community is hungry for this research

Amplifying Tools (Deliverables)

Regulatory Science Tool Catalog

- iMRMC software package
  - Software to do multi-reader multi-case analysis of reader studies

Medical Device Development Tools

- HTT dataset may reduce burden to sponsors
  - Skip the design of the clinical trial
  - Know performance evaluation methods FDA will accept
  - Replace 40-70 pages of a submission with,

"We used the MDDT dataset and our algorithm performance was …"

- Reduce burden to FDA
  - Qualify data and analysis methods once to support multiple sponsors

State of the Project

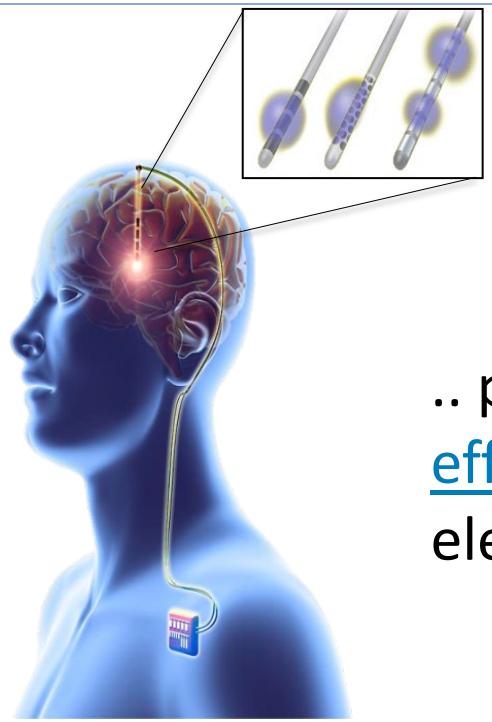
- We are on the home stretch to launch the pivotal study
- Pivotal study slide and metadata sourcing
  - **Huge effort**
  - RCAs with 2 sites, one more in process (1 year to execute, 1 year to receive data)
  - Chart reviews to find TNBC and metadata
  - Received first set of slides and metadata (n=86)
  - Target n=200
- Statistical analysis plan under development
- Future: “Share pivotal data”
  - Create a pipeline for doing AI/ML performance assessment
  - Host data on https://grand-challenge.org/ platform or https://precision.fda.gov/

Collaborative Community: Engage with FDA

- Pathology Innovation Collaborative Community
  - Face-to-face Meeting
  - June 27,28
  - Arlington, VA

https://pathologyinnovationcc.org/events/picc23-unlocking-the-potential-of-digital-pathology-and-ai-through-regulatory-science

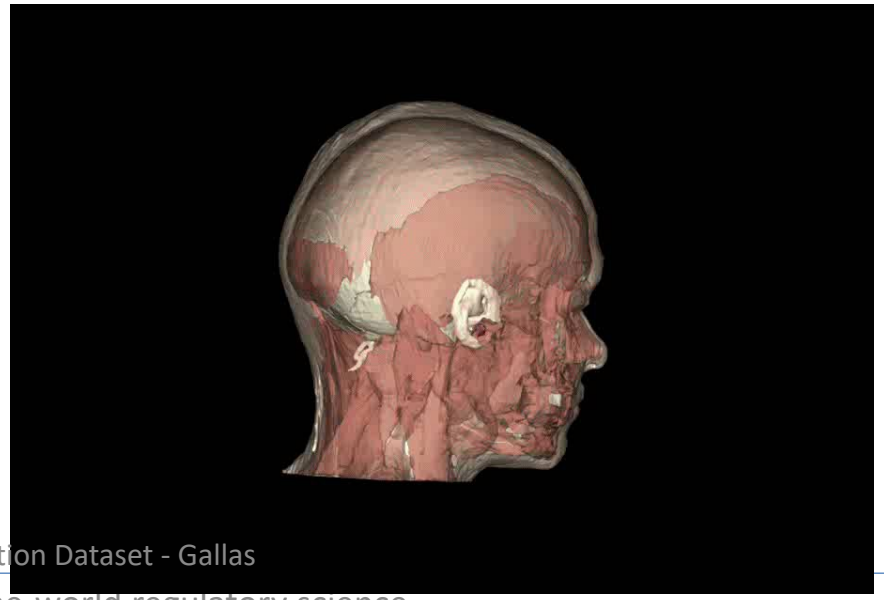
Title and content (black background)



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 in Perspective

1900
EMPLOYEES

18k
Medical Device Manufacturers

183k
Medical Devices
On the U.S. Market

22k/year
Premarket Submissions
includes 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 in Perspective

183
FEDERAL EMPLOYEES
Up to 180 visiting scientists

140 Projects
In 27 Laboratories and Program Areas

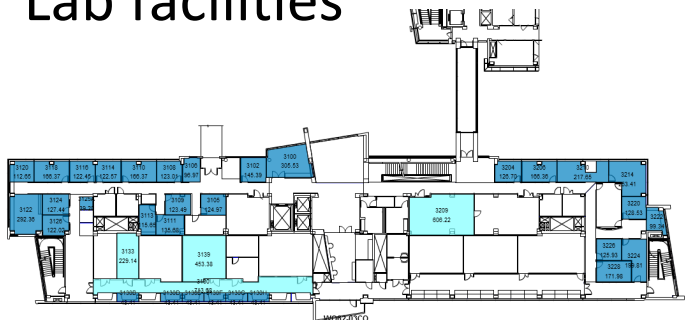
400/year
Peer reviewed presentations, articles, and other public disclosures

2,500k/year
Premarket Regulatory consults

75
Standards and conformity assessment committees

70%
Staff with post graduate degree

55,000 ft²
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 in Perspective

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 ft²
DIDSR Lab and facilities