

Computational Pathology: Deep Learning, Classification and Predicting the Future

Joel Saltz MD, PhD

Chair and Professor Department of Biomedical Informatics

Director Center for Engineering Driven Medicine

Professor Department of Pathology

Cherith Endowed Chair

Stony Brook University

6th Digital Pathology and AI Congress USA

November 20th 2020

Faster Computing – 1993 to 2020

**1993: Size of Small
Ant 1mm**

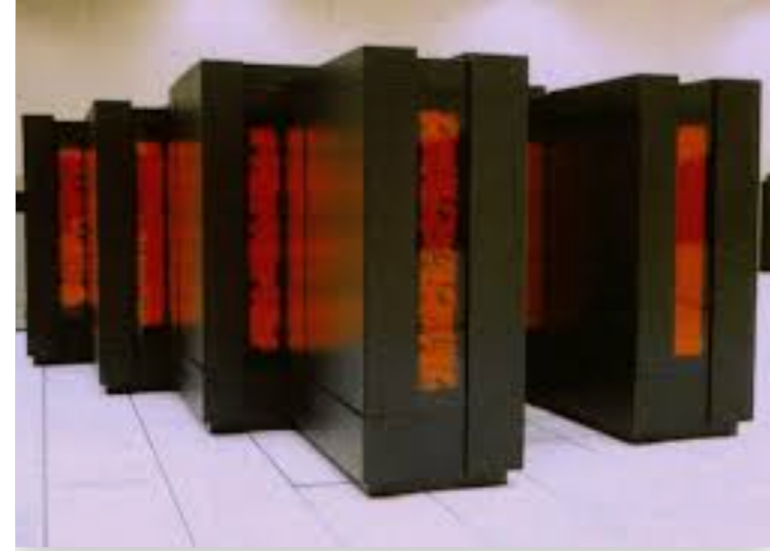


Ratio

**2020: Altitude of Plane
7km**



Thinking Machines
CM 5 1993
60 GFLOPS



Fujitsu Fugaku 2020
416 PFLOPS



Tissue Characterization for every patient

- Every diagnostic WSI contains vast amounts of data
 - Tumor, lymphocyte infiltration
 - Cells – normal, dysplastic, lymphocytes, necrotic
 - Spatial distribution and morphology of glands, ducts
- Clinical decision support
- Predict treatment response and outcome
- Patient stratification
- Treatment selection
- Cancer Epidemiology

Public TIL, nuclear segmentation and tumor models

- *public* TIL, tumor segmentation and nuclear models and datasets
- I will present the methods and results in the remainder of this talk

https://github.com/SBU-BMI/histopathology_analysis/blob/main/README.md

Coordinated Tumor/TIL maps (Breast Cancer)

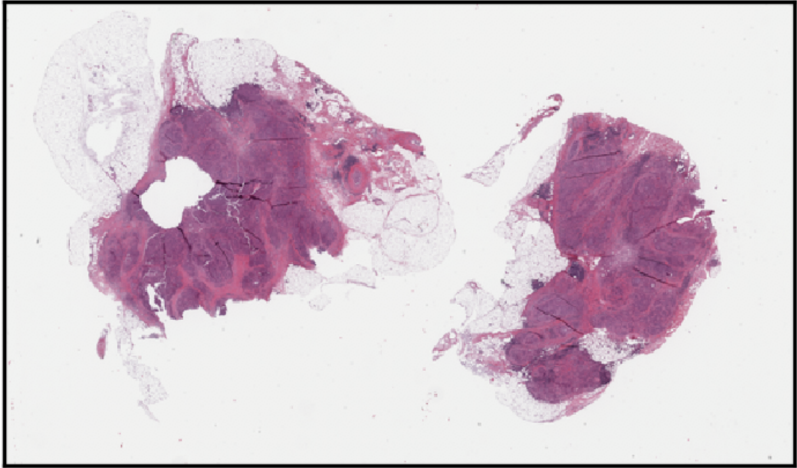
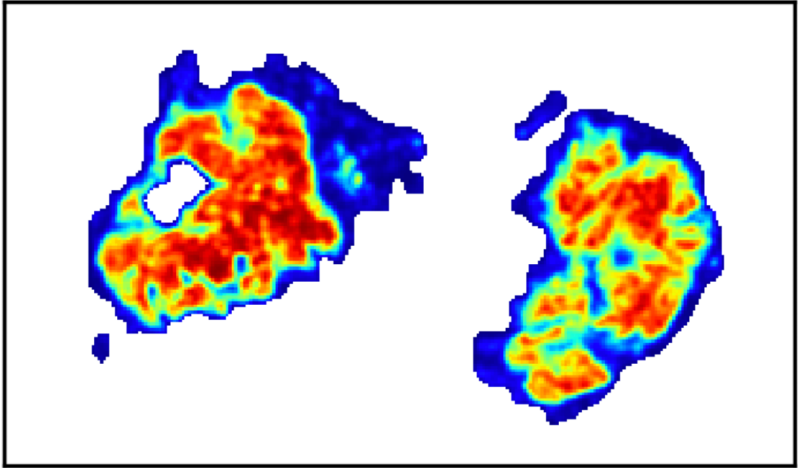
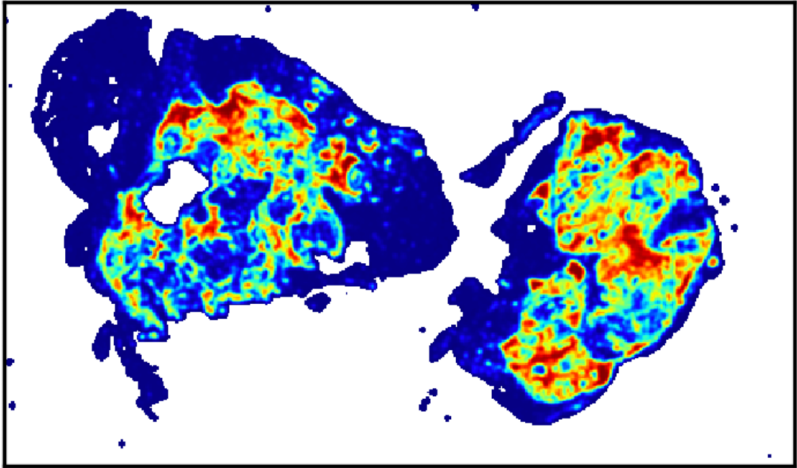
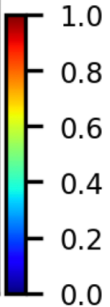


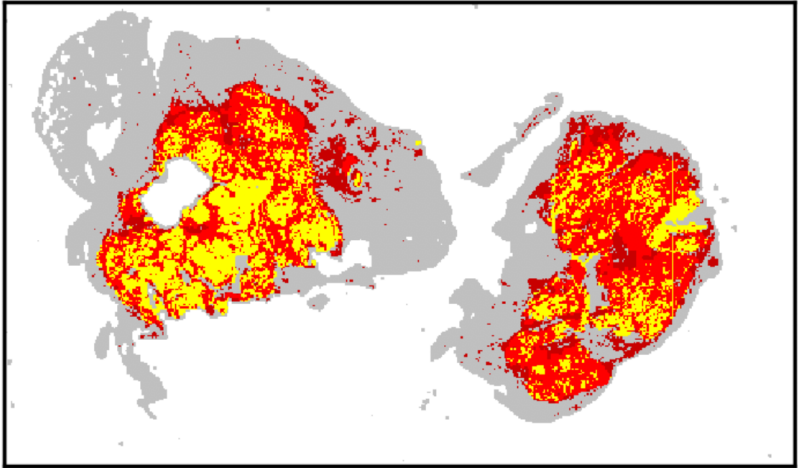
Image of Hematoxylin-Eosin tissue sample



Automated tumor detection (as probability)



Automated lymphocyte detection (as probability)



Combined Tumor-TIL map



Tumor/TIL maps – Characterize variations in TIL patterns

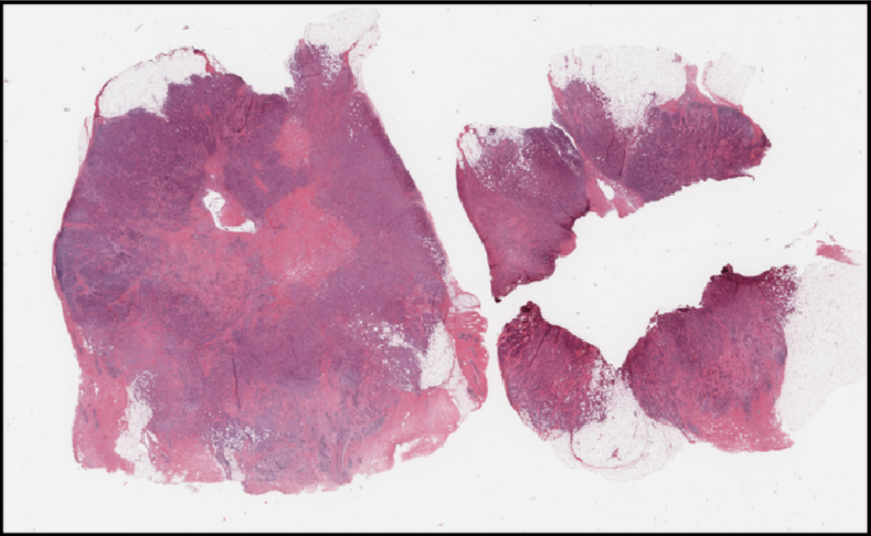
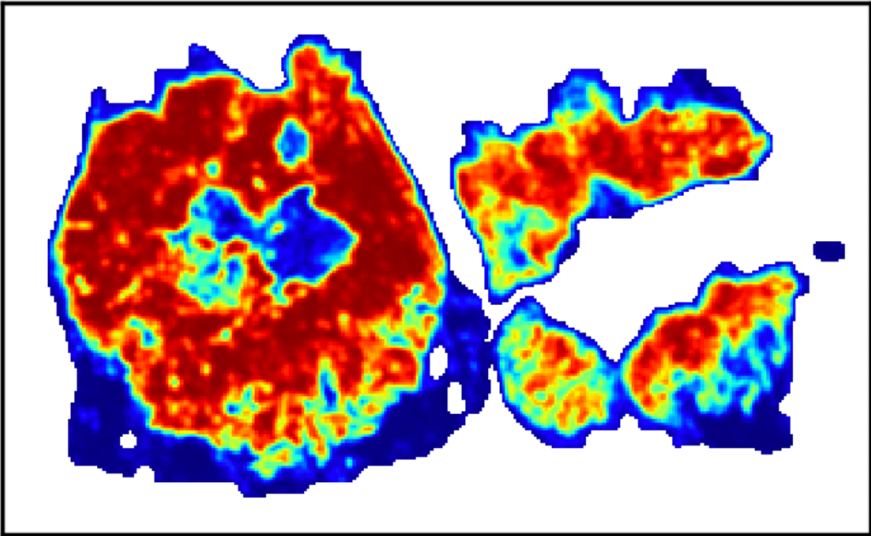
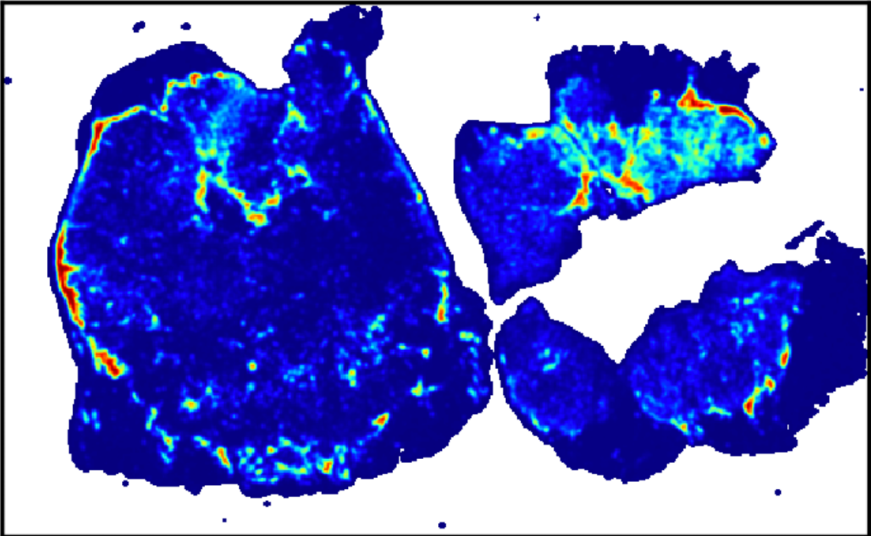


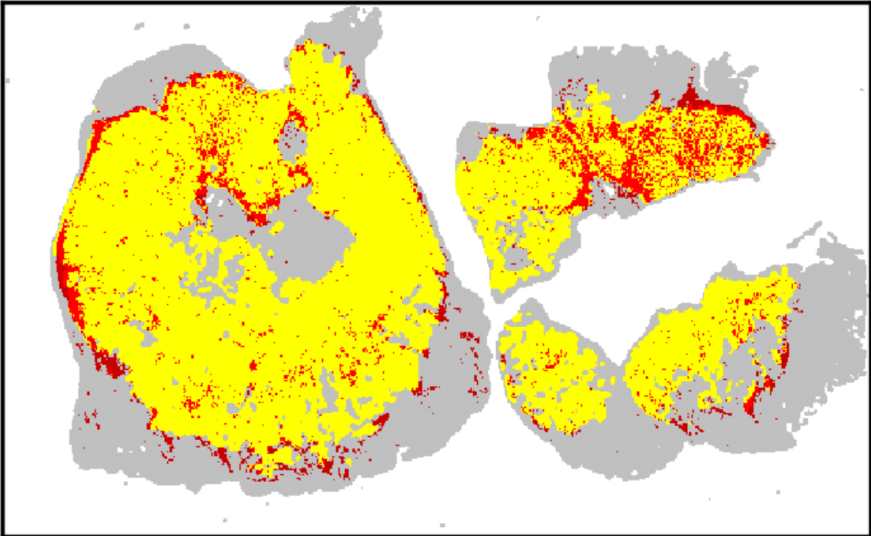
Image of Hematoxylin-Eosin tissue sample



Automated tumor detection (as probability)



Automated lymphocyte detection (as probability)



Combined Tumor-TIL map

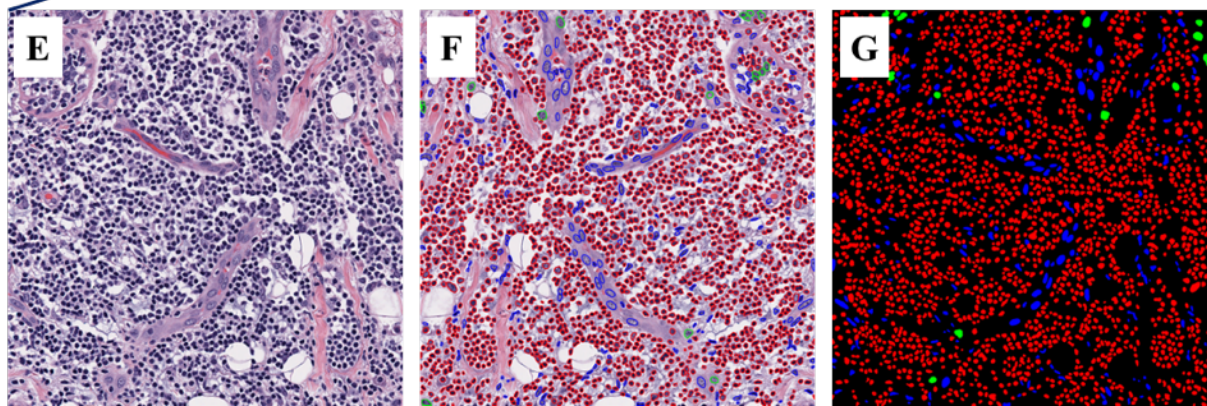
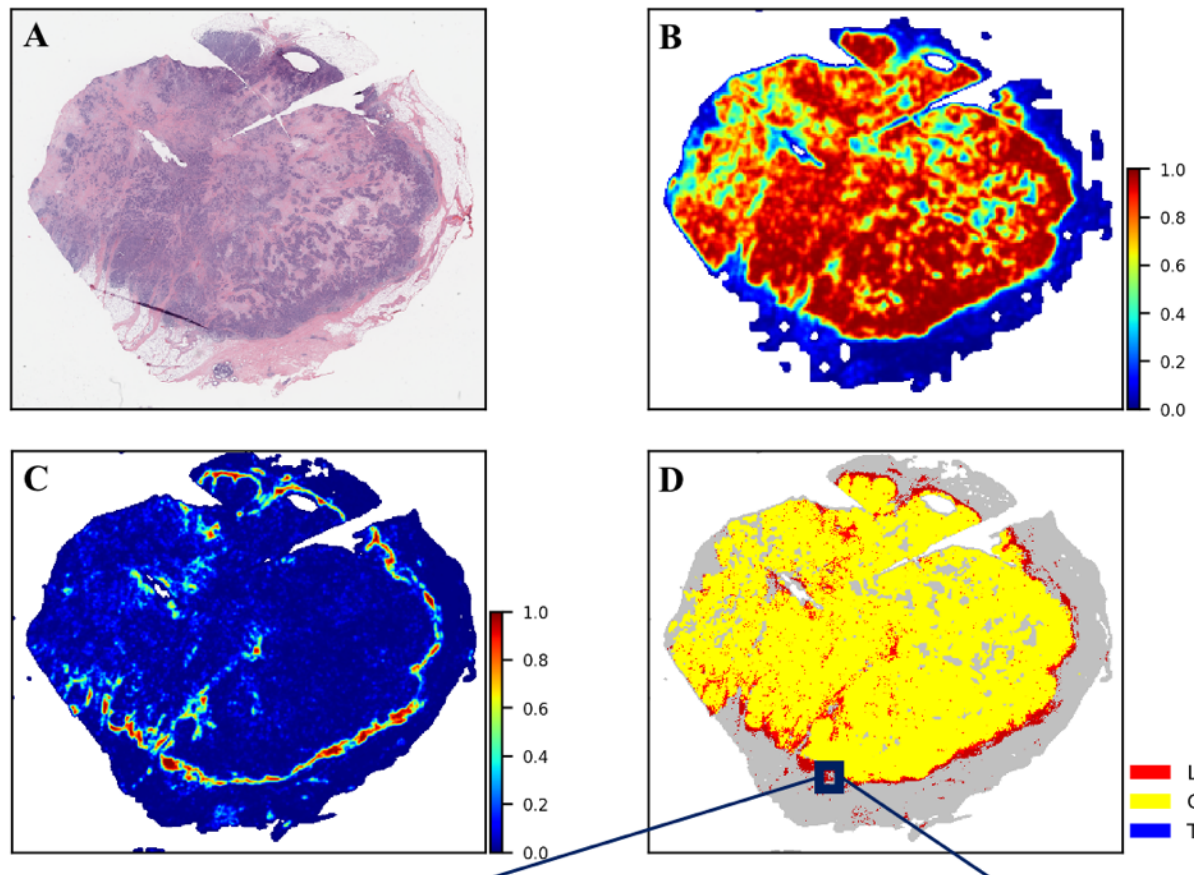
Pathomics tissue analytics for Precision Medicine

(1) Tumor-TILs analyses

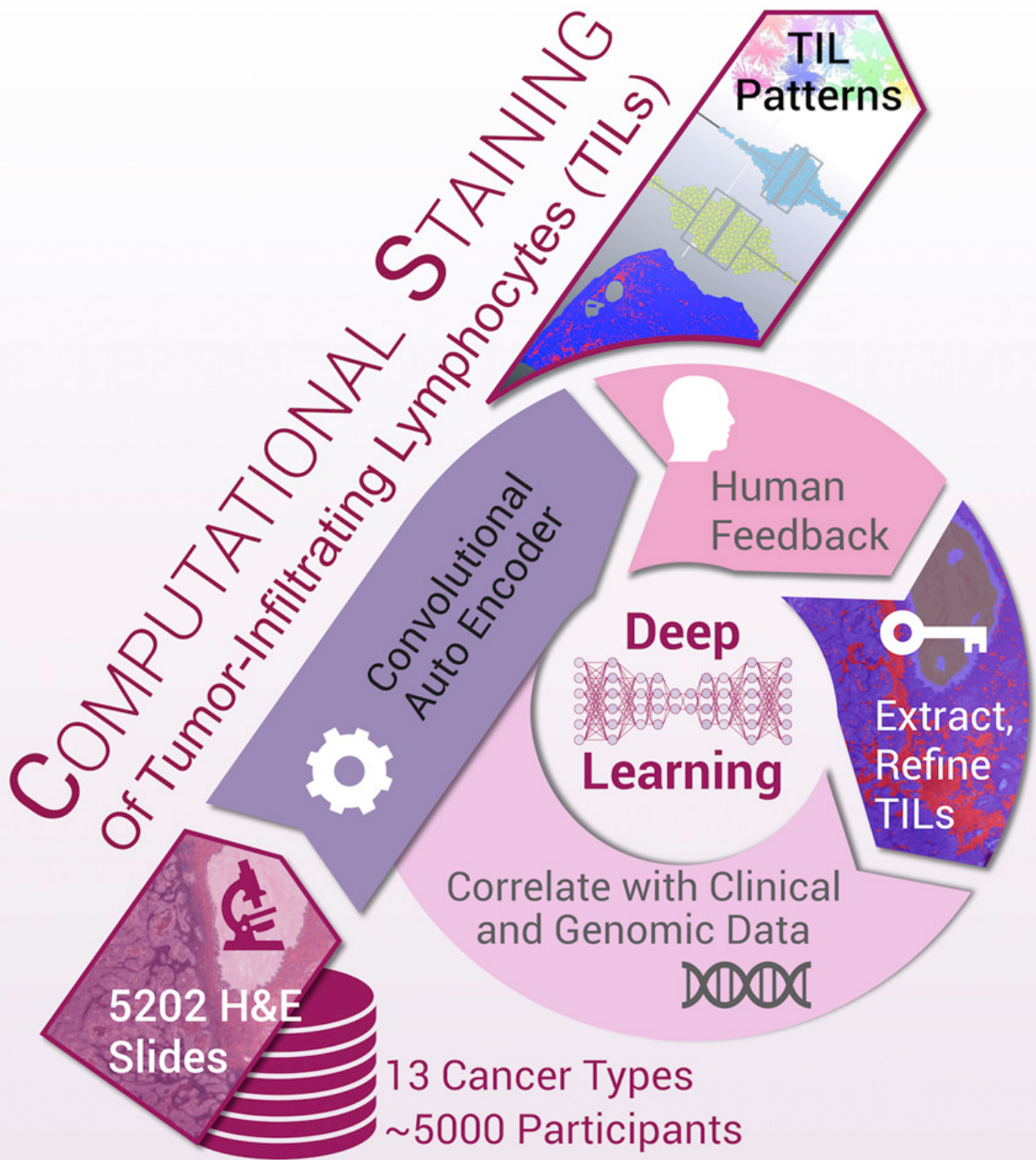
(2) High-resolution detection and classification of tumor cells, lymphocytes, and stromal cells in the entirety of whole slide images

(3) Support the scoring of the number of TILs in microscopic regions of interest chosen by pathologists

(4) Analysis of cell composition and features in different tumor niches

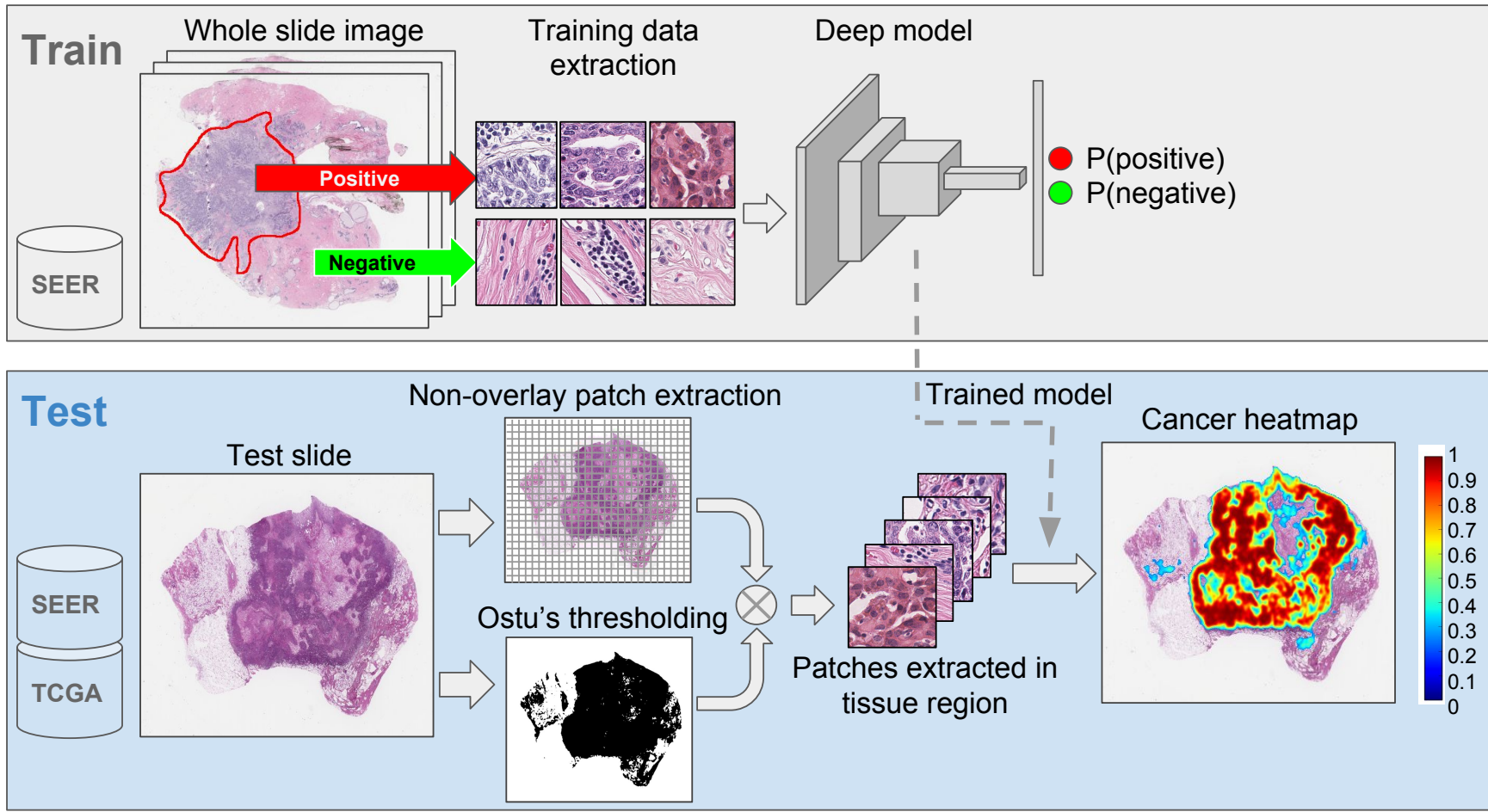


Legend
red = lymphocytes
blue = stromal cells
green = tumor cells



- Deep learning based computational stain for staining tumor infiltrating lymphocytes (TILs)
- TIL patterns generated from 4,759 TCGA subjects (5,202 H&E slides), 13 cancer types
- Computationally stained TILs correlate with pathologist eye and molecular estimates
- TIL patterns linked to tumor and immune molecular features, cancer type, and outcome
- This work has been refined and extended to roughly 20 cancer types

Deep Learning Pipeline

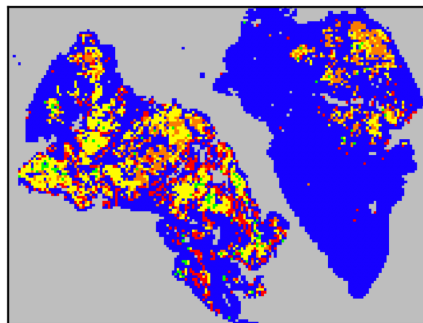
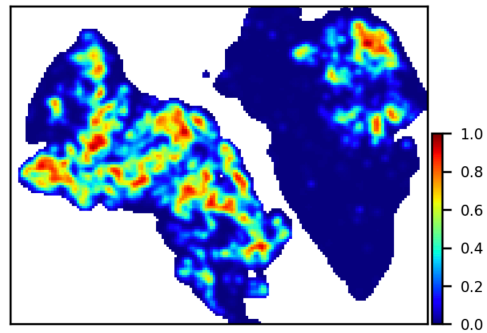
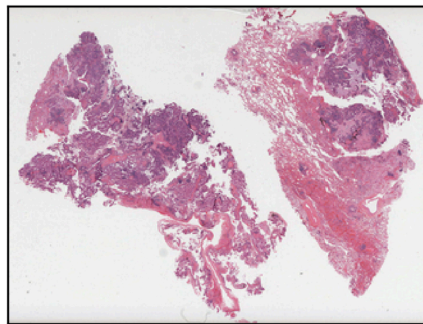


Deep learning analysis workflows that produce *maps* of TIL and tumor distributions in whole slide tissue images. The TIL analysis models can resolve TIL regions at the level of 50x50 square micron patches. The tumor analysis models classify tumor and non-tumor regions at the level of 88x88 square micron patches.

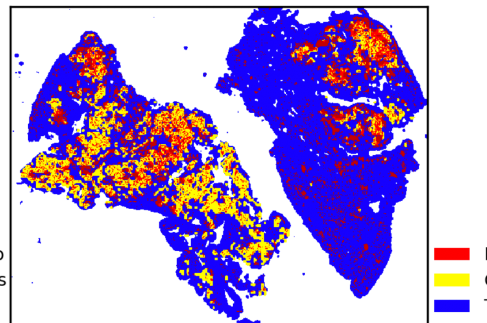
Tumors and Subtypes

- Effective tumor segmentation results currently requires substantial training on each tumor type
- We are creating *public* tumor and TIL segmentation models
- Status:
 - Prostate, breast, LUNG: models for tumor segmentation
 - Pancreas: successful model for medium resolution segmentation (MICCAI `19), high resolution methods in progress
 - Prostate, Gleason – computational gleason and NSCLC adenocarcinoma subtypes

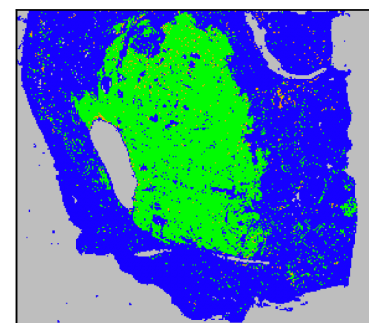
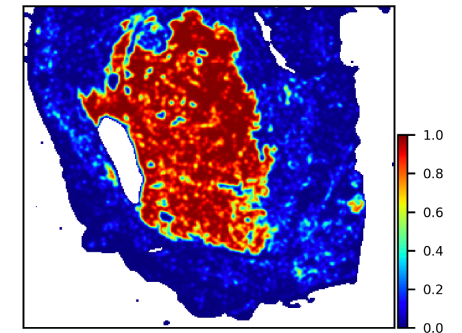
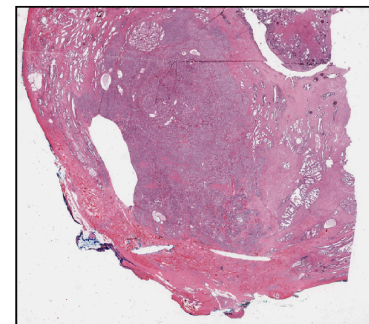
NSCLC Adeno and Prostate – integrated subtype and Tumor/TILS Analyses



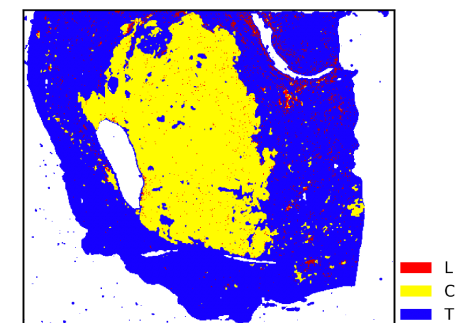
- Lepidic
- Benign
- Acinar
- Micropap
- Mucinous
- Solid



- L
- C
- T



- G3
- G4+5
- Benign



- L
- C
- T

Validation Methodology (Arvind Rao Cell Reports, Saltz et al 2018)

- Human Annotated Super-patches – multiple Pathologists
- Count number of machine learning positive patches per super-patch
- Statistical analysis to quantitate agreement between Pathologists and between Pathologists and algorithm

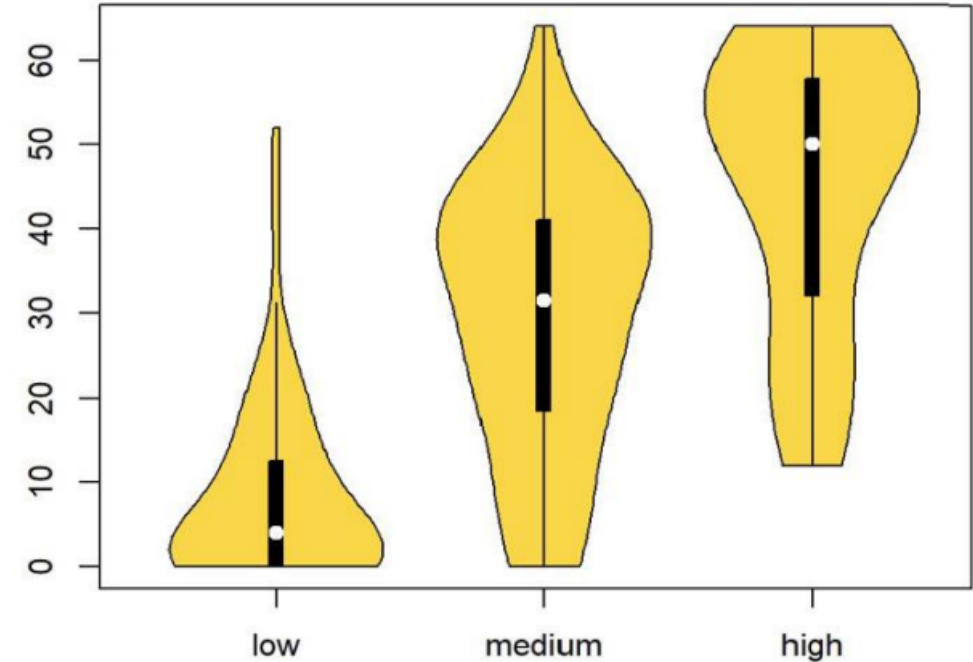




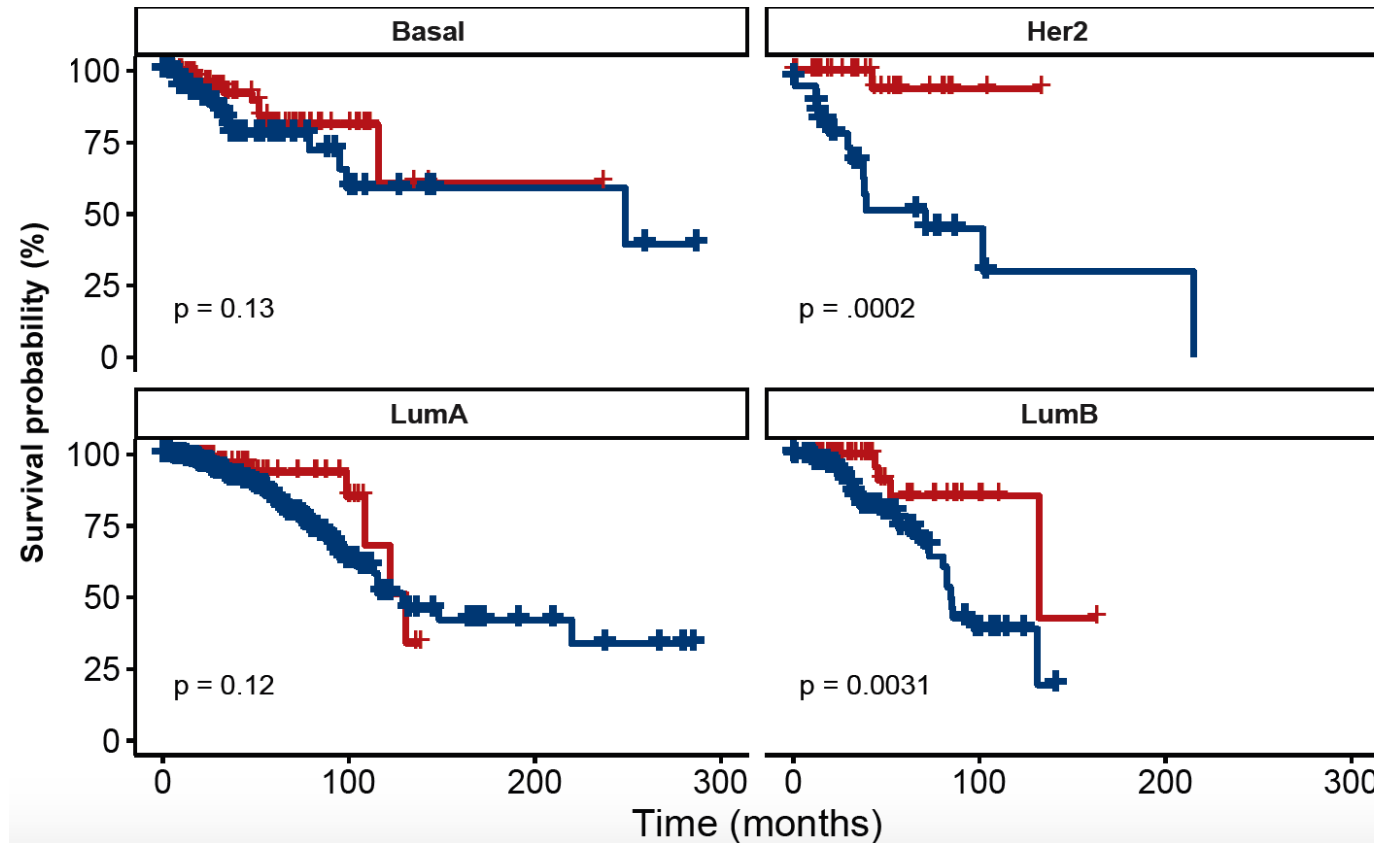
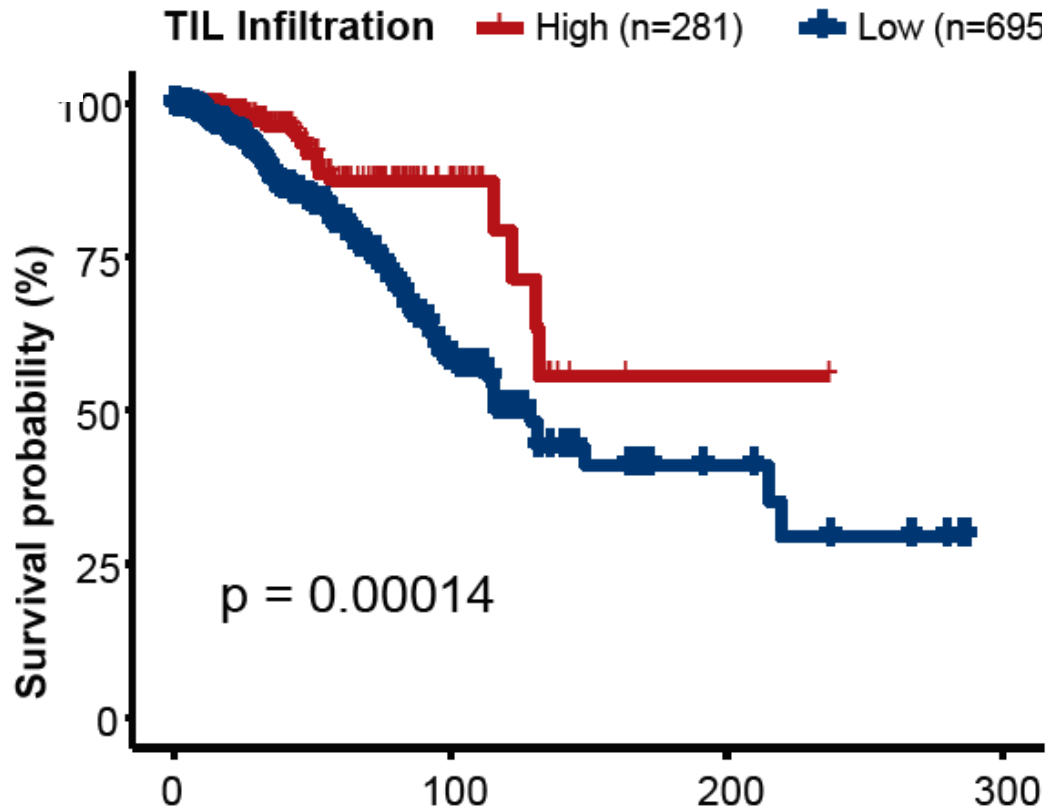
Fig. 4. Comparison of TIL scores of super-patches between pathologists and computational stain. x axis: median scores from three pathologists assessing 500 super-patches as having low, medium, or high lymphocyte infiltrate. y axis: scores from deep learning predictions on a scale from 0 to 64.

The American Journal of Pathology

Volume 190, Issue 7, July 2020, Pages 1491-1504

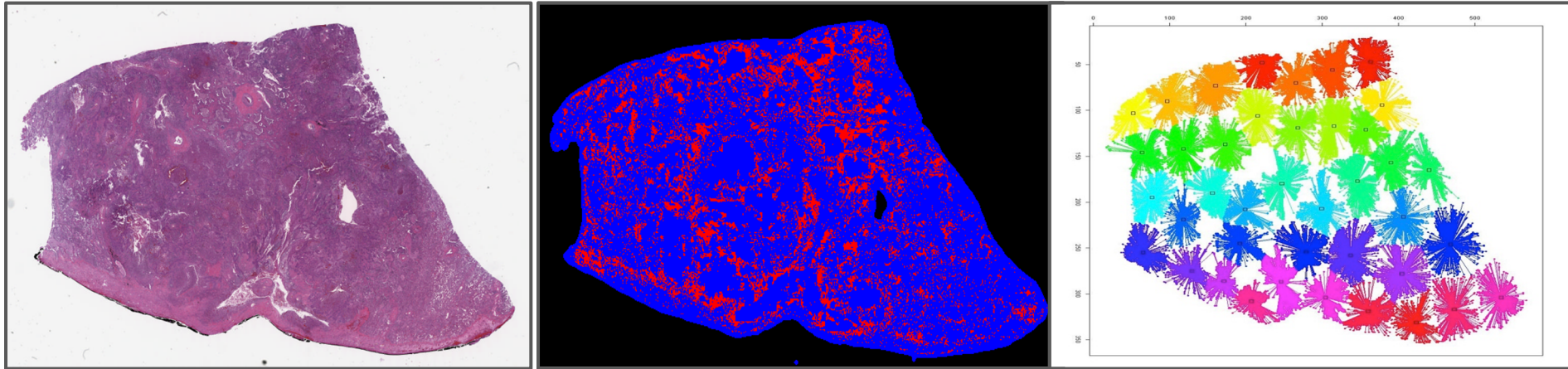
Utilizing Automated Breast Cancer Detection to Identify Spatial Distributions of Tumor-Infiltrating Lymphocytes in Invasive Breast Cancer

Han Le ^{*}  , Rajarsi Gupta ^{†, ‡}, Le Hou ^{*}, Shahira Abousamra ^{*}, Danielle Fassler [‡], Luke Torre-Healy [†], Richard A. Moffitt ^{†, ‡}, Tahsin Kurc [†], Dimitris Samaras ^{*}, Rebecca Batiste [‡], Tianhao Zhao [‡], Arvind Rao [§], Alison L. Van Dyke [¶], Ashish Sharma ^{||}, Erich Bremer [†], Jonas S. Almeida ^{**}, Joel Saltz [†]

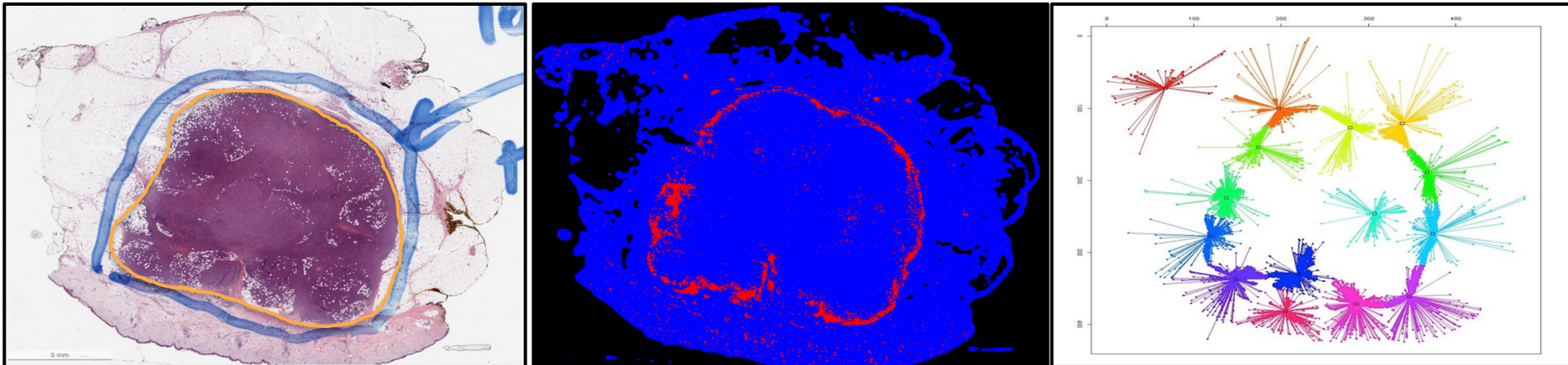


TIL Map Structural Patterns

Squamous Cell Carcinoma of the Lung, TIL map with “Brisk, Diffuse” structural pattern



Cutaneous Malignant Melanoma, TIL map with “Brisk, Band-like” structural pattern



TIL Pattern Descriptions

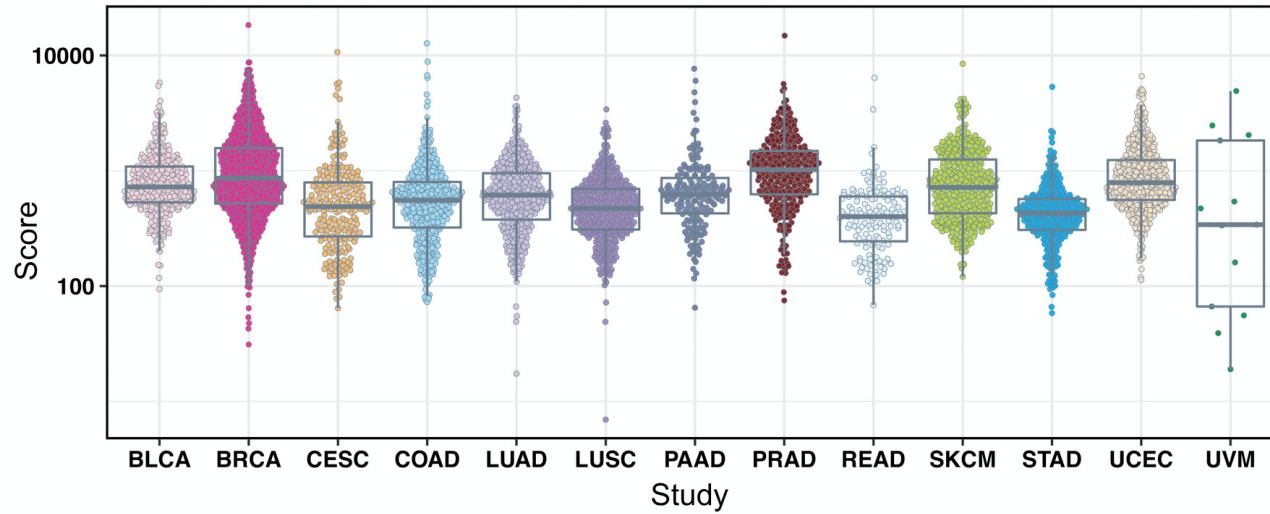
- **Qualitative (Alex Lazar, Raj Gupta)**

- “Brisk, diffuse” diffusely infiltrative TILs scattered throughout at least 30% of the area of the tumor (1,856 cases);
- “Brisk, band-like” - band-like boundaries bordering the tumor at its periphery (1,185);
- “Nonbrisk, multi-focal” loosely scattered TILs present in less
- than 30% but more than 5% of the area of the tumor (1,083);
- “Non-brisk, focal” for TILs scattered throughout less than 5% but greater than 1% of the area of the tumor (874);
- “None” < 1% TILS - in 143 cases

- **Quantitative – Arvind Rao**

- Agglomerative clustering
- Cluster indices representing cluster number, density, cluster size, distance between clusters
- Traditional spatial statistics measures
- R package clusterCrit by Bernard Desgraupes - Ball-Hall, Banfield-Raftery, C Index, and Determinant Ratio indices

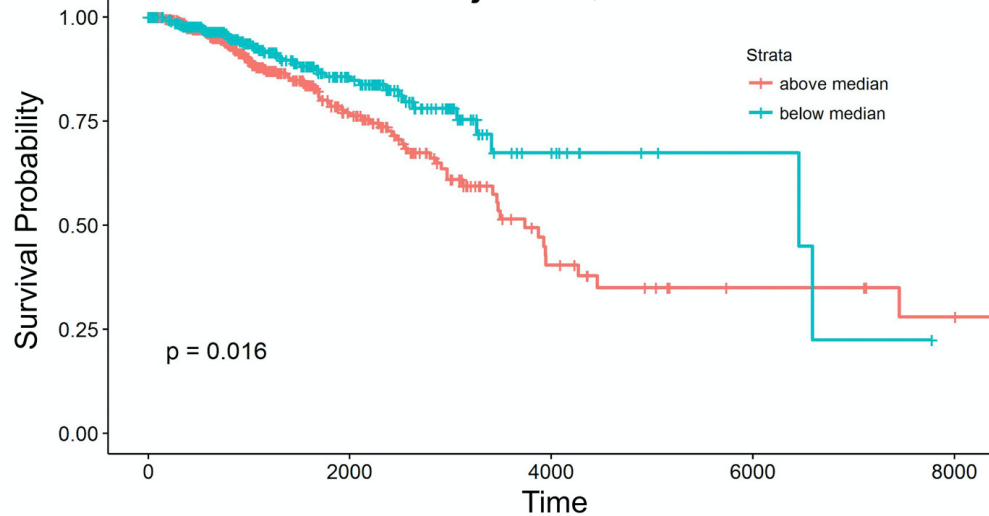
A Ball-Hall Index by Tumor Type



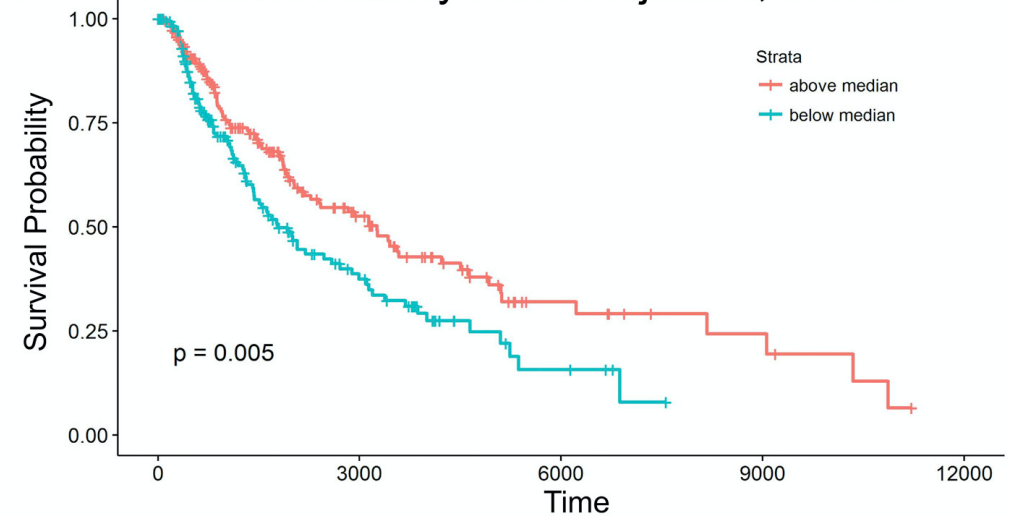
B

Index	Tumor Type	p-value
Ball Hall Adjusted	BRCA	0.007223
C index Adjusted	LUAD	0.002552
Banfield Raftery Adjusted	PRAD	0.013075
Det Ratio Adjusted	PRAD	0.012113
Banfield Raftery Adjusted	SKCM	0.001349

C Ball-Hall Index Adjusted, BRCA



D Banfield-Raftery Index Adjusted, SKCM



Training data crucial - with excellent training data, off the shelf Networks do a great job at detecting TILS!!!

Model Name	Training set (see Section 2.1)
Baseline-Youd & Baseline-HITL [13, 6]	Subsets (at most 52K) of 86K manually annotated
VGG-manual & Incep-manual	86K manually annotated

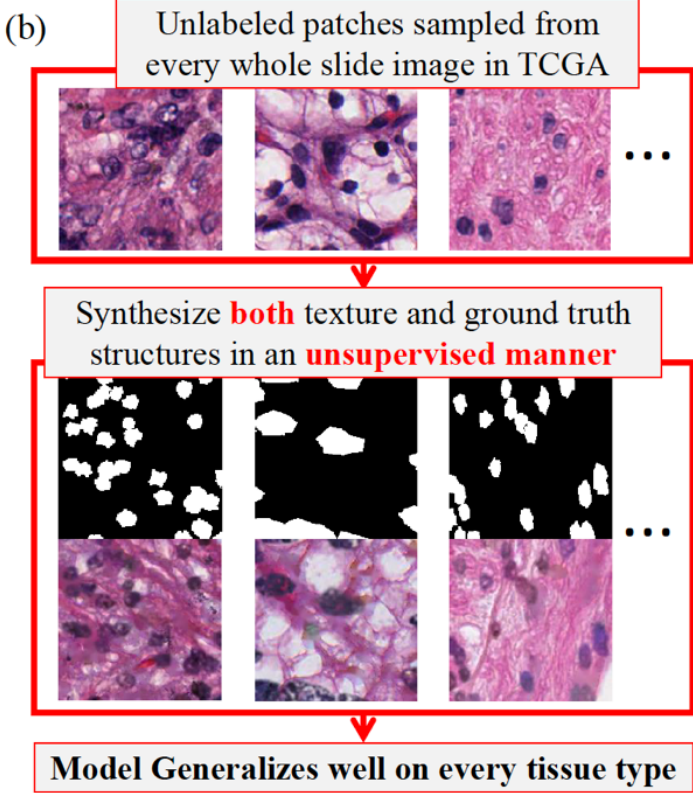
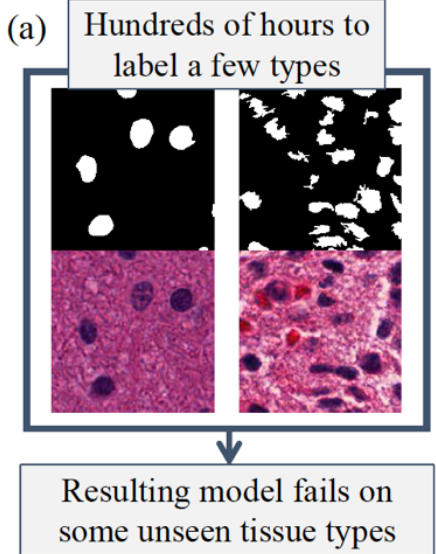
Learning from Thresholds: Fully Automated Classification of Tumor Infiltrating Lymphocytes for Multiple Cancer Types

Shahira Abousamra , Le Hou, Rajarsi Gupta, Chao Chen, Dimitris Samaras, Tahsin Kurc , Rebecca Batiste, Tianhao Zhao, Shroyer Kenneth, Joel Saltz

Stony Brook University, NY, USA

Model Name	Overall		
	F1-score	Accuracy	AUC
Baseline-HITL	0.85	79.56%	0.798
Baseline-Youd	0.85	78.22%	0.798
VGG-manual	0.87	82.44%	0.899
Incep-manual	0.87	82.44%	0.890

CVPR '19 GAN based nuclear segmentation – tissue synthesis and training augmentation



Data Descriptor | [Open Access](#) | Published: 19 June 2020

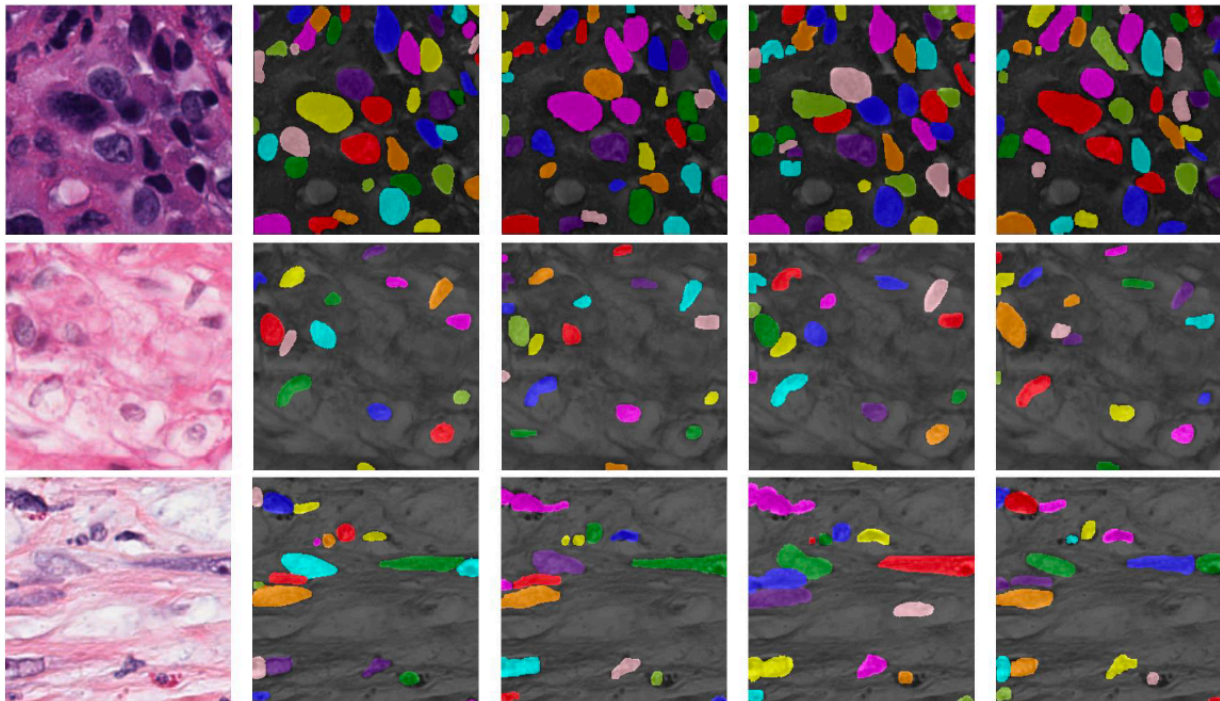
Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types

Le Hou, Rajarsi Gupta, John S. Van Arnam, Yuwei Zhang, Kaustubh Sivalenka, Dimitris Samaras, Tahsin M. Kurc & Joel H. Saltz

Nuclear Segmentation at Scale – 5 billion segmented TCGA Nuclei from 10 cancer types – deposited in TCIA

- Validated using Human Annotation

- Mask R-CNN run on 1300 patches, iteratively corrected by Two Pathologists



H&E patch

Manual segmentation by correcting Mask

Manual segmentation labeled from scratch

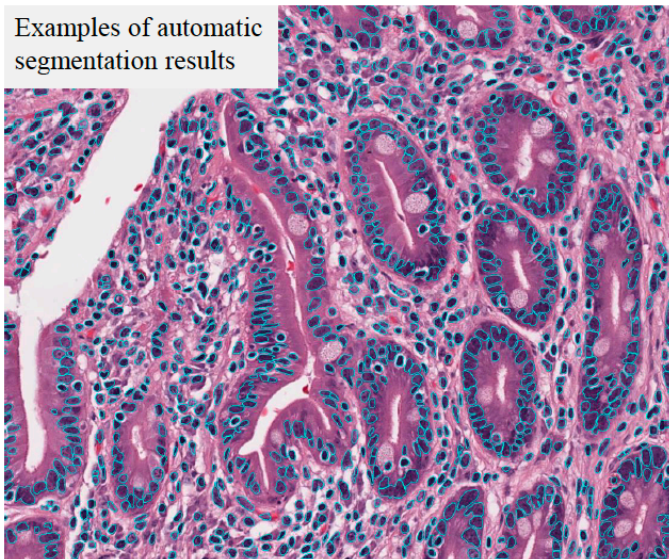
Manual segmentation labeled from scratch

Manual segmentation labeled from scratch

WSI groups	Percentage of patches with bad segmentations	#. slides
Best	0%	2,346
Good	0.01 - 6.67%	1,246
Adequate	6.68 - 13.3%	593
Problematic	13.4 - 20.0%	302
Unacceptable	> 20.0% or failed WSI QC	573

Nuclear Segmentation and Classification

Examples of automatic segmentation results

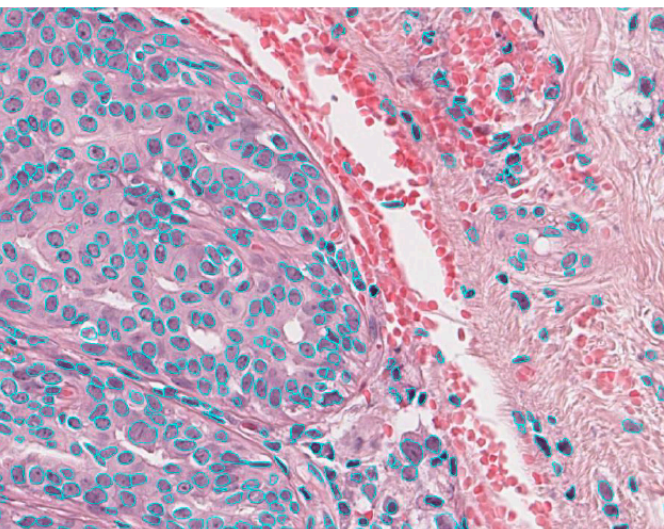


Annotator	Dice	Instance- Nuclei count		
		Dice	Correlat.	MAE%
Annotator A	0.803	0.664	0.962	12.4%
Annotator B	0.793	0.631	0.984	11.2%
Annotator C	0.780	0.683	0.973	9.5%

Comparing labeling from scratch vs. correcting Mask R-CNN's results.

Inter-annotator	Dice	Instance- Nuclei count		
		Dice	Correlat.	MAE%
Annotator A vs. B	0.760	0.600	0.959	10.8%
Annotator B vs. C	0.752	0.622	0.959	15.5%
Annotator C vs. A	0.774	0.697	0.954	12.2%

Table 5: Agreements between annotations from different human annotators. This is the performance upper bond of any automatic segmentation method.



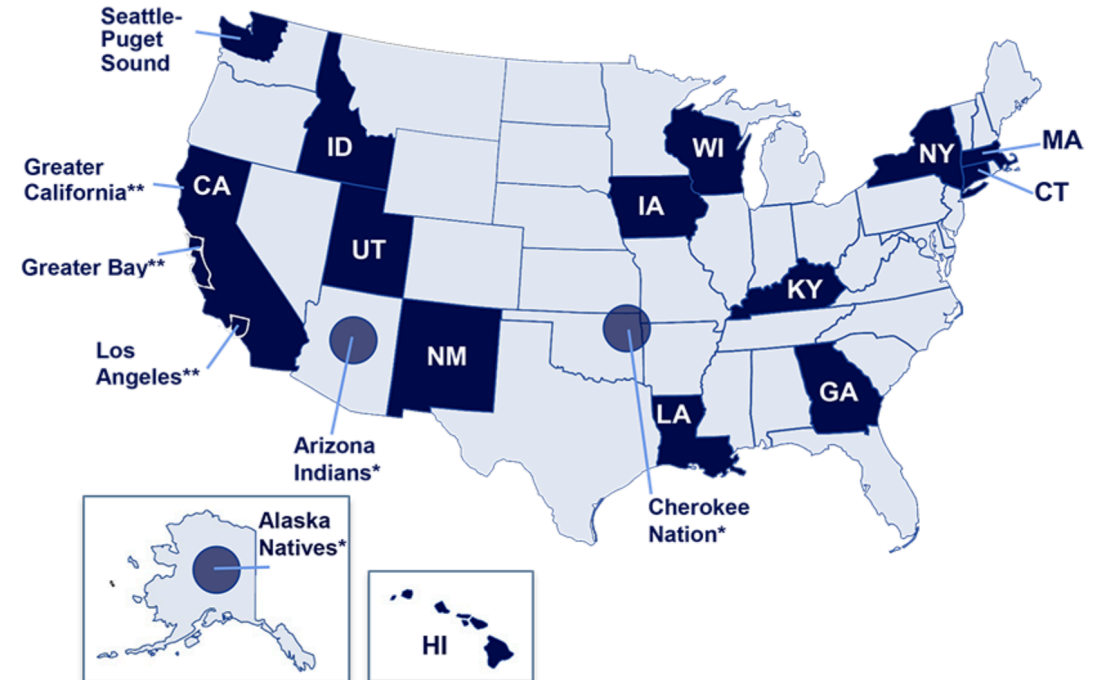
Methods and tools for integrating pathomics data into cancer registries

Saltz, Sharma, Foran and Durban

- Tissue stained with Hematoxylin and Eosin (H&E), slides created for virtually all cancer patients
- Pathologist reports describe findings
- Pathomics quantitatively characterize cells and tissue
- Pathomic features include **spatial characterization of tumor and stroma; shapes, textures of nuclei; classifications of cell type and quantitative characterization of lymphocytic infiltration**
- Potential for Deep Biological Cancer insights gathered from existing specimens for all SEER cancer patients

Computable Cancer Phenotype Information from Whole Slide Images

- Enhance SEER registry data with Pathomics
- The New Jersey State Cancer Registry, Georgia and Kentucky State Cancer Registries
- Prostate Cancer, Lymphoma and NSCLC
- Creation of population cohorts containing linked Pathomics and clinical attributes
- Response to checkpoint inhibitor therapy



Value of Pathomics

- Treatment and clinical course by linked claims data
- Real World Study: Tumor immune map for each patient

Cancer Site	Total Unique patients receiving at least one administration of a checkpoint inhibitor			
	Nivolumab	Pembrolizumab	Ipilimumab	Combined
All	1178	735	237	2150
Tongue	12	13		25
Oral Cavity	26	25	1	52
Esophagus	12	17	2	31
Stomach	7	19	1	27
Colon	15	18	4	37
Rectum	3	14	3	20
Anus, Anal Canal and Anorectum	10	5	2	17
Liver	31	1	1	33
Intrahepatic				
Bile Duct/GB/Other Biliary	3	4	1	8
Pancreas	11	4	5	20
Other Digestive Organs	1	5		6
Larynx	4	13		17
Lung and Bronchus	573	354	26	953
Melanoma of the Skin	136	78	137	351
Other Non-Epithelial Skin	2	2	1	5
Breast	18	15	2	35
Cervix Uteri	2	7		9
Corpus Uteri	5	15	1	21
Ovary	10	1	1	12
Prostate	19	23	2	44
Urinary Bladder	20	36	2	58
Kidney and Renal Pelvis	190	8	30	228
Ureter	2	7		9
Thyroid	2	8		10
Hodgkins	10	3		13
Non-Hodgkins	4	4	1	9
Mesothelioma	8	8		16

Example longitudinal claims from oncology practices (Unlimited Systems):
 Understanding approved and off label use of Checkpoint Inhibitors by cancer site - (2013-March 31, 2019)

Lynne Penberthy, NCI

Virtual Tissue Repository (VTR)

Collaboration between SEER, NCI CBIIT, Emory and Stony Brook

- **Problem:**

- Current tissue-based research often performed on limited population enrolled in clinical trials

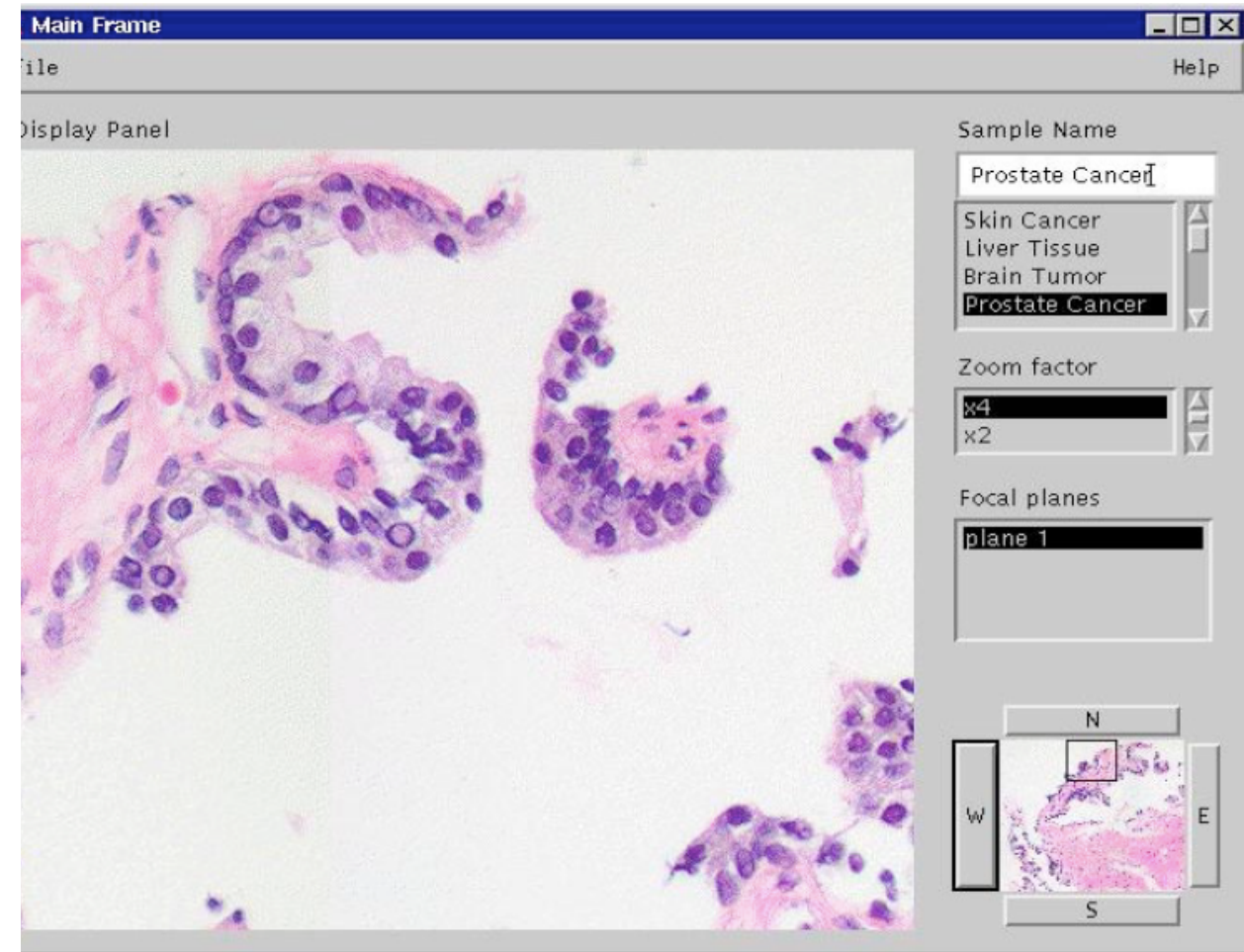
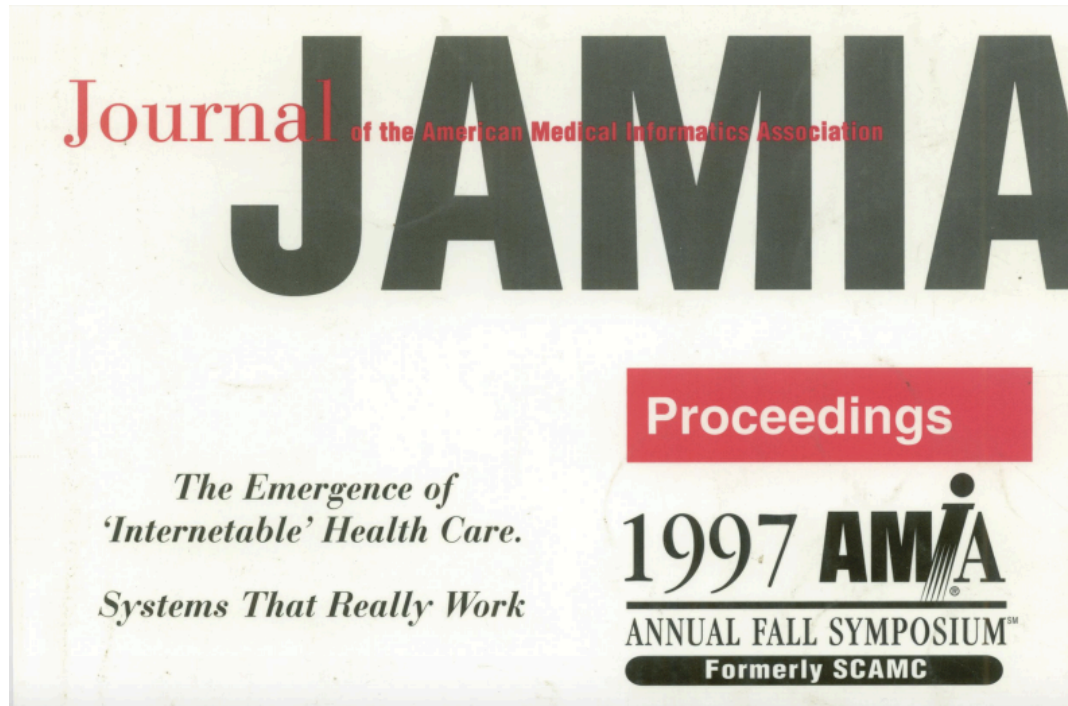
- **Objectives:**

- Provide infrastructure for tissue & data collection on a population level
- Enable use of community-based tissue specimens for biomedical research

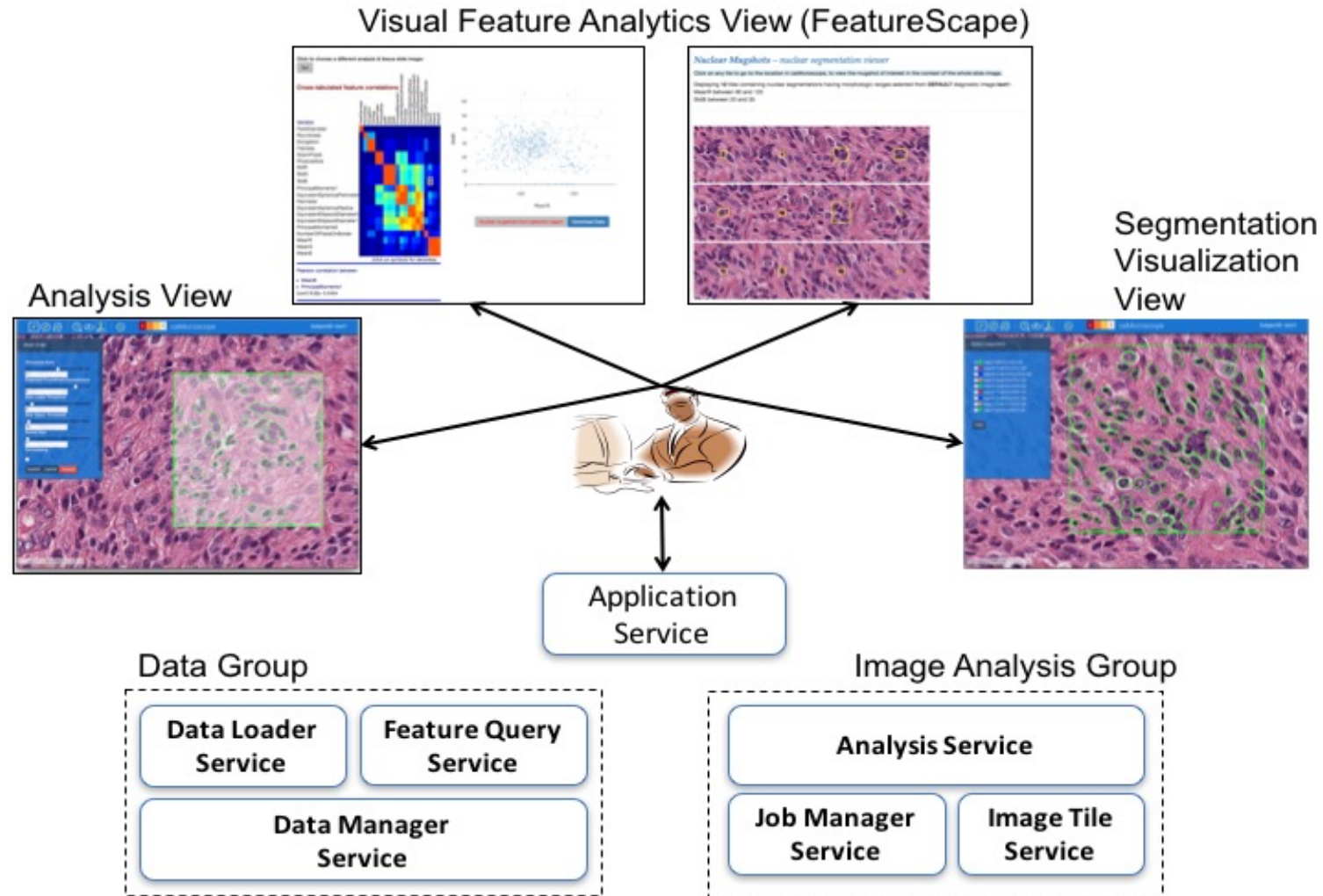
- Access to rare cancers & exceptional outcomes
- Linked to long-term outcomes
- Existing annotation with clinical & demographic data
- Potential for custom annotation
- Renewable with > 450,000 incident cases annually

The Virtual Microscope *

Renato Ferreira[†], Bongki Moon, Ph.D.[†], Jim Humphries[†], Alan Sussman, Ph.D.[†],
Joel Saltz, M.D., Ph.D.^{† ‡}, Robert Miller, M.D.[†], Angelo Demarzo, M.D.[†]



Tools: Quantitative Imaging Pathology - QuIP Tool Set



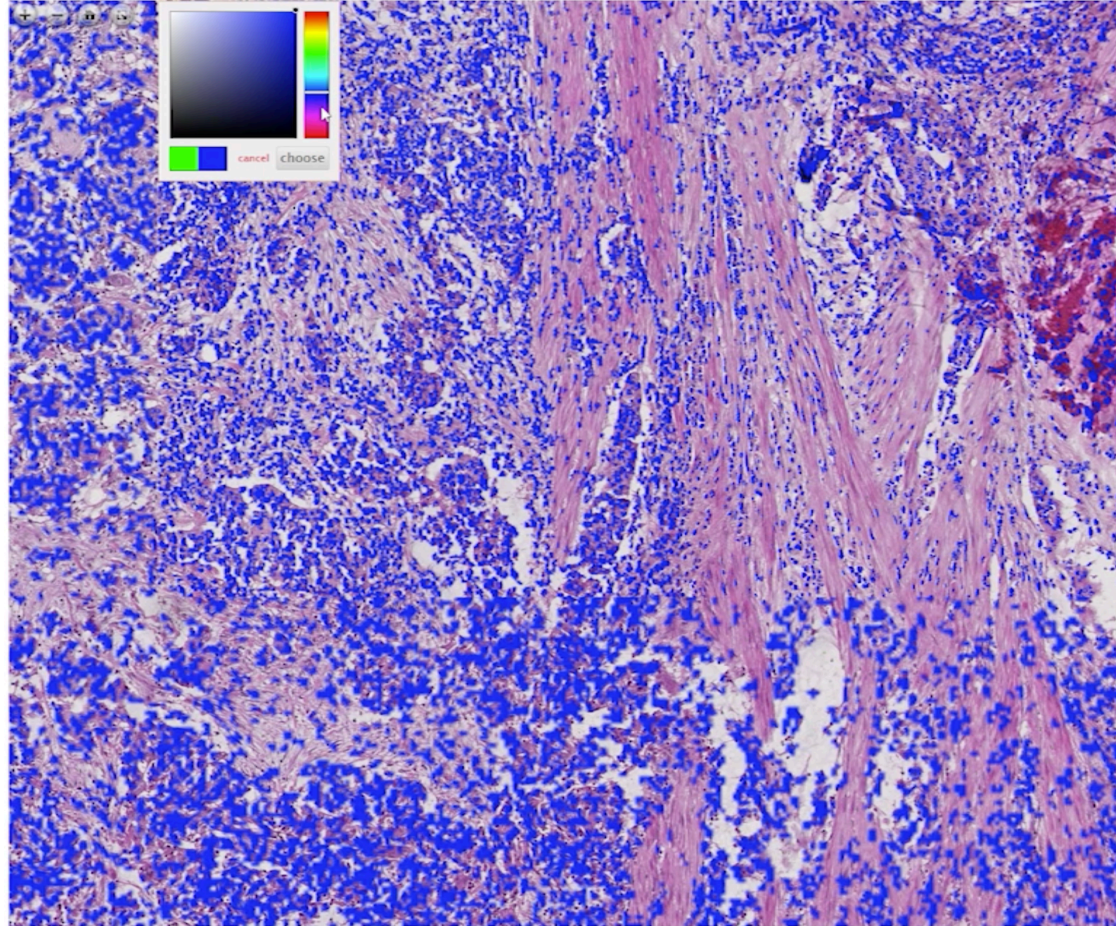
Scalable Digital Pathology - Halcyon

Halcyon Demo 1 - October 5, 2020

Opacity (whole slide image)

Opacity (segmentations)

Color (segmentations) ■



Original Article

ImageBox 2 – Efficient and Rapid Access of Image Tiles from Whole-Slide Images Using Serverless HTTP Range Requests

Erich Bremer¹, Joel Saltz¹, Jonas S Almeida²

¹Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA, ²Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Maryland, USA

Submitted: 09-Apr-2020

Revised: 20-Apr-2020

Accepted: 03-Jul-2020

Published: 10-Sep-2020

- New software stack designed for scalability
- Images and metadata can be distributed across sites
- Machine generated annotations are managed using Hilbert Curves

Annotation management with Hilbert Curves: [arXiv:2005.06469v](https://arxiv.org/abs/2005.06469v)

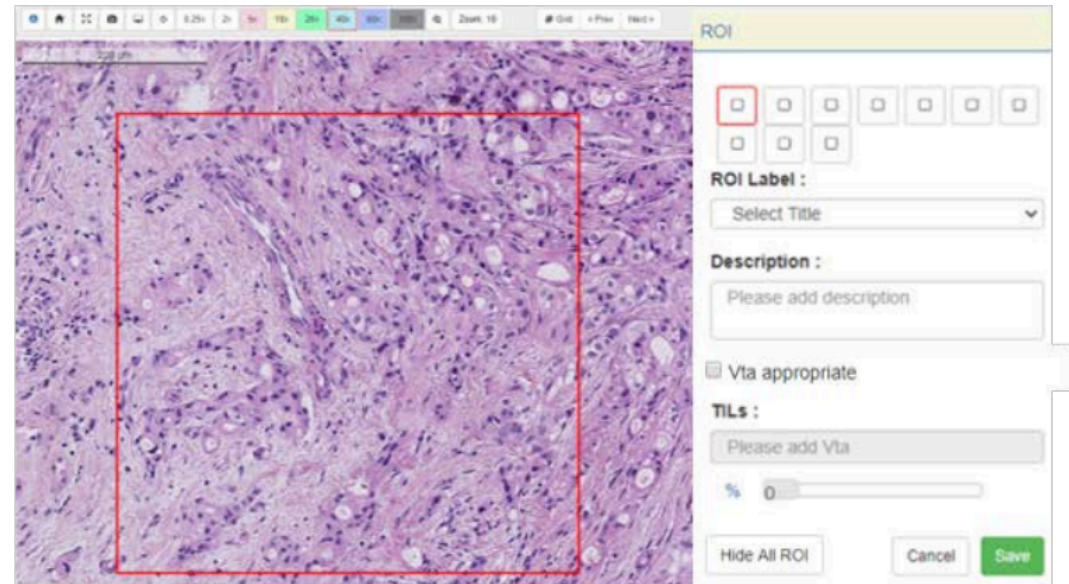
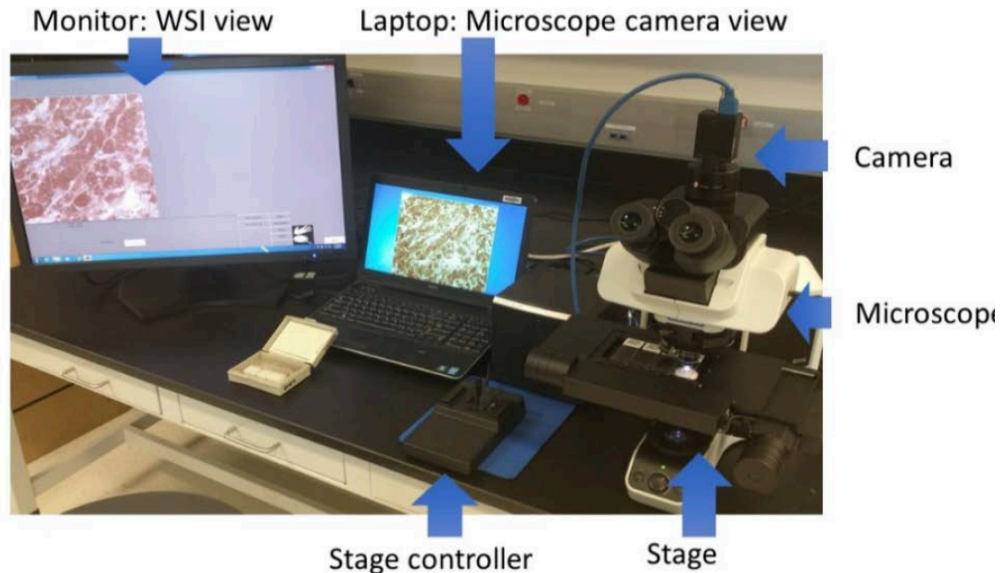
Multi-site Access and Management of WSIs: [Journal Digital Pathology](https://www.nature.com/articles/s41746-020-0100-0)

Validating Artificial Intelligence Methods for Clinical Use

Title

A Pathologist-Annotated Dataset for Validating Artificial Intelligence: A Project Description and Pilot Study

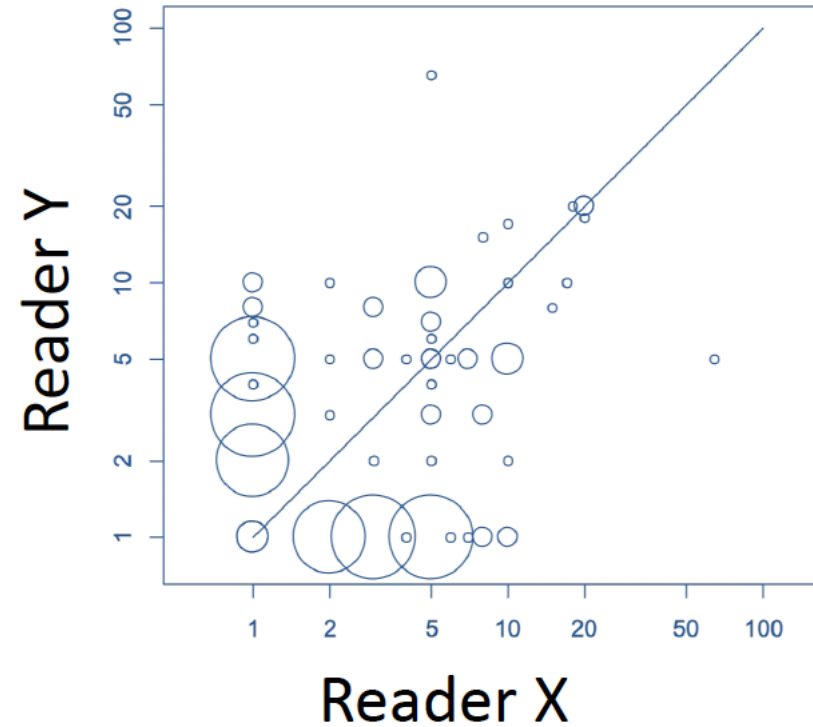
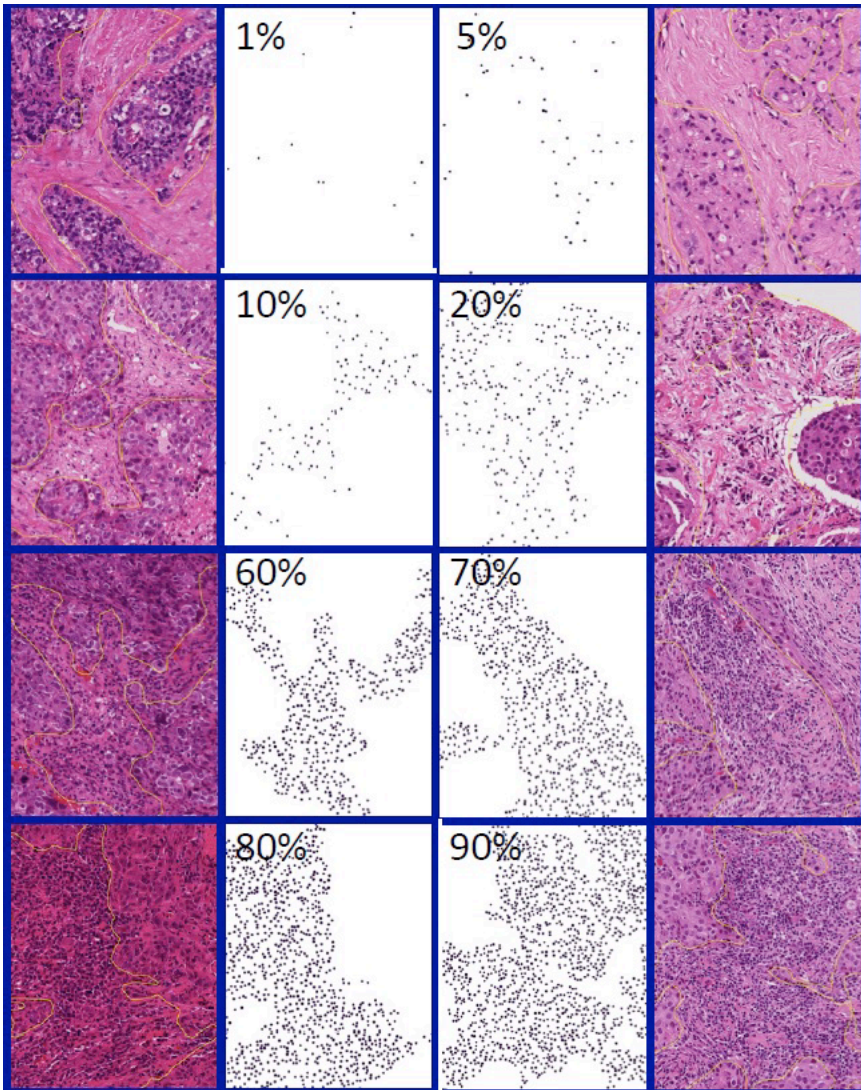
Sarah N Dudgeon (1), Si Wen (1), Matthew G Hanna (2), Rajarsi Gupta (3), Mohamed Amgad (4), Manasi Sheth (5), Hetal Marble (6), Richard Huang (6), Markus D Herrmann (7), Clifford H. Szu (8), Darick Tong (8), Bruce Werness (8), Evan Szu (8), Denis Larsimont (9), Anant Madabhushi (10), Evangelos Hytopoulos (11), Weijie Chen (1), Rajendra Singh (12), Steven N. Hart (13), Joel Saltz (3), Roberto Salgado (14), Brandon D Gallas (1)



Validation dataset of pathologist annotations for algorithms that process whole slide images

- Researchers from FDA, academic colleagues in collaboration with the Alliance for Digital Pathology, are collecting pathologist annotations as data for AI/ML algorithm validation
- Application - tumor infiltrating lymphocyte (TIL) detection and quantitation..
- Training materials and workflows to crowdsource pathologist image annotations on two modes: an optical microscope and two digital platforms.
- The microscope platform allows the same ROIs to be evaluated in both modes.
- Pursue an FDA Medical Device Development Tool Qualification

Systematic study to compare human and algorithmic TIL evaluations



HTT materials, pointers and presentations:

<https://ncihub.org/groups/eedapstudies/wiki/HTTupdate20201023>

<https://digitalpathologyalliance.org/jun-2020-alliance-at-siim-2020>

Thanks to Brandon Gallas, FDA

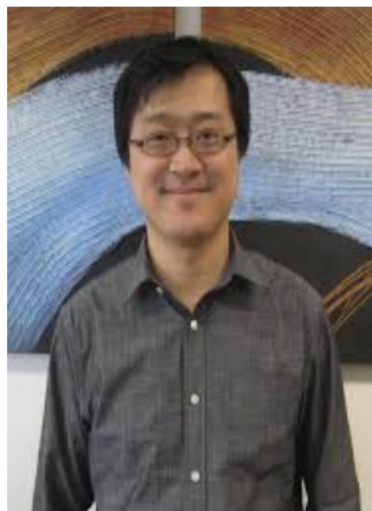
**Prediction is very
difficult, especially
about the future. –**

**Niels Bohr, 1885-1962,
Danish Physicist.**

Forty Year Span

- **From** Virtual Microscope – Hopkins/U Maryland Computer Science: 1995
- **Through** Deep Learning for quantitation – Now
- **To** combined deep learning/statistical methods to integrate clinical, “omic”, Radiology, Pathology data to steer treatment – 2025
- **To** ability to employ deep learning and integrative data to develop new treatments – 2035

Stony Brook Deep Learning Pathology Faculty



Dimitis Samaras, Chao Chen, Tahsin Kurc, Prateek Prasanna, Raj Gupta

Adjunct Faculty Pathology Deep Learning Team Members

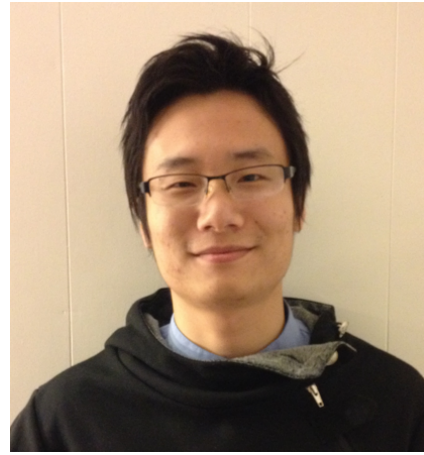
John and Beatrice are practicing Anatomic Pathologists



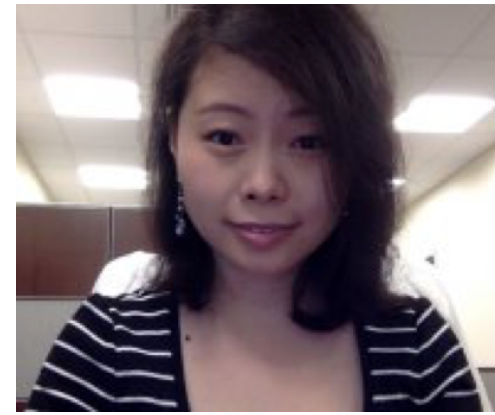
John Van Arnam, Beatrice Knudsen (University of Utah)



Han Le– Software Development Engineer II, Amazon, Seattle



Le Hou– software engineer, google core, Mountain View



Anne Zhao – Assistant Professor, Dept Pathology Stony Brook



Vu Nguyen– Graduate Student Computer Science

Deep Learning and Tissue Characterization Trainee Team circa 2018



Raj Gupta – Department of BMI Stony Brook



**The Stony Brook Multi-Scale Student Group:
David Belinsky, Mahmudul Hasan
Prantik Howlader**

SEER UG3 Team

- Stony Brook
 - Joel Saltz MD, PhD
 - Tahsin Kurc PhD
 - Dimitris Samara PhD
 - Erich Bremer
 - Le Hou
 - Shahira Abousamra
 - Raj Gupta
 - Han Le
 - Bridge Wang
- Emory
 - Ashish Sharma PhD
 - Ryan Birmingham
 - Nan Li
- Georgia SEER Registry
 - Kevin Ward MPH, PhD
- Rutgers
 - David J. Foran PhD
 - Evita Sadimin, MD
 - Wenjin Chen, PhD
 - Doreen Loh M.S.
 - Jian Ren, Ph.D
 - Christine Minerowicz, MD
- Rutgers SEER/ NJ State Cancer Registry
 - Antoinette Stroup, PhD
 - Adrian Botchway, CTR
 - Gerald Harris, PhD
- University Kentucky; Kentucky SEER Registry
 - Eric B. Durbin, DrPH, MS
 - Isaac Hands, MPH
 - John Williams, MA
 - Justin Levens

QuIP ITCR Team

Stony Brook University

Joel Saltz

Tahsin Kurc

Raj Gupta

Dimitris Samaras

Erich Bremer

Fusheng Wang

Tammy DiPrima

Le Hou

NCI / DCEG

Jonas S Almeida

Emory University

Ashish Sharma

Ryan Birmingham

Nan Li

University of Tennessee Knoxville

Jeremy Logan

Scott Klasky

Harvard University

Rick Cummings

ITCR/IMAT Supplement

Richard Levinson

Raj Gupta

Tahsin Kurc

Joel Saltz

Han Le

Maozheng Zhao

Dimitris Samaras

High Throughput Truthing Team

Brandon Gallas, FDA Initiative Leader

Project mgmt.

- **Sarah Dudgeon, MPH**
 - FDA/CDRH/OSEL/DIDSR

caMicroscope team

- **Ashish Sharma, PhD**
 - Emory University Department of Biomedical Informatics
- **Joel Saltz, MD PhD**
 - Dept. of Biomedical Informatics, Stony Brook Medicine
- **Nan Li, MS**
 - Dept. of Biomedical Informatics, Stony Brook Medicine

PathPresenter team

- **Matthew Hanna, MD**
 - Memorial Sloan Kettering, New York, NY
- **Rajendra Singh, MD**
 - Icahn School of Medicine at Mt Sinai
- **Krushnavadan Acharya, MCA**
 - PathPresenter

Slides and Clinical

- **Roberto Salgado**
 - Peter Mac Callum Cancer Center; GZA-ZBA Hospitals
 - International Working Group for TILs in Breast cancer
- **Denis Larismont**
 - Institut Jules Bordet

Statistics

- **Si Wen**
 - FDA/CDRH/OSEL/DIDSR
- **Manasi Sheth**
 - FDA/CDRH/OPEQ/OCEA/Biostatistics
- **Chava Zibman**
 - FDA/CDRH/OPEQ/OCEA/Biostatistics
- **Weijie Chen, PhD**
 - FDA/CDRH/OSEL/DIDSR

Committee

- **Mohamed Amgad, MSc**
 - Emory University School of Medicine, Atlanta, GA
- **Rajarsi Gupta, MD, PhD**
 - Renaissance School of Medicine and Dept. of Biomedical Informatics, Stony Brook Medicine
- **Steven N. Hart, PhD**
 - Mayo Clinic, Rochester, MN
- **Joe Lennerz, MD, PhD**
 - Massachusetts General Hospital, Boston, MA
- **Richard Huang, MD, MS**
 - Massachusetts General Hospital, Boston, MA
- **Anant Madabhushi, PhD**
 - Case Western Reserve University
- **Kyle J. Myers, PhD**
 - FDA/CDRH/OSEL/DIDSR
- Open door policy



EMORY
UNIVERSITY



Stony Brook
University

- **Fred Prior, PhD**
- Jonathan Bona, PhD
- Kirk Smith
- Lawrence Tarbox, PhD
- Mathias Brochhausen, PhD
- Roosevelt Dobbins
- Tracy Nolan
- William Bennett

- **Ashish Sharma, PhD**
- Annie Gu
- Mohanapriya Narapareddy
- Monjoy Saha, PhD
- Pradeeban Kathiravelu, PhD

- **Joel Saltz, MD, PhD**
- Erich Bremer
- Rajrisi Gupta MD
- Tahsin Kurc, PhD
- Tammy DiPrima



- TJ Fitzgerald, MD
- Fran Laurie



The PRISM Team

Thanks - Funding

- U24CA180924, U24CA215109, NCIP/Leidos 14X138 and HHSN261200800001E, UG3CA225021-01 from the NCI; R01LM011119-01 and R01LM009239 from the NLM, Bob Beals and Betsy Barton, Stony Brook Mount Sinai Seed Funding, Pancreatic Cancer Action Network
- This research used resources provided by the National Science Foundation XSEDE Science Gateways program and the Pittsburgh Supercomputer Center BRIDGES-AI facility