



Information Technology Powering Cancer Research for Discover and Novel Hypothesis Generation -- *Pathomics*

Joel Saltz

Department of Biomedical Informatics Stony Brook University

Feb 2017 BHI





- Pathology data is employed in care guidelines and clinical settings for virtually all cancer disease sites.
- Treatment decisions frequently hinge on subjective assessments -- poor inter-observer reproducibility.
- Widespread clinical adoption of Digital Pathology platforms in coming years
- Combination of Digital Pathology platforms and maturing of machine learning and artificial intelligence methodology will make possible adoption of image data driven decision support systems.
- Development and adoption of such systems will have tremendous impact on improving quality and consistency of clinical decision making.

Tools to Analyze Morphology and Spatially Mapped Molecular Data - U24 CA180924

- **Specific Aim 1** Analysis **pipelines** for multi- scale, integrative image analysis.
- Specific Aim 2: Database infrastructure to manage and query Pathomics features.
- Specific Aim 3: HPC software that targets clusters, cloud computing, and leadership scale systems.
- Specific Aim 4: Develop visualization middleware to relate Pathomics feature and image data and to integrate Pathomics image and "omic" data.



SEER Virtual Tissue Repository

Vision – Enable population/epidemiological cancer research that leverages rich cancer phenotype information available from Pathology tissue studies

NCIP/Leidos 14X138 and HHSN261200800001E - NCI

- Lynne Penberthy MD, MPH NCI SEER
- Ed Helton PhD NCI CBIIT Clinical Imaging Program
- Ulrike Wagner CBIIT Clinical Imaging Program
- Radim Moravec NCI PhD, NCI SEER
- Ashish Sharma PhD Biomedical Informatics Emory
- Joel Saltz MD, PhD Biomedical Informatics Stony Brook
- Tahsin Kurc PhD Biomedical Informatics Stony Brook
- Georgia Tourassi, Oak Ridge National Laboratory



SEER Virtual Tissue Repository

- SEER registries are a potential source of information about unusual outcomes and rare cancers
- Leverage Pathology labs which store FFPE tumors, slides and digital images
- Link to SEER data track long term outcomes
- SEER: 500K Cancer patients per year
- Accrue linked clinical data, Pathology slides from SEER sites



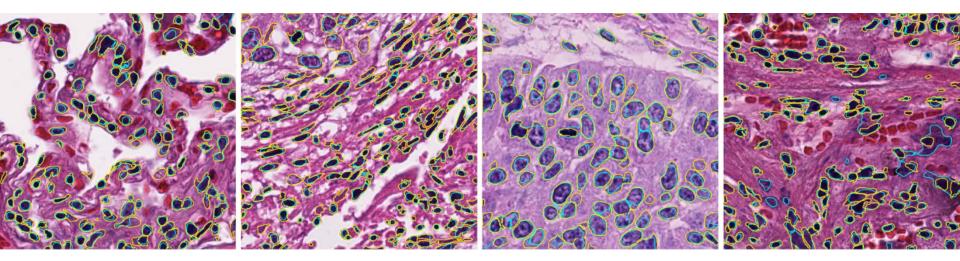
SEER VIRTUAL TISSUE REPOSITORY

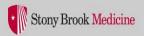
- Create linked collection of de-identified clinical data and whole slide images
- Extract features from a sample set of images (pancreas and breast cancer).
- Enable search, analysis, epidemiological characterization
- Pilot focus on extreme outcome Breast Cancer, Pancreatic Cancer cases
- Display images and analyzed features



Robust Nuclear Segmentation

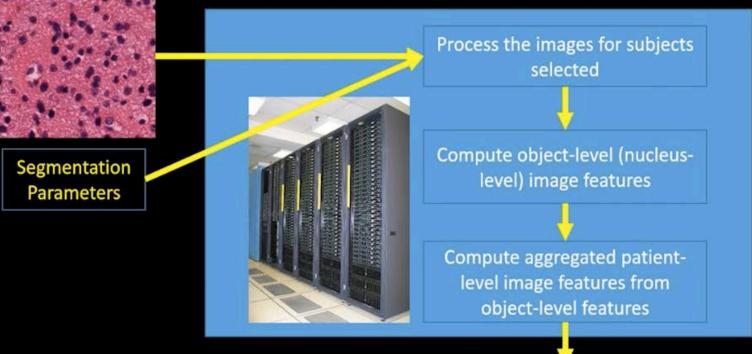
- Robust ensemble algorithm to segment nuclei across tissue types
- Optimized algorithm tuning methods
- Parameter exploration to optimize quality
- Systematic Quality Control pipeline encompassing tissue image quality, human generated ground truth, convolutional neural network critique
- Yi Gao, Allen Tannenbaum, Dimitris Samaras, Le Hou, Tahsin Kurc





Whole Slide Images (WSI)

Compute Cluster



FeatureDB

Load object-level imaging features and segmentation results
Load patient-level imaging features along with a selected subset of clinical and genomic data (e.g. gene mutations, days to death, vital status)



Feature Viz Suite

- Explore Relationship Between Imaging Features, Outcome, "omics"
- Explore relationships between features and explore how features relate to images

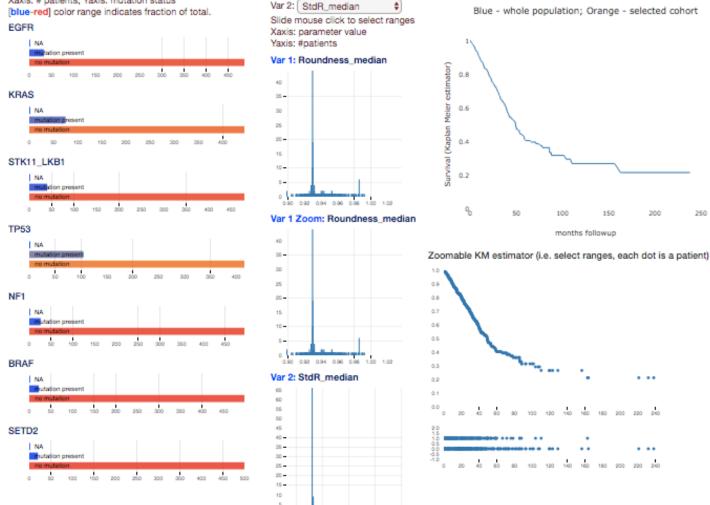


Feature Explorer - Integrated Pathomics Features, Outcomes and "omics" - TCGA NSCLC Adeno Carcinoma Patients

Survival

Gene Mutation

Click on bars to select molecular cohorts, Xaxis: # patients; Yaxis: mutation status [blue-red] color range indicates fraction of total.



100 120 140 160 180 200 220 240

o I=

Morphology, Epi, etc

Var 1: Roundness median 🛊

Blue - whole population; Orange - selected cohort

250

200

. ..

220 24

- 200



Feature Explorer - Integrated Pathomics Features, Outcomes and "omics" – TCGA NSCLC Adeno Carcinoma Patients

Gene Mutation

Click on bars to select molecular cohorts, Xaxis: # patients; Yaxis: mutation status [blue-red] color range indicates fraction of total.

2.0

1.8

1.6

14

12

1.0

0.0 0.6

0.4

0.2 0.0

1.0

0.9

0.8

0.7

0.6

0.5

0.4

0.3

0.2

0.1

2.0

1.0

1.6 14

12

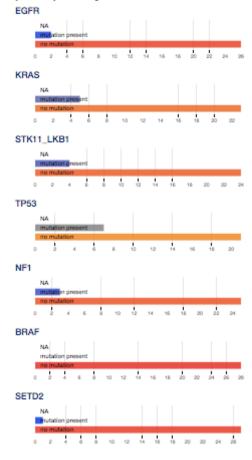
1.0

0.8

0.6 0.4 0.2

100 120 140

00



Morphology, Epi, etc Survival Var 1: Roundness median 🛊 0 B Q 🕂 B B X # 🖛 🚍 🛝 Var 2: StdR_median Slide mouse click to select ranges Xaxis: parameter value Yaxis: #patients 1 0.9638548 trace 0 Var 1: Roundness median estimator) 0.8 Meler 0.6 (Kaplan 0.4 1 0.2 0.94 0.96 0.92 100 150 200 3.975359 Var 1 Zoom: Roundness_median months followup Zoomable KM estimator (i.e. select ranges, each dot is a patient) 10 0.9 0.8 0.7 0.6 0.5 0.4 0.90 0.92 0.94 0.96 0.98 1.00 0.3 Var 2: StdR median 0.1 40 60 80 100 120 140 160 180 200 220 240 80 100 120 140 160 180 200 220 240

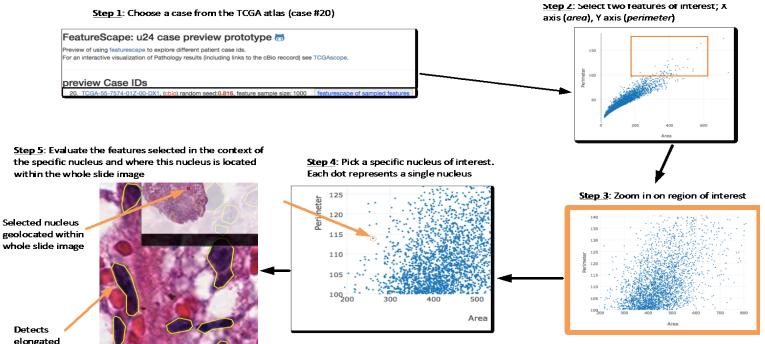
Blue - whole population; Orange - selected cohort

250



Pathomics

Relationship Between Image and Features



Detects elongated nucleus

The tool provides visual context for feature evaluation. This technique maps both intuitive features (i.e. size, shape, color) and non-intuitive features (i.e. wavelets, texture) to the ground truth of source images through an interactive web-based user interface.



Select Feature Pair – dots correspond to nuclei

FeatureScape a reliminary demo of integrative use of multiple FeatureScape tools

1000 entries sampled from https://tahsin175.informatics.stonybrook.edu:4500/?limit=1000&find={%22randval%22: {%22\$gte%22:0.149},%22provenance.analysis_execution_id%22:%22lung-features-v4%22,%22image.caseid%22:%22TCGA-38-4628-01Z-00-DX1%22}

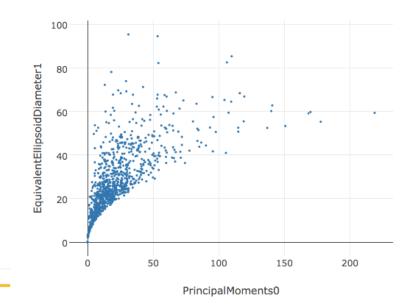
+ Load Data

Cross-tabulated feature correlations

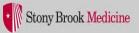
E 101 1	
FeretDiameter	000000000000000000000000000000000000000
MeanR	000000000000000000000000000000000000000
MeanG	000000000000000000000000000000000000000
MeanB	000000000000000000000000000000000000000
StdR	000000000000000000000000000000000000000
StdG	000000000000000000000000000000000000000
StdB	000000000000000000000000000000000000000
EquivalentSphericalRadius	000000000000000000000000000000000000000
EquivalentEllipsoidDiameter	000000000000000000000000000000000000000
EquivalentEllipsoidDiameter1	00000000000000000000000000000000000000
Perimeter	000000000000000000000000000000000000000
EquivalentSphericalPerimete	rooooooooooooooooooooooooo
PrincipalMoments0	000000000000000000000000000000000000000
PhysicalSize	000000000000000000000000000000000000000
PrincipalMoments1	000000000000000000000000000000000000000
Area	000000000000000000000000000000000000000
NumberOfPixels	000000000000000000000000000000000000000
NumberOfPixelsOnBorder	000000000000000000000000000000000000000
Roundness	000000000000000000000000000000000000000
Elongation	000000000000000000000000000000000000000
Flatness	000000000000000000000000000000000000000



- PrincipalMoments0
- EquivalentEllipsoidDiameter1



Resample from selected region (under development)



Subregion selected – form of gating analogous to flow cytometry

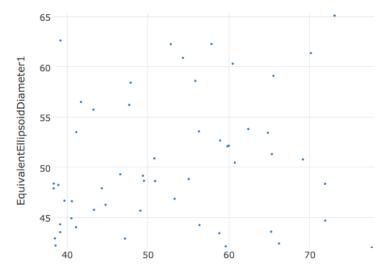
FeatureScape a reliminary demo of integrative use of multiple FeatureScape tools

1000 entries sampled from https://tahsin175.informatics.stonybrook.edu:4500/?limit=1000&find={%22randval%22: {%22\$gte%22:0.149},%22provenance.analysis_execution_id%22:%22lung-features-v4%22,%22image.caseid%22:%22TCGA-38-4628-01Z-00-DX1%22}

+ Load Data

Cross-tabulated feature correlations

FeretDiameter	000000000000000000000000000000000000000
MeanR	0 000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
MeanG	000000000000000000000000000000000000000
MeanB	000000000000000000000000000000000000000
StdR	000000000000000000000000000000000000000
StdG	000000000000000000000000000000000000000
StdB	000000000000000000000000000000000000000
EquivalentSphericalRadius	000000000000000000000000000000000000000
EquivalentEllipsoidDiameter0	000000000000000000000000000000000000000
EquivalentEllipsoidDiameter1	00000000000000000000000000000000000000
Perimeter	000000000000000000000000000000000000000
EquivalentSphericalPerimete	rooooooooooooooooooooo
PrincipalMoments0	000000000000000000000000000000000000000
PhysicalSize	000000000000000000000000000000000000000
PrincipalMoments1	000000000000000000000000000000000000000
Area	000000000000000000000000000000000000000
NumberOfPixels	000000000000000000000000000000000000000
NumberOfPixelsOnBorder	000000000000000000000000000000000000000
Roundness	000000000000000000000000000000000000000
Elongation	000000000000000000000000000000000000000
Flatness	000000000000000000000000000000000000000



PrincipalMoments0

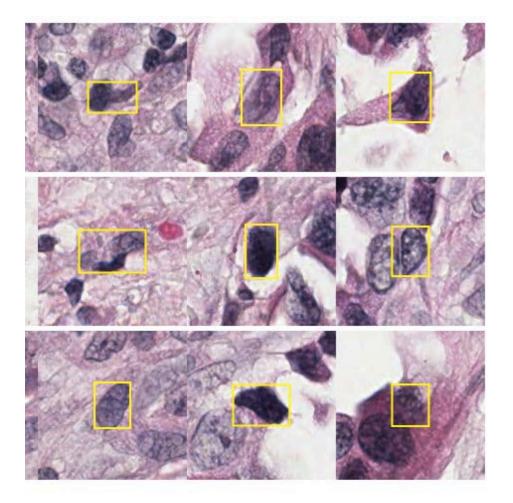
Resample from selected region (under development)

Pearson correlation between

- PrincipalMoments0
- EquivalentEllipsoidDiameter1

* Stony Brook Medicine

Sample Nuclei from Gated Region





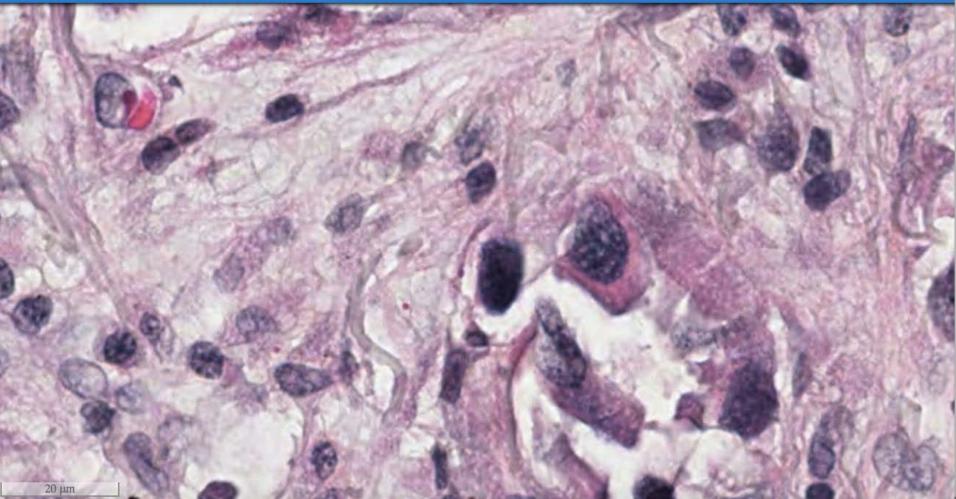
Gated Nuclei in Context

🖉 🖉 🖉 🖉 📃 📔 🚺

caMicroscope

1

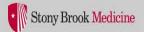
SubjectID :TCGA-38-4628-01Z-00-DX1





Docker and Virtual Machine Distributions

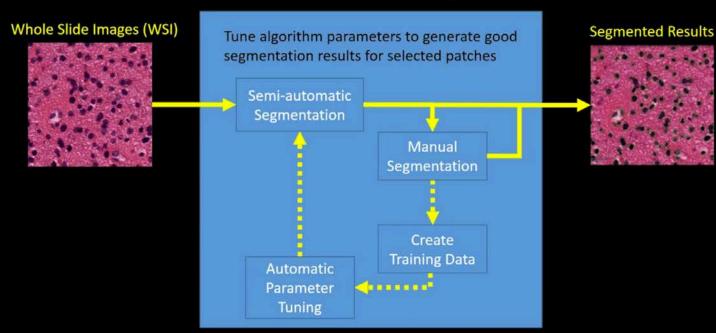
K VMware Fusion File Edit View Virtual Machine Window Help ① 😤 ◀ 66% 🗈 Thu Feb 16 9:32 AM Q 🔚 misc e quip [User: quip -- QuIP login] - Google Chrome t1 40) 6:32 AM ₫ D [User: quip - QuiP lo × Q C () localhost ☆ codar_example QuIP: Quantitative Imaging for Pathology Ulick on any of the colored buttons to launch the associated tool. SEER implement ation Image and Results Viewer R You can view whole slide tissue images and nuclear segmentations. Click on the magnifier icon to choose algorithm results. You may zoom in, zoom out, and pan the images. Mouse Click: Zoom in, Shift-Click: Zoom out. Visual Feature Analytics A visual analytics platform for exploring slide-level imaging features generated by analysis of whole slide tissue images to OuIP Upload Images Web application for loading whole slide tissue images to QuIP **QuIP Repository**



3D Slicer Pathology – Generate High Quality Ground Truth

ITCR - Tools to Analyze Morphology and Spatially Mapped Molecular Data

3D Slicer Pathology

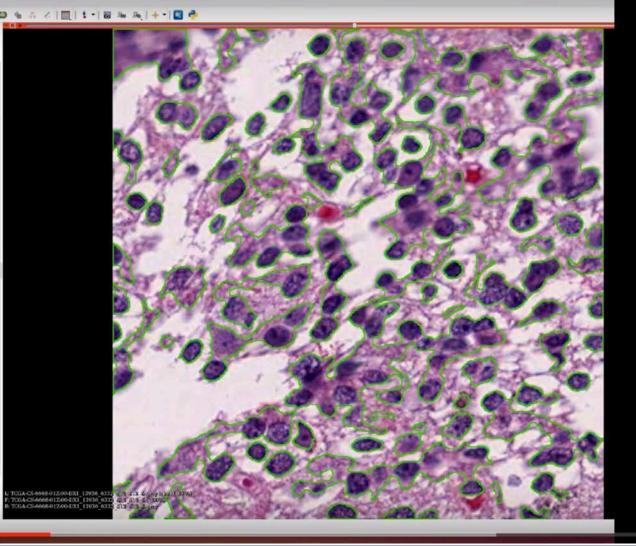




Apply Segmentation Algorithm

ITCR - Tools to Analyze Morphology and Spatially Mapped Molecular Data

305Hcer				
Create and Select	Label Breez			
create and select	Label Maps			
faster Volume: T	CGA C5 8668 012	00-DX1_13936_6332_5	12_512_6_gray	
ferge Volume: TC	GA CS 6668-01Z	00-DX1_13936_6332_5	12_512_6_gray-label	•
Per-Structure Vo	lumes			
Edit Selected Labe	el Map			
	<u>12</u> 2 2			
Active Tool:	1. **	QuickTCGAE#	ec:	
				-
abet • Terminology • Paint Over	Nucleus	1		
Terminology K Paint Over Threshold Paint	matic segmentati	on on the current image	rusing given parame	ters.
Terminology K Paint Over Threshold Paint	matic segmentati	on on the current image Quick TCGA Segmenter	rusing given parame	ters.
Terminology K Paint Over Threshold Paint	matic segmentati	on on the current image Quick TCGA Segmenter Clear Selection	using given parame	ters.
Terminology K Paint Over Threshold Paint ress Y to run autor	matic segmentati Stop	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	rusing given parame	ters.
Terminology Paint Over Threshold Paint ress Y to run aidor Nucleus Segme	matic segmentati Stop	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given parame	ters.
Terminology C Paint Over Threshold Paint ress 4 to run aidor Nucleus Segme Dtsu Threshold	matic segmentati Stop Intation Paramete	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given parame	
Terminology C Paint Over Threshold Paint ress 4 to run aided Vaulleus Segme Dtsu Threshold Curvature Weight:	matic segmentati Stop Intation Paramete	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	0.8 ¢
	matic segmentati Stop mation Paramete	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	
Terminology K Paint Over Threshold Paint reas 4 to run aider Nucleus Segme Otsu Threshold Curvature Weight: Size Threshold Size Upper Threshold	matic segmentati Stop mation Paramete	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	- 0.8 C - 8.0 C - 8.0 C
Terminology Teint Over Threshold Paint Threshold Paint Threshold Paint Nucleus Segme Nucleus Segme Otsu Threshold Otrusture Weight: Size Threshold Size Upper Threshold Size:	matic segmentati Stop mation Paramete	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	
Terminology Teint Over Threshold Paint Treshold Paint Nucleus Segme Otsu Threshold. Curvature Weight: Size Upper Threshold Size Upper Threshold Size (Upper Thr	matic segmentati Stop mation Paramete	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	
Reminology Paint Over Threshold Paint treas ¥ to run aider Nucleus Segme Dtsu Threshold Curvature Weight: Size Upper Threshold Kernel Size ? rameters	matic segmentati Stop Intation Parameter	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	
Terminology	matic segmentati Step	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	= 08 0 = 80 0 30 0 = 100 0 = 15 0
Serminology Anint Over Threshold Paint treas Y to run aiddo Nucleus Segme Nucleus Segme Otsu Threshold Curvature Weight: Size Threshold Size Upper Threshold Size Upper Threshold Size Segme Size Threshold Size topology rameters e Screenshots rishot scale factor	matic segmentati Step	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	
Terminology Paint Over Threshold Paint treas Y to run aider Nucleus Segme Otsu Threshold Curvature Weight: Size Upper Threshold Size Upper Threshold reshold rameters	natic segmentati Step Intation Paramete Content of the Content of	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	= 08 0 = 80 0 30 0 = 100 0 = 15 0
Terminology Teint Over Threshold Paint treas Y to run aidda Voucleus Segme Otsu Threshold Curvature Weight: Size Threshold Size Upper Threshold Size Upper Threshold Size Upper Threshold Size Screenshots erstreamshots enshot scale factor ta Probe	natic segmentati Step Intation Paramete Content of the Content of	on on the current image Quick TCGA Segmenter Clear Selection Run Segmentation	using given pacame	= 08 0 = 80 0 30 0 = 100 0 = 15 0

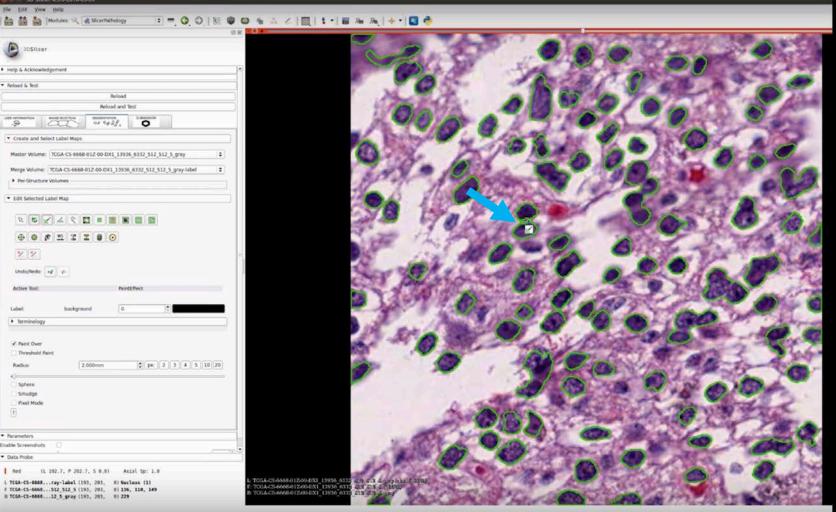


Stony Brook Medicine

U

Adjust algorithm parameters, manual fine tuning

ITCR - Tools to Analyze Morphology and Spatially Mapped Molecular Data





Classification

- Automated or semi-automated identification of tissue or cell type
- Variety of machine learning and deep learning methods
- Quantification of lymphocyte infiltration
 - Collaboration with TCGA Pan Can Atlas Immune Group
- Classification of Neuroblastoma
- Classification of Gliomas



Team

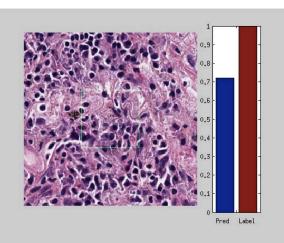
- TCGA Pan Can Immune:
 - Vesteinn Thorsson
 - Iya Shmulevich
- TIL Project Leads
 - Joel Saltz
 - Dimitris Samaras
 - Tahsin Kurc
 - Alex Lazar
- caMicroscope Lead
 - Ashish Sharma

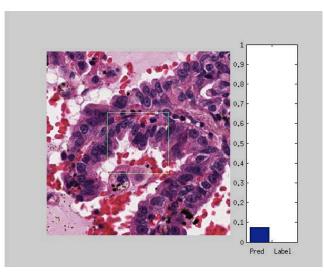
- Deep Learning Graduate Students
 - Le Hou
 - Vu Nyugen
- Pathology Fellows/ Faculty
 - Anne Zhao
 - John Van Arnam
 - Rebecca Batiste
- Biostatistics
 - Arvind Rao
- Active Learning Collaborator
 - Lee Cooper

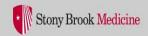


TCGA PanCan TILS Collaborative Effort

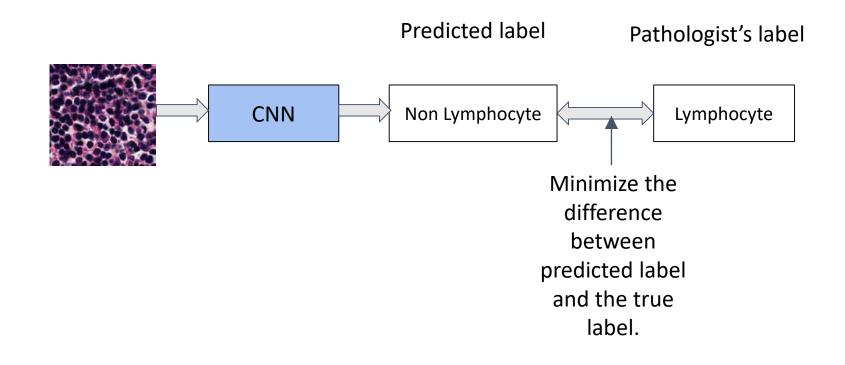
- Deep learning algorithm trained on 20K+ patches
- Pathologist correction is essential to reduce false positives as there are many patches.
- GUI developed to accomplish this rolled out to TCGA Pathologists
- TCGA TIL rich tumors including NSCLC adenocarcinoma, breast, pancreatic, colorectal, skin and uveal melanoma
- Working group of TCGA Pathologists - leverage tool to generate TCGA TIL data and TIL maps (Alex Lazar)
- CNN Algorithm presented at USCAP 2017 – Zhao et al





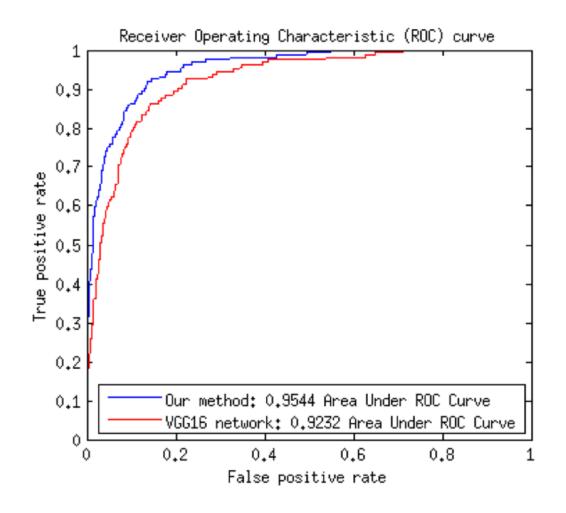


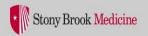
Training a CNN





Patch Based Performance Evaluation of CNN Classification – TCGA Non Small Cell Lung Cancer





Lymphocyte Classification Heat Map

Trained with 22.2K image patches Pathologist corrects and edits

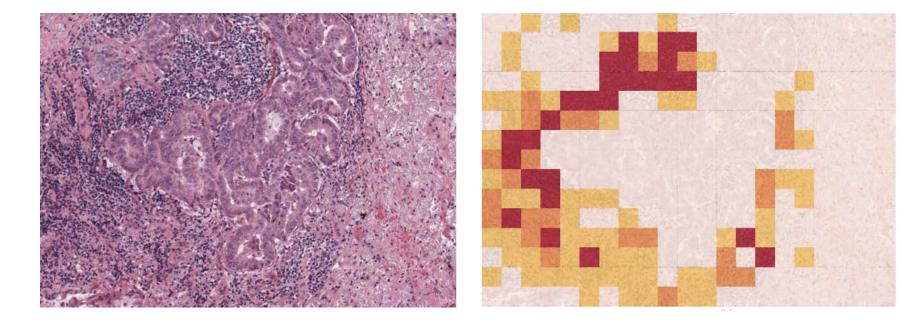


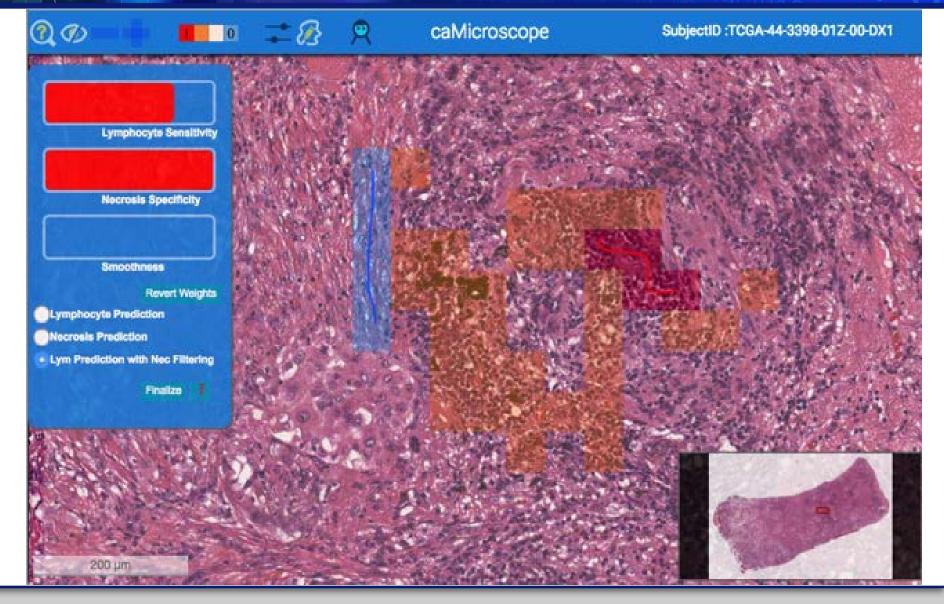


Image based TIL prediction

- Initial unsupervised training step (autoencoder) initialize CNN
- Train CNN on initial supervised dataset
- Apply CNN to obtain predicted lymphcyte heatmaps
- Pathologists edit heatmaps using caMicroscope
- Extract new training data from edited heatmaps
- Sampling algorithm to adjust thresholds
 - USCAP 2017 Zhao et al, submitted publication to International Conference on Computer Vision

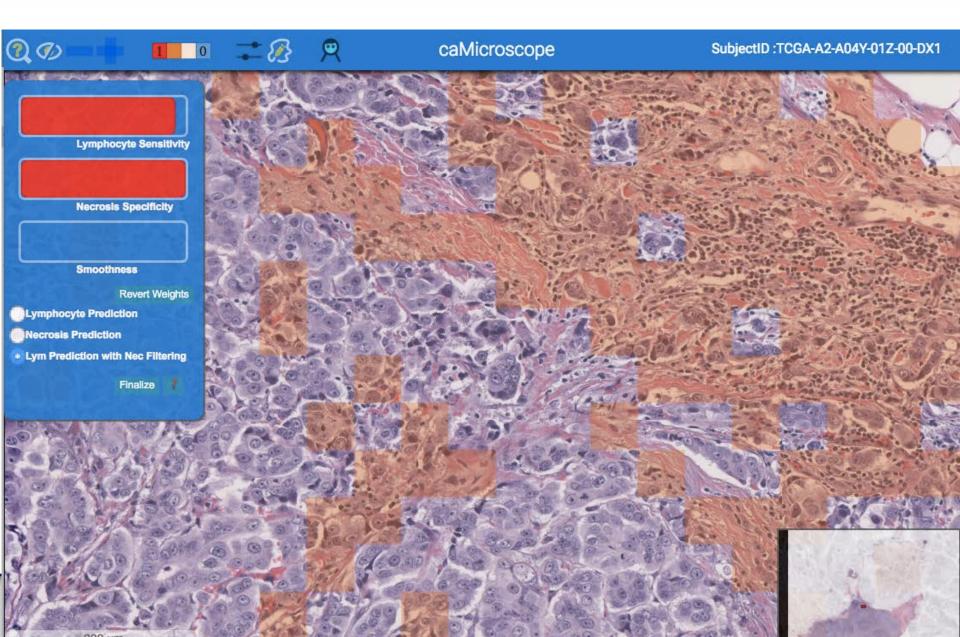


caMicroscope with TIL heatmap

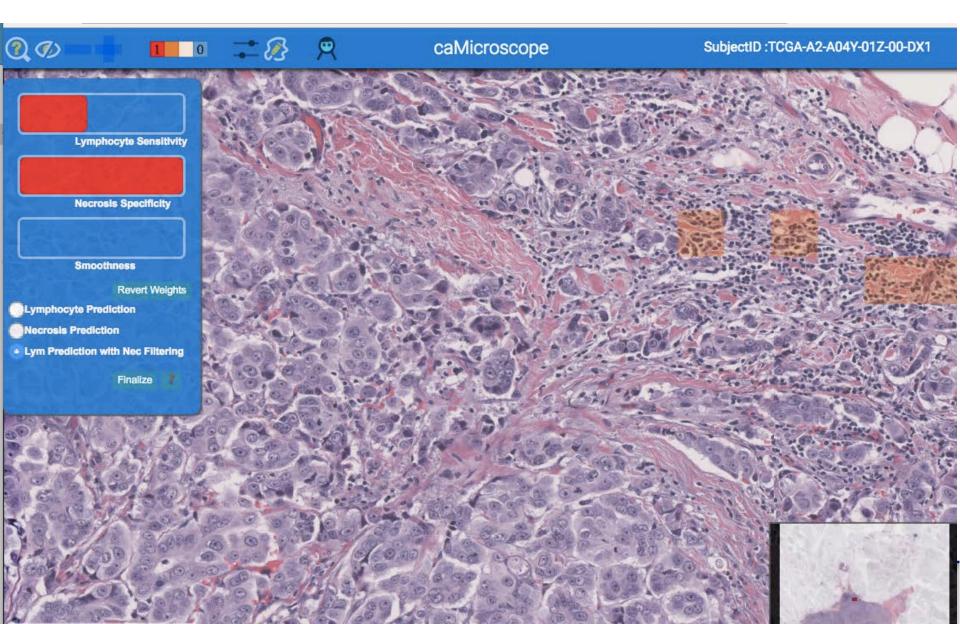




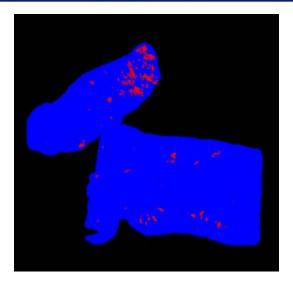
Adjust sensitivity - High

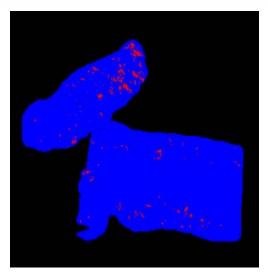


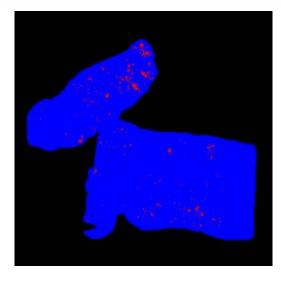
Adjust sensitivity - Low



TIL Distribution Maps







Prediction edited by Pathologist 1 Prediction without editing

5 mm

Prediction edited by Pathologist 2

Tissue Specimen



TCGA Pan Can Immune

- Roughly 3.5K TIL maps generated to date with pipeline on track to complete roughly 10K by mid-April
- Comparison with TIL molecular epigenetic and RNA seq data
- Initial draft manuscript completed March 24th
- Deeper dive into TCGA "omics" analytics



Dissemination

- Containers
 - Containerized segmentation algorithm/FeatureDB Employed to support TIES, MICCAI, and competitions supported through Kalpathy-Kramer ITCR
 - Full containerized implementation of caMicroscope/FeatureDB/Segmentation algorithm/Feature Scape - Feb 1 2017
- Cloud Pilots
- TCIA
- HPC via NSF and DOE
- TCGA PanCanAtlas Lymphocyte characterization
- Integrated Features/NLP joint with TIES



ITCR Team

Stony Brook University

Joel Saltz Tahsin Kurc Yi Gao Allen Tannenbaum **Erich Bremer** Jonas Almeida Alina Jasniewski Fusheng Wang Tammy DiPrima Andrew White Le Hou **Furqan Baig** Mary Saltz

Emory University

Ashish Sharma Adam Marcus

Oak Ridge National Laboratory

Scott Klasky Dave Pugmire Jeremy Logan

Yale University Michael Krauthammer

Harvard University Rick Cummings



Funding – Thanks!

- This work was supported in part by U24CA180924-01, NCIP/Leidos 14X138 and HHSN261200800001E from the NCI; R01LM011119-01 and R01LM009239 from the NLM
- This research used resources provided by the National Science Foundation XSEDE Science Gateways program under grant TG-ASC130023 and the Keeneland Computing Facility at the Georgia Institute of Technology, which is supported by the NSF under Contract OCI-0910735.



Thanks!

