Digital Pathology – Precision Medicine, Pathomics and Decision Support

Joel Saltz MD, PhD

Chair and Professor Department of Biomedical Informatics

Professor Department of Pathology

Cherith Endowed Chair

Stony Brook University

RSNA Session 553

November 28 2018

Stony Brook University

NO CONFLICTS TO DISCLOSE

Pathology Image Driven Decision Support

Stony Brook University

- Improve reproducibility in traditional Pathology assessments (e.g. Gleason grade, NSCLC subtypes)
- Precise scoring of well known criteria (tumor infiltrating lymphocytes, mitoses and IHC staining)
- Development of novel computational methods to employ Pathology image information to predict response to cancer treatment and outcomes.

What is the Gleason grade or Gleason score? What do the numbers in the Gleason score mean, for example 3+4=7 or 3+3=6?

Pathologists grade prostate cancers using numbers from 1 to 5 based on how much the cells in the cancerous tissue look like normal prostate tissue under the microscope. This is called the *Gleason system*. Grades 1 and 2 are not often used for biopsies – most biopsy samples are grade 3 or higher.

- If the cancerous tissue looks much like normal prostate tissue, a grade of 1 is assigned.
- If the cancer cells and their growth patterns look very abnormal, a grade of 5 is assigned.
- Grades 2 through 4 have features in between these extremes.

Since prostate cancers often have areas with different grades, a grade is assigned to the 2 areas that make up most of the cancer. These 2 grades are added to yield the Gleason score (also called the *Gleason sum*). The highest a Gleason score can be is 10.

Major treatment decisions can hinge on subjective judgements

Stony Brook University

The ASCO Post

A Gleason 6 Tumor: Is It Cancer, and Should It Be Treated?

cancernetwork

Gleason 6 Prostate Cancer: Serious Malignancy or Toothless Lion?

By Herbert Lepor, MD and Nicholas M. Donin, MD

Jan 15, 2014

PROSTATE CANCER

DISCOVERY

"Tumors with a Gleason score of 4 + 3 are more aggressive and predictive of advanced disease at the time of surgery, compared to Gleason 3 + 4 tumors," explains Mark L. Gonzalgo, M.D., Ph.D., assistant professor of urology and oncology. In a recent study, published in the journal Urology, Gonzalgo and urologists Alan W. Partin, M.D., Ph.D., and Patrick C. Walsh, M.D., investigated the relationship between a man's biopsy Gleason score, the Gleason score in the entire prostate (the specimen removed during radical prostatectomy) and the recurrence of PSA among men who were diagnosed with Gleason 7 cancer in a needle biopsy.

BJUI
BJU International

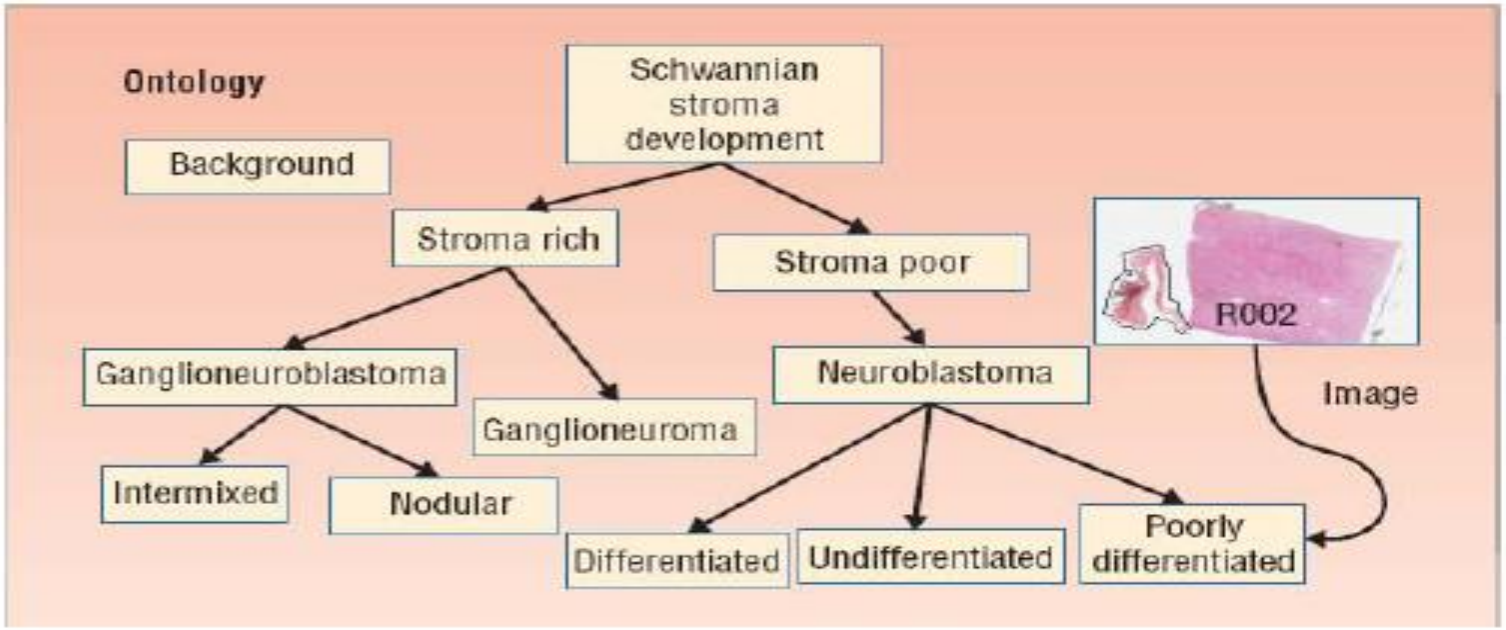
Editorial: Current Gleason score 3 + 4 = 7: has it lost its significance compared with its historical counterpart?

01 Jun 2016

Early Steps to Pathology Computer Aided Classification 2005-2010

BISTI/NIBIB Center for Grid Enabled Image Analysis - P20 EB000591, PI Saltz

- Analyze images by computer
- Analyze the whole tissue, several slides
- Provide quantitative information to the pathologist
- Reduce inter- and intra-reader variability

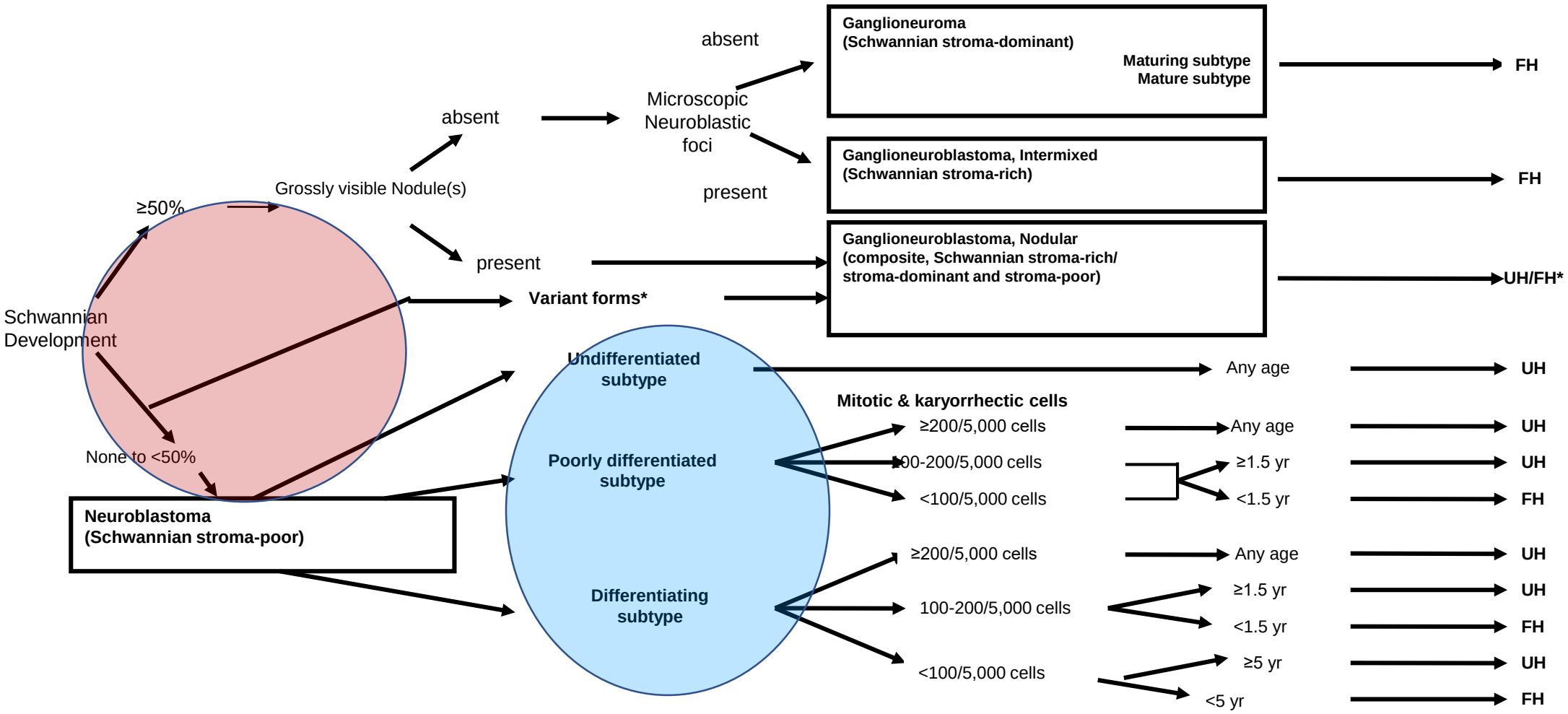


Morphological characterization of tissue used for prognosis

Hiro Shimada, Metin Gurcan, Jun Kong, Lee Cooper Joel Saltz

Gurcan, Shamada, Kong, Saltz

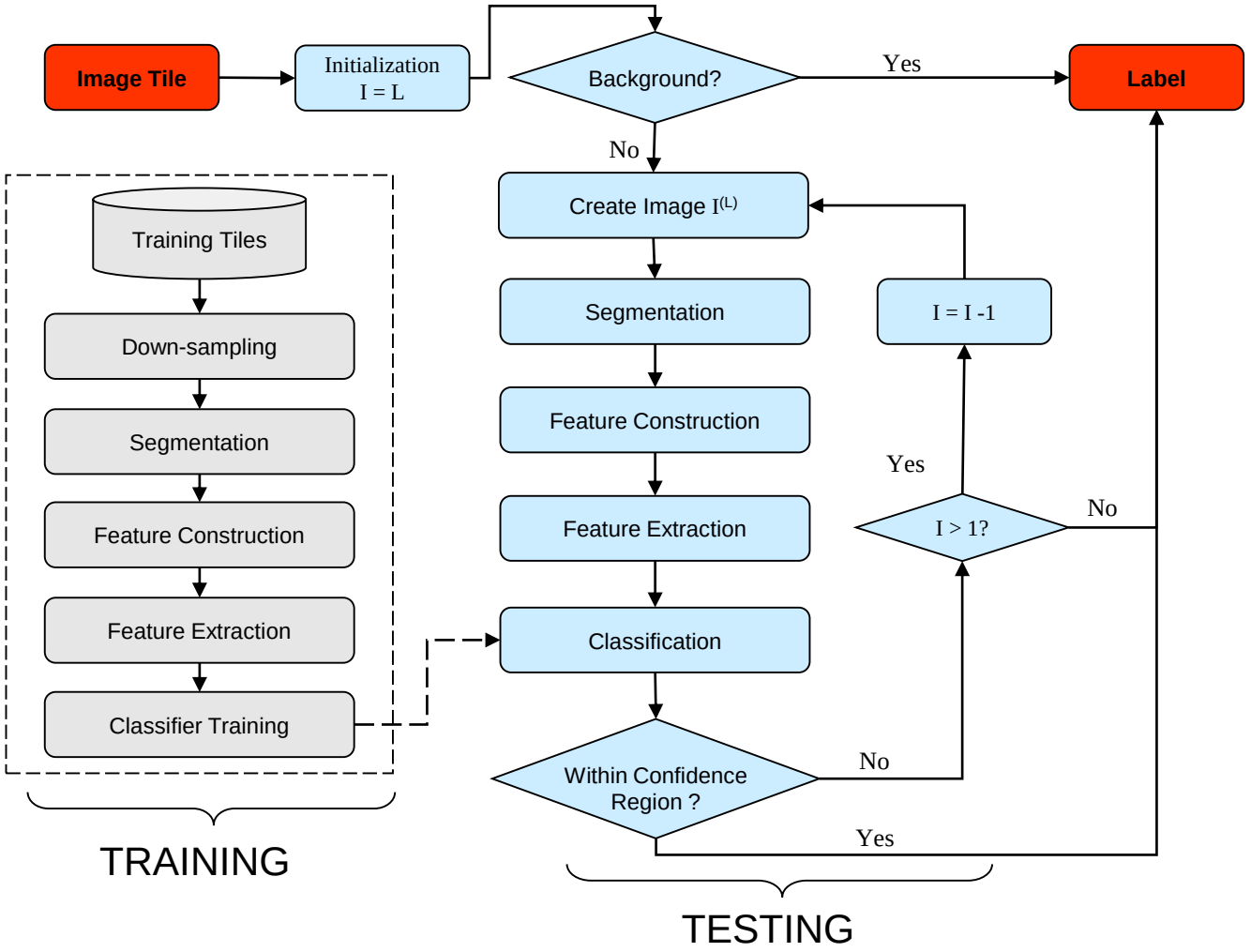
Neuroblastoma Classification



FH: favorable histology **UH:** unfavorable histology

CANCER 2003; 98:2274-81

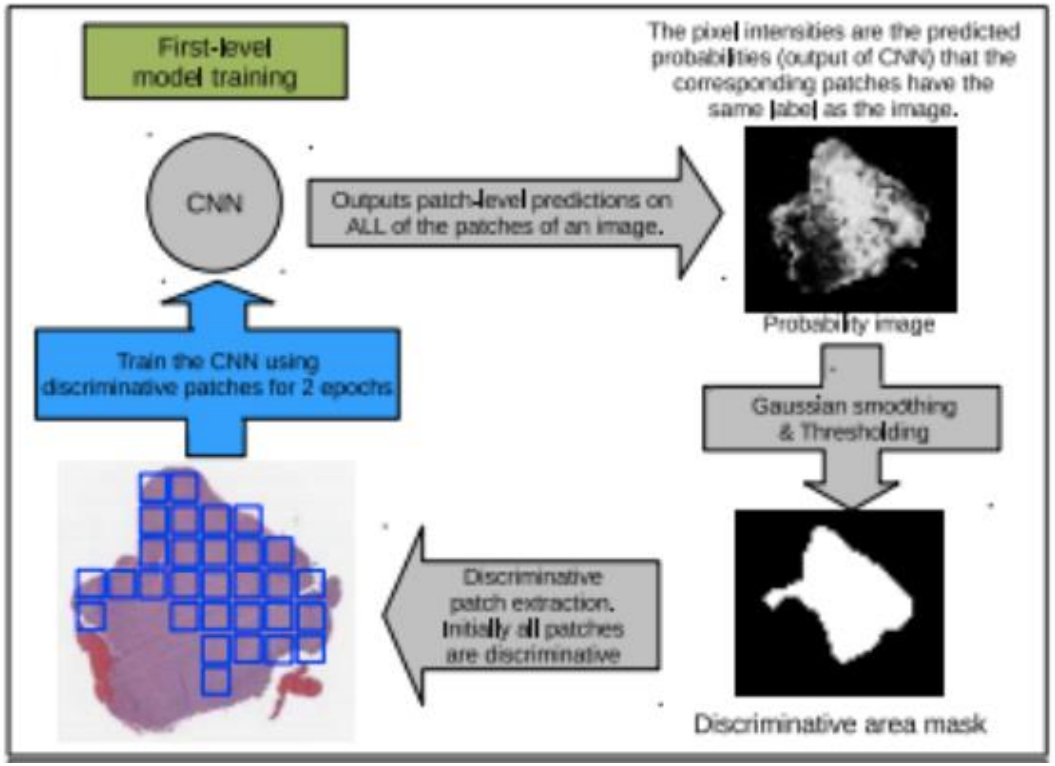
Multi-Scale Machine Learning Based Shimada Classification System

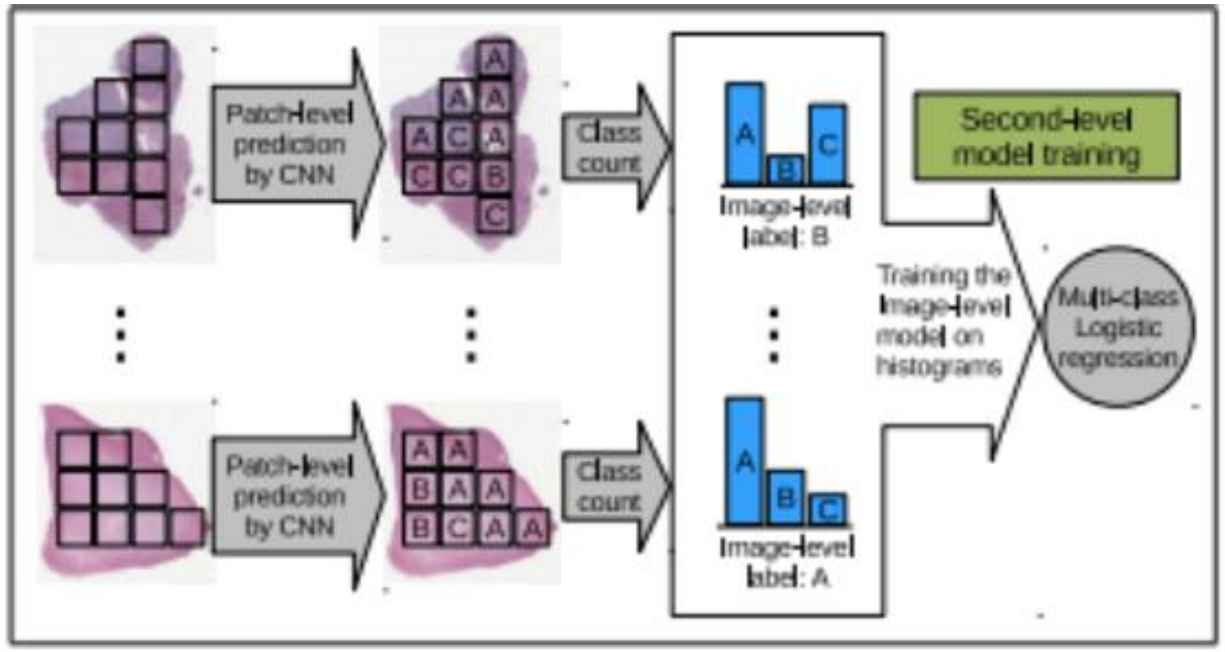


- Background Identification
- Image Decomposition (Multi-resolution levels)
- Image Segmentation (EMLDA)
- Feature Construction (2nd order statistics, Tonal Features)
- Feature Extraction (LDA) + Classification (Bayesian)
- Multi-resolution Layer Controller (Confidence Region)

Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification

Le Hou, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, James E. Davis, Joel H. Saltz; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2424-2433





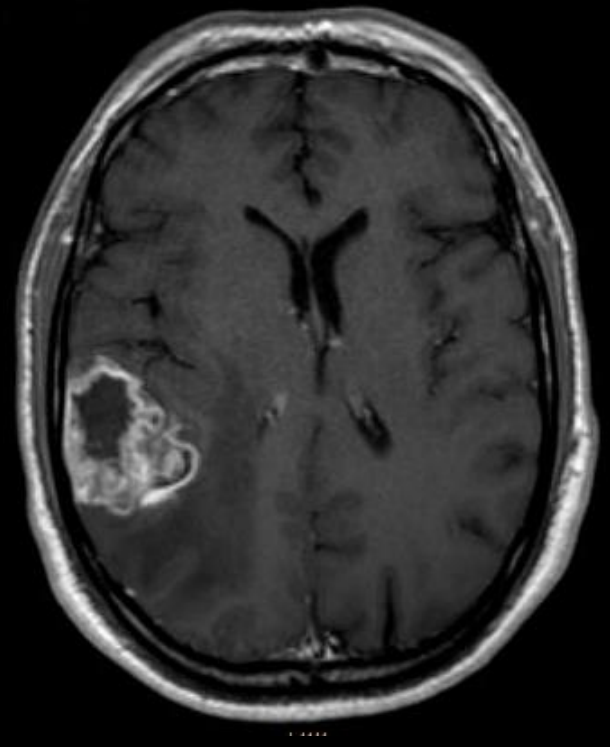
Confusion Matrix: OA is very hard even for pathologists	GBM	OD	OA	DA	AA	AO
Glioblastoma, Grade IV (GBM)	214		2		1	
Oligodendroglioma, Grade II (OD)	1	47	22	2		1
Oligoastrocytoma, Grade II & III (OA)	1	18	40	8	3	1
Diffuse Astrocytoma, Grade II (DA)	3	9	6	20		1
Anaplastic Astrocytoma, Grade III (AA)	3	2	3	3	4	
Anaplastic Oligodendroglioma, Grade III (AO)	2	2	3			1

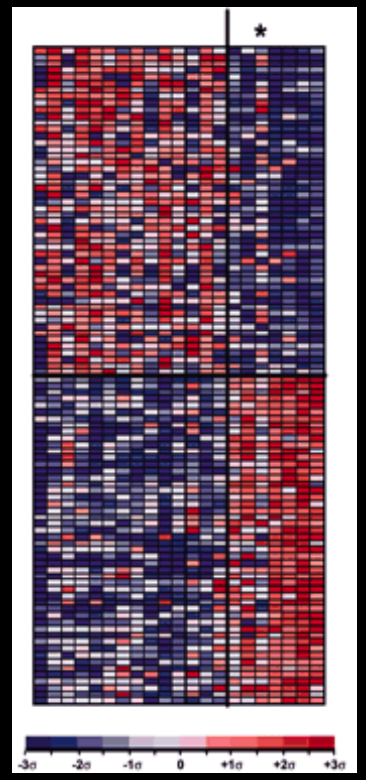
Le Hou, Dimitris Samaras, Tahsin Kurc, Yi Gao, Liz Vanner, James Davis, Joel Saltz

Digital Pathology as Precision Medicine

Stony Brook University

- Statistical analyses and machine learning to link Radiology/Pathology features to “omics” and outcome biological phenomena
- Image analysis and deep learning methods to extract features from images
- Support queries against ensembles of features extracted from multiple datasets
- Identify and segment trillions of objects – nuclei, glands, ducts, nodules, tumor niches
- Analysis of integrated spatially mapped structural/”omic” information to gain insight into cancer mechanism and to choose best intervention

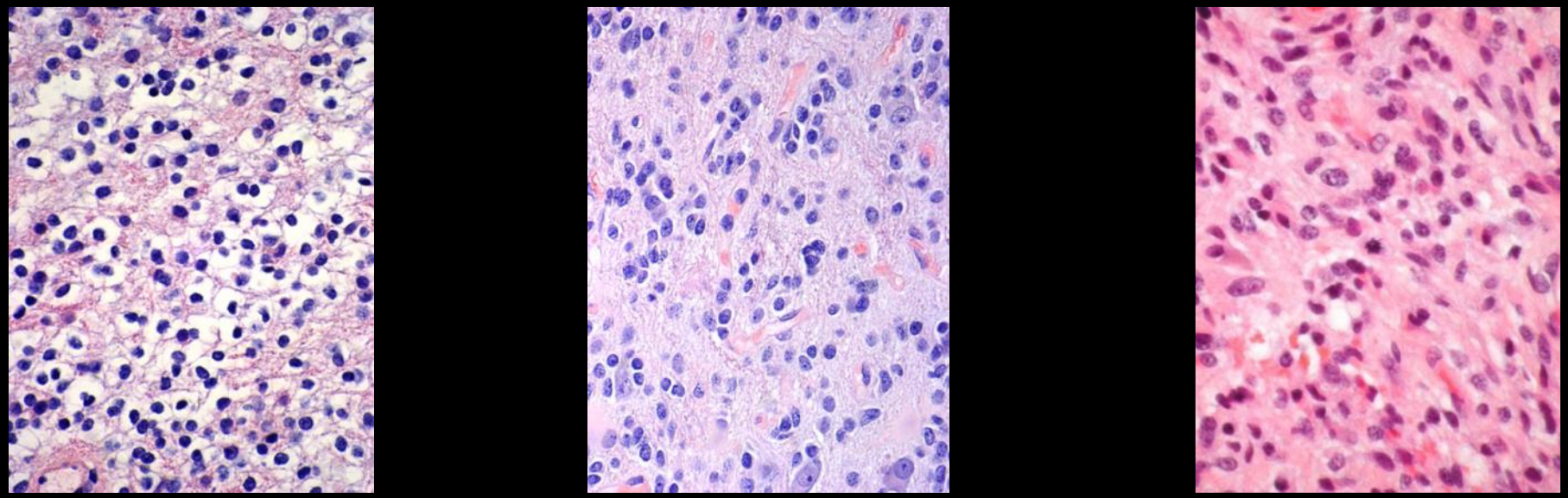




Quantitative Feature Analysis in Pathology: Emory In Silico Center for Brain Tumor Research (PI = Dan Brat, PD= Joel Saltz) 2009 - 2013

Can we use image analysis of TCGA GBMs TO INFORM diagnostic criteria based on molecular or clinical endpoints?

Nuclear Qualities



Oligodendroglioma **Astrocytoma**

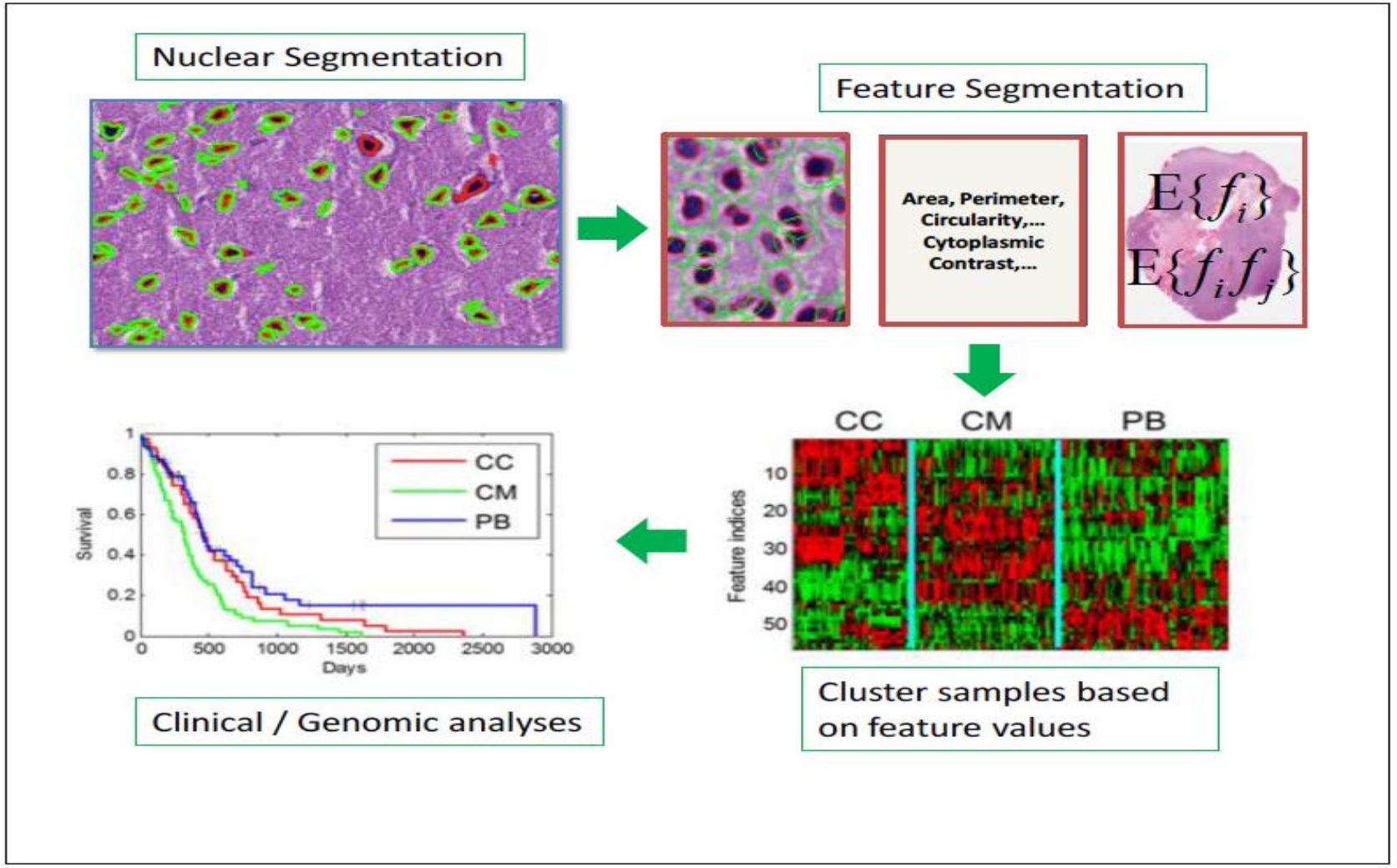
Application: Oligodendroglioma Component in GBM

Integrative Morphology/"omics"

Quantitative Feature Analysis in Pathology: Emory In Silico Center for Brain Tumor Research (PI = Dan Brat, PD= Joel Saltz)

NLM/NCI: Integrative Analysis/Digital Pathology R01LM011119, R01LM009239 (Dual PIs Joel Saltz, David Foran)

Marcus Foundation Grant – Ari Kaufman, Joel Saltz



Stony Brook

Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images

Joel Saltz,[1,*] Rajarsi Gupta,[1,4] Le Hou,[2] Tahsin Kurc,[1] Pankaj Singh,[3] Vu Nguyen,[2] Dimitris Samaras,[2] Kenneth R. Shroyer,[4] Tianhao Zhao,[4] Rebecca Batiste,[4] John Van Arnam,[5] The Cancer Genome Atlas Research Network, Ilya Shmulevich,[6] Arvind U.K. Rao,[3,7] Alexander J. Lazar,[8] Ashish Sharma,[9] and Vésteinn Thorsson[6,10,*]

http://www.cell.com/cell-reports/pdf/S2211-1247(18)30447-9.pdf

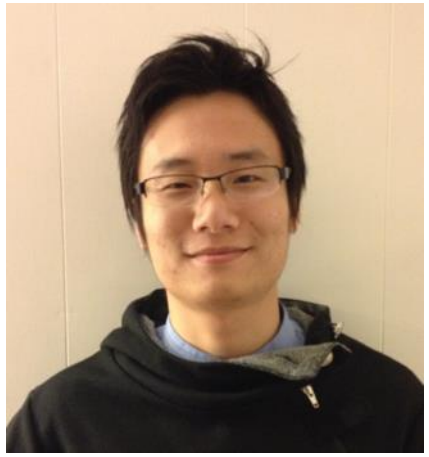
- Stony Brook, Institute for Systems Biology, MD Anderson, Emory group
- TCGA Pan Cancer Immune Group – led by ISB researchers
- Deep dive into linked molecular and image based characterization of cancer related immune response

- 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



Le Hou – Graduate Student
Computer Science

Anne Zhao – Pathology Informatics
Biomedical Informatics, Pathology
(now Surg Path Fellow SBM)

Vu Nguyen– Graduate Student
Computer Science

Deep Learning and Lymphocytes: Stony Brook Digital Pathology Trainee Team

Raj Gupta – Pathology Informatics
Biomedical Informatics, Pathology

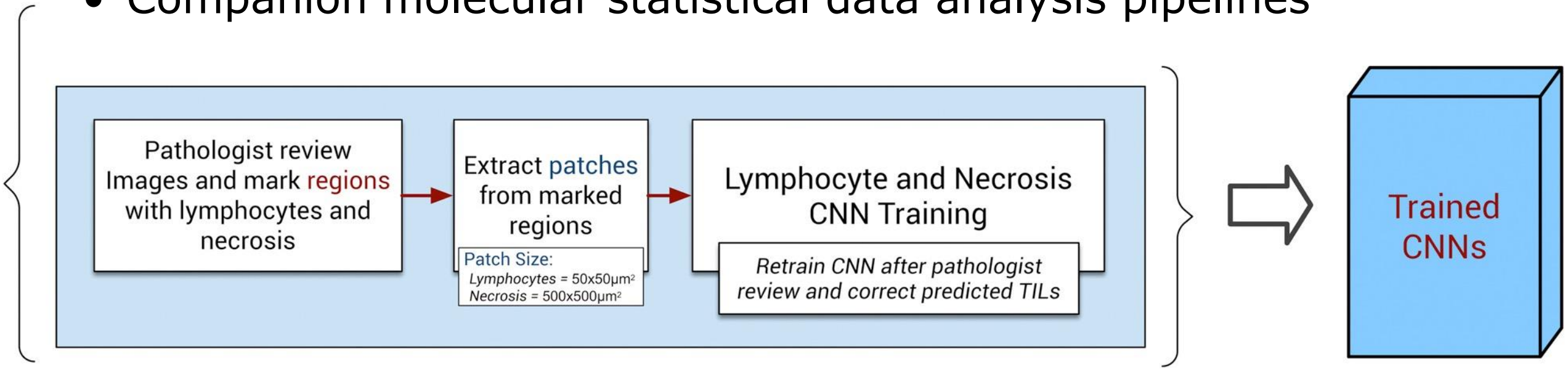
The future of Digital Pathology

Importance of Immune System in Cancer Treatment and Prognosis

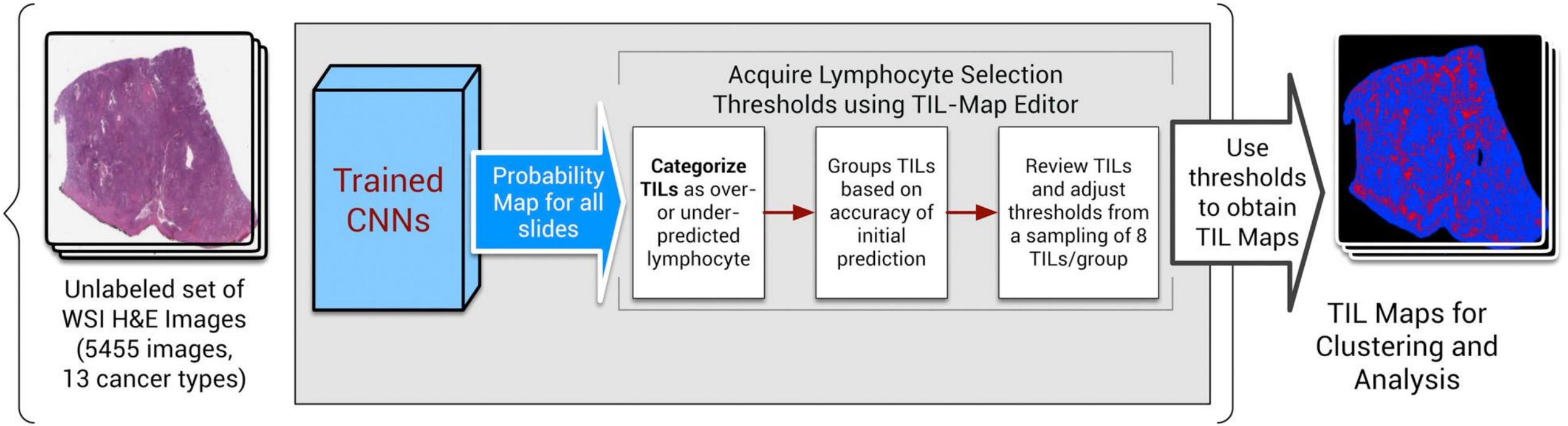
- Tumor spatial context and cellular heterogeneity are important in cancer prognosis
- Spatial TIL densities in different tumor regions have been shown to have high prognostic value – they may be superior to the standard TNM classification
- Immune related assays used to determine Checkpoint Inhibitor immune therapy in several cancer types
- Strong relationships with molecular measures of tumor immune response – results to soon appear in TCGA Pan Cancer Immune group publications
- TIL maps being computed for SEER Pathology studies and will be routinely computed for data contributed to TCIA archive
- Ongoing study to relate TIL patterns with immune gene expression groups and patient response

Training, Model Creation

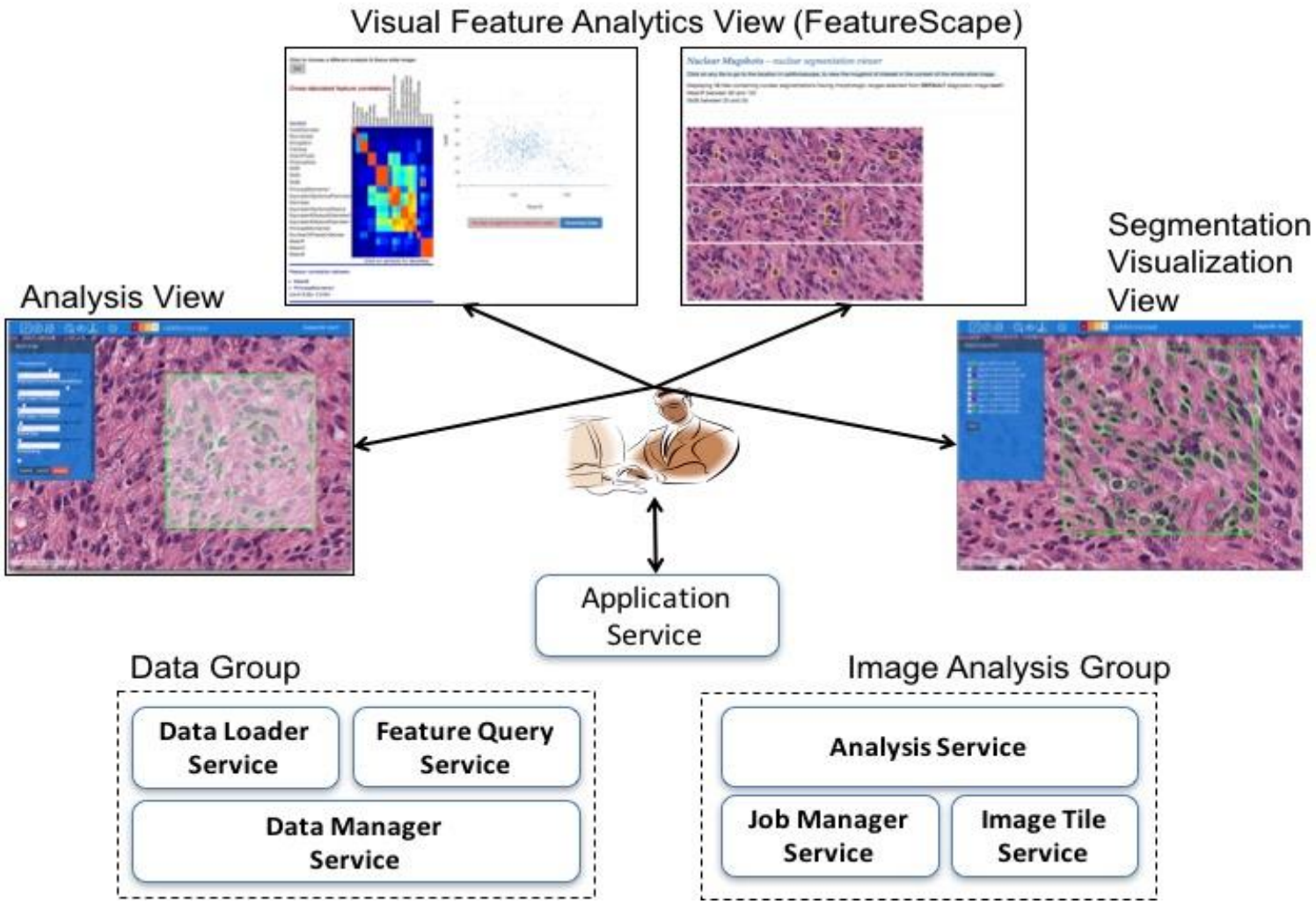
- Algorithm first trained on image patches
- Several cooperating deep learning algorithms generate heat maps
- Heat maps used to generate new predictions
- Companion molecular statistical data analysis pipelines



Training, threshold adjustment, quality control

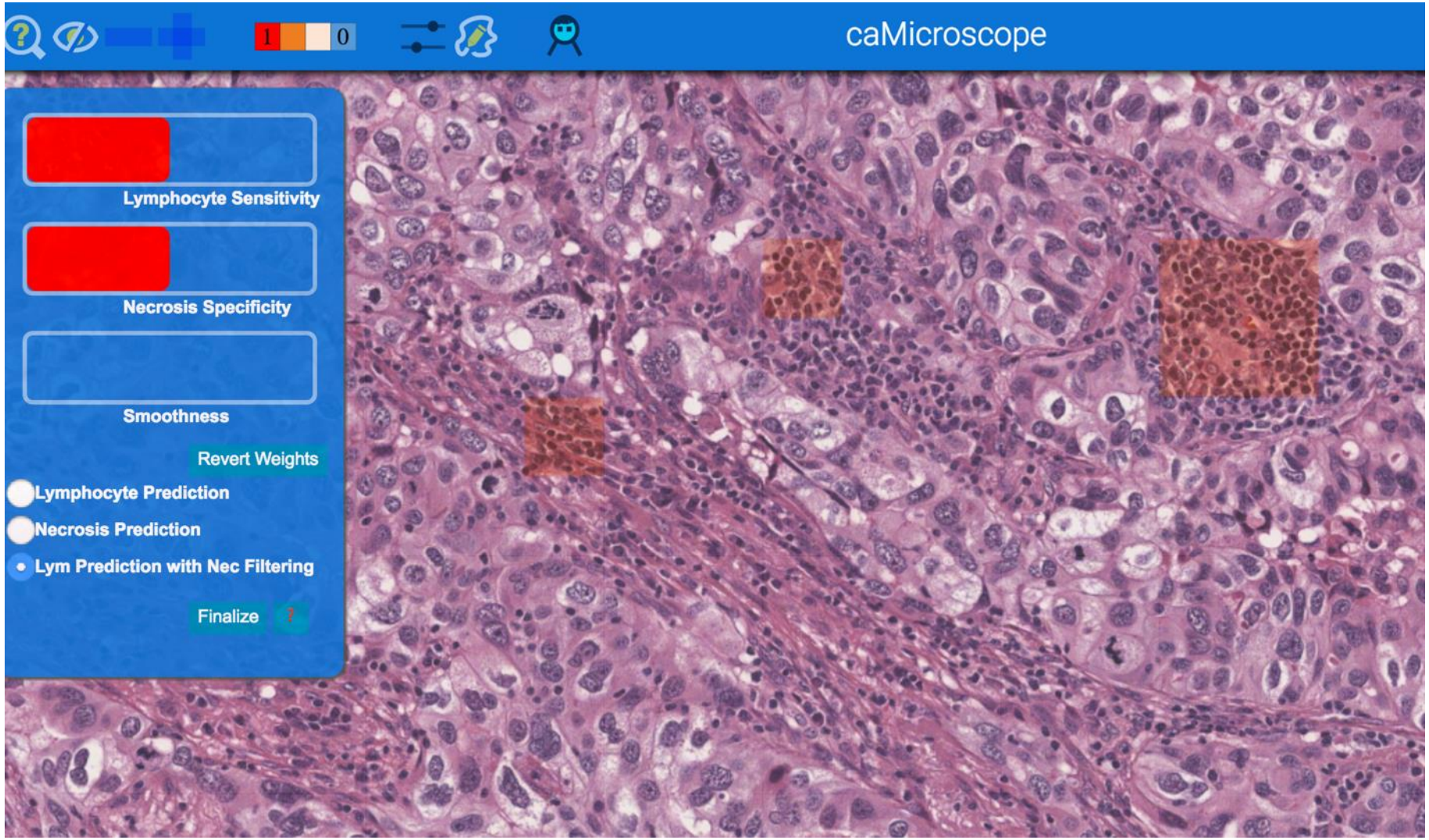


Tools: Quantitative Imaging Pathology - QuIP Tool Set



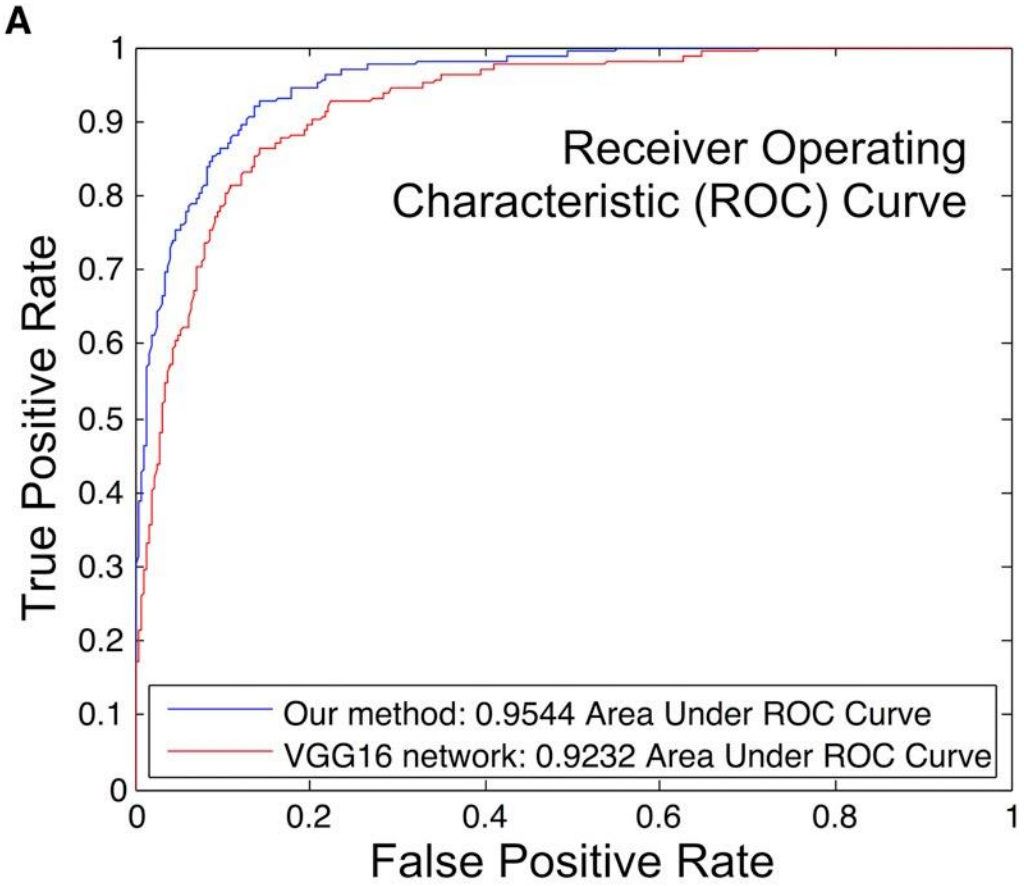
Interactive Deep Learning Training Tool

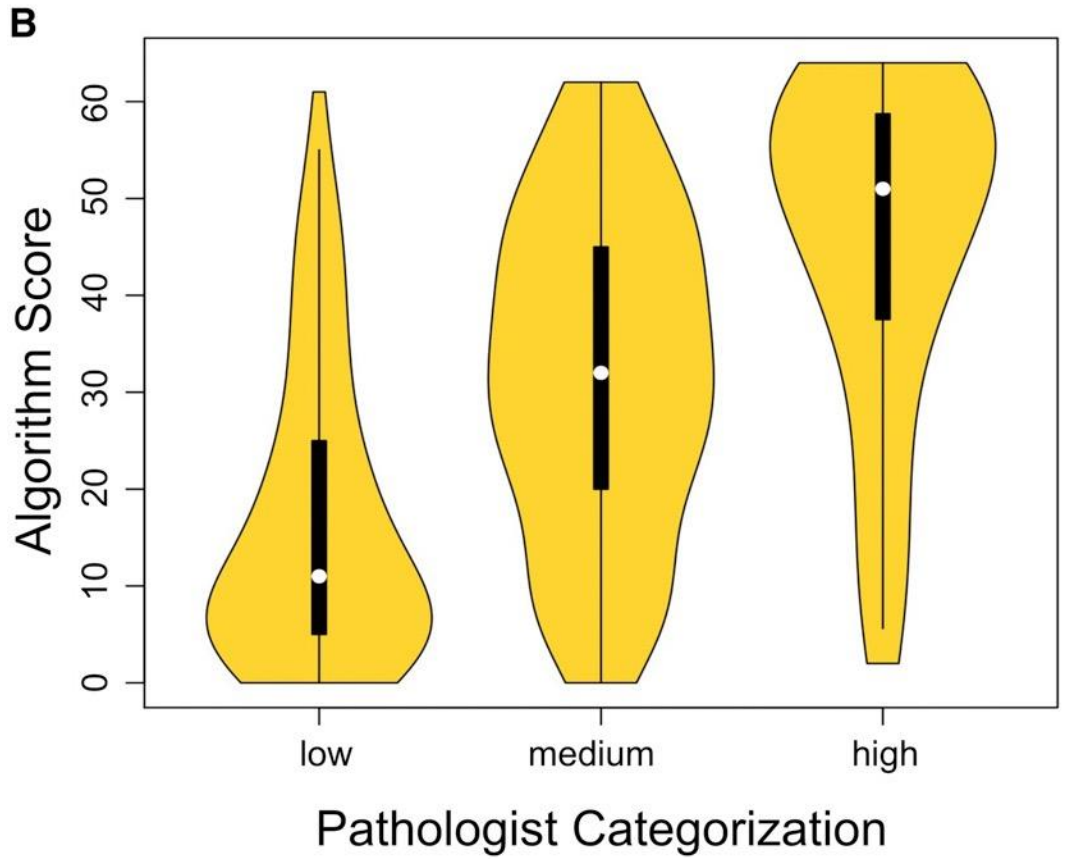
Stony Brook University



Validation – Stratified sampling from 5K whole slide images

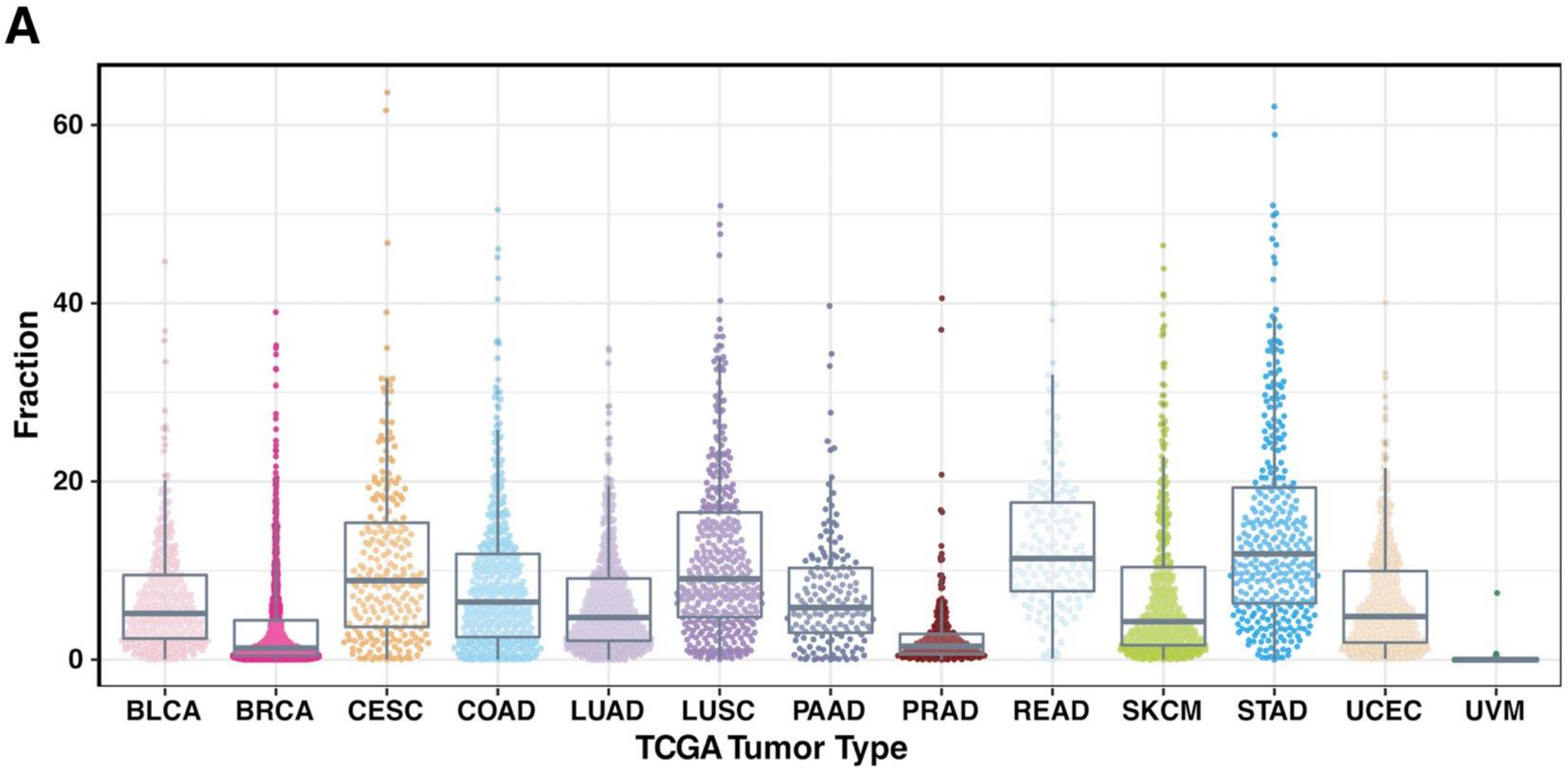
Arvind Rao, expert in spatial biostatistics (U Michigan)





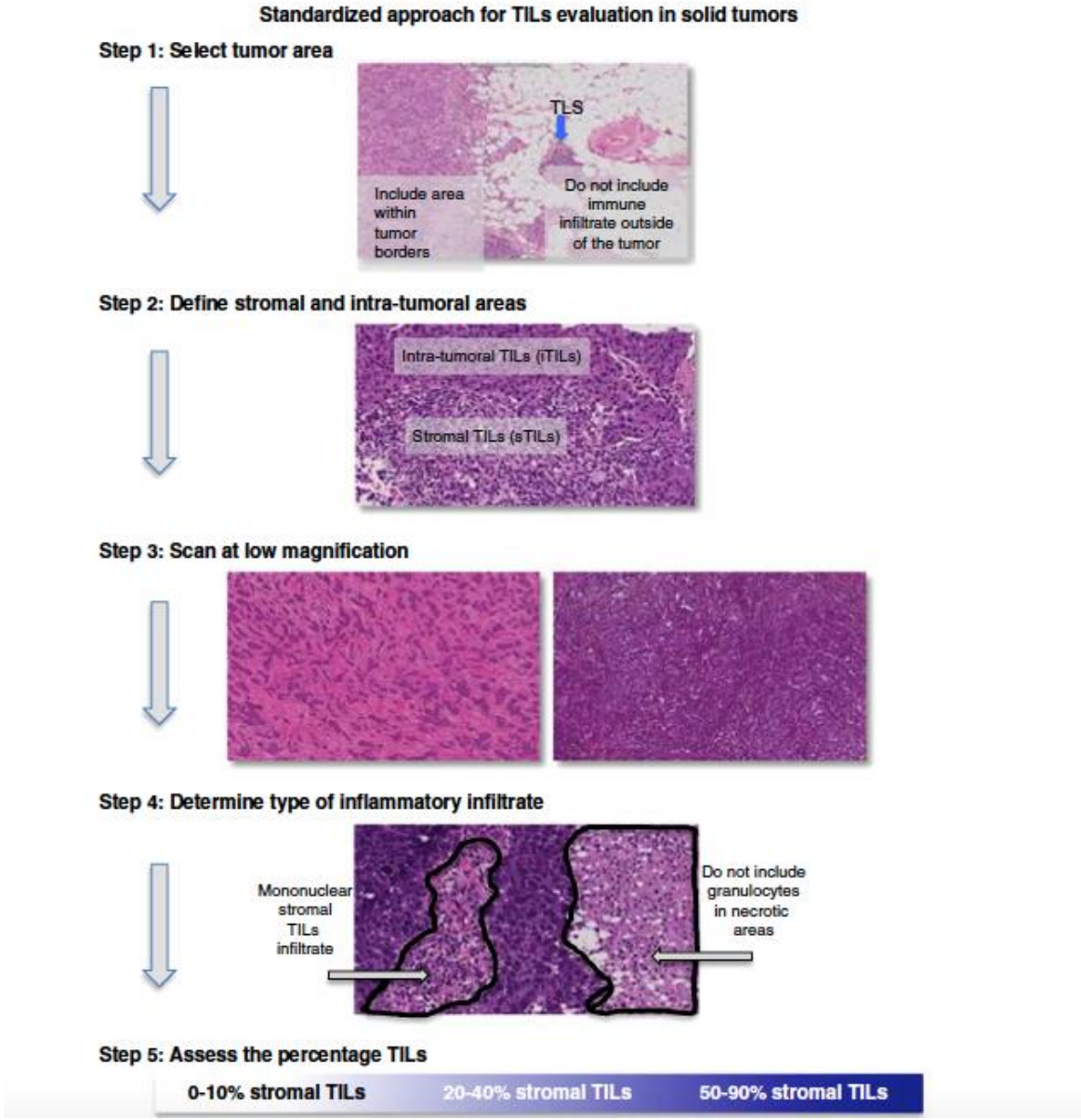
Quantitative Assessment of TIL Fractions

Stony Brook University

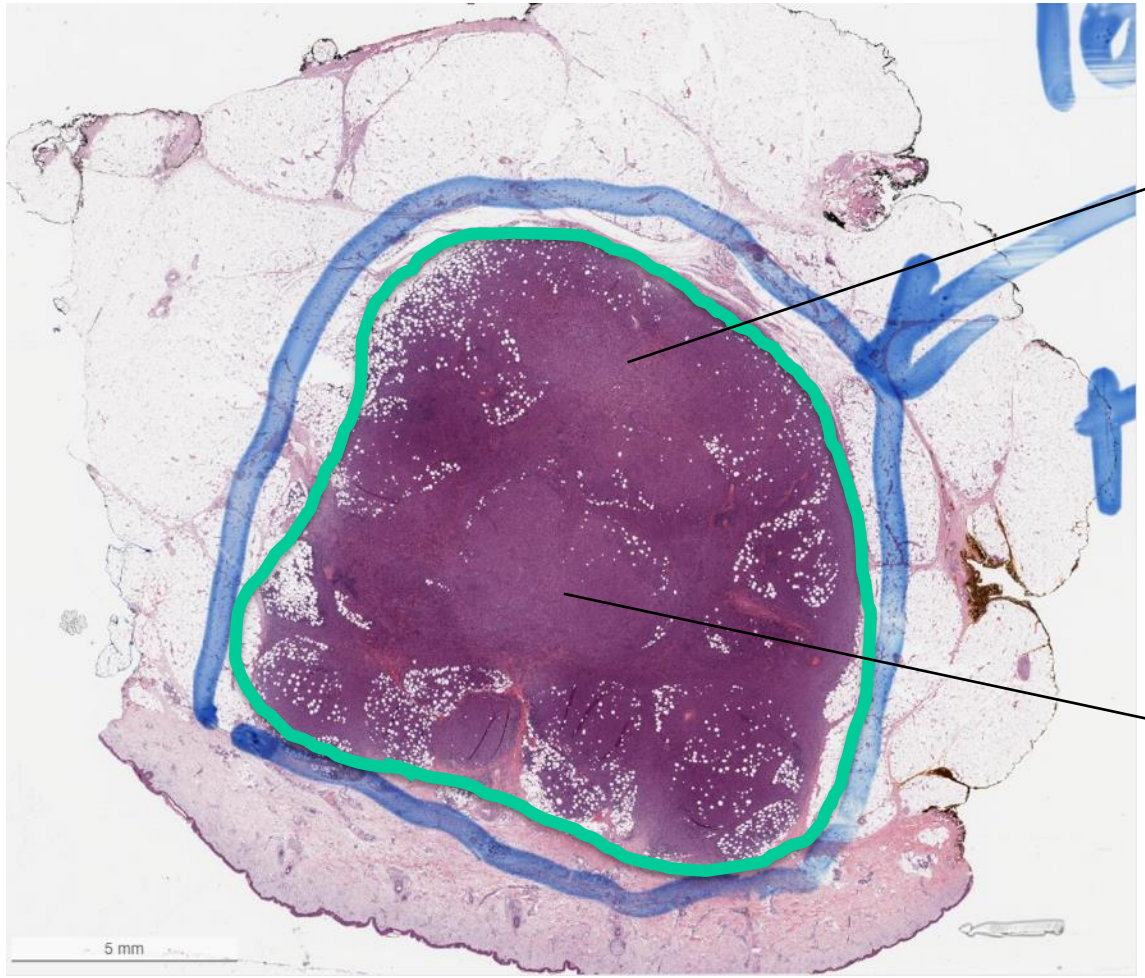


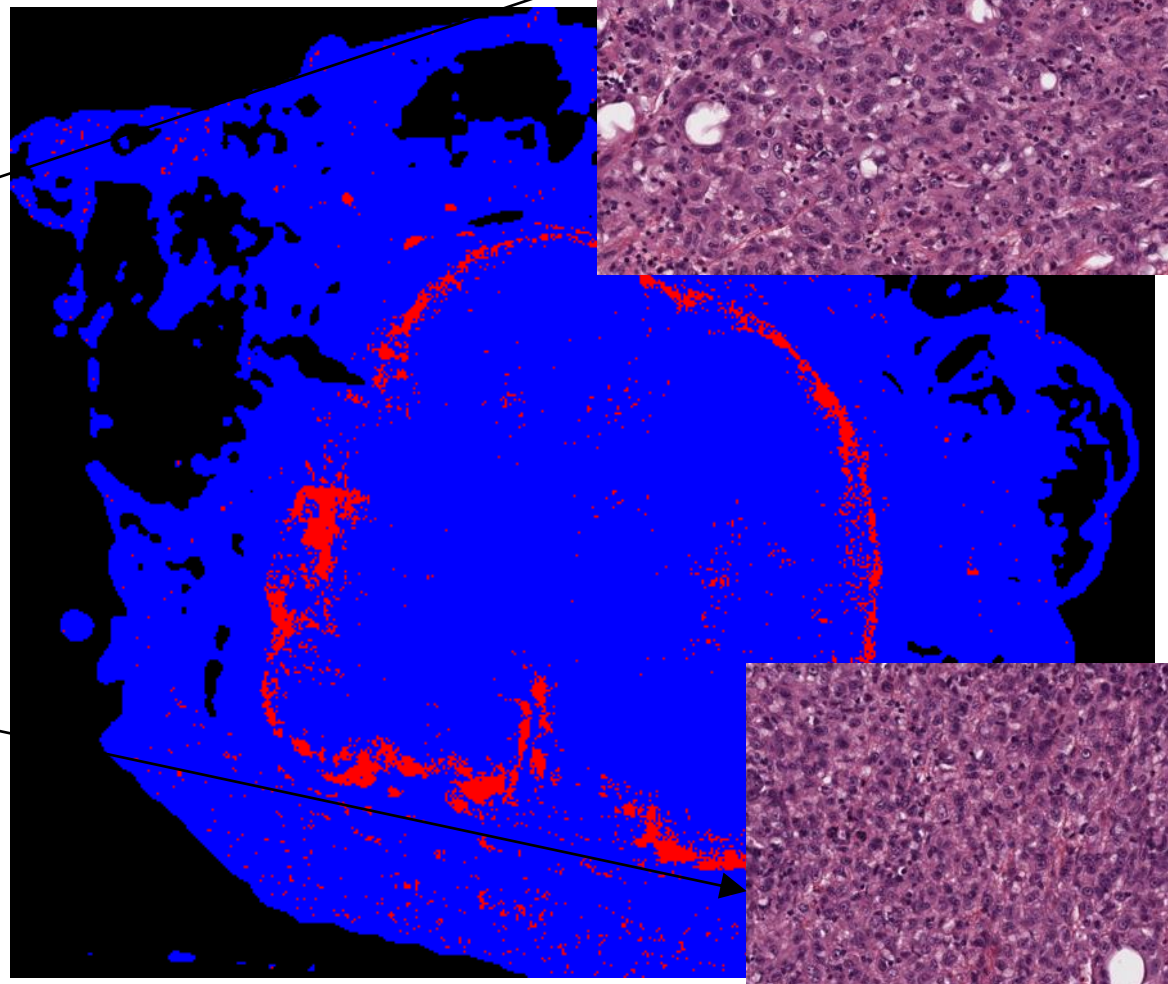
Characterization of TIL Pattern and Relationship to Molecular Immune Subtype

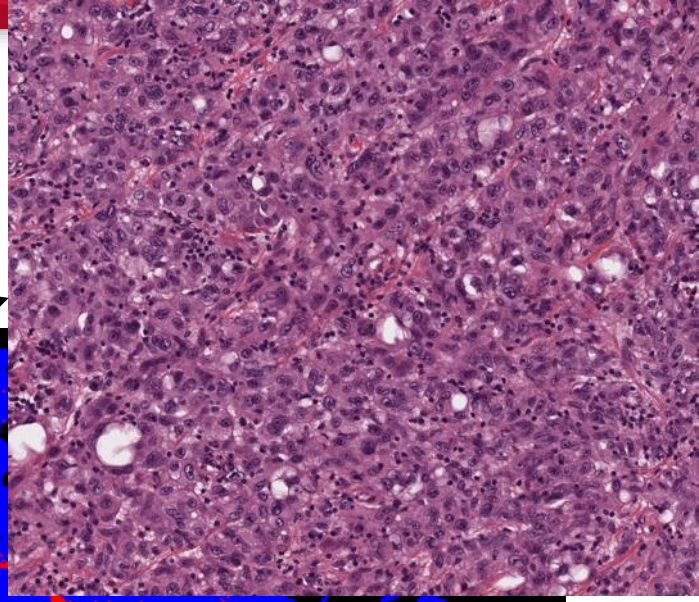
- The ***pattern*** of immune infiltrate
- Division of immune infiltrate between different compartments
- Does it surround tumor region? Present in tumor, invasive margin?
- Assessing Tumor-infiltrating Lymphocytes in Solid Tumors:
- A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group – part 1 and 2 - Adv Anat Pathol Volume 24, Number 5, September 2017 (figure to right from that reference)



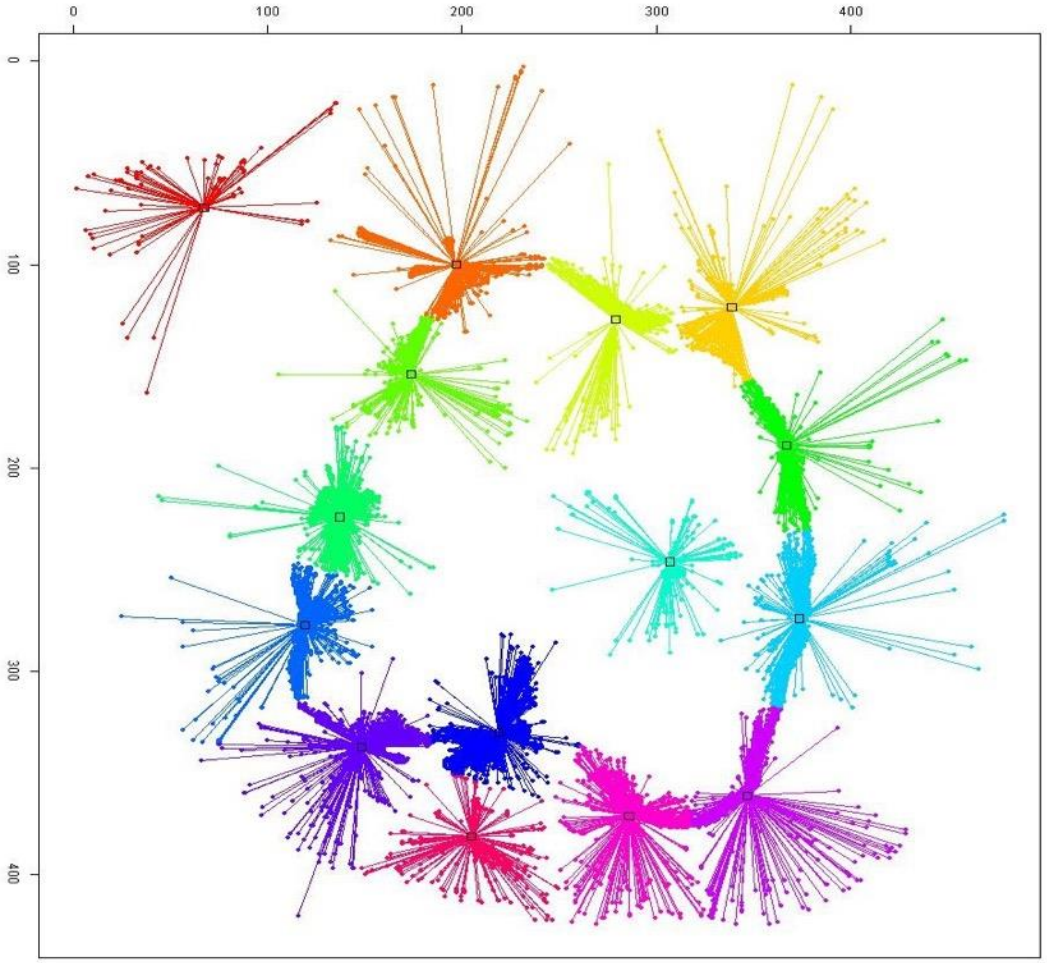
SKCM TCGA-D3-A2JF-06Z-00-DX1

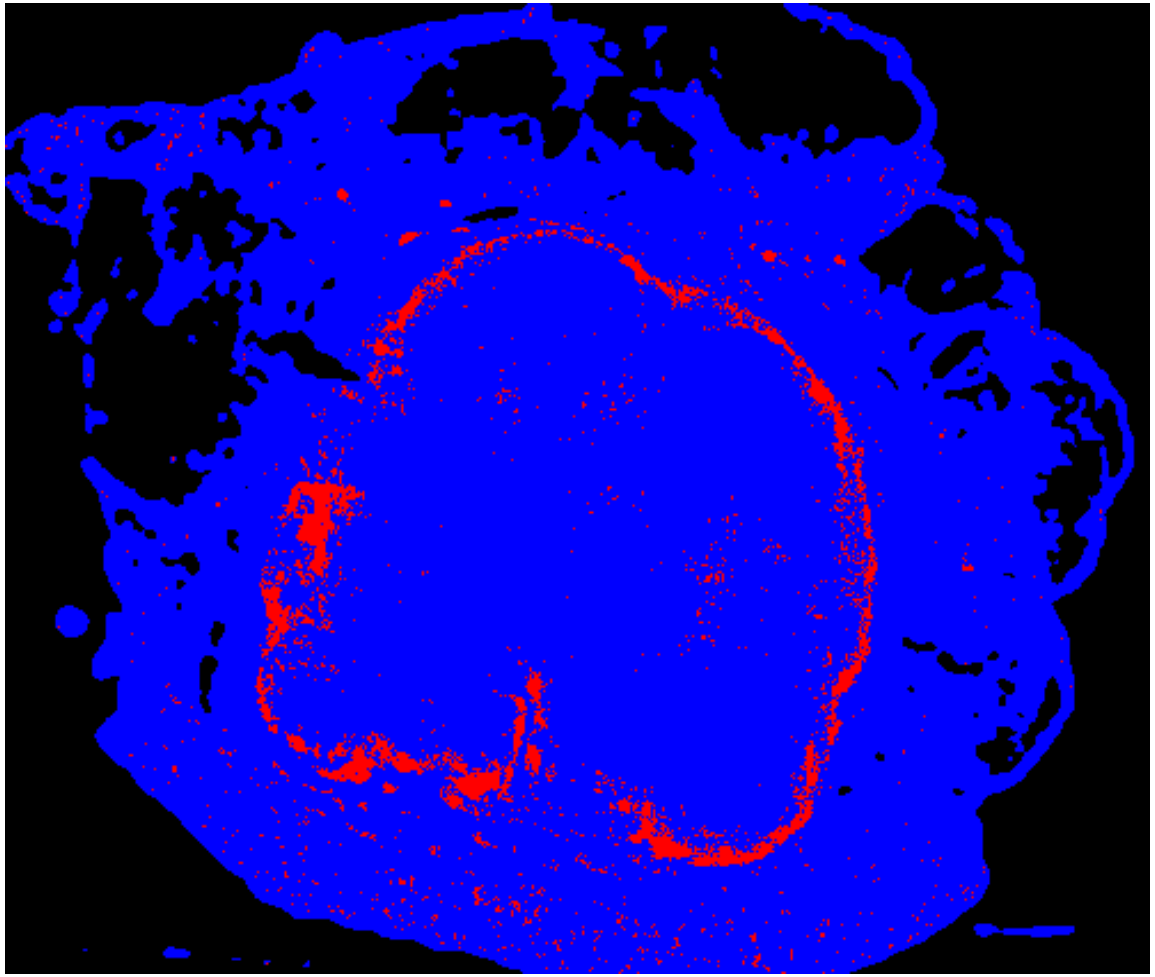




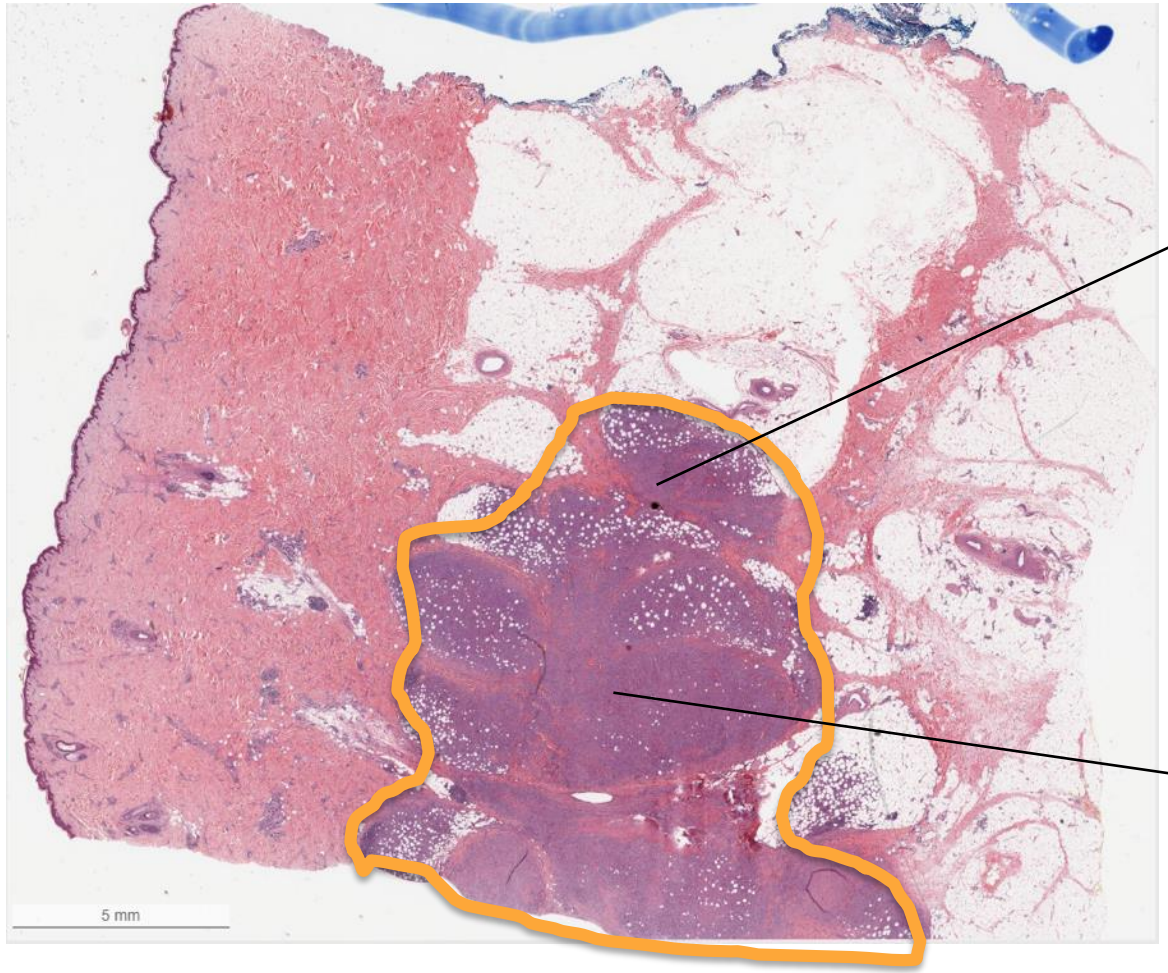


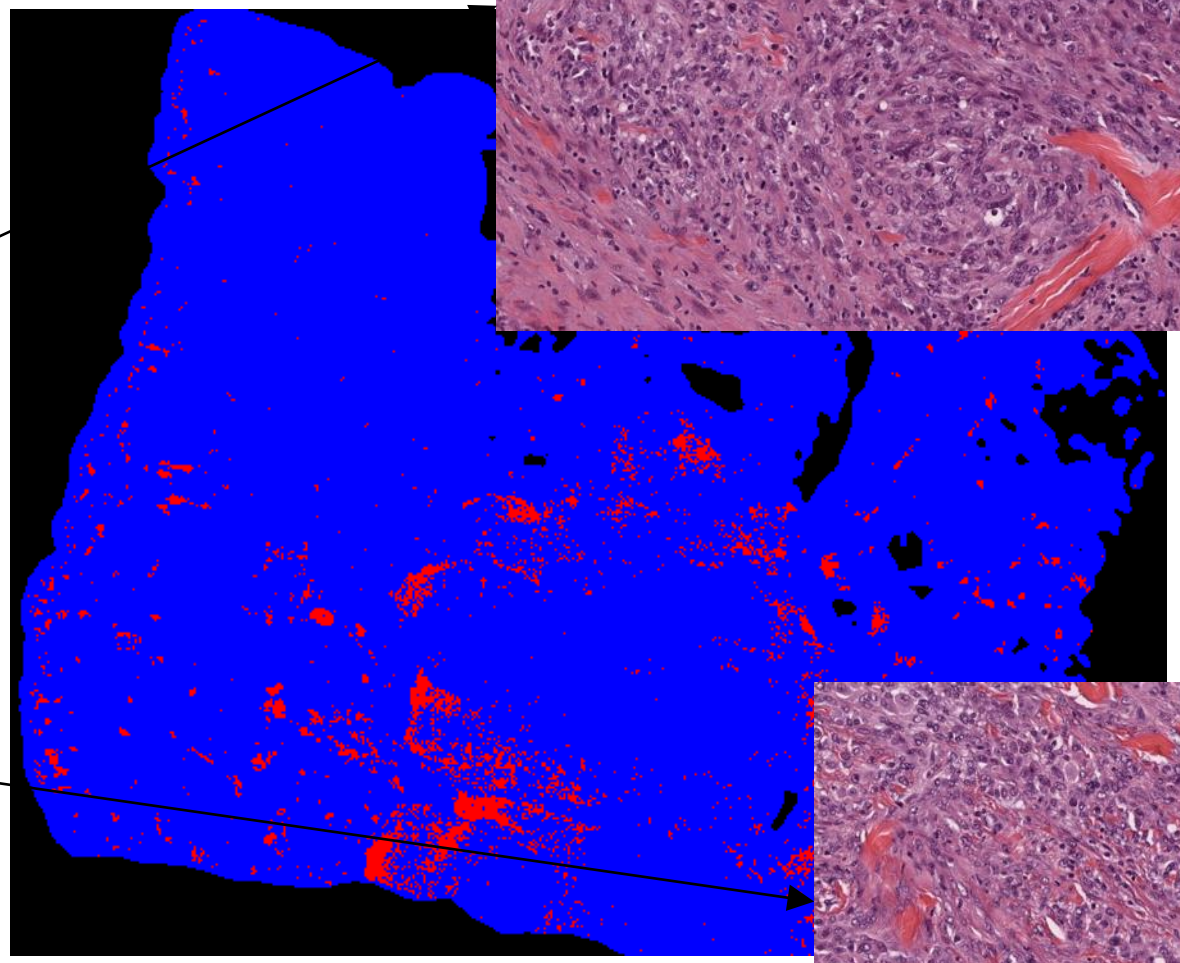
SKCM TCGA-D3-A2JF-06Z-00-DX1

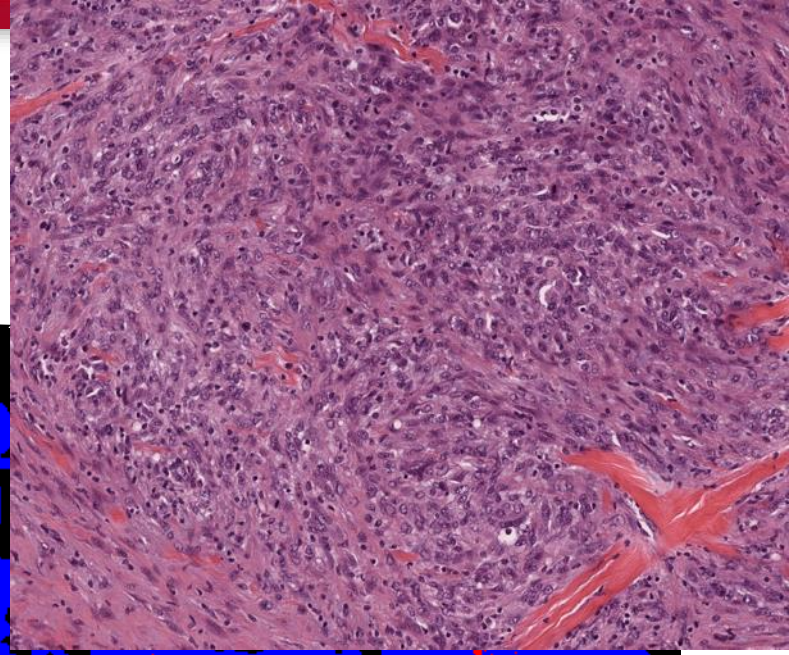


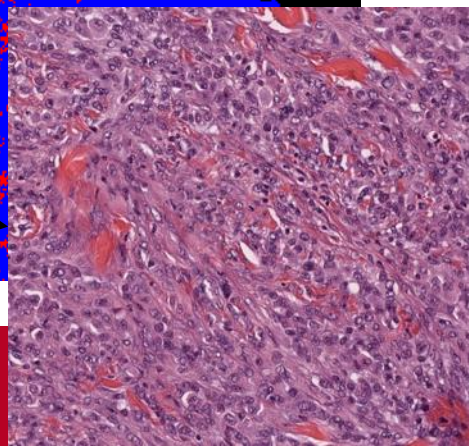


SKCM TCGA-D3-A2JA-06Z-00-DX1

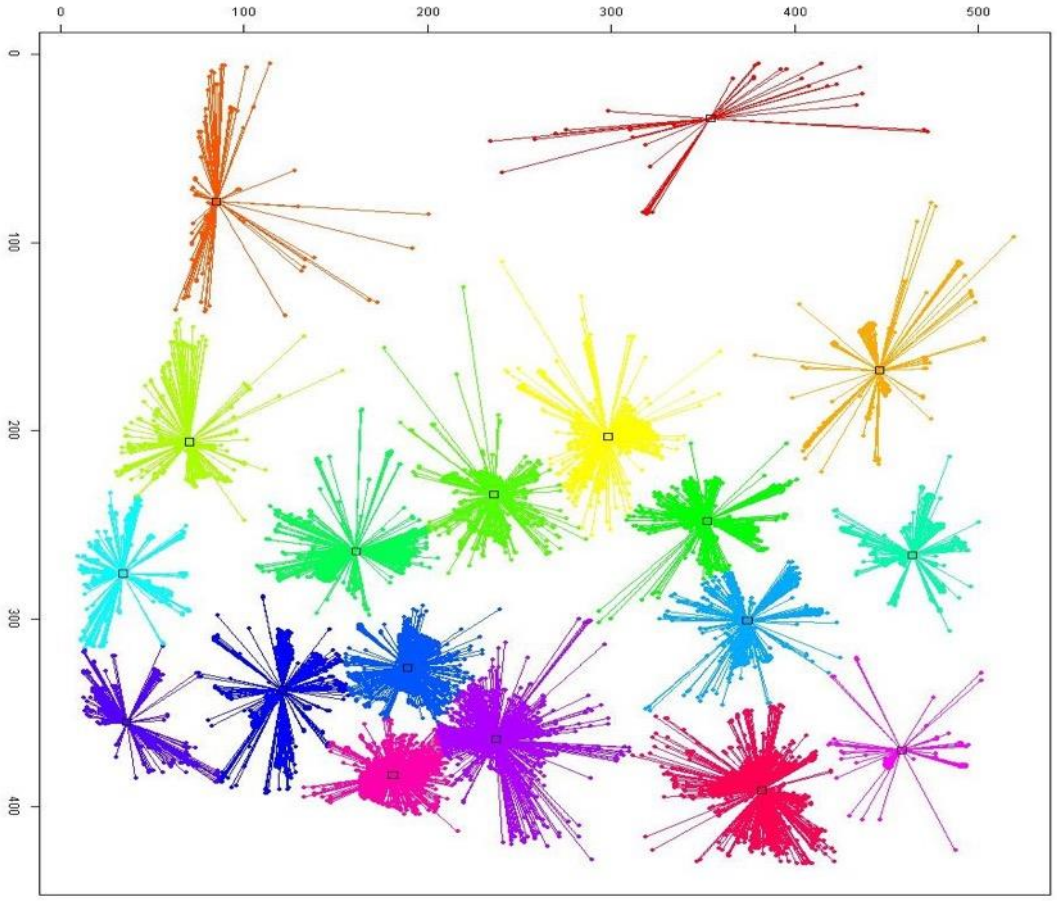


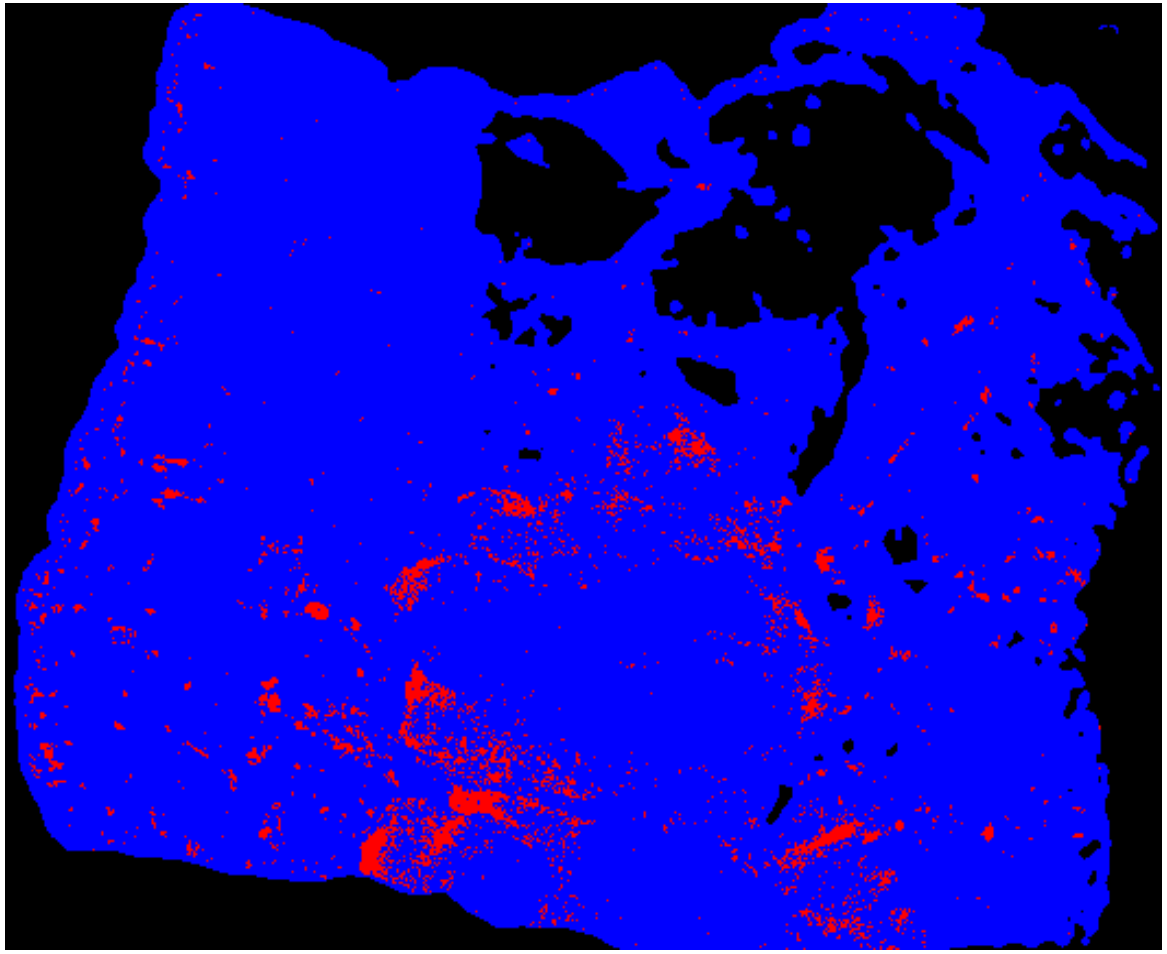






SKCM TCGA-D3-A2JA-06Z-00-DX1





TIL Pattern Descriptions

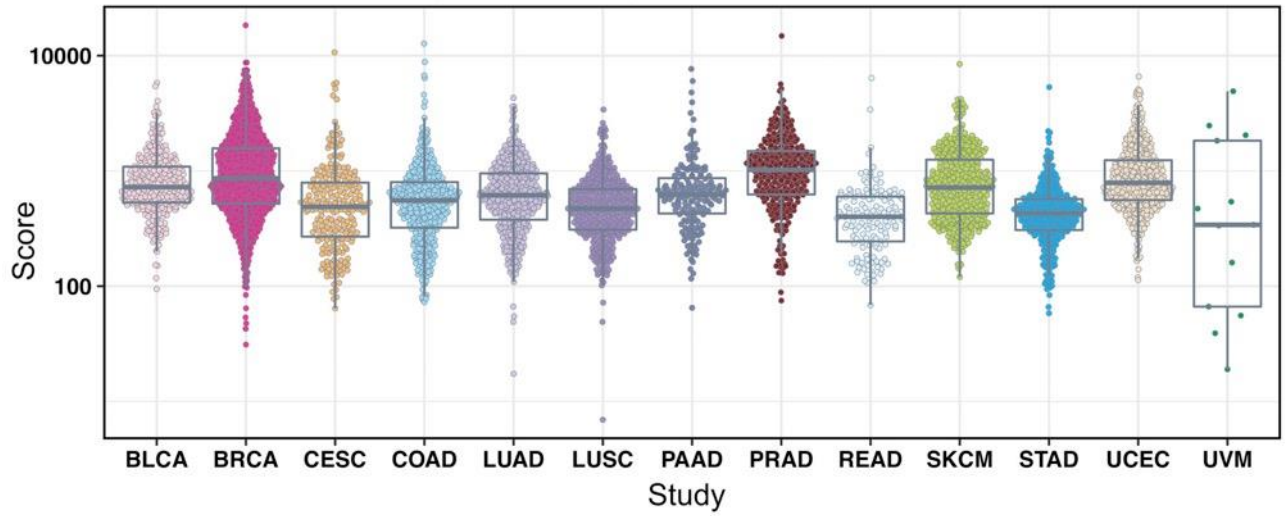
Qualitative (Alex Lazar, Raj Gupta)

- **"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

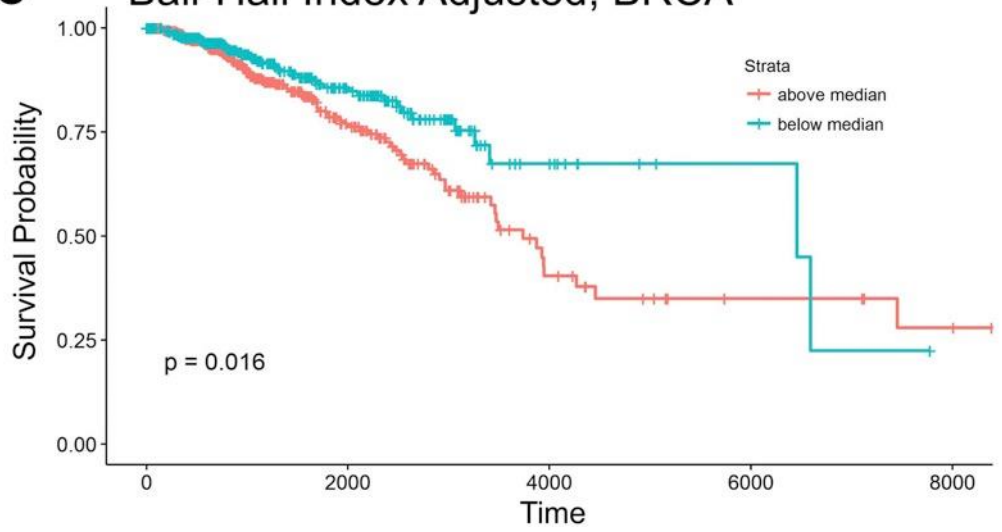
A



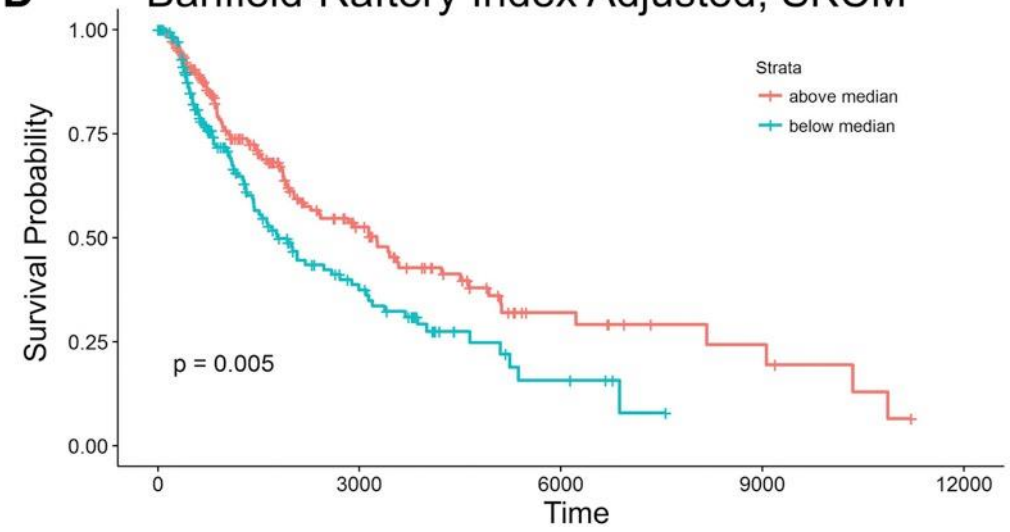
B

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



D



TCGA Pan Cancer Atlas – Immune Landscape of Cancer

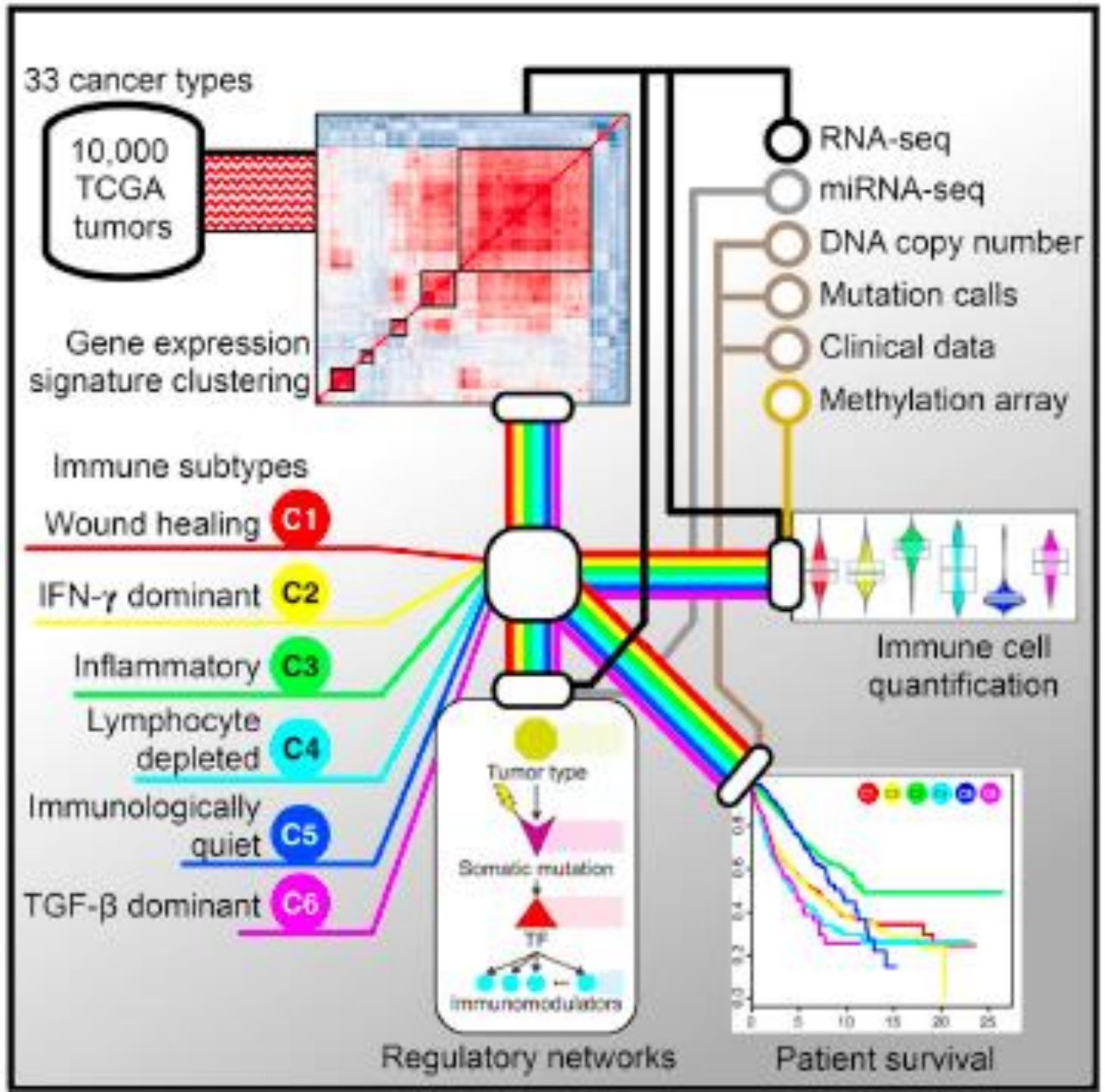
Stony Brook University

Immunity

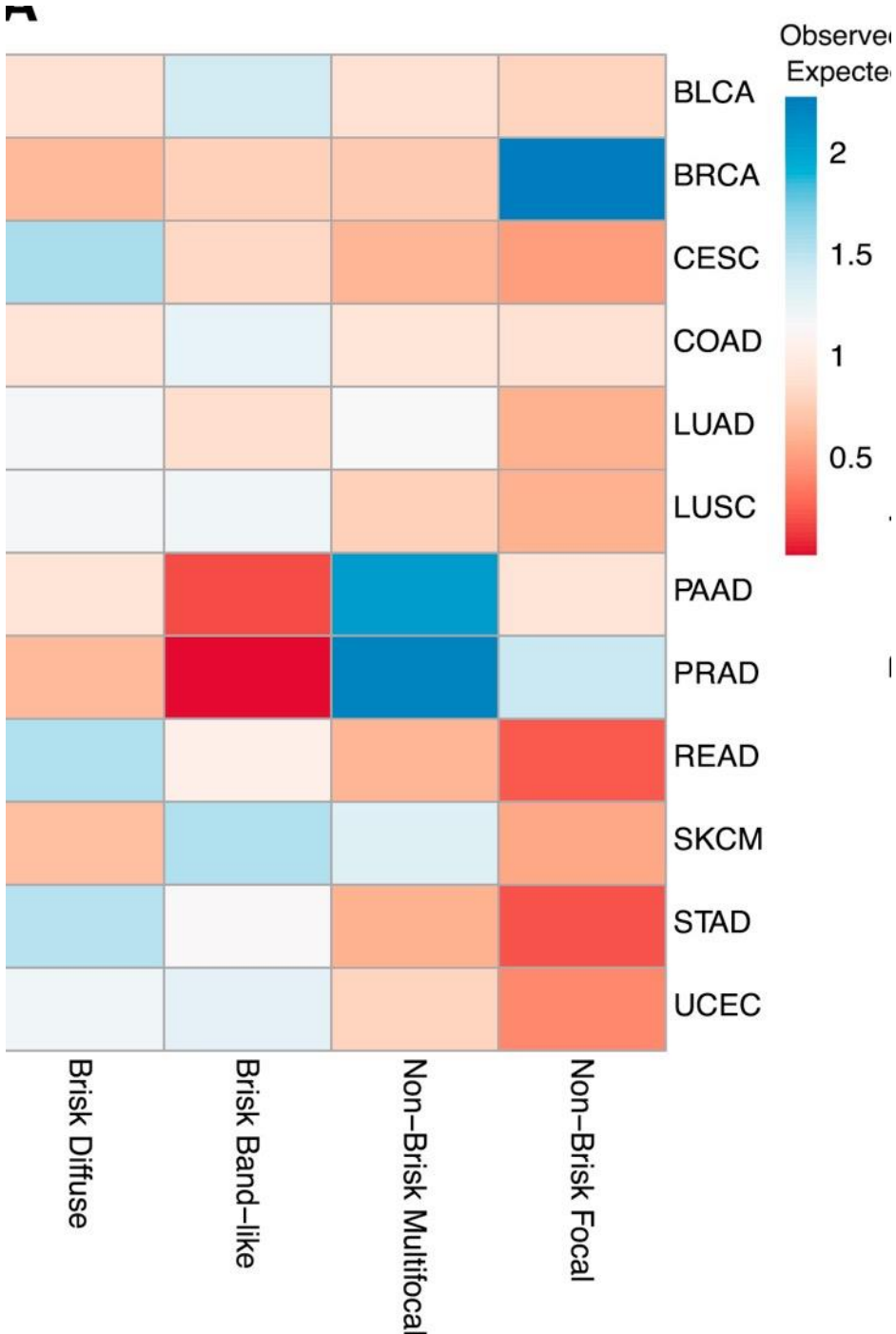
The Immune Landscape of Cancer

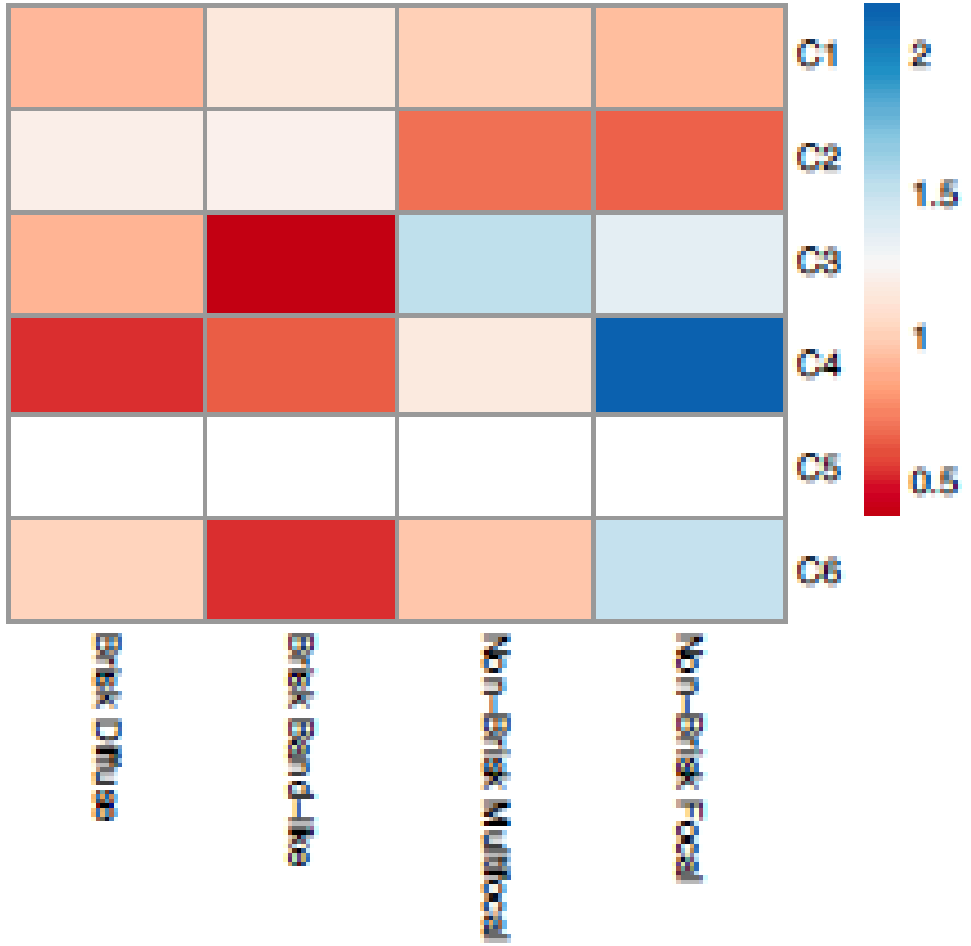
- Six identified immune subtypes span cancer tissue types and molecular subtypes
- Immune subtypes differ by somatic aberrations, microenvironment, and survival
- Multiple control modalities of molecular networks affect tumor-immune interactions
- These analyses serve as a resource for exploring immunogenicity across cancer types

http://www.cell.com/immunity/fulltext/S1074-7613(18)30121-3



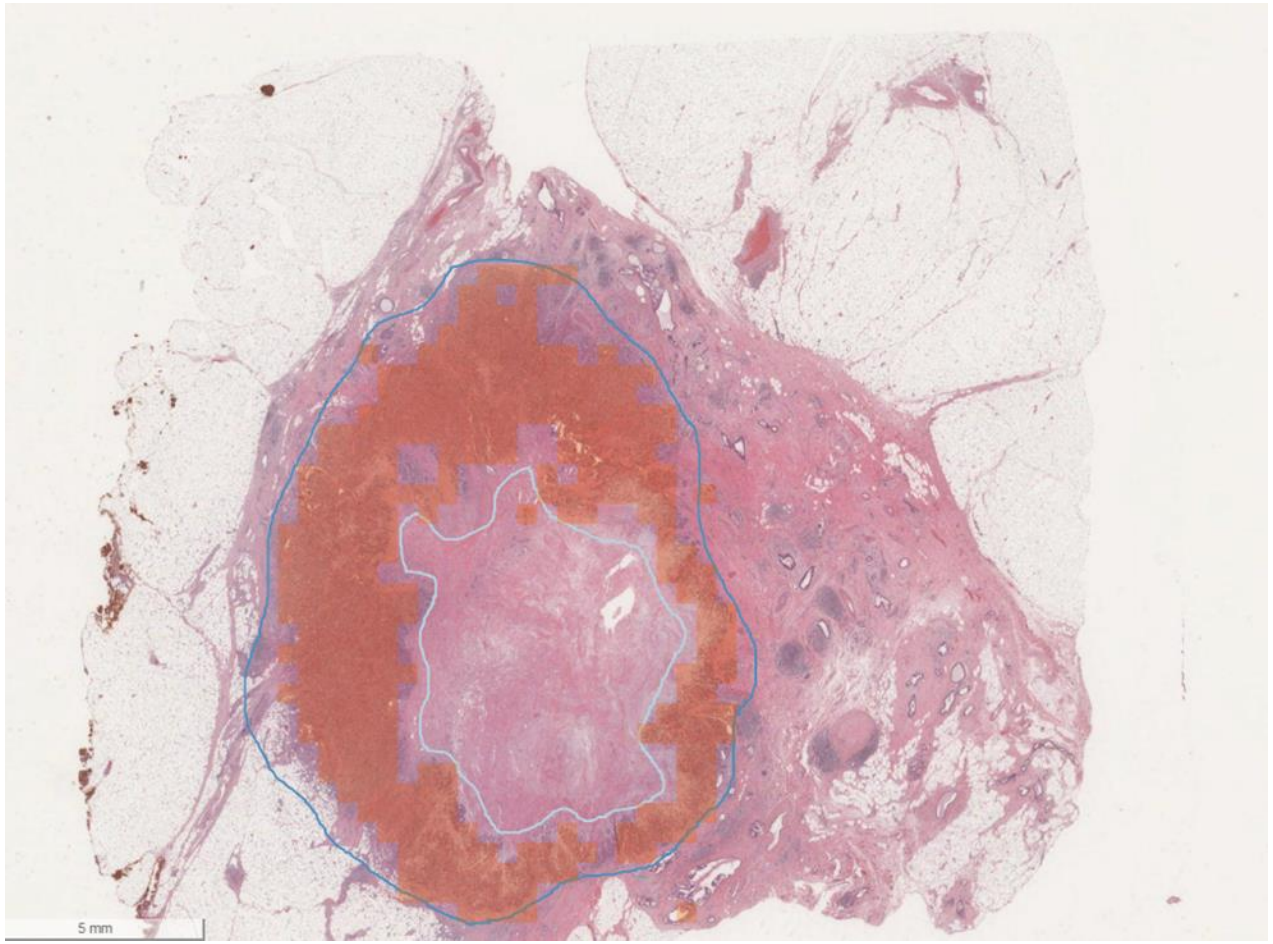
Spatial Patterns vs TCGA Tumor, Molecular Subtypes





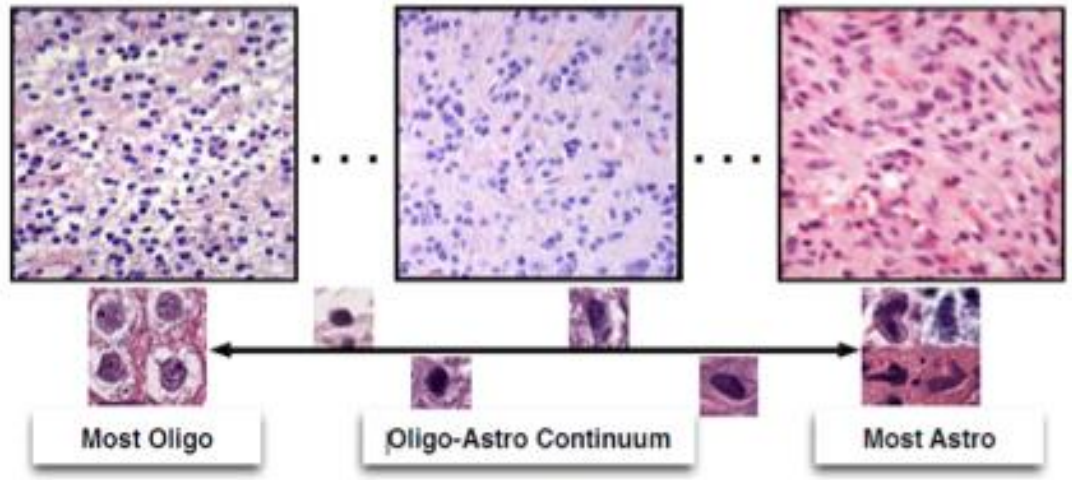
New Results: CNN – Tumor Segmentation

Stony Brook University

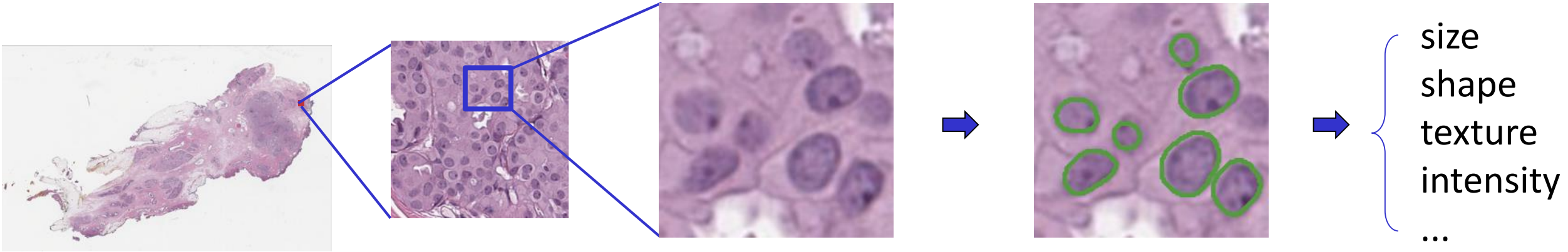


Pathomics – Nuclear Features

- Cell morphology and architectural patterns of tumor growth are critical in cancer diagnosis
- The nucleus-level features, such as size, shape and intensity, are used in cancer diagnosis and classification of cancer subtypes



Nuclear Pathomics Features



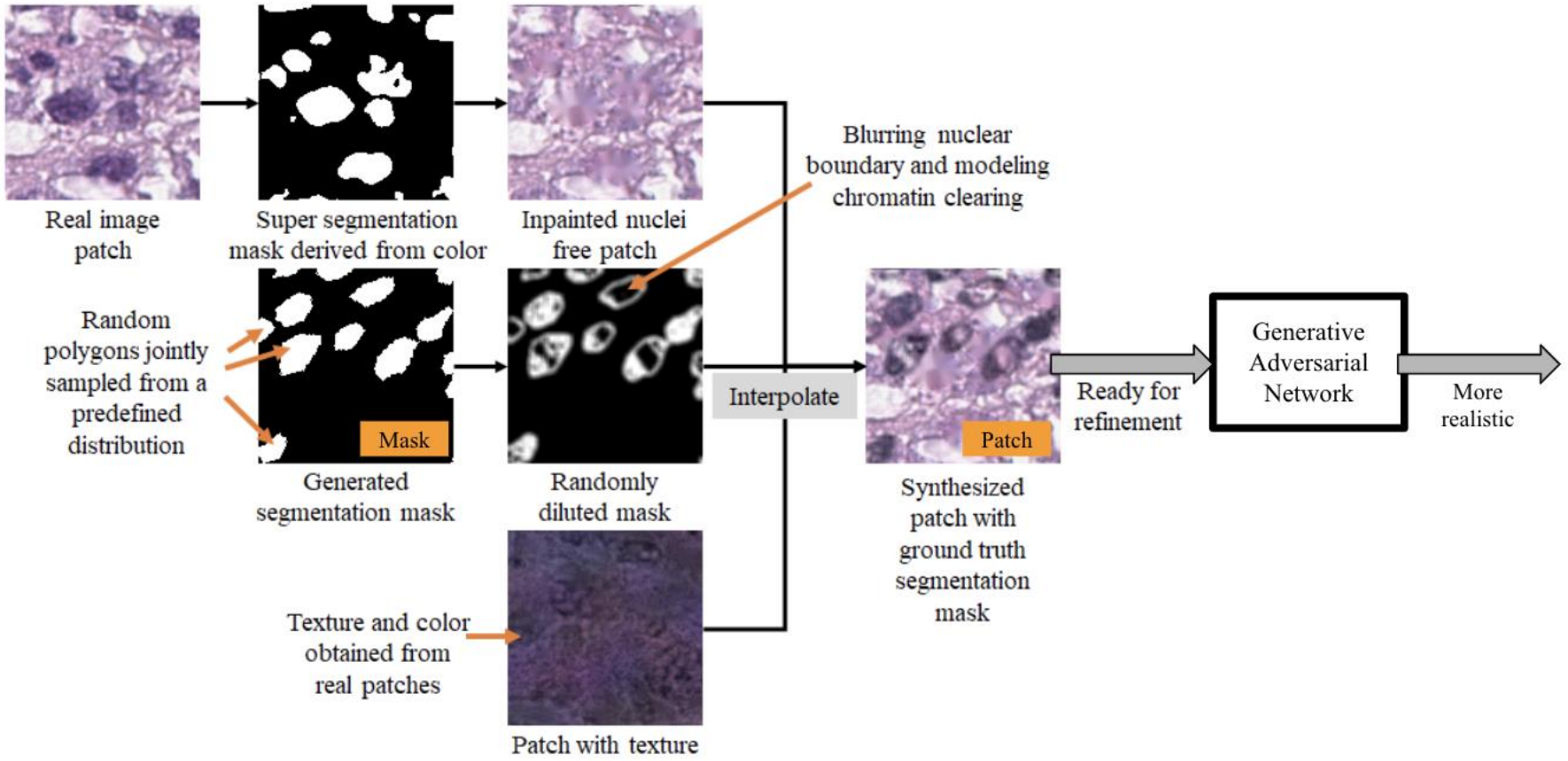
- H&E stained tissue slides: Nuclei are colored blue/purple
- Nucleus segmentation is applied to delineate the boundary of the nuclei
- Nucleus-level features are extracted from segmented object for downstream quantitative analyses

Nuclear Segmentation/Labeling Methods Development

Stony Brook University

- Baseline – traditional numerical methods: level set/mean shift
- Convolutional Neural Network nuclear segmentation algorithms
- GAN based CNN nuclear segmentation algorithm designed to minimize training requirements
- Deep learning based nuclear classification methods

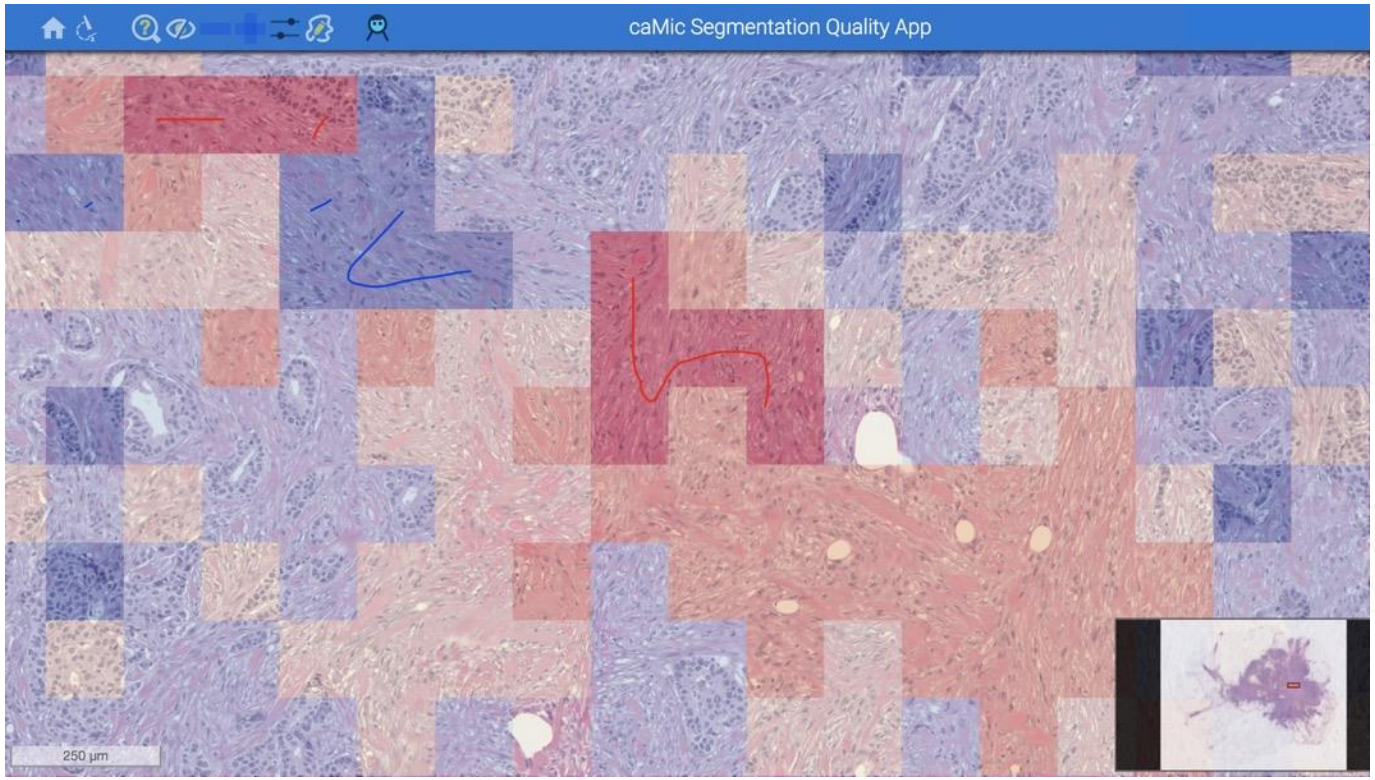
Deep Learning: Synthetic Tissue/GAN based Nuclear Segmentation Algorithm

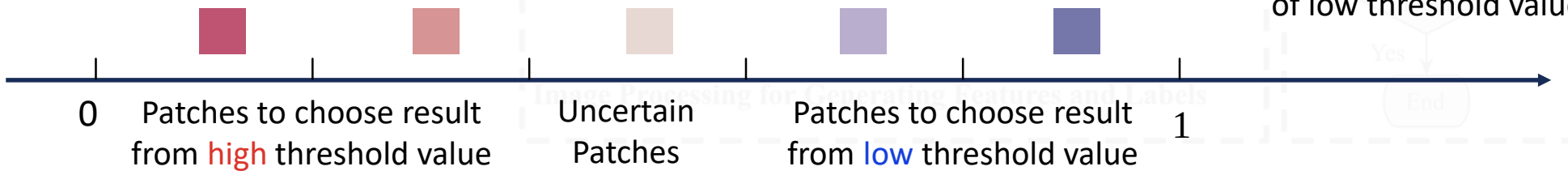


Using Machine Learning to Critique Segmentation Results

- Automatically select low & high gain values for each tissue region
- Investigate machine learning algorithms:
  - Random Forests
  - Support Vector Machine
  - Convolutional Neural Network
- Investigate impact of active learning process
- **PhD Thesis – Si Wen co supervised by Joel Saltz and Tahsin Kurc**

ACTIVE Learning





Label the uncertain patches with high threshold value

Label the uncertain patches with low threshold value

Label Selected Uncertain Patches

the predicted probability of low threshold value

Active Learning Iterations

Consortia

- NCI Quantitative Imaging for Pathology (QuIP): Stony Brook, Emory, MD Anderson, Institute for Systems Biology, Oak Ridge
- NCI SEER Pathology: Stony Brook, Emory, Rutgers, University of Kentucky (three Cancer registries)
- Cancer Imaging Archive: Arkansas, Stony Brook, Emory (Stony Brook leads Pathology component)
- Virtual Tissue Repository: Led by NCI SEER; Stony Brook, Emory
- TIES Research Network - Integrated Pathology text and imaging: Pittsburgh, Stony Brook main sites, 6+ other sites (Stony Brook leads digital Pathology)

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

Stony Brook University

- **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.

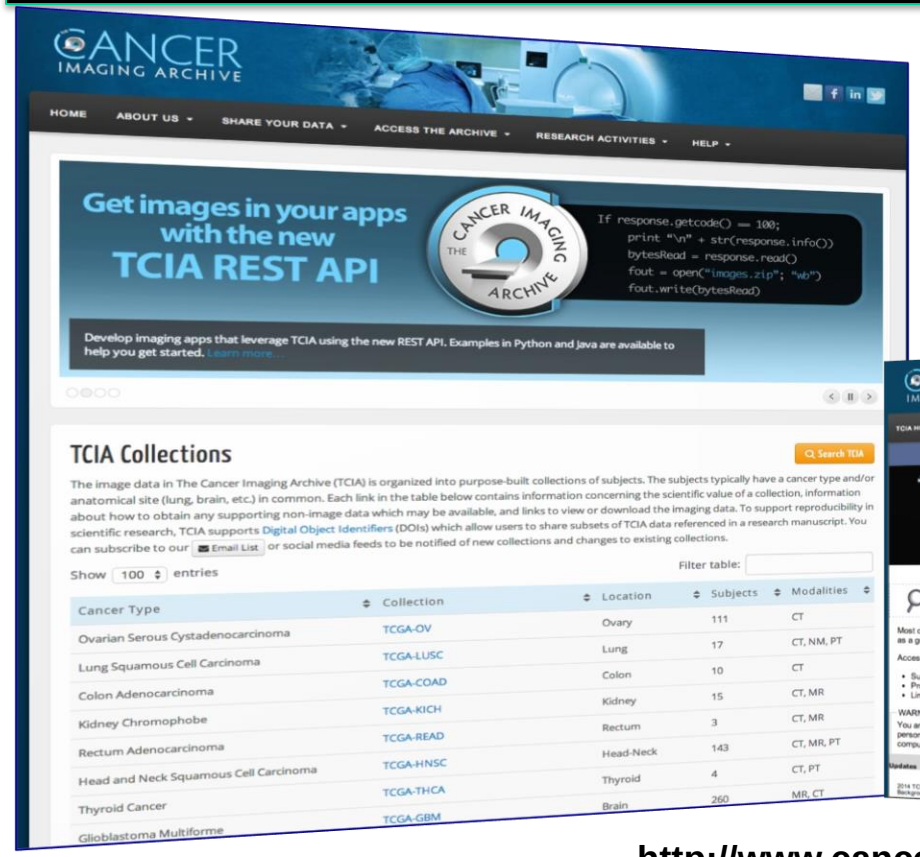
Methods and tools for integrating pathomics data into cancer registries
Saltz, Sharma, Foran and Durban

- Enhance SEER registry data with machine learning based classifications and quantitative pathomics feature sets.
- The New Jersey State Cancer Registry, Georgia and Kentucky State Cancer Registries
- Prostate Cancer, Lymphoma and NSCLC
- Repository of high-quality digitized pathology images for subjects whose data is being collected by the registries.
- Extract computational features and establish deep linkages with registry data, thus enabling the creation of information-rich, population cohorts containing objective imaging and clinical attributes

Cancer Imaging Archive – Integration of Pathology and Radiology for Community Clinical Studies

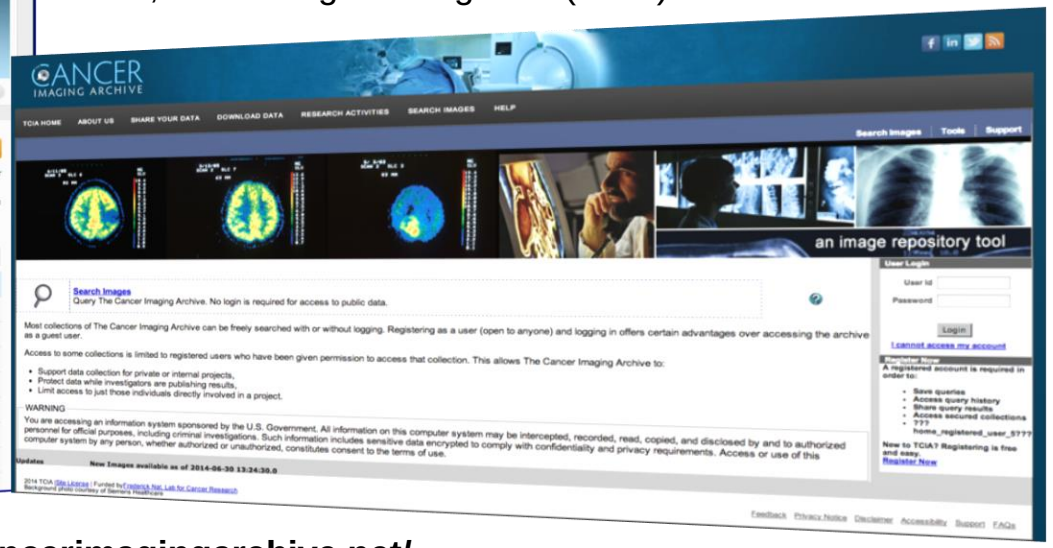
Stony Brook University

THE CANCER IMAGING ARCHIVE



TCIA encourages and supports the cancer imaging open science community by hosting and managing Findable Accessible, Interoperable, and Reusable (FAIR) images and related data.

Clark, et al. J Digital Imag 26.6 (2013): 1045-1057.



http://www.cancerimagingarchive.net/

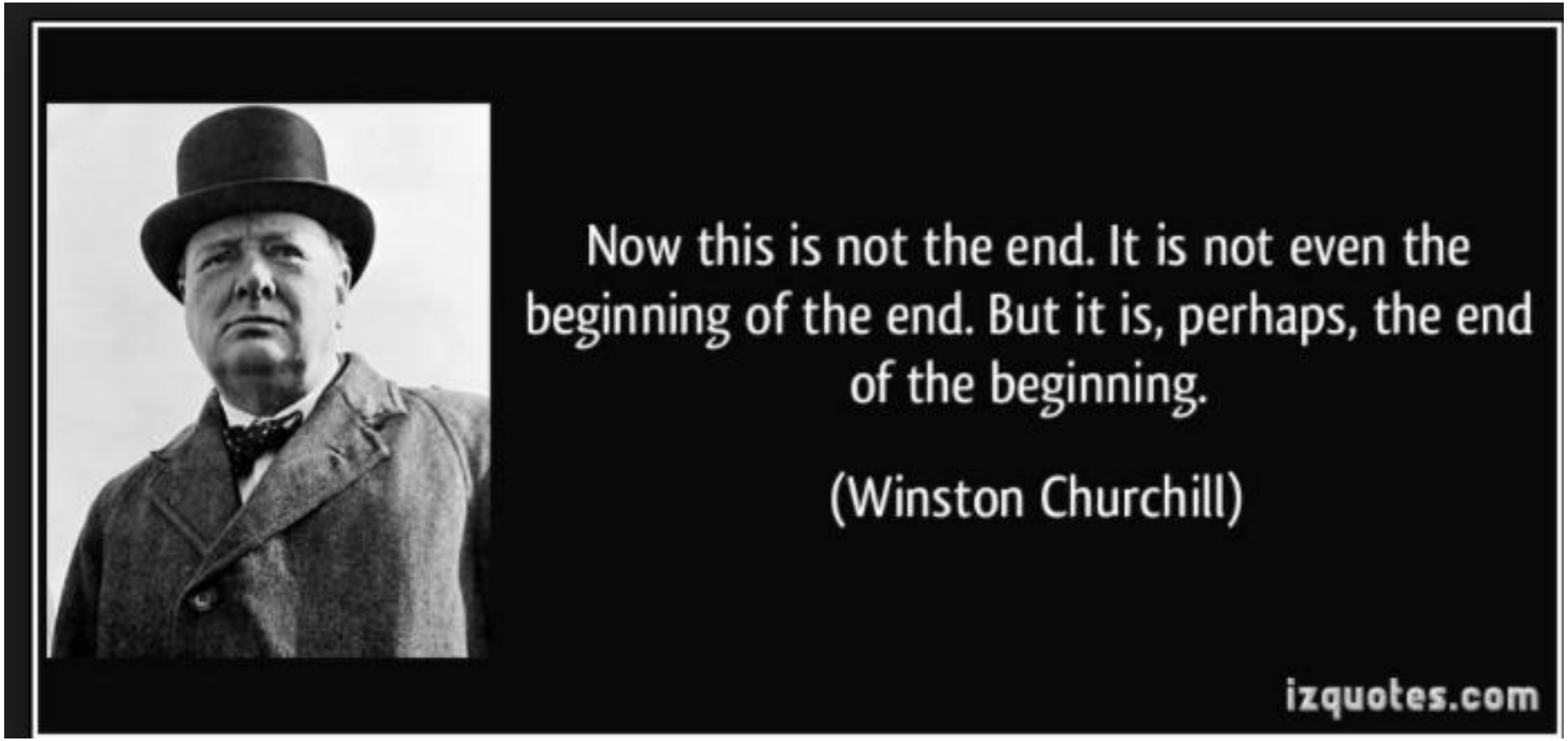
TCIA sustainment and scalability
Platforms for quantitative imaging informatics in precision medicine
Prior, Saltz, Sharma -- U24CA215109-01

Stony Brook University

- Identify quantitative imaging phenotypes across scale through the use of Radiomic/Pathomic analyses
- Well-curated data for algorithm testing and validation.
- Integrative Radiology/Pathology Image-Omics studies
- Extend TCIA to support its rapidly growing user community and continue to promote research reproducibility and data reuse in cancer precision medical research.

Thanks!

Stony Brook University



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
Raj Gupta

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, U24CA215109, NCIP/Leidos 14X138 and HHSN261200800001E, UG3CA225021-01 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.