Stony Brook University

TOOLS TO ANALYZE MORPHOLOGY AND SPATIALLY MAPPED MOLECULAR DATA

Tahsin Kurc
Department of Biomedical Informatics
Stony Brook University

Research and Development Team

Stony Brook University
Joel Saltz
Tahsin Kurc
Erich Bremer
Jonas Almeida
Tammy DiPrima
Feiqiao Wang
Joseph Balsamo
Le Hou
Maozheng Zhao
Han Le
Vu Nguyen
Rajarsi R Gupta
Anne Zhao
Dimitris Samaras

Emory University
Ashish Sharma
Ryan Birmingham
Annie Gu
Whitney Hogg

University of Tennessee Knoxville
Scott Klasky
Dave Pugmire
Jeremy Logan

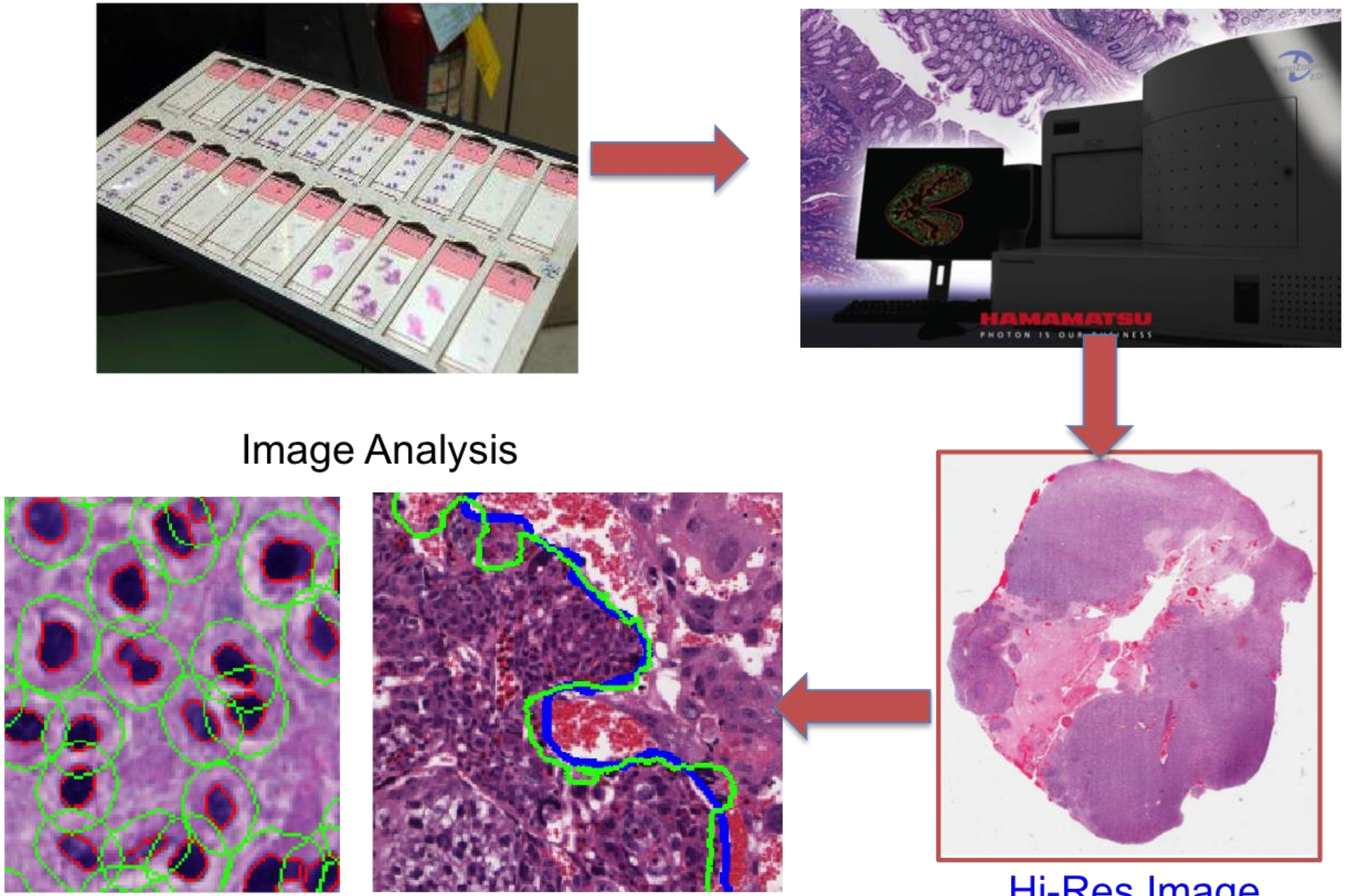
Tools to Analyze Morphology and Spatially Mapped Molecular Data - U24 CA180924

- **Aim 1:** Analysis pipelines for multi-scale, integrative analysis of histopathology images
- **Aim 2:** Database infrastructure to manage and query Pathomics features
- **Aim 3:** Software that targets local machines as well as cloud computing, and high performance computing systems
- **Aim 4:** Develop visualization middleware to relate Pathomics feature and image data and to integrate Pathomics image and "omic" data.
- Combine and deploy tools and methods in an integrated software platform

Using Tissue Slides

- **Traditional Approach**
  - Tissue taken from patients is fixed on a glass slide and stained
  - Pathologist examines the tissue slide under high-power microscope to classify the tissue and render a diagnosis
  - Hard (or impossible) to generate (detailed) quantitative information
  - Labor intensive, not feasible for studies using large numbers of slides
- **Quantitative Analysis/Assessment**
  - Whole slide tissue imaging has progressed significantly in recent years
  - Time required to scan a slide at high-resolution has reduced from multiple hours to several minutes
  - Storage space is getting cheaper

Whole Slide Tissue Imaging



Glass Slides

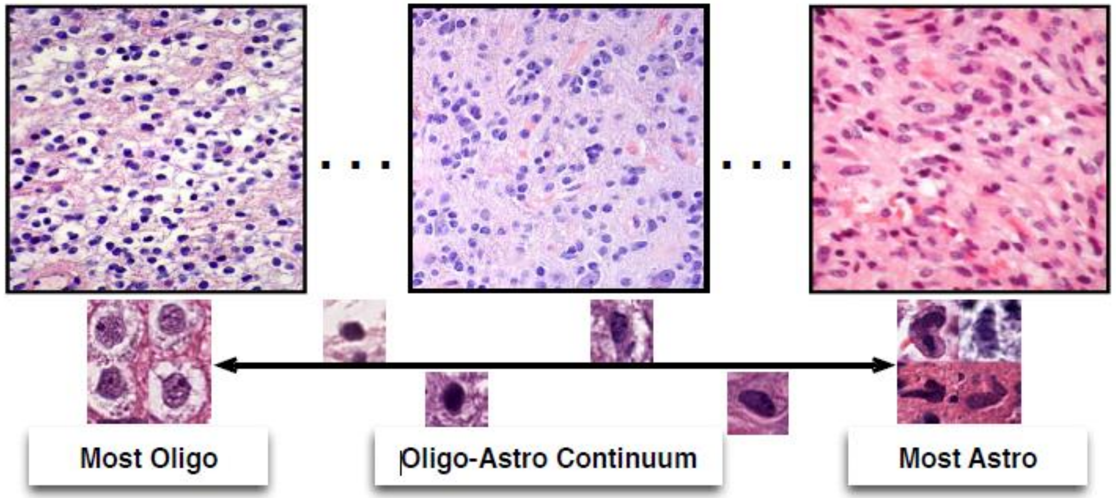
Whole Slide Imaging

Image Analysis

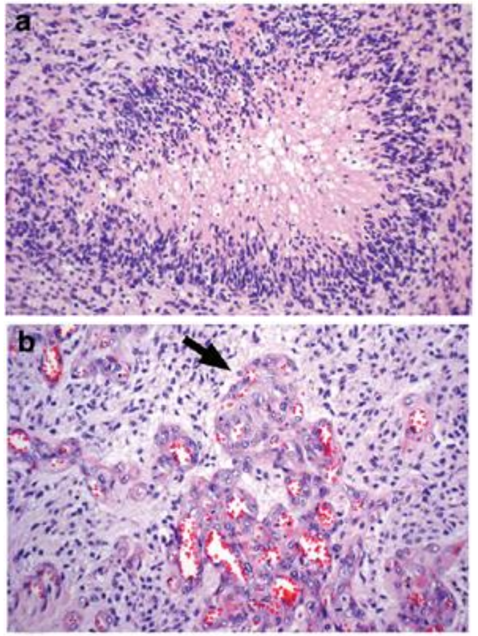
Hi-Res Image

Quantitative Assessment: Nuclear Morphology

Different types of nuclei and cells

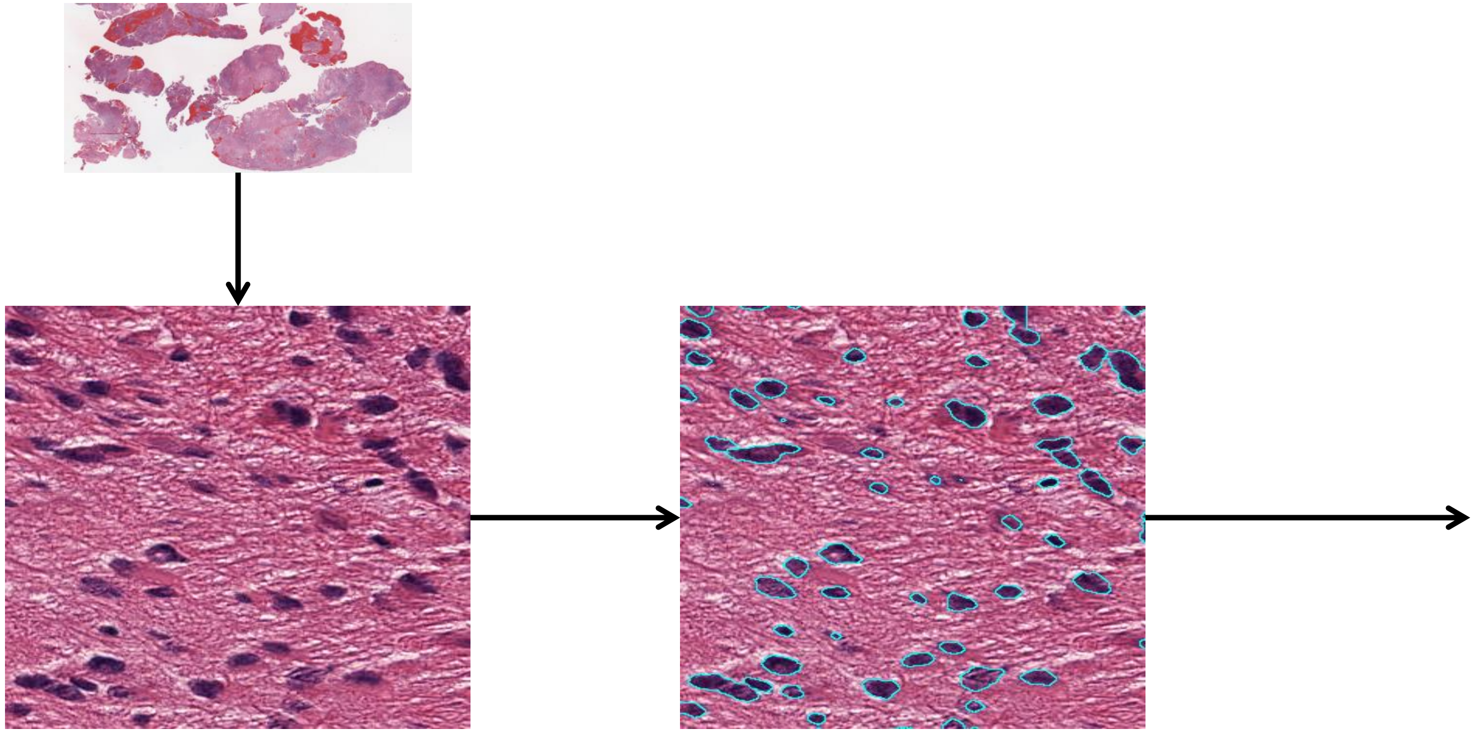


Regions of necrosis
and angiogenesis



- Datasets of rich nuclear morphological information
- *Some Research Questions:*
  - *Are there natural clusters of tumor morphology?*
  - *Are there links to patient outcome and molecular/genomic characteristics?*

Quantitative Assessment: Nuclear Features



Nuclear Morphometry

- Nuclei Area
- Nuclei Perimeter
- Eccentricity
- Circularity
- Major Axis
- Minor Axis
- Extent Ratio
- Fourier Shape Descriptor

Intensity Information

- Avg Inty
- Std Inty
- Max Inty
- Min Inty

Texture Information

- Entropy
- Energy
- Skewness
- Kurtosis

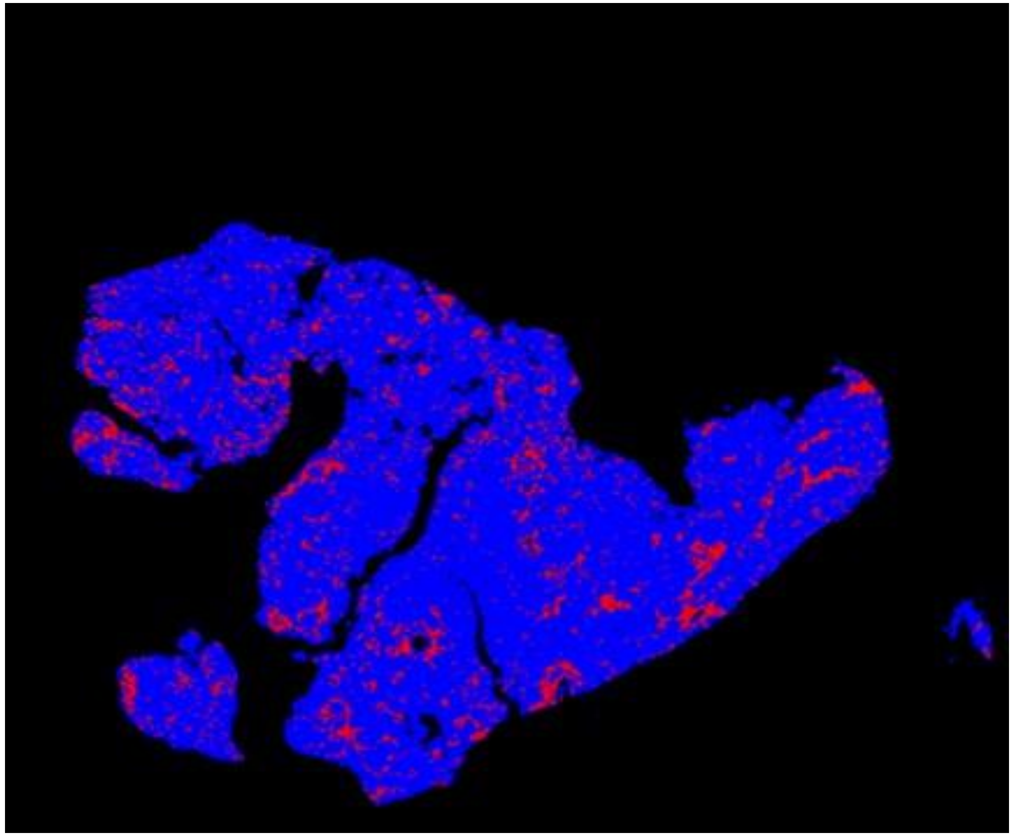
Gradient Statistics

- Avg GM
- Std GM
- Etropy GM
- Skewness GM
- Energy GM
- Kurtosis GM
- Edge Pixel Summation
- Edge Pixel Percentage

- Big Data Problem
  - Thousands of images
  - Millions of regions
  - Billions of nuclei and features

Quantitative Assessment: Classification and Spatial Patterns

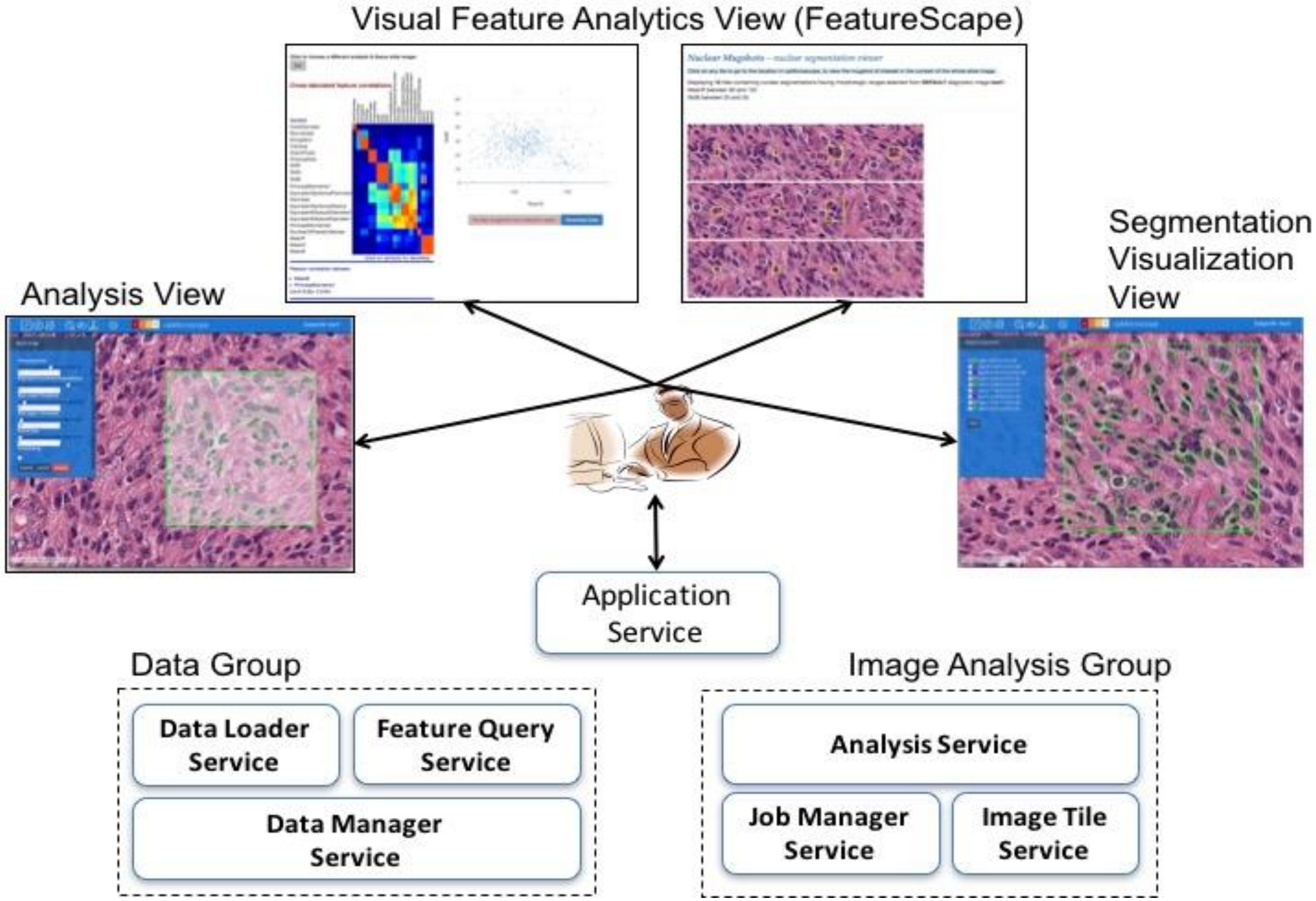
- Spatial characterization – mapping of patterns
- Tumor Infiltrating Lymphocytes by CNN Deep Learning Method
- Classification of images and image regions



Quantitative Imaging in Pathology (QuIP) Software Platform

Software: https://github.com/SBU-BMI/quip_distro.git
Description: https://sbu-bmi.github.io/quip_distro/

QuIP Software Platform



Visual Feature Analytics View (FeatureScape)

Analysis View

Segmentation Visualization View

Application Service

Data Group

Data Loader Service

Feature Query Service

Data Manager Service

Image Analysis Group

Analysis Service

Job Manager Service

Image Tile Service

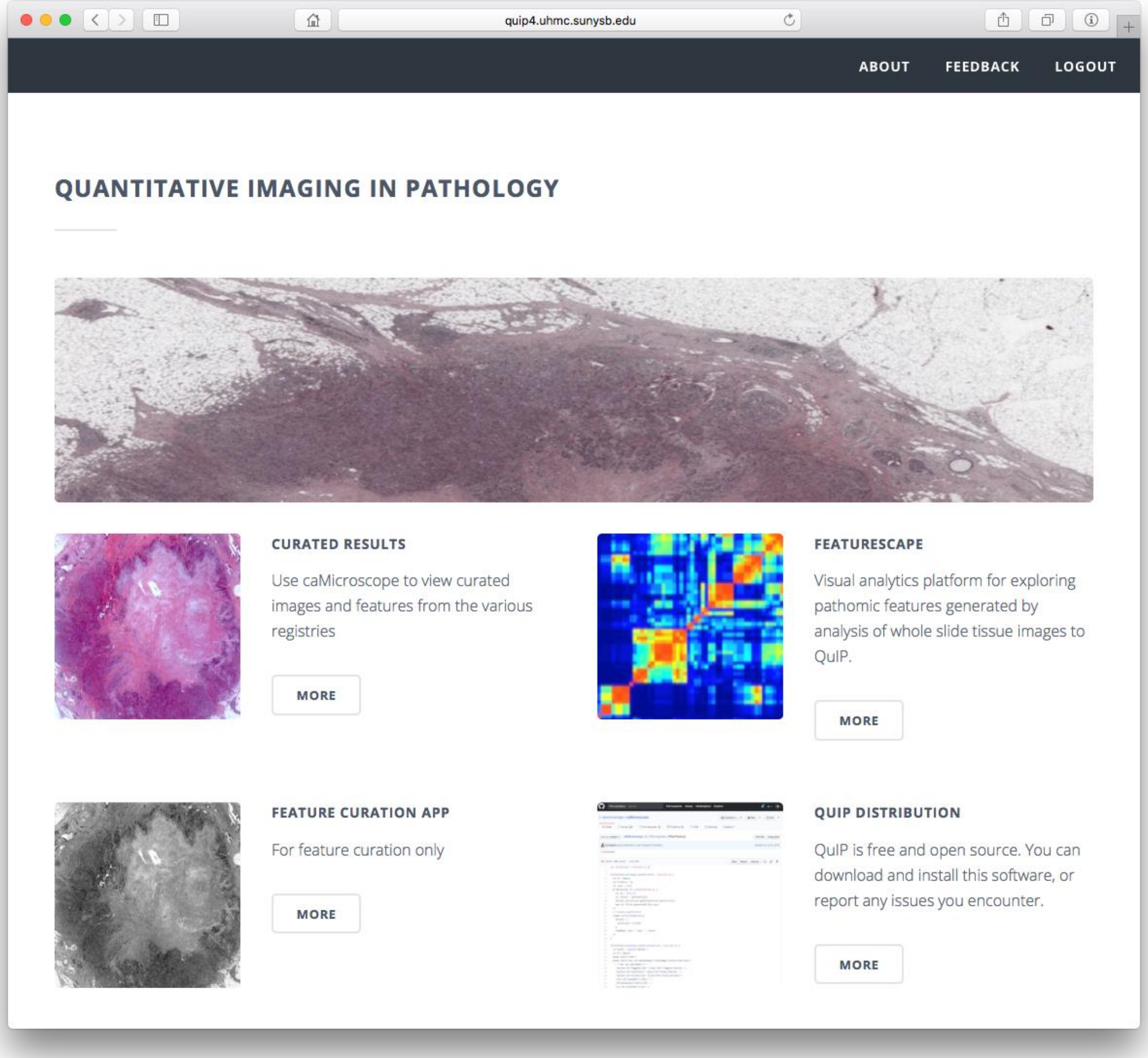
Containerized Software for QuIP Services

- The **application service group** is a single container that hosts a suite of Web applications to view images and interact with analysis results.
  - Visualization – caMicroscope, FeatureScape
  - Higher-level Apps – TIL/Heatmap Editor, Segmentation Curation App
- The **image analysis group** is made up of three containers, which collectively execute image analysis requests.
  - *Analysis service* – hosts analysis pipeline
  - *Job Manager service* – tracks jobs
  - *Image tile service* – services image tiles for analysis
- The **data management service group** is implemented as a set of three containers for data loading, data management, and query processing.
  - *Data loader service* -- load image metadata and analysis results
  - *Data manager service* – manage and index image data, analysis results, features
  - *Feature query service* – query feature data for visualization and exploration

Image Analysis Libraries

- Deep learning workflow for analysis of TIL patterns
  - A set of programs and scripts to train, re-train, execute prediction
  - https://github.com/SBU-BMI/u24_lymphocyte
  - Graphical user interfaces via QuIP for review and refinement of prediction results and iterative learning
  - Docker container for prediction and training steps
- Nuclear segmentation pipeline
  - A C++ library and programs to carry out segmentation of nuclear material in whole slide tissue images
  - https://github.com/SBU-BMI/pathomics_analysis.git
  - Integrated in QuIP as a Docker container

QuIP Web Interface



QUANTITATIVE IMAGING IN PATHOLOGY

CURATED RESULTS

Use caMicroscope to view curated images and features from the various registries

FEATURESCAPE

Visual analytics platform for exploring pathomic features generated by analysis of whole slide tissue images to QuIP.

FEATURE CURATION APP

For feature curation only

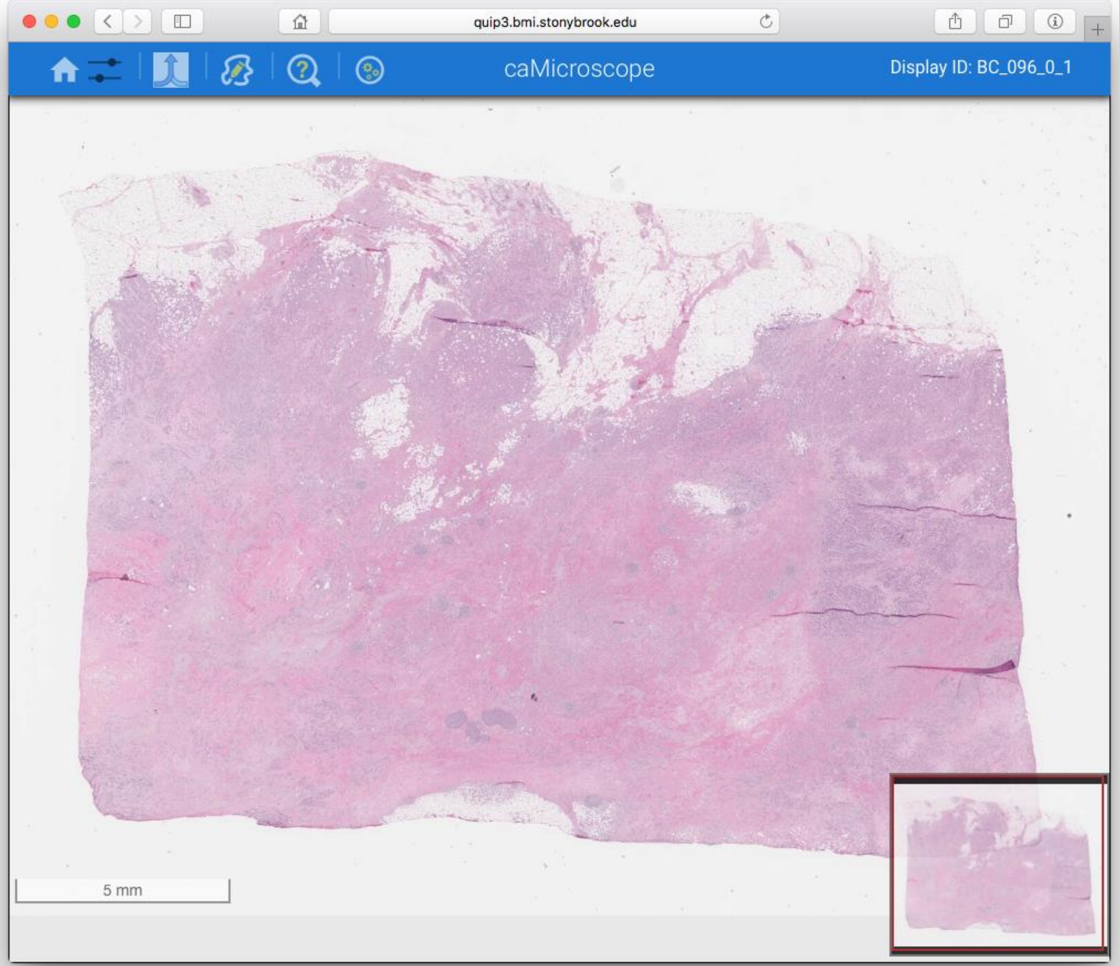
QUIP DISTRIBUTION

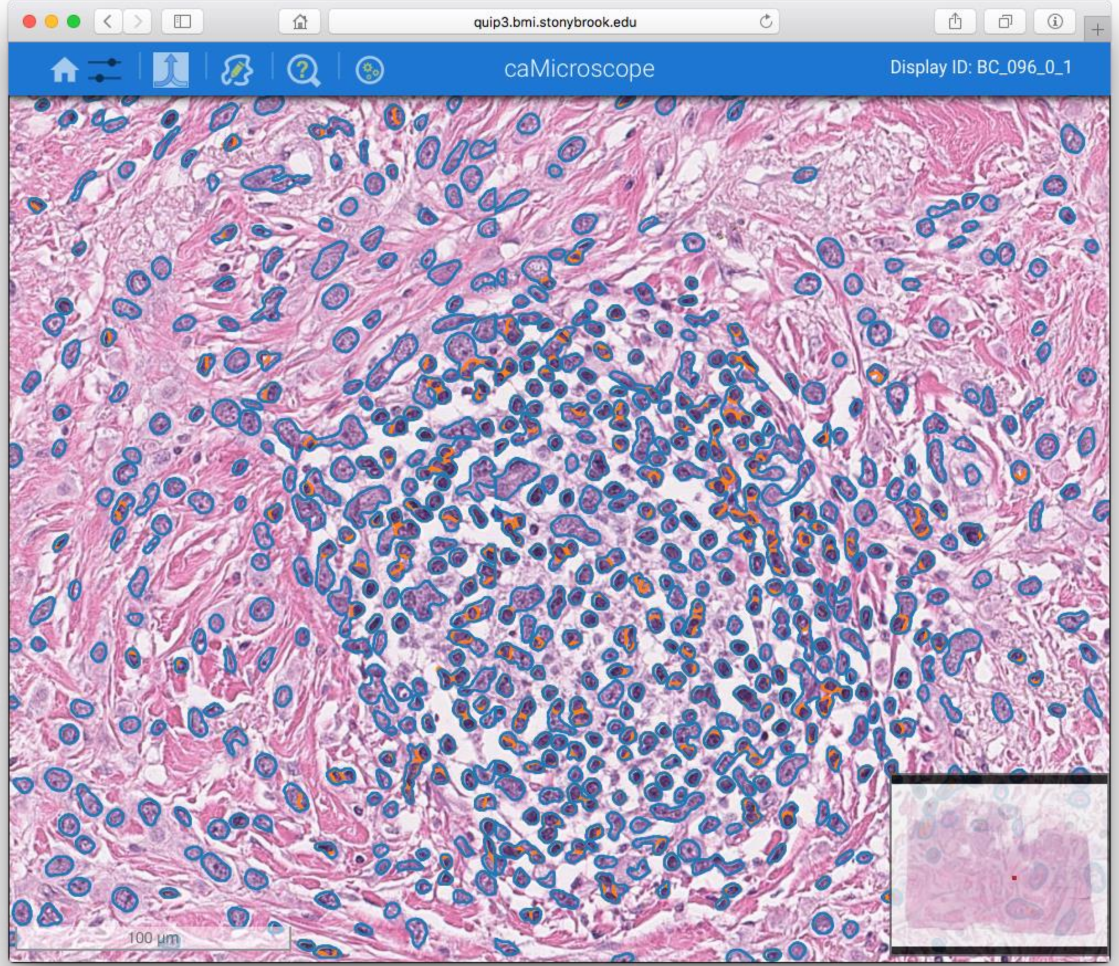
QuIP is free and open source. You can download and install this software, or report any issues you encounter.

Visualization: caMicroscope

- Web-based platform for visualizing digital pathology images with segmentation results and features that are overlaid on the images
- Smooth web-based viewing of high resolution whole slide tissue images
  - Panning, Zooming
  - Overlay of one or more segmentation results
  - Human markups
- Provides APIs that allow interactive back-and-forth between feature exploration and image visualization

caMicroscope

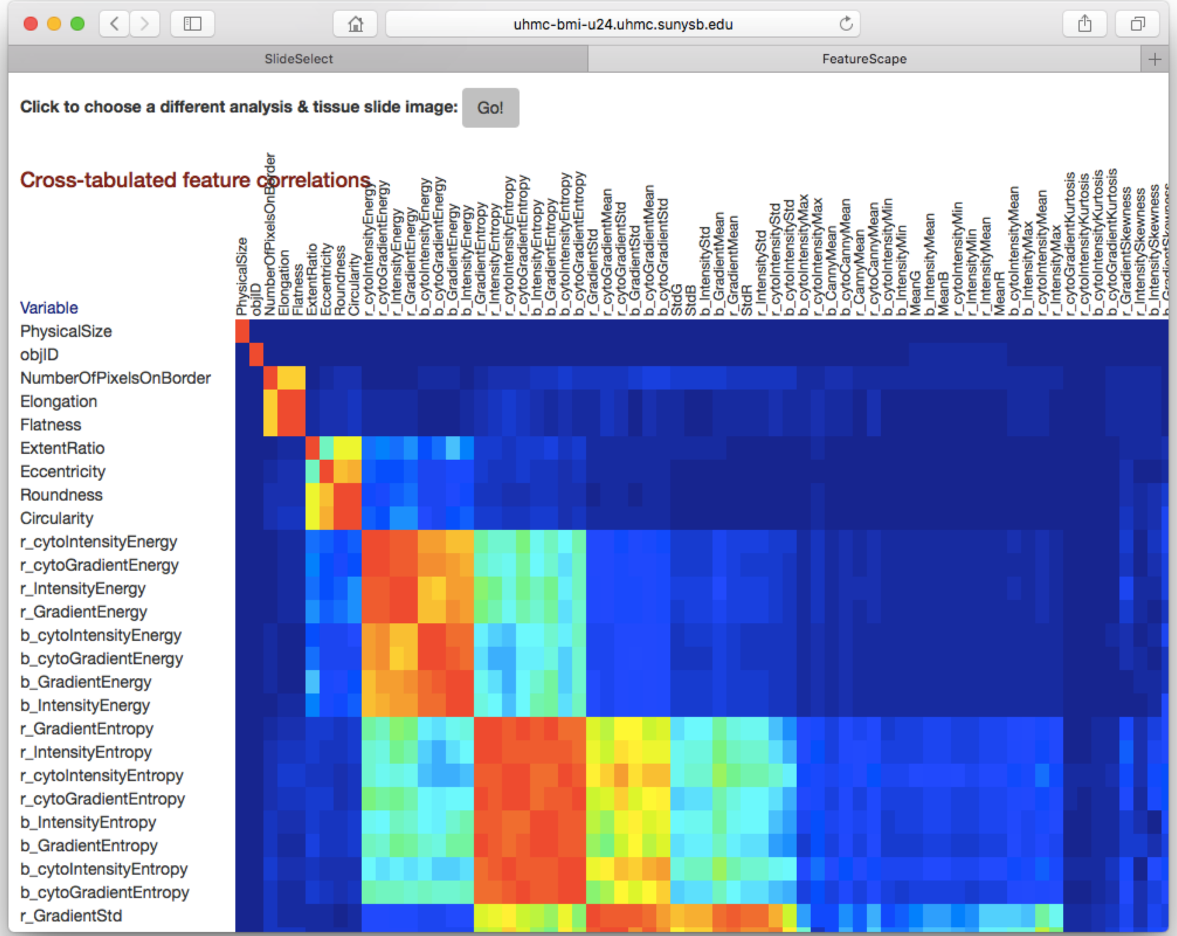

quip3.bmi.stonybrook.edu


quip3.bmi.stonybrook.edu

Visualization: FeatureScape

- Web-based apps for feature based visual analytics
- Visualization of inter-related imaging features
- Go from feature level to image level

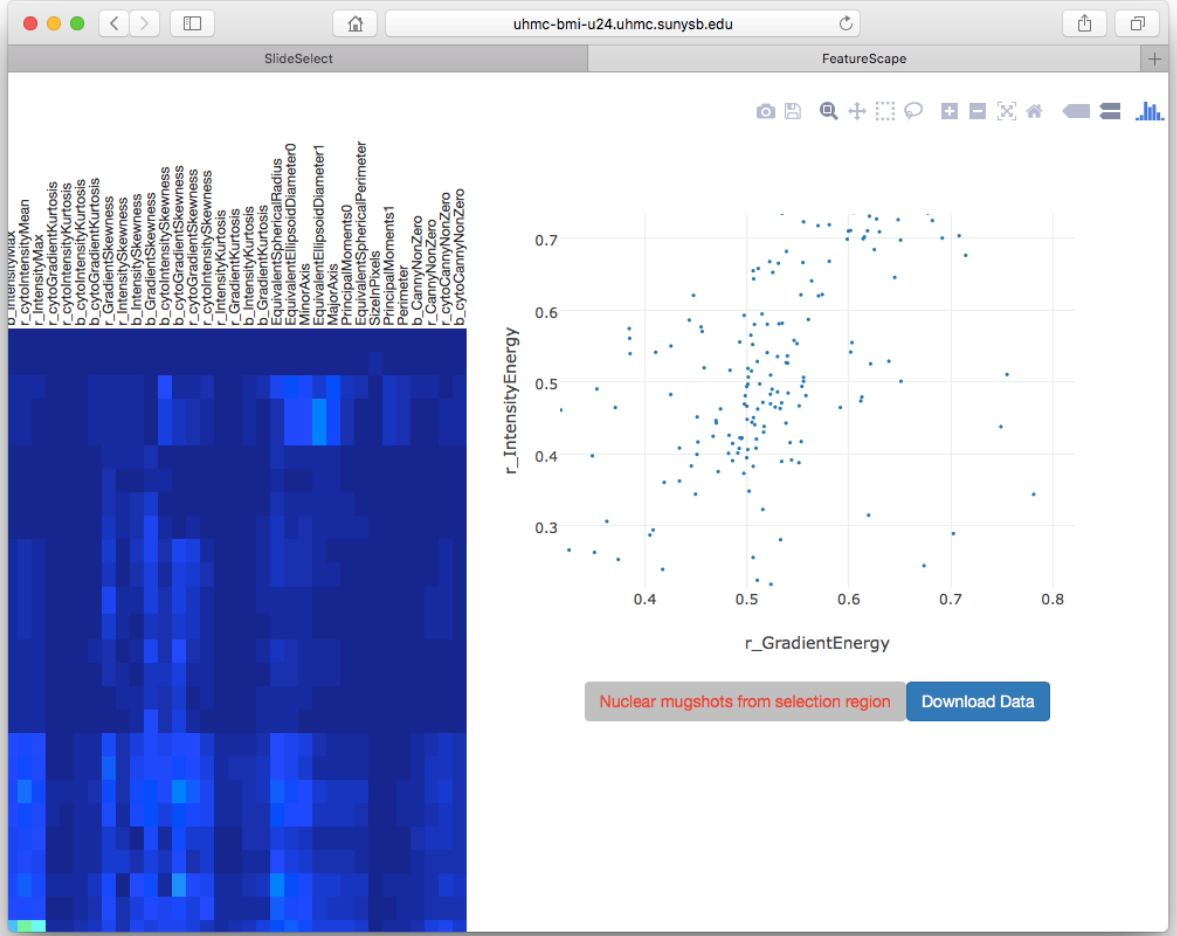
FeatureScape



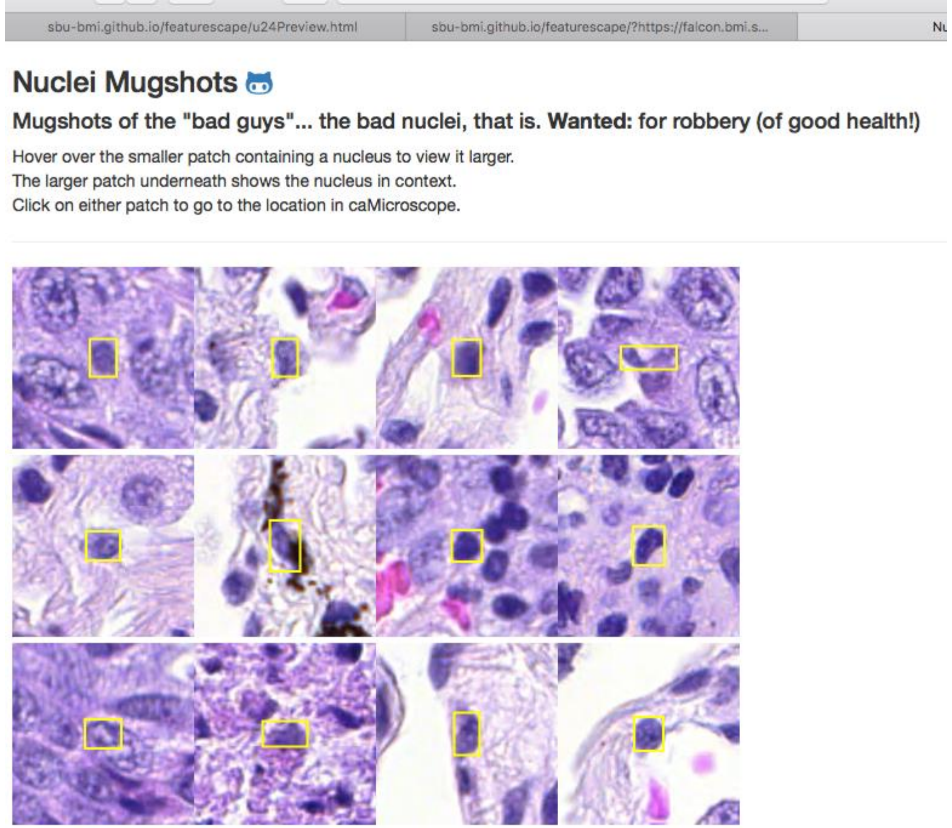
Click to choose a different analysis & tissue slide image:

Cross-tabulated feature correlations

Variable
PhysicalSize
objID
NumberOfPixelsOnBorder
Elongation
Flatness
ExtentRatio
Eccentricity
Roundness
Circularity
r_cytoIntensityEnergy
r_cytoGradientEnergy
r_IntensityEnergy
r_GradientEnergy
b_cytoIntensityEnergy
b_cytoGradientEnergy
b_GradientEnergy
b_IntensityEnergy
r_GradientEntropy
r_IntensityEntropy
r_cytoIntensityEntropy
r_cytoGradientEntropy
b_IntensityEntropy
b_GradientEntropy
b_cytoIntensityEntropy
b_cytoGradientEntropy
r_GradientStd



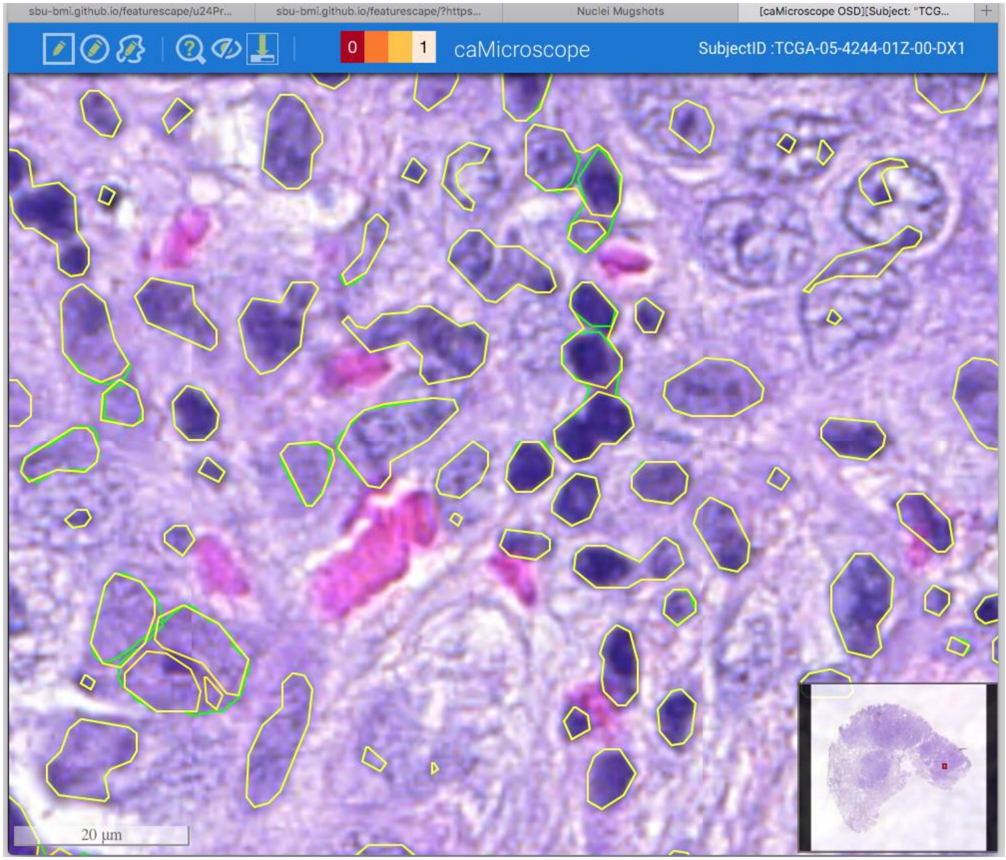
caMicroscope + FeatureScape



Nuclei Mugshots

Mugshots of the "bad guys"... the bad nuclei, that is. Wanted: for robbery (of good health!)

Hover over the smaller patch containing a nucleus to view it larger.
The larger patch underneath shows the nucleus in context.
Click on either patch to go to the location in caMicroscope.



Database: FeatureDB

- Based on NoSQL document database
- Flexible data model in JSON format
  - Borrows elements from AIM and PAIS models
- Organizes segmentation and feature data into GeoJSON compatible format
  - Segmentation results as polygons
  - Features as key-value pairs
    - Shape, intensity, texture and size features

FeatureDB – GeoJSON-based Data Model

```
{
    "type" : "Feature",
    "parent_id" : "self",
    "randval" : 0.43199294805526733,
    "object_type" : "nucleus",
    "x" : 0.017050,
    "y" : 0.358914,
    "normalized" : "true",
    "bbox" : [
        0.016993561910465638,
        0.35885439819583437,
        0.017108055978792176,
        0.35897366395246716
    ],
    "geometry" : {
        "type" : "Polygon",
        "coordinates" : [
            [
                [0.017042001709342003,0.3589736521244049],
                [0.01703319512307644,0.3589736521244049],
                …
            ]
        ]
    },
    "footprint" : 93,
```

FeatureDB – GeoJSON-based Data Model

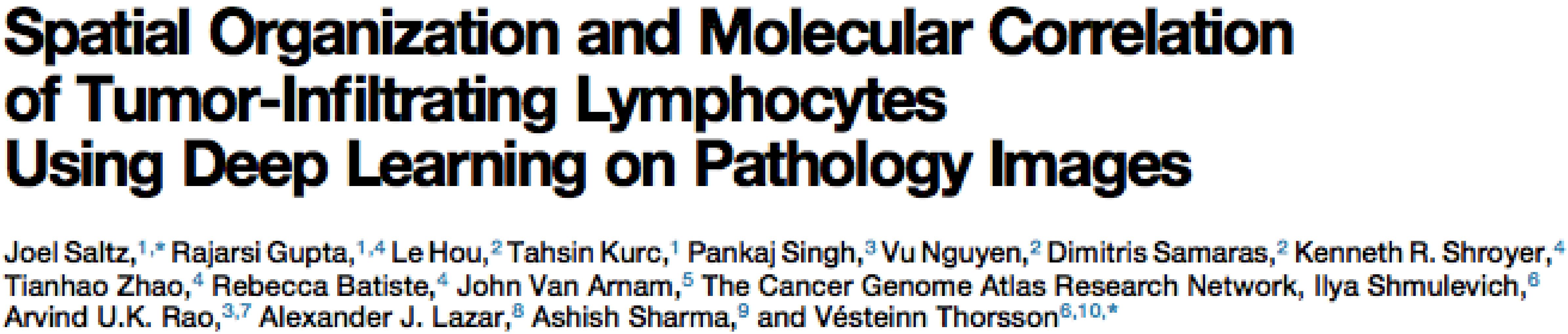
```
        "properties" : {
        "scalar_features" : [
            {
                "ns" : "http://u24.bmi.stonybrook.edu/v1",
                "nv" : [
                    {"name" : "b_cytoGradientMean","value" : 10.4891996383667},
                    …
                ]
            }
        ]
    },
    "provenance" : {
        "image" : {
            "case_id" : "0001-215633",
            "subject_id" : "0001-212"
        },
        "analysis" : {
            "execution_id" : "Mean_Shift_1.0",
            "study_id" : "small",
            "source" : "computer",
            "computation" : "segmentation"
        },
    },
    "submit_date" : ISODate("2017-05-24T19:48:37.776Z")
}
```

Higher-level Web Applications

- TIL/Heatmap Editor
  - Used in study of Tumor Infiltrating Lymphocytes (TILs)
- Segmentation Curation Application
  - Used in SEER Virtual Repository Pilot/Project

TIL/Heatmap Editor

- Support for Iterative Learning/Refinement Workflows
- Motivated by and used in study of TIL patterns in 13 TCGA cancer types



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,*]

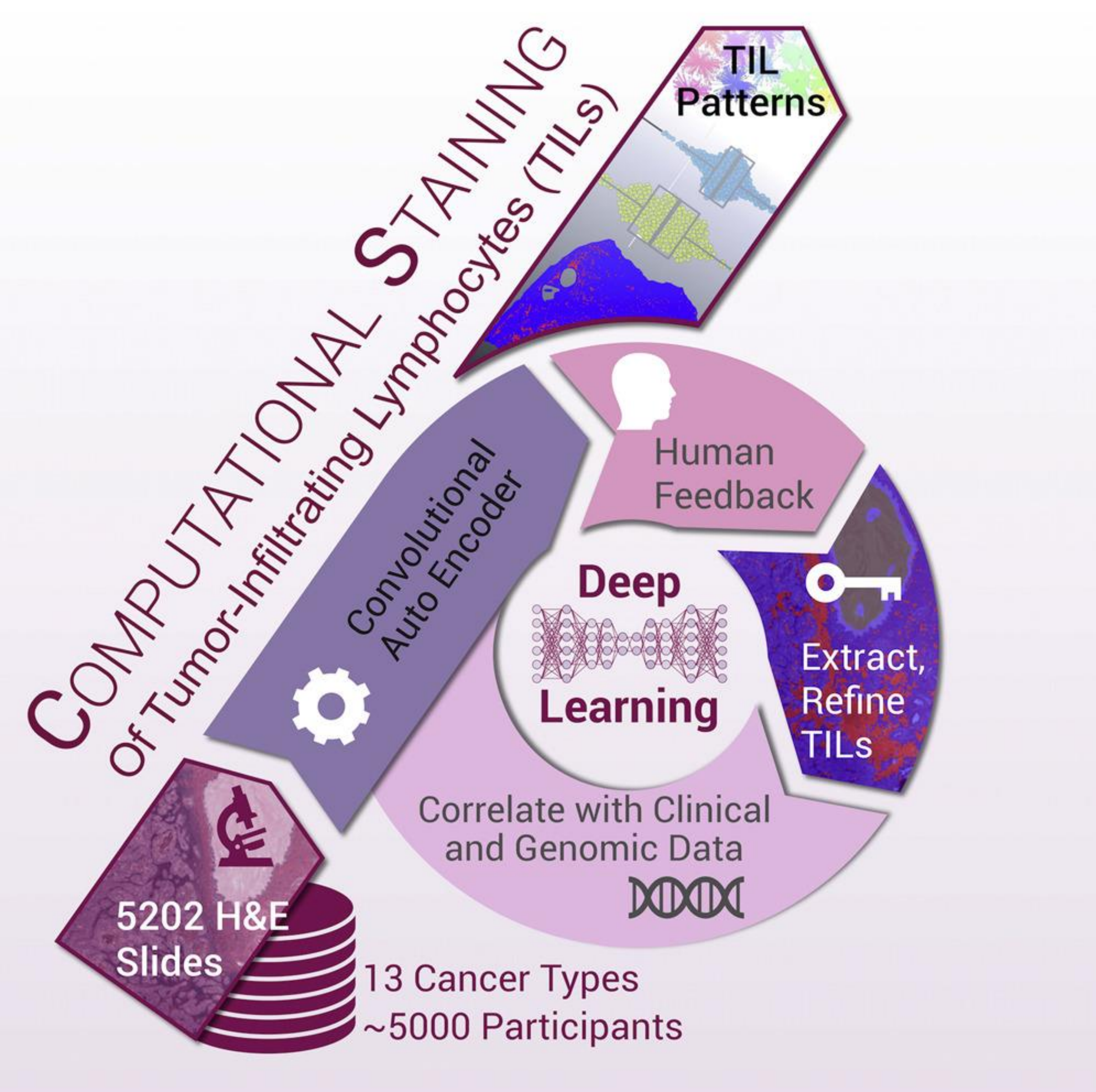
http://www.cell.com/cell-reports/pdf/S2211-1247(18)30447-9.pdf

- 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

Importance of Immune System in Cancer Treatment and Prognosis

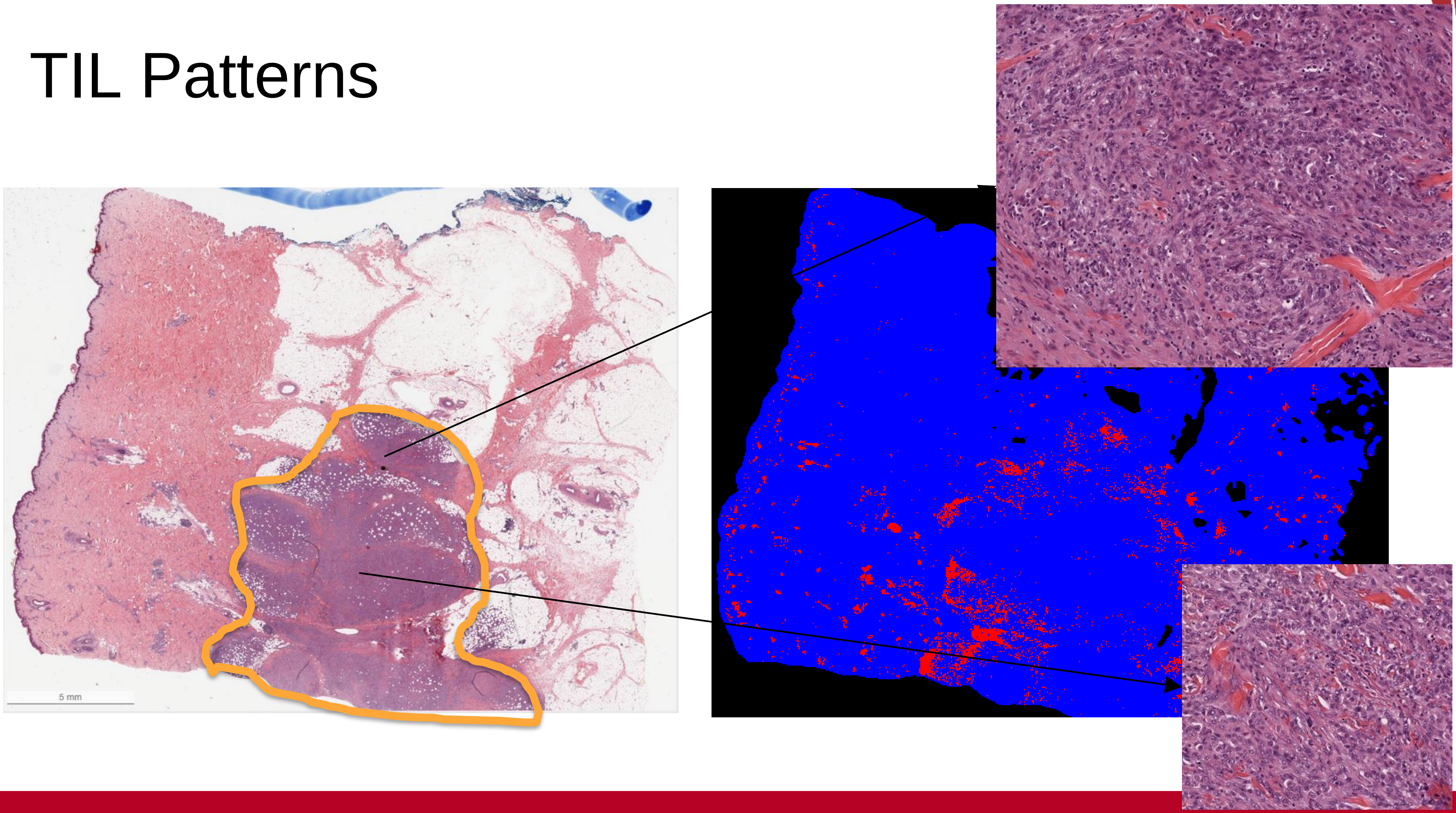
Stony Brook University

- 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

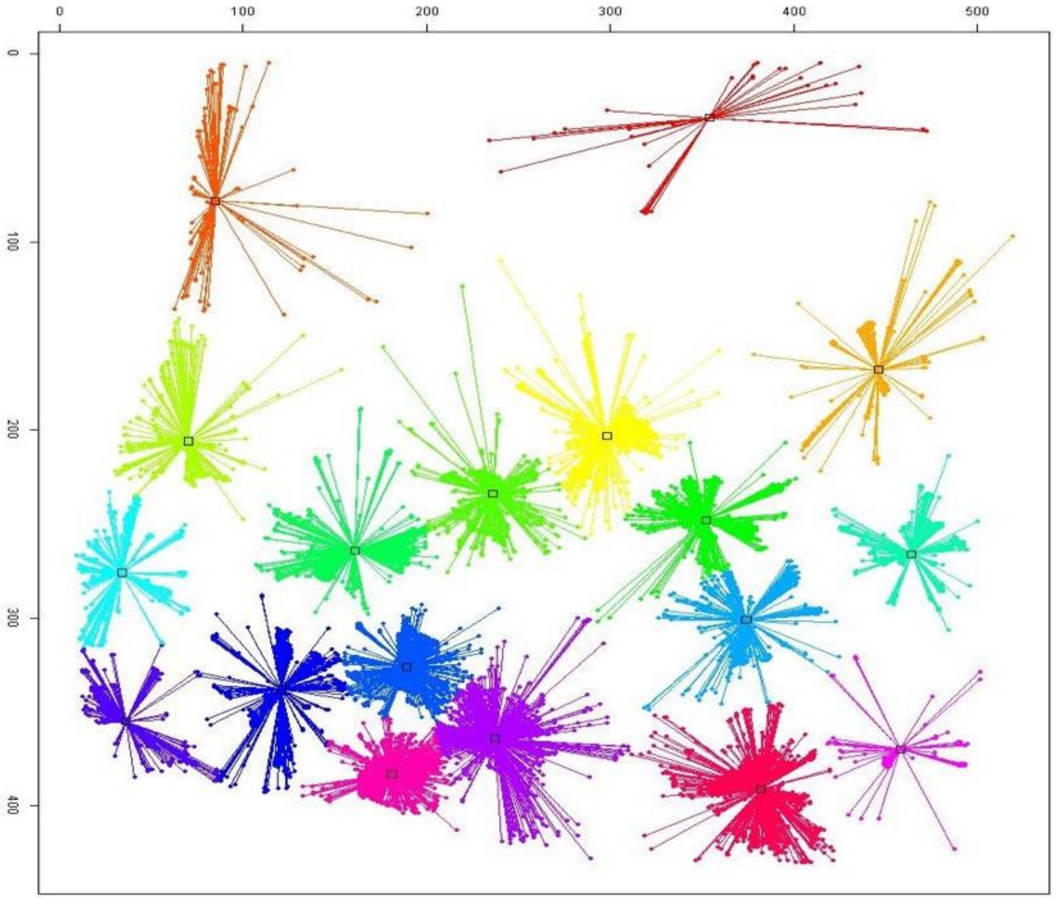


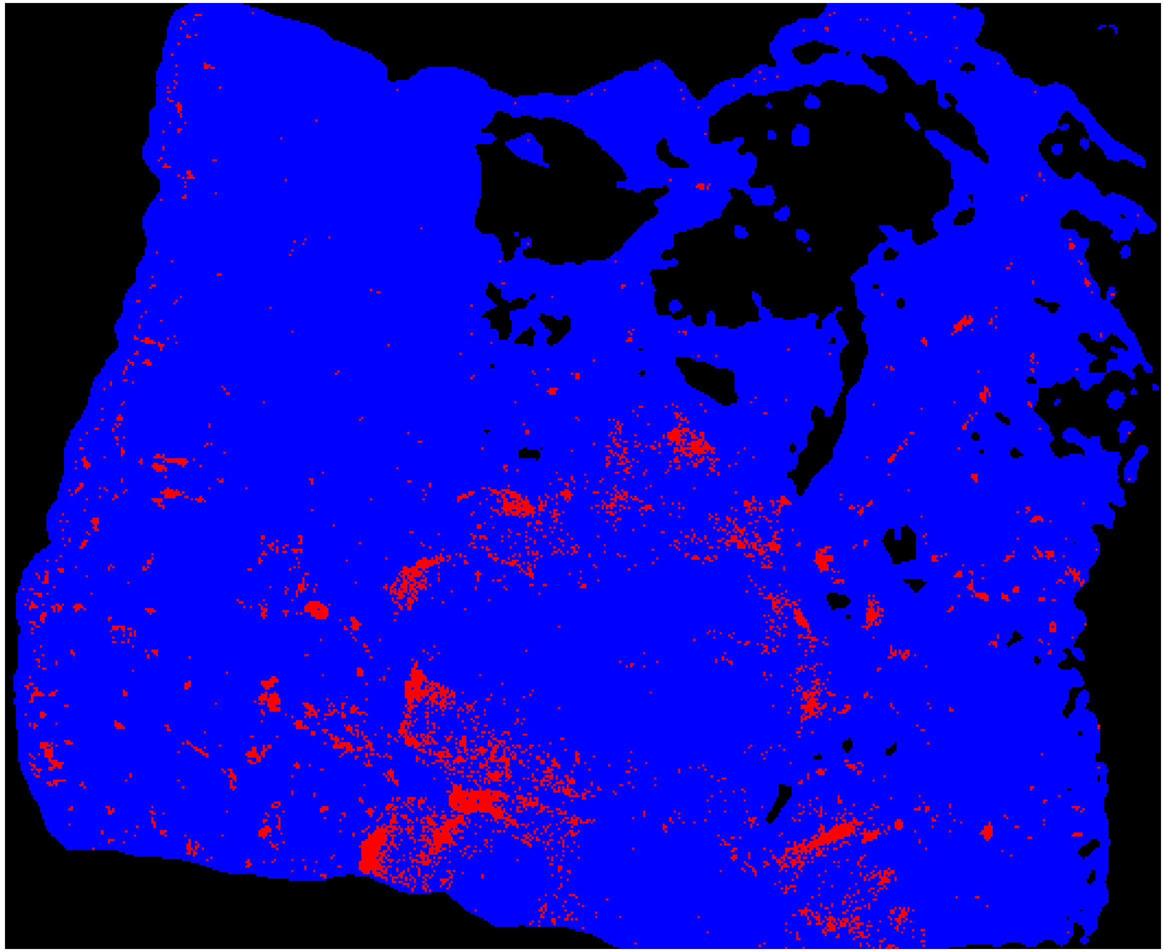
- Deep learning based analysis of patterns of tumor infiltrating lymphocytes (TILs)
- TIL patterns generated from 4,759 TCGA subjects (5,202 H&E slides), 13 cancer types
- TIL patterns correlate with pathologist eye and molecular estimates
- TIL patterns linked to tumor and immune molecular features, cancer type, and outcome

TIL Patterns



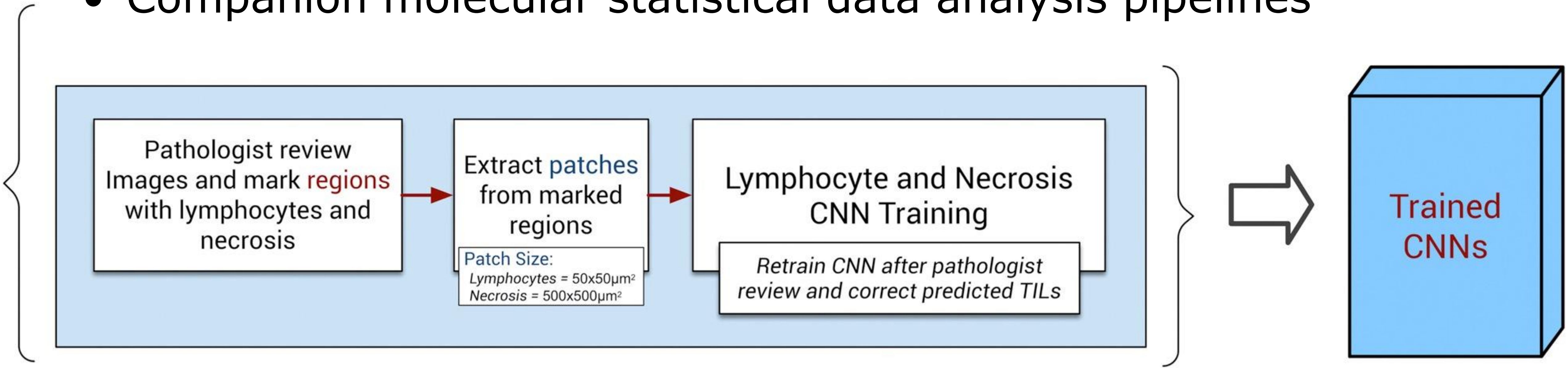
TIL Patterns – Clustering Indices





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



Pathologist review Images and mark regions with lymphocytes and necrosis

Extract patches from marked regions

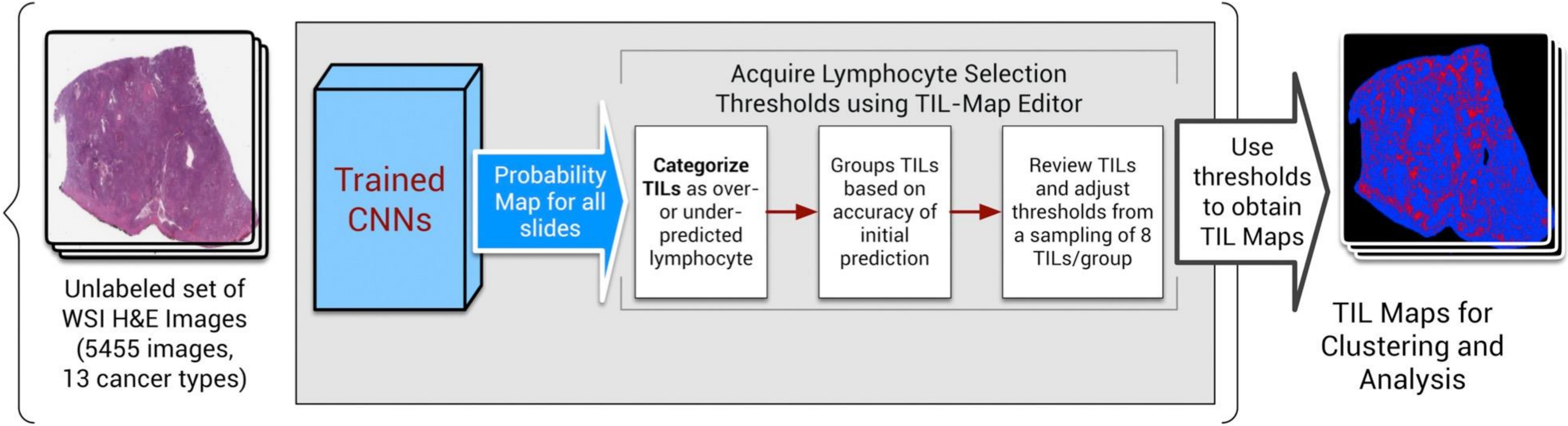
Patch Size:
Lymphocytes = 50x50μm^2
Necrosis = 500x500μm^2

Lymphocyte and Necrosis CNN Training

Retrain CNN after pathologist review and correct predicted TILs

Trained CNNs

Training, threshold adjustment, quality control



Unlabeled set of WSI H&E Images (5455 images, 13 cancer types)

Trained CNNs

Probability Map for all slides

Acquire Lymphocyte Selection Thresholds using TIL-Map Editor

Categorize TILs as over- or under-predicted lymphocyte

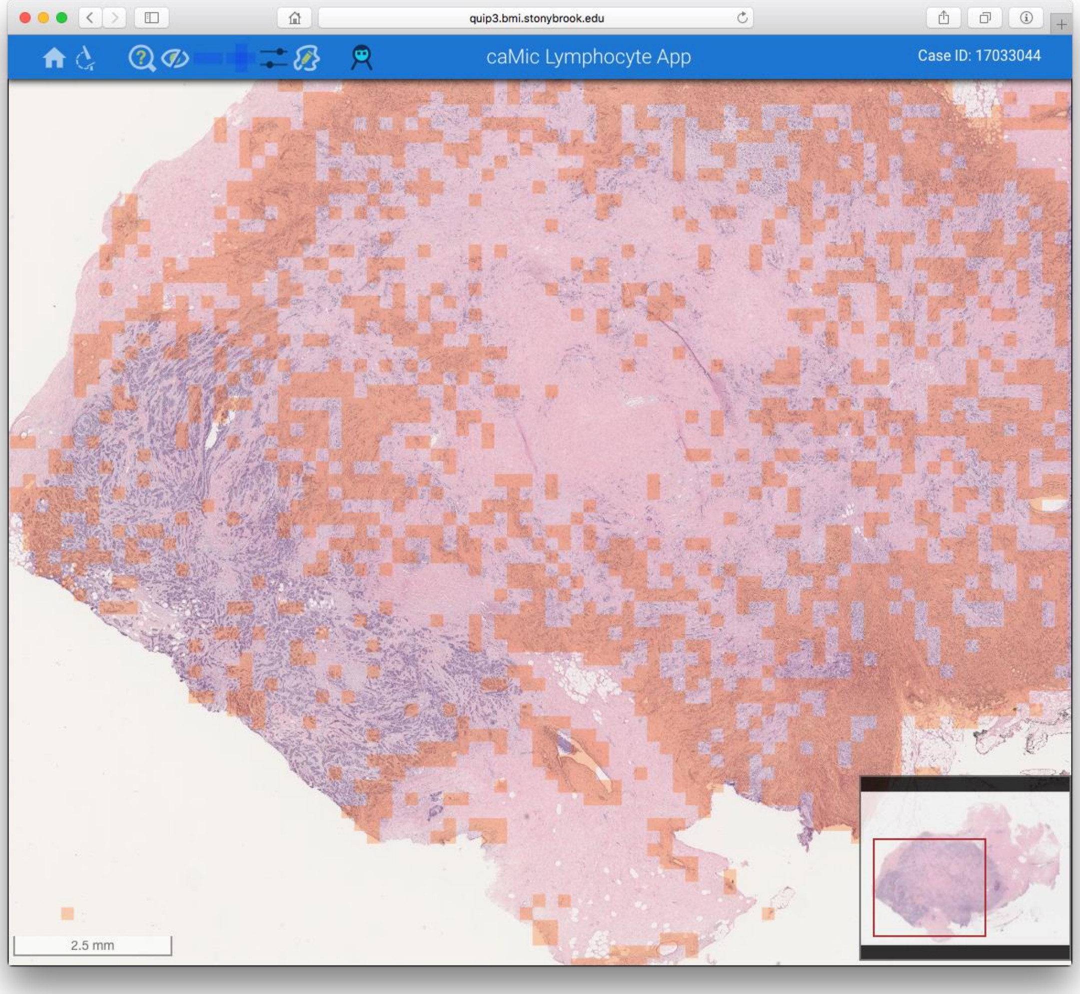
Groups TILs based on accuracy of initial prediction

Review TILs and adjust thresholds from a sampling of 8 TILs/group

Use thresholds to obtain TIL Maps

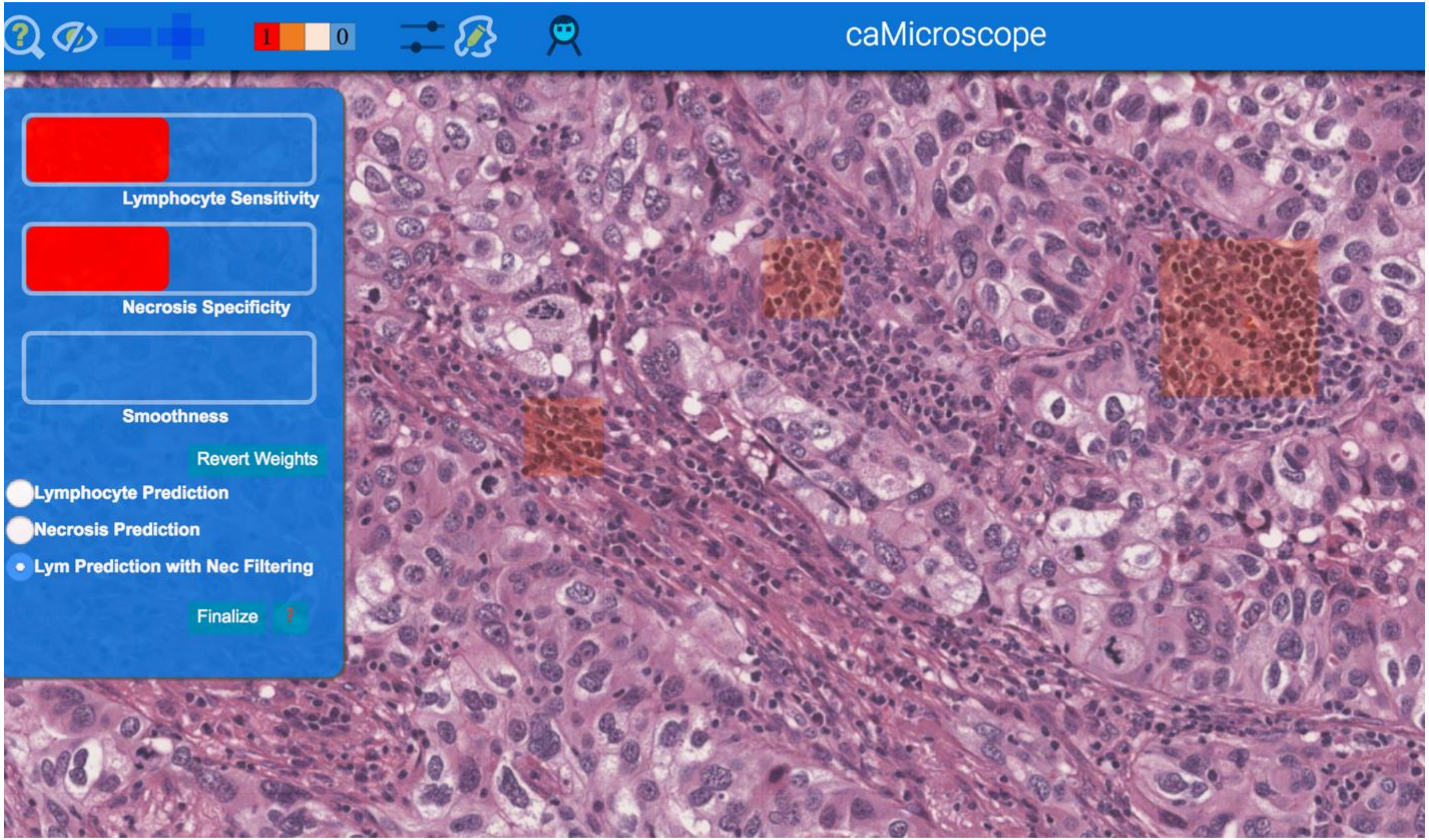
TIL Maps for Clustering and Analysis

Heatmap View of TIL Probability Maps



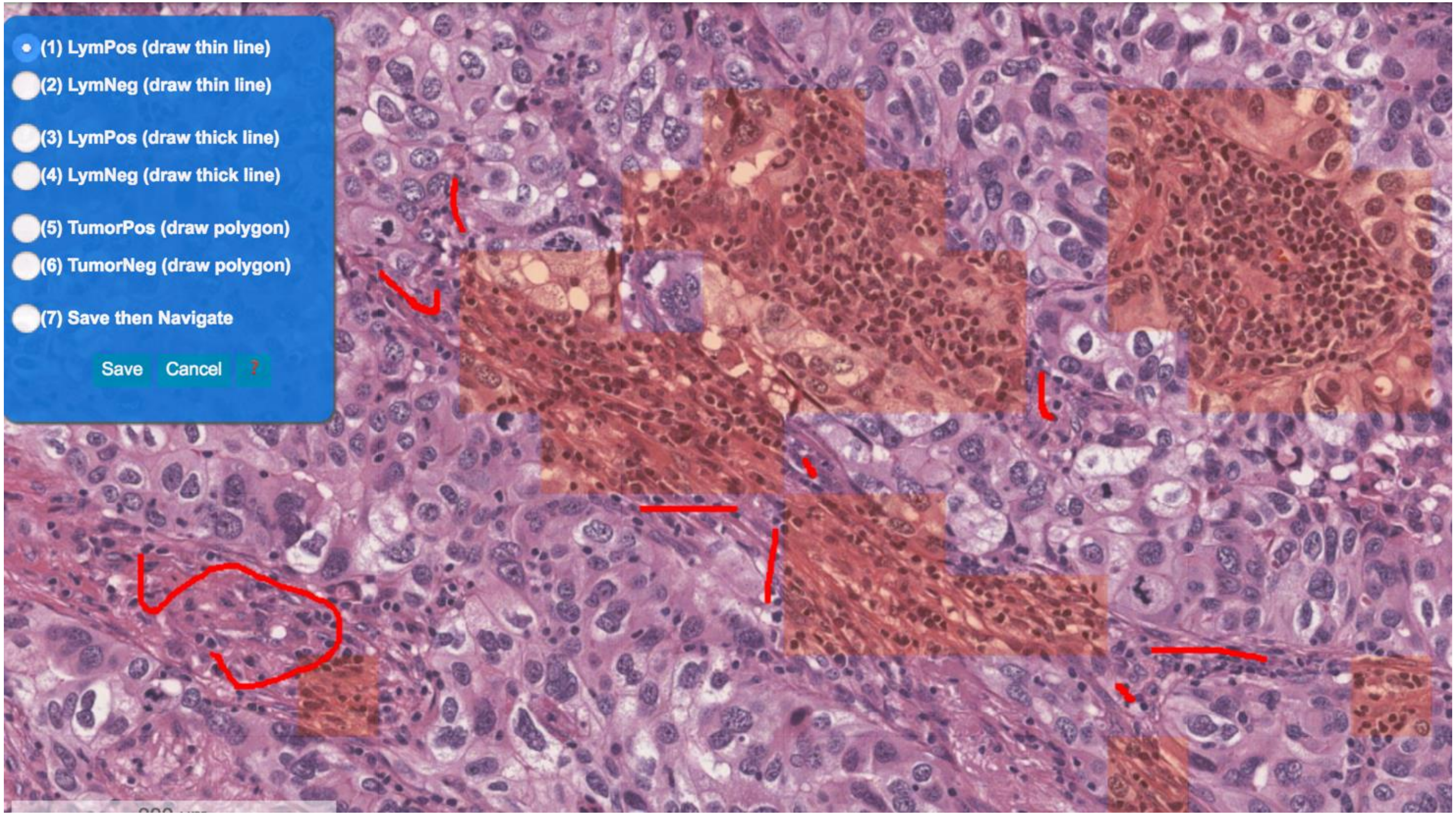
TIL Map Editor

Stony Brook University



○ Lymphocyte Prediction
○ Necrosis Prediction
● Lym Prediction with Nec Filtering

TIL Map Editor

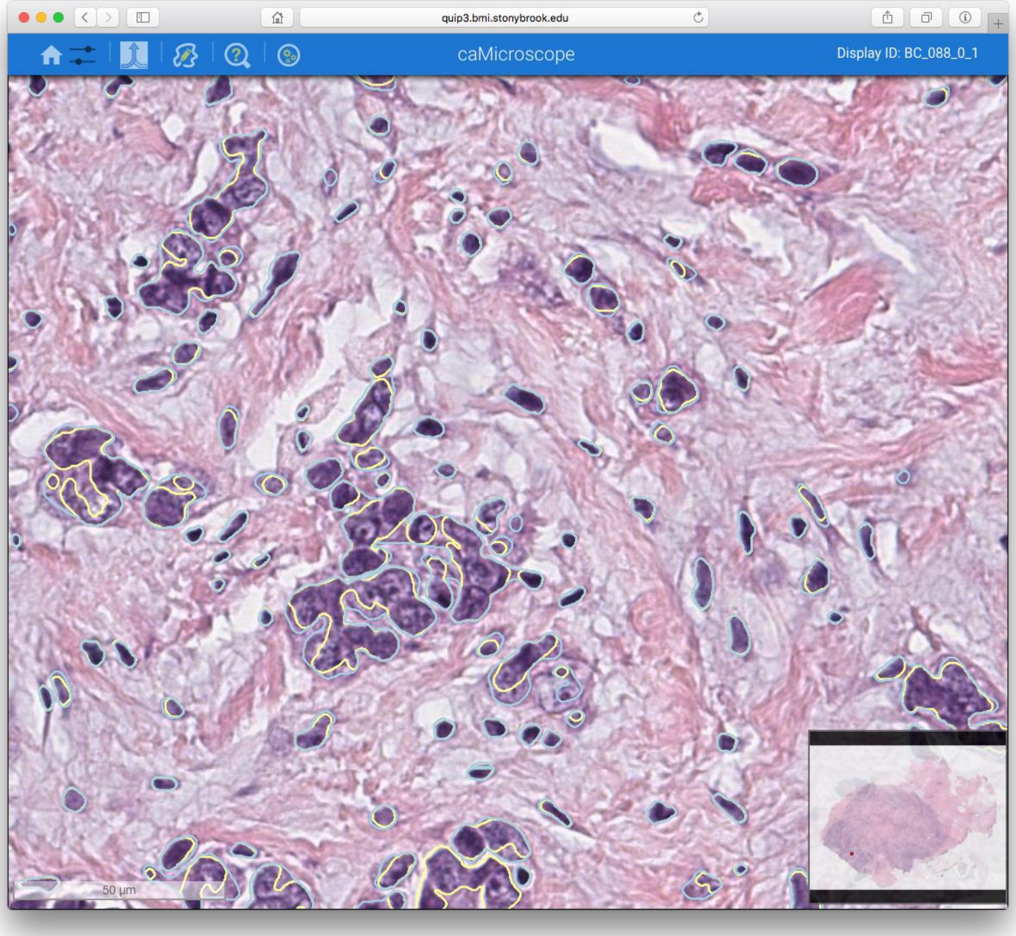


Curation Tool for Nuclear Segmentation Results

- Motivated by sensitivity of analysis methods to input parameters
- Used in SEER Virtual Repository Pilot/Project

Analysis Sensitivity: Generating Imaging Feature Sets

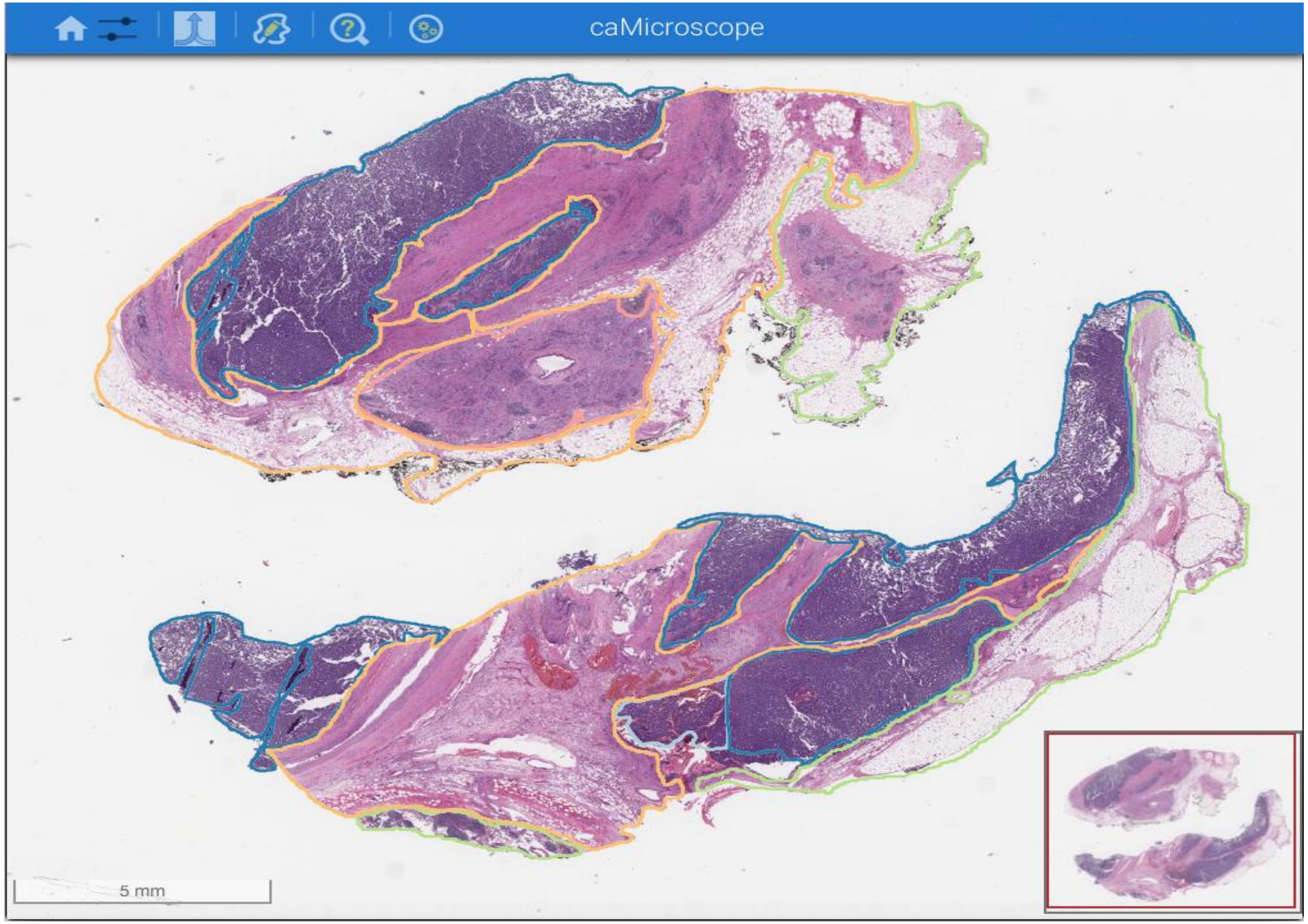
- Image analysis pipelines are sensitive to input parameters



Generating Robust Feature Sets

- Run multiple analyses
- Store, index, interact with results
- Curation: Visual comparison and selection of results

Segmentation Results Curation Tool



Methods and tools for integrating pathomics data into cancer registries - UG3CA225021

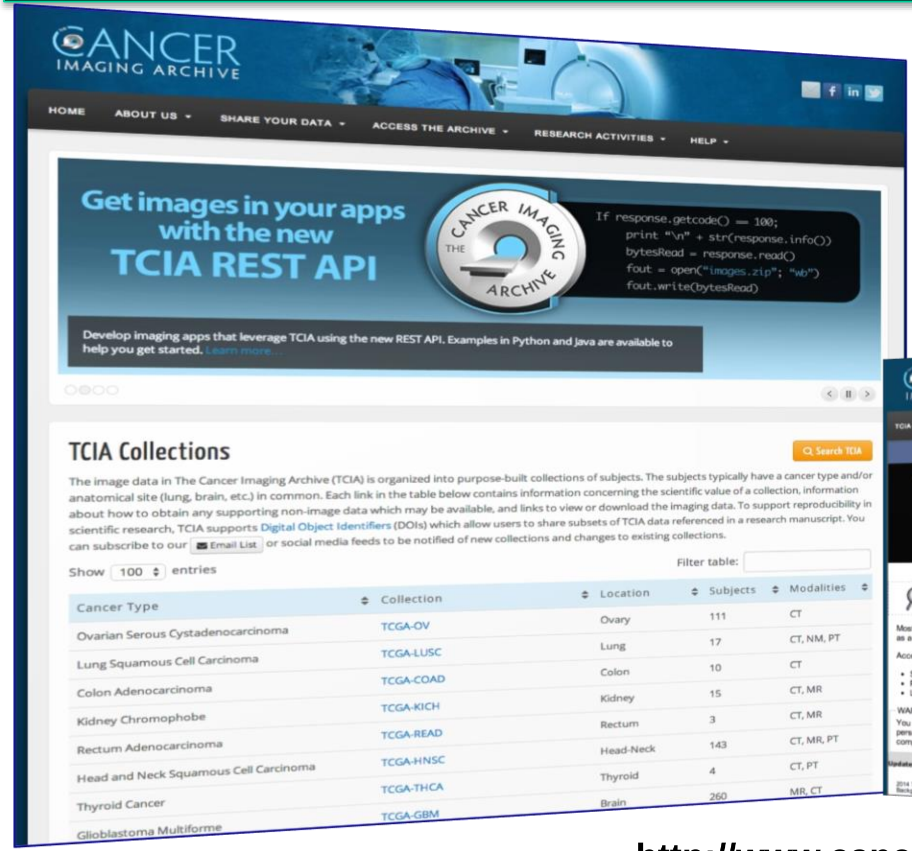
Saltz, Sharma, Foran and Durban

- 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



Get images in your apps with the new TCIA REST API

```
If response.getcode() == 100;
    print "\n" + str(response.info())
    bytesRead = response.read()
    fout = open("images.zip", "wb")
    fout.write(bytesRead)
```

Develop imaging apps that leverage TCIA using the new REST API. Examples in Python and Java are available to help you get started. Learn more...

TCIA Collections

The image data in The Cancer Imaging Archive (TCIA) is organized into purpose-built collections of subjects. The subjects typically have a cancer type and/or anatomical site (lung, brain, etc.) in common. Each link in the table below contains information concerning the scientific value of a collection, information about how to obtain any supporting non-image data which may be available, and links to view or download the imaging data. To support reproducibility in scientific research, TCIA supports Digital Object Identifiers (DOIs) which allow users to share subsets of TCIA data referenced in a research manuscript. You can subscribe to our Email List or social media feeds to be notified of new collections and changes to existing collections.

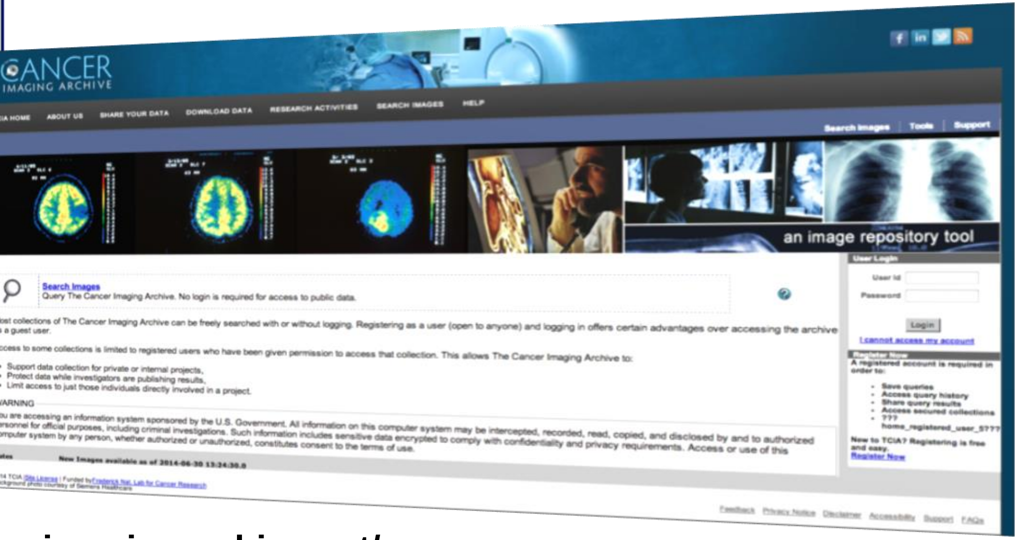
Show 100 entries

Filter table:

Cancer Type	Collection	Location	Subjects	Modalities
Ovarian Serous Cystadenocarcinoma	TCGA-OV	Ovary	111	CT
Lung Squamous Cell Carcinoma	TCGA-LUSC	Lung	17	CT, NM, PT
Colon Adenocarcinoma	TCGA-COAD	Colon	10	CT
Kidney Chromophobe	TCGA-KICH	Kidney	15	CT, MR
Rectum Adenocarcinoma	TCGA-READ	Rectum	3	CT, MR
Head and Neck Squamous Cell Carcinoma	TCGA-HNSC	Head-Neck	143	CT, MR, PT
Thyroid Cancer	TCGA-THCA	Thyroid	4	CT, PT
Glioblastoma Multiforme	TCGA-GBM	Brain	260	MR, CT

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.



Search Images
Query The Cancer Imaging Archive. No login is required for access to public data.

Most collections of The Cancer Imaging Archive can be freely searched with or without logging. Registering as a user (open to anyone) and logging in offers certain advantages over accessing the archive as a guest user.

Access to some collections is limited to registered users who have been given permission to access that collection. This allows The Cancer Imaging Archive to:

- Support data collection for private or internal projects,
- Protect data while investigators are publishing results,
- Limit access to just those individuals directly involved in a project.

WARNING

You are accessing an information system sponsored by the U.S. Government. All information on this computer system may be intercepted, recorded, read, copied, and disclosed by and to authorized personnel for official purposes, including criminal investigations. Such information includes sensitive data encrypted to comply with confidentiality and privacy requirements. Access or use of this computer system by any person, whether authorized or unauthorized, constitutes consent to the terms of use.

A registered account is required in order to:

- Save queries
- Access query history
- Share query results
- Access secured collections
- ???home_registered_user_5???

New to TCIA? Registering is free and easy.
Register Now

Updates New Images available as of 2014-06-30 13:24:30.0

2014 TCIA Site License | Funded by Frederick Nat. Lab for Cancer Research
Background photo courtesy of Siemens Healthcare

http://www.cancerimagingarchive.net/

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

- 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.

Quantitative Imaging in Pathology (QuIP) Platform

- Software: https://github.com/SBU-BMI/quip_distro.git
- Containerized, service oriented
- Several key services are available
- Actively developed
  - More services and applications are being developed

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.