



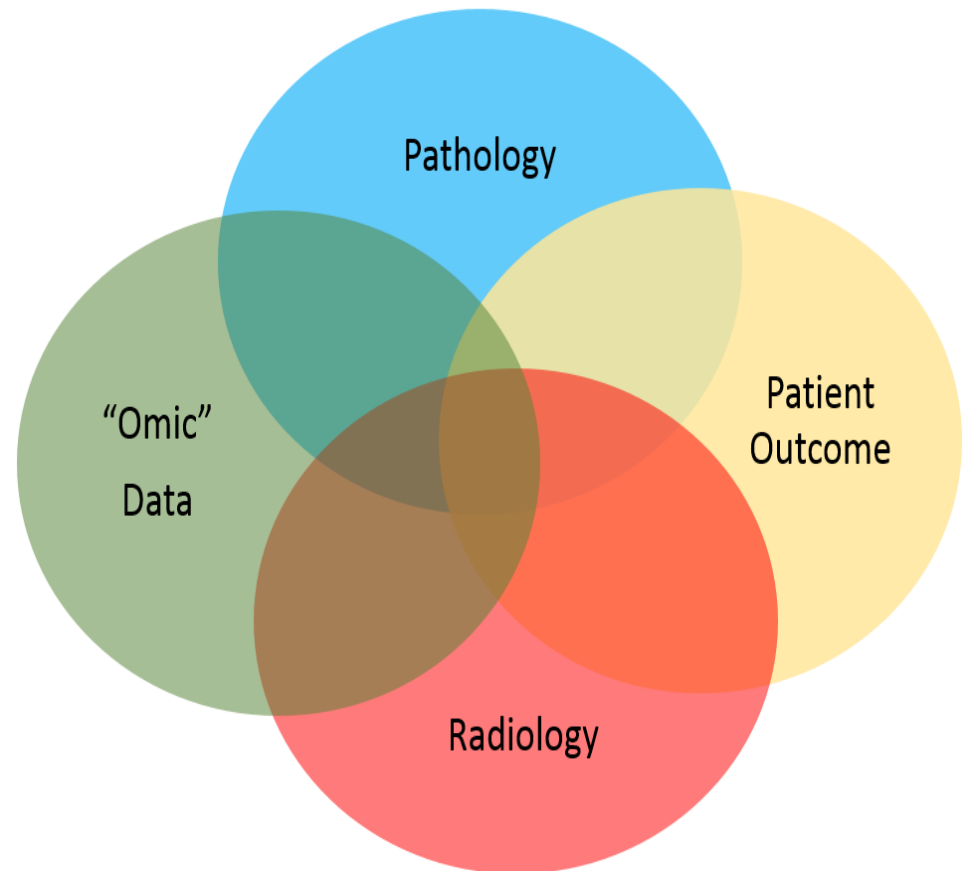
Tools to Analyze Morphology and Spatially Mapped Molecular Data

Joel Saltz

Department of Biomedical Informatics
Stony Brook University

June 1 2017
ITCR PI Meeting

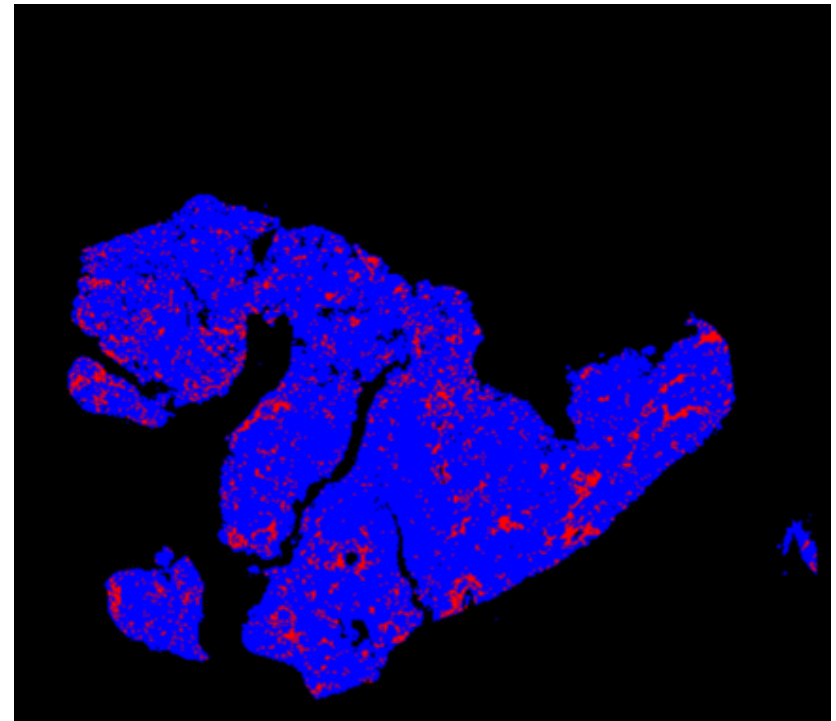
- ***Predict treatment outcome, select, monitor treatments***
- Reduce inter-observer variability in diagnosis
- Computer assisted exploration of new classification schemes
- Tumor heterogeneity, Immune response



Quantitative Assessment of Pathology

- Pathology is ***basis*** of cancer classification
- Inconsistent Pathology classification will confound any clinical research study
- Quantitation increasingly important- TILS, Pathology subtype %, mitoses
- Use of histological features to predict outcome, response to treatment
- Novel microscopy imaging modalities
- Continuum between Radiology and Pathology

Tumor Infiltrating
Lymphocytes
By CNN Deep Learning Method



Pathomics, Radiomics

Identify and segment trillions of **objects** – nuclei, glands, ducts, nodules, tumor niches ... from Pathology, Radiology imaging datasets

Identify and classify tissue **regions** – e.g. tumor infiltrated by lymphocytes

Extract features from objects and spatio-temporal regions

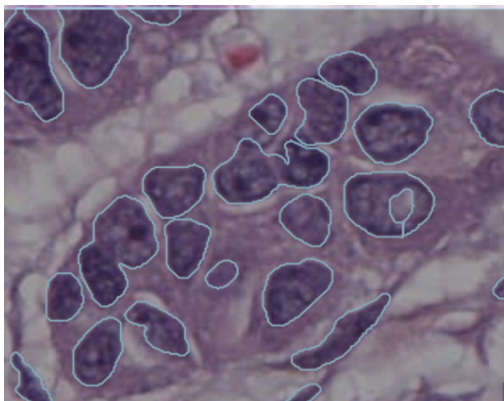
Support queries against ensembles of features extracted from multiple datasets

Statistical analyses and machine learning to link Radiology/Pathology features to “omics” and outcome biological phenomena

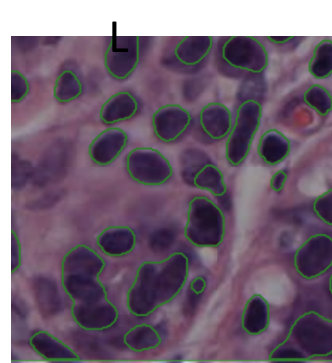
Principle based analyses to bridge spatio-temporal scales – linked Pathology, Radiology studies

EXTRACTED PATHOMIC FEATURES HAVE REAL CLINICAL MEANING

CANCER

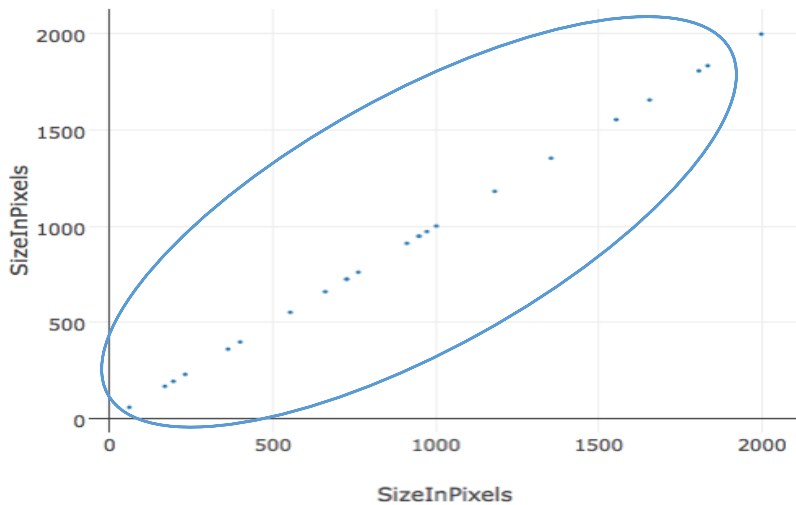


NORMA

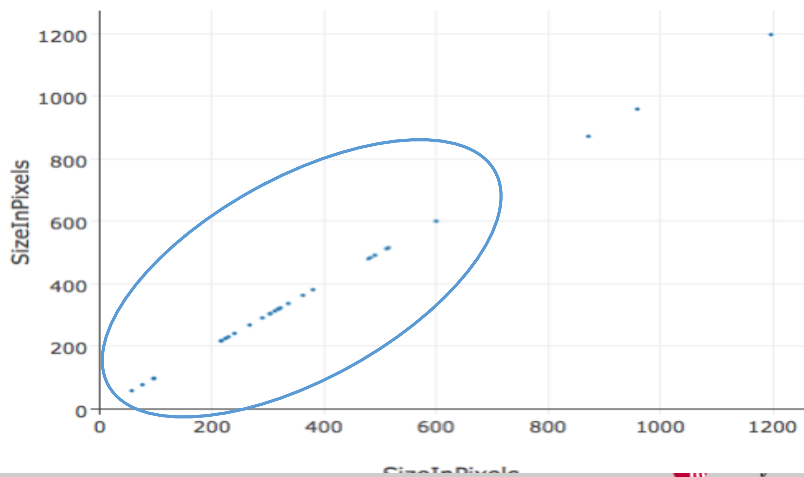


IMAGES FROM THE SAME PATIENT, AS THE SAME MAGNIFICATION

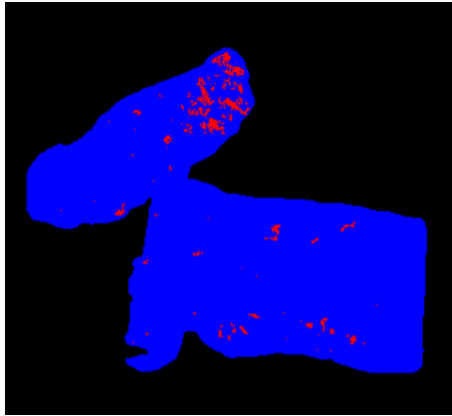
CANCER NUCLEI larger size range



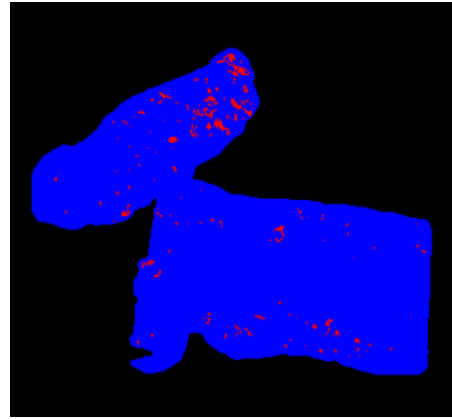
NORMAL NUCLEI smaller size range



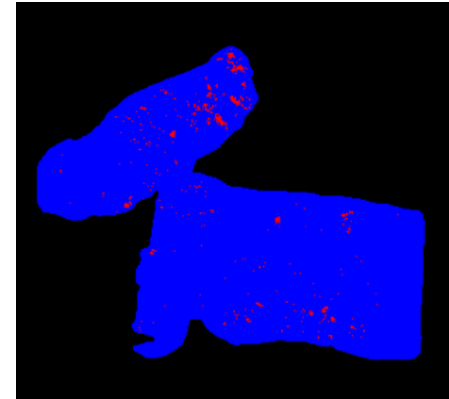
Tissue Infiltrating Maps – TCGA Pan Cancer Atlas Project



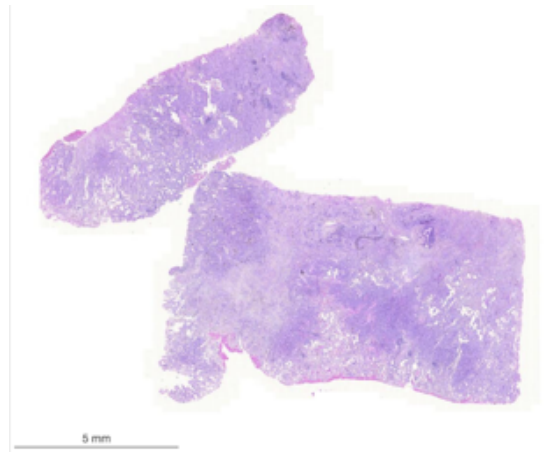
Prediction edited by
Pathologist 1



Prediction without
editing



Prediction edited by
Pathologist 2



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

- **Specific Aim 1** Analysis **pipelines** for multi- scale, integrative image analysis, classification.
- **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.

Software Tools

- Suite of software tools
 - caMicroscope: Visualization of tissue images and tissue segmentation results
 - FeatureDB: Management of imaging features
 - FeatureVis: Visual analytics for exploration of imaging features
 - Deep Viz: Deep Learning Visual Training Environment

SEER Virtual Tissue Repository

- FDA approval for diagnosis use of WSIs will enable collection and analysis of WSI information *for all cancer patients*
- Current project -- 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
- Accrue linked clinical data, Pathology slides from SEER sites

SEER Virtual Tissue Repository

- 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

Vision –Population/epidemiological cancer research using rich cancer phenotype information available from Pathology tissue studies

Pathology is central to cancer diagnosis – primary cancer phenotype information

NCIP/Leidos 14X138 and HHSN261200800001E - NCI

SEER VIRTUAL TISSUE REPOSITORY

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

Image analysis methods for estimating TCGA tumor lymphocyte infiltration: TCGA Pan Cancer Atlas

Stony Brook ITCR

- Joel Saltz
- Tahsin Kurc
- Ashish Sharma
- Dimitri Samaras

MD Anderson

- Alex Lazar
- Arvind Rao

Emory ITCR Collaborator

- Lee Cooper

Institute for Systems Biology

- Vesteinn Thorsson
- Ilya Shmulevich
- Sheila Reynolds

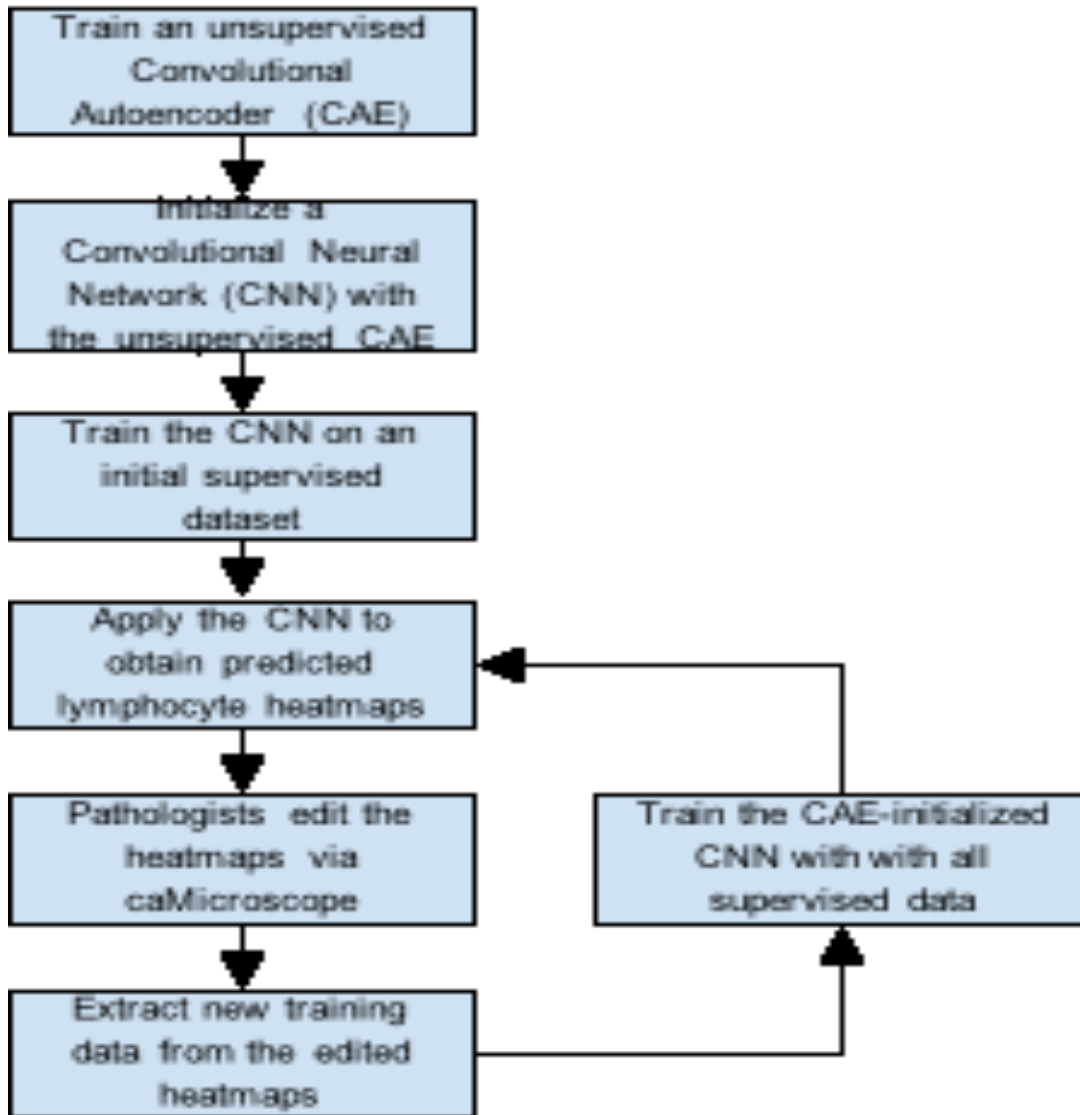
Anatomic Pathologists

- John Van Arnam
- Anne Zhao
- Rebecca Batiste

Graduate Students

- Le Hou
- Vu Nyugen

Iterative Image Based Deep Learning Training



- Training results generalize across many tissue types
- Initial training on patches (20K)
- Iterative training using visual tool
- Algorithm generates prediction, pathologist corrects and feeds back
- Patch based sampling for threshold adjustment

Comparison with Molecular lymphocyte estimates

Estimate leukocyte fraction from methylation

CIBERSORT to identify lymphocyte subset:

B.cells.naive, B.cells.memory, T.cells.CD8,
T.cells.CD4.naive, T.cells.CD4.memory.resting,
T.cells.CD4.memory.activated,
T.cells.follicular.helper, T.cells.regulatory..Tregs,
T.cells.gamma.delta, NK.cells.resting,
NK.cells.activated

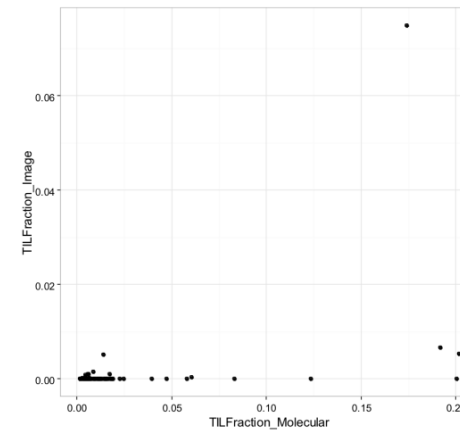
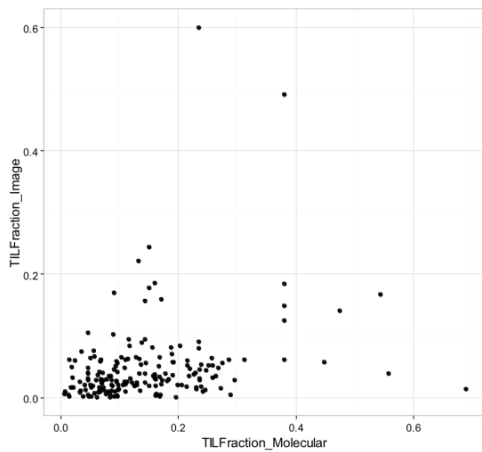
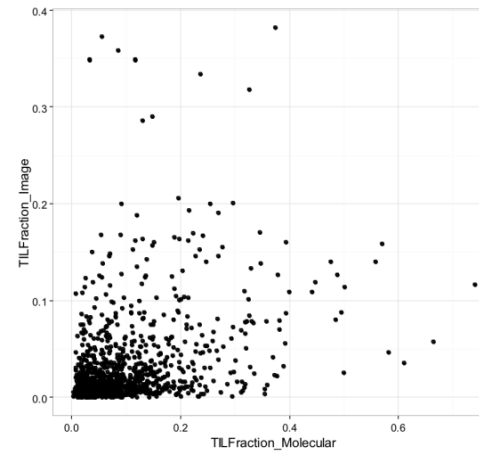
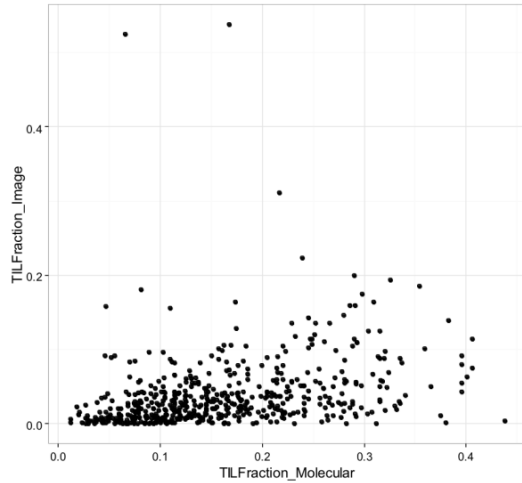
Compare with lymphocyte fraction obtained from TIL
image analysis

Comparisons between Molecular and Imaging TIL

Data subset – we have now completed TIL analysis for 14 TCGA tumor types

Tumor Type	Spearman Rank	P value
LUAD	0.46	2.6×10^{-24}
LUSC	0.32	7.6×10^{-12}
BRCA	0.36	1.2×10^{-31}
PAAD	0.33	5.7×10^{-6}
COAD	0.28	1.1×10^{-17}
SKCM	0.49	1.9×10^{-29}
UVM	NOT Significant	
PRAD	0.16	0.002

Molecular vs Imaging TIL Estimates



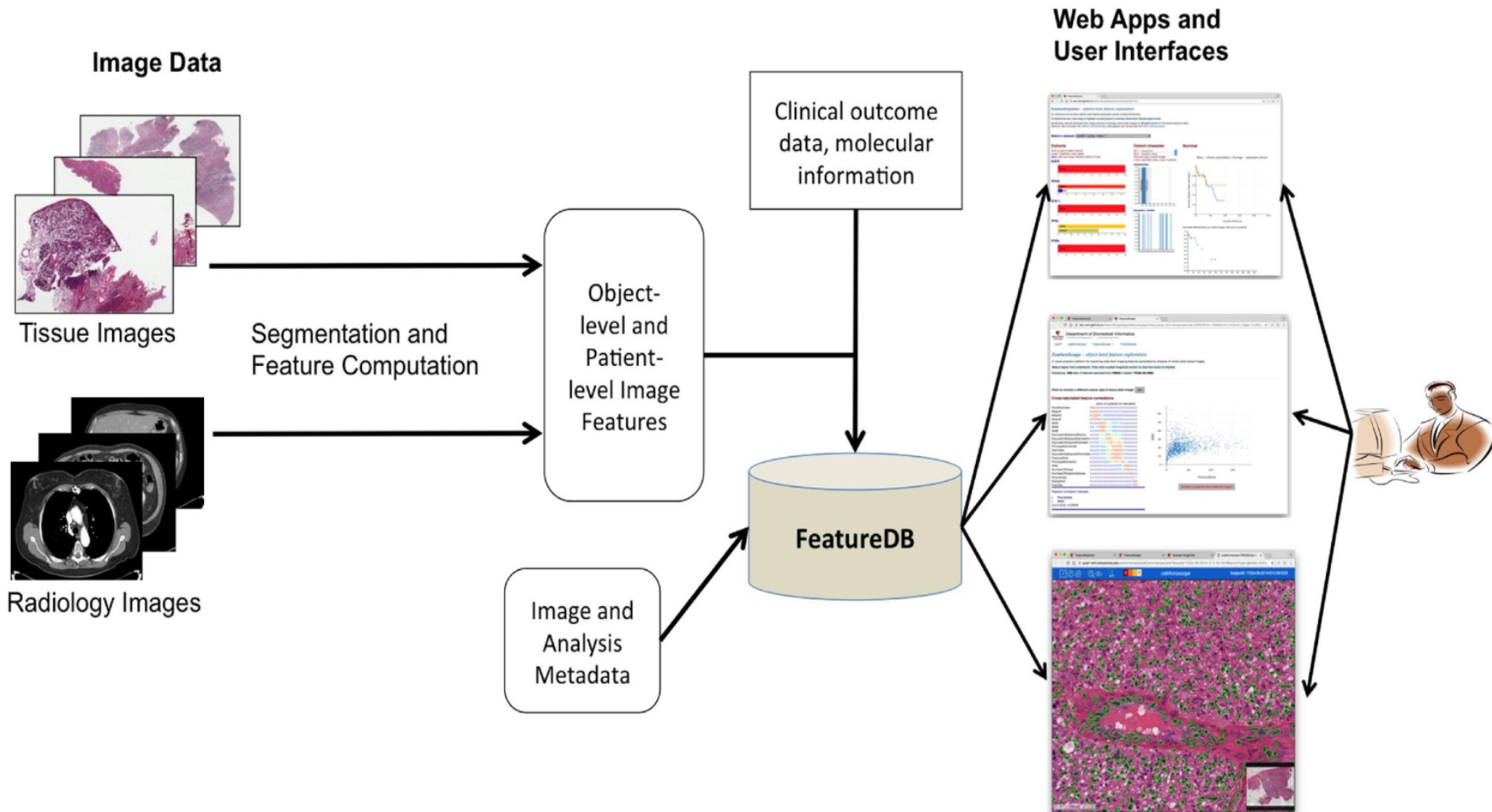
LUAD, BRCA, PAAD and UVM - TIL Image vs TIL Molecular Fraction (left to right by row).

Towards Generation, Management, and Exploration of Combined Radiomics and Pathomics Datasets for Cancer Research

Joel Saltz, Jonas Almeida, Yi Gao, Ashish Sharma, Erich Bremer, Tammy DiPrima, **Mary Saltz**, Jayashree Kalpathy-Cramer, Tahsin Kurc

Won 2017 AMIA Summit Marco Ramoni Award

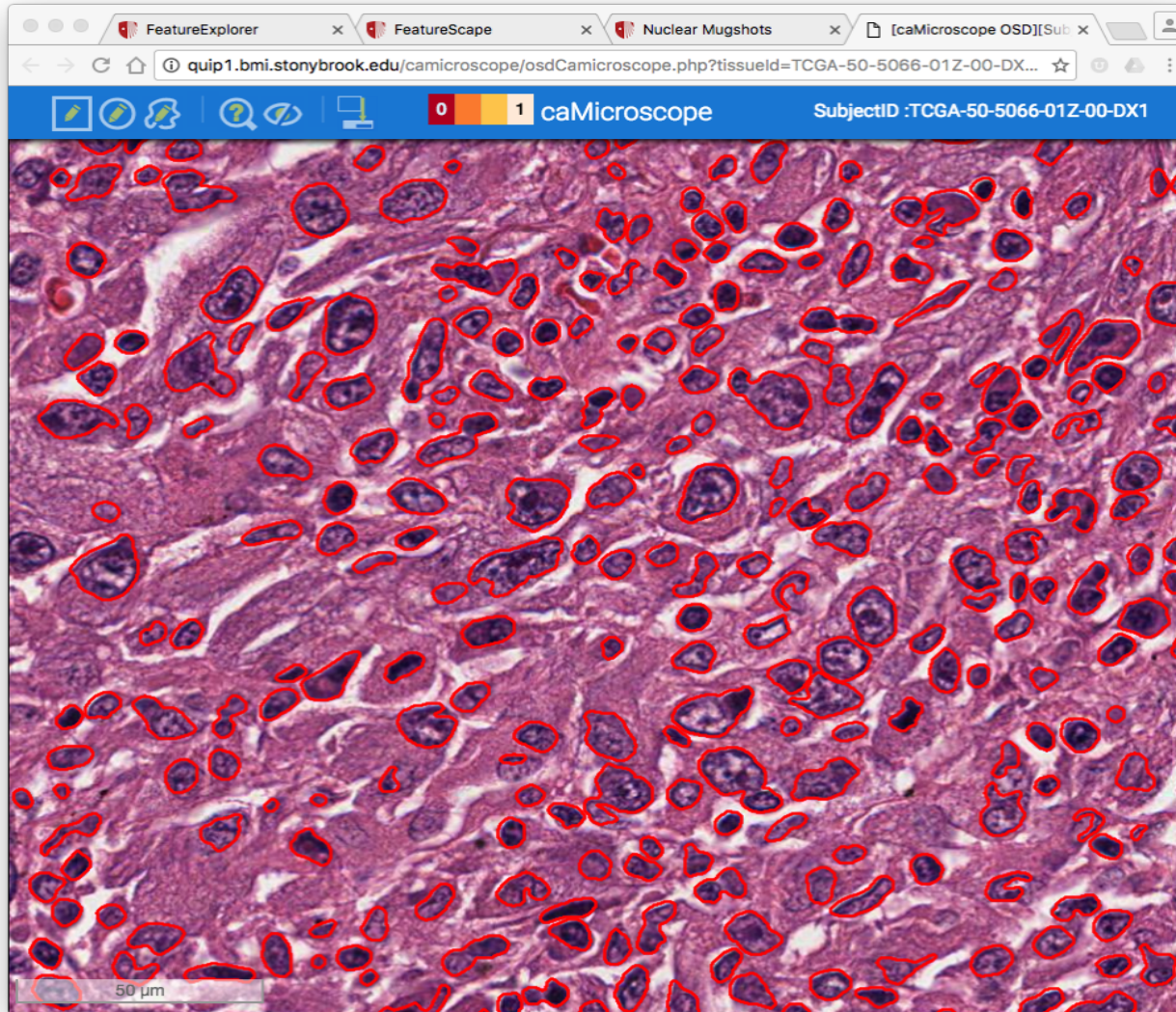
Methods



caMicroscope

- Web-based platform for visualizing digital pathology images with segmentation results and features that are overlaid on the images
- Provides APIs that allow interactive back-and-forth between feature exploration and image visualization

caMicroscope



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

FeatureVis

- Web-based apps for coordinated spatial and feature based visual analytics
- visualization of inter-related imaging features
- Inter-relate collections of features with images and non-imaging data such as gene alteration and survival
- Apps go from feature level to population level and back to individual patients and features

FeatureVis

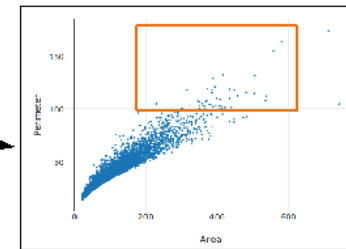
Leveraging Visualization to Aid in Feature Management

- Going from the whole slide data set to selected features and back to the image
- Adding a visual perspective by using a live web-based interactive tool (<http://sbu-bmi.github.io/featurescape/u24/Preview.html>)

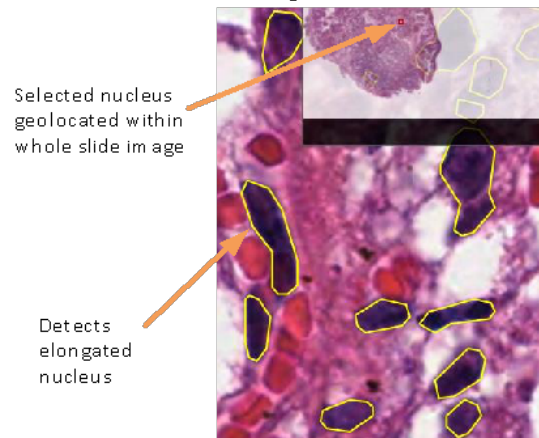
Step 1: Choose a case from the TCGA atlas (case #20)



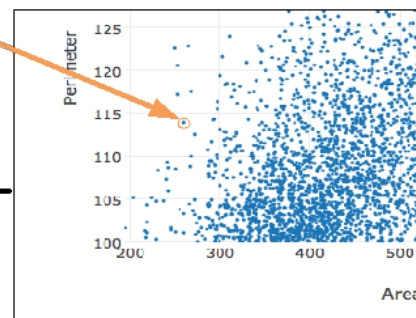
Step 2: Select two features of interest; X axis (*area*), Y axis (*perimeter*)



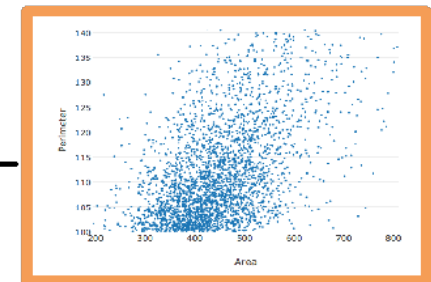
Step 5: Evaluate the features selected in the context of the specific nucleus and where this nucleus is located within the whole slide image



Step 4: Pick a specific nucleus of interest. Each dot represents a single nucleus



Step 3: Zoom in on region of interest



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.

Feature Explorer - Integrated Pathway 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.

EGFR



KRAS



STK11_LKB1



TP53



NF1



BRAF



SETD2



Morphology, Epi, etc

Var 1:

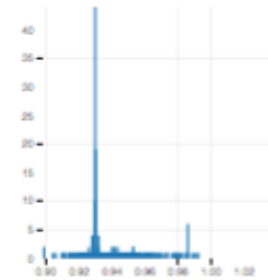
Var 2:

Slide mouse click to select ranges

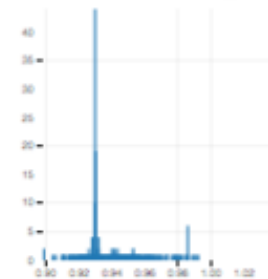
Xaxis: parameter value

Yaxis: #patients

Var 1: Roundness_median



Var 1 Zoom: Roundness_median

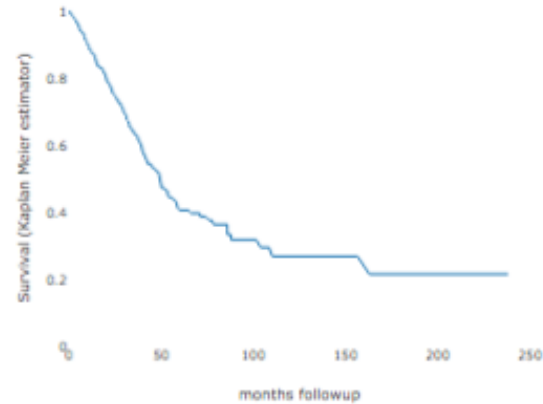


Var 2: StdR_median

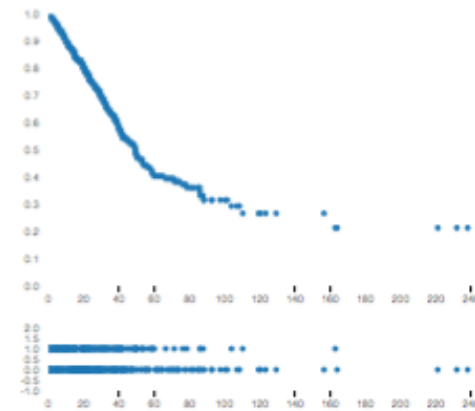


Survival

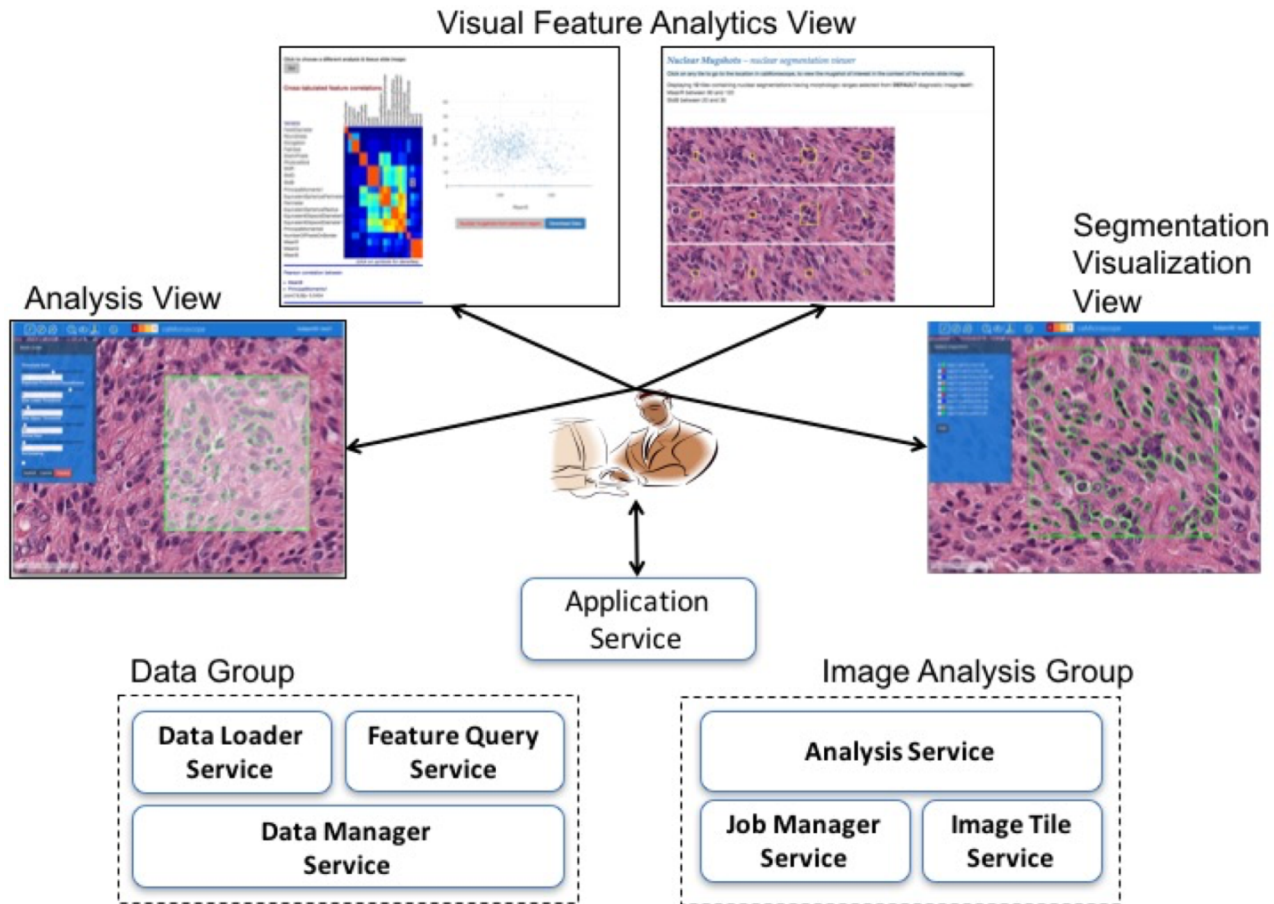
Blue - whole population; Orange - selected cohort



Zoomable KM estimator (i.e. select ranges, each dot is a patient)



Containerization



Container Architecture

- ***The application service group*** - single container that hosts a Web applications to visualize and curate images and analysis results. ***caMicroscope, FeatureScape, Classification Editing Tool, Segmentation Curation Tool***
- ***The data management service group*** is implemented as a set of three containers. These containers are responsible for data loading, data management, and query processing.
- ***The image analysis group*** is made up of three containers, which collectively execute image analysis requests

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!