



Tools to Analyze Morphology and Spatially Mapped Molecular Data

ITCR U24CA180924

Joel Saltz

Department of Biomedical Informatics

Stony Brook University

PI Meeting June 13-14 2016

Video Overview/Demonstration –

<https://www.youtube.com/watch?v=kShwls0O5o4>

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
- M Reda El-MKhantar
- Tara Maloney

Emory University

- Ashish Sharma
- Adam Marcus
- Dan Brat
- Roberd Bostick
- Yue Hou
- Ganesh Iyer

Harvard University

- Richard Cummings

UTK

- Jeremy Logan
- Scott Klasky
- Dave Pugmire

Yale University

- Michael Krauthammer

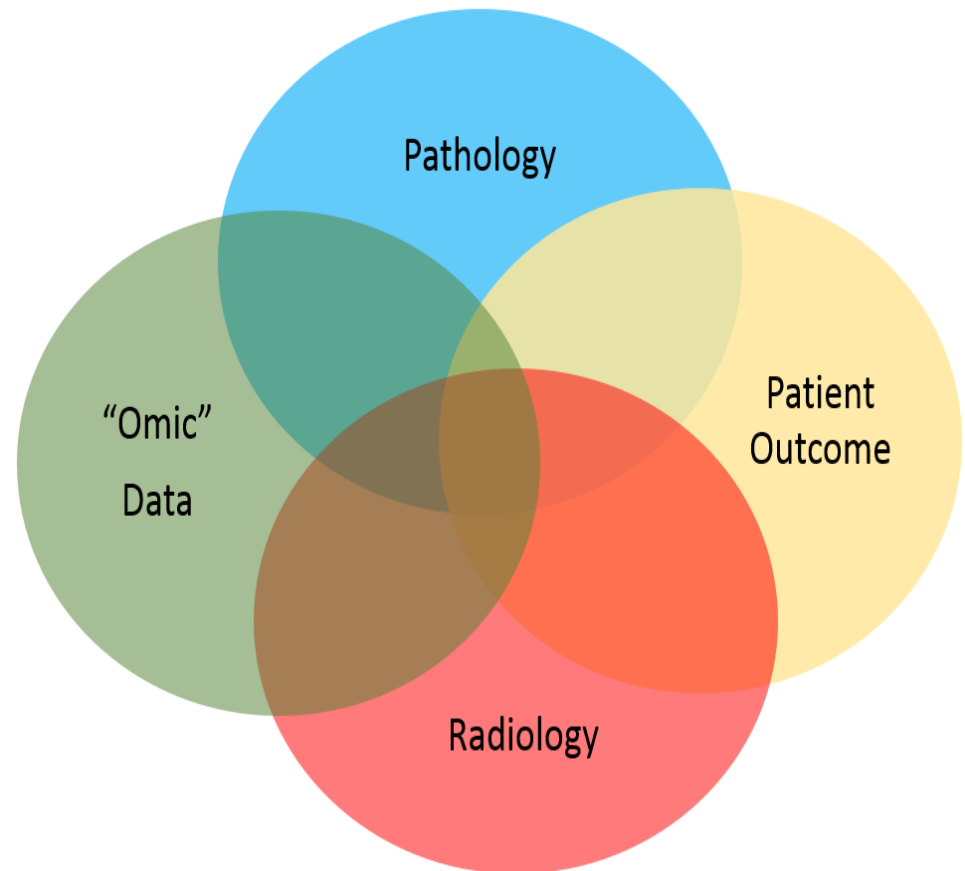
3DSlicer Pathology and FeatureScape for Linked Radiology/Pathology Image Analysis Competitions

Collaboration between SBU, BWH and MGH Groups

- Jonas Almeida, Stony Brook
- Nicole Aucoin, Brigham and Women's Hospital
- Erich Bremer, Stony Brook
- Tammy DiPrima, Stony Brook
- Andrey Fedorov, Brigham and Women's Hospital
- Yi Gao, Stony Brook
- Jayashree Kalpathy-Cramer, Massachusetts General Hospital
- Ron Kikinis, Brigham and Women's Hospital
- Tahsin Kurc, Stony Brook
- Arten Mamonov, Massachusetts General Hospital
- Steve Pieper, Isomics, Inc.
- Jean-Christophe Fillion-Robin, Kitware Inc.
- Bruce Rosen Massachusetts General Hospital
- Allen Tannenbaum, Stony Brook

Multi-scale Integrative Analysis in Biomedical Informatics

- Predict treatment outcome, select, monitor treatments
- Reduce inter-observer variability in diagnosis
- Computer assisted exploration of new classification schemes
- Multi-scale cancer simulations



Pathomics, Radiomics

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

Extract features

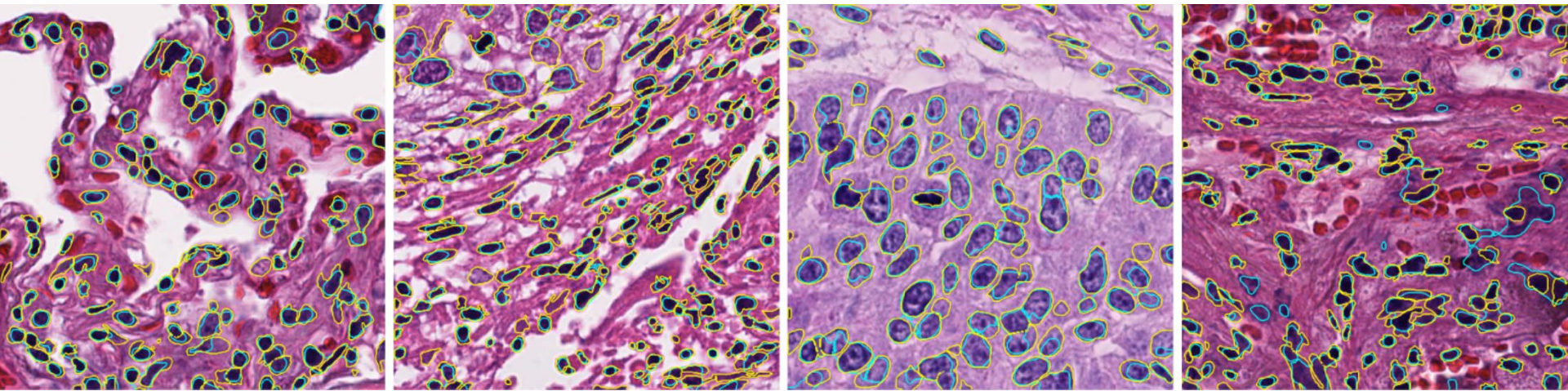
Queries against ensembles of features

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

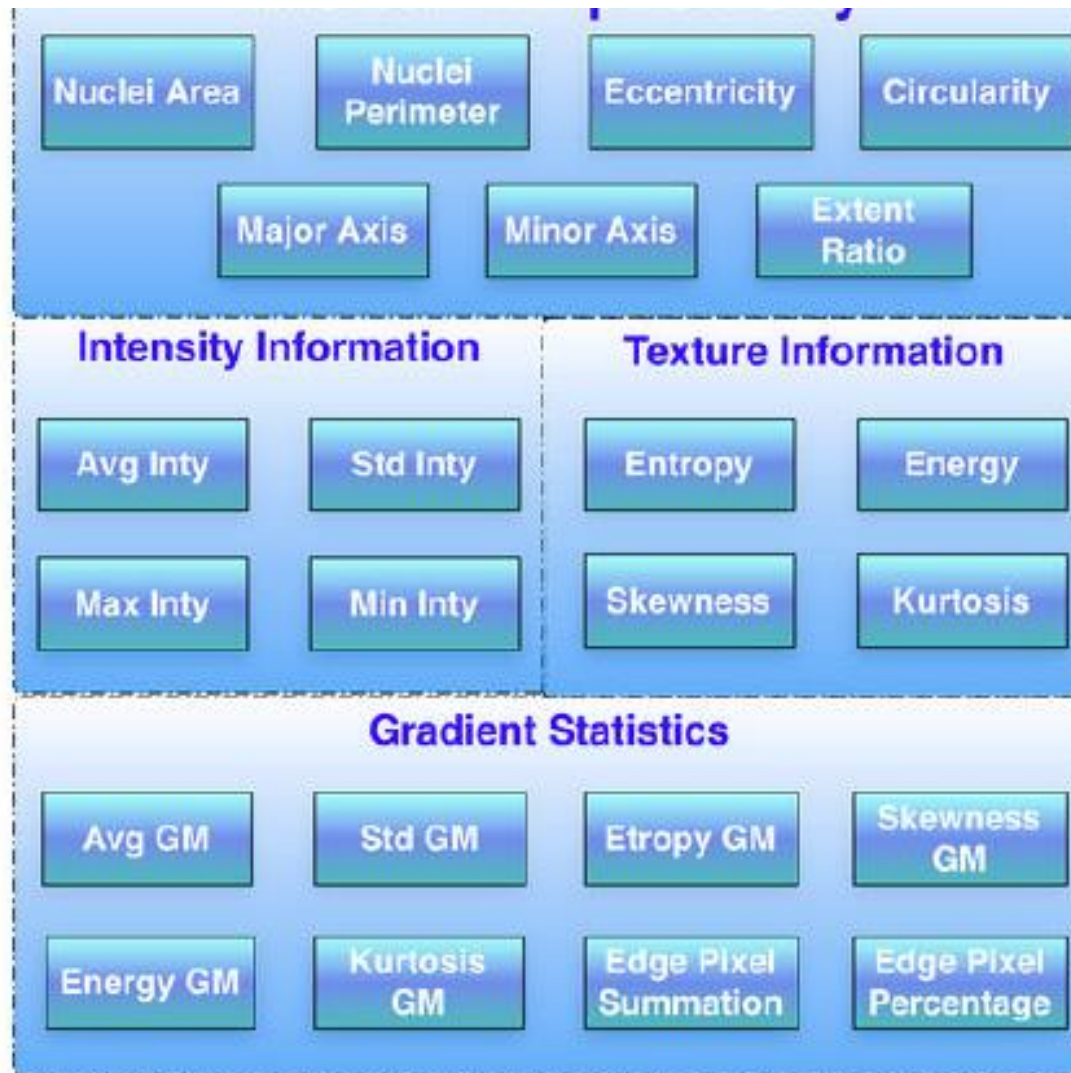
- **Specific Aim 1** Analysis **pipelines** for multi- scale, integrative image analysis.
- **Specific Aim 2:** **Database** infrastructure to manage and query Pathomics features.
- **Specific Aim 3:** HPC software that **targets clusters, cloud computing, and leadership scale systems.**
- **Specific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

Robust Nuclear Segmentation

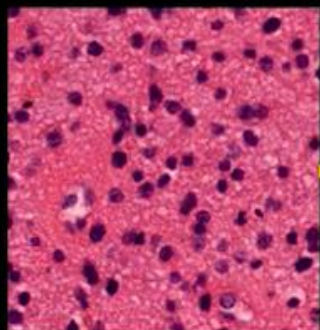
- Robust ensemble algorithm to segment nuclei across tissue types
- Parameter exploration to optimize quality
- Systematic Quality Control pipeline encompassing tissue image quality, human generated ground truth, convolutional neural network critique
- Yi Gao, Allen Tannenbaum, Dimitris Samaras, Le Hou,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



Cell Morphometry Features



Whole Slide Images (WSI)



Segmentation
Parameters

Compute Cluster



Process the images for subjects
selected

Compute object-level (nucleus-
level) image features

Compute aggregated patient-
level image features from
object-level features

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)

FeatureExplorer

Explore relationships between patient-level imaging features and gene mutations and survival patterns.

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.

EGFR



KRAS



STK11_LKB1



TP53



NF1



BRAF



SETD2



Morphology, Epi, etc

Var 1: Roundness_median

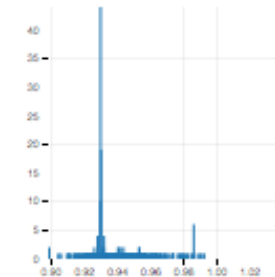
Var 2: StdR_median

Slide mouse click to select ranges

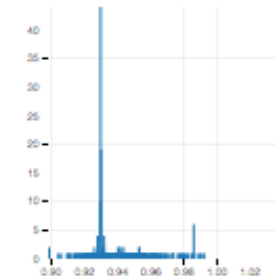
Xaxis: parameter value

Yaxis: #patients

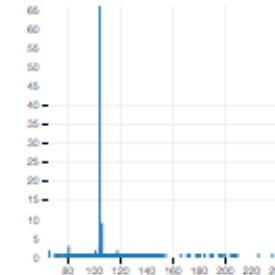
Var 1: Roundness_median



Var 1 Zoom: Roundness_median

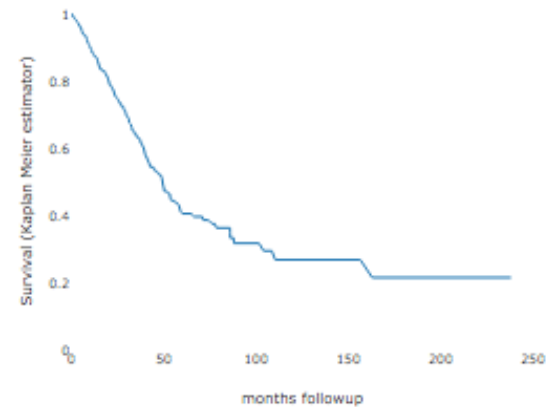


Var 2: StdR_median

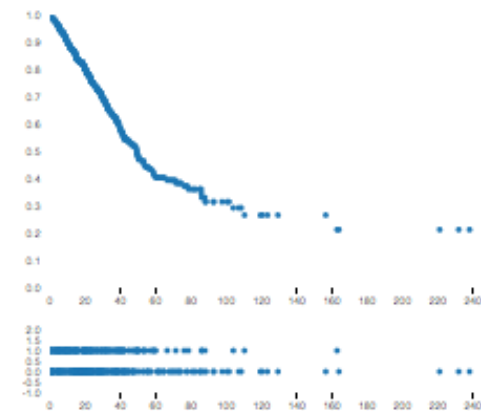


Survival

Blue - whole population; Orange - selected cohort



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



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.

EGFR



KRAS



STK11_LKB1



TP53



NF1



BRAF



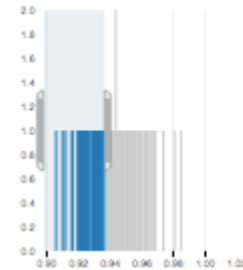
SETD2



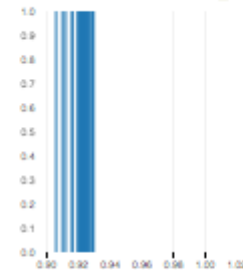
Morphology, Epi, etc

Var 1: Roundness_median
Var 2: StdR_median
Slide mouse click to select ranges
Xaxis: parameter value
Yaxis: #patients

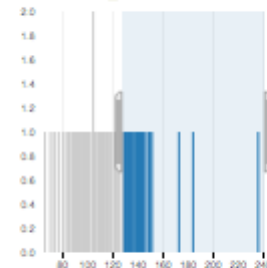
Var 1: Roundness_median



Var 1 Zoom: Roundness_median

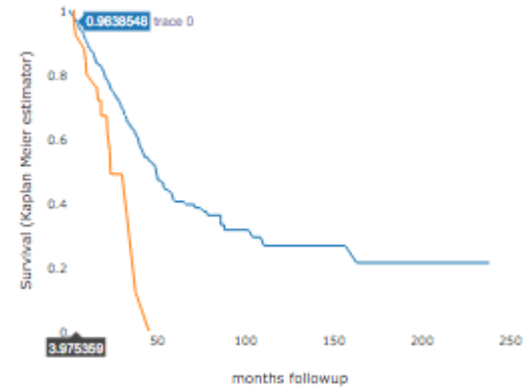


Var 2: StdR_median

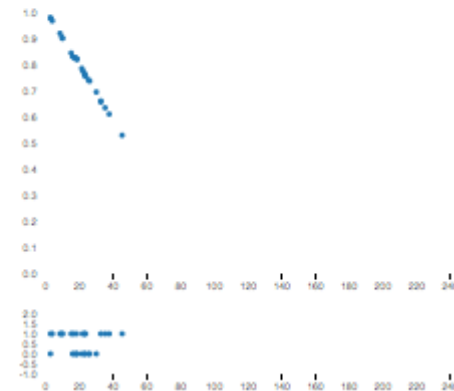


Survival

Blue - whole population; Orange - selected cohort



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



Collaboration with MGH – Feature Explorer – Radiology Brain MR/Pathology Features

Select a dataset: GBM - Glioblastoma Multiforme - rad:path:ver1

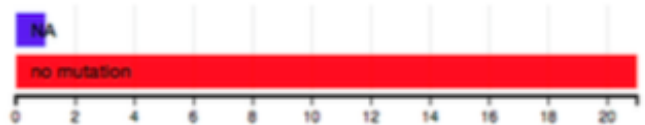
Gene Mutation

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

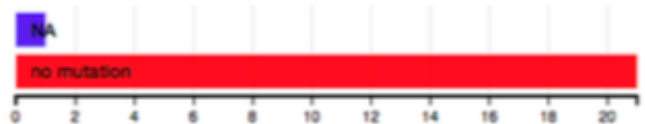
EGFR



KRAS



STK11

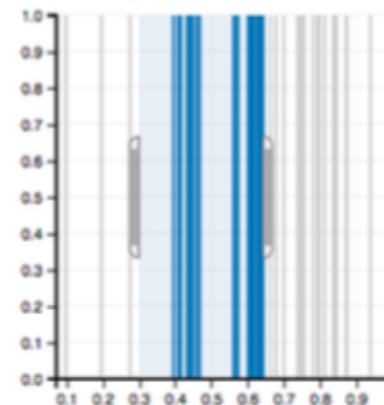


TP53

Cohort character

Var 1: edemaRatio_4_T2
Var 2: Flatness_median
Slide mouse click to select ranges
Xaxis: parameter value
Yaxis: #patients

edemaRatio_4_T2

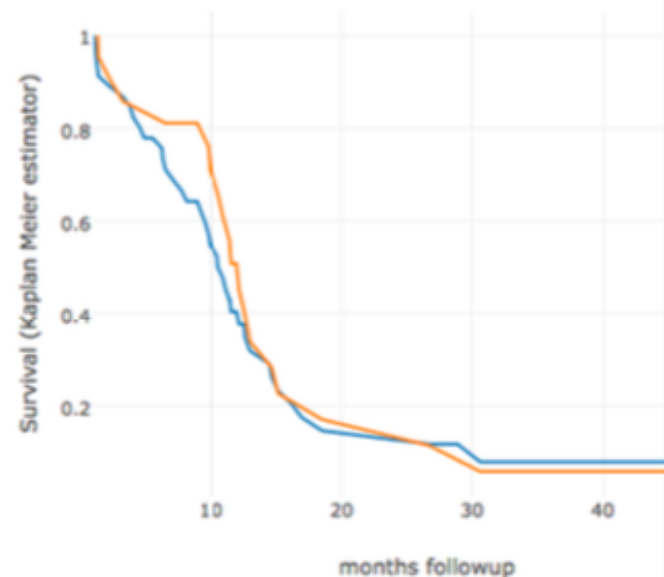


Flatness_median



Survival

Blue - whole population; Orange - selected



Collaboration with SBU Radiology – TCGA NSCLC Adeno Carcinoma

Integrative Radiology, Pathology, “omics”, outcome

Gene Mutation

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

EGFR



KRAS



PTEN



STK11



TP53



Rad, Path, Epi, etc

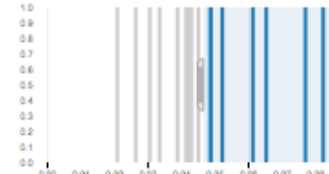
Var 1: Roundness_50

Var 2: compactness

Var 3: age_at_initial_pathologic

Slide mouse click to select ranges
Xaxis: parameter value, Yaxis: #patients

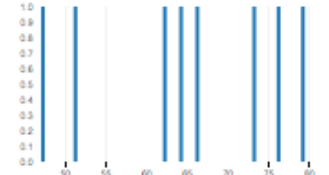
Var 1: Roundness_50



Var 2: compactness

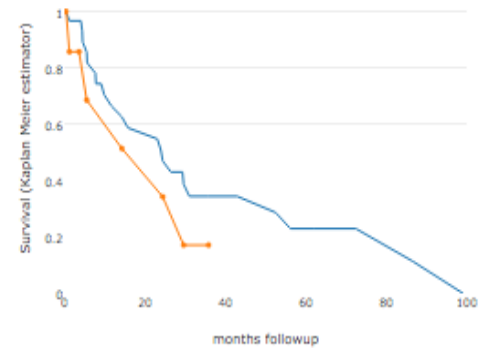


Var 3: age_at_initial_pathologic_diagnosis

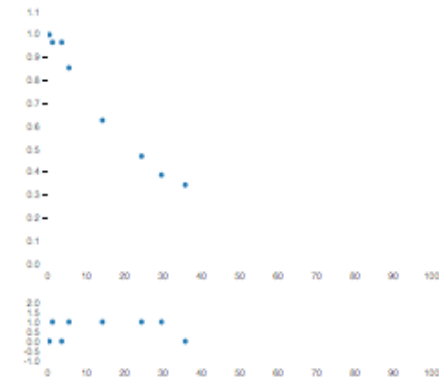


Survival

Blue - whole population; Orange - selected cohort



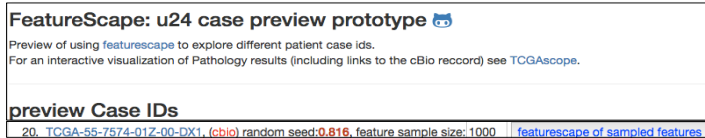
Zoomable KM estimator (i.e. select ranges, each dot is a patient)



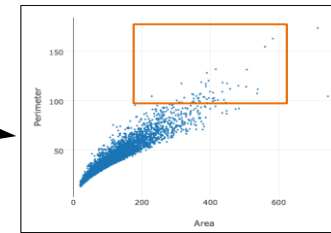
Feature Scope

Relationship Between Image and Features

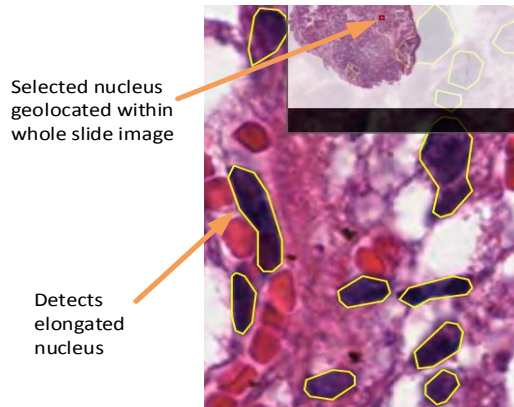
Step 1: Choose a case from the TCGA atlas (case #20)



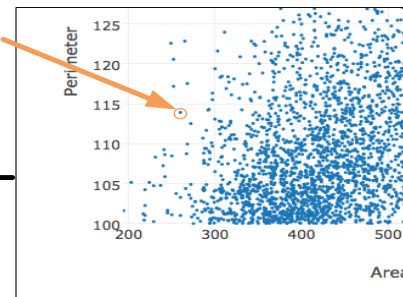
Step 2: Select two features of interest; x axis (area), Y axis (perimeter)



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.

FeatureScape: u24 case preview prototype

Preview of using [featurescape](#) to explore different patient case ids.

For an interactive visualization of Pathology results (including links to the cBio record) see [TCGAscope](#).

preview Case IDs

1. TCGA-34-2605-01Z-00-DX1, (cbio) random seed:0.559, feature sample size: 1000	featurescape of sampled features
2. TCGA-38-4625-01Z-00-DX1, (cbio) random seed:0.628, feature sample size: 1000	featurescape of sampled features
3. TCGA-38-4626-01Z-00-DX1, (cbio) random seed:0.700, feature sample size: 1000	featurescape of sampled features
4. TCGA-38-4628-01Z-00-DX1, (cbio) random seed:0.016, feature sample size: 1000	featurescape of sampled features
5. TCGA-38-4629-01Z-00-DX1, (cbio) random seed:0.185, feature sample size: 1000	featurescape of sampled features
6. TCGA-38-6178-01Z-00-DX1, (cbio) random seed:0.317, feature sample size: 1000	featurescape of sampled features
7. TCGA-38-A44F-01Z-00-DX1, (cbio) random seed:0.906, feature sample size: 1000	featurescape of sampled features
8. TCGA-50-5044-01Z-00-DX1, (cbio) random seed:0.055, feature sample size: 1000	featurescape of sampled features
9. TCGA-50-5045-01Z-00-DX1, (cbio) random seed:0.946, feature sample size: 1000	featurescape of sampled features
10. TCGA-50-5045-01Z-00-DX2, (cbio) random seed:0.551, feature sample size: 1000	featurescape of sampled features
11. TCGA-50-5055-01Z-00-DX1, (cbio) random seed:0.127, feature sample size: 1000	featurescape of sampled features
12. TCGA-34-5232-01Z-00-DX1, (cbio) random seed:0.208, feature sample size: 1000	featurescape of sampled features
13. TCGA-50-5055-01Z-00-DX2, (cbio) random seed:0.321, feature sample size: 1000	featurescape of sampled features
14. TCGA-50-5066-01Z-00-DX1, (cbio) random seed:0.711, feature sample size: 1000	featurescape of sampled features
15. TCGA-50-5066-02Z-00-DX1, (cbio) random seed:0.008, feature sample size: 1000	featurescape of sampled features
16. TCGA-50-5942-01Z-00-DX1, (cbio) random seed:0.031, feature sample size: 1000	featurescape of sampled features
17. TCGA-50-5946-01Z-00-DX1, (cbio) random seed:0.768, feature sample size: 1000	featurescape of sampled features
18. TCGA-50-6590-01Z-00-DX1, (cbio) random seed:0.668, feature sample size: 1000	featurescape of sampled features
19. TCGA-50-6591-01Z-00-DX1, (cbio) random seed:0.498, feature sample size: 1000	featurescape of sampled features

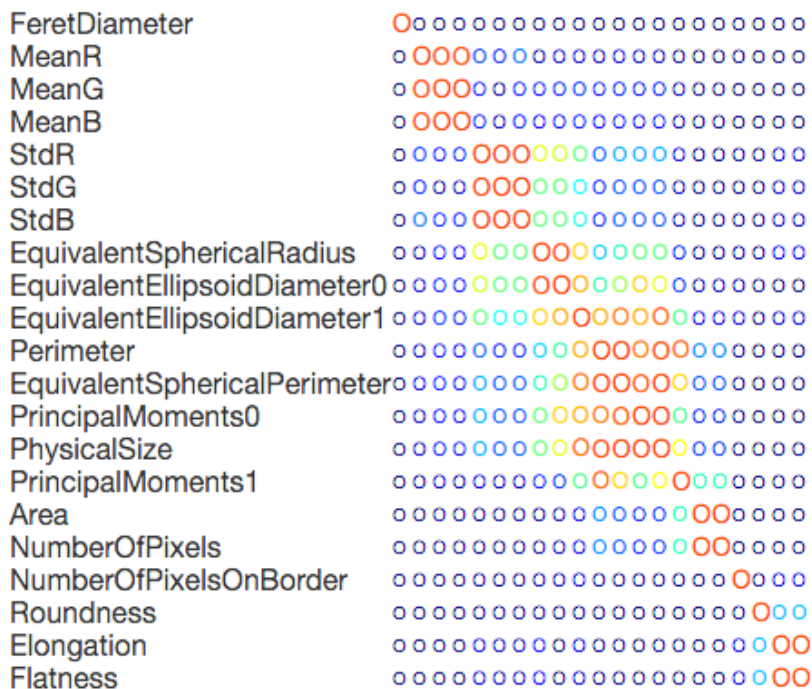
← Preliminary demo of integrative use of multiple FeatureScape tools

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}](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

(click on symbols for densities)



FeatureScape

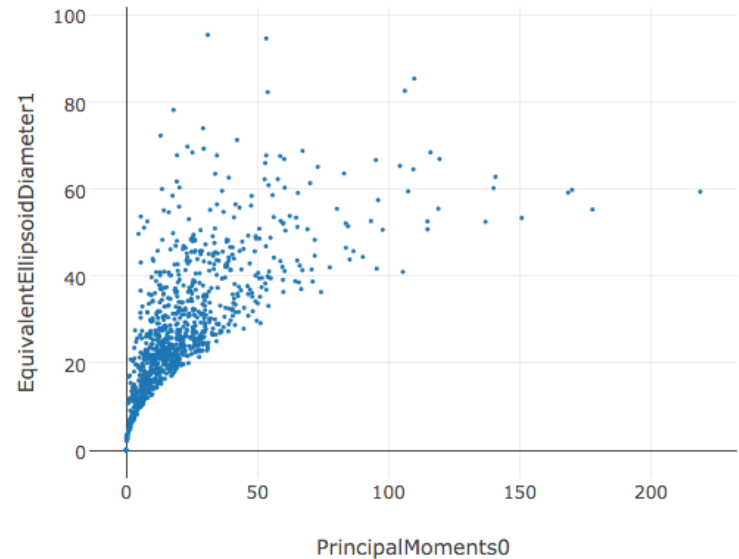
1000 entries sampled from <https://tahsin175.informatics.stonybrook.edu:4500/?limit=1000&find={%22randval%22:>

```
{%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

Feature	Color
FeretDiameter	Orange
MeanR	Orange
MeanG	Orange
MeanB	Orange
StdR	Blue
StdG	Blue
StdB	Blue
EquivalentSphericalRadius	Blue
EquivalentEllipsoidDiameter0	Blue
EquivalentEllipsoidDiameter1	Blue
Perimeter	Blue
EquivalentSphericalPerimeter	Blue
PrincipalMoments0	Blue
PhysicalSize	Blue
PrincipalMoments1	Blue
Area	Blue
NumberOfPixels	Blue
NumberOfPixelsOnBorder	Blue
Roundness	Blue
Elongation	Blue
Flatness	Blue



Pearson correlation between

- PrincipalMoments0
- EquivalentEllipsoidDiameter1

Resample from selected region (under development)

FeatureScape

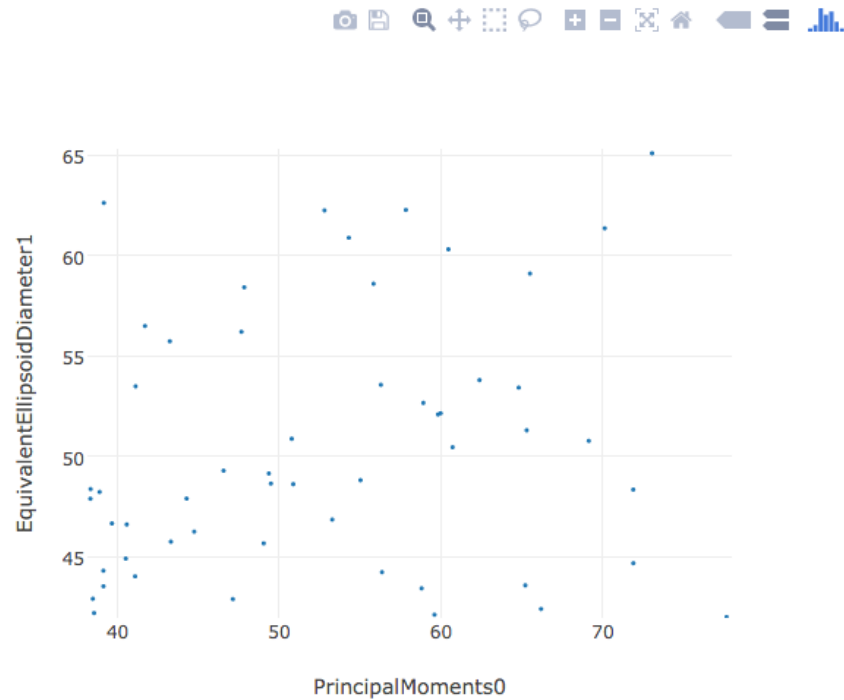
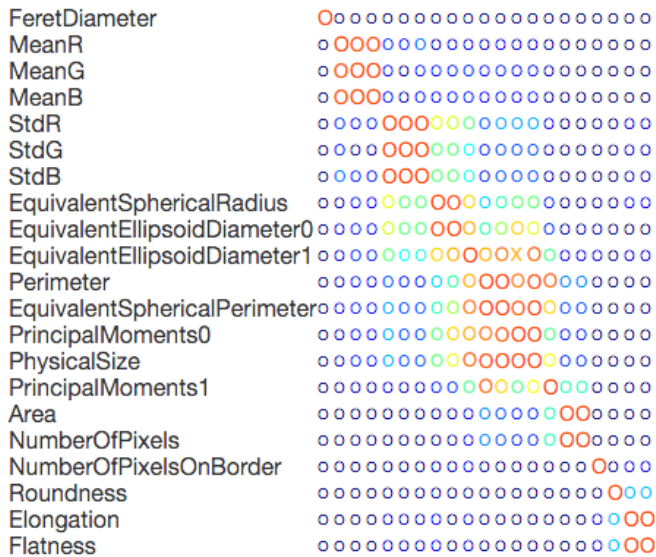
← Preliminary demo of integrative use of multiple FeatureScape tools

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

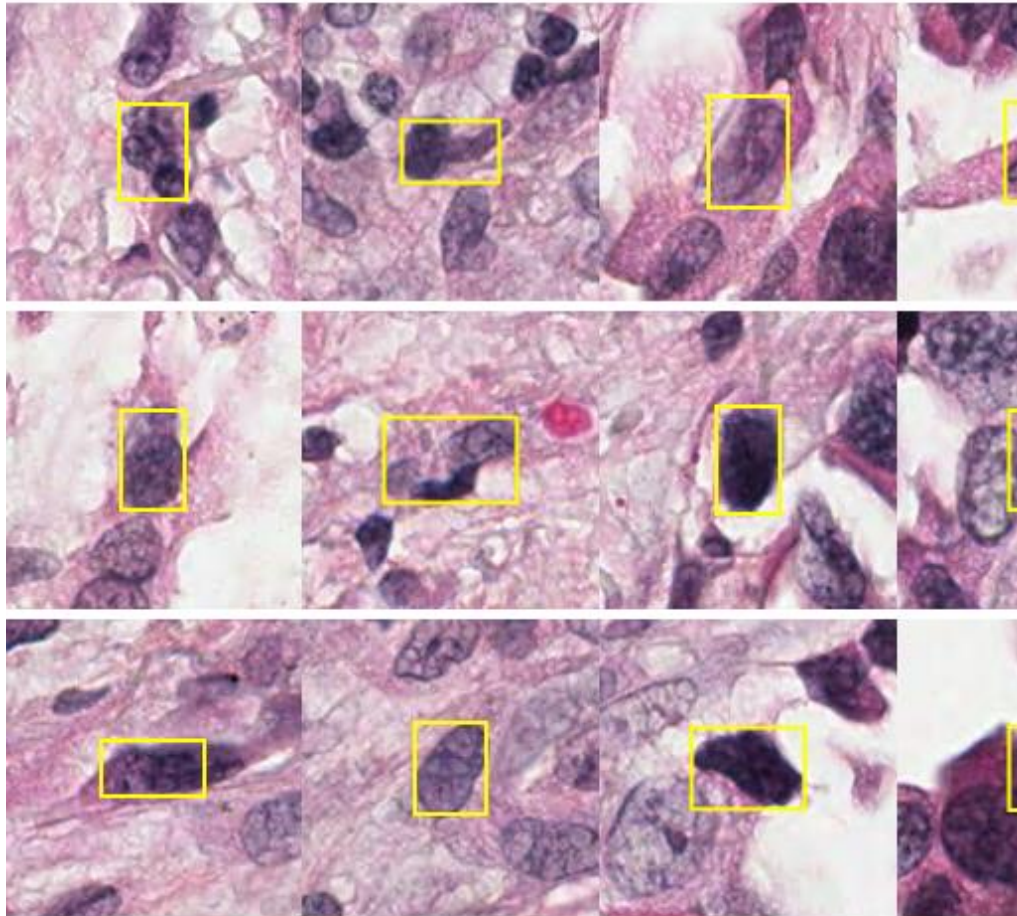


Pearson correlation between

- PrincipalMoments0
- EquivalentEllipsoidDiameter1

Resample from selected region (under development)

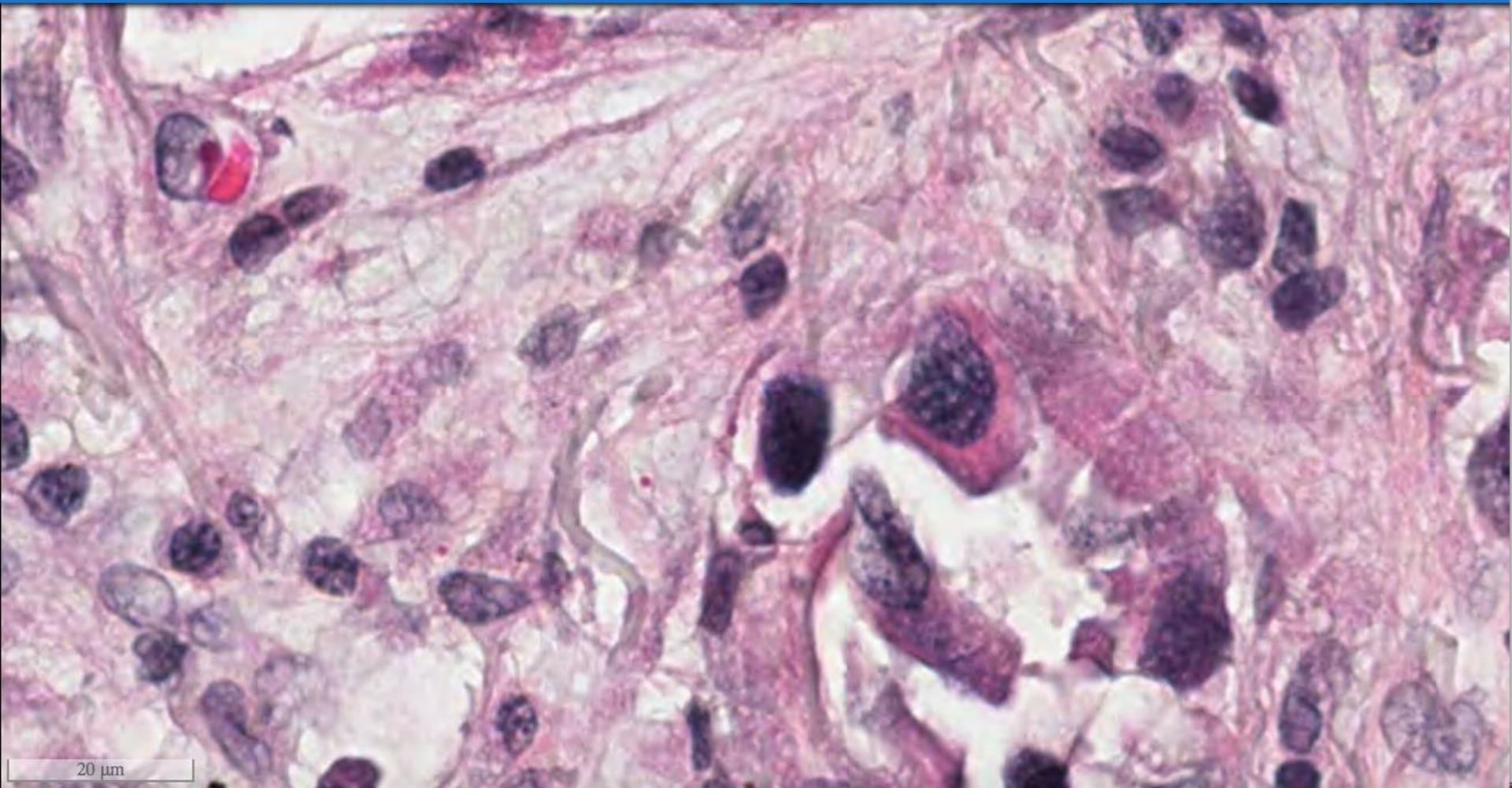
Hover over the smaller patch containing a nucleus to view it larger.
The larger patch underneath shows the nucleus in context.
Click on either patch to go to the location in caMicroscope.





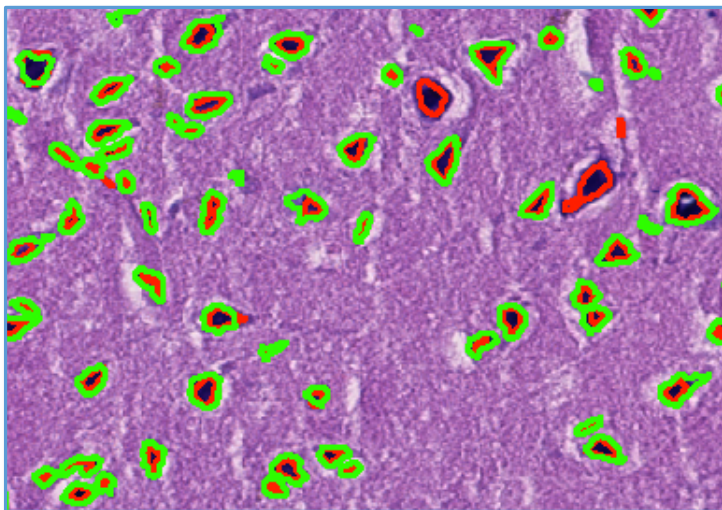
caMicroscope

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



- High quality image analysis algorithms are essential to support biomedical research and diagnosis
 - Validate algorithms with human annotations
 - Compare and consolidate different algorithm results

e.g.: what are the distances and overlap ratios between markup boundaries from two algorithms?



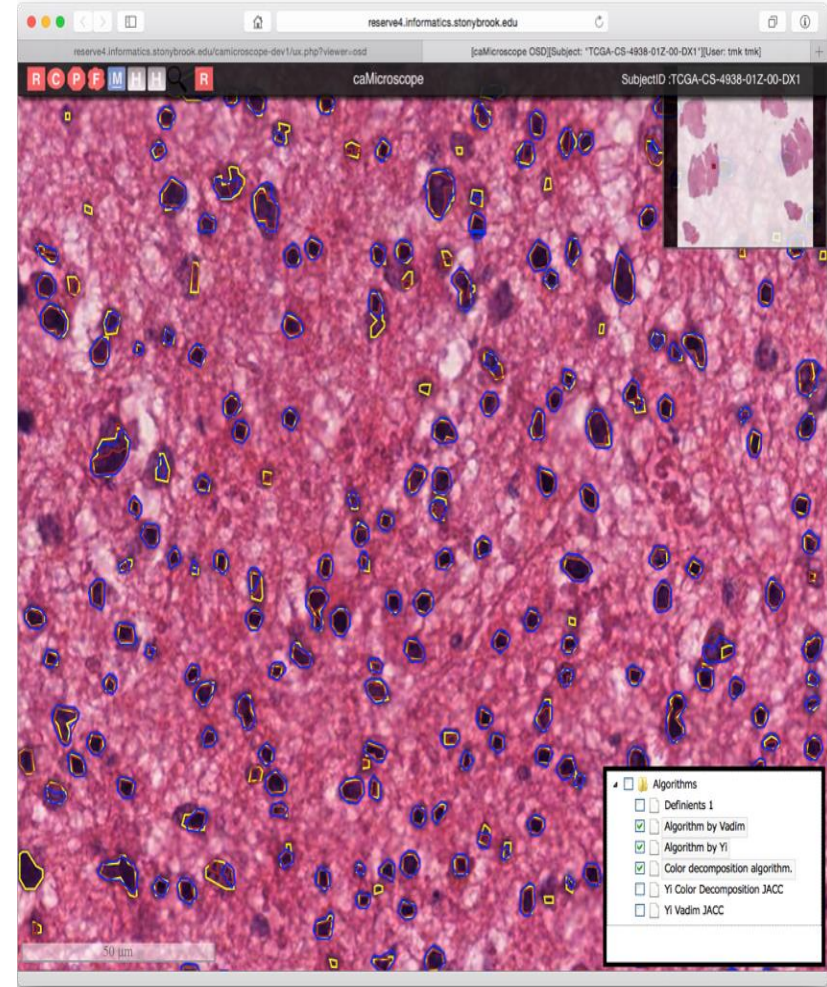
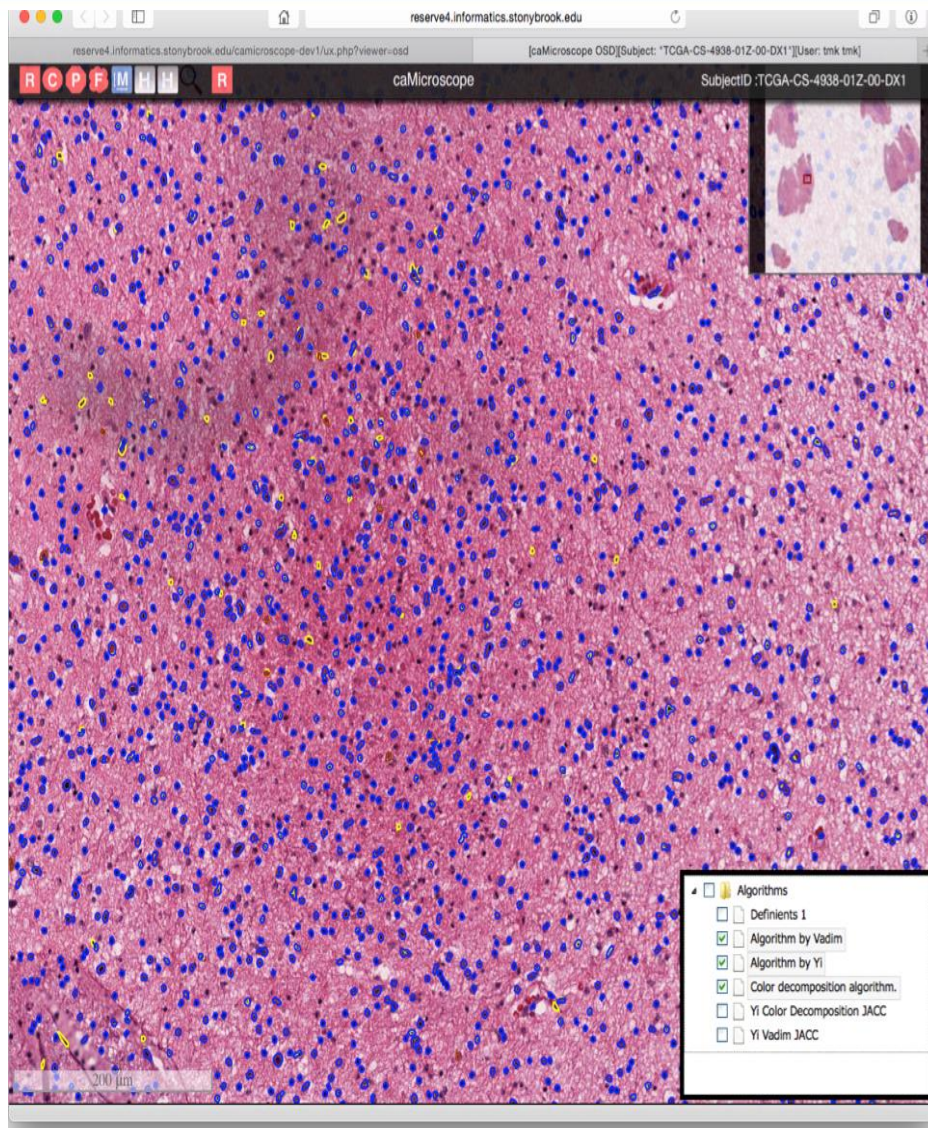
Cross matching of two spatial data sets

Green: algorithm 1

Red: algorithm 2



caMicroscope/MongoDB - Multiple Algorithm Comparison; Generate and Curate Pathomics Feature set

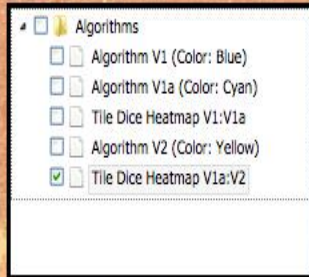


Heatmap – Depicts Agreement Between Algorithms

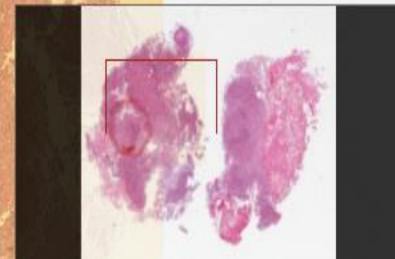


caMicroscope

SubjectID :TCGA-12-1090-01Z-00-DX1



1 mm



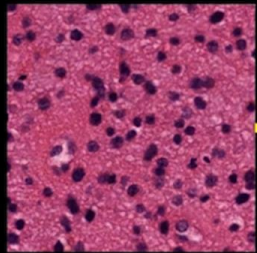
3D Slicer Pathology – Generate High Quality Ground Truth

ITCR - Tools to Analyze Morphology and Spatially Mapped Molecular Data



3D Slicer Pathology

Whole Slide Images (WSI)



Tune algorithm parameters to generate good segmentation results for selected patches

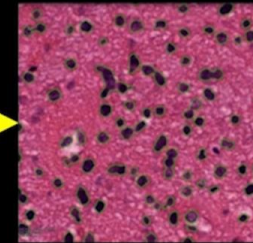
Semi-automatic Segmentation

Manual Segmentation

Create Training Data

Automatic Parameter Tuning

Segmented Results

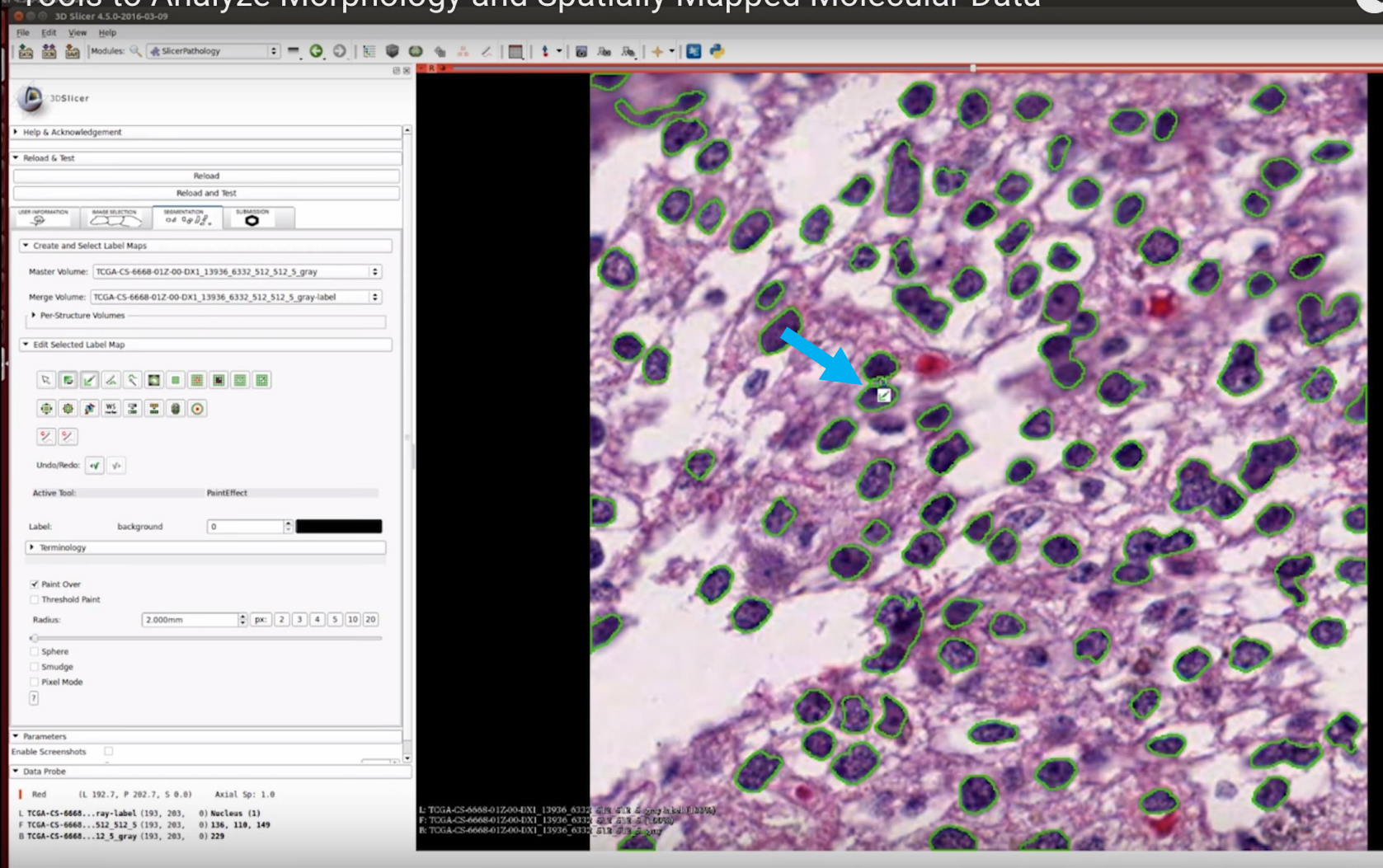


ITCR: Tools to Analyze Morphology and Spatially Mapped Molecular Data



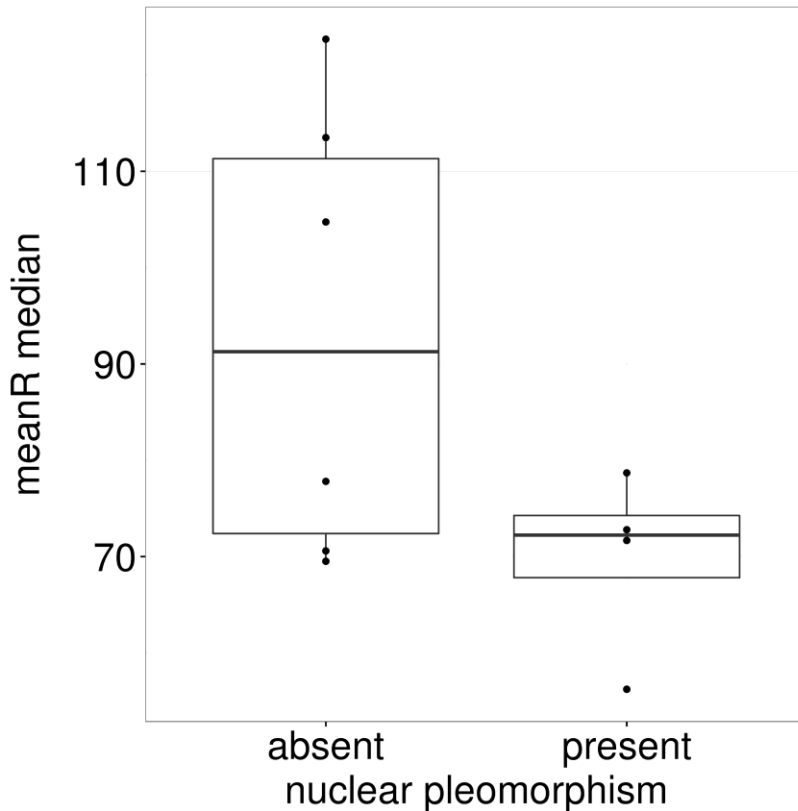
Adjust algorithm parameters, manual fine tuning

ITCR - Tools to Analyze Morphology and Spatially Mapped Molecular Data



Integrated feature, text extraction

GBM – Text/Image Features



Collaboration with TIES group

1. Docker based Segmentation/FeatureDB
2. TIES plus text feature extractor
3. SBU ITCR + Balasubramanian, Lynn

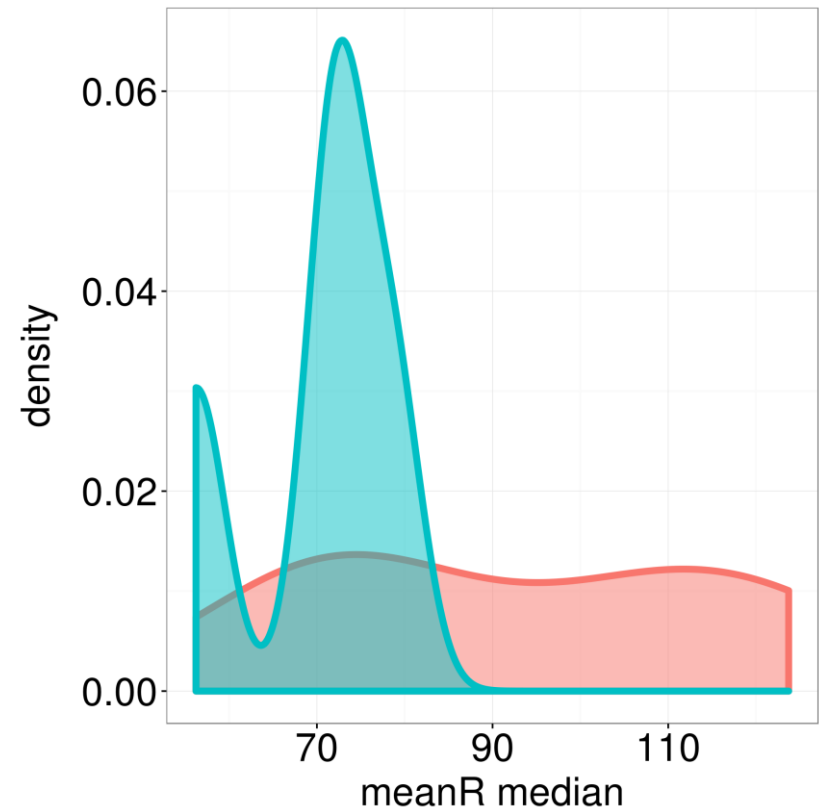
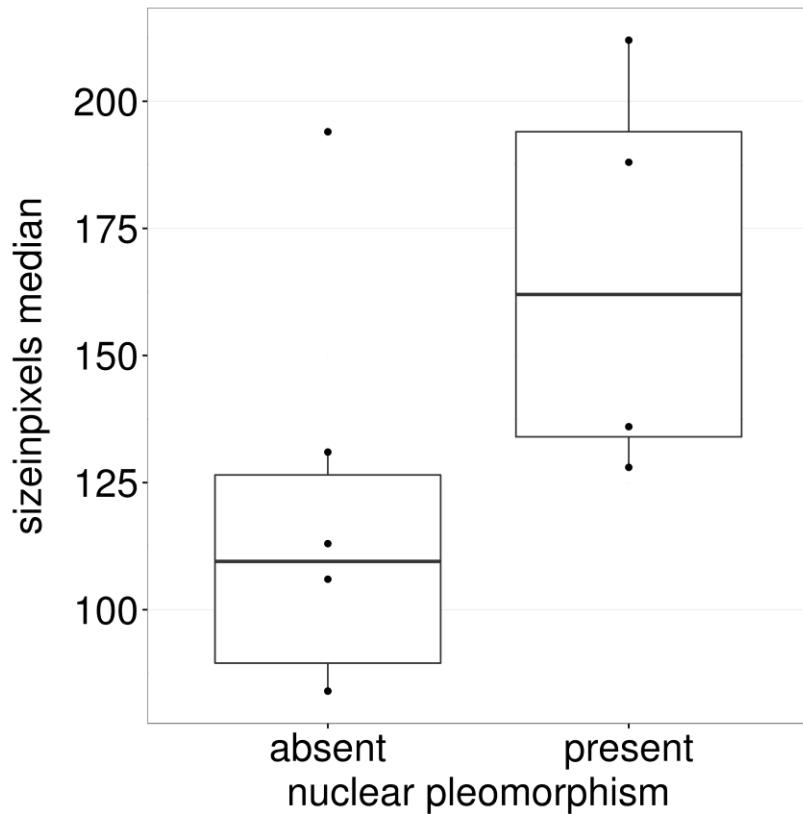
Relate text features to image features

1. Demonstrate relationships
2. Create integrated human/algorithm feature metadata
3. Relate to “omics”, Radiology, outcome
4. In silico classifications

Dockers Run at U Pittsburgh; extracted data analyzed at Stony Brook

Relationship between Text Nuclear Pleomorphism Feature and Imaging Features

Blue – presence of pleomorphism,
Red – absence of pleomorphism



Dissemination

- <http://quip1.bmi.stonybrook.edu>
- Containers
- Cloud
- 3D Slicer Pathology
- TCIA – Integration between Radiology, Pathology images and features
- TCIA - plugins will fetch GDC data for particular TCIA subject
- HPC via NSF and DOE
- TCGA – PanCanAtlas – Joint with Cooper and Gutman - Lymphocyte characterization
- Integrated Features/NLP joint with TIES

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!

[Movie!](#)