Tools and Methods for Leveraging Digital Pathology in Integrative Cancer Research

Stony Brook: Joel Saltz, Tahsin Kurc, Yi Gao, Allen Tannenbaum, Fusheng Wang, Jonas Almeida, George Teodoro, Erich Bremer, Kevin Katcher, Yangyang Zhu, Elizabeth Vanner, Andrew White
Emory University: Ashish Sharma, Adam Marcus, Richard Cummings, Daniel Brat, Roberd Bostick **Oak Ridge National Laboratory:** Scott Klasky, David Pugmire, Jeremy Logan **Yale University:** Michael Krauthammer

Introduction

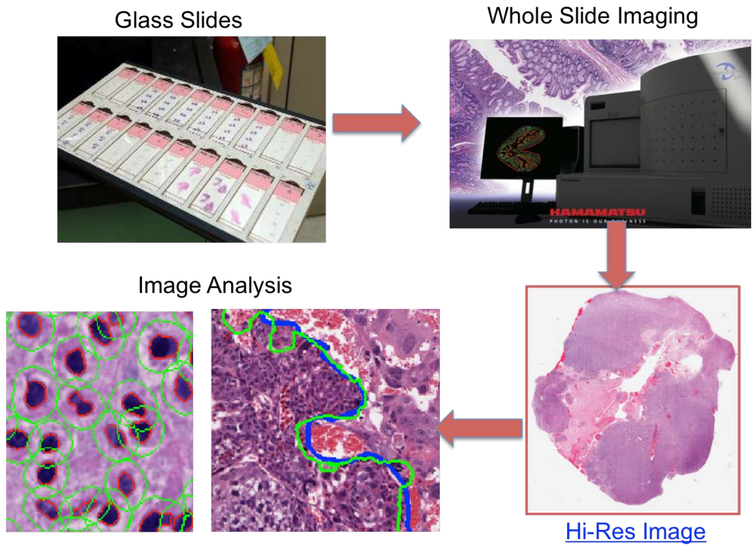
This project Is developing a suite of open source tools and integrated informatics platform that will facilitate multi-scale, correlative analyses of high resolution whole slide tissue image data, spatially mapped genetics and molecular data for cancer research.

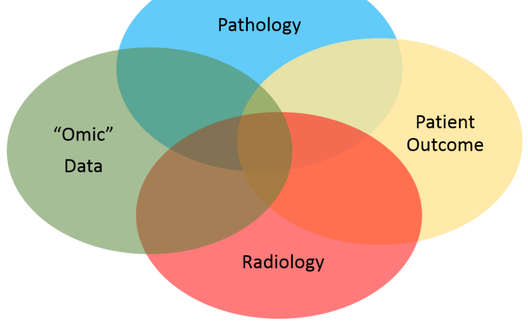
These tools will allow quantitative analyses of the interplay between morphology and spatially mapped genetics and molecular data and will be used in studies that predict outcome and response to treatment, in radiogenomic and quantitative Radiology imaging studies and in studies to identify cancer targets.

The software and methods will enable researchers to assemble and visualize detailed, multi-scale descriptions of tissue morphologic changes originating from a wide range of microscopy instruments and make it possible to efficiently manage, interrogate, and explore microscopy imaging data at multiple scales and to identify and analyze features across individuals and cohorts.

Specific Aims

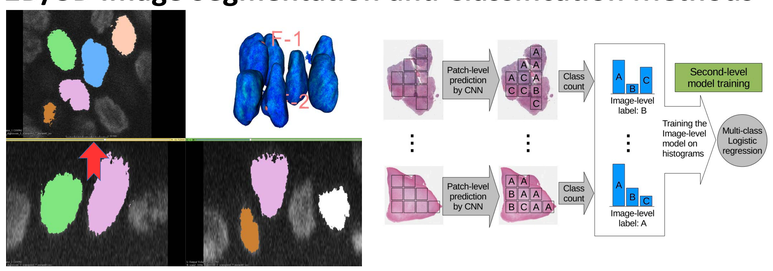
- Develop, deploy, and evaluate robust and scalable methods and analysis pipelines for multi- scale, integrative image analysis.
- Develop database infrastructure to manage and query image data, image analysis results.
- Develop high performance software that targets clusters, cloud computing, and leadership scale systems.
- Develop visualization middleware for 2D/3D image and feature data and for integrated image and “omic” data.
- Drive continuing development of the tools using a suite of cancer driving biomedical problem, and provide collaborative support and training to the cancer research community.





Tools and Methods

2D/3D Image Segmentation and Classification Methods

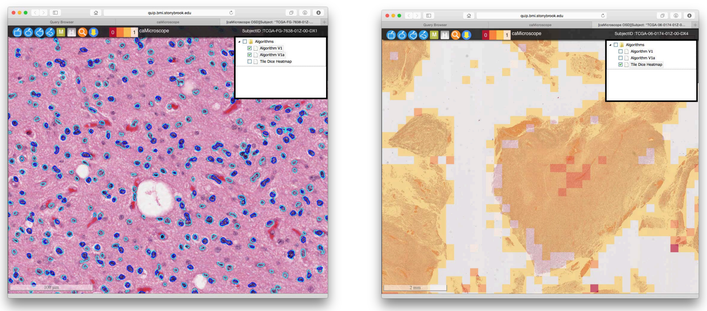


MongoDB based management of segmentation results and features

Methods for segmenting nuclei from whole slide tissue images; 3D nuclear segmentation from con-focal and super-resolution tissue images. CNN based methods for classification of tissues and patients.

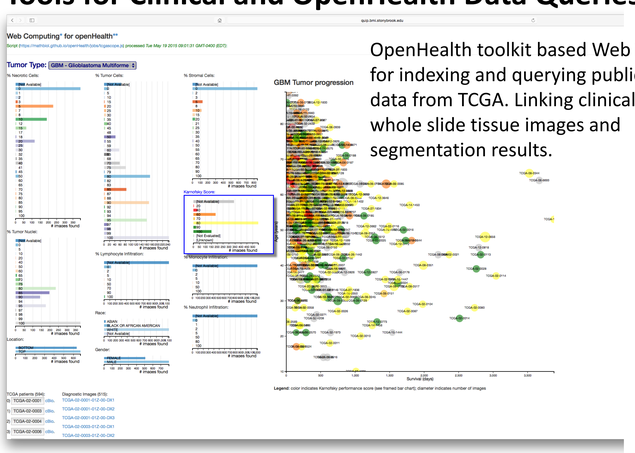
- Tens of Thousands to Millions of nuclei per slide
- Multiple features per nucleus
- Analysis of thousands of images
- 1.2 Billion segmented nuclei
- Need to store, index, and query data quickly

caMicroscope-based Web Viewing of Images and Segmentation Results



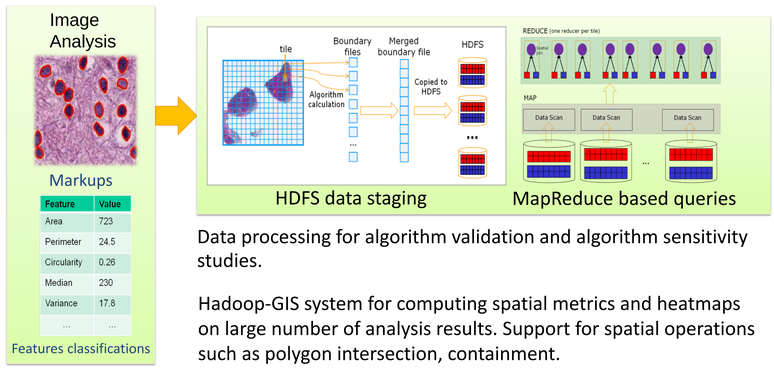
View whole slide images and segmentation results from multiple algorithms. Heat maps for visual inspection of agreement and disagreement between results from two algorithm.

Tools for Clinical and OpenHealth Data Queries



OpenHealth toolkit based Web interfaces for indexing and querying public clinical data from TCGA. Linking clinical data with whole slide tissue images and segmentation results.

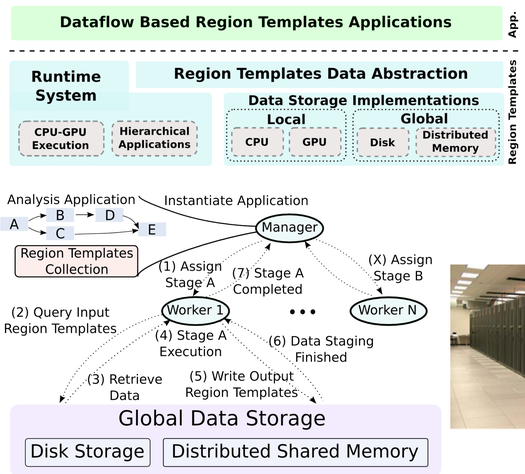
Hadoop-GIS for spatial analytics



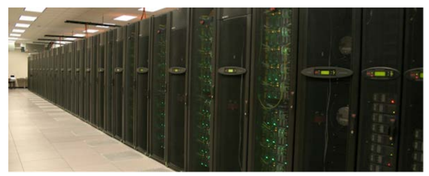
Data processing for algorithm validation and algorithm sensitivity studies.

Hadoop-GIS system for computing spatial metrics and heatmaps on large number of analysis results. Support for spatial operations such as polygon intersection, containment.

High Performance Computing for Analysis

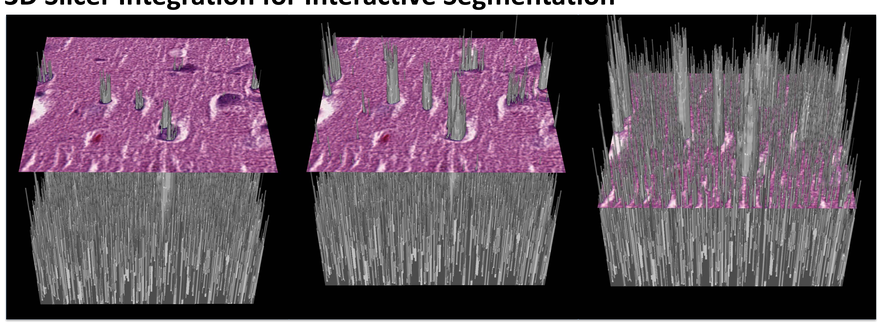


SBU Region Templates to take advantage of large computational clusters and hybrid systems with multiple GPUs and CPUs



ORNL ADIOS system for high performance I/O for scientific applications running on machines with hundreds of thousands of compute cores. EAVL library for high performance scientific data visualization.

3D Slicer Integration for Interactive Segmentation



Ability to adjust intensity thresholds. Interactively view variation of intensity and segmentation of nuclei on image patches.

This work was supported in part by 1U24CA180924-01A1 from the NCI and by R01LM011119-01 and R01LM009239 from the NLM and NCIP/Leidos 14X138