

Informatics Tools for Optimized Imaging Biomarkers for Cancer Research & Discovery

Bruce Rosen/Jayashree Kalpathy-Cramer/Artem
Mamonov/Karl Helmer

Athinoula A. Martinos Center for Biomedical Imaging,
Massachusetts General Hospital, Boston, MA

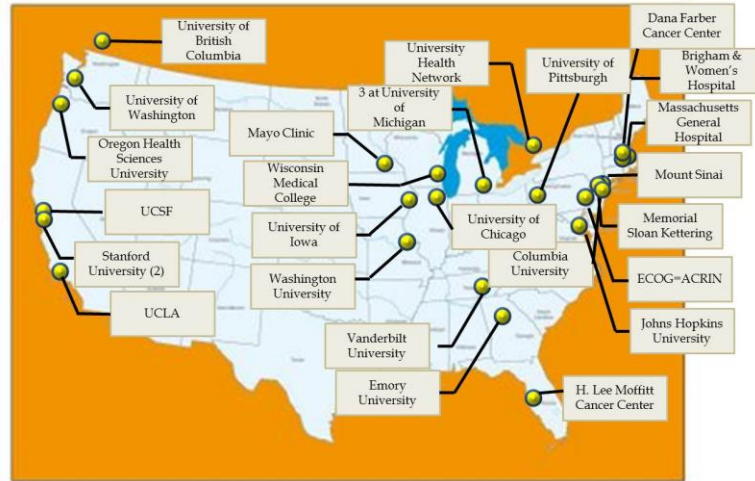


MGH/HST Athinoula A. Martinos
Center for Biomedical Imaging

Collaborators

- ▣ Robert Gillies, PhD
 - ▣ Moffitt Cancer Center
- ▣ Dmitry Goldgof, PhD
 - ▣ University of South Florida
- ▣ Sandy Napel, PhD
 - ▣ Stanford University
- ▣ Binsheng Zhao, PhD
 - ▣ Columbia University

Quantitative Imaging Network



The network is designed to promote research and development of **quantitative imaging** methods for the **measurement of tumor response** to therapies in clinical trial settings, with the overall goal of facilitating clinical decision making.

Grant arose from the collaborative projects as part of QIN

ITCR U24 : QIN as test-bed for C-BIBOP

- ▣ QIN is engaged in challenges in a fairly substantial manner
- ▣ Working on previous QIN challenges highlighted need for resources
 - ▣ Infrastructure (platform)
 - ▣ Resources to coordinate and conduct challenges
 - ▣ Metrology tools
 - ▣ Visualization tools
- ▣ Paradigm shift
 - ▣ Share code, not just results
- ▣ Support reproducibility in image analysis
- ▣ **Make imaging data more accessible to non-imaging scientists**
 - ▣ Note: Also funded by Leidos contract

Cloud-based Image Biomarker Optimization Platform (C-BIBOP)

- ▣ Goal to create an open-source platform to support algorithm comparison/benchmarking
 - ▣ Supports many use cases including uploading results or running algorithms in the cloud (VMs, Docker)
 - ▣ Live at all times, not just at conferences
 - ▣ Can be used for benchmarking and algorithm comparison
 - ▣ Can be used to share image analysis workflows
- ▣ Initial use cases
 - ▣ CT volumetry
 - ▣ Brain tumor segmentation
- ▣ Build on following:
 - ▣ QIN challenge infrastructure
 - ▣ CodaLab
 - ▣ VISCERAL (EU project)
 - ▣ TCIA API

Why “challenges”?

- Reproducibility is an issue in all aspects of medicine
 - Algorithm performance often not replicated by other sites
- Access to clinical data of sufficient variety can be a challenge for (computational) scientists developing algorithms
- Can evaluate the performance of techniques on real, noisy clinical data
- Test data (sequestered) can provide indication of algorithm generalizability to unseen data
- Allows for cross-pollination of methods from other domains
- Best algorithms can be translated into commercial products

Coding4Cancer

Challenges for improving cancer screening

In the second competition,
The National Cancer Institute,



part of the National Institutes of Health, and the FNIH will support competitions that will ask coders to create the best computer algorithm that can identify a person as having lung cancer based on two sets of low dose computed tomography images taken 1 to 2 years apart—when the scans may show the growth of a tumor. Another challenge will be more



difficult: It will ask coders to create algorithms that can spot lung cancer based on one set of images taken from a single scan or study.

What is a “challenge”?

- Impartial group of scientists (govt./commercial org) organizes a ‘challenge’ to solve a (clinical relevant) problem
 - Meaningful question
 - Well curated, representative dataset
 - Well established evaluation metrics
- Typically split into a training dataset, a validation dataset (optional) and test dataset
 - Test data withheld from challenge participants and used for final evaluation
- Leaderboards can provide real-time feedback to participants based on the validation dataset
- Final results based on (gold-standard, preferably independent) test dataset
- Such a design closely reflects the actual difficulties faced by real-world users trying to determine whether **an algorithm generalizes to unseen cases**

Where are challenges conducted?

- At annual conferences sponsored by scientific societies
 - MICCAI has held “grand challenges” since 2007
 - ISBI
 - SPIE
- On commercial platforms
 - Kaggle
 - TopCoder
 - Sage/Synapse
- Within Organizations such as QIN and QIBA
 - Lung nodule segmentation challenges
 - DCE
 - PET

How are challenges typically conducted?

- ▣ Organizers identify a “challenge”
- ▣ Organizers generate/identify “ground truth” for training and test data
- ▣ Organizers define evaluation metrics
- ▣ Organizers announce challenge
- ▣ Interested participants register and download data
- ▣ Participants apply their methods to the training data and obtain results. If satisfactory, participants apply methods to test data
- ▣ Upload results test data results
- ▣ Results made available to community (“leader board”), perhaps at conference
- ▣ Prizes can be awarded

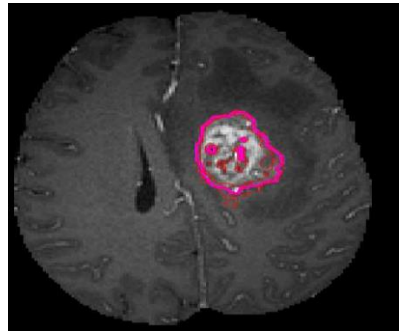
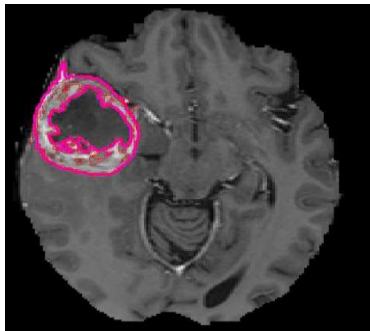
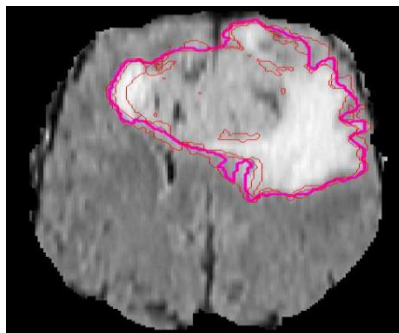
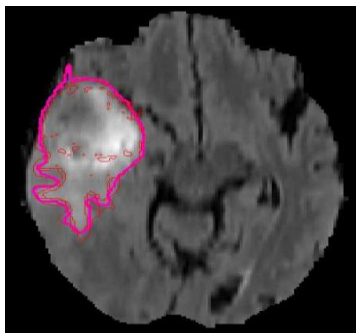
MICCAI Brain tumor segmentation challenge

- Organized by academics, NCI
- Being run since 2012
 - Had 3-4 experts label volumes
- In 2014, added data from TCIA-GBM collection
 - Test labels were machine generated (not optimal)
- 2015 added expert labels
 - 2 experts

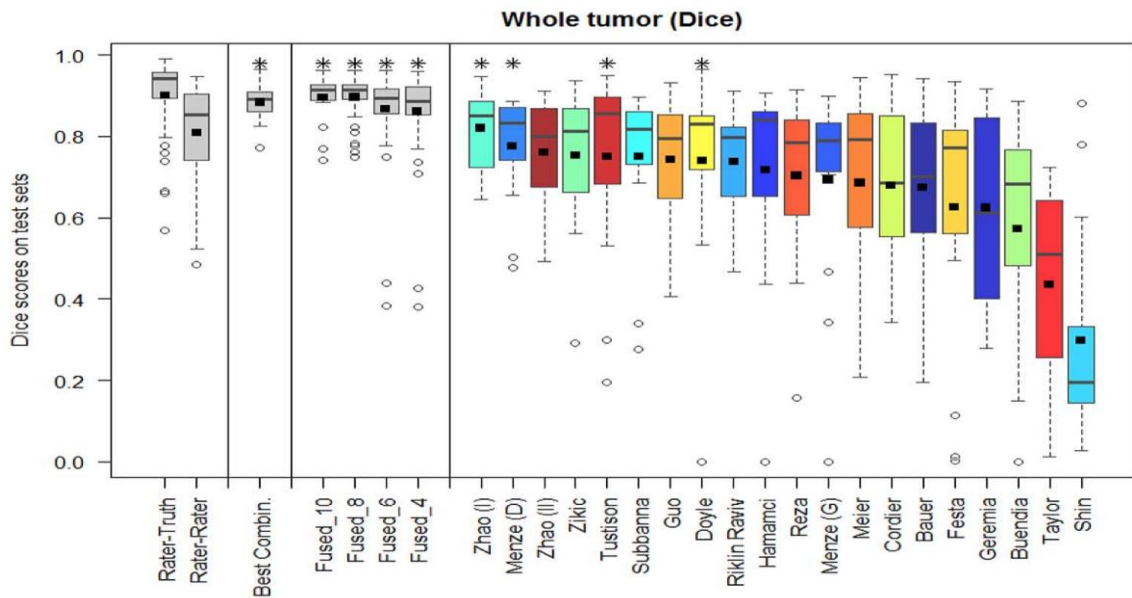
NCI MICCAI-BraTS

- Segmentation of brain tumors into 3 (4) regions
 - Necrosis
 - Enhancing tumor
 - Edema
- Based on
 - T1 (pre and post contrast)
 - T2
 - FLAIR
- Run over 3 years
- 10-20 groups participated

Example slices with output of segmentation algorithms



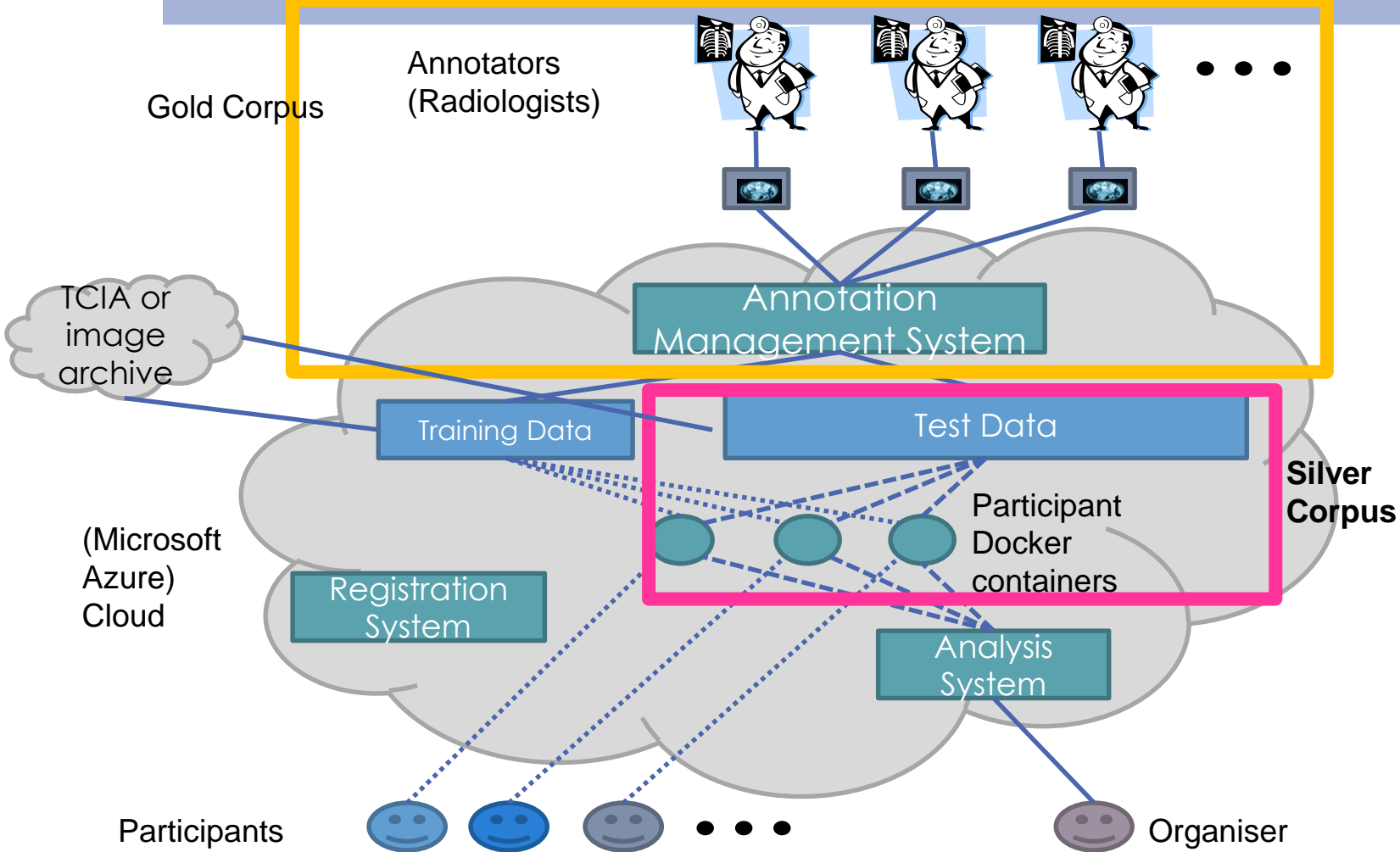
Results of computer generated labels



Improvement by combining best methods

A new paradigm

- ▣ Limitations of traditional model:
 - ▣ Datasets too large to be moved around (in “Big Data” era)
 - ▣ Some data may be too sensitive to share (PHI)
 - ▣ Limited reproducibility when participants provide just results and not code
 - ▣ Cannot compare algorithm efficiency (and how should we compare alg A with 95% Dice, 24 hours run time with alg B at 93% Dice, 2.4s run time)
 - ▣ Cannot compare algorithms performance on unseen, new data
- ▣ Need to transition to cloud-based evaluation
 - ▣ Algorithms move to data
 - ▣ Share executables/code/VM/Docker
 - ▣ Participants never see test data
 - ▣ Code can be run on new (prospective?) data

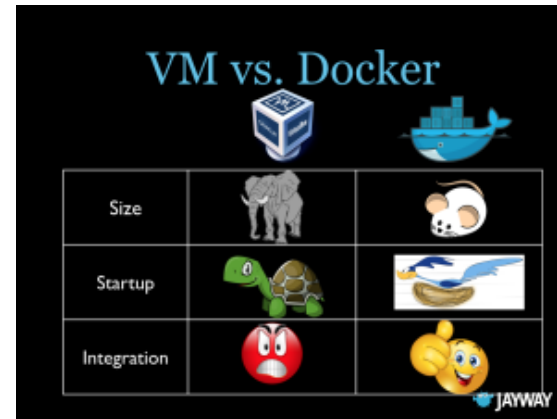
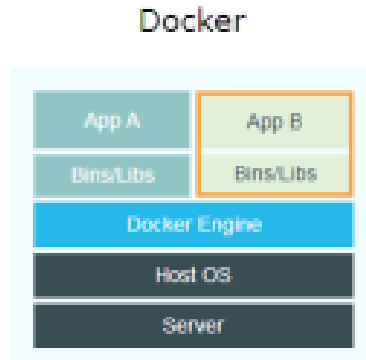
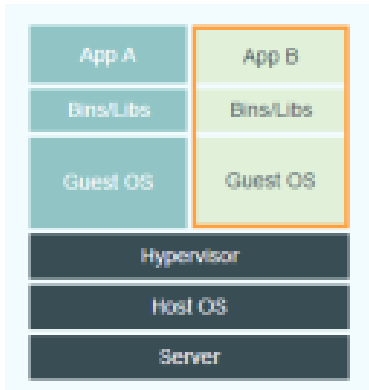


System components

- ▣ CodaLab for challenge management
- ▣ Shiny/R integration for statistical analysis
- ▣ Integration with TCIA (Rest API)
- ▣ ePad for visualization
- ▣ caMicroscope for supporting digital pathology challenges
- ▣ Containers for sharing code

Also funded through Leidos contract

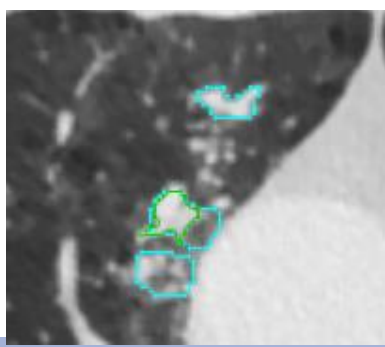
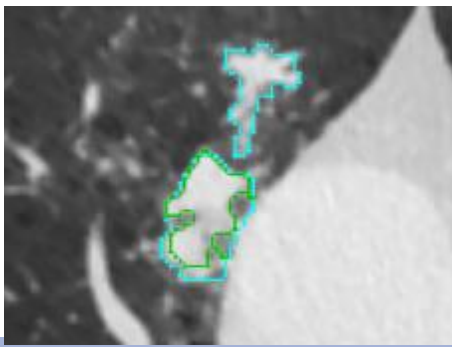
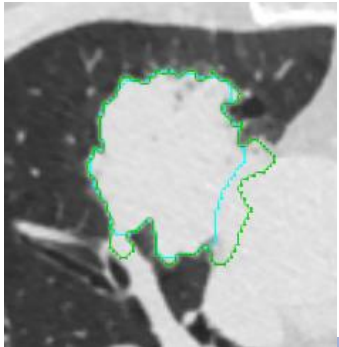
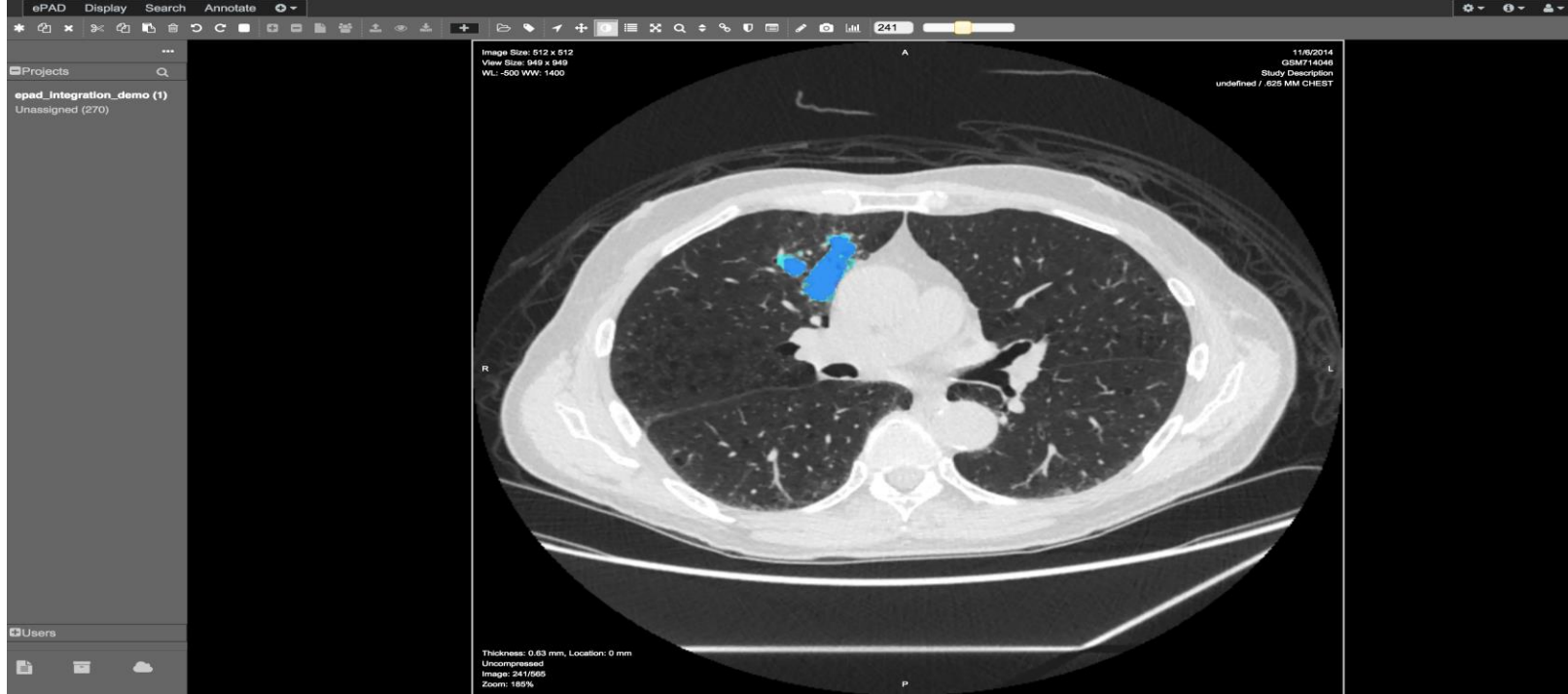
Docker vs. Virtual Machines



- <http://www.jayway.com/2015/03/21/a-not-very-short-introduction-to-docker/>

QIN challenge: Lung nodule segmentation

- 52 lesions from 41 CT studies
- 33 to μl to 57 ml and demonstrated a diversity of shapes from round through spiculated.
- Three algorithms, each submitted 3 repeat segmentations per nodule



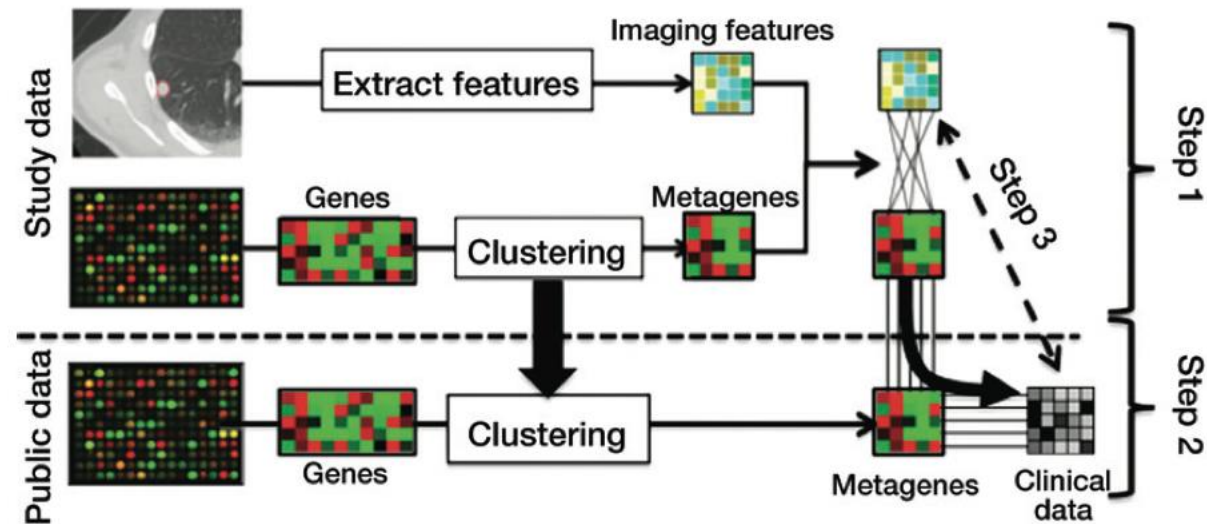
Output of challenge

- CT volumes in TCIA (existing collections)
- Segmentations in TCIA (in DICOM-SEG format)
- Segmentations can be used for radiomics and radiogenomic studies (underway in QIN)
 - Stability of features
 - Correlation between features
 - Identify “habitats” or sub volumes based on features

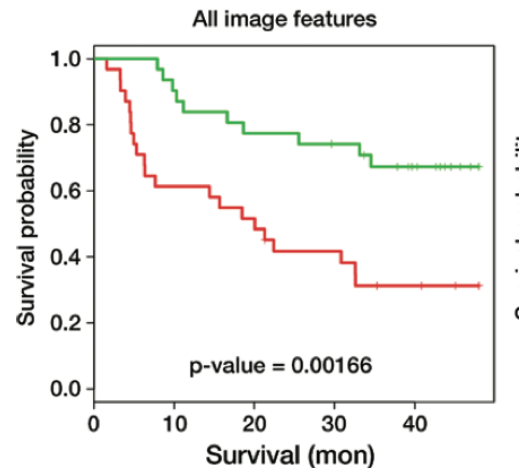
Radiogenomics (Radiopathagenomics...)

■ Radiomics

■ Genomics

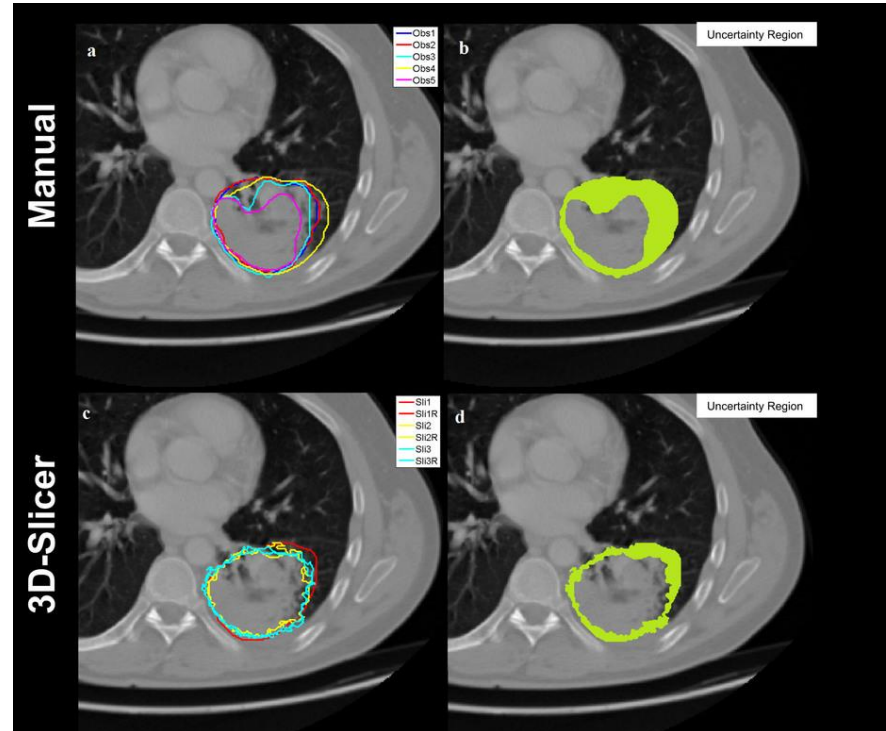


■ Clinical prediction/out comes



Features can be sensitive to segmentation


- Reduced uncertainty with machine assisted segmentation



QIN Feature comparison challenge

- “Radiomics” pipelines allow for the quantification of imaging characteristics
 - Can be used in outcomes research
 - Radiogenomics
 - However, features can be sensitive to segmentation
- 8 QIN sites participated
- 10-300 features per site

C-BIBOP used for management



CT Feature Comparison Study

Organized by artem - Current server time: June 9, 2016, 11:42 p.m. UTC

[Current](#)[End](#)

[Evaluate](#)[Competition Ends](#)

Oct. 16, 2015, midnight UTCNever

[Learn the Details](#)[Phases](#)[Participate](#)[Results](#)

[Overview](#)[Evaluation](#)[Terms and Conditions](#)

Welcome!


Background: As part of the CT Segmentation Challenge, we evaluated the results of 3 repeated runs of segmentation algorithms from 3 participating institutions. Various measures were developed to compare the results of each run of a specified algorithm, comparison to ground truth (if known) as well as pair-wise comparison of algorithms from each institution.

Goal: The goal of the CT Feature Comparison Study is to evaluate features generated from each of these segmentations for repeatability between repeated runs of each algorithm, and reproducibility across segmentation algorithms.

We will use the 5 collections of DICOM CT images that were used for the "moist run" from the following sources: CUMC_FDA Phantom, Moffitt Cancer Center, Rider, Stanford, and UDC. The collection consists of volumes of 52 nodules and 9 segmentations of each nodule. These collections and full descriptions are available at NCI TCIA under "QIN multi-site collection of Lung CT data with Nodule Segmentations" with DOI <http://dx.doi.org/10.7937/K9/TCIA.2015.1BUVJR7>. (If you have issues reading DICOM-SEG files, you can find segmentations in NIfTI formats at <https://ncipub.org/projects/ctfeature> in new, "nii_segmentations" subdirectory (under Files)).

We will expand on the work of the segmentation challenge as well as the work on feature reproducibility on test-retest data and numerous papers on various measures for feature stability measures, including repeatability and reproducibility. Note that R-scripts will be run on C-BIBOP to compute commonly used metrics. We will coordinate these submissions with separate QIN effort to compare and harmonize features across sites.

Each participating institution would upload a comma separated value file containing their feature computations for each of the 450 segmentations (5*10*3*3) to the NCIP HUB using a standard format (likely csv). Sample output file is provided via NCI-HUB "CT feature



CT Feature Comparison Study

Organized by artem - Current server time: Jan. 12, 2016, 3:03 p.m. UTC

[Current](#)[End](#)

[Evaluate](#)[Competition Ends](#)

Oct. 16, 2015, midnight UTCNever

[Learn the Details](#)[Phases](#)[Participate](#)[Results](#)

[Evaluate](#)

Phase description
None

[Download CSV](#)[Download all submissions on leaderboard](#)

Results				
	User	N features	Avg Repeatability	Avg Reproducibility
1	JohannaUthoff	304.0	0.985 (1)	0.845 (2)
2	ivan.yeung@mp.uhn.on.ca	10.0	0.983 (2)	0.675 (8)
3	lhadjjiiski	49.0	0.974 (3)	0.863 (1)
4	mmcnittgray	15.0	0.972 (4)	0.834 (3)
5	LinLu	71.0	0.947 (5)	0.796 (4)
6	sechegaray	198.0	0.928 (6)	0.722 (6)
7	sechegaray-2	774.0	0.894 (7)	0.748 (5)
8	sechegar	762.0	0.861 (8)	0.713 (7)
9	cherezov	186.0	0.842 (9)	0.616 (9)

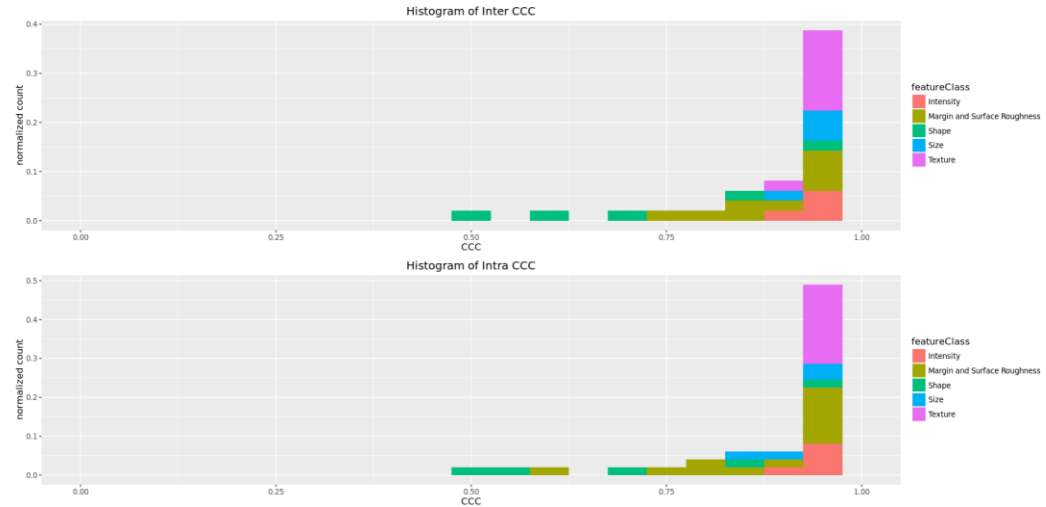
Results

Site Selection

Select site

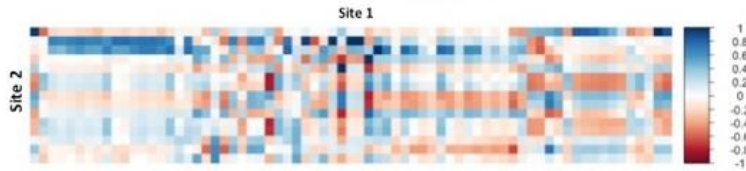
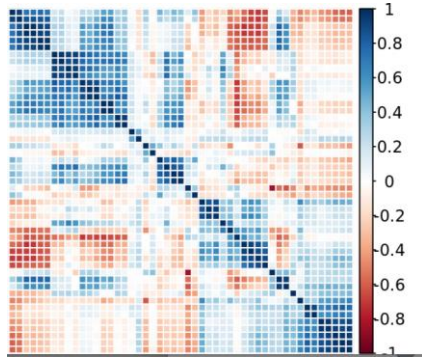
UMICH

CUMC
NStanford
PM
Stanford
UCLA
UMICH
Iowa
Uof

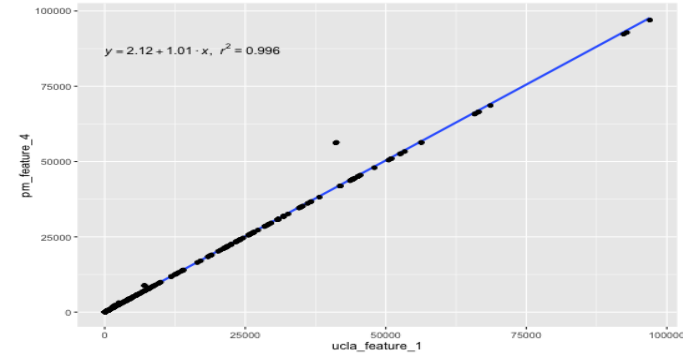


- Feature stability with respect to segmentation

Results



c. Example of heat map of correlation between features submitted between two sites highlighting the high correlation of features across implementation



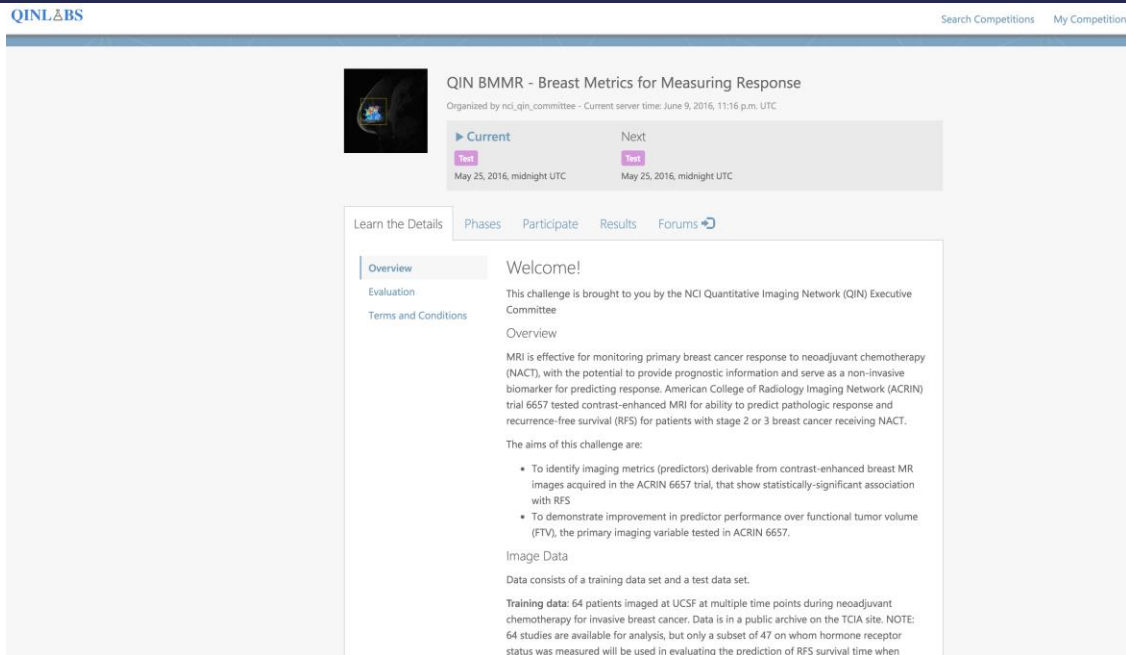
4	feature_4	PM	Size	Volume	3D	no	1
1	feature_1	UCLA	Size	Volume	volume	3D	No

- Inter and intra-site correlation of features

QIN BMMR challenge (clinical trial data)

- The aims of this challenge are:
 - To identify imaging metrics (predictors) derivable from contrast-enhanced breast MR images acquired in the ACIN 6657 trial, that show statistically-significant association with RFS
 - To demonstrate improvement in predictor performance over functional tumor volume (FTV), the primary imaging variable tested in ACIN 6657.

QIN BMMR challenge



The screenshot shows the QIN BMMR challenge website. At the top, the QIN LABS logo is on the left, and links for 'Search Competitions' and 'My Competitions' are on the right. The main header area features a small image of a brain scan on the left and the title 'QIN BMMR - Breast Metrics for Measuring Response' on the right. Below the title, it says 'Organized by nci_qin_committee - Current server time: June 9, 2016, 11:16 p.m. UTC'. A navigation bar shows 'Current' and 'Next' phases, both with 'Test' buttons and dates of 'May 25, 2016, midnight UTC'. Below this is a 'Learn the Details' section with tabs for 'Overview', 'Evaluation', and 'Terms and Conditions'. The 'Overview' tab is active, displaying a 'Welcome!' message and an 'Overview' section. The overview text states that the challenge is brought by the NCI Quantitative Imaging Network (QIN) Executive Committee and describes the goal of predicting pathologic response and recurrence-free survival (RFS) using MRI. It lists two aims: to identify imaging metrics (predictors) and to demonstrate improvement in predictor performance over functional tumor volume (FTV). The 'Image Data' section mentions that data consists of a training data set and a test data set. The 'Training data' section notes that 64 patients were imaged at UCSF, and data is available in a public archive on the TCIA site, with a note that only a subset of 47 patients had hormone receptor status measured.

QIN LABS

Search Competitions My Competitions

QIN BMMR - Breast Metrics for Measuring Response

Organized by nci_qin_committee - Current server time: June 9, 2016, 11:16 p.m. UTC

Current Next

Test Test

May 25, 2016, midnight UTC May 25, 2016, midnight UTC

Learn the Details Phases Participate Results Forums

Overview Evaluation Terms and Conditions

Welcome!

This challenge is brought to you by the NCI Quantitative Imaging Network (QIN) Executive Committee

Overview

MRI is effective for monitoring primary breast cancer response to neoadjuvant chemotherapy (NACT), with the potential to provide prognostic information and serve as a non-invasive biomarker for predicting response. American College of Radiology Imaging Network (ACRIN) trial 6657 tested contrast-enhanced MRI for ability to predict pathologic response and recurrence-free survival (RFS) for patients with stage 2 or 3 breast cancer receiving NACT.

The aims of this challenge are:

- To identify imaging metrics (predictors) derivable from contrast-enhanced breast MR images acquired in the ACRIN 6657 trial, that show statistically-significant association with RFS
- To demonstrate improvement in predictor performance over functional tumor volume (FTV), the primary imaging variable tested in ACRIN 6657.

Image Data

Data consists of a training data set and a test data set.

Training data: 64 patients imaged at UCSF at multiple time points during neoadjuvant chemotherapy for invasive breast cancer. Data is in a public archive on the TCIA site. NOTE: 64 studies are available for analysis, but only a subset of 47 on whom hormone receptor status was measured will be used in evaluating the prediction of RFS survival time when

- QIN challenge led by Nola Hylton, data from ISPY/ACRIN trial

Integrated Challenges (imaging/pathology/omics)

▣ MICCAI 2015

- ▣ Joint radiology pathology brain tumor challenge
- ▣ Classify glioma grade using pathology and radiology images

▣ MICCAI 2016

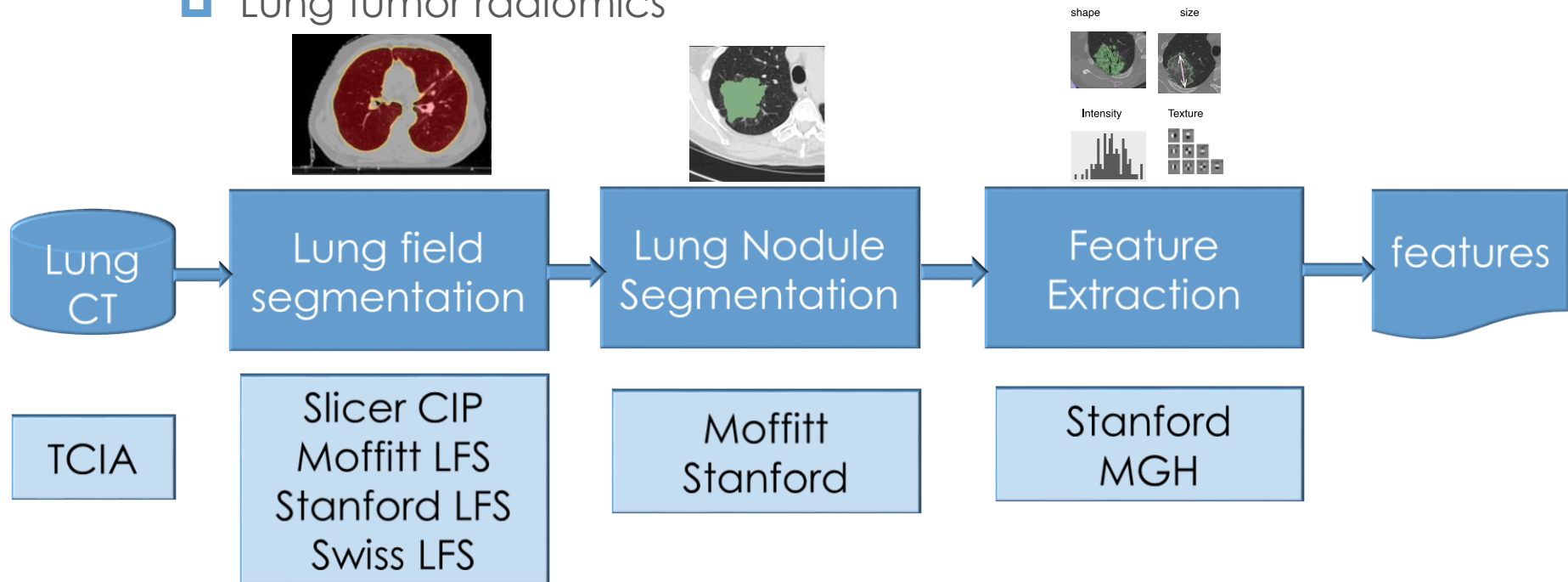
- ▣ Joint radiology pathology challenge
- ▣ TCIA-TCGA data

Cloud-based imaging workflows

- ▣ Tools are shared in Docker containers
- ▣ A worksheet can be used to share a pipeline of tools
- ▣ Supports reproducibility by allowing sharing of workflows with configurations, data, results

Example workflow

■ Lung tumor radiomics



Stanford tools (dockerized)



riipl/feature
public

0
STARS

38
PULLS

>
DETAILS



riipl/tumor
public

0
STARS

7
PULLS

>
DETAILS



riipl/lung
public

0
STARS

6
PULLS

>
DETAILS

■ Available on Dockerhub

Stanford tools

- 3D image feature pipeline (DSOs + DICOM series in; features out)
- Lung tumor segmentation (AIM file with seed pixels + DICOM series + DSO for lung field in; DSO of nodule out)
- Lung field segmentation (DICOM series in; DOS for lung field out).
- Stanford features for the feature challenge were computed using the Docker version of our QIFP.

Pathology nuclear segmentation

CBIBOP

Worksheets Competitions Help Sign In

Click here or press '/' to start Go

pathology-nuclear-seg

Permission: you(read) public(read)

Upload the data bundle

uuid	name	description	bundle_type	created	dependencies	command	data_size	state
0x6613b2	input	Upload /home/azureuser/docker-segmentation/input	dataset	2016-02-08 20:40:44			1.0G	ready

Run segmentation on the first image

uuid	name	description	bundle_type	created	dependencies	command	data_size	state
0x8d046c	run-seg-1st-image	run	run	2016-02-17 17:31:43	input	mkdir output; mainTileAndSegmentWS/NucleiOutputFeature input/TCGA-02-0006-01Z-00-DX2.F55eb13f-5816-4490-b2aa-675b20676e5c.output/TCGA-02-0006-01Z-00-DX2.F55eb13f-5816-4490-b2aa-675b20676e5c 4096 1.2	170M	ready

Run segmentation on the second image

uuid	name	description	bundle_type	created	dependencies	command	data_size	state
0x1a7ffb	run-seg-2nd-image	run	run	2016-02-22 23:35:38	input	mkdir output; mainTileAndSegmentWS/NucleiOutputFeature input/TCGA-06-0141-01Z-00-DX1.ec301079-2d58-4079-b2ab-d8fdbc2fb7.output/TCGA-06-0141-01Z-00-DX1.ec301079-2d58-4079-b2ab-d8fdbc2fb7 4096 1.2	632M	ready

Run segmentation on the whole directory

uuid	name	description	bundle_type	created	dependencies	command	data_size	state
0xc32a0b	run-seg-dir	run	run	2016-02-25 21:49:55	input	mkdir output; run-seg-dir.sh input output	1009M	failed

pathology-nuclear-seg

0x14dc56d29df843b4a317b932d0efb7bb

Permissions: you(read) public(read)

Bundles

Name	UUID
input	0x6613b2
run-seg-1st-image	0x8d046c
run-seg-2nd-image	0x1a7ffb
run-seg-dir	0xc32a0b

□ Docker container from Stony Brook

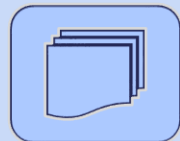
Intel Personalized Care Platform

- Data organized and registered – TCIA-GBM across 2 “partner” sites
- Tools containerized – AFNI, FSL, DRAMMS
- Workflow converted from shell script to Workflow Description Language
- Job submitted to Execution Engine

Slide courtesy Jonathan Lefman

Partner site 1

110 images



load raw

unify

FOV

Strip

Register

Merge

Central site

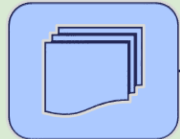


Brain
tumor
atlas

Atlas calc

Partner site 2

274 images



load raw

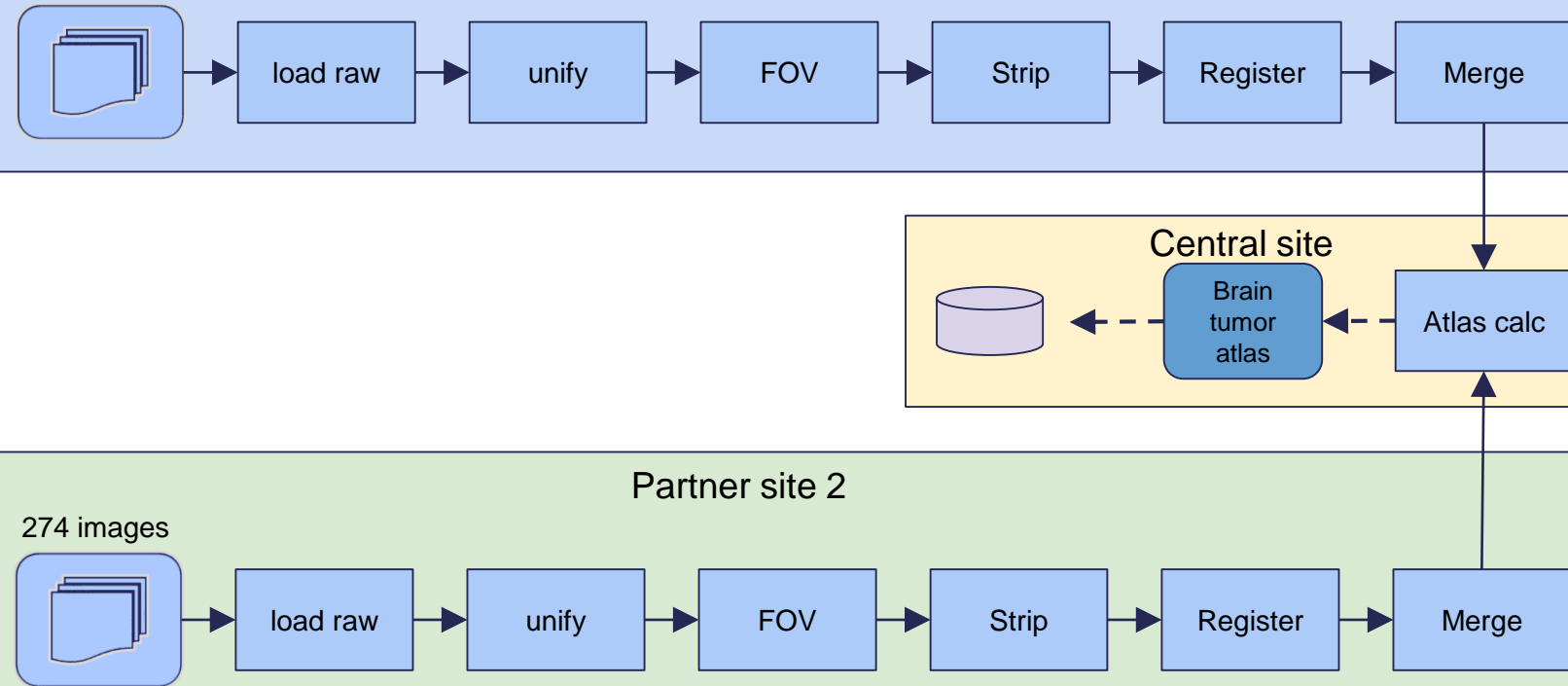
unify

FOV

Strip

Register

Merge



Summary

- Challenges and benchmarks can be important in image analysis, radiomics and radiogenomics.
- The C-BIBOP facilitates conducting of challenges and benchmarks
- Moving algorithms to data is a new paradigm
- Containerization of algorithms facilitates sharing of code and workflows
 - Exploring CWL and WDL

Acknowledgments

- ▣ MGH
 - ▣ QTIM lab
- ▣ MICCAI BraTS Org
- ▣ Stanford
- ▣ Moffitt/USF
- ▣ CUMC
- ▣ ITCR
- ▣ QIN
- ▣ TCIA
- ▣ BWH
- ▣ Stony Brook
- ▣ Grant support: U01CA154601, U24CA180927, U24CA180918, Leidos contract, ITCR administrative supplement