Data Science Methodology Transfer: Space Science to Biomedicine

Dan Crichton
PI, JPL Informatics Center
Leader, Center for Data Science & Technology
Manager, JPL Data Science Office

Rich Doyle
Manager, Information and Data Science Program Office
Manager, HPSC

June 27, 2019

NATIONAL CANCER INSTITUTE®

UNC SCHOOL OF MEDICINE

FRED HUTCHINSON CANCER RESEARCH CENTER
A LIFE OF SCIENCE

CALIFORNIA INSTITUTE OF TECHNOLOGY 1891

NASA

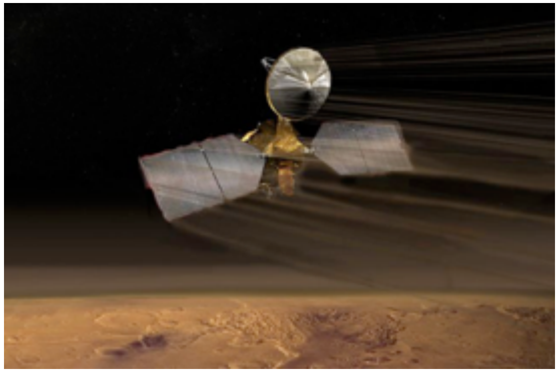
JPL
Jet Propulsion Laboratory
California Institute of Technology

Data Lifecycle Model for NASA Missions

From Onboard Computing to Scalable Data Analytics

Emerging Solutions

- *Next-Generation Flight Computing*
- *Onboard Data Analytics*



Observational Platforms and Flight Computing

Scaling Pressures Expose the Need for an Integrated End-to-End Data and Computational Architecture

Emerging Solutions

- *Intelligent Ground Stations*
- *Agile Mission Operations*

Ground-based Mission Systems

Emerging Solutions

- *Data-Driven Discovery from Archives*
- *Scalable Computation and Storage*

Interactive Analytics and Visualization and Decision Support

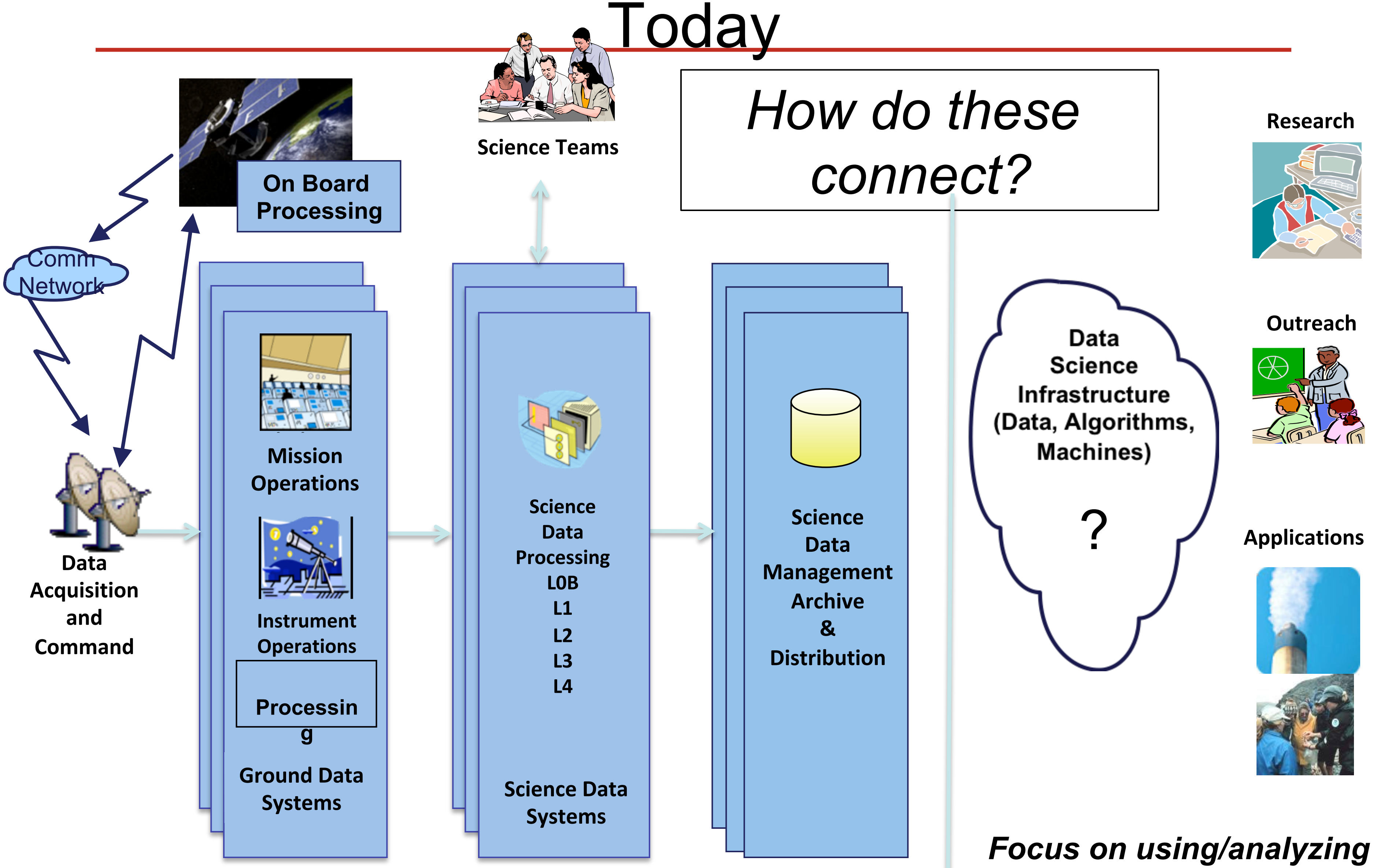
NASA Data Archives

- **NASA captures robust scientific archives of data to support long-term analysis**
  - A requirement on every mission
  - Adheres to specific standards (both in terms of structure + description/metadata)
  - Mandates public access after a specified time period

- **The capture of well-curated, organized data collections is the basis for enabling data analysis**
  - This is a critical precursor step to analysis
  - The scientific community uses these collections as the basis for studies (often grant funded)
  - The development of robust architectures and systems to support data management and data analysis services are emerging

- ***Evolving from research preservation and stewardship towards driving analysis from these archives…***

Planetary Science Data Analytics Support Today



Science Teams

How do these connect?

On Board Processing

Comm Network

Data Acquisition and Command

Mission Operations

Instrument Operations

Processin
g

Ground Data Systems

Science Data Processing
L0B
L1
L2
L3
L4

Science Data Systems

Science Data Management Archive & Distribution

Data Science Infrastructure (Data, Algorithms, Machines)

?

Research

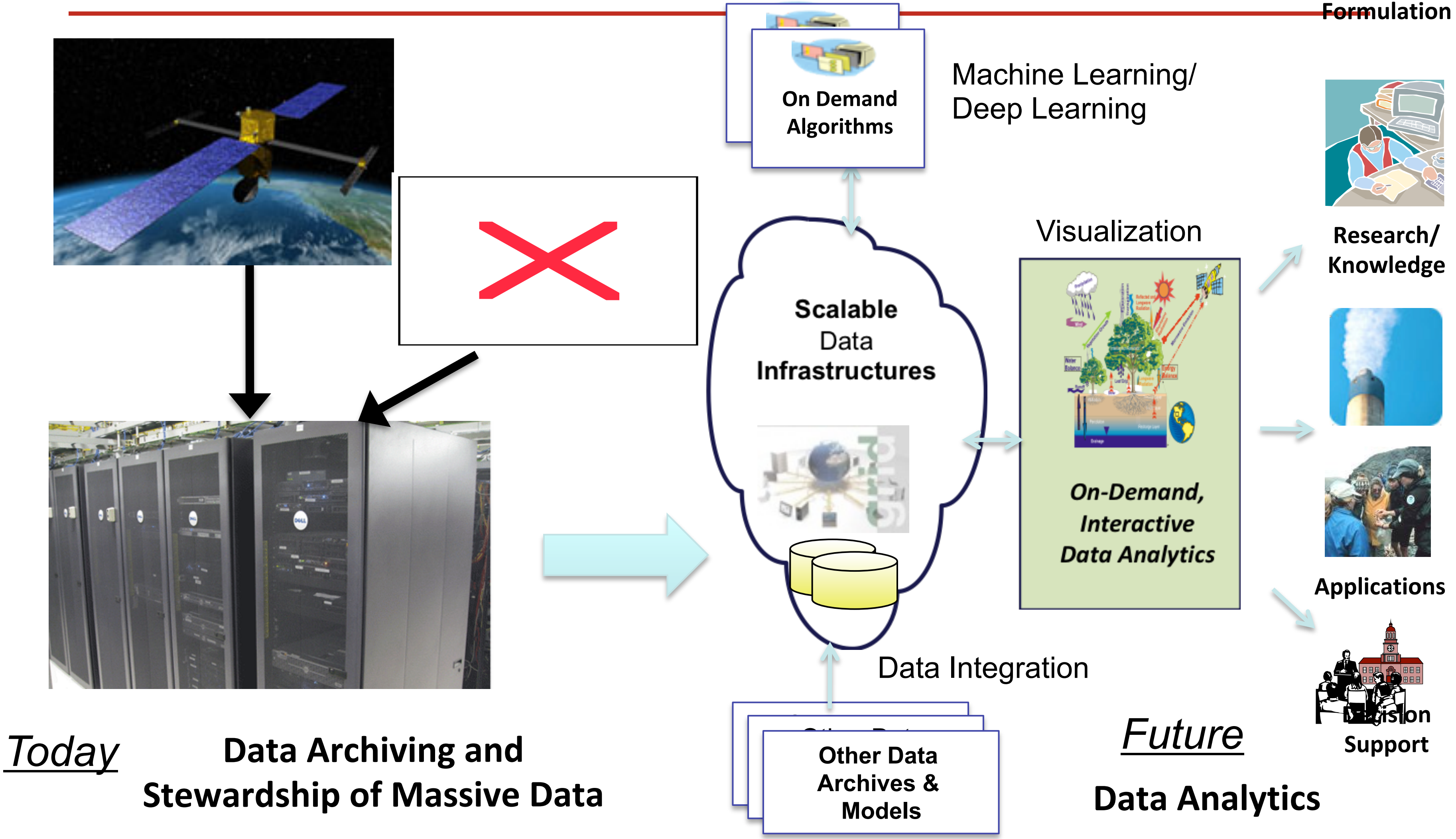
Outreach

Applications

Focus on generating, capturing, managing data

Focus on using/analyzing data

Expanding to Data-Driven Analytics to Enable Science



Formulation

On Demand Algorithms

Machine Learning/ Deep Learning

Scalable Data Infrastructures

Visualization

On-Demand, Interactive Data Analytics

Research/ Knowledge

Applications

Decision Support

Data Integration

Other Data Archives & Models

Today Data Archiving and Stewardship of Massive Data

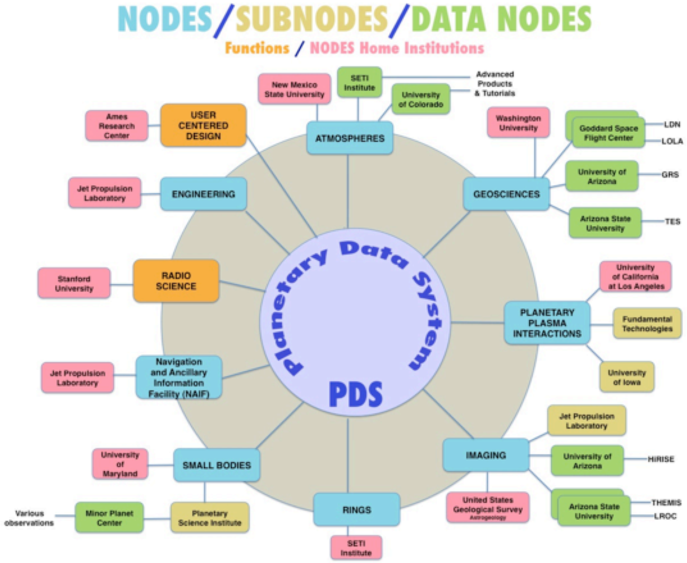
Future Data Analytics

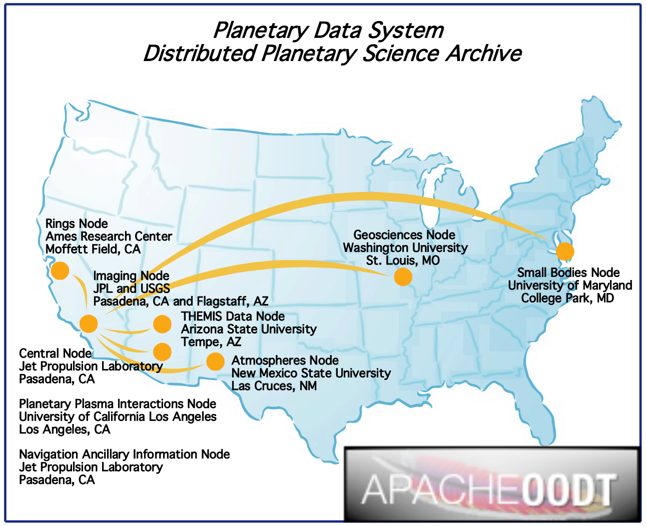
Reducing Data Wrangling: “There is a major need for the development of software components… that link high-level data analysis-specifications with low-level distributed systems architectures.” *Frontiers in the Analysis of Massive Data*, National Research Council, 2013.

Planetary Data System

- **Purpose:** **To collect, archive and make accessible digital data and documentation produced from NASA's exploration of the solar system from the 1960s to the present.**

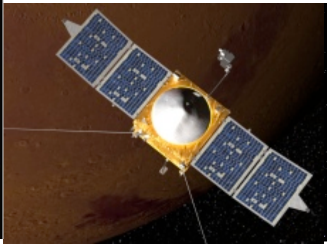
- **Infrastructure:** **A highly distributed infrastructure with planetary science data repositories implemented at major government labs and academic institutions**
  - System driven by a well defined planetary science ontology
  - Approximately 1.7 PB of data
  - About 4000 different types of data and 40M data products
  - International adoption
  - NASA's de facto archive for all planetary data

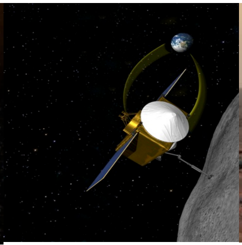



Planetary Data System
Distributed Planetary Science Archive

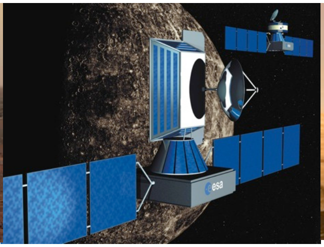
International Collaboration on PDS4 Through IPDA

LADEE
(NASA)


MAVEN
(NASA)

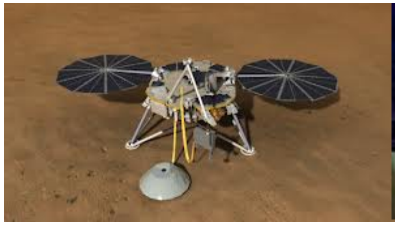

Osiris-REx
(NASA)

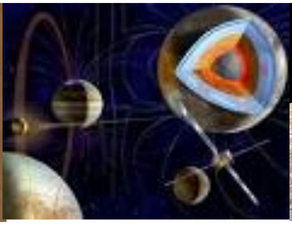
ExoMars
(ESA/Russia)

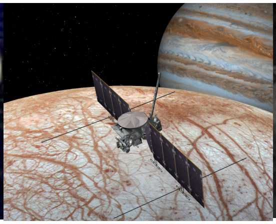

BepiColombo
(ESA/JAXA)

Mars 2020
(NASA)

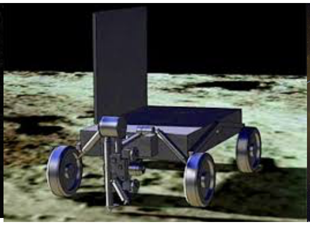
Psyche
(NASA)

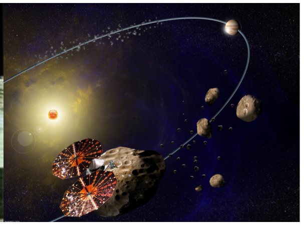

InSight
(NASA)


JUICE
(ESA)


Europa
(NASA)

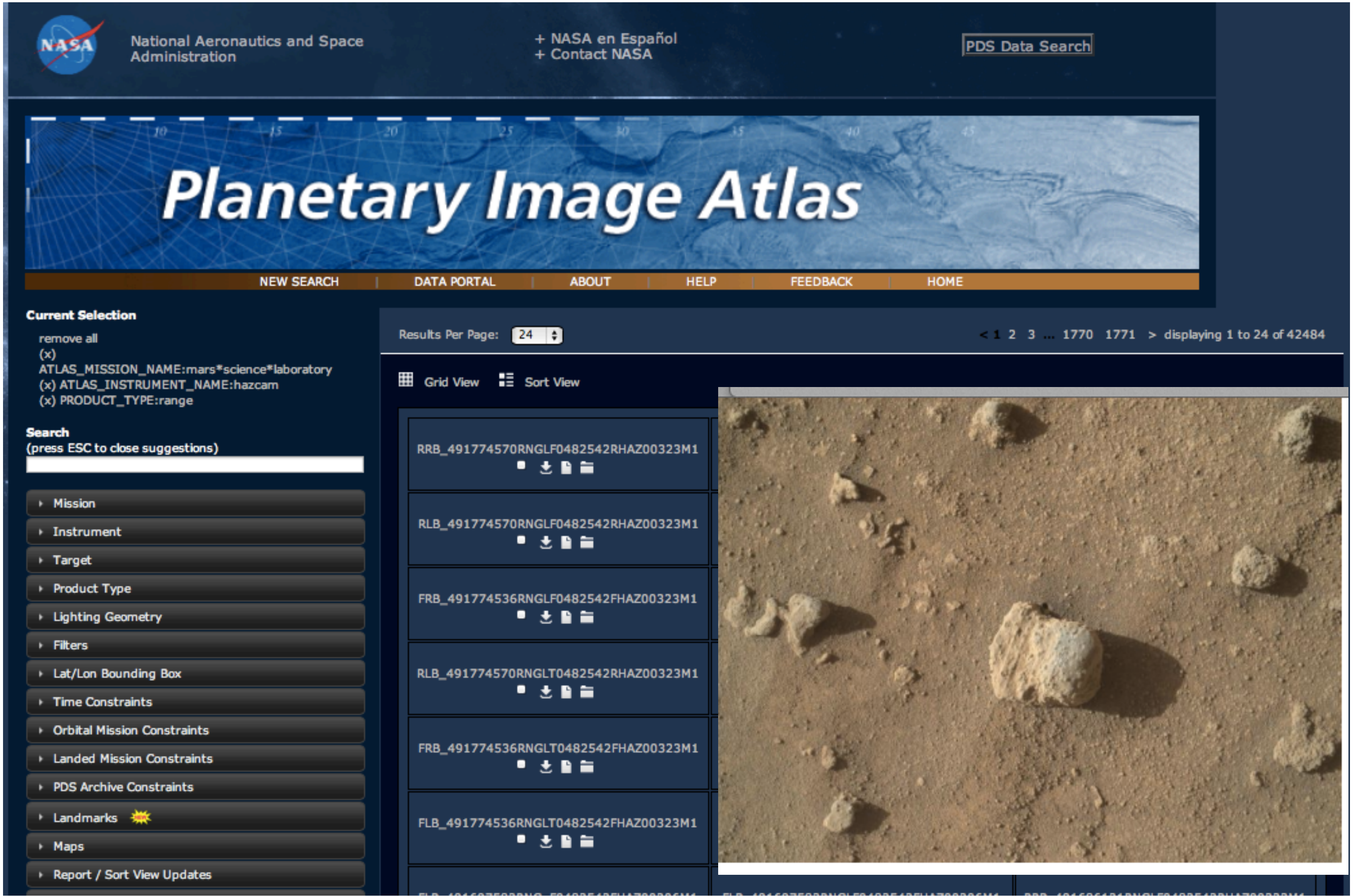
Hyabussa-2
(JAXA)


Chandrayaan-2
(ISRO)


Lucy
NASA

Endorsed by the **International Planetary Data Alliance** in July 2012 –
https://planetarydata.org/documents/steering-committee/ipda-endorsements-recommendations-and-actions

Planetary Image Archiving



National Aeronautics and Space Administration
+ NASA en Español
+ Contact NASA
PDS Data Search
Planetary Image Atlas
NEW SEARCH | DATA PORTAL | ABOUT | HELP | FEEDBACK | HOME

Current Selection

remove all
(x) ATLAS_MISSION_NAME:mars*science*laboratory
(x) ATLAS_INSTRUMENT_NAME:hazcam
(x) PRODUCT_TYPE:range

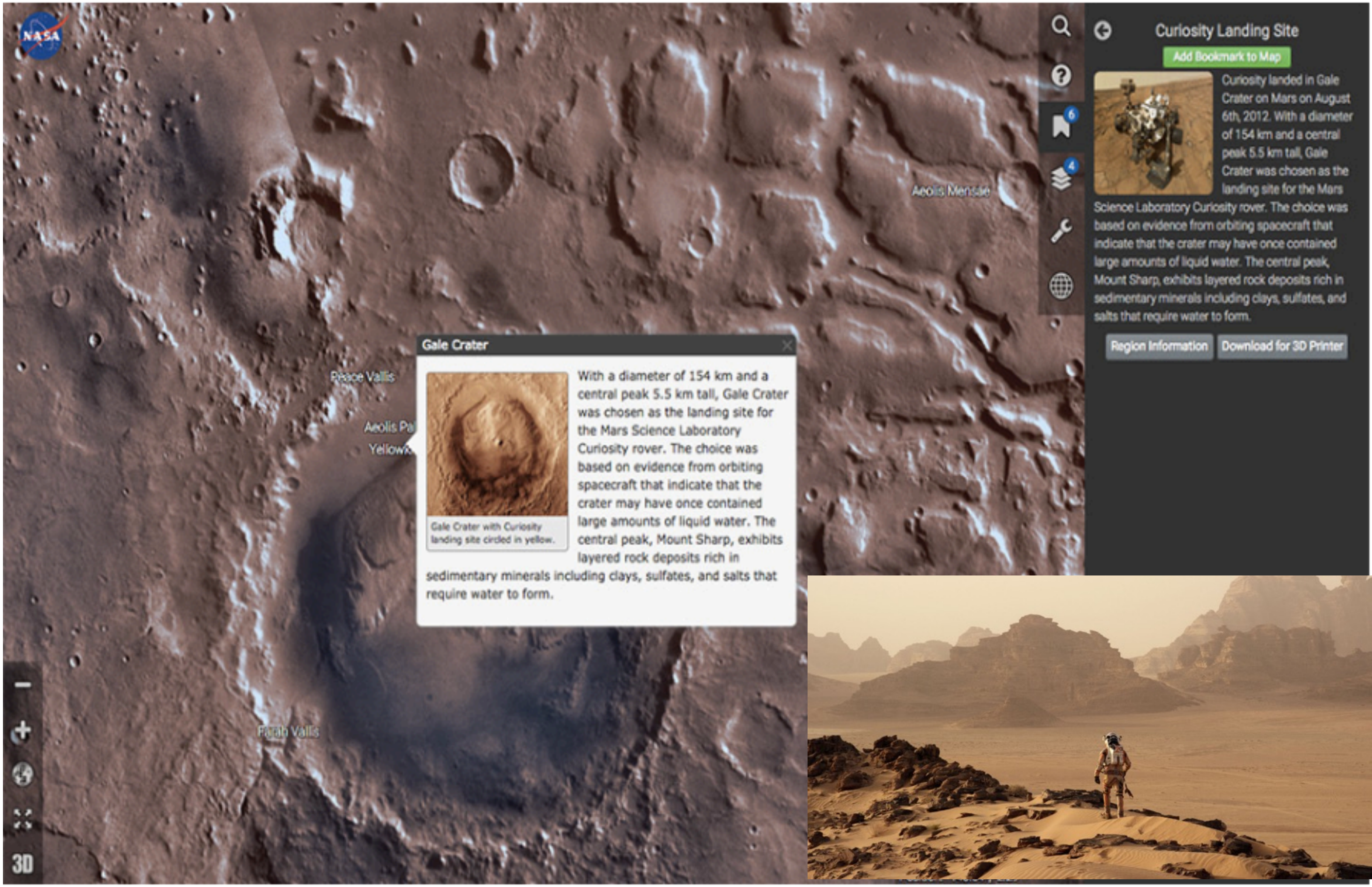
Search
(press ESC to close suggestions)

Results Per Page: 24

< 1 2 3 ... 1770 1771 > displaying 1 to 24 of 42484

RRB_491774570RNGLF0482542RHAZ00323M1
RLB_491774570RNGLF0482542RHAZ00323M1
FRB_491774536RNGLF0482542FHAZ00323M1
RLB_491774570RNGLT0482542RHAZ00323M1
FRB_491774536RNGLT0482542FHAZ00323M1
FLB_491774536RNGLT0482542FHAZ00323M1

Mars Trek: The Google Earth of Mars



Gale Crater

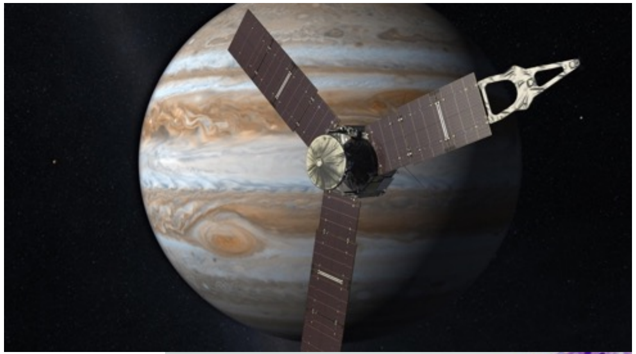
With a diameter of 154 km and a central peak 5.5 km tall, Gale Crater was chosen as the landing site for the Mars Science Laboratory Curiosity rover. The choice was based on evidence from orbiting spacecraft that indicate that the crater may have once contained large amounts of liquid water. The central peak, Mount Sharp, exhibits layered rock deposits rich in sedimentary minerals including clays, sulfates, and salts that require water to form.

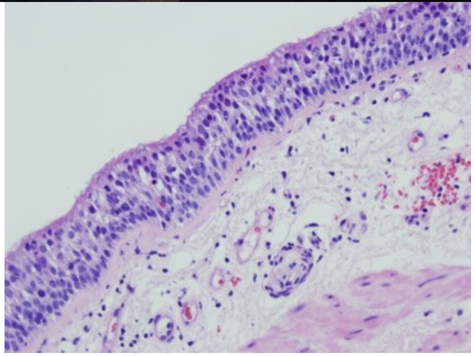
Curiosity Landing Site

Curiosity landed in Gale Crater on Mars on August 6th, 2012. With a diameter of 154 km and a central peak 5.5 km tall, Gale Crater was chosen as the landing site for the Mars Science Laboratory Curiosity rover. The choice was based on evidence from orbiting spacecraft that indicate that the crater may have once contained large amounts of liquid water. The central peak, Mount Sharp, exhibits layered rock deposits rich in sedimentary minerals including clays, sulfates, and salts that require water to form.

NCI/JPL Informatics Collaboration: Crossing Disciplines to Support Scientific Research

- Development of an advanced Knowledge System to *capture*, process, *share* and support *reproducible analysis* for biomarker research
  - Genomics, Proteomics, Imaging, etc data types of data
- NASA-NCI partnership, leveraging informatics and data science technologies from planetary and Earth science
  - Reproducible, Big Data Systems for exploring the universe
  - Software and data science methodology transfer based on JPL open source technologies and architectures





Biomarkers Knowledge Environment

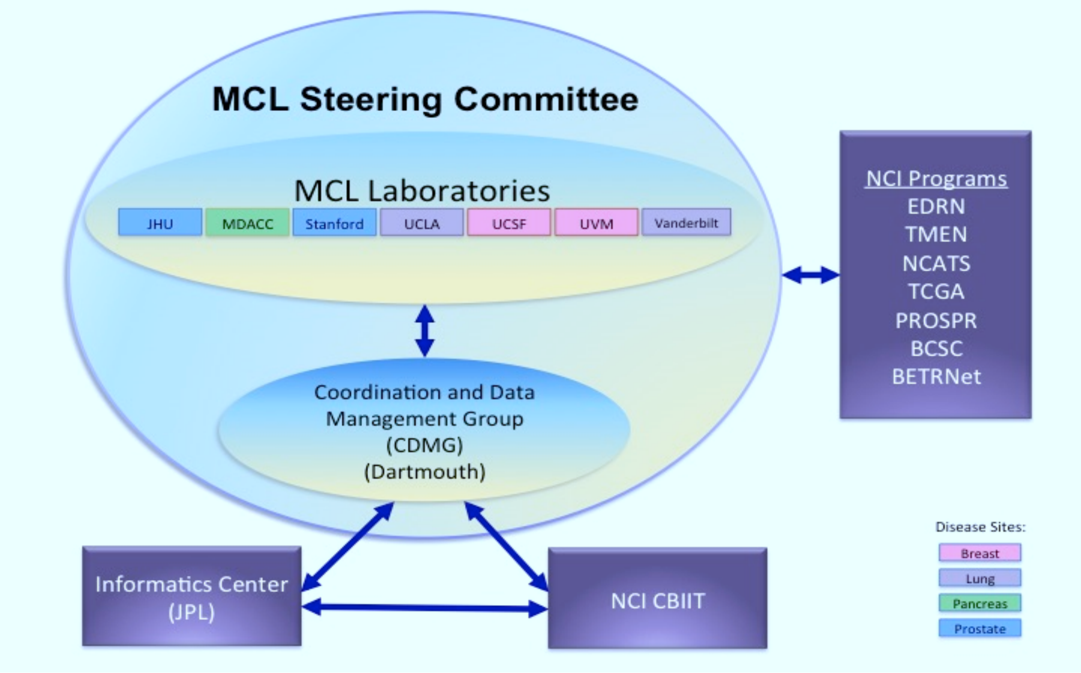
- Integrate diverse research into an online data environment
  - Integrate data as opposed to re-managing data
  - Be agnostic to data formats and structure

- Provide a well architected data management environment that captures, integrates, and shares data for biomarker research including:
  - Biomarkers
  - Biospecimens
  - Validation Study Information
  - Protocol Information
  - Study Results and Data Sets
  - Publications
  - Artifacts supporting collaboration
  - Curated metadata

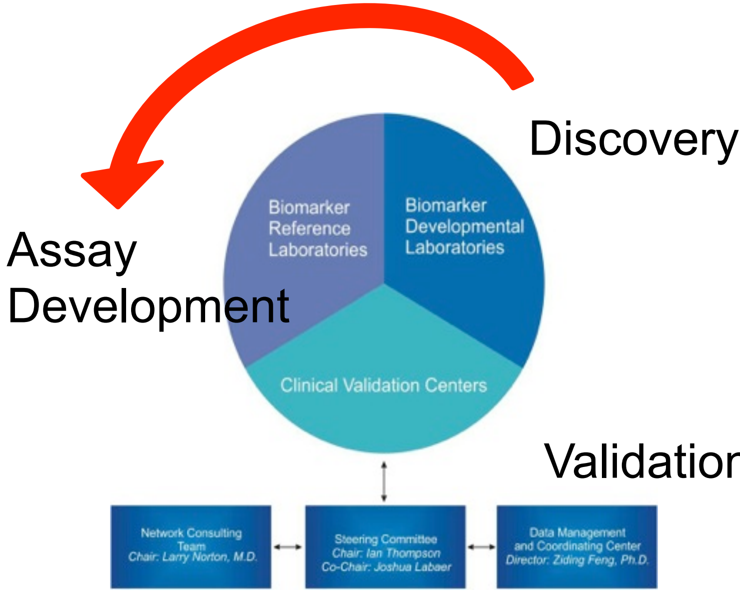
- Support diverse community needs

Two Key Programs

Organizational Structure of the MCL Consortium

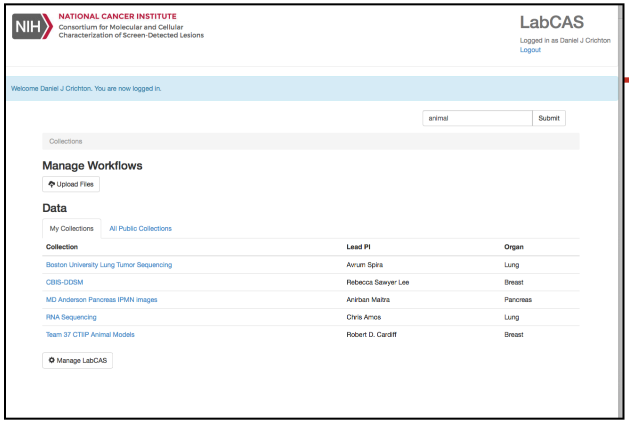


EDRN Organizational Structure


Discovery
Assay Development
Validation
Biomarker Reference Laboratories
Biomarker Developmental Laboratories
Clinical Validation Centers
Network Consulting Team
Chair: Larry Norton, M.D.
Steering Committee
Chair: Ian Thompson
Co-Chair: Joshua Labaer
Data Management and Coordinating Center
Director: Ziding Feng, Ph.D.

Develop cross-cutting informatics capabilities to support the capture, curation, management, distribution, and analysis of diverse data and results

Capture – Process – Search – Access – Analyze - Visualize



NIH NATIONAL CANCER INSTITUTE
Consortium for Molecular and Cellular Characterization of Screen-Detected Lesions

LabCAS
Logged in as Daniel J Crichton
Logout

Welcome Daniel J Crichton. You are now logged in.

animal | Submit

Collections

Manage Workflows

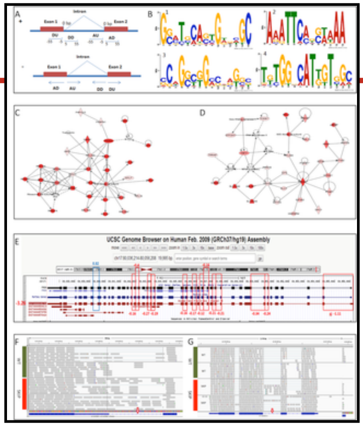
Upload Files

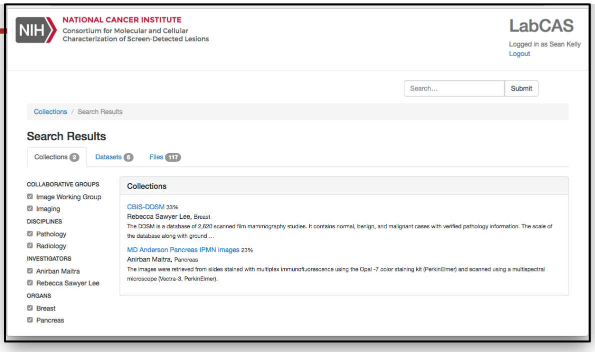
Data

My Collections | All Public Collections

Collection	Lead PI	Organ
Boston University Lung Tumor Sequencing	Avrum Spira	Lung
CBIS-DDSM	Rebecca Sawyer Lee	Breast
MD Anderson Pancreas IPMN images	Anirban Maitra	Pancreas
RNA Sequencing	Chris Amos	Lung
Team 37 CTIIP Animal Models	Robert D. Cardiff	Breast

Manage LabCAS





NIH NATIONAL CANCER INSTITUTE
Consortium for Molecular and Cellular Characterization of Screen-Detected Lesions

LabCAS
Logged in as Sean Kelly
Logout

Search... | Submit

Collections / Search Results

Search Results

Collections 2 | Datasets 6 | Files 117

COLLABORATIVE GROUPS
- ☑ Image Working Group
- ☑ Imaging

DISCIPLINES
- ☑ Pathology
- ☑ Radiology

INVESTIGATORS
- ☑ Anirban Maitra
- ☑ Rebecca Sawyer Lee

ORGANS
- ☑ Breast
- ☑ Pancreas

Collections

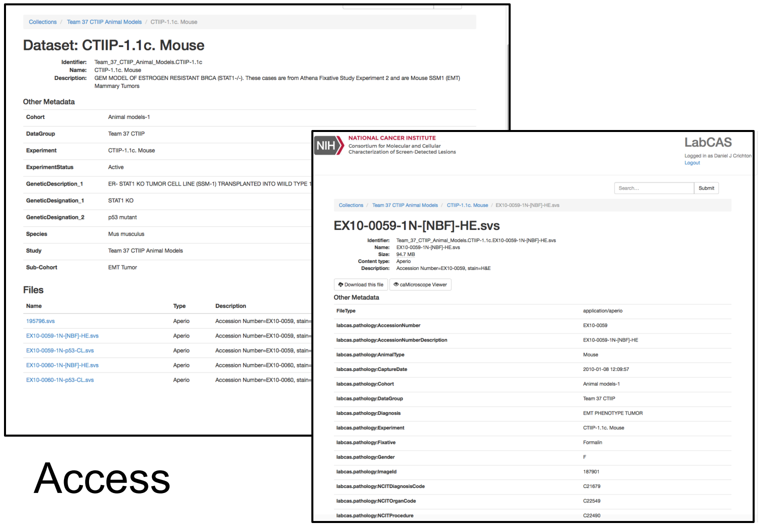
CBIS-DDSM 33%
Rebecca Sawyer Lee, Breast
The DDSM is a database of 2,620 scanned film mammography studies. It contains normal, benign, and malignant cases with verified pathology information. The scale of the database along with ground …

MD Anderson Pancreas IPMN images 23%
Anirban Maitra, Pancreas
The images were retrieved from slides stained with multiplex immunofluorescence using the Opal -7 color staining kit (PerkinElmer) and scanned using a multispectral microscope (Vectra-3, PerkinElmer).

Capture data and metadata

Process reproducibly

Search Metadata



Collections / Team 37 CTIIP Animal Models / CTIIP-1.1c. Mouse

Dataset: CTIIP-1.1c. Mouse

Identifier: Team_37_CTIIP_Animal_Models.CTIIP-1.1c
Name: CTIIP-1.1c. Mouse
Description: GEM MODEL OF ESTROGEN RESISTANT BRCA (STAT1-/-). These cases are from Athena Fixative Study Experiment 2 and are Mouse SSM1 (EMT) Mammary Tumors

Other Metadata

Cohort	Animal models-1
DataGroup	Team 37 CTIIP
Experiment	CTIIP-1.1c. Mouse
ExperimentStatus	Active
GeneticDescription_1	ER- STAT1 KO TUMOR CELL LINE (SSM-1) TRANSPLANTED INTO WIILD TYPE 1
GeneticDesignation_1	STAT1 KO
GeneticDesignation_2	p53 mutant
Species	Mus musculus
Study	Team 37 CTIIP Animal Models
Sub-Cohort	EMT Tumor

Files

Name	Type	Description
195796.svs	Aperio	Accession Number=EX10-0059, stain=
EX10-0059-1N-[NBF]-HE.svs	Aperio	Accession Number=EX10-0059, stain=
EX10-0059-1N-p53-CL.svs	Aperio	Accession Number=EX10-0059, stain=
EX10-0060-1N-[NBF]-HE.svs	Aperio	Accession Number=EX10-0060, stain=
EX10-0060-1N-p53-CL.svs	Aperio	Accession Number=EX10-0060, stain=

NIH NATIONAL CANCER INSTITUTE
Consortium for Molecular and Cellular Characterization of Screen-Detected Lesions

LabCAS
Logged in as Daniel J Crichton
Logout

Search... | Submit

Collections / Team 37 CTIIP Animal Models / CTIIP-1.1c. Mouse / EX10-0059-1N-[NBF]-HE.svs

EX10-0059-1N-[NBF]-HE.svs

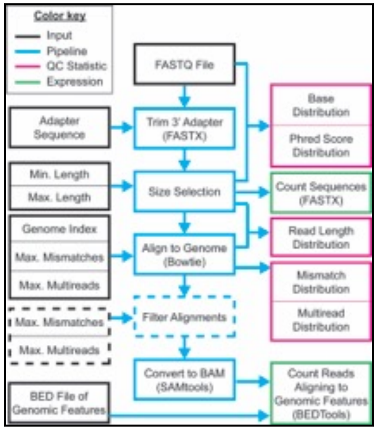
Identifier: Team_37_CTIIP_Animal_Models.CTIIP-1.1c.EX10-0059-1N-[NBF]-HE.svs
Name: EX10-0059-1N-[NBF]-HE.svs
Size: 94.7 MB
Content type: Aperio
Description: Accession Number=EX10-0059, stain=H&E

Download this file | caMicroscope Viewer

Other Metadata

FileType	application/aperio
labcas.pathology:AccessionNumber	EX10-0059
labcas.pathology:AccessionNumberDescription	EX10-0059-1N-[NBF]-HE
labcas.pathology:AnimalType	Mouse
labcas.pathology:CaptureDate	2010-01-08 12:09:57
labcas.pathology:Cohort	Animal models-1
labcas.pathology:DataGroup	Team 37 CTIIP
labcas.pathology:Diagnosis	EMT PHENOTYPE TUMOR
labcas.pathology:Experiment	CTIIP-1.1c. Mouse
labcas.pathology:Fixative	Formalin
labcas.pathology:Gender	F
labcas.pathology:ImageId	187901
labcas.pathology:NCITDiagnosisCode	C21679
labcas.pathology:NCITOrganCode	C22549
labcas.pathology:NCITProcedure	C22490

Access



Color key
- Input
- Pipeline
- QC Statistic
- Expression

FASTQ File

Adapter Sequence

Trim 3' Adapter (FASTX)

Min. Length

Max. Length

Size Selection

Genome Index

Max. Mismatches

Max. Multireads

Align to Genome (Bowtie)

Max. Mismatches

Max. Multireads

Filter Alignments

BED File of Genomic Features

Convert to BAM (SAMtools)

Base Distribution

Phred Score Distribution

Count Sequences (FASTX)

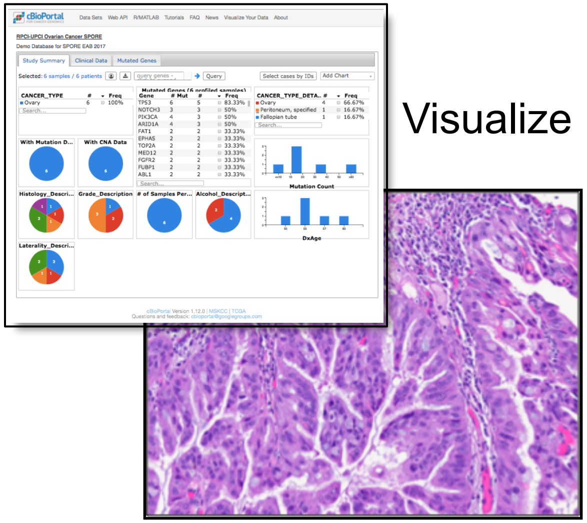
Read Length Distribution

Mismatch Distribution

Multiread Distribution

Count Reads Aligning to Genomic Features (BEDTools)

Analyze with algorithms, pipelines



cBioPortal FOR CANCER GENOMICS

Data Sets | Web API | R/MATLAB | Tutorials | FAQ | News | Visualize Your Data | About

RPCI-UPCI Ovarian Cancer SPORE

Demo Database for SPORE EAB 2017

Study Summary | Clinical Data | Mutated Genes

Selected: 6 samples / 6 patients | query genes | Query | Select cases by IDs | Add Chart

CANCER_TYPE	#	Freq
Ovary	6	100%

Mutated Genes (6 profiled samples)

Gene	# Mut	#	Freq
TP53	6	5	83.33%
NOTCH3	3	3	50%
PIK3CA	4	3	50%
ARID1A	4	3	50%
FAT1	2	2	33.33%
EPHA5	2	2	33.33%
TOP2A	2	2	33.33%
MED12	2	2	33.33%
FGFR2	2	2	33.33%
FUBP1	2	2	33.33%
ABL1	2	2	33.33%

CANCER_TYPE_DETA..	#	Freq
Ovary	4	66.67%
Peritoneum, specified	1	16.67%
Fallopian tube	1	16.67%

cBioPortal Version 1.12.0 | MSKCC | TCGA
Questions and feedback: cbioportal@googlegroups.com

Visualize

NIH NATIONAL CANCER INSTITUTE

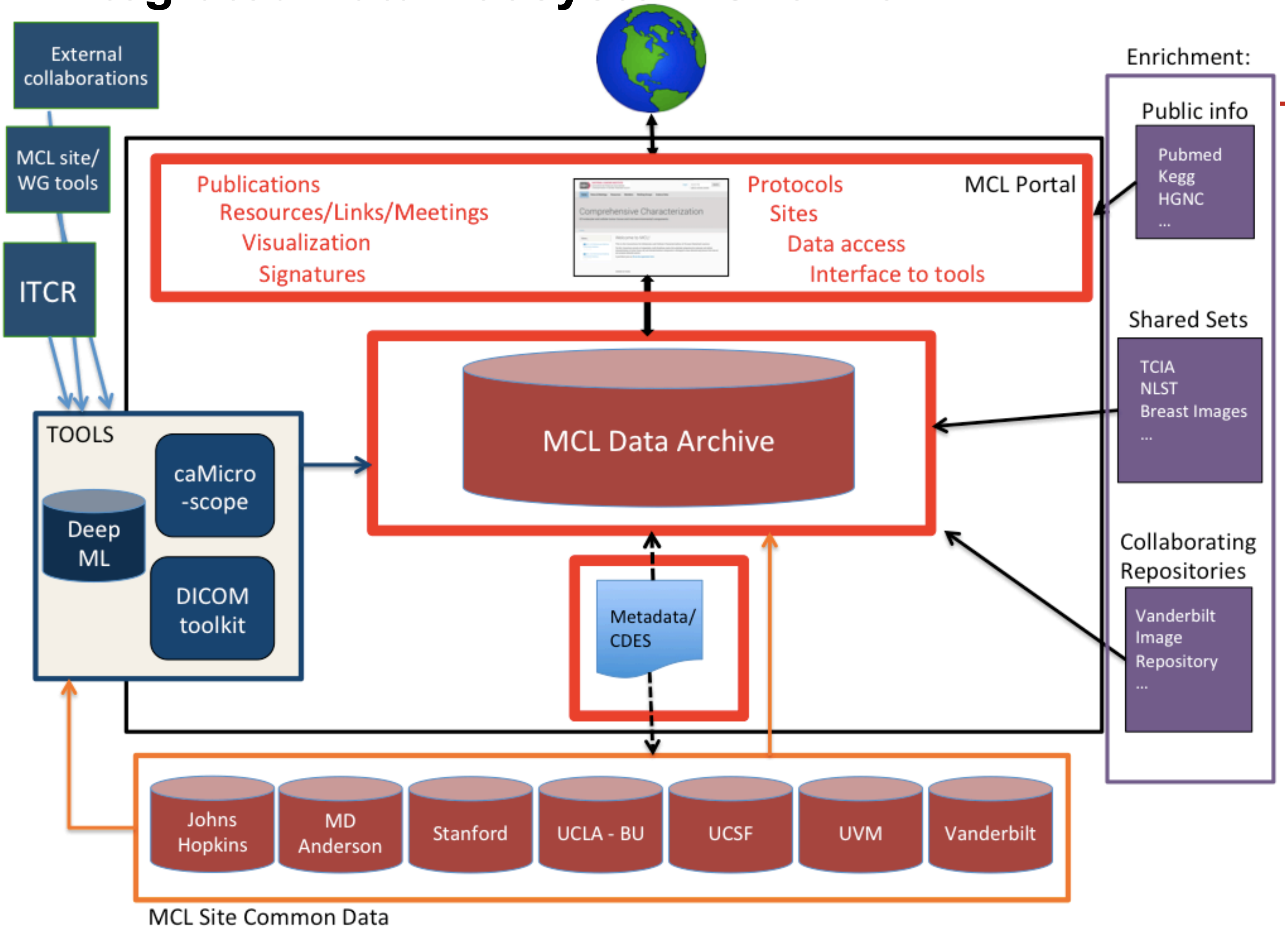
NASA JPL

UNC SCHOOL OF MEDICINE

BCM Baylor College of Medicine

Dartmouth GEISEL SCHOOL OF MEDICINE

MCL Integrated Data Ecosystem Overview



External collaborations

MCL site/ WG tools

ITCR

TOOLS

caMicro-scope

Deep ML

DICOM toolkit

MCL Portal

Publications

Resources/Links/Meetings

Visualization

Signatures

Protocols

Sites

Data access

Interface to tools

Comprehensive Characterization

MCL Data Archive

Metadata/ CDES

Johns Hopkins

MD Anderson

Stanford

UCLA - BU

UCSF

UVM

Vanderbilt

MCL Site Common Data

Enrichment:

Public info

Pubmed
Kegg
HGNC
…

Shared Sets

TCIA
NLST
Breast Images
…

Collaborating Repositories

Vanderbilt Image Repository
…

NIH NATIONAL CANCER INSTITUTE

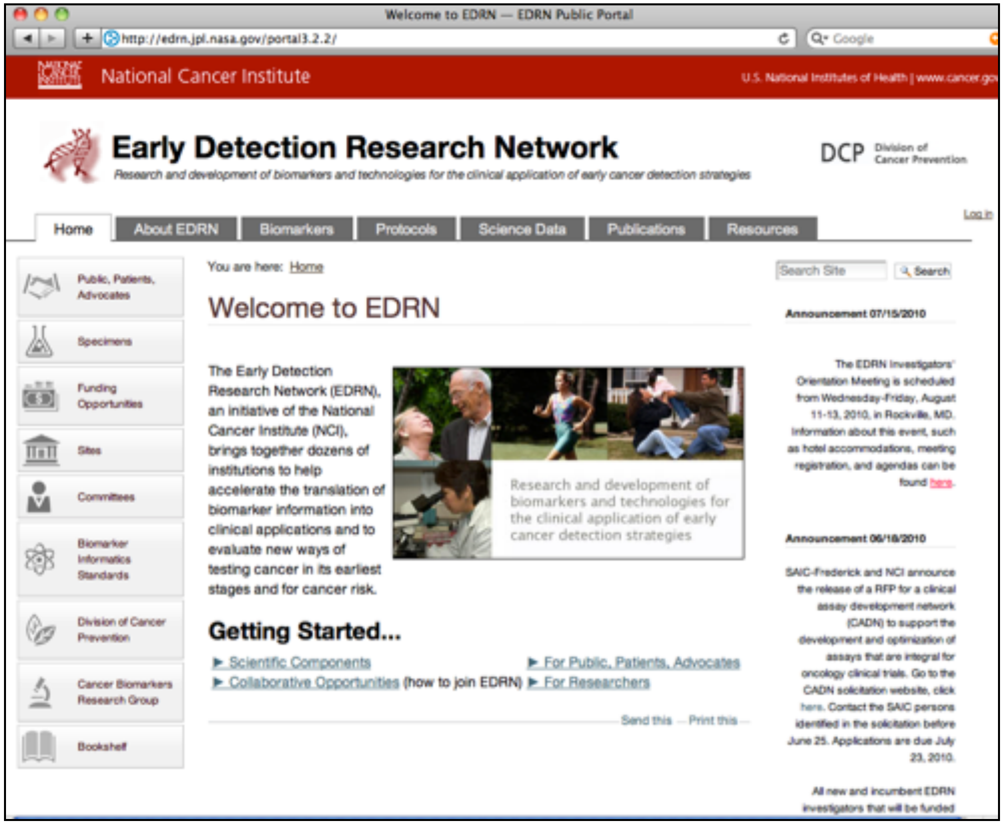
NASA JPL

UNC SCHOOL OF MEDICINE

BCM Baylor College of Medicine

Dartmouth GEISEL SCHOOL OF MEDICINE

Portal: Dissemination and Access to Biomarker Data


Welcome to EDRN — EDRN Public Portal
http://edrn.jpl.nasa.gov/portal3.2.2/
Google
National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov
DCP Division of Cancer Prevention
Log in
Home
About EDRN
Biomarkers
Protocols
Science Data
Publications
Resources
Public, Patients, Advocates
Specimens
Funding Opportunities
Sites
Committees
Biomarker Informatics Standards
Division of Cancer Prevention
Cancer Biomarkers Research Group
Bookshelf
Search Site
Search
Send this
Print this

Early Detection Research Network

Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies

You are here: Home

Welcome to EDRN

The Early Detection Research Network (EDRN), an initiative of the National Cancer Institute (NCI), brings together dozens of institutions to help accelerate the translation of biomarker information into clinical applications and to evaluate new ways of testing cancer in its earliest stages and for cancer risk.

Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies

Getting Started...

- ▶ Scientific Components
- ▶ Collaborative Opportunities (how to join EDRN)
- ▶ For Public, Patients, Advocates
- ▶ For Researchers

Announcement 07/15/2010

The EDRN Investigators' Orientation Meeting is scheduled from Wednesday-Friday, August 11-13, 2010, in Rockville, MD. Information about this event, such as hotel accommodations, meeting registration, and agendas can be found here.

Announcement 06/18/2010

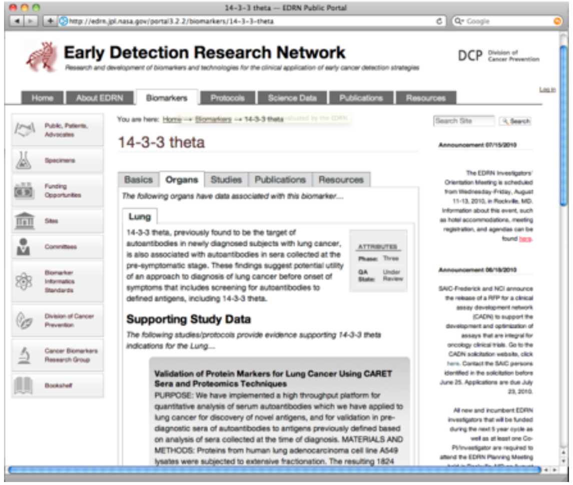
SAIC-Frederick and NCI announce the release of a RFP for a clinical assay development network (CADN) to support the development and optimization of assays that are integral for oncology clinical trials. Go to the CADN solicitation website, click here. Contact the SAIC persons identified in the solicitation before June 25. Applications are due July 23, 2010.

All new and incumbent EDRN investigators that will be funded

- Gateway to information
- Information managed both within and outside the knowledge system
- Initial starting point for community to get to research data
- Google-like search to access the wealth of data
- Multi-level Security

http://cancer.gov/edrn
http://mcl.jpl.nasa.gov

Navigating the Knowledge System: Data Semantically Linked


14-3-3 theta — EDRN Public Portal
http://edrn.jpl.nasa.gov/portal3.2.2/biomarkers/14-3-3-theta
Google
Early Detection Research Network
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
DCP Division of Cancer Prevention
Home
About EDRN
Biomarkers
Protocols
Science Data
Publications
Resources
Log in
Public, Patients, Advocates
Specimens
Funding Opportunities
Sites
Committees
Biomarker Informatics Standards
Division of Cancer Prevention
Cancer Biomarkers Research Group
Bookshelf
Search Site
Search
Basics
Organs
Studies
Publications
Resources

You are here: Home → Biomarkers → 14-3-3 theta

14-3-3 theta

The following organs have data associated with this biomarker....

Lung

14-3-3 theta, previously found to be the target of autoantibodies in newly diagnosed subjects with lung cancer, is also associated with autoantibodies in sera collected at the pre-symptomatic stage. These findings suggest potential utility of an approach to diagnosis of lung cancer before onset of symptoms that includes screening for autoantibodies to defined antigens, including 14-3-3 theta.

ATTRIBUTES	
Phase:	Three
QA State:	Under Review

Supporting Study Data

The following studies/protocols provide evidence supporting 14-3-3 theta indications for the Lung....

Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques

PURPOSE: We have implemented a high throughput platform for quantitative analysis of serum autoantibodies which we have applied to lung cancer for discovery of novel antigens, and for validation in pre-diagnostic sera of autoantibodies to antigens previously defined based on analysis of sera collected at the time of diagnosis. MATERIALS AND METHODS: Proteins from human lung adenocarcinoma cell line A549 lysates were subjected to extensive fractionation. The resulting 1824

Announcement 07/15/2010

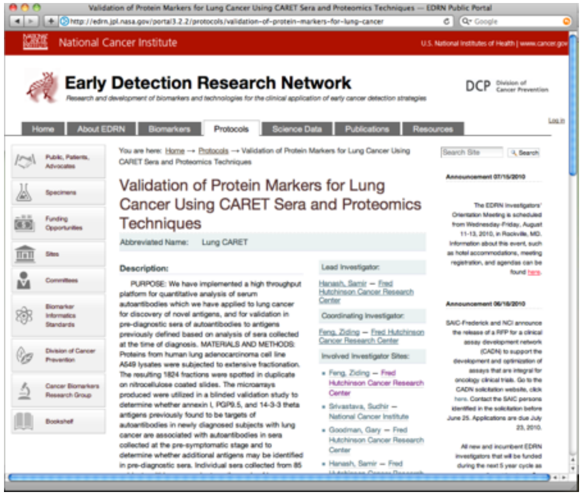
The EDRN Investigators' Orientation Meeting is scheduled from Wednesday-Friday, August 11-13, 2010, in Rockville, MD. Information about this event, such as hotel accommodations, meeting registration, and agendas can be found here.

Announcement 06/18/2010

SAIC-Frederick and NCI announce the release of a RFP for a clinical assay development network (CADN) to support the development and optimization of assays that are integral for oncology clinical trials. Go to the CADN solicitation website, click here. Contact the SAIC persons identified in the solicitation before June 25. Applications are due July 23, 2010.

All new and incumbent EDRN investigators that will be funded during the next 5 year cycle as well as at least one Co-PI/investigator are required to attend the EDRN Planning Meeting

Biomarker Annotations


Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques — EDRN Public Portal
http://edrn.jpl.nasa.gov/portal3.2.2/protocols/validation-of-protein-markers-for-lung-cancer
Google
National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov
Early Detection Research Network
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
DCP Division of Cancer Prevention
Home
About EDRN
Biomarkers
Protocols
Science Data
Publications
Resources
Log in
Public, Patients, Advocates
Specimens
Funding Opportunities
Sites
Committees
Biomarker Informatics Standards
Division of Cancer Prevention
Cancer Biomarkers Research Group
Bookshelf
Search Site
Search

You are here: Home → Protocols → Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques

Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques

Abbreviated Name: Lung CARET

Description:

PURPOSE: We have implemented a high throughput platform for quantitative analysis of serum autoantibodies which we have applied to lung cancer for discovery of novel antigens, and for validation in pre-diagnostic sera of autoantibodies to antigens previously defined based on analysis of sera collected at the time of diagnosis. MATERIALS AND METHODS: Proteins from human lung adenocarcinoma cell line A549 lysates were subjected to extensive fractionation. The resulting 1824 fractions were spotted in duplicate on nitrocellulose coated slides. The microarrays produced were utilized in a blinded validation study to determine whether annexin I, PGP9.5, and 14-3-3 theta antigens previously found to be targets of autoantibodies in newly diagnosed subjects with lung cancer are associated with autoantibodies in sera collected at the pre-symptomatic stage and to determine whether additional antigens may be identified in pre-diagnostic sera. Individual sera collected from 85

Lead Investigator:

Hanash, Samir — Fred Hutchinson Cancer Research Center

Coordinating Investigator:

Feng, Ziding — Fred Hutchinson Cancer Research Center

Involved Investigator Sites:

- Feng, Ziding — Fred Hutchinson Cancer Research Center
- Srivastava, Sudhir — National Cancer Institute
- Goodman, Gary — Fred Hutchinson Cancer Research Center
- Hanash, Samir — Fred

Announcement 07/15/2010

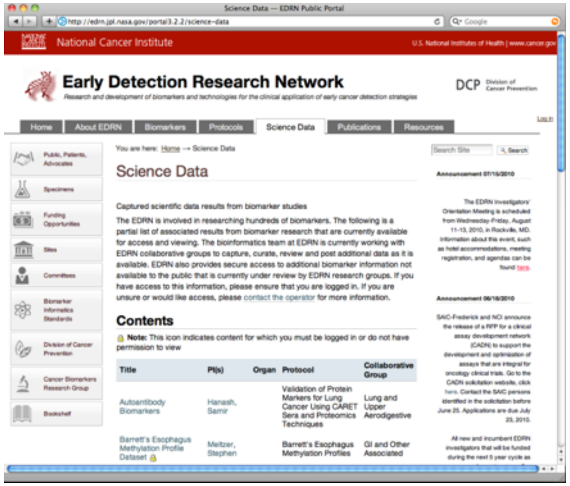
The EDRN Investigators' Orientation Meeting is scheduled from Wednesday-Friday, August 11-13, 2010, in Rockville, MD. Information about this event, such as hotel accommodations, meeting registration, and agendas can be found here.

Announcement 06/18/2010

SAIC-Frederick and NCI announce the release of a RFP for a clinical assay development network (CADN) to support the development and optimization of assays that are integral for oncology clinical trials. Go to the CADN solicitation website, click here. Contact the SAIC persons identified in the solicitation before June 25. Applications are due July 23, 2010.

All new and incumbent EDRN investigators that will be funded during the next 5 year cycle as

Protocols


Science Data — EDRN Public Portal
http://edrn.jpl.nasa.gov/portal3.2.2/science-data
Google
National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov
Early Detection Research Network
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
DCP Division of Cancer Prevention
Home
About EDRN
Biomarkers
Protocols
Science Data
Publications
Resources
Log in
Public, Patients, Advocates
Specimens
Funding Opportunities
Sites
Committees
Biomarker Informatics Standards
Division of Cancer Prevention
Cancer Biomarkers Research Group
Bookshelf
Search Site
Search

You are here: Home → Science Data

Science Data

Captured scientific data results from biomarker studies

The EDRN is involved in researching hundreds of biomarkers. The following is a partial list of associated results from biomarker research that are currently available for access and viewing. The bioinformatics team at EDRN is currently working with EDRN collaborative groups to capture, curate, review and post additional data as it is available. EDRN also provides secure access to additional biomarker information not available to the public that is currently under review by EDRN research groups. If you have access to this information, please ensure that you are logged in. If you are unsure or would like access, please contact the operator for more information.

Contents

🔒 **Note:** This icon indicates content for which you must be logged in or do not have permission to view

Title	PI(s)	Organ	Protocol	Collaborative Group
Autoantibody Biomarkers	Hanash, Samir		Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques	Lung and Upper Aerodigestive
Barrett's Esophagus Methylation Profile Dataset 🔒	Meltzer, Stephen		Barrett's Esophagus Methylation Profiles	GI and Other Associated

Announcement 07/15/2010

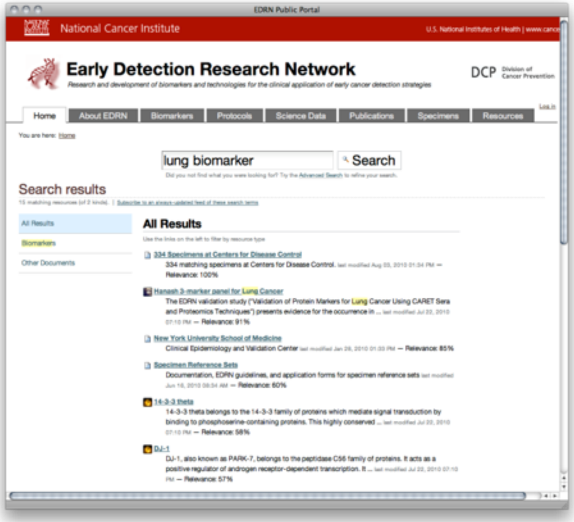
The EDRN Investigators' Orientation Meeting is scheduled from Wednesday-Friday, August 11-13, 2010, in Rockville, MD. Information about this event, such as hotel accommodations, meeting registration, and agendas can be found here.

Announcement 06/18/2010

SAIC-Frederick and NCI announce the release of a RFP for a clinical assay development network (CADN) to support the development and optimization of assays that are integral for oncology clinical trials. Go to the CADN solicitation website, click here. Contact the SAIC persons identified in the solicitation before June 25. Applications are due July 23, 2010.

All new and incumbent EDRN investigators that will be funded during the next 5 year cycle as

Biomarker Data Results


EDRN Public Portal
National Cancer Institute
U.S. National Institutes of Health | www.cance
Early Detection Research Network
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
DCP Division of Cancer Prevention
Home
About EDRN
Biomarkers
Protocols
Science Data
Publications
Specimens
Resources
Log in
Search
All Results
Biomarkers
Other Documents

You are here: Home

lung biomarker

Did you not find what you were looking for? Try the Advanced Search to refine your search.

Search results

15 matching resources (of 2 kinds). | Subscribe to an always-updated feed of these search terms

All Results

Use the links on the left to filter by resource type

334 Specimens at Centers for Disease Control
334 matching specimens at Centers for Disease Control. last modified Aug 03, 2010 01:34 PM — Relevance: 100%

Hanash 3-marker panel for Lung Cancer
The EDRN validation study ("Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques") presents evidence for the occurrence in ... last modified Jul 22, 2010 07:10 PM — Relevance: 91%

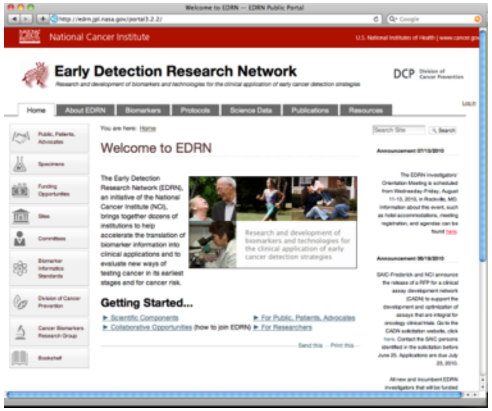
New York University School of Medicine
Clinical Epidemiology and Validation Center last modified Jan 28, 2010 01:33 PM — Relevance: 85%

Specimen Reference Sets
Documentation, EDRN guidelines, and application forms for specimen reference sets last modified Jun 16, 2010 08:34 AM — Relevance: 60%

14-3-3 theta
14-3-3 theta belongs to the 14-3-3 family of proteins which mediate signal transduction by binding to phosphoserine-containing proteins. This highly conserved ... last modified Jul 22, 2010 07:10 PM — Relevance: 58%

DJ-1
DJ-1, also known as PARK-7, belongs to the peptidase C56 family of proteins. It acts as a positive regulator of androgen receptor-dependent transcription. It ... last modified Jul 22, 2010 07:10 PM — Relevance: 57%

Specimens


Welcome to EDRN — EDRN Public Portal
http://edrn.jpl.nasa.gov/portal3.2.2/
Google
National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov
Early Detection Research Network
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
DCP Division of Cancer Prevention
Home
About EDRN
Biomarkers
Protocols
Science Data
Publications
Resources
Log in
Public, Patients, Advocates
Specimens
Funding Opportunities
Sites
Committees
Biomarker Informatics Standards
Division of Cancer Prevention
Cancer Biomarkers Research Group
Bookshelf
Search Site
Search
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
Send this
Print this

You are here: Home

Welcome to EDRN

The Early Detection Research Network (EDRN), an initiative of the National Cancer Institute (NCI), brings together dozens of institutions to help accelerate the translation of biomarker information into clinical applications and to evaluate new ways of testing cancer in its earliest stages and for cancer risk.

Getting Started...

- Scientific Components
- Collaborative Opportunities (how to join EDRN)
- For Public, Patients, Advocates
- For Researchers

Announcement 07/15/2010

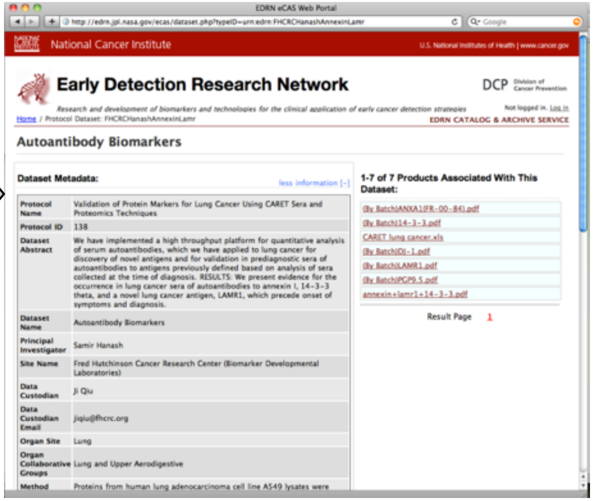
The EDRN Investigators' Orientation Meeting is scheduled from Wednesday-Friday, August 11-13, 2010, in Rockville, MD. Information about this event, such as hotel accommodations, meeting registration, and agendas can be found here.

Announcement 06/18/2010

SAIC-Frederick and NCI announce the release of a RFP for a clinical assay development network (CADN) to support the development and optimization of assays that are integral for oncology clinical trials. Go to the CADN solicitation website, click here. Contact the SAIC persons identified in the solicitation before June 25. Applications are due July 23, 2010.

All new and incumbent EDRN investigators that will be funded

Linked through Public Portal


EDRN eCAS Web Portal
http://edrn.jpl.nasa.gov/ecas/dataset.php?typeID=urn:edrn:FHCRCHanashAnnexinLamr
Google
National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov
Early Detection Research Network
Research and development of biomarkers and technologies for the clinical application of early cancer detection strategies
DCP Division of Cancer Prevention
Not logged in. Log in
EDRN CATALOG & ARCHIVE SERVICE
less information [-]

Home / Protocol Dataset: FHCRCHanashAnnexinLamr

Autoantibody Biomarkers

Dataset Metadata:

Protocol Name	Validation of Protein Markers for Lung Cancer Using CARET Sera and Proteomics Techniques
Protocol ID	138
Dataset Abstract	We have implemented a high throughput platform for quantitative analysis of serum autoantibodies, which we have applied to lung cancer for discovery of novel antigens and for validation in prediagnostic sera of autoantibodies to antigens previously defined based on analysis of sera collected at the time of diagnosis. RESULTS: We present evidence for the occurrence in lung cancer sera of autoantibodies to annexin I, 14-3-3 theta, and a novel lung cancer antigen, LAMR1, which precede onset of symptoms and diagnosis.
Dataset Name	Autoantibody Biomarkers
Principal Investigator	Samir Hanash
Site Name	Fred Hutchinson Cancer Research Center (Biomarker Developmental Laboratories)
Data Custodian	Ji Qiu
Data Custodian Email	jiqiu@fhcrc.org
Organ Site	Lung
Organ Collaborative Groups	Lung and Upper Aerodigestive
Method	Proteins from human lung adenocarcinoma cell line A549 lysates were

1-7 of 7 Products Associated With This Dataset:

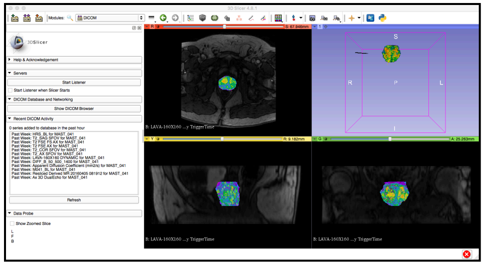
- (By Batch)ANXA1(FR-00-84).pdf
- (By Batch)14-3-3.pdf
- CARET lung cancer.xls
- (By Batch)DJ-1.pdf
- (By Batch)LAMR1.pdf
- (By Batch)PGP9.5.pdf
- annexin+lamr1+14-3-3.pdf

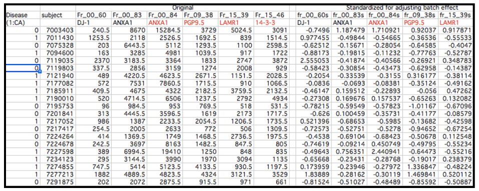
Result Page 1

Access to download data

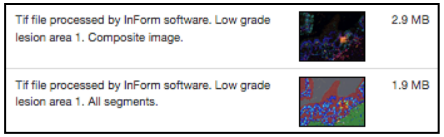
Data Models and Elements

Data generated in the course of research from any source



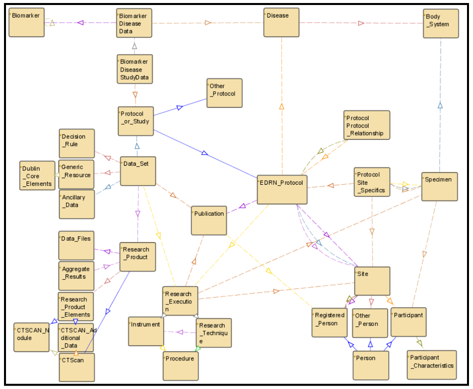


Disease	subject	Original							Standardized for adjusting batch effect				
Disease	subject	Fr_00_60	Fr_00_83	Fr_00_84	Fr_09_38	Fr_15_39	Fr_15_46		fr_00_60s	fr_00_83s	fr_00_84s	fr_09_38s	fr_15_39s
(1:CA)		DJ-1	ANXA1	ANXA1	PGP9.5	LAMR1	14-3-3		DJ-1	ANXA1	ANXA1	PGP9.5	LAMR1
0	7003403	240.5	8670	15284.5	3729	5024.5	3091		-0.7496	1.187479	1.710921	0.92037	0.917871
1	7011430	1253.5	2118	2526.5	1692.5	839	1514.5		0.977455	-0.49844	-0.54665	-0.36536	-0.55533
0	7075328	203	6443.5	5112	1293.5	1100	2598.5		-0.62512	-0.15671	-0.28054	-0.64585	-0.4047
1	7094600	163	3285	4981	1039.5	917	1722		-0.88173	-0.19815	-0.11232	-0.77763	-0.52787
0	7119035	2370	3183.5	3364	1833	2747	3872		2.555053	-0.41874	-0.40566	-0.26921	0.348783
0	7119803	337.5	2856	3159	1274	2008	929		-0.58423	-0.30854	-0.43473	-0.62958	-0.14387
1	7121940	489	4220.5	4623.5	2671.5	1151.5	2028.5		-0.2054	-0.33539	-0.3155	0.316177	-0.38114
1	7177082	572	7531	7860.5	1715.5	910	1066.5		-0.0836	-0.0693	-0.08381	-0.35124	-0.49162
1	7185911	409.5	4675	4322	2182.5	3759.5	2132.5		-0.46147	0.159512	-0.22893	-0.056	0.47262
1	7190010	520	4714.5	6506	1237.5	2792	4934		-0.27308	0.169676	0.157537	-0.65263	0.132082
0	7195753	96	984.5	953	769.5	518	531.5		-0.78215	-0.59549	-0.57823	-1.01167	-0.67096
0	7201841	313	4445.5	3596.5	1619	2173	1717.5		-0.626	0.100459	-0.35731	-0.41177	-0.08579
1	7217052	986	1387	2233.5	2054.5	1206.5	1735.5		0.521396	-0.68653	-0.5985	-0.13682	-0.42598
0	7217417	254.5	2005	2633	772	506	1309.5		-0.72573	-0.52751	-0.5278	-0.94652	-0.67254
0	7224264	414	1369.5	1749	1468.5	2736.5	1975.5		-0.4538	-0.69104	-0.68423	-0.50678	0.112548
0	7224678	242.5	3697	8163	1482.5	847.5	805		-0.74619	-0.09214	0.450749	-0.49795	-0.55234
1	7227598	389	6994.5	19410	1250	848	835		-0.49643	0.756351	2.440941	-0.64473	-0.55216
1	7234123	295	3144.5	3990	1970	3094	1135		-0.65668	-0.23431	-0.28768	-0.19017	0.238379
1	7274855	747.5	5414	5123.5	4133.5	930.5	1197.5		0.173959	-0.23946	-0.27972	1.336847	-0.48224
1	7277213	1882	4889.5	4823.5	4324	3121.5	3529		1.83889	-0.28162	-0.30119	1.469841	0.520112
0	7291875	202	2072	2875.5	915.5	971	661		-0.81524	-0.51027	-0.48489	-0.85592	-0.50887

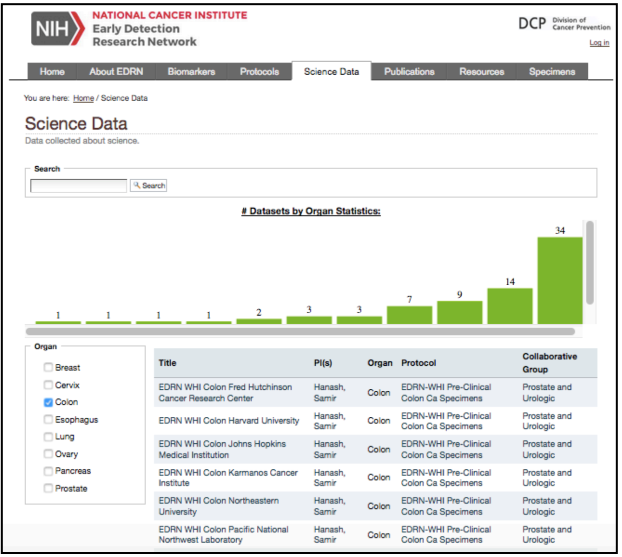


Tif file processed by InForm software. Low grade lesion area 1. Composite image.	2.9 MB
Tif file processed by InForm software. Low grade lesion area 1. All segments.	1.9 MB

Instruments, etc.



Data linked across the Data Commons



You are here: Home / Science Data

Science Data

Data collected about science.

Datasets by Organ Statistics:

Organ

□ Breast
□ Cervix
☑ Colon
□ Esophagus
□ Lung
□ Ovary
□ Pancreas
□ Prostate

Title	PI(s)	Organ	Protocol	Collaborative Group
EDRN WHI Colon Fred Hutchinson Cancer Research Center	Hanash, Samir	Colon	EDRN-WHI Pre-Clinical Colon Ca Specimens	Prostate and Urologic
EDRN WHI Colon Harvard University	Hanash, Samir	Colon	EDRN-WHI Pre-Clinical Colon Ca Specimens	Prostate and Urologic
EDRN WHI Colon Johns Hopkins Medical Institution	Hanash, Samir	Colon	EDRN-WHI Pre-Clinical Colon Ca Specimens	Prostate and Urologic
EDRN WHI Colon Karmanos Cancer Institute	Hanash, Samir	Colon	EDRN-WHI Pre-Clinical Colon Ca Specimens	Prostate and Urologic
EDRN WHI Colon Northeastern University	Hanash, Samir	Colon	EDRN-WHI Pre-Clinical Colon Ca Specimens	Prostate and Urologic
EDRN WHI Colon Pacific National Northwest Laboratory	Hanash, Samir	Colon	EDRN-WHI Pre-Clinical Colon Ca Specimens	Prostate and Urologic

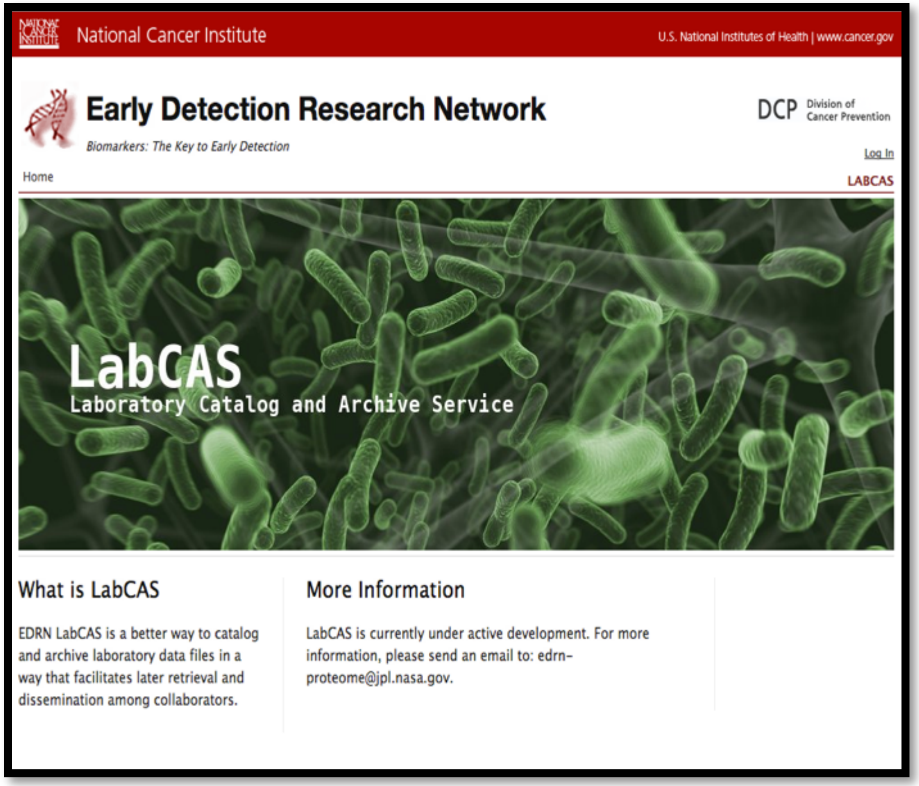
Data is linked and available for integration of different analytical methods and tools to drive data-driven discovery.

Addition of data sets over time – any data type generated

Drives a consistent data architecture across the consortium.

LabCAS: Capturing and Sharing Science Data Analysis Data and Pipelines

- A secure, reliable means to capture, process and manage data
- Plug in analytical methods
  - Repeatable data processing pipelines
- Integrate visualization



National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov

Early Detection Research Network

Biomarkers: The Key to Early Detection

DCP Division of Cancer Prevention

Log In

Home

LABCAS

LabCAS

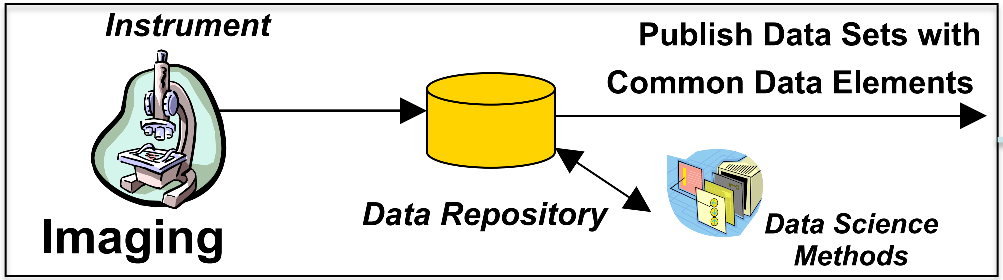
Laboratory Catalog and Archive Service

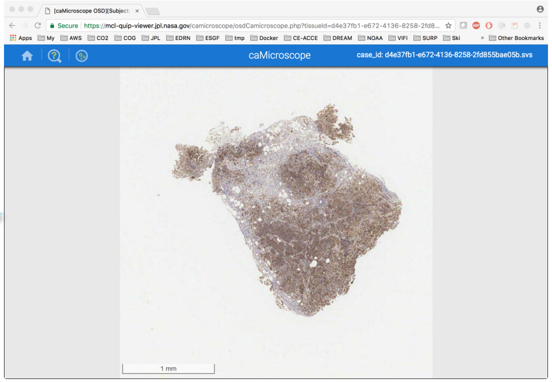
What is LabCAS

EDRN LabCAS is a better way to catalog and archive laboratory data files in a way that facilitates later retrieval and dissemination among collaborators.

More Information

LabCAS is currently under active development. For more information, please send an email to: edrn-proteome@jpl.nasa.gov.





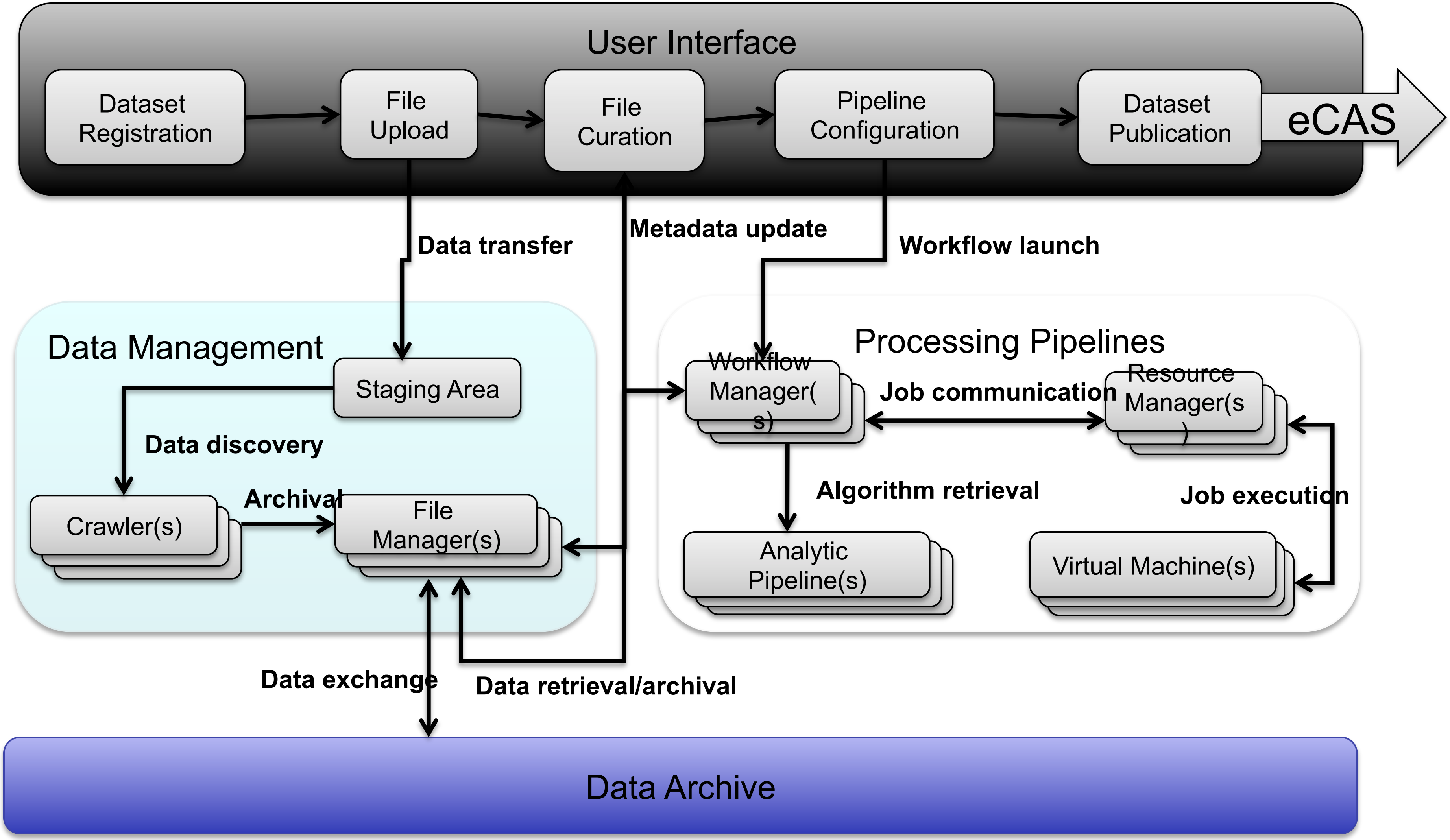
[caMicroscope OSD][Subject:

Secure https://mcl-quip-viewer.jpl.nasa.gov/camicroscope/osdCamicroscope.php?tissueId=d4e37fb1-e672-4136-8258-2fd8...

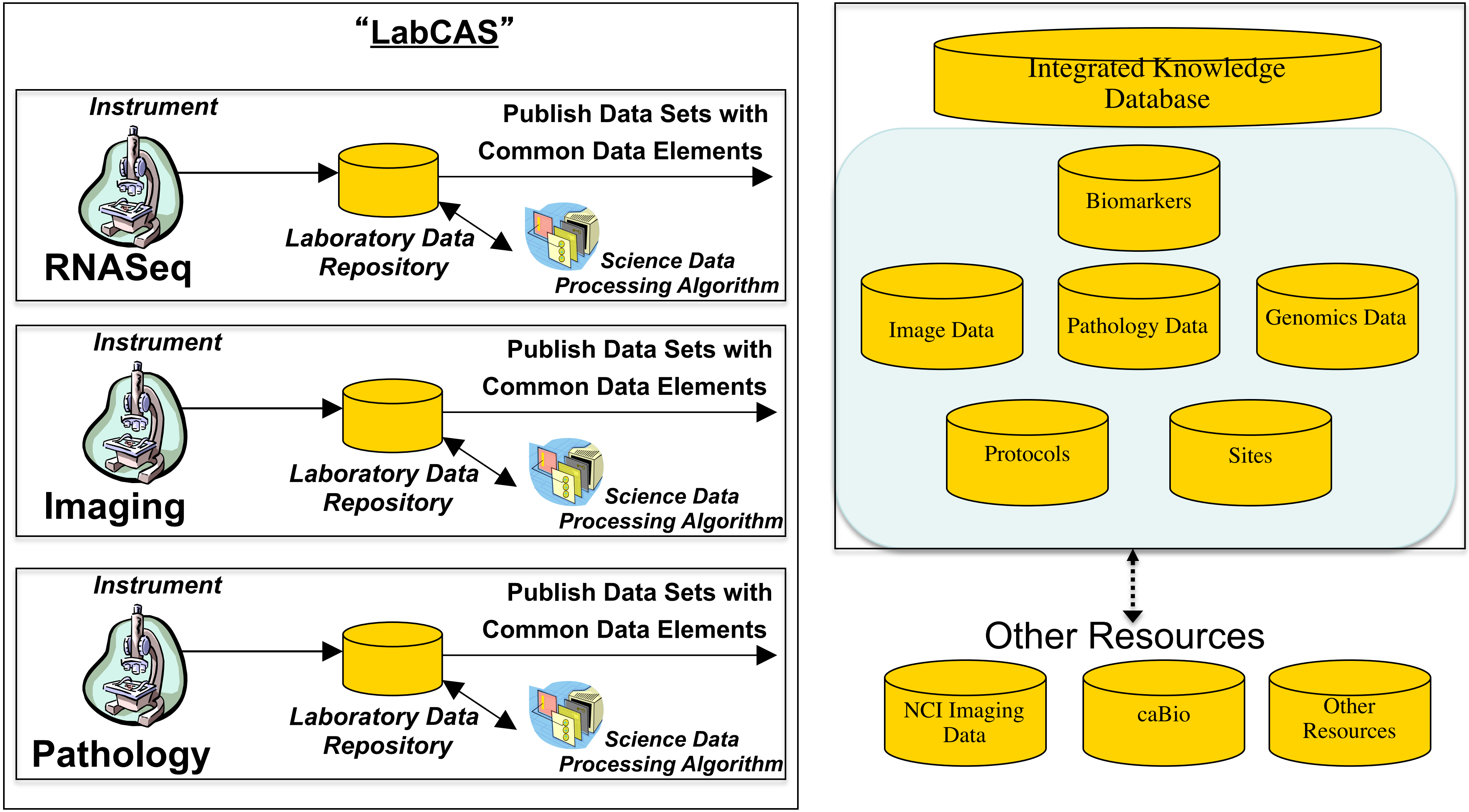
caMicroscope

case_id: d4e37fb1-e672-4136-8258-2fd855bae05b.svs

Data Capture, Processing and Ingestion



Data Pipelines and the Knowledge Environment



"LabCAS"

Instrument

RNASeq

Laboratory Data Repository

Publish Data Sets with Common Data Elements

Science Data Processing Algorithm

Instrument

Imaging

Laboratory Data Repository

Publish Data Sets with Common Data Elements

Science Data Processing Algorithm

Instrument

Pathology

Laboratory Data Repository

Publish Data Sets with Common Data Elements

Science Data Processing Algorithm

Integrated Knowledge Database

Biomarkers

Image Data

Pathology Data

Genomics Data

Protocols

Sites

Other Resources

NCI Imaging Data

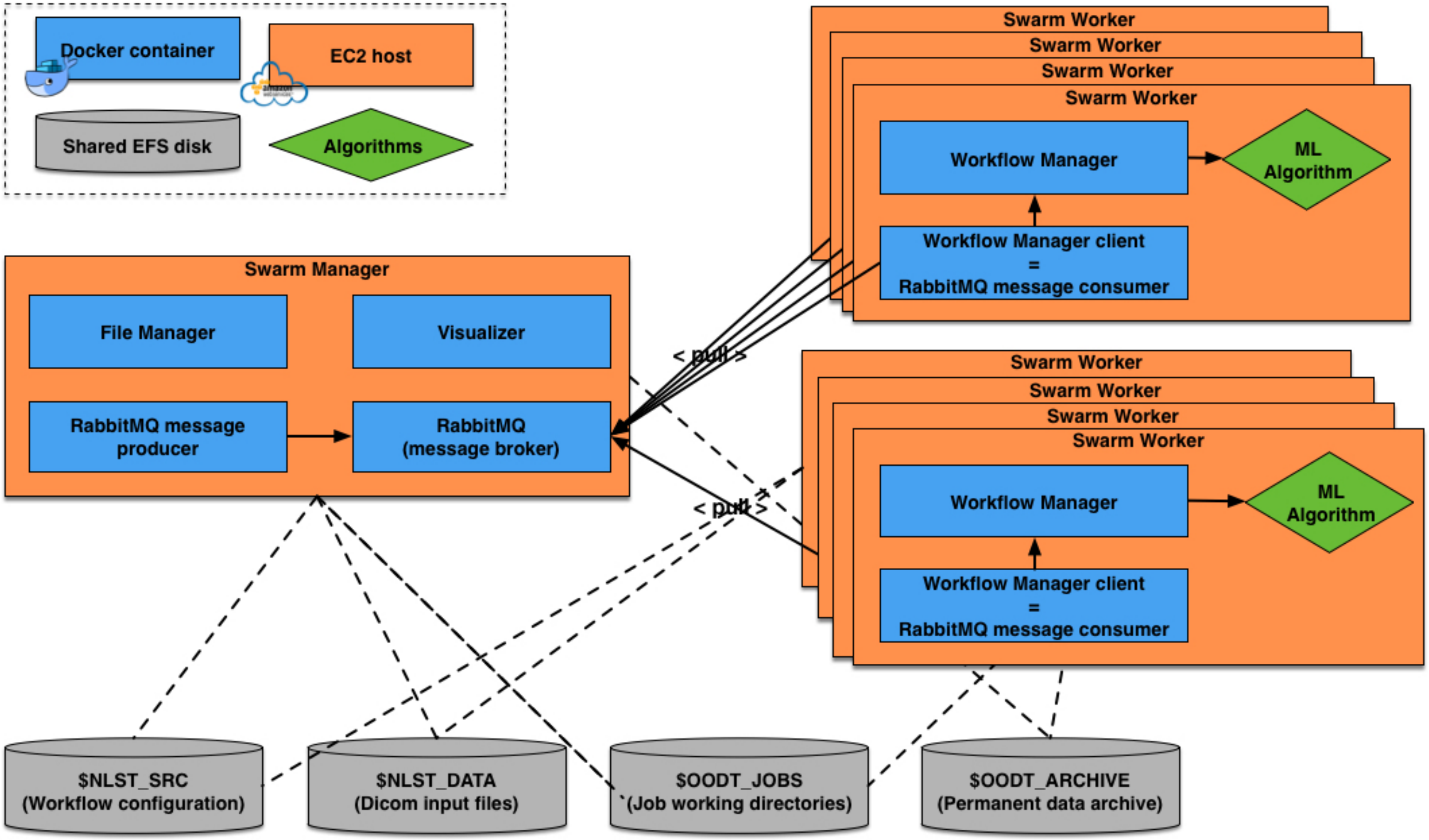
caBio

Other Resources

Scalable computing, common data elements, computational methods

Scalable architecture for ML on AWS

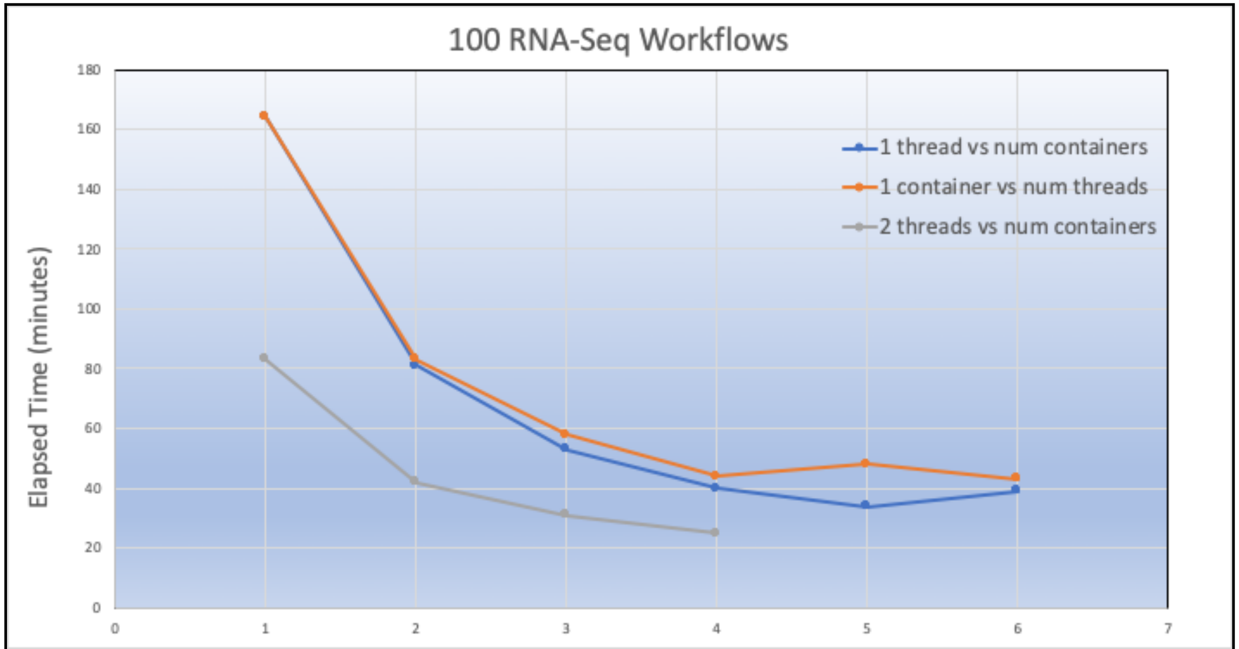
- Uses Docker Swarm, Apache OODT workflows (from NASA/JPL), RabbitMQ messaging
- Can scale/auto-scale to any number of EC2 nodes



Docker container
EC2 host
Shared EFS disk
Algorithms
Swarm Worker
Swarm Worker
Swarm Worker
Swarm Worker
Workflow Manager
ML Algorithm
Workflow Manager client = RabbitMQ message consumer
Swarm Manager
File Manager
Visualizer
RabbitMQ message producer
RabbitMQ (message broker)
< pull >
Swarm Worker
Swarm Worker
Swarm Worker
Swarm Worker
Workflow Manager
ML Algorithm
Workflow Manager client = RabbitMQ message consumer
< pull >
$NLST_SRC (Workflow configuration)
$NLST_DATA (Dicom input files)
$OODT_JOBS (Job working directories)
$OODT_ARCHIVE (Permanent data archive)

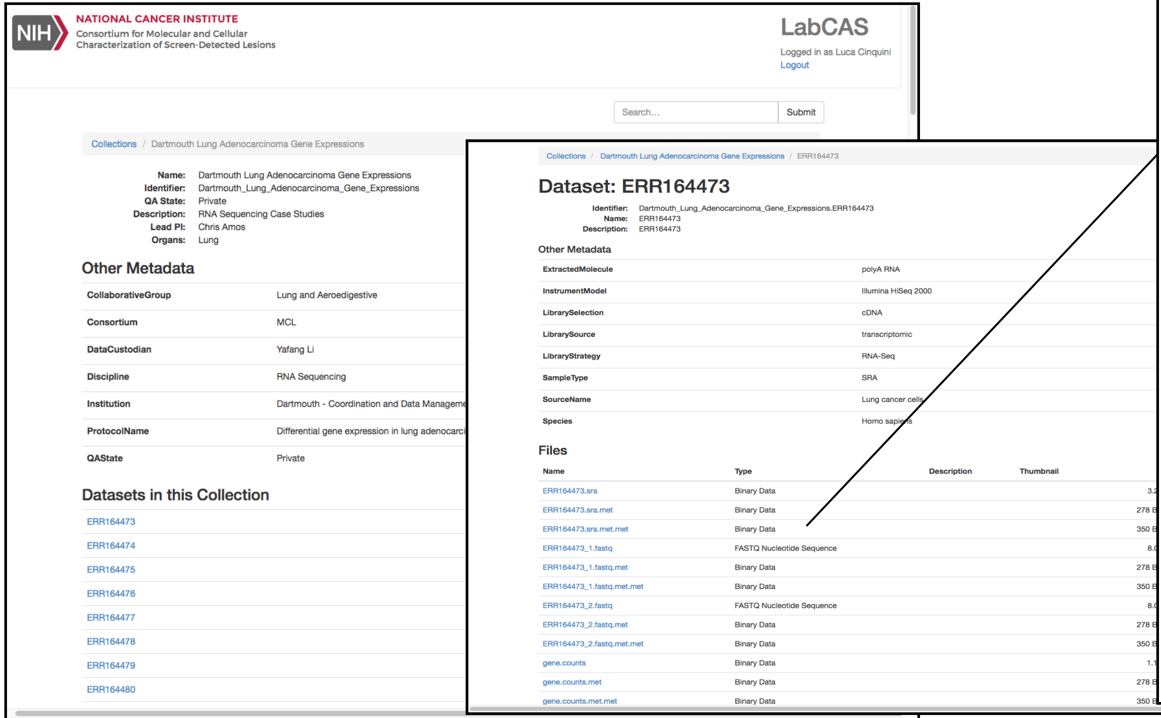
Benchmarking on AWS

- Executing genomics workflow on single EC2 server (r5.2xlarge, 8 CPUs, 64 GB memory)
- Measuring time to execute 100 workflows vs # of containers, # of processes/container
- Can scale horizontally until processes start to compete for resources



Lung Adenocarcinoma Gene Expressions

- In collaboration with Dr. Chris Amos, Yafang Lee (Dartmouth)
- 10 TB of data for full study comparing gene expression profiles between smoking and non-smoking patients.
- Integrated Dartmouth analysis tools with 99% accuracy into LabCAS pipelines



NATIONAL CANCER INSTITUTE
Consortium for Molecular and Cellular
Characterization of Screen-Detected Lesions

LabCAS
Logged in as Luca Cinquini
Logout

Search... Submit

Collections / Dartmouth Lung Adenocarcinoma Gene Expressions

Name: Dartmouth Lung Adenocarcinoma Gene Expressions
Identifier: Dartmouth_Lung_Adenocarcinoma_Gene_Expressions
QA State: Private
Description: RNA Sequencing Case Studies
Lead PI: Chris Amos
Organs: Lung

Other Metadata

CollaborativeGroup	Lung and Aeroedigestive
Consortium	MCL
DataCustodian	Yafang Li
Discipline	RNA Sequencing
Institution	Dartmouth - Coordination and Data Manageme
ProtocolName	Differential gene expression in lung adenocarci
QAState	Private

Datasets in this Collection

- ERR164473
- ERR164474
- ERR164475
- ERR164476
- ERR164477
- ERR164478
- ERR164479
- ERR164480

Collections / Dartmouth Lung Adenocarcinoma Gene Expressions / ERR164473

Dataset: ERR164473

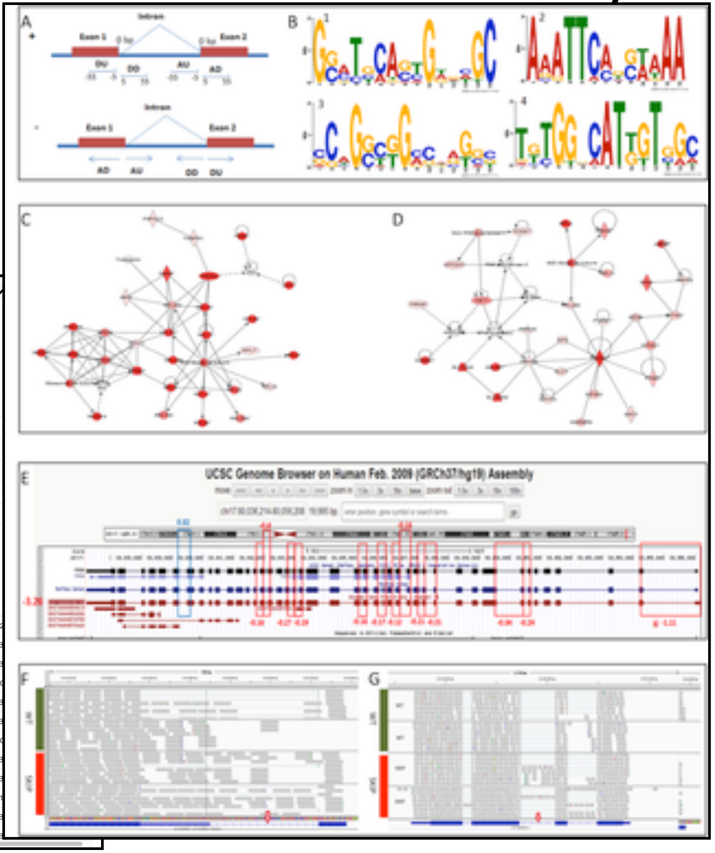
Identifier: Dartmouth_Lung_Adenocarcinoma_Gene_Expressions.ERR164473
Name: ERR164473
Description: ERR164473

Other Metadata

ExtractedMolecule	polyA RNA
InstrumentModel	Illumina HiSeq 2000
LibrarySelection	cDNA
LibrarySource	transcriptomic
LibraryStrategy	RNA-Seq
SampleType	SRA
SourceName	Lung cancer cells
Species	Homo sapiens

Files

Name	Type	Description	Thumbnail	
ERR164473.sra	Binary Data			3.2
ERR164473.sra.met	Binary Data			278 B
ERR164473.sra.met.met	Binary Data			350 B
ERR164473_1.fastq	FASTQ Nucleotide Sequence			8.0
ERR164473_1.fastq.met	Binary Data			278 B
ERR164473_1.fastq.met.met	Binary Data			350 B
ERR164473_2.fastq	FASTQ Nucleotide Sequence			8.0
ERR164473_2.fastq.met	Binary Data			278 B
ERR164473_2.fastq.met.met	Binary Data			350 B
gene.counts	Binary Data			1.1
gene.counts.met	Binary Data			278 B
gene.counts.met.met	Binary Data			350 B



Smart-3SEQ

- In collaboration with Robert West, Joseph Foley, Sujen Vennam @ Stanford School of Medicine

- Smart-3SEQ: new method for quantifying gene expressions in small RNA samples including single cells (see: https://github.com/jwfoley/3SEQtools)

- Smart-3SEQ pipeline is composed of 3 steps:
  - FastQ generation and alteration (w/ Illumina bcl2fastq)
  - Gene alignment (w/ Samtools, STAR, UMI-dedup)
  - Read counting (w/ Bioconductors

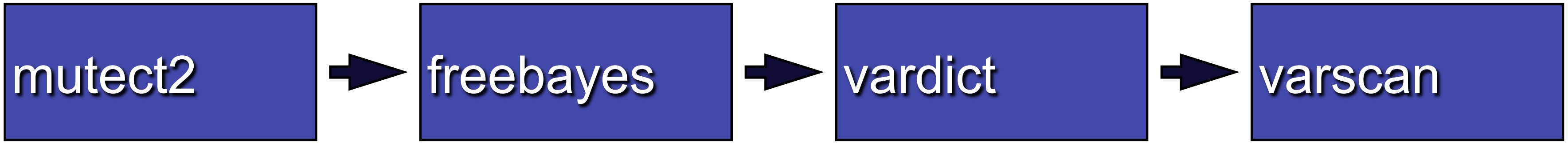
FastQ generation → Gene alignment → Read counting

DNA-Sequencing

- In collaboration with Olivier Harismendy at UCSD
  - See https://github.com/bcbio, http://bcbio.io

- Using bcbio (“Bue Collar Bioinformatics”): Python framework and community tools for analysis of biological data (variant calling, RNA-SEQ and small RNA pipelines)

- Experimented with running sample pipelines (big and small)



mutect2 → freebayes → vardict → varscan

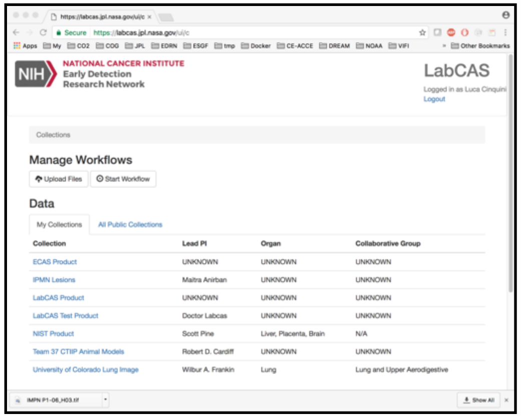
Multi-dimensional Integrated Data

Collaboration with Sandy Borowsky/UC Davis and Anirban Maitra/ MD Anderson

- Developed a pathology archive for MCL
- Dr. Maitra goal: multi-dimensional data – IPMN ppts

Collaboration with Radka Stoyanova and Alan Pollack

- MAST (Mapped Active Surveillance Trial). Longitudinal multivariate data (mpMRI, pathology and gene expression) is obtained from patients on Active Surveillance for prostate cancer.



LabCAS

Logged in as Luca Cinquini
Logout

Collections

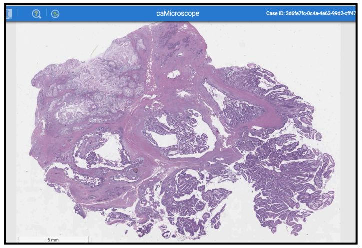
Manage Workflows

Upload Files | Start Workflow

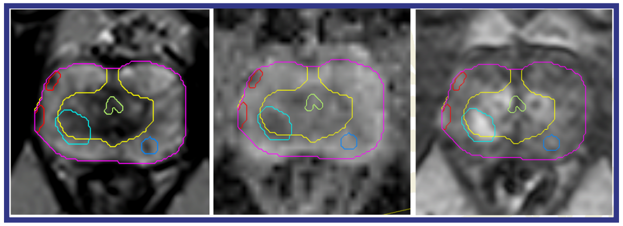
Data

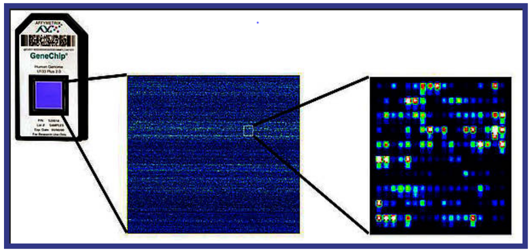
My Collections | All Public Collections

Collection	Lead PI	Organ	Collaborative Group
ECAS Product	UNKNOWN	UNKNOWN	UNKNOWN
IPMN Lesions	Maitra Anirban	UNKNOWN	UNKNOWN
LabCAS Product	UNKNOWN	UNKNOWN	UNKNOWN
LabCAS Test Product	Doctor Labcas	UNKNOWN	UNKNOWN
NIST Product	Scott Pine	Liver, Placenta, Brain	N/A
Team 37 CTIIP Animal Models	Robert D. Cardiff	UNKNOWN	UNKNOWN
University of Colorado Lung Image	Wilbur A. Frankin	Lung	Lung and Upper Aerodigestive

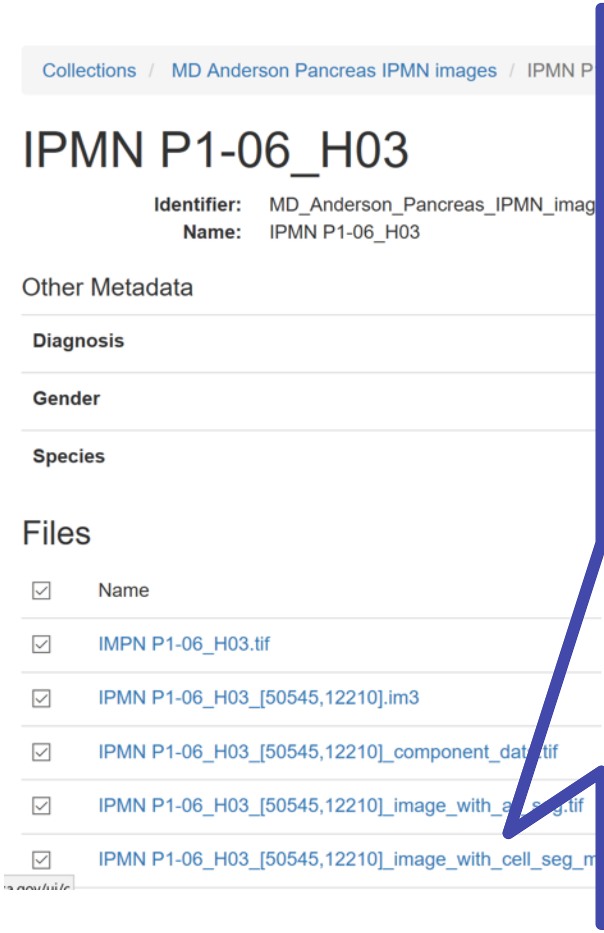


Path Viz Tool Credit:
caMicroscope - Dr. Joel Saltz Lab
ITCR Program





Support for Image Archiving



Collections / MD Anderson Pancreas IPMN images / IPMN P

IPMN P1-06_H03

Identifier: MD_Anderson_Pancreas_IPMN_imag

Name: IPMN P1-06_H03

Other Metadata

Diagnosis

Gender

Species

Files

☑ Name

☑ IMPN P1-06_H03.tif

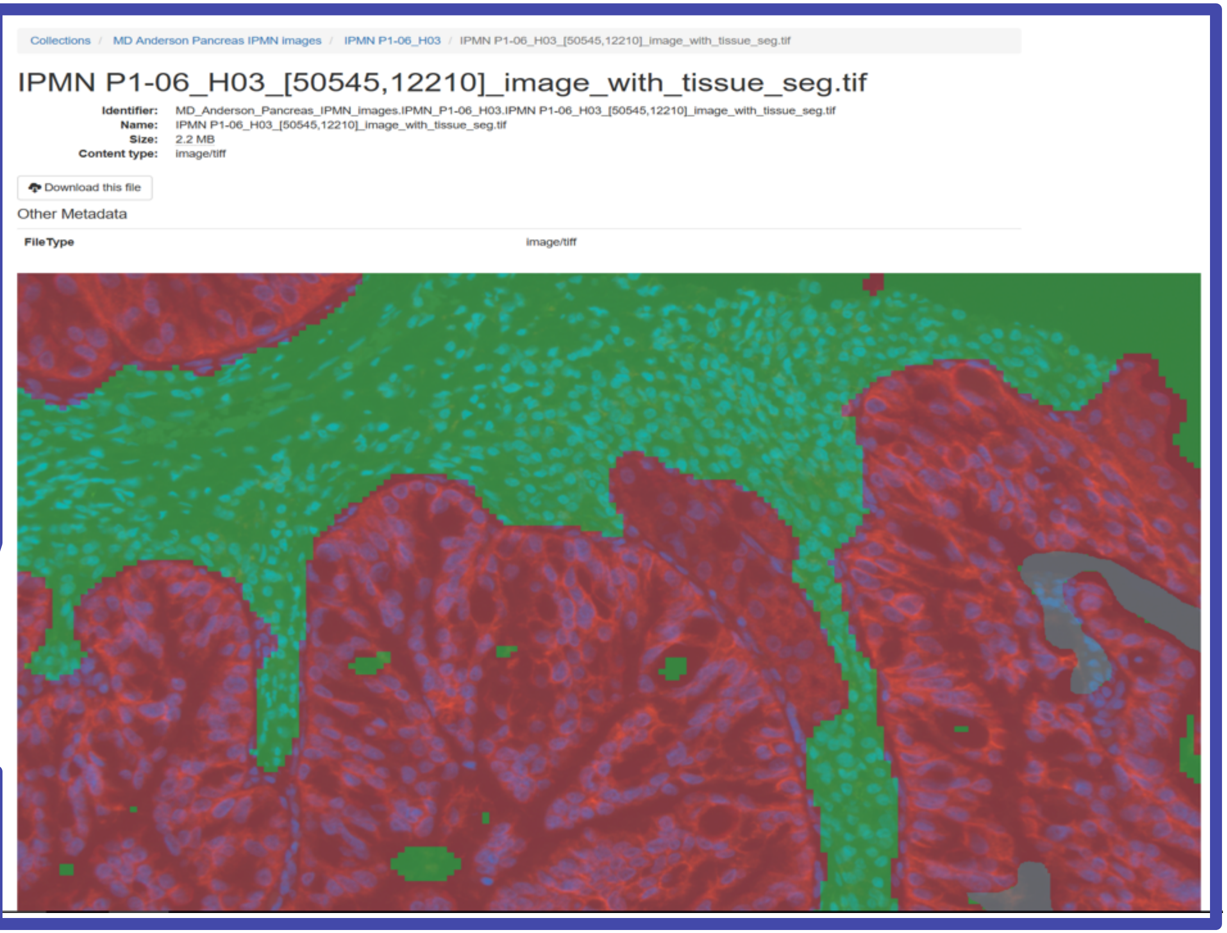
☑ IPMN P1-06_H03_[50545,12210].im3

☑ IPMN P1-06_H03_[50545,12210]_component_dat tif

☑ IPMN P1-06_H03_[50545,12210]_image_with_a seg.tif

☑ IPMN P1-06_H03_[50545,12210]_image_with_cell_seg_m

a.gov/ui/c



Collections / MD Anderson Pancreas IPMN images / IPMN P1-06_H03 / IPMN P1-06_H03_[50545,12210]_image_with_tissue_seg.tif

IPMN P1-06_H03_[50545,12210]_image_with_tissue_seg.tif

Identifier: MD_Anderson_Pancreas_IPMN_images.IPMN_P1-06_H03.IPMN P1-06_H03_[50545,12210]_image_with_tissue_seg.tif

Name: IPMN P1-06_H03_[50545,12210]_image_with_tissue_seg.tif

Size: 2.2 MB

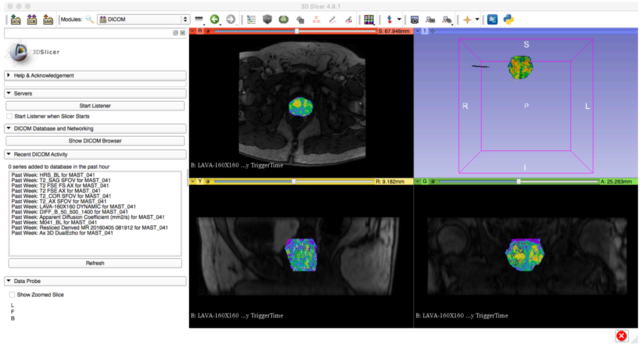
Content type: image/tiff

Other Metadata

FileType	image/tiff

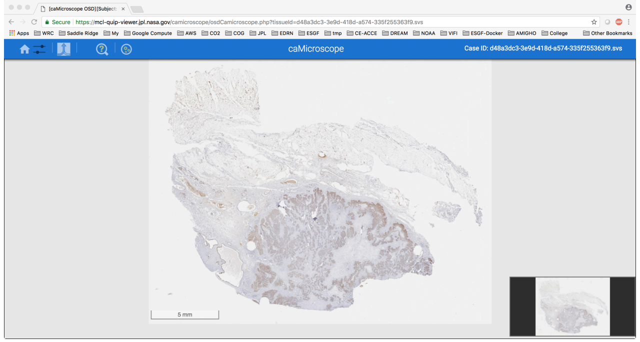
Integrate different analytical tools

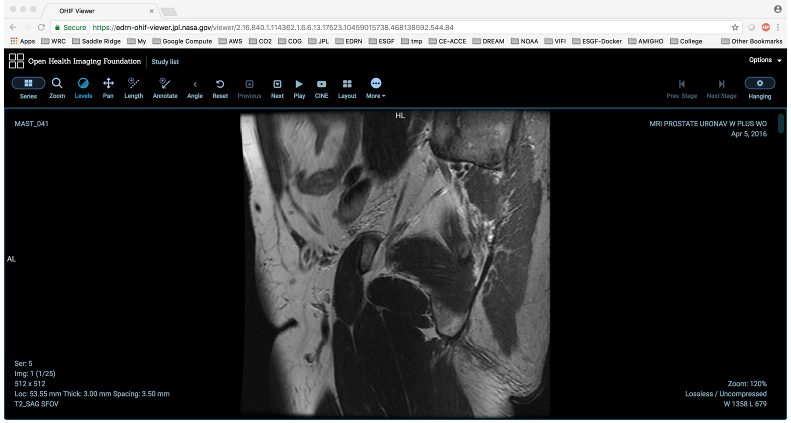
- Framework to plug-in many open-source viewers (client and server side)
  - caMicroscope
  - OHIF DICOM Viewer
  - 3D Slicer
- Capture additional metadata
- User specified organization of data
- Explore VR and other approaches for multii-dimensional imaging


Modules:
DICOM
3DSlicer
Help & Acknowledgement
Servers
Start Listener
Start Listener when Slicer Starts
DICOM Database and Networking
Show DICOM Browser
Recent DICOM Activity
Refresh
Data Probe
Show Zoomed Slice
L
F
B
S
R
P
L
I
B: LAVA-160X160 ...y TriggerTime
B: LAVA-160X160 ...y TriggerTime
B: LAVA-160X160 ...y TriggerTime

0 series added to database in the past hour

Past Week: HRS_BL for MAST_041
Past Week: T2_SAG SFOV for MAST_041
Past Week: T2 FSE FS AX for MAST_041
Past Week: T2 FSE AX for MAST_041
Past Week: T2_COR SFOV for MAST_041
Past Week: T2_AX SFOV for MAST_041
Past Week: LAVA-160X160 DYNAMIC for MAST_041
Past Week: DIFF_B_50_500_1400 for MAST_041
Past Week: Apparent Diffusion Coefficient (mm2/s) for MAST_041
Past Week: M041_BL for MAST_041
Past Week: Resliced Derived MR 20160405 081912 for MAST_041
Past Week: Ax 3D DualEcho for MAST_041


[caMicroscope OSD][Subject:
Secure
https://mcl-quip-viewer.jpl.nasa.gov/camicroscope/osdCamicroscope.php?tissueId=d48a3dc3-3e9d-418d-a574-335f255363f9.svs
Apps
WRC
Saddle Ridge
My
Google Compute
AWS
CO2
COG
JPL
EDRN
ESGF
tmp
CE-ACCE
DREAM
NOAA
VIFI
ESGF-Docker
AMIGHO
College
Other Bookmarks
caMicroscope
Case ID: d48a3dc3-3e9d-418d-a574-335f255363f9.svs
5 mm



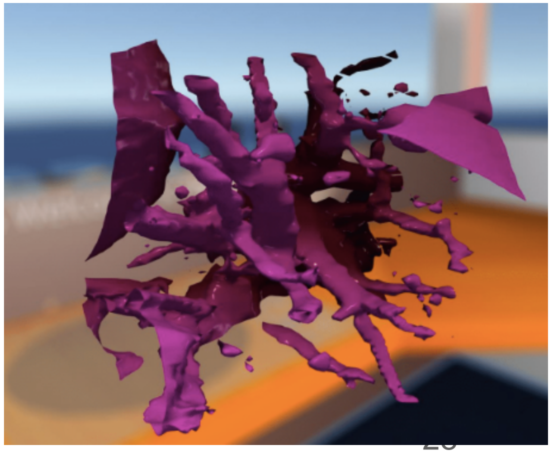
MAST_041

MRI PROSTATE URONAV W PLUS WO
Apr 5, 2016

Ser: 5
Img: 1 (1/25)
512 x 512
Loc: 53.55 mm Thick: 3.00 mm Spacing: 3.50 mm
T2_SAG SFOV

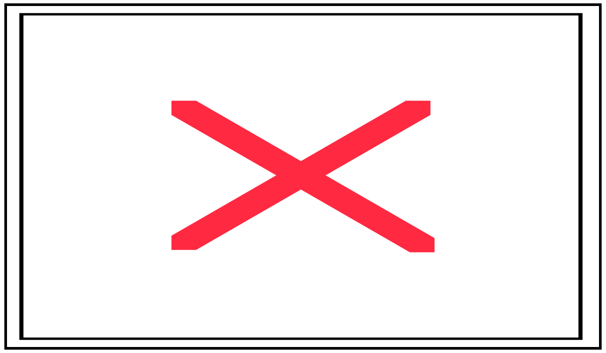
Zoom: 120%
Lossless / Uncompressed
W 1358 L 679

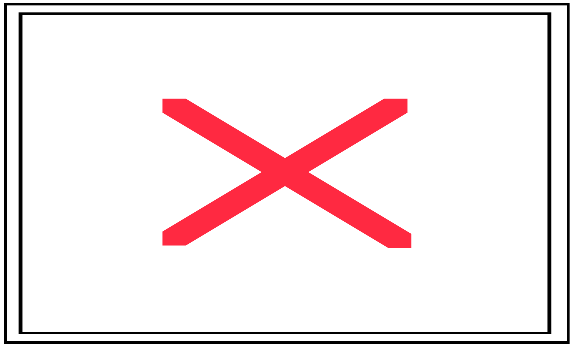
3D rendered isosurfaces: section of lung tissue



Amazon Workspaces

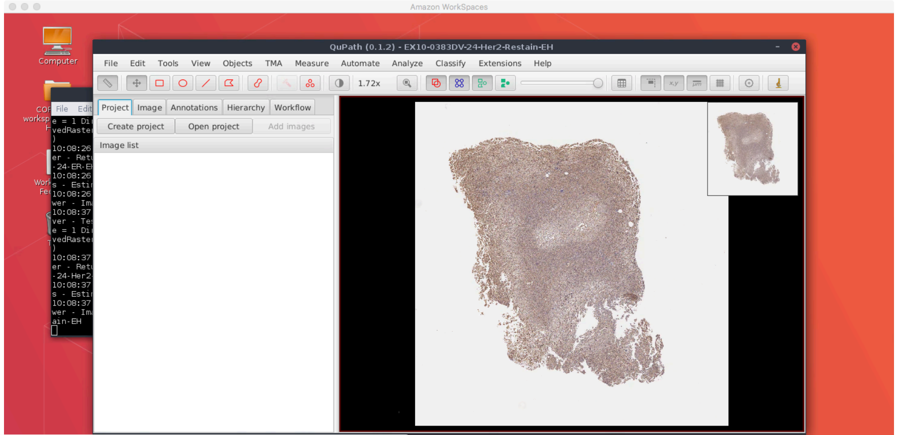
- – Remote desktops running on the AWS Cloud
- – Support analysis tools that must be installed on client machines, to access local data:
  - QuPath for use by Sandy Borowsky’s group (UC Davis)
  - 3D Slicer for use by Radka Stoyanava’s group (Univ. Of Miami)

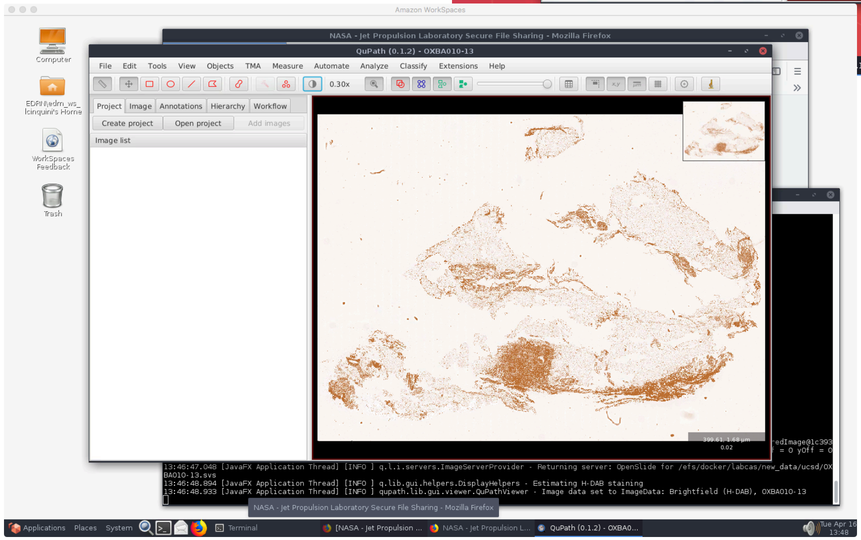




AWS Workspaces: QuPath

UC Davis Pathology data


Amazon WorkSpaces
QuPath (0.1.2) - EX10-0383DV-24-Her2-Restain-EH
File Edit Tools View Objects TMA Measure Automate Analyze Classify Extensions Help
1.72x
Project Image Annotations Hierarchy Workflow
Create project Open project Add images
Image list
Computer

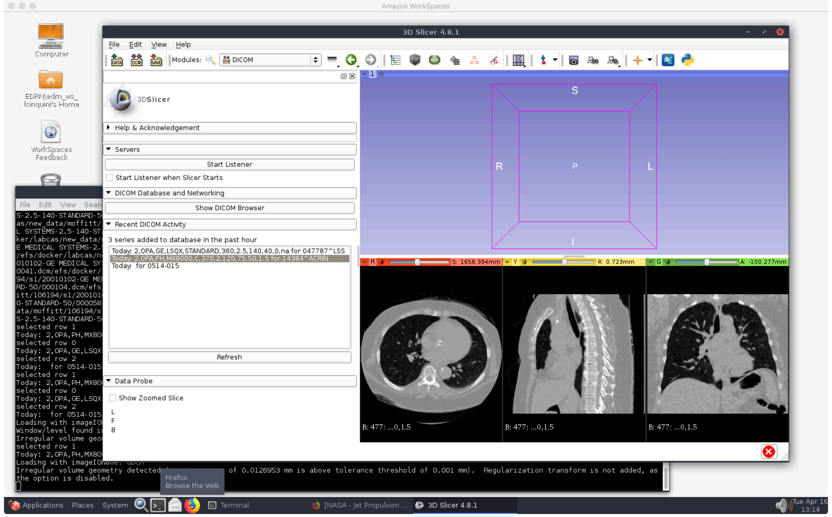

Amazon WorkSpaces
NASA - Jet Propulsion Laboratory Secure File Sharing - Mozilla Firefox
QuPath (0.1.2) - OXBA010-13
File Edit Tools View Objects TMA Measure Automate Analyze Classify Extensions Help
0.30x
Project Image Annotations Hierarchy Workflow
Create project Open project Add images
Image list
Computer
EDRN\edrn_ws_lcinquini's Home
WorkSpaces Feedback
Trash
Applications Places System
Terminal
[NASA - Jet Propulsion ...
NASA - Jet Propulsion L...
QuPath (0.1.2) - OXBA0...
Tue Apr 16 13:48

```
13:46:47.048 [JavaFX Application Thread] [INFO ] q.l.i.servers.ImageServerProvider - Returning server: OpenSlide for /efs/docker/labcas/new_data/ucsd/OX
BA010-13.svs
13:46:48.894 [JavaFX Application Thread] [INFO ] q.lib.gui.helpers.DisplayHelpers - Estimating H-DAB staining
13:46:48.933 [JavaFX Application Thread] [INFO ] qupath.lib.gui.viewer.QuPathViewer - Image data set to ImageData: Brightfield (H-DAB), OXBA010-13
```

UCSD DNA-Sequencing data

AWS Workspaces: 3dSlicer

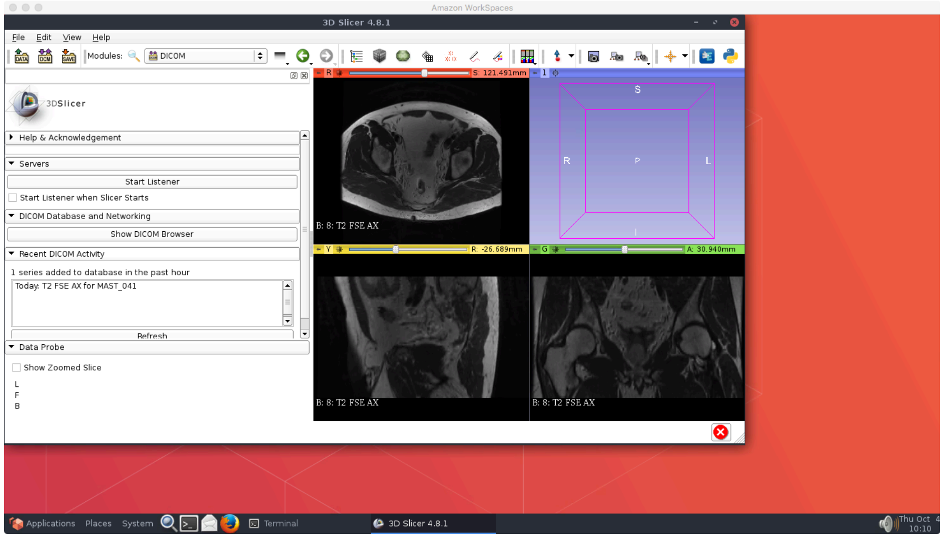
Moffitt lung data



3 series added to database in the past hour

Today: 2,OPA,GE,LSQX,STANDARD,360,2.5,140,40,0,na for 047787^LSS
Today: 2,OPA,PH,MX8000,C,370,2,120,75,50,1.5 for 14384^ACRIN
Today: for 0514-015

```
File  Edit  View  Sear
S-2.5-140-STANDARD-5
as/new_data/moffitt/
L SYSTEMS-2.5-140-ST
ker/labcas/new_data/
E MEDICAL SYSTEMS-2.
/efs/docker/labcas/n
010102-GE MEDICAL SY
0041.dcm/efs/docker/
94/s1/20010102-GE ME
RD-50/000104.dcm/efs
itt/106194/s1/200101
0-STANDARD-50/000058
ata/moffitt/106194/s
S-2.5-140-STANDARD-5
selected row 1
Today: 2,OPA,PH,MX80
selected row 0
Today: 2,OPA,GE,LSQX
selected row 2
Today:  for 0514-015
selected row 1
Today: 2,OPA,PH,MX80
selected row 0
Today: 2,OPA,GE,LSQX
selected row 2
Today:  for 0514-015
Loading with imageIO
Window/level found i
Irregular volume geo
selected row 1
Today: 2,OPA,PH,MX80
Loading with imageIOName: GDCM
Irregular volume geometry detected                 of 0.0126953 mm is above tolerance threshold of 0.001 mm).  Regularization transform is not added, as
the option is disabled.
```



1 series added to database in the past hour

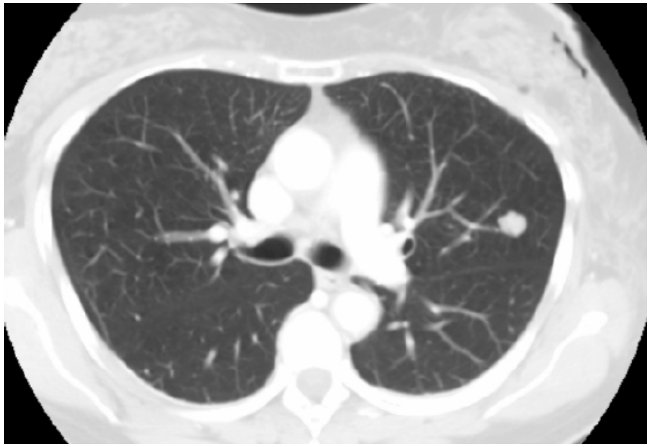
Today: T2 FSE AX for MAST_041

Univ. of Miami prostate data

Feature classification in images

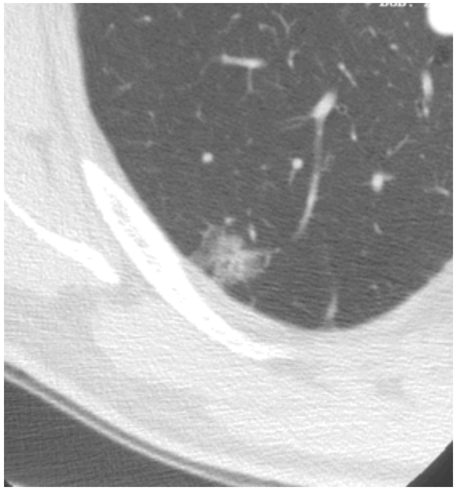
Pulmonary Nodule

Solid Nodule

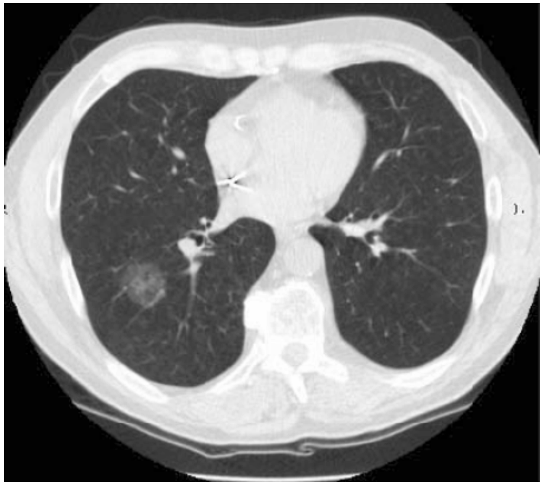


Sub-solid Nodule

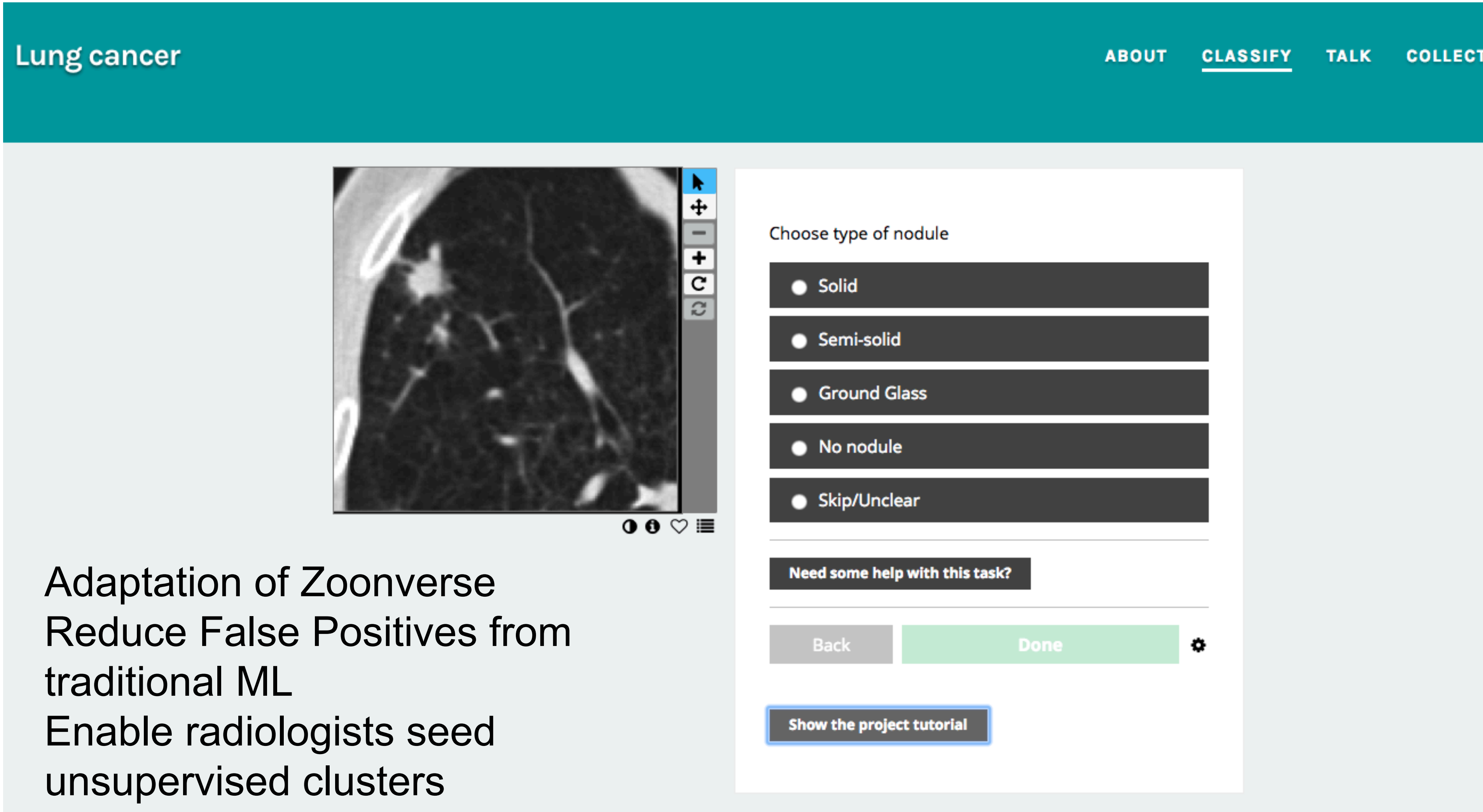
Part-solid



Ground Glass



Crowd Sourcing Image Analysis



Choose type of nodule

- ○ Solid
- ○ Semi-solid
- ○ Ground Glass
- ○ No nodule
- ○ Skip/Unclear

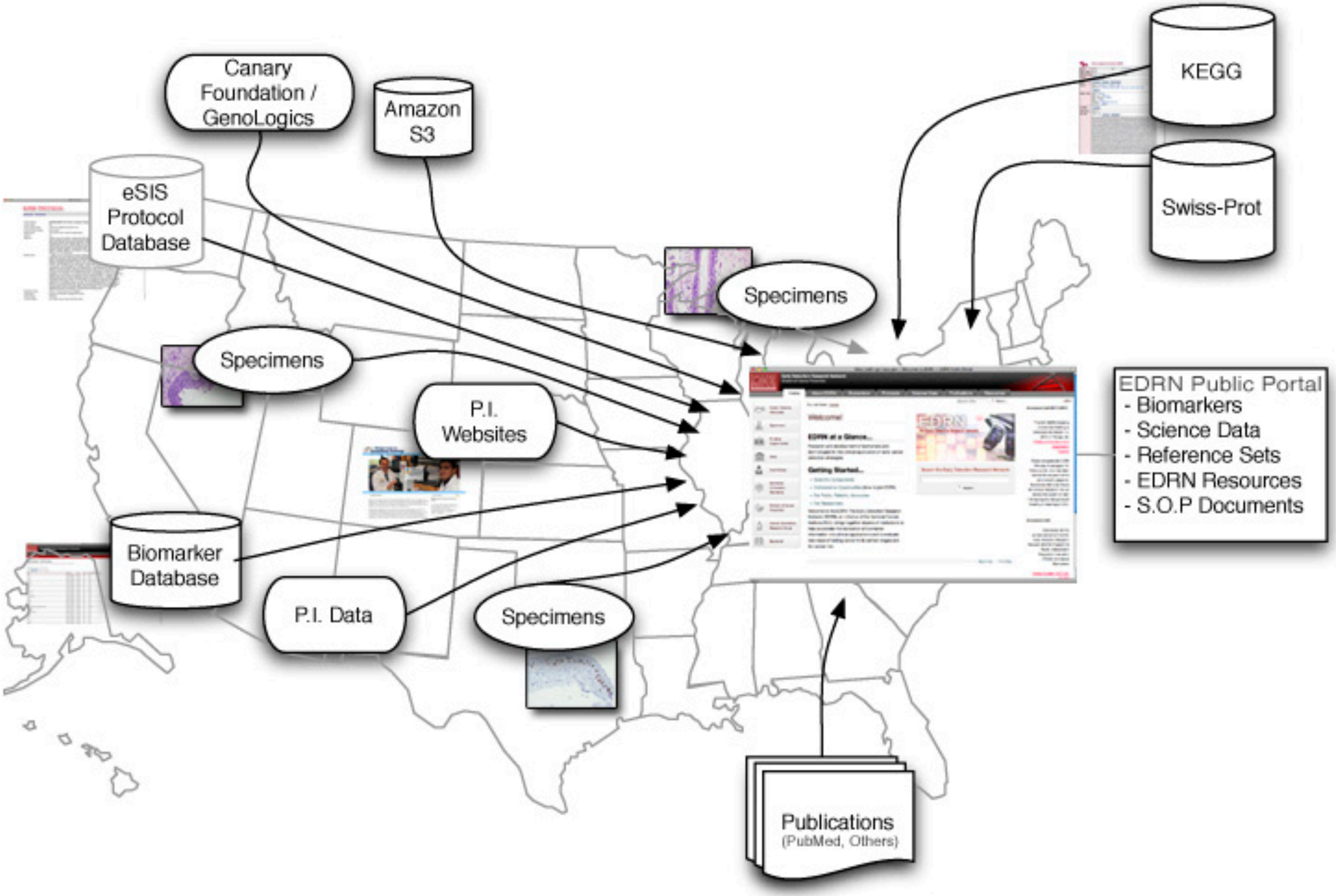
Adaptation of Zoonverse

Reduce False Positives from traditional ML

Enable radiologists seed unsupervised clusters

Drive towards ML pipelines

What’s Emerged: A Virtual, National Integration Biomarkers Knowledge System



Canary Foundation / GenoLogics

Amazon S3

eSIS Protocol Database

KEGG

Swiss-Prot

Specimens

Specimens

P.I. Websites

Biomarker Database

P.I. Data

Specimens

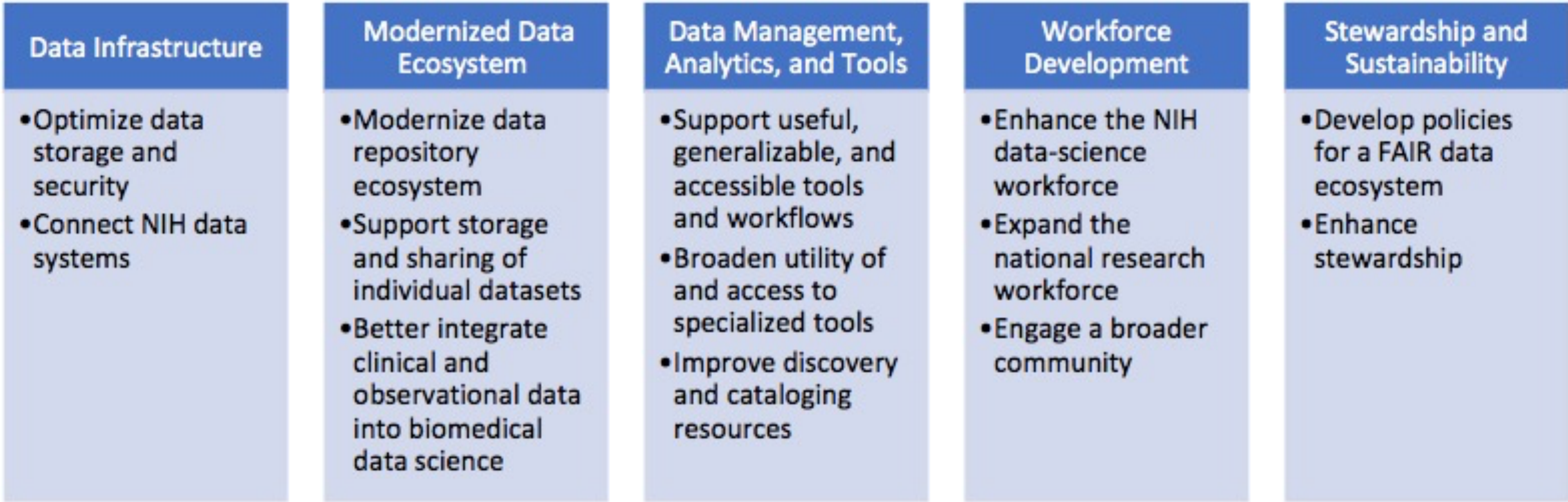
EDRN Public Portal
- Biomarkers
- Science Data
- Reference Sets
- EDRN Resources
- S.O.P Documents

Publications
(PubMed, Others)

What’s in Place Today

- A national, biomarker knowledge system serving multiple programs
- A biomarker data infrastructure consisting of ~1000 biomarkers, ~200 protocols, 1500 publications, 100 TB data
- Tools for laboratories to support the processing, capture, curation and sharing of data before publications
- Pilot projects in imaging, scalable workflows, data integration, etc
- Support for data-driven approaches for data discovery and analysis
- Common portals to access the knowledge environment

NIH Data Science Strategic Plan*



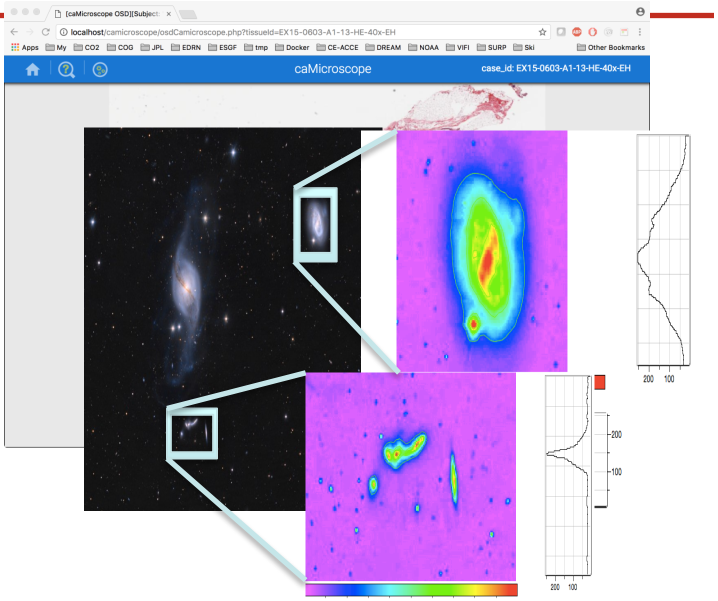
Data Infrastructure	Modernized Data Ecosystem	Data Management, Analytics, and Tools	Workforce Development	Stewardship and Sustainability
•Optimize data storage and security •Connect NIH data systems	•Modernize data repository ecosystem •Support storage and sharing of individual datasets •Better integrate clinical and observational data into biomedical data science	•Support useful, generalizable, and accessible tools and workflows •Broaden utility of and access to specialized tools •Improve discovery and cataloging resources	•Enhance the NIH data-science workforce •Expand the national research workforce •Engage a broader community	•Develop policies for a FAIR data ecosystem •Enhance stewardship

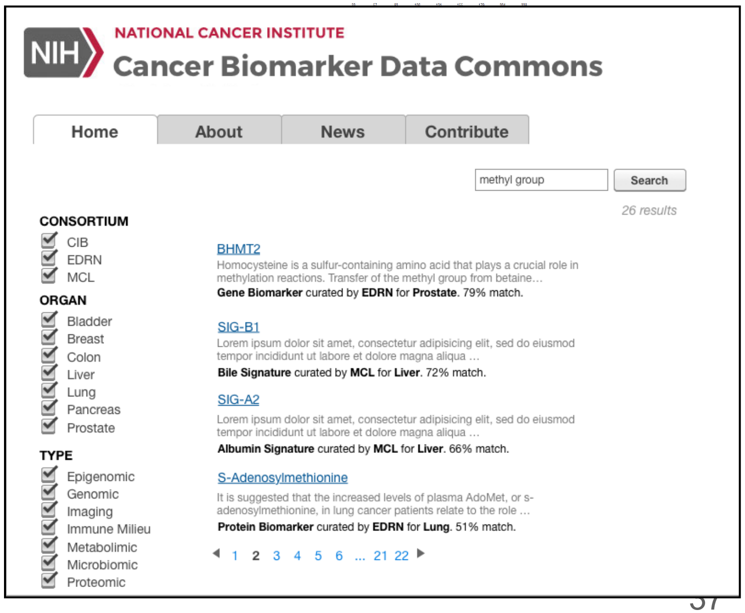
Figure 2. NIH Strategic Plan for Data Science: Overview of Goals and Objectives

* By 2025, the total amount of genomics data alone is expected to equal or exceed totals from the three other major producers of large amounts of data: astronomy, YouTube, and Twitter.

Future Directions

- Systematize the capture of data end-to-end into a knowledge environment
- Integrate data-driven techniques and tools such as machine learning as part of the end-to-end knowledge environment
- Enable collaborative analysis that scales
- Unify consortium enterprises and data
- Enable science through an explicit and well architected data science strategy and platform


[caMicroscope OSD][Subject
localhost/camicroscope/osdCamicroscope.php?tissueId=EX15-0603-A1-13-HE-40x-EH
Apps
My
CO2
COG
JPL
EDRN
ESGF
tmp
Docker
CE-ACCE
DREAM
NOAA
VIFI
SURP
Ski
Other Bookmarks
caMicroscope
case_id: EX15-0603-A1-13-HE-40x-EH
200
100
200
100
200
100


NATIONAL CANCER INSTITUTE

Cancer Biomarker Data Commons

26 results

CONSORTIUM
☑ CIB
☑ EDRN
☑ MCL

ORGAN
☑ Bladder
☑ Breast
☑ Colon
☑ Liver
☑ Lung
☑ Pancreas
☑ Prostate

TYPE
☑ Epigenomic
☑ Genomic
☑ Imaging
☑ Immune Milieu
☑ Metabolimic
☑ Microbiomic
☑ Proteomic

BHMT2
Homocysteine is a sulfur-containing amino acid that plays a crucial role in methylation reactions. Transfer of the methyl group from betaine…
Gene Biomarker curated by **EDRN** for **Prostate**. 79% match.

SIG-B1
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua …
Bile Signature curated by **MCL** for **Liver**. 72% match.

SIG-A2
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua …
Albumin Signature curated by **MCL** for **Liver**. 66% match.

S-Adenosylmethionine
It is suggested that the increased levels of plasma AdoMet, or s-adenosylmethionine, in lung cancer patients relate to the role …
Protein Biomarker curated by **EDRN** for **Lung**. 51% match.

◀ 1 2 3 4 5 6 … 21 22 ▶

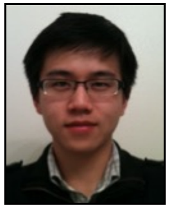
JPL Informatics Center
Data Science Team

Dan Crichton
NASA/JPL
Principal Investigator

Data Architecture ⟹

- Luca Cinquini NASA/JPL
- Asitang Mishra NASA/JPL

Portal
System Engineering ⟹



Sean Kelly David Liu
NASA/JPL

Project Management
Data Coordination ⟹

Heather Kincaid
NASA/JPL

Machine Learning
Visualization ⟹

- Alphan Altinok
NASA JPL
- Santiago Lombeyda
Caltech

Ashish Mahabal
Caltech

Bioinformatics ⟹
Biomarker
Curation/CDEs

Kristen Anton Maureen Colbert
University of North Carolina

System Admin ⟹
Cloud Computing

Paul Zimdars
NASA/JPL

- Susan Neely NAS
JPL
- Rojeh Yaghoobi
NASA/JPL

Backup