Navigating CCDI Hub's Explore Dashboard and Data Access

Childhood Cancer Data Initiative Webinar Series

NIH NATIONAL CANCER INSTITUTE

November 13, 2023

1. *Introductions*
2. *Overview and Demonstration of the CCDI Hub's Explore Dashboard*
3. *Controlled Data Access Process*
4. *Navigating the Cancer Genomics Platform*
5. *Q&A*

Introductions

Gregory Reaman

Peter Gilbertson

Technical Project Manager

Sean Burke

Bioinformatics Manager

Zélia Worman

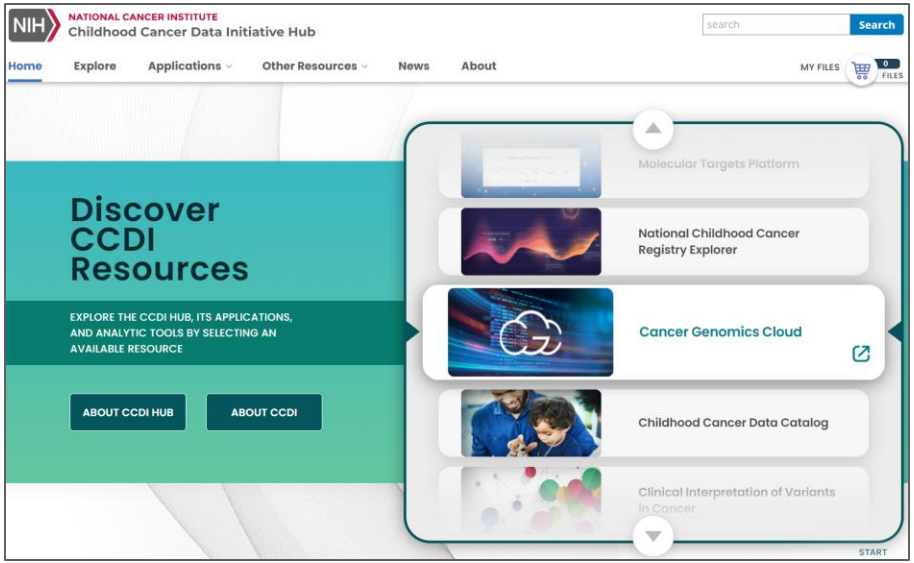
Director of Researcher Engagement and Education

Overview of the CCDI Hub's Explore Dashboard

Peter Gilbertson

CCDI Hub

- The CCDI Hub is an entry point for researchers, data scientists, and citizen scientists looking to use and connect with CCDI-related data.
- It provides information and direct links to CCDI platforms, tools, and resources, along with additional technical information.
- The Explore Dashboard (discussed in upcoming slides) brings together CCDI-supported data and allows exploration of data in new ways.



ccdi.cancer.gov

Evolution of the CCDI Hub

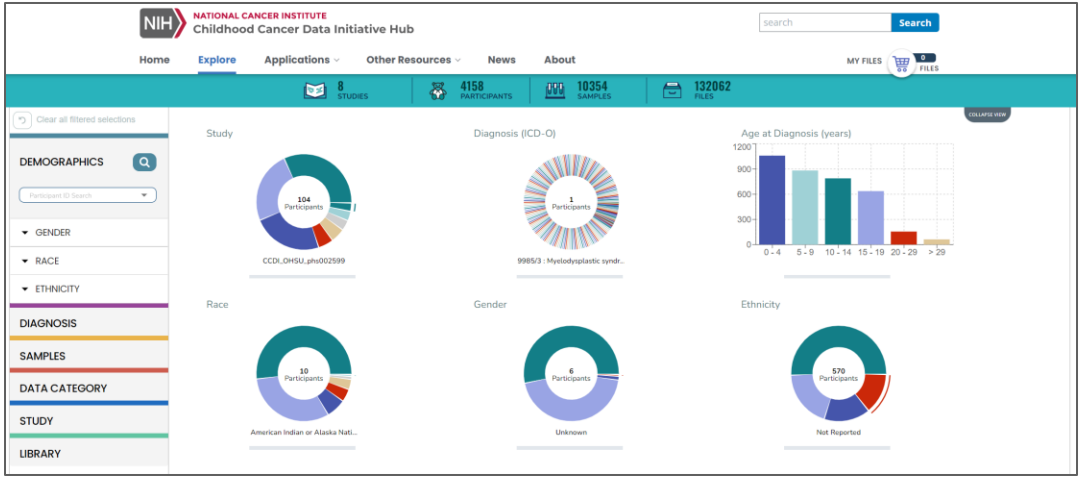
Explore Dashboard: Brings together CCDI-managed data & allows exploration of data in new ways.

What it is:

- An inventory of CCDI-managed childhood cancer data

What it does:

- Makes it easier for researchers to find data
- Faceted search
- Visualize results
- Exportable results



ccdi.cancer.gov/explore

CCDI-Managed Data

- Genomic Characterization: Juvenile Myelo Monocytic Leukemia (PHS002504)
- Molecular Characterization: Pediatric Brain Tumors & Other Cancers (PHS002517)
- OncoKids Cancer Panel: Pediatric Cancers (PHS002518)
- Comprehensive Genomic Sequencing: Pediatric Cancers (PHS002529)
- Genomic Landscape: Acute Myeloid Leukemia (PHS002599)
- Whole Genome & Transcriptome Profiling: Pediatric and Young Adult Cancers (PHS002620)
- CCDI's Molecular Characterization Initiative (PHS002790)
- Molecular Characterization during Clonal Evolution: High-Risk Neuroblastoma (PHS003111)

https://ccdi.cancer.gov/explore

Demonstration of the Explore Dashboard

Peter Gilbertson

CCDI Data Access Portals

Study-level directories

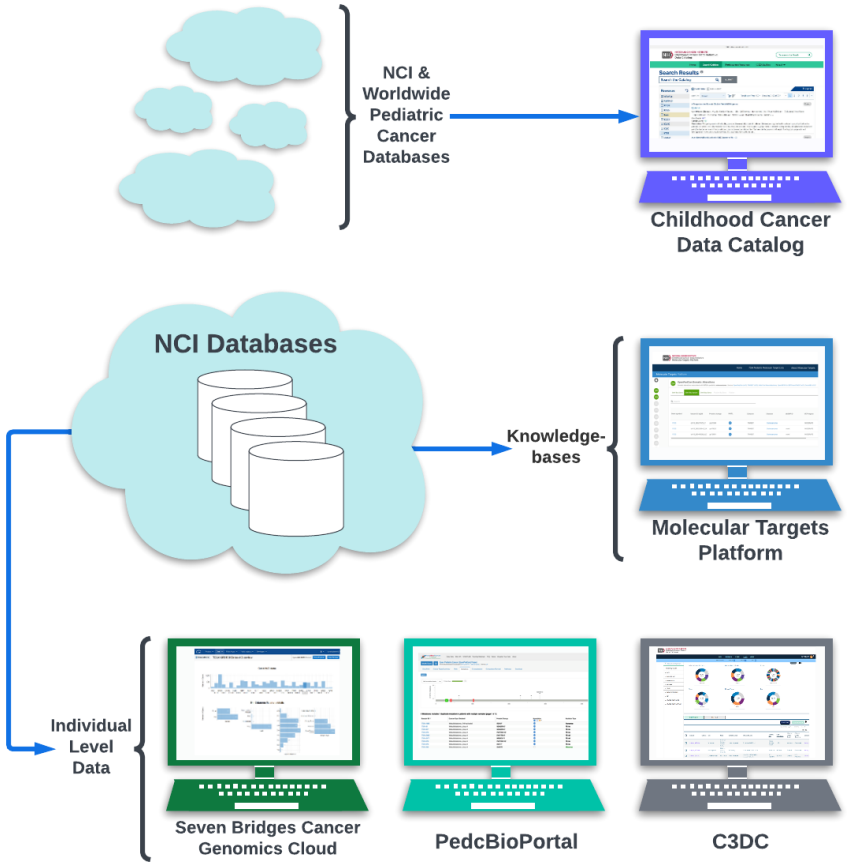
- Childhood Cancer Data Catalog (open access)

Aggregations and knowledge bases

- Molecular Targets Platform (open access)

Individual-level data

- *Custom analyses:* Cancer Genomics Cloud
- *Clinical:* Childhood Cancer Clinical Data Commons (C3DC)
- *Genomics:* PedcBioPortal



Controlled Data Access Process

Sean Burke

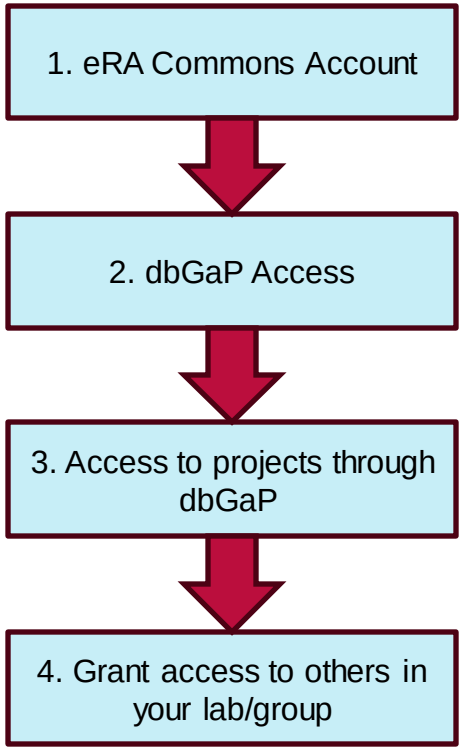
Obtain Controlled Access Files

1. The database of Genotypes and Phenotypes (dbGaP) access is given using eRA Commons accounts.
   - Go to the eRA Commons site and create an account under your organization or institution

2. Go to the dbGaP Controlled Access Data section and select Authorized Access. Login with your eRA Commons Account.

3. Create a Research Project.
   - Select the projects you would like controlled access to
   - Create a Research Use Statement explaining the need for the projects
   - Confirm project structure and send off for review to the Data Access Committee

4. Go to the My Requests tab to see all current access that is linked to your eRA Commons Account.
   - Go to the Downloaders tab and search for other members in your lab/group and add them to the selected Research Projects



How to Apply for Controlled Access on dbGaP: https://www.youtube.com/watch?v=m0xp_cCO7kA

Controlled Data Access Process

- For the CCDI studies, genomic data is hosted in the Cancer Data Service (CDS), which is a data repository under the Cancer Research Data Commons infrastructure.
- dbGaP maintains a list of subject IDs, sample IDs, and consents.
- Study-level metadata, demographic, and diagnosis details are available on the CCDI Hub Explore Dashboard as open access.
- Accessing controlled-access data and clinical/phenotypic files requires authorization through dbGaP.
- Users can analyze CCDI data on the Cancer Genomics Cloud (CGC) through the Cancer Data Service Explorer.

- CCDI CGC Data Access Instructions: datacatalog.ccdi.cancer.gov/CCDI_CGC_Data_Access_Instructions_1.0.pdf
- Tutorial on how to import CDS data: docs.cancergenomicscloud.org/docs/import-cds-data

Navigating the Cancer Genomics Platform

Zélia Worman

CANCER GENOMICS CLOUD
SEVEN BRIDGES

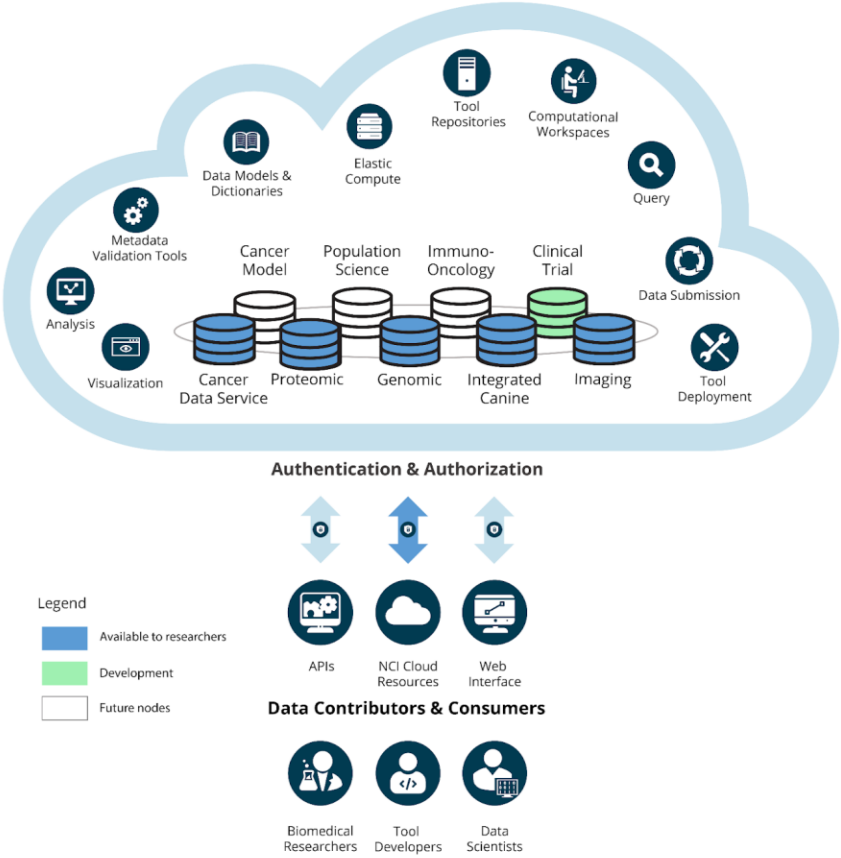
3+	1600+	800+	8000+	80000+
Petabytes Public Data	Years of Compute	Public Tools & Workflows	Users	User-Created Workflows

Provides powerful, yet easy to use interfaces to empower cancer researchers to draw new insights from petabyte scale data.

Stable, secure, and highly customizable cloud storage and computing platform.

**Data Hub, CDA, & CTDC have not reached go-live*

NCI Cancer Research Data Commons (CRDC)



Who are the CGC Users?

The CGC is designed to serve a wide range of scientists and users with varying skill sets.

ADMINISTRATORS

- Manage and control users
- Monitor and control institutional assets
- Manage and monitor projects
- Monitor and control costs
- Create reports

BIOINFORMATICIANS

- Store, manage, and share data
- Access public and proprietary data sets
- Query, build, and investigate cohorts of interest
- Access optimized tools and workflows
- Create, optimize, maintain, and distribute new tools and workflows
- Create push-button automation solutions
- Analyze data at scale with tools and workflows
- Conduct interactive exploratory analyses
- Explore and visualize results and gather insights
- Easily collaborate with other stakeholders
- Integrate with external systems

BENCH SCIENTISTS

- Store, manage, and share data
- Run optimized tools/ workflows at scale
- Conduct defined analyses via push-button solutions
- Investigate and visualize results
- Easily collaborate with other stakeholders

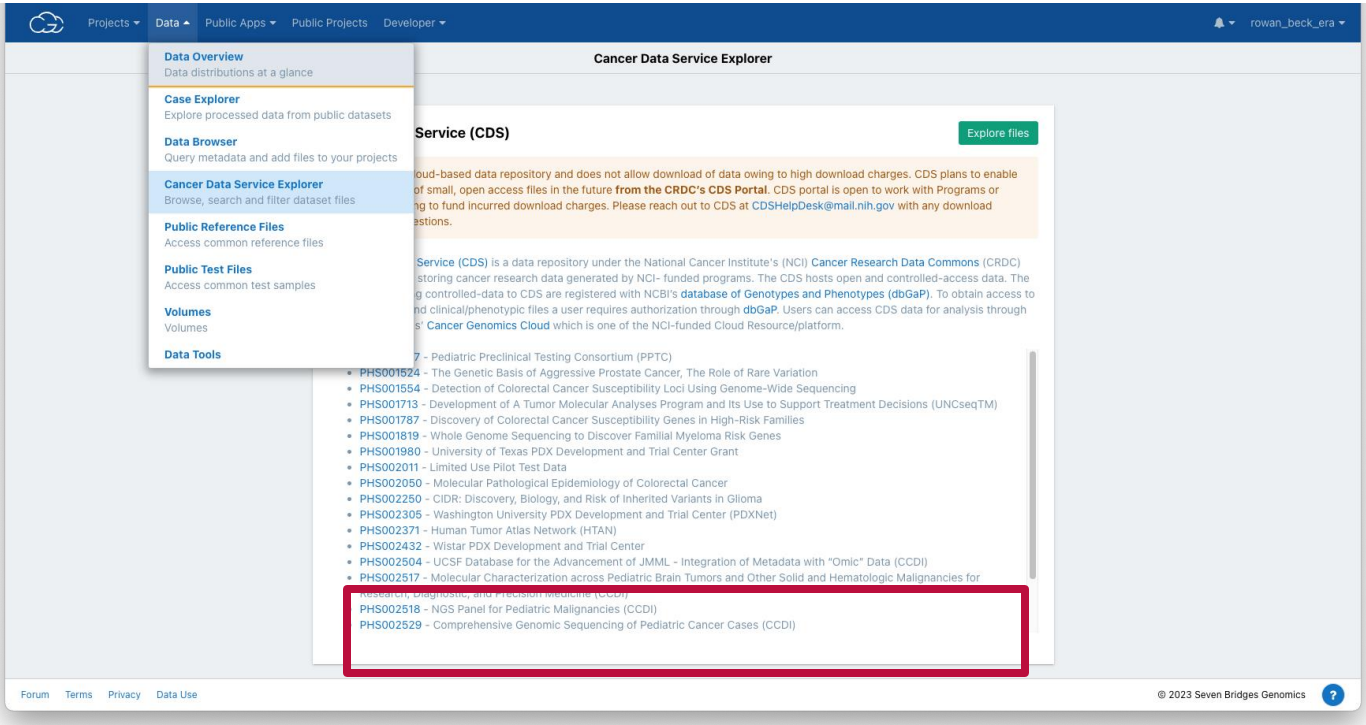
CLINICIANS

- Conduct validated analyses via push-button solutions
- Query, build, and investigate cohorts of interest
- Create reports
- Investigate and visualize results
- Easily collaborate with other stakeholders

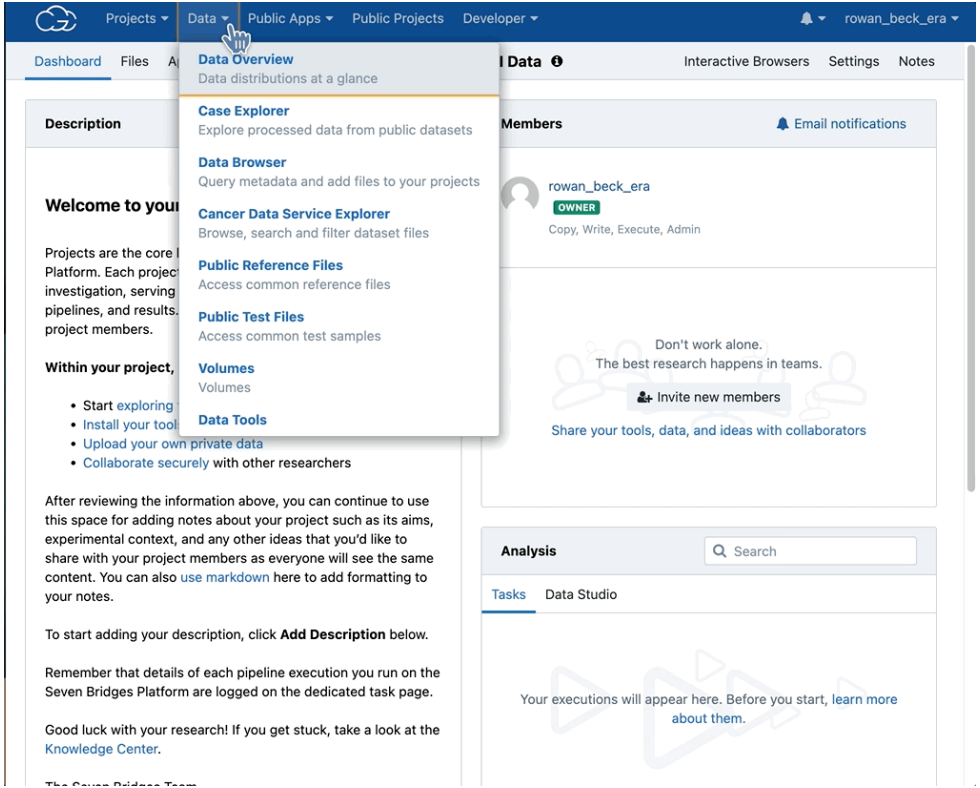
DEVELOPERS

- Create, optimize, and maintain new tools and workflows
- Create push-button automation solutions
- Create custom interfaces for specific use cases
- Distribute proprietary tools and workflows
- Integrate with upstream and downstream systems

Access CCDI Data through the Cancer Data Service Explorer



Access CCDI Data through the Cancer Data Service Explorer (cont.)



Petabytes of Public Data, at your Fingertips

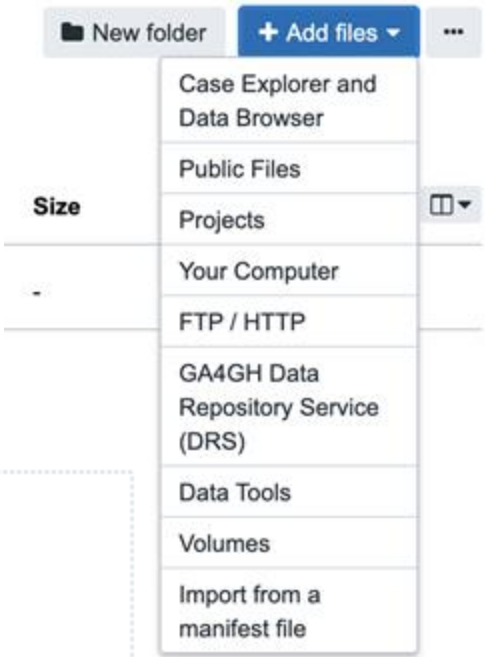
Researchers have access to various analysis tools, with the option to upload data either through the user interface portal or the command line.

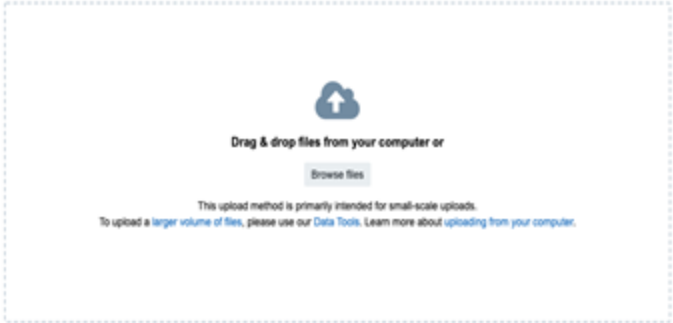
NCI Resources:

- The Cancer Image Archive
- Childhood Cancer Data Initiative
- Clinical Proteomic Tumor Analysis Consortium
- Therapeutically Applicable Research to Generate Effective Treatments (TARGET)
- The Cancer Genome Atlas Program (TCGA)
- Human Tumor Atlas Network

Additional Resources:

- ICGC Data Portal
- The Personal Genome Project





Access to a Public Applications Gallery

Platform Tool/Workflow Repository

900+

Curated tools & workflows

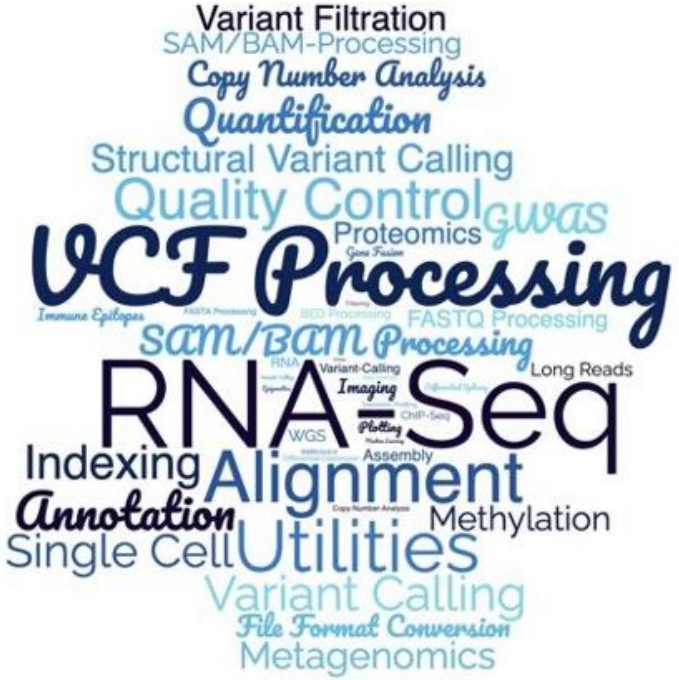
Secure

High quality apps & documentation

Optimized to run on cloud

Updated regularly

Customized user tools/workflows



Integrated Custom Tertiary Analysis Tools

Data Science Workbench

Derive new insights using interactive analysis environments with JupyterLab and RStudio environments. Code in Python and R and create Jupyter Notebooks to record and share your analyses.

jupyter

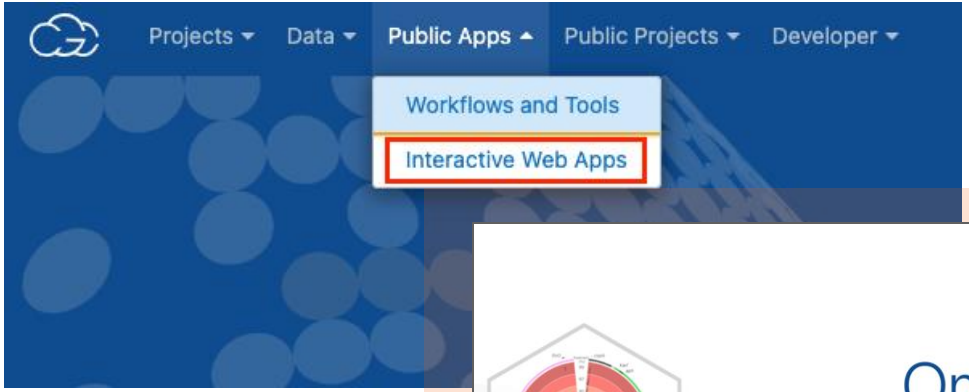
RStudio

Galaxy PROJECT

Open Health Imaging Foundation

COMING SOON

sas



NIH NATIONAL CANCER INSTITUTE
Center for Biomedical Informatics & Information Technology

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

GitHub

Making an Impact

112

Publications + citations

30+

Webinars + video tutorials

100+

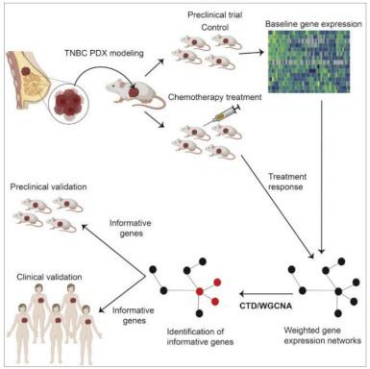
Office hour sessions

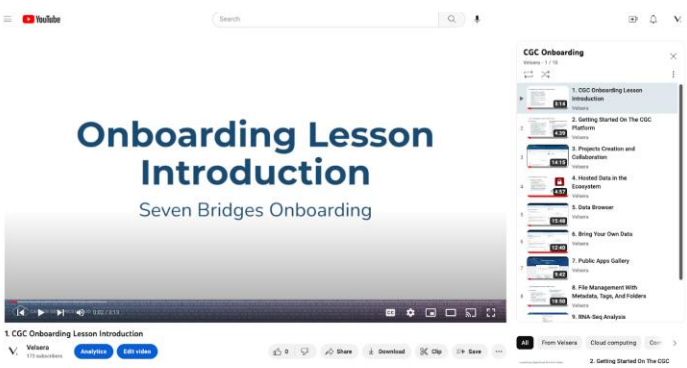
LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint

Wilson McKerrow[a,b], Xuya Wang[a,b], Carlos Mendez-Dorantes[c,d], Paolo Mita[a,b], Song Cao[e,f], Mark Grivainis[a,b], Li Ding[e,f], John LaCava[g,h], Kathleen H. Burns[c,d], Jef D. Boeke[a,b,i,1], and David Fenyö[a,b,1]

[a]Institute for Systems Genetics, New York University Grossman School of Medicine, New York, NY 10016; [b]Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY 10016; [c]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215; [d]Department of Pathology, Harvard Medical School, Boston, MA 02115; [e]Department of Medicine and Genetics, Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63110; [f]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108; [g]Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065; [h]European Research Institute for the Biology of Ageing, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands; and [i]Department of Biomedical Engineering, Tandon School of Engineering, Brooklyn, NY11201

Contributed by Jef D. Boeke; received August 31, 2021; accepted January 14, 2022; reviewed by Molly Gale Hammell and Michael MacCoss





Register for a Cancer Genomics Cloud webinar: bit.ly/CGCwebinar

Join Our Office Hours

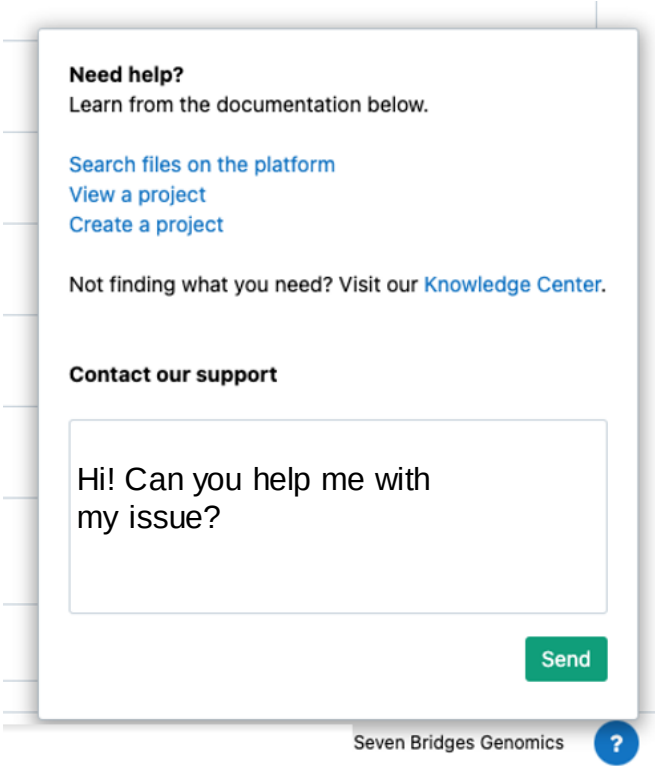
Questions? Need help?

We hold sessions twice a week:
Tuesdays at 10AM and
Thursdays at 2PM ET

Come chat with us about your research!

Scan the QR code to join!

Learn more at:
cancergenomicscloud.org



Find Out More About CCDI

Visit the CCDI Hub and access the CCDI Data Ecosystem.
ccdi.cancer.gov

Subscribe to our monthly newsletter.
cancer.gov/CCDI

Questions? Email us.
NCIChildhoodCancerDataInitiative@mail.nih.gov

Q&A

Navigating St. Jude's PeCan v2 & Survivorship Data Sharing Tools

Tuesday, January 23rd, 1pm - 2pm EST

Clay McLeod

Director of Product and Engineering
St. Jude Department of
Computational Biology

Dr. Xin Zhou

Faculty Member
St. Jude Department of
Computational Biology

Register Here: https://cbiit.webex.com/weblink/register/rd6032b43f1af64ba0c577f6095ce8709

Thank you!

U.S. Department of Health & Human Services

National Institutes of Health | National Cancer Institute

cancer.gov

1-800-4-CANCER