

Navigating CCDI Hub's Explore Dashboard and Data Access

Childhood Cancer Data Initiative Webinar Series

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

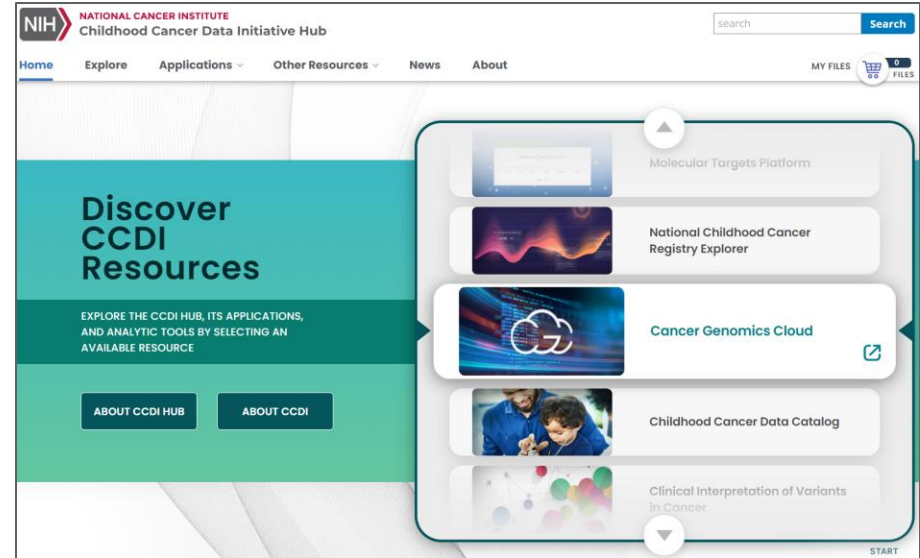
Overview of the CCDI Hub's Explore Dashboard

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



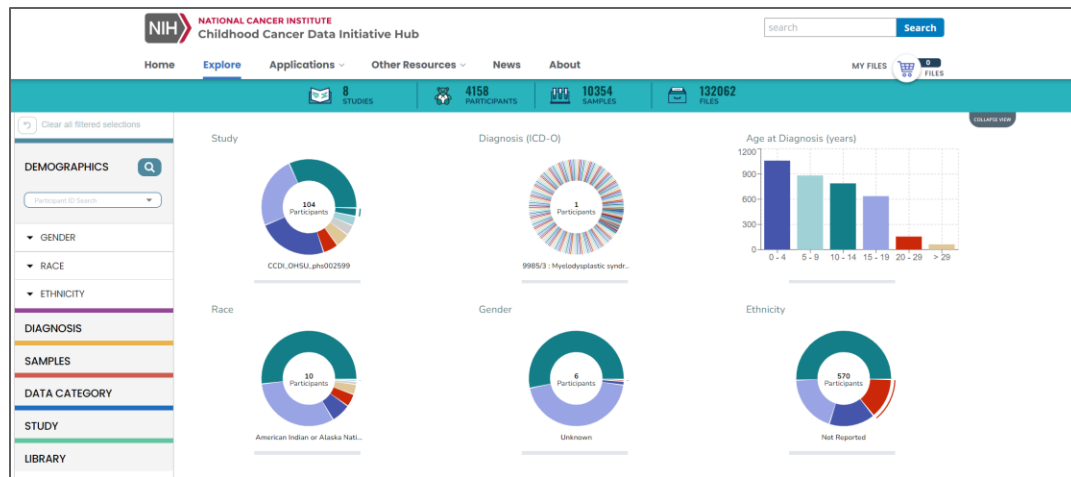
Explore Dashboard: Brings together CCDI-managed data & allows exploration of data in new ways.

What it is:

- An inventory of CCDI-managed childhood cancer data

What it does:

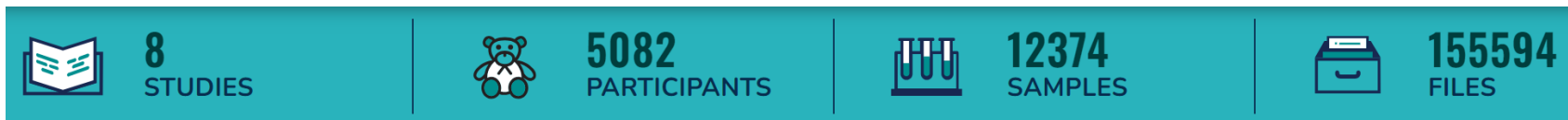
- Makes it easier for researchers to find data
- Faceted search
- Visualize results
- Exportable results



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>

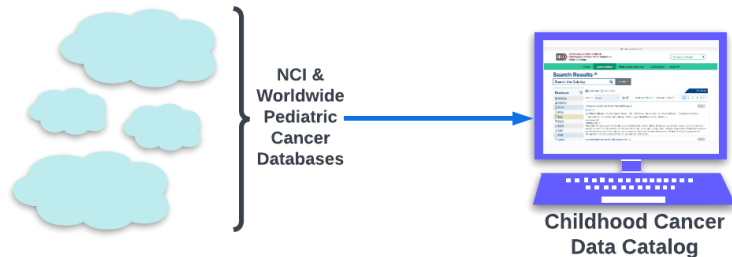
Demonstration of the Explore Dashboard

Peter Gilbertson

CCDI Data Access Portals

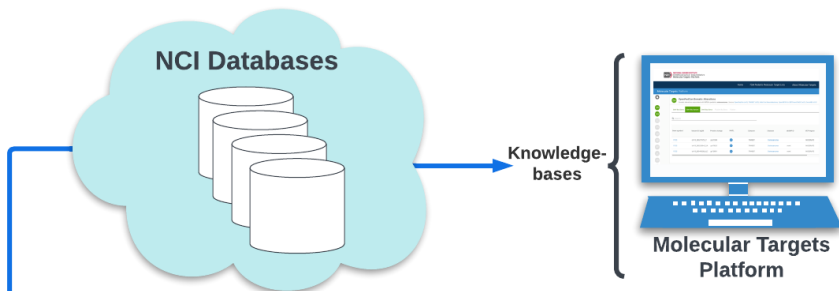
Study-level directories

- Childhood Cancer Data Catalog (open access)



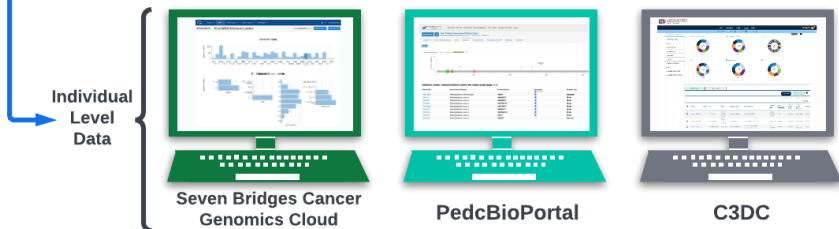
Aggregations and knowledge bases

- Molecular Targets Platform (open access)



Individual-level data

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

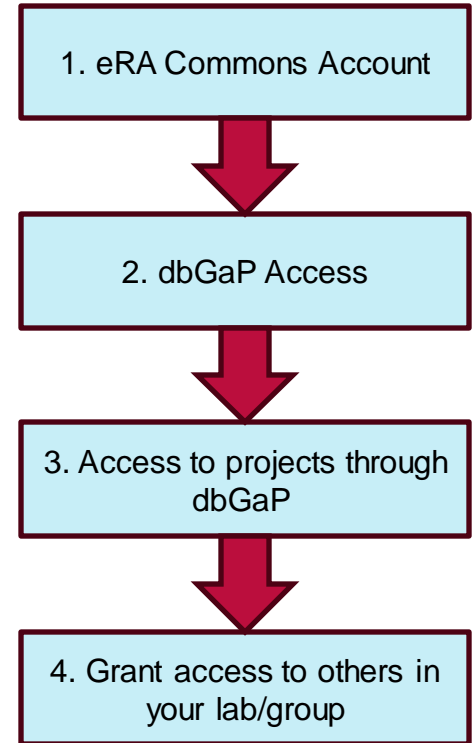


Controlled Data Access Process

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](#).



This Photo by Unknown Author is licensed under [CC BY-NC](#)

- 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.cancer-genomics-cloud.org/docs/import-cds-data

Navigating the Cancer Genomics Platform

Zélia Worman



CANCER GENOMICS CLOUD

SEVEN BRIDGES

3+

Petabytes
Public Data

1600+

Years of
Compute

800+

Public Tools
& Workflows

8000+

Users

80000+

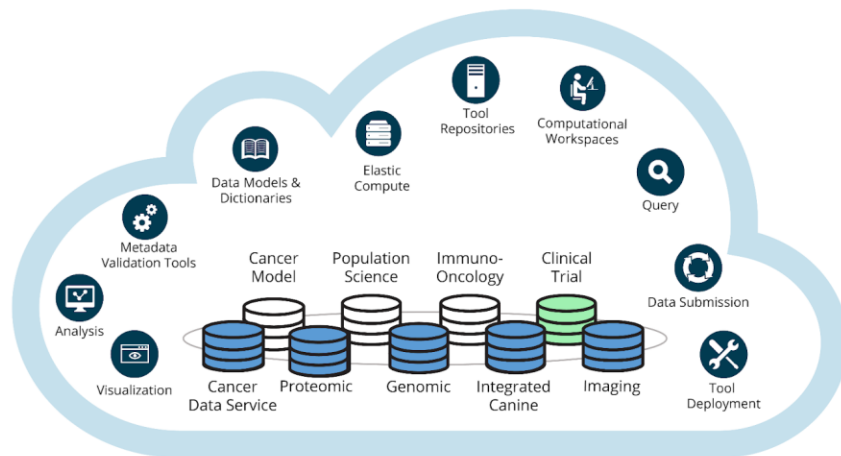
User-Created
Workflows

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

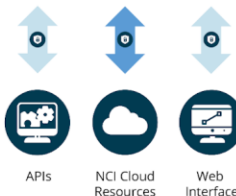
Stable, secure, and highly customizable cloud storage and computing platform.

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

NCI Cancer Research Data Commons (CRDC)



Authentication & Authorization



Legend

- Available to researchers
- Development
- Future nodes

Data Contributors & Consumers



Who are the CGC Users?

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



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



ADMINISTRATORS

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

Access CCDI Data through the Cancer Data Service Explorer

The screenshot displays the Cancer Data Service Explorer web application. The top navigation bar includes 'Projects', 'Data', 'Public Apps', 'Public Projects', and 'Developer'. A user profile 'rowan_beck_era' is visible in the top right. A sidebar on the left contains a menu with options: 'Data Overview', 'Case Explorer', 'Data Browser', 'Cancer Data Service Explorer' (selected), 'Public Reference Files', 'Public Test Files', 'Volumes', and 'Data Tools'. The main content area is titled 'Cancer Data Service Explorer' and features a 'Service (CDS)' section with an 'Explore files' button. Below this, there is a list of projects. Two projects are highlighted with a red box: PHS002518 - NGS Panel for Pediatric Malignancies (CCDI) and PHS002529 - Comprehensive Genomic Sequencing of Pediatric Cancer Cases (CCDI). The footer includes 'Forum', 'Terms', 'Privacy', 'Data Use', and '© 2023 Seven Bridges Genomics'.

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

The screenshot displays the Cancer Data Service Explorer interface. At the top, a navigation bar includes 'Projects', 'Data', 'Public Apps', 'Public Projects', and 'Developer'. A hand cursor points to the 'Data' menu, which is open, showing options: 'Data Overview' (Data distributions at a glance), 'Case Explorer' (Explore processed data from public datasets), 'Data Browser' (Query metadata and add files to your projects), 'Cancer Data Service Explorer' (Browse, search and filter dataset files), 'Public Reference Files' (Access common reference files), 'Public Test Files' (Access common test samples), 'Volumes' (Volumes), and 'Data Tools'.

The main content area is titled 'Data Overview' and includes a 'Description' section with a 'Welcome to you' message. It explains that projects are the core of the platform and lists actions like 'Start exploring', 'Install your tool', 'Upload your own private data', and 'Collaborate securely with other researchers'. Below this, there is a section for adding project descriptions and a note about pipeline execution logging.

On the right side, the 'Members' section shows the user 'rowan_beck_era' as the 'OWNER' with permissions to 'Copy, Write, Execute, Admin'. It also features a 'Don't work alone' message and an 'Invite new members' button. Below the members section is an 'Analysis' section with a search bar and a 'Data Studio' task.

Petabytes of Public Data, at your Fingertips

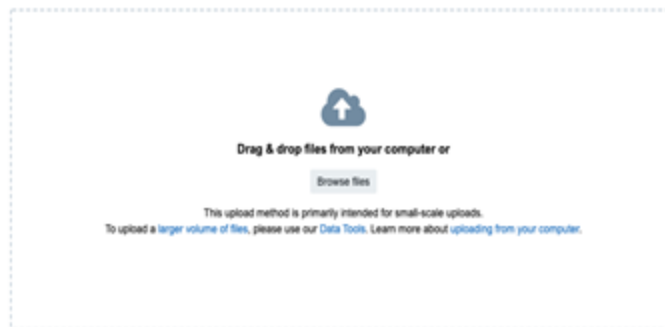
Researchers have access to various analysis tools, with the option to upload data either through the user interface portal or the command line.

NCI Resources:

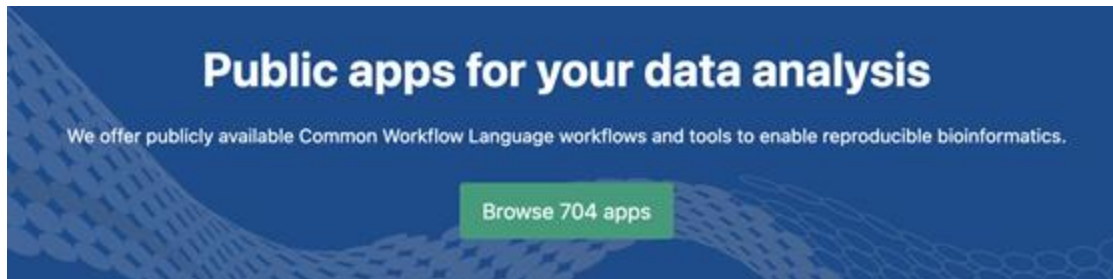
- [The Cancer Image Archive](#)
- [Childhood Cancer Data Initiative](#)
- [Clinical Proteomic Tumor Analysis Consortium](#)
- [Therapeutically Applicable Research to Generate Effective Treatments \(TARGET\)](#)
- [The Cancer Genome Atlas Program \(TCGA\)](#)
- [Human Tumor Atlas Network](#)

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



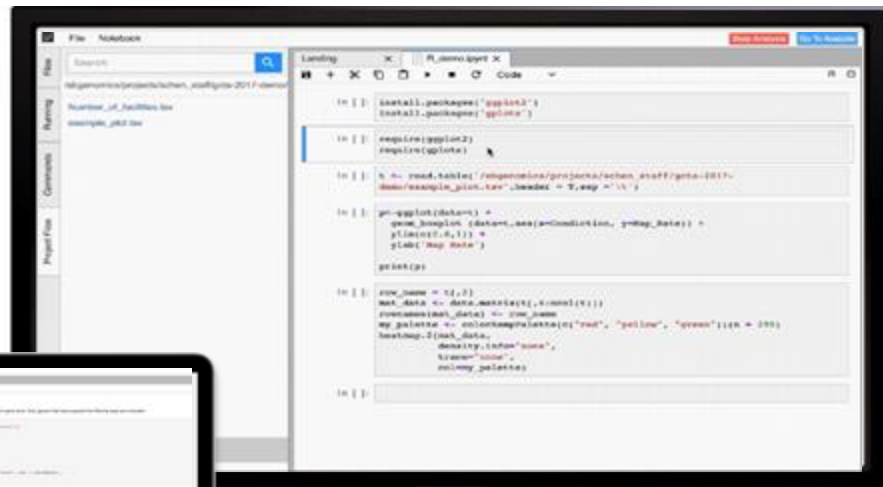
Customized user tools/workflows



Integrated Custom Tertiary Analysis Tools

Data Science Workbench

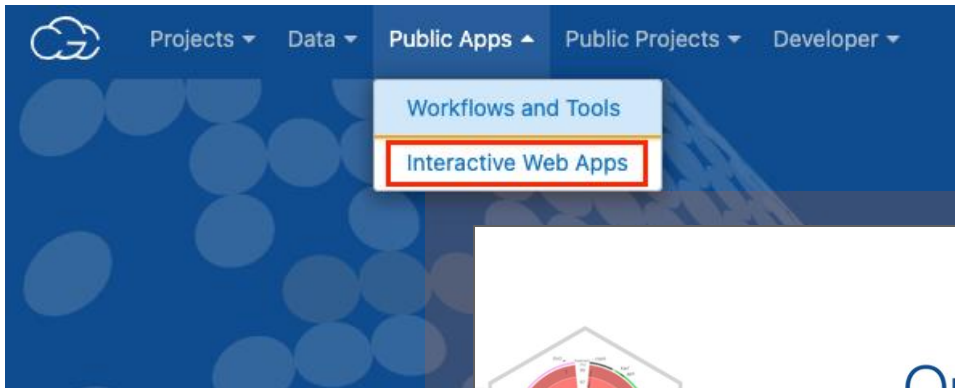
Derive new insights using interactive analysis environments with JupyterLab and RStudio environments. Code in Python and R and create Jupyter Notebooks to record and share your analyses.



- Open Health
- Imaging Foundation

COMING SOON





OmicCircos

Powerful, Informative,
Publishable

Circos Plots, Made Easy

[Start Discovery](#)

[User Guide](#) [References](#)

CANCER GENOMICS CLOUD
NIH NATIONAL CANCER INSTITUTE

Making an Impact

112

Publications + citations

30+

Webinars + video tutorials

100+

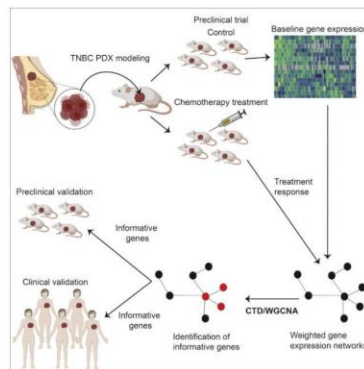
Office hour sessions

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 Fenyo^{a,b,1}

^aInstitute for Systems Genetics, New York University Grossman School of Medicine, New York, NY 10016; ^bDepartment of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY 10016; ^cDepartment of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215; ^dDepartment of Pathology, Harvard Medical School, Boston, MA 02115; ^eDepartment of Medicine and Genetics, Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63110; ^fMcDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108; ^gLaboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065; ^hEuropean Research Institute for the Biology of Ageing, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands; and ⁱDepartment of Biomedical Engineering, Tandon School of Engineering, Brooklyn, NY 11201

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

Need help?

Learn from the documentation below.

[Search files on the platform](#)

[View a project](#)

[Create a project](#)

Not finding what you need? Visit our [Knowledge Center](#).

Contact our support

Hi! Can you help me with my issue?

Send

Seven Bridges Genomics



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](https://www.cancer.gov)

1-800-4-CANCER

