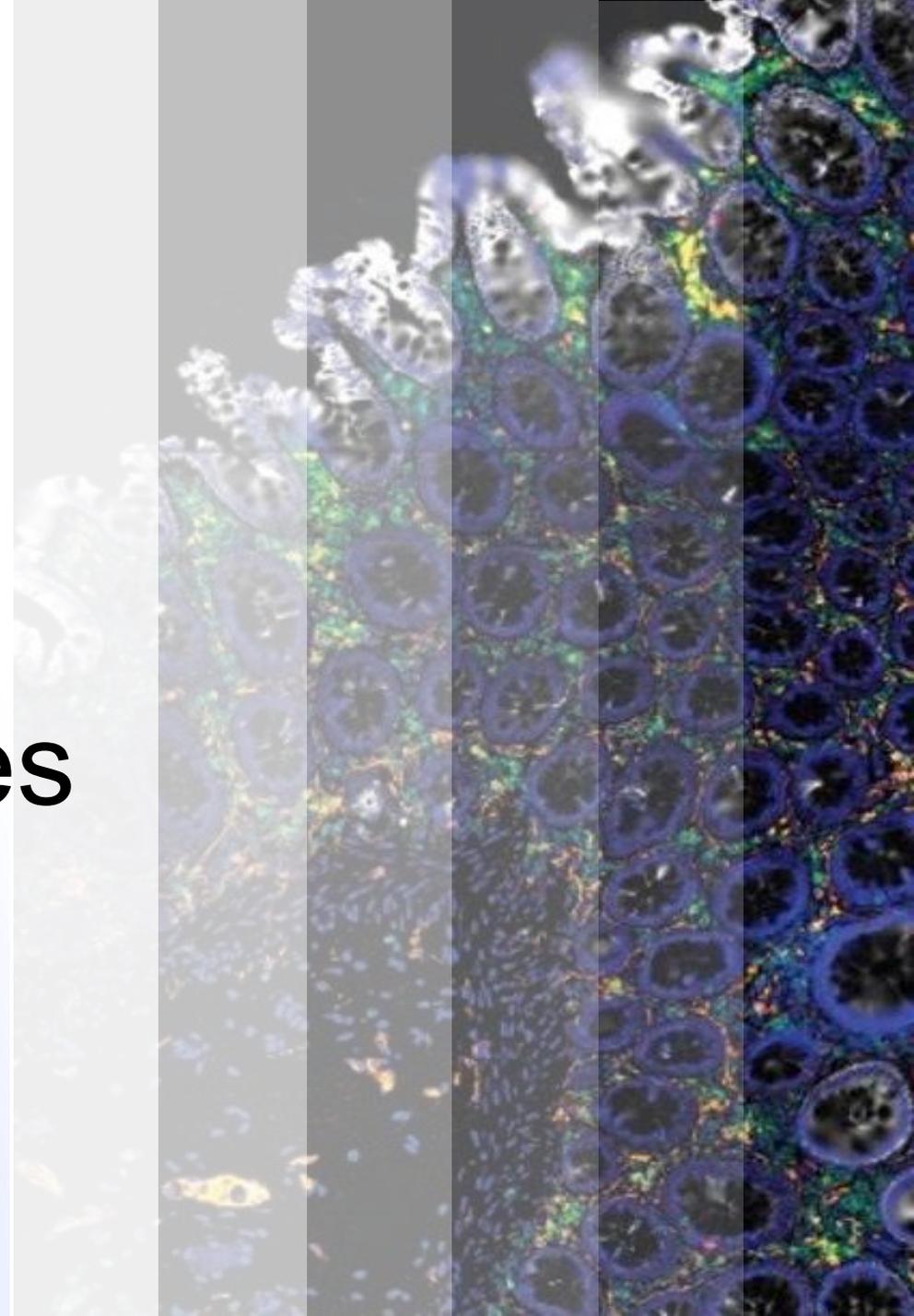


VELSERA.

HTAN via The CRDC's Seven Bridges Cancer Genomics Cloud

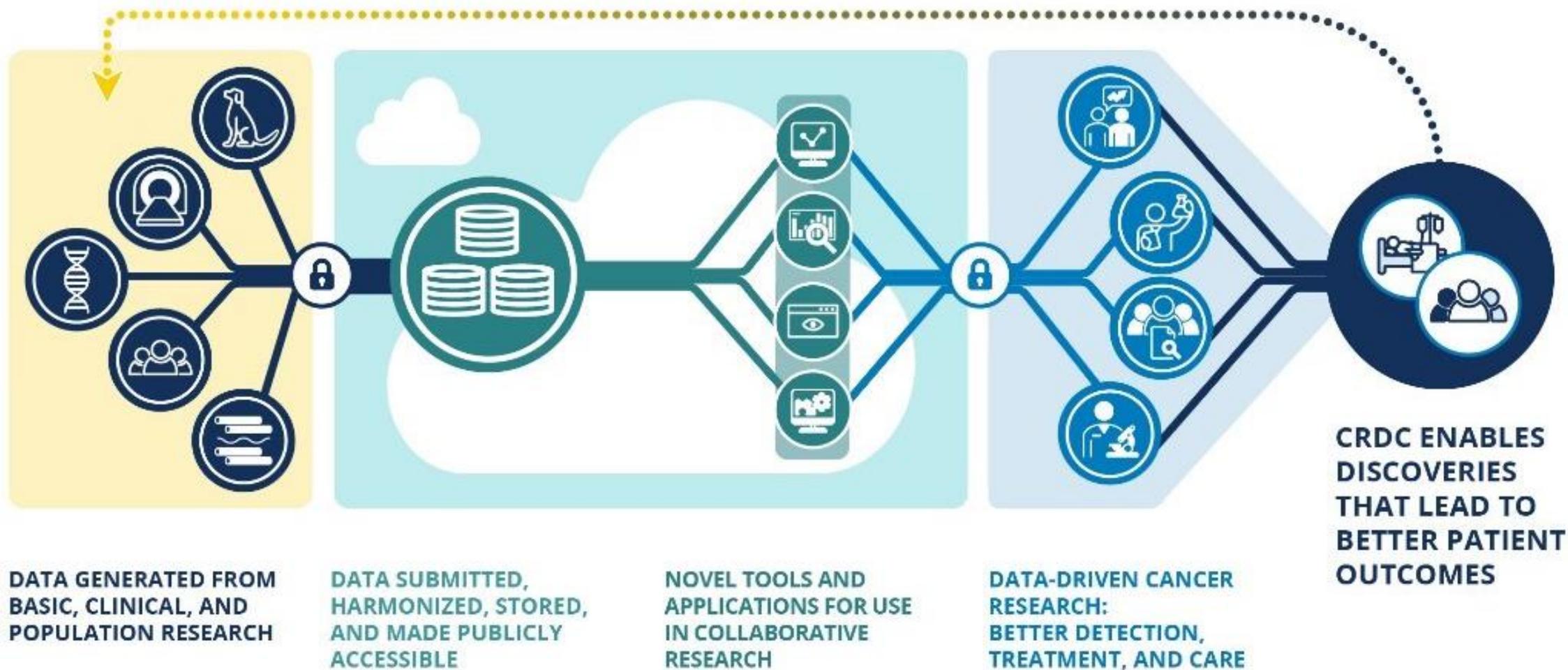
Rowan Beck
Seven Bridges SB-CGC
Bioinformatics Trainers and
Education Program



Agenda

- **Overview of the Seven Bridges Cancer Genomics Cloud (SB-CGC), powered by Velsera**
- **How to use SB-CGC for HTAN Integration**

Cancer Research Data Commons



 = Secure data storage and access

datacommons.cancer.gov



CANCER GENOMICS CLOUD

SEVEN BRIDGES

3+

Petabytes Public
Data

1600+

Years of
Compute

1000+

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.

Who are the SB-CGC Users?

The SB-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 Datasets
- 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/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/Visualize Results
- Easily Collaborate with Other Stakeholders



ADMINISTRATORS

- Manage and Control Users
- Monitor and Control Institutional Assets
- Manage and Monitor Projects
- Monitor and Control Costs
- Create Reports



CLINICIANS

- Conduct Validated Analyses via Push-button Solutions
- Query, Build, and Investigate Cohorts of Interest
- Create Reports
- Investigate/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/Workflows
- Integrate with Upstream/Downstream Systems

Access Public Data, or Use Your Own

The image displays two overlapping screenshots of a data management web application. The left screenshot shows a 'Cart > Selected Files' page with a table of selected files. The right screenshot shows a file browser interface with a dropdown menu open, listing various data sources and tools.

Cart > Selected Files

AVAILABLE EXPORT OPTIONS DOWNLOAD MANIFEST

Study Name	Accession	Participant Id	Sample Id	Study Access	File Type	Remove
GECCO OICR: Molecular Pathological Epidemiology of Colorectal Cancer	phs002050	GECCO_6002	GECCO_6002_Li_P_TS_1000.bam_DNA GECCO_6002_Ly_R_TS_1000.bam_DNA	Controlled	BAM	

File Browser

Case Explorer and Data Browser
Public Files
Projects
Your Computer
FTP / HTTP
GA4GH Data Repository Service (DRS)
Data Tools
Volumes
Import from a manifest file

Showing 1-3 of 3

Access HTAN Data Directly on the SB-CGC

Level 1 & 2 on the SB-CGC

- Access Controlled Sequencing data and Open Access Imaging data ([CC BY 4.0](#)).

Synapse - Open Access Processed level 3 and level 4 data

Imaging Data Commons (IDC) - Open Access Imaging data ([CC BY 4.0](#)) in [DICOM-TIFF](#) format

The screenshot shows the SB-CGC interface with a 'Cart > Selected Files' view. At the top right, there are buttons for 'AVAILABLE EXPORT OPTIONS' and 'EXPORT TO CANCER GENOMICS CLOUD', along with a 'DOWNLOAD MANIFEST' button. Below this is a table with columns: File Name, Study Name, Accession, Participant Id, Sample Id, Study Access, File Type, and Remove. The table contains several rows of data, including HTAN primary sequencing data and HTAN imaging data.

File Name ↑	Study Name	Accession	Participant Id	Sample Id	Study Access	File Type	Remove ↓
HT056P-S1PAA1-possorted_genome_bam.bam	Human Tumor Atlas Network (HTAN) primary sequencing data	phs002371	HTA12_1	HTA12_1_1_Tissue, HTA12_1_2_Tissue, HTA12_1_3_Tissue	Controlled	BAM	<input type="checkbox"/>
HT056P-S1PBA3-possorted_genome_bam.bam	Human Tumor Atlas Network (HTAN) primary sequencing data	phs002371	HTA12_1	HTA12_1_2_Tissue, HTA12_1_3_Tissue, HTA12_1_3_Tissue	Controlled	BAM	<input type="checkbox"/>
HT056P-S1RTA1-possorted_genome_bam.bam	Human Tumor Atlas Network (HTAN) primary sequencing data	phs002371	HTA12_1	HTA12_1_1_Tissue, HTA12_1_2_Tissue, HTA12_1_3_Tissue	Controlled	BAM	<input type="checkbox"/>
HT056P1Pavs	Human Tumor Atlas Network (HTAN) imaging data	phs002371_image	HTA12_1	HTA12_1_101_Tissue Biospecimen Type, HTA12_1_102_Tissue Biospecimen Type, HTA12_1_103_Tissue Biospecimen Type, HTA12_1_104_Tissue Biospecimen Type, HTA12_1_105_Tissue Biospecimen Type, HTA12_1_101_Tissue	Open	SVS	<input type="checkbox"/>

The screenshot shows the 'Importing DRS data' form. It includes a 'Destination project' dropdown menu (currently set to 'No project selected'), a 'Create new project' link, a 'Resolve naming conflicts' dropdown menu (set to 'Skip'), and an 'Add tags' field. At the bottom, there is a checkbox for a disclaimer and an 'Import data' button.

Browse Hundreds of Tools and Workflows

Public Apps for your data analysis

We offer publicly available Common Workflow Language workflows and tools to enable reproducible bioinformatics.

[Browse 939 apps](#)

GRAF Germline Variant Detection Workflow

The GRAF Germline Variant Detection Workflow enables accurate alignment and variant calling by utilizing a genome graph reference that can address the bias and other limitations inherent in linear genome references. Seven Bridges has constructed a comprehensive pan-genome graph that incorporates the...

[Alignment](#) [Variant Calling](#) [Graph](#)

[Copy](#) [Run](#)

Public Apps

Search: MCMICRO

Search result: 1 item

MCMICRO

MCMICRO is an end-to-end processing pipeline for multiplexed whole slide imaging and tissue microarrays. It comprises ...

[Copy](#) [Run](#)

MCMICRO

Enables the processing of multiplexed tissue images.

Transform whole-slide images into single-cell data using this simple workflow.

No Coding Required to Run an Analysis

The screenshot displays the Velsera web interface for configuring a task. The top navigation bar includes 'Projects', 'Data', 'Public Apps', 'Public Projects', and 'Developer'. The user 'rowan_beck_era' is logged in. The current task is titled 'DRAFT Differential Expression - Salmon + DESeq2 run - 11-30-23 17:12:19'. Below the title, there are buttons for 'Get support', 'Discard', and 'Run'. The task was last updated on Nov. 30, 2023 at 12:12. The app is identified as 'Differential Expression - Salmon + DESeq2 - Revision: 1'. The interface is divided into three main sections: 'Task Inputs', 'App Settings', and 'Output Settings'. 'Task Inputs' shows 'Batching' set to 'Off' and lists several FASTQ read files and a GTF annotation file. 'App Settings' includes 'DESeq2' parameters such as 'Covariate of interest' (Genotype) and 'Factor level' (WT for reference, KD for test). 'Output Settings' lists various output options, all currently set to 'No value'. An 'Activity Monitor' button is located at the bottom center.

Task Inputs

- Batching: Off
- FASTQ read files *
 - SRR9058997_1.fastq
 - SRR9058993_1.fastq
 - SRR9058992_2.fastq
 - SRR9058992_1.fastq
 - SRR9058991_2.fastq
 - ...and 25 more items
- GTF annotation *
 - GRCh38ERCC.ensembl95.gtf
- Genome FASTA *
 - No files selected

App Settings

- DESeq2 (#deseq2_1_26_0)
 - Covariate of interest *
 - Genotype
 - Factor level - reference *
 - WT
 - Factor level - test *
 - KD

Output Settings

DESeq2 analysis results	No value
Expression matrix genes	No value
Expression matrix transcripts	No value
Gene-level quantification	No value
HTML report	No value
HTML reports	No value
Normalized counts	No value
RData file	No value
Report zip	No value
Salmon Quant archive	No value
Salmon quant log	No value
Transcript-level quantification	No value
pheno out	No value

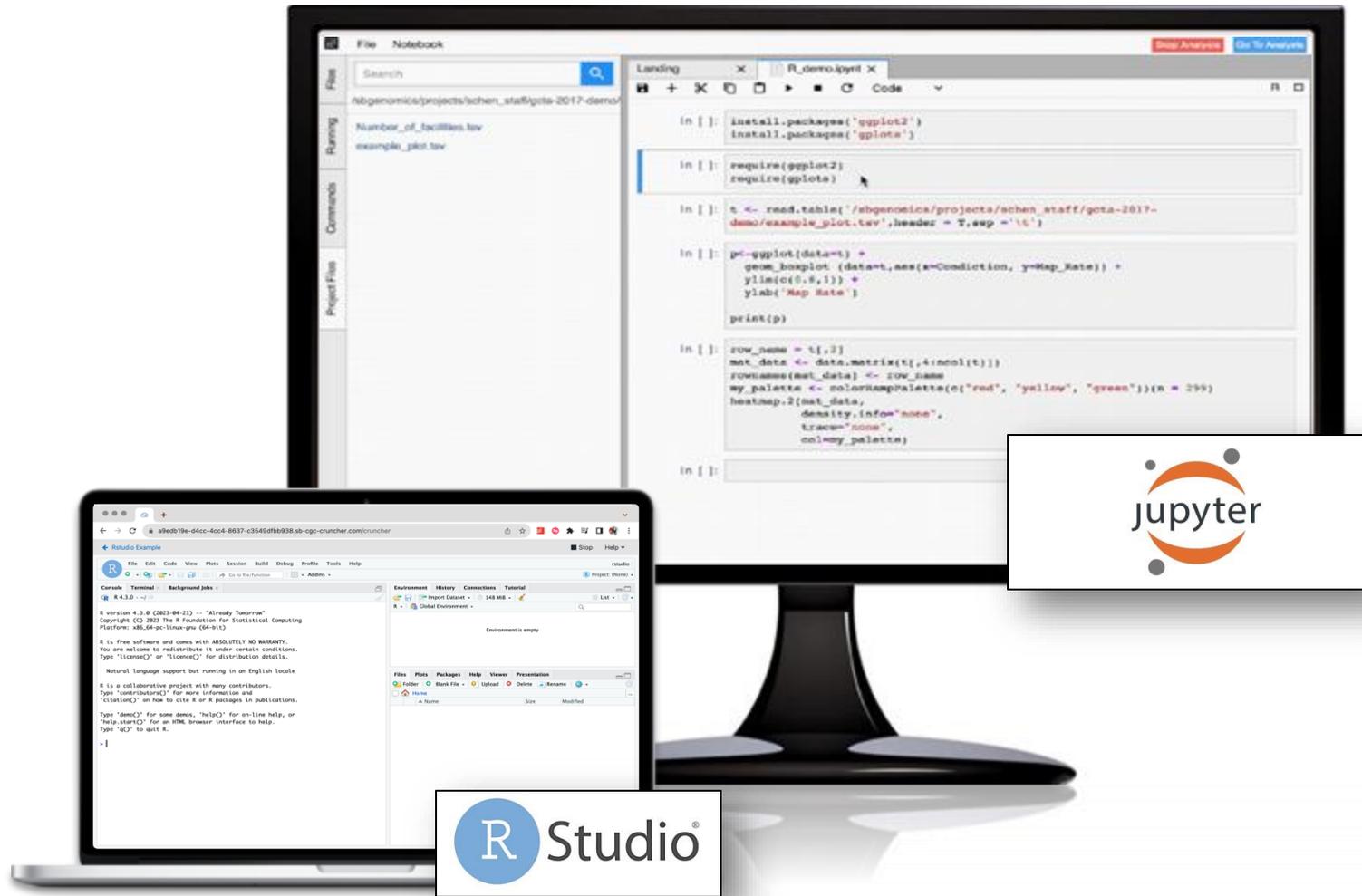
Integrated Custom Tertiary Analysis Tools

Data Science Workbench

Derive new insights using interactive analysis environments with **JupyterLab**, **SAS**, **Galaxy**, and **RStudio** environments.

Code in Python and create **Jupyter Notebooks** to record and share your analyses.

Similarly, turn your analyses into high quality documents, reports, presentations and dashboards using **R Markdown**.



Estimate Cloud Costs

Performance Benchmarking

- Runtimes
- Task Costs
- Various file sizes

Experiment type	Input size	Paired-end	# of reads	Read length	Duration	Cost	Instance (AWS)
RNA-Seq	2 x 230 MB	Yes	1M	101	18min	\$0.40	c4.8xlarge
RNA-Seq	2 x 4.5 GB	Yes	20M	101	30min	\$0.60	c4.8xlarge
RNA-Seq	2 x 17.4 GB	Yes	76M	101	64min	\$1.20	c4.8xlarge

Cloud Cost Estimator

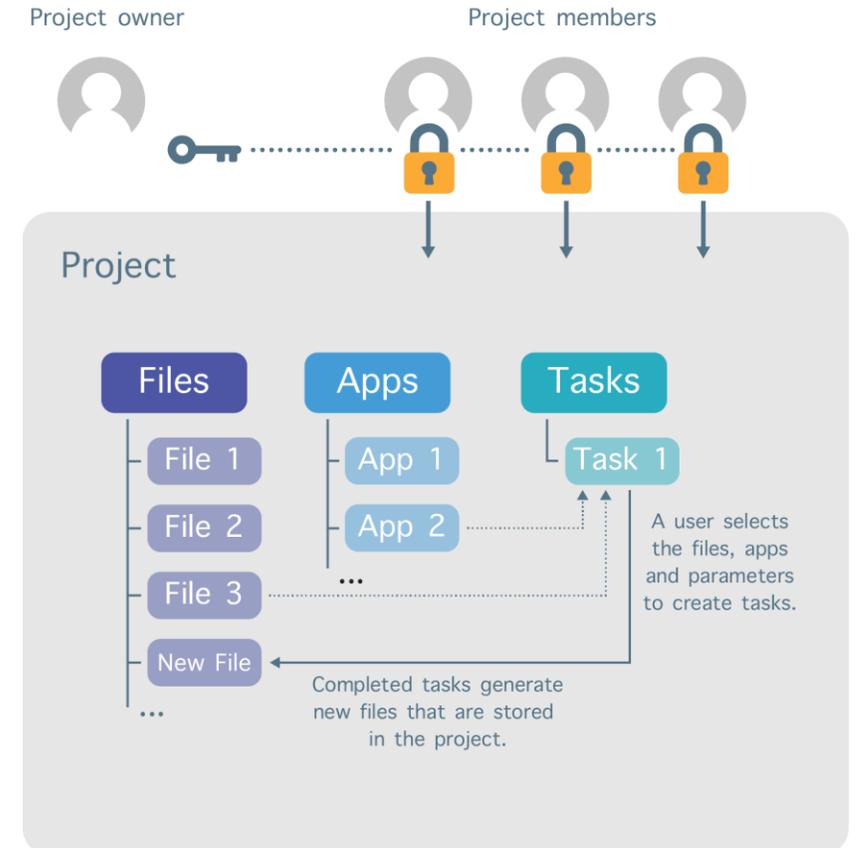
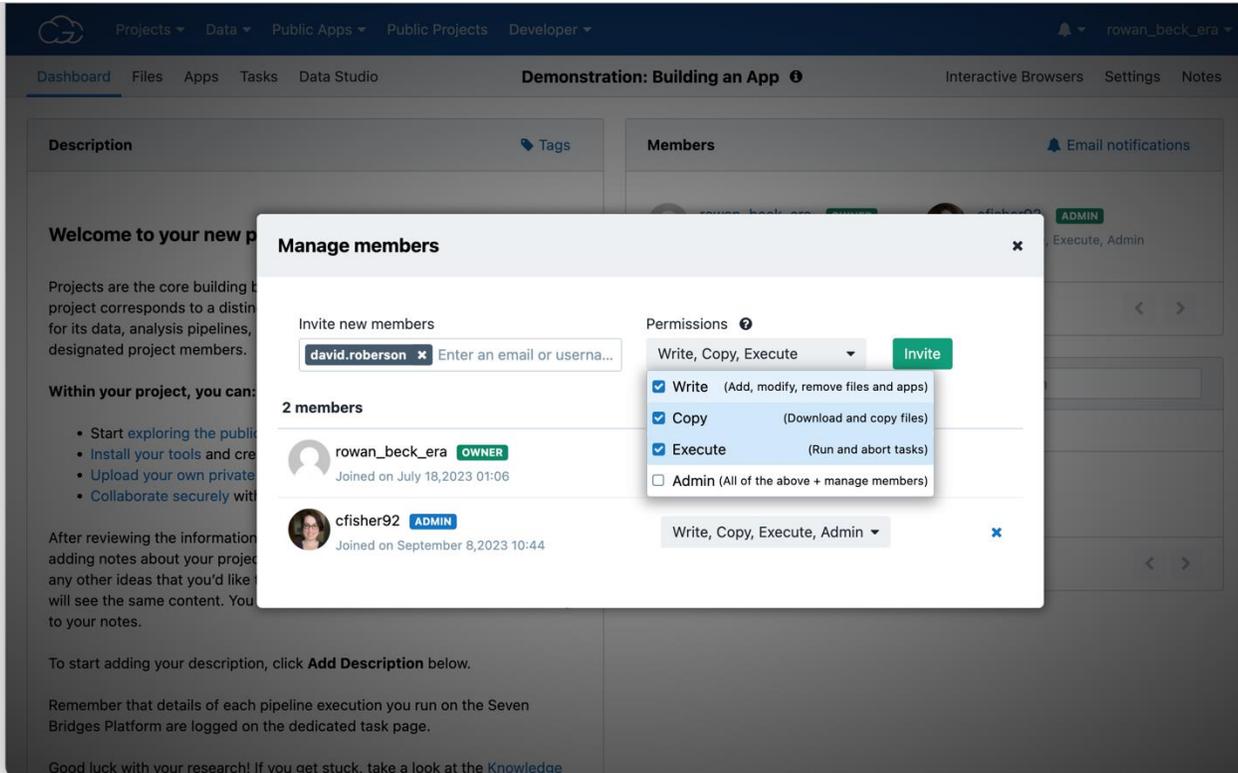
- Available for a limited number of apps
- Create an estimate for the cost of your specific use case
- Compare workflows to save on computing costs

Cost estimator: STAR

The cost is estimated based on these parameters:

- Spot Instances**
On
- File size**
119.75 GB
- Instance type**
Default
- Cost Estimation**
\$0.32 - 0.58

Collaborating Has Never Been Easier



Collaborating Has Never Been Easier

Pre-loaded with

- input and output files
- pre-run tasks
- recommended settings

Detailed descriptions on what the workflow does and how to use the pipeline

The screenshot displays the MCMICRO project page in the Cancer Genomics Cloud interface. The top navigation bar includes 'Projects', 'Data', 'Public Apps', 'Public Projects' (highlighted with an orange box and a red arrow), and 'Developer'. The main content area is divided into two columns. The left column, titled 'Description', contains the following text:

MCMICRO - End to End Microscopy Image Processing Public Project

MCMICRO is an end-to-end processing pipeline for multiplexed whole slide imaging and tissue microarrays. It comprises stitching and registration, segmentation, and single-cell feature extraction. Note that this is a CWL wrapper designed by Seven Bridges in order to cover default features of the original Nextflow app built by the analysis' authors. The original, Nextflow App is available on the [Github](#) or via the project's [homepage](#).

This Project demonstrates the usage of two CWL versions of MCMICRO, namely **MCMICRO** and **MCMICRO for HTAN** on four different samples. **MCMICRO for HTAN** is a trimmed version of the **MCMICRO** workflow which skips illumination correction, background subtraction and registration as it is tuned for HTAN data which is already registered.

HTAN data is available on the CGC through the [Cancer Data Service Explorer](#). For access to CWL Apps used in these workflows, do not hesitate to contact CGC support team.

Citation

Project participants agree to acknowledge the funding for the CGC in all publications and external presentations, as follows:

"The Seven Bridges Cancer Research Data Commons Cloud Resource has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Contract No. HHSN261201400008C and ID/IQ Agreement No. 17X146 under Contract No. HHSN261201500003I and 75N91019D00024."

and to cite the following in all publications:

Lau *et al* (2017) The Cancer Genomics Cloud: Collaborative, Reproducible, and Democratized—A New Paradigm in Large-Scale Computational Research. *Cancer Res.* 77(21):e3-e6. doi: 10.1158/0008-5472.CAN-17-0387.

The right column, titled 'Analysis', contains a search bar and a list of completed tasks:

- COMPLETED **MCMICRO run - Exemplar 002**
Submitted by: sevenbridges · Apr 13, 2023 16:16
- COMPLETED **MCMICRO run - Exemplar 001**
Submitted by: sevenbridges · Apr 13, 2023 16:13
- COMPLETED **MCMICRO for HTAN run - WD-76845**
Submitted by: sevenbridges · Apr 13, 2023 15:59
- COMPLETED **MCMICRO for HTAN run - HTMA402**
Submitted by: sevenbridges · Apr 13, 2023 15:56

Collaborating Has Never Been Easier

Harvard Medical School

Artem Sokolov
Clarence Yapp
Jeremy Muhlich
Yu-An Chen
Clemens Hug
Greg Baker
Juha Ruukonen
Edward Novikov
Robert Krueger

Heidelberg University

Denis Schapiro
Florian Wünnemann
Miguel Ibarra
Krešimir Beštak

Oregon Health and Sciences University

Allison Creason
Jeremy Goecks
Daniel Persson
Qiang Gu
Luke Sargent
Cameron Watson
Luke Strgar

Dana-Farber Cancer Institute

Ajit Johnson

Vanderbilt University

Darren Tyson

Brigham and Women's Hospital

Giorgio Gaglia

The Jackson Laboratory

Brian White

Broad Institute

Huan Wang

Brigham Young University

Matthew Hodgman

Indica Labs

Erik Burlingame

University of Macau

Yimin Zheng

Sage Bionetworks

Adam Taylor

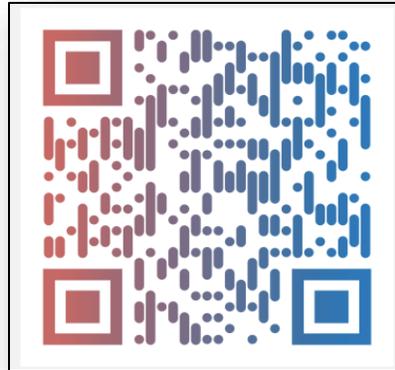
Leidos Biomedical Research

Shirish (Sam) Pathak

Cancer Research Data Commons (CRDC)

Erin Beck

Get The Support You Need



Every Week:

- 10:00 am ET Tuesday
- 2:00 pm ET Thursday

FAILED Task 1 - StringTie run [Get support](#) [View stats & logs](#)

Executed on June 12, 2020 07:53 by sevenbridges

Spot Instances: **On** | Memoization (WorkReuse): **Off** | Price: \$0.01 | Duration: 5 minutes

App: StringTie - Revision: 0

Error:
This task ran into a problem during execution and did not finish.
[Show details](#)

Inputs	App Settings	Output Settings
Aligned reads HCC1143-CCLE-RNASeq-subset01.genome_aligned...	Create input files for Ballgown and DESeq2: No value	Archived ballgown input tables: No value
Reference annotation file Homo_sapiens.GRCh38.84.gtf	Disable trimming: True	Assembled transcripts: No value
	Ignore alignments on the specified sequence: No value	Covered reference transcripts: No value
	Keep annotated transcripts only: No value	DESeq2 gene count matrix: No value
	Maximum fraction of multiply mapped reads: 0.95	DESeq2 transcript count matrix: No value
	Minimum anchor length for junctions: 10	Gene abundance estimation: No value
	Minimum isoform abundance: 0.1	
	Minimum isoform length: 200	
	Minimum junction coverage: 1	
	Minimum locus gap separation value: 50	
	Minimum read coverage: 2.5	
	Number of threads: 2	
	Output covered reference transcripts: False	
	Output gene abundance: False	
	Transcripts name prefix: STRG	

Funding Is Available on the SB-CGC

Diverse approaches and engagement strategies tailored to community needs

Pilot Credit Funds

- **\$300 of cloud credits**
- Free for new SB-CGC users
- Easy to request when signing up
- Fast approval

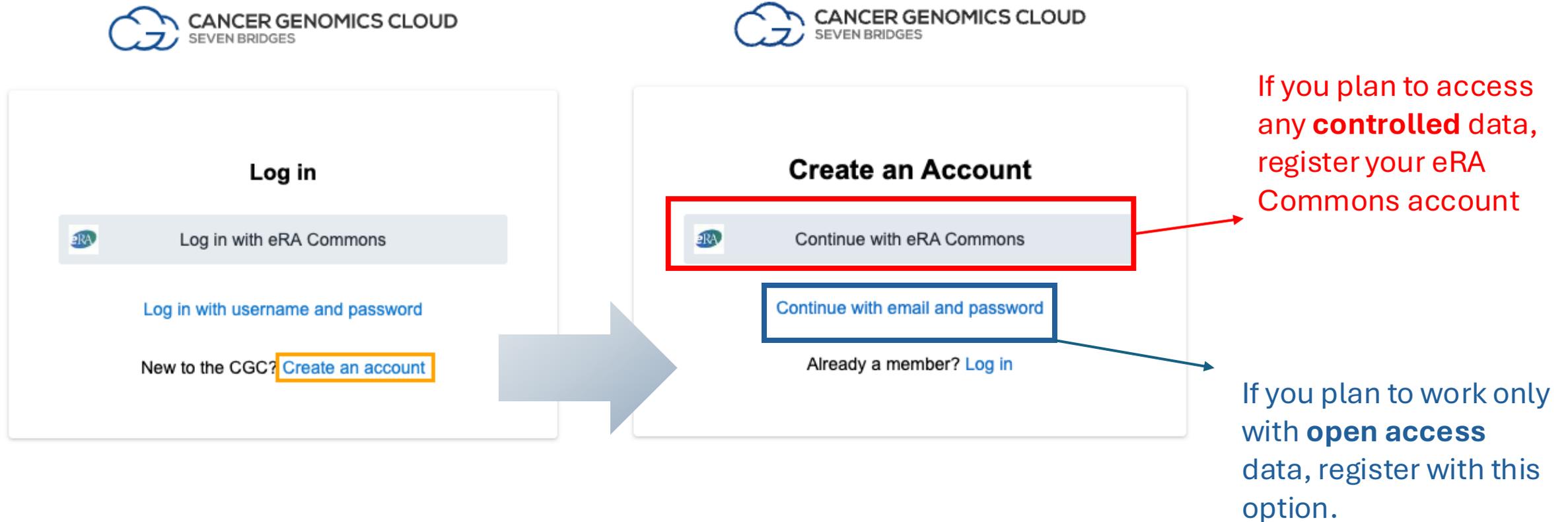
Collaborative Projects

- Cost estimation, optimization, and planning support
- Great for researchers new to bioinformatics and cloud approaches
- **Up to \$10k compute/storage costs**
- Fast, rolling applications
- To date > 60 projects



How to use the SB-CGC for HTAN Integration

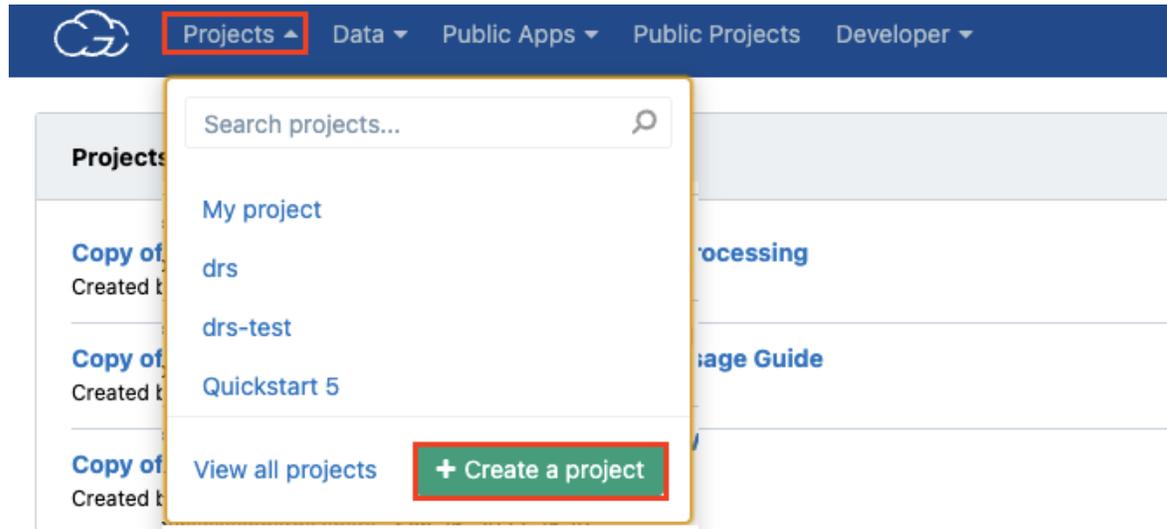
Register An Account



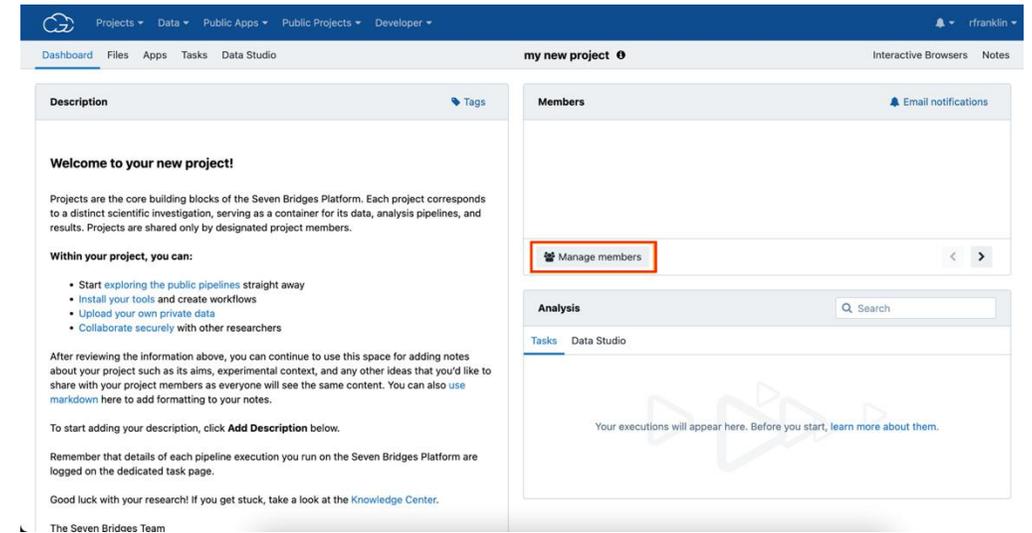
Register at:
CGC.sbgenomics.com

Analysis Flow

1. Create a Project

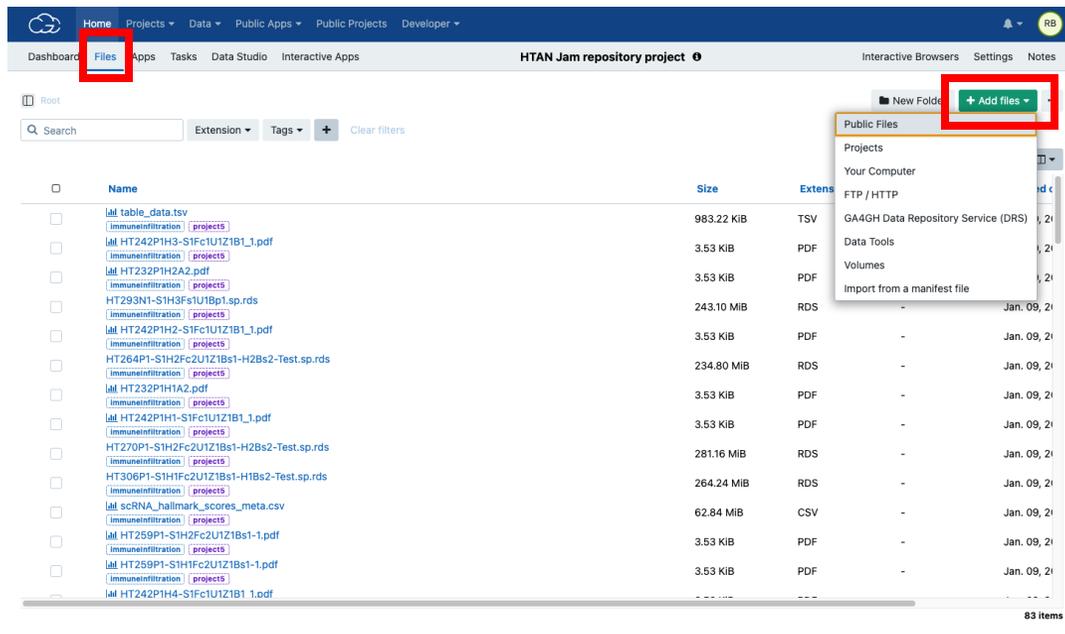


2. Add members

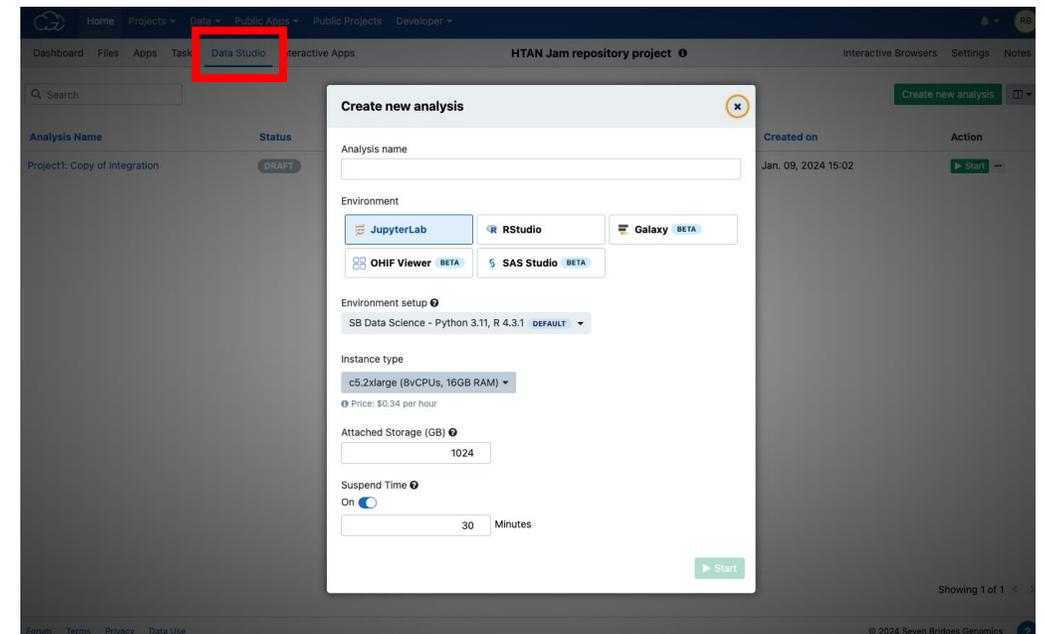


Analysis Flow

3. Add data (More information on next slide)



4. Conduct your analysis (More information available: Helpful Information – Accessing Files in Data Studio)



Data Access

The HTAN Portal leverages several repositories to provide access to data:

- **Synapse** - Open Access Processed level 3 and level 4 data
- **Imaging Data Commons (IDC)** - Open Access Imaging data (CC BY 4.0) in DICOM-TIFF format
- **Seven Bridges Cancer Genomics Cloud** - Level 1 & 2 Access- Controlled Sequencing data and Open Access Imaging data (CC BY 4.0). Access control for the sequencing data is managed through dbGaP (Study Accession: phs002371).

Instructions for accessing data from these repositories can be found directly from the [Explore Page](#) by making a selection of files and clicking on the download button. More information about accessing data can be found [here](#).

Data Access

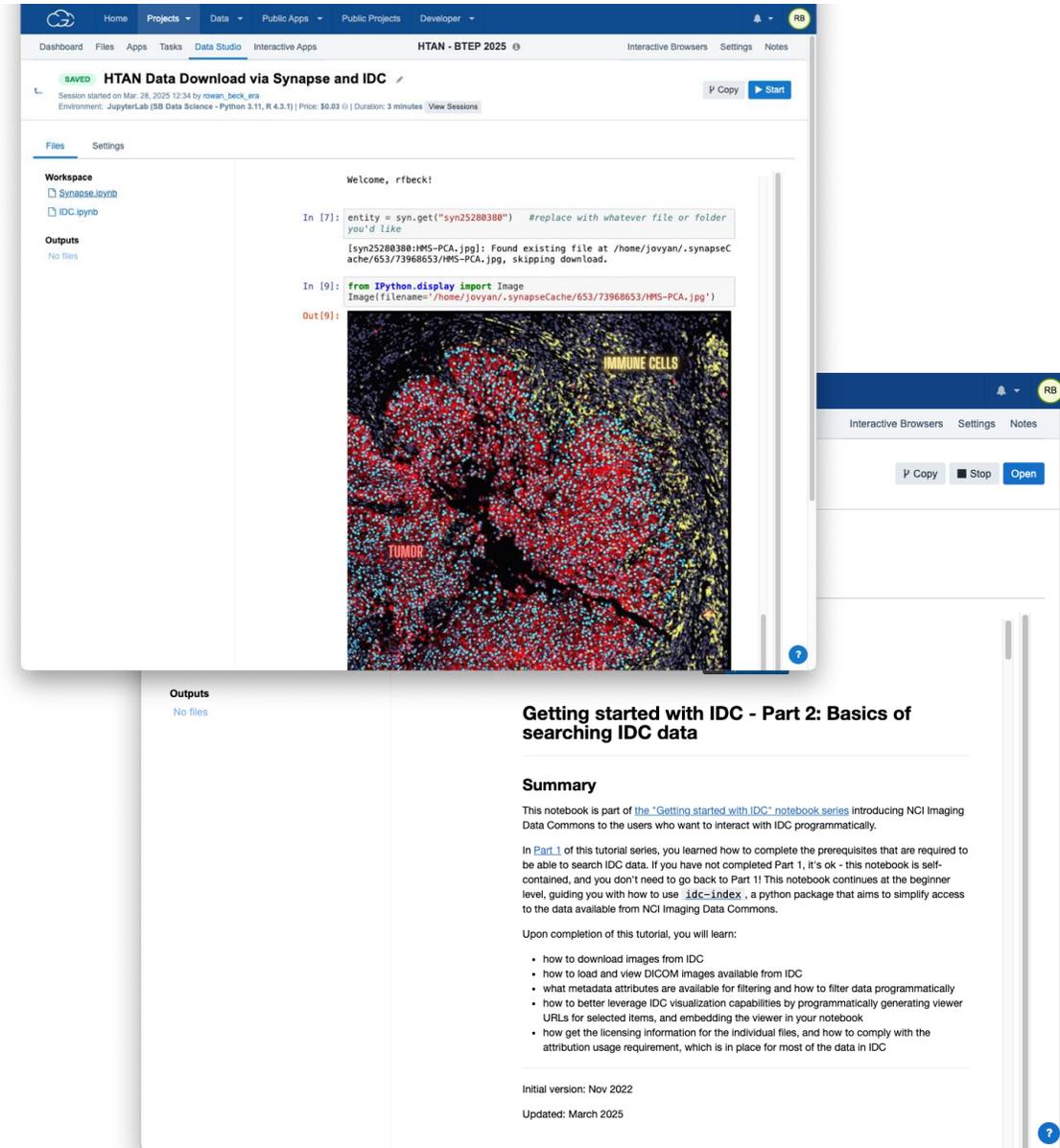
[Synapse](#) - Open Access
Processed [level 3](#) and [level 4](#) data

[Imaging Data Commons \(IDC\)](#) - Open Access
Imaging data ([CC BY 4.0](#)) in [DICOM-TIFF](#) format

Pre-built Jupyter notebooks to obtain data through Synapse or the Imaging Data Commons

<https://learn.canceridc.dev/data/downloading-data>

https://docs.humantumoratlas.org/open_access/synapse_to_cds/



The image shows a Jupyter notebook interface with a dark theme. The top navigation bar includes 'Home', 'Projects', 'Data', 'Public Apps', 'Public Projects', and 'Developer'. The notebook title is 'HTAN Data Download via Synapse and IDC'. The left sidebar shows a workspace with files 'Synapse.ipynb' and 'IDC.ipynb'. The main area displays code execution output:

```
Welcome, rfbec!
```

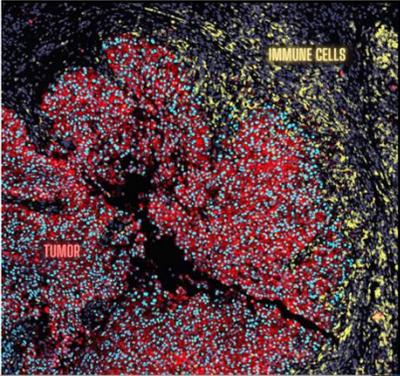
```
In [7]: entity = syn.get("syn25280388") #replace with whatever file or folder you'd like
```

```
[syn25280388:HMS-PCA.jpg]: Found existing file at /home/jovyan/.synapseCache/653/73968653/HMS-PCA.jpg, skipping download.
```

```
In [9]: from IPython.display import Image
```

```
Image(filename='/home/jovyan/.synapseCache/653/73968653/HMS-PCA.jpg')
```

```
Out [9]:
```



The visualization shows a histology image with red and blue regions. The red region is labeled 'TUMOR' and the blue region is labeled 'IMMUNE CELLS'. The notebook interface also shows a 'Getting started with IDC - Part 2: Basics of searching IDC data' section with a summary and a list of learning objectives.

Getting started with IDC - Part 2: Basics of searching IDC data

Summary

This notebook is part of the ["Getting started with IDC" notebook series](#) introducing NCI Imaging Data Commons to the users who want to interact with IDC programmatically.

In [Part 1](#) of this tutorial series, you learned how to complete the prerequisites that are required to be able to search IDC data. If you have not completed Part 1, it's ok - this notebook is self-contained, and you don't need to go back to Part 1! This notebook continues at the beginner level, guiding you with how to use `idc-index`, a python package that aims to simplify access to the data available from NCI Imaging Data Commons.

Upon completion of this tutorial, you will learn:

- how to download images from IDC
- how to load and view DICOM images available from IDC
- what metadata attributes are available for filtering and how to filter data programmatically
- how to better leverage IDC visualization capabilities by programmatically generating viewer URLs for selected items, and embedding the viewer in your notebook
- how get the licensing information for the individual files, and how to comply with the attribution usage requirement, which is in place for most of the data in IDC

Initial version: Nov 2022
Updated: March 2025



Explore

Programs Projects Studies **Data** Publications People Experimental Models Computational Tools Target Enabling Resources Results

DATA (242,360) Hide Filters

[Send 3 Results To CAVATICA](#) Search Download Print

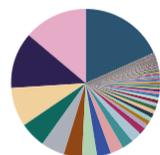
Available Filters

- Analysis Type + **Assay** ✓ Cell Type +
- Chromosome + Consortium + Data Subtype +
- Data Type** ✓ File Format +
- Individual Id Source + Is Model System +
- Is Multi Specimen + Library Prep +
- Metabolite Type + Metadata Type +
- Model System Name + Model System Type +
- Nucleic Acid Source + Organ +
- Resource Type + Sex + Species + **Study** ✓
- Tissue +

Study

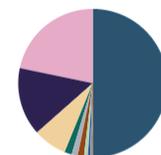
- All 242,360
- ACOM 45
- ACT 72
- AD-BXD 481
- AD_CrossSpecies 20,630
- AD_CrossSpecies 470
- Show all (100)

Study



AMP-AD_DiverseCohorts	44,246
ROSMAP	33,472
WGS_Harmonization	29,432
AD_CrossSpecies	20,630

Data Type



geneExpression	12,100
genomicVariants	52,000
epigenetics	35,000
proteomics	17,000

[View All Charts](#)

		Id	Name	Study	Data Type	Assay
<input checked="" type="checkbox"/>		syn2426151	chr1.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input checked="" type="checkbox"/>		syn2426152	chr2.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input checked="" type="checkbox"/>		syn2426153	chr4.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input type="checkbox"/>		syn2426154	chr8.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input type="checkbox"/>		syn2426155	chr9.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input type="checkbox"/>		syn2426156	chr6.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input type="checkbox"/>		syn2426157	chr5.chop.dosage.gz	ROSMAP	genomicVariants	snpArray
<input type="checkbox"/>		syn2426182	chr3.chop.dosage.gz	ROSMAP	genomicVariants	snpArray

[Clear Selection](#)

3 Rows Selected

[Send To CAVATICA](#)

[Add To Download Cart](#)

**Synapse integration
with the Seven Bridges
Cancer Genomics Cloud**

3 Rows Selected

**EXPORT TO CANCER
GENOMICS CLOUD**

Add To Download Cart

Data Access through SB-CGC

Seven Bridges Cancer Genomics Cloud - Level 1 & 2 Access- Controlled Sequencing data and Open Access Imaging data ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)). Access control for the sequencing data is managed through [dbGaP](https://dbgap.ncbi.nlm.nih.gov/) (Study Accession: [phs002371](https://dbgap.ncbi.nlm.nih.gov/STUDY/study.cgi?study_id=phs002371)).

Direct export from the General Commons (formerly CDS) portal.

The screenshot displays the NIH Cancer Research Data Commons interface. At the top, there is a navigation bar with 'HOME', 'DATA', 'PROGRAMS', 'STUDIES', and 'ABOUT'. A search bar is located in the top right corner. Below the navigation bar, there is a 'Cart > Selected Files' section with a 'README' button. The main content area features a table of files with columns for File Name, Study Name, Accession, Participant Id, Sample Id, Study Access, File Type, and a 'Remove' button. The table lists several files, including sequencing data (BAM) and imaging data (SVS). A callout box highlights the 'AVAILABLE EXPORT OPTIONS' menu, which includes 'EXPORT TO CANCER GENOMICS CLOUD' and 'DOWNLOAD MANIFEST'. A second callout box shows a detailed view of the 'EXPORT TO CANCER GENOMICS CLOUD' option, including a 'DOWNLOAD MANIFEST' button and a 'Remove' button. A third callout box shows the 'Importing DRS data' form, which includes fields for 'Destination project', 'Resolve naming conflicts', and 'Add tags', along with a checkbox for 'I understand that data accessible via DRS...' and a 'Submit' button.

File Name ↑	Study Name	Accession	Participant Id	Sample Id	Study Access	File Type	Remove
HT056P-S1PAA1-possorte_d_genome_bam.bam	Human Tumor Atlas Network (HTAN) primary sequencing data	phs002371	HTA12_1	HTA12_1_1_Tissue, HTA12_1_2_Tissue, HTA12_1_3_Tissue	Controlled	BAM	<input type="checkbox"/>
HT056P-S1PBA3-possorte_d_genome_bam.bam	Human Tumor Atlas Network (HTAN) primary sequencing data	phs002371	HTA12_1	HTA12_1_1_Tissue, HTA12_1_2_Tissue, HTA12_1_3_Tissue	Controlled	BAM	<input type="checkbox"/>
HT056P-S1R1A1-possorte_d_genome_bam.bam	Human Tumor Atlas Network (HTAN) primary sequencing data	phs002371	HTA12_1	HTA12_1_1_Tissue, HTA12_1_2_Tissue, HTA12_1_3_Tissue	Controlled	BAM	<input type="checkbox"/>
HT056P1Psys	Human Tumor Atlas Network (HTAN) imaging data	phs002371_image	HTA12_1	HTA12_1_101_Tissue Biospecimen Type, HTA12_1_102_Tissue Biospecimen Type, HTA12_1_103_Tissue Biospecimen Type, HTA12_1_104_Tissue Biospecimen Type, HTA12_1_105_Tissue Biospecimen Type, HTA12_1_101_Tissue	Open	SVS	<input type="checkbox"/>

https://docs.humantumoratlas.org/open_access/cds_imaging/

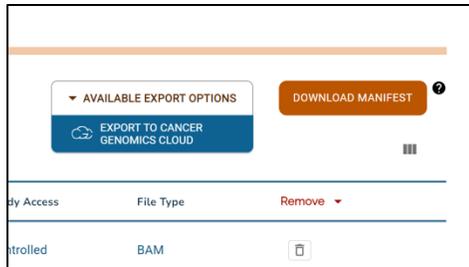
Data Access through SB-CGC

[Seven Bridges Cancer Genomics Cloud](#) - Level 1 & 2 Access- Controlled Sequencing data and Open Access Imaging data ([CC BY 4.0](#)). Access control for the sequencing data is managed through [dbGaP](#) (Study Accession: [phs002371](#)).

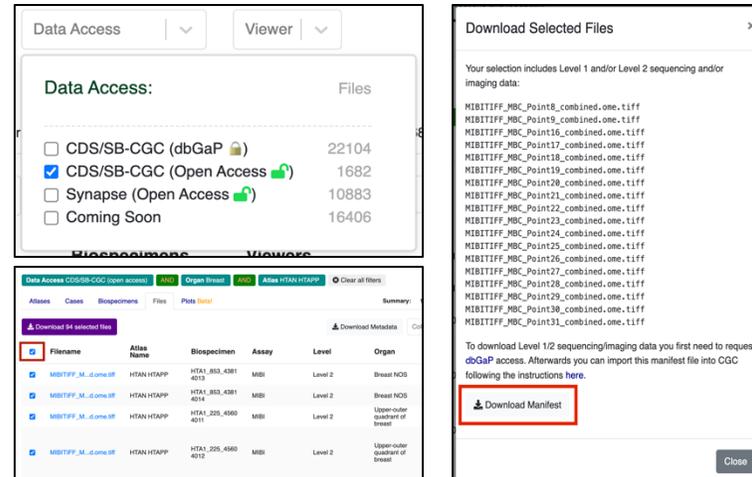
Export via a Data Repository Service (DRS) Manifest.

1. Generate a DRS Manifest

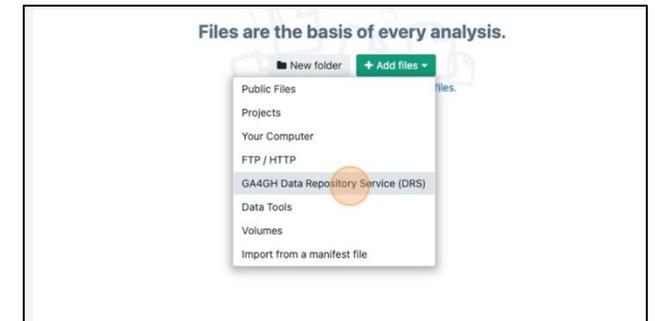
General Commons
(formerly CDS) Portal



HTAN Data Portal



2. Import Manifest into SB-CGC



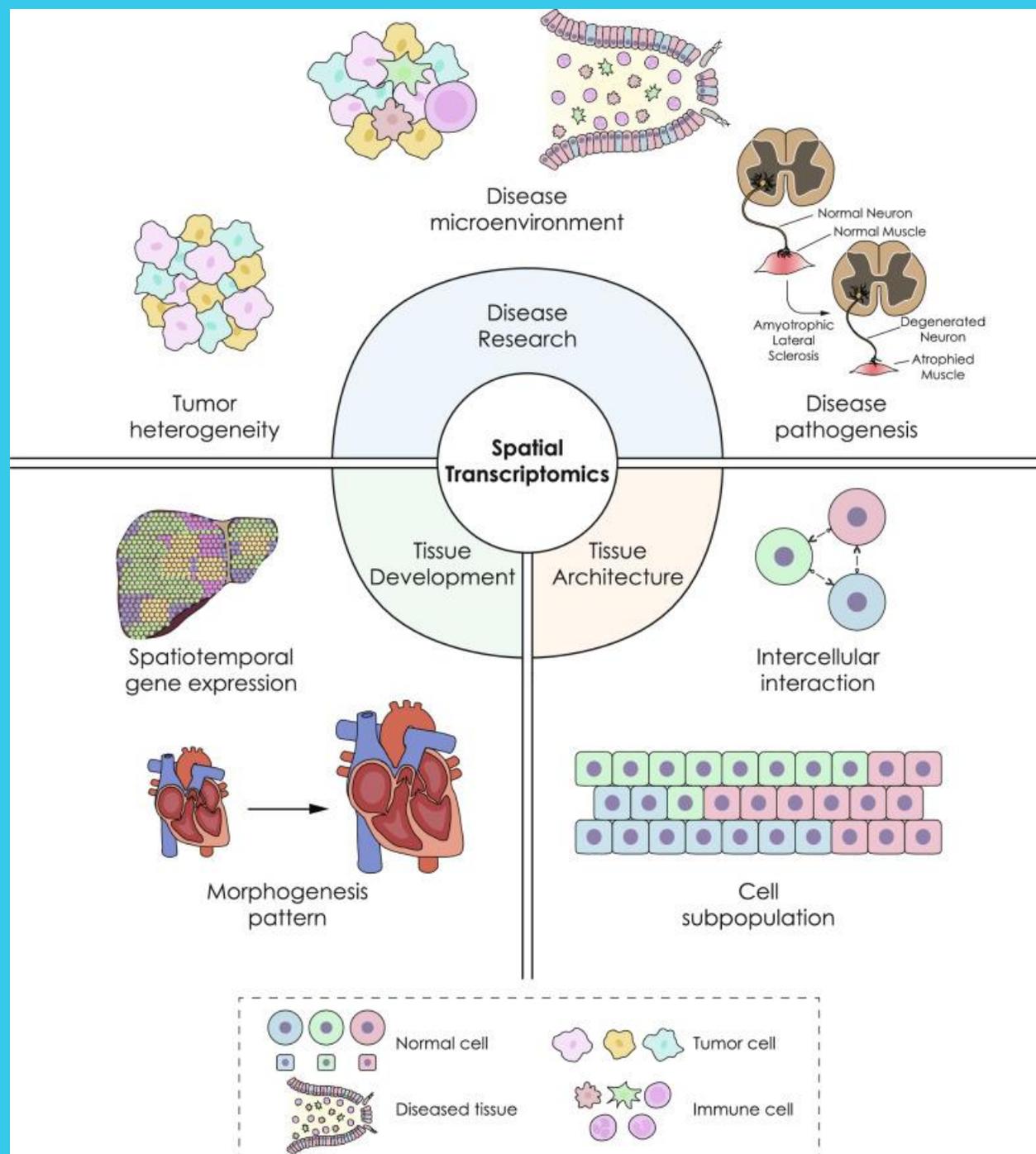
State of the art.

- **134** abstracts in AACR 2023 used spatial transcriptomic analysis
- **>160** abstracts in AACR 2024
- Focus of topics of discussion in sessions and posters
- **25k+** research papers mentioning spatial omics (transcripts, proteomics, metabolomics)

“Single-cell RNA sequencing (scRNA-seq) cannot provide spatial information, while spatial transcriptomics technologies allow gene expression information to be obtained from intact tissue sections in the original physiological context at a spatial resolution.

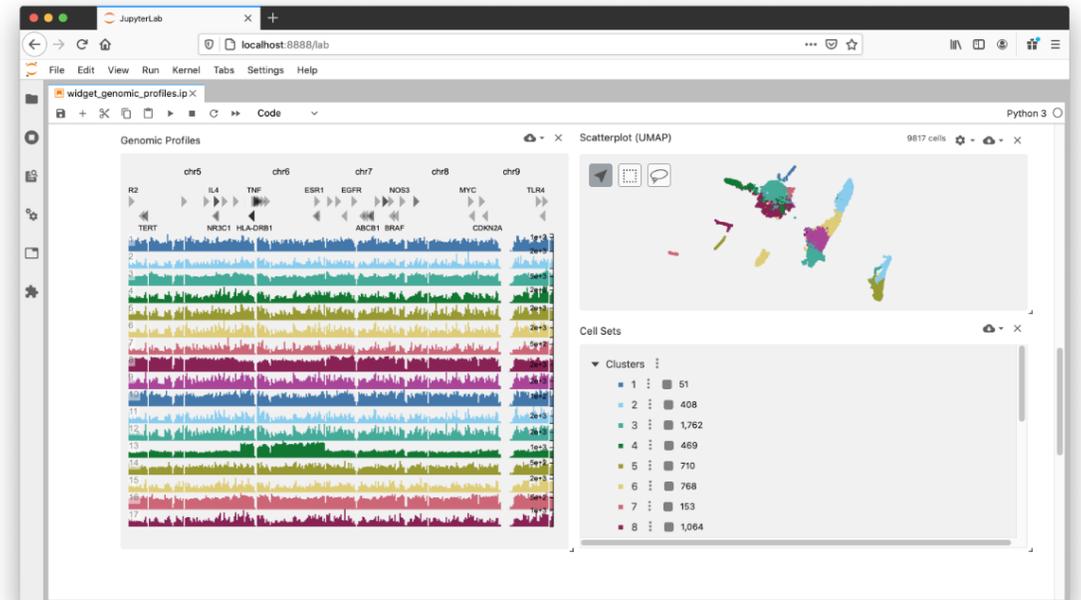
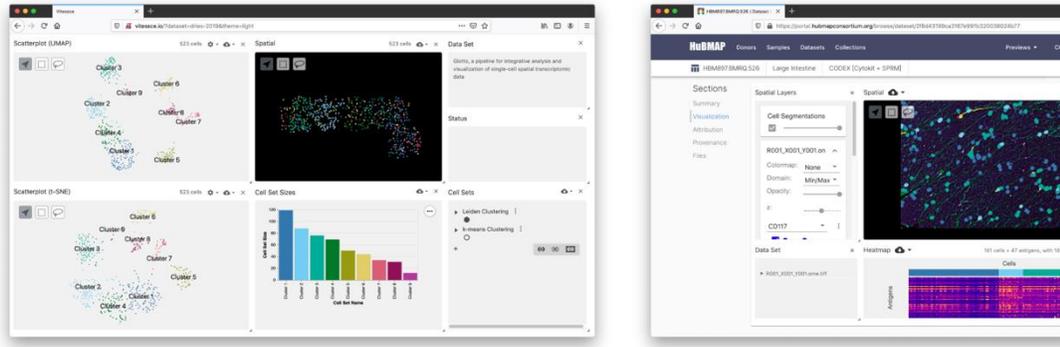
Various biological insights can be generated into tissue architecture and further the elucidation of the interaction between cells and the microenvironment.

Thus, we can gain a general understanding of histogenesis processes and disease pathogenesis, etc.”



What is Vitesse?

A visual integration tool for exploration of spatial single cell experiments - <http://vitesse.io/>



Vitesse as an ipywidget in JupyterLab

Vitessce on the SB-CGC

Public projects you can access to analyse your data

Start your analysis by copying one of our publicly available projects with all of the required resources

Integrative Single-cell Data Visualization with Vitessce: User Guide

Integrative Single-cell Data Visualization with Vitessce: User Guide
This project serves as a comprehensive tutorial for users interested in leveraging Vitessce for the visualization and analysis of single-cell data. It features one Data Studio interactive analysis, written in Python, with step-by-step demonstrations and examples showcasing the integrative capabilities of Vitessce Python API.

Copy project

Cancer Data Aggregator (CDA) with CGC: Usage Guide

Cancer Data Aggregator (CDA) with CGC: Usage Guide
This project contains two data studio analyses for querying and processing data from the CDA database and importing them to your CGC project(s). - The __CDA Release 3 - Import Data to CGC__ analysis demonstrates usage of cda-python release 3 library, SBG python API and

Copy project

MCMICRO - End to End Microscopy Image Processing

MCMICRO - End to End Microscopy Image Processing Public Project
MCMICRO is an end-to-end processing pipeline for multiplexed whole slide imaging and tissue microarrays. It comprises stitching and registration, segmentation, and single-cell feature extraction. Note that this is a CWL wrapper designed by Seven

Copy project

Integrative Single-cell Data Visualization with Vitessce: User Guide

This project serves as a comprehensive tutorial for users interested in leveraging Vitessce for the visualization and analysis of single-cell data. It features one Data Studio interactive analysis, written in Python, with step-by-step demonstrations and examples showcasing the integrative capabilities of Vitessce Python API.

Vitessce is a powerful tool for visualizing and analyzing single-cell data, offering integrative capabilities for exploring multimodal and spatially-resolved datasets. With Vitessce, users can interactively visualize their single-cell data in a spatial context, enabling deeper insights into cellular interactions and spatial relationships.

More information about Vitessce can be found in the [official documentation](#), while the latest `vitessce-python` package, along with additional tutorial notebooks, can be found on the [GitHub page](#).

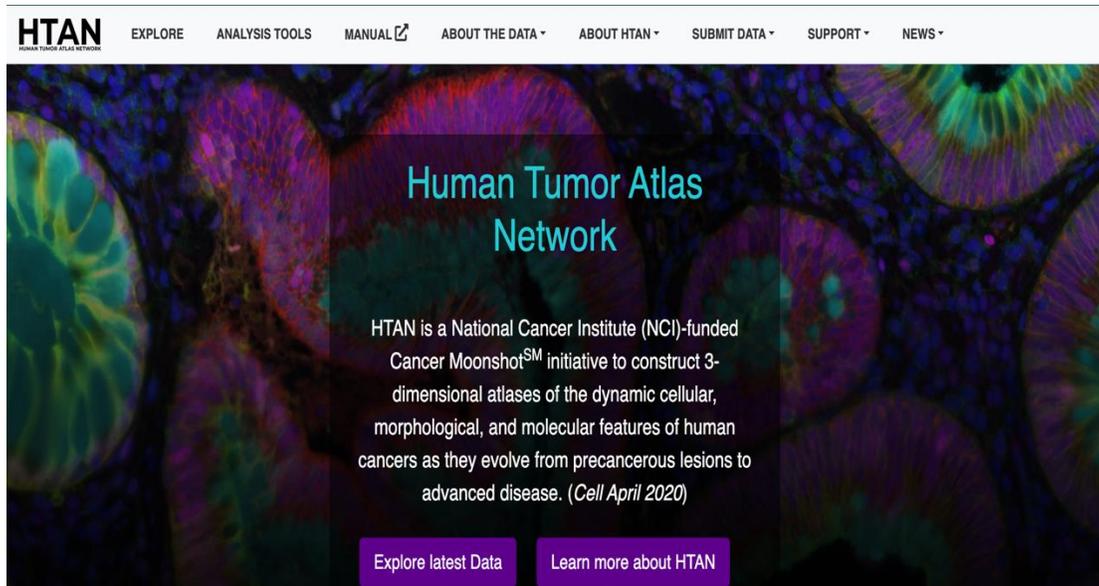
Copying the analysis to your project.

Data Studio analyses are available under the Data Studio tab in this project. To copy the analysis to one of your projects, click the additional actions (three dots) button to the right of **Vitessce Demo Notebook**, click **Copy** and select the target project. You should be able to choose any of the projects you are a part of and in which you have at least *Write* access.

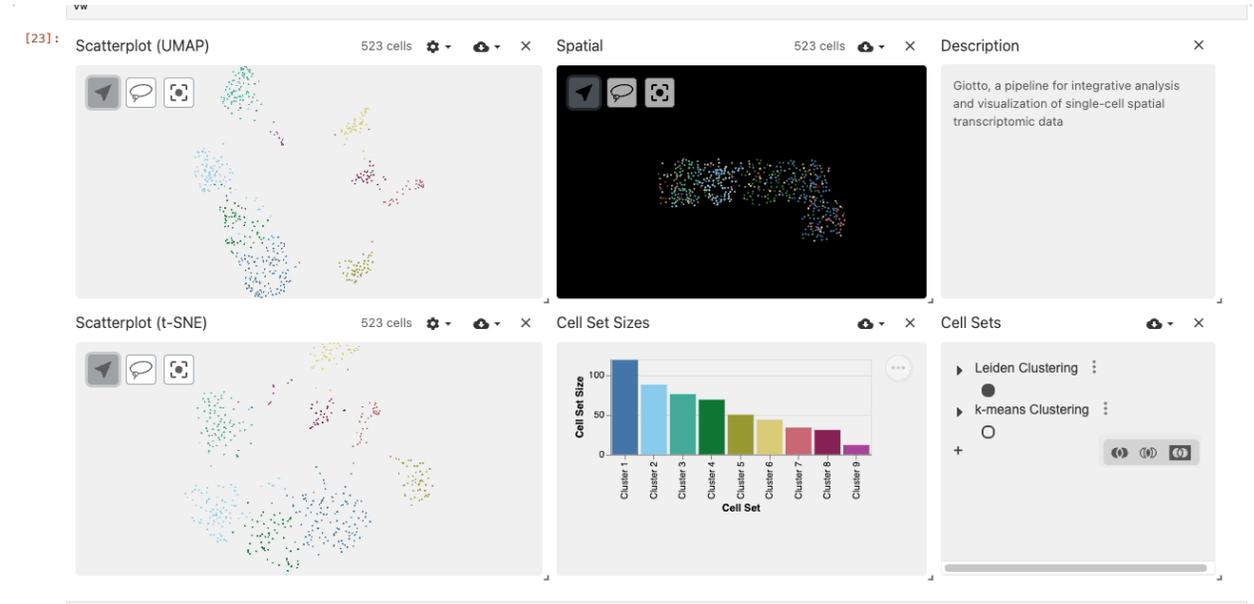
Copying the entire project.

It is also possible to copy the entire project. That way, all Data Studio analyses, as well as Apps, Tasks and Files will be copied. To copy the entire project, click on the information circle next to the public project title (in the top-center of the screen). Then, enter the name for your copy of the project, and select the billing group it will be assigned to.

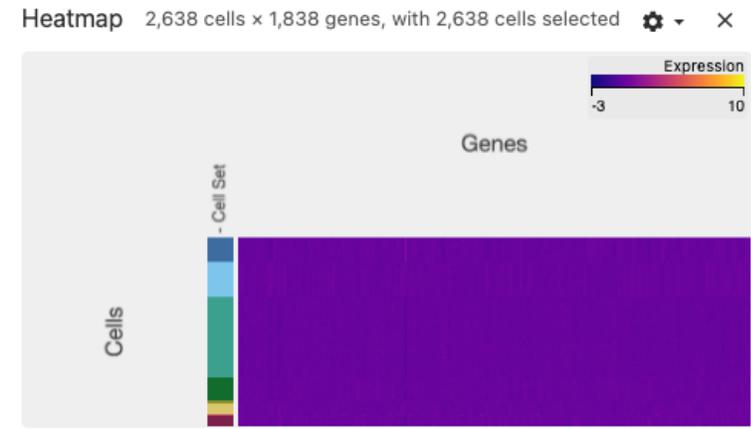
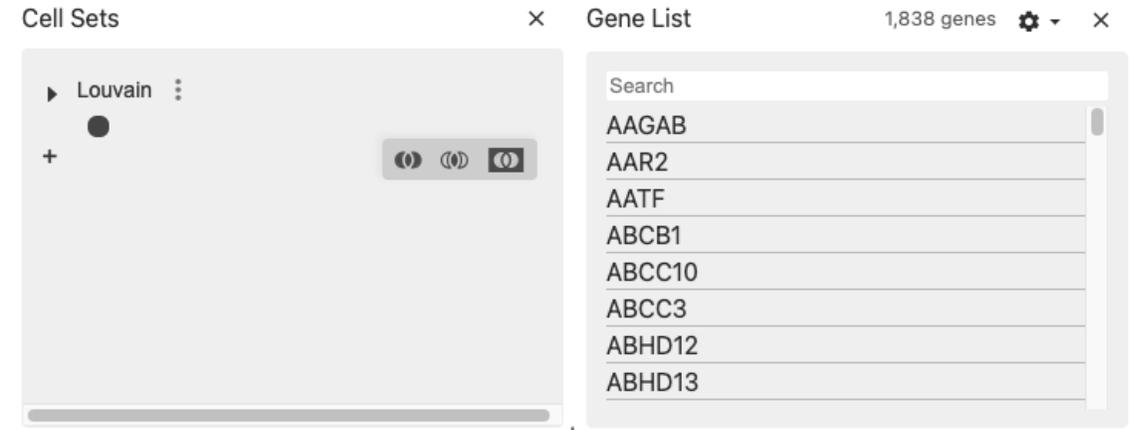
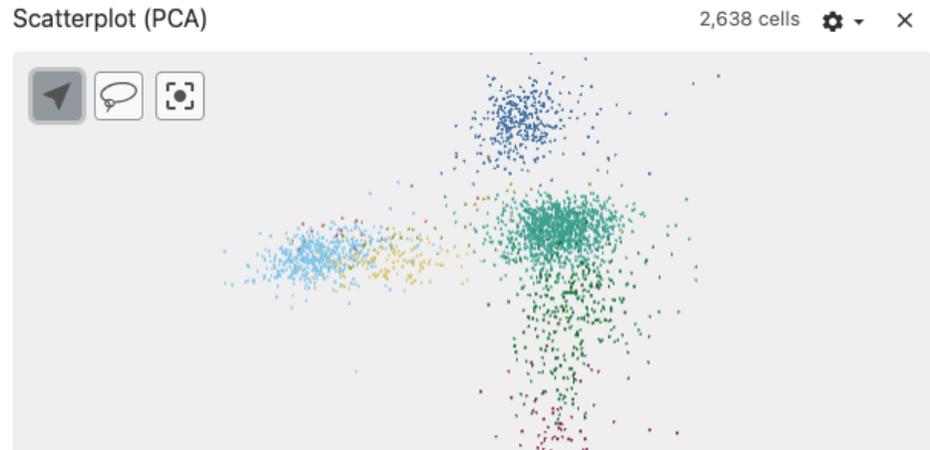
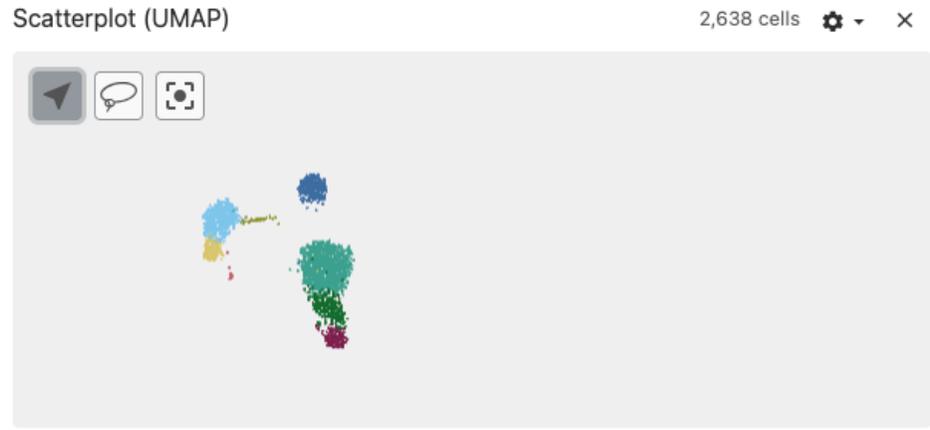
What can we use Vitessece for?



The screenshot shows the HTAN (Human Tumor Atlas Network) website. The header includes navigation links: EXPLORE, ANALYSIS TOOLS, MANUAL, ABOUT THE DATA, ABOUT HTAN, SUBMIT DATA, SUPPORT, and NEWS. The main content area features a large image of a tumor section with the text "Human Tumor Atlas Network" and a description: "HTAN is a National Cancer Institute (NCI)-funded Cancer MoonshotSM initiative to construct 3-dimensional atlases of the dynamic cellular, morphological, and molecular features of human cancers as they evolve from precancerous lesions to advanced disease. (Cell April 2020)". At the bottom, there are two buttons: "Explore latest Data" and "Learn more about HTAN".

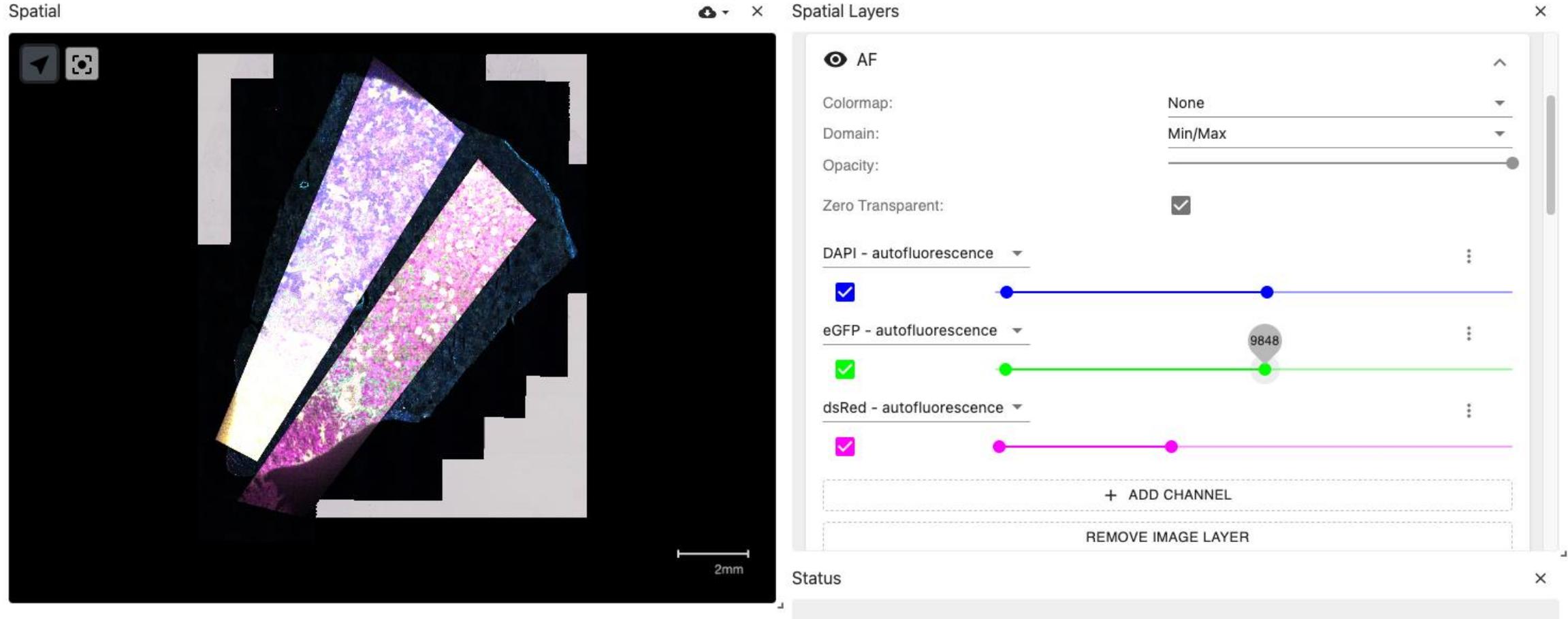


3K PBMC UMAP, PCA, and Scatter Plots



Multimodal Data Images (Spriggins Lab)

Spatial ☁️ × Spatial Layers ×



2mm

AF

Colormap: None

Domain: Min/Max

Opacity:

Zero Transparent:

DAPI - autofluorescence ⋮

eGFP - autofluorescence ⋮

9848

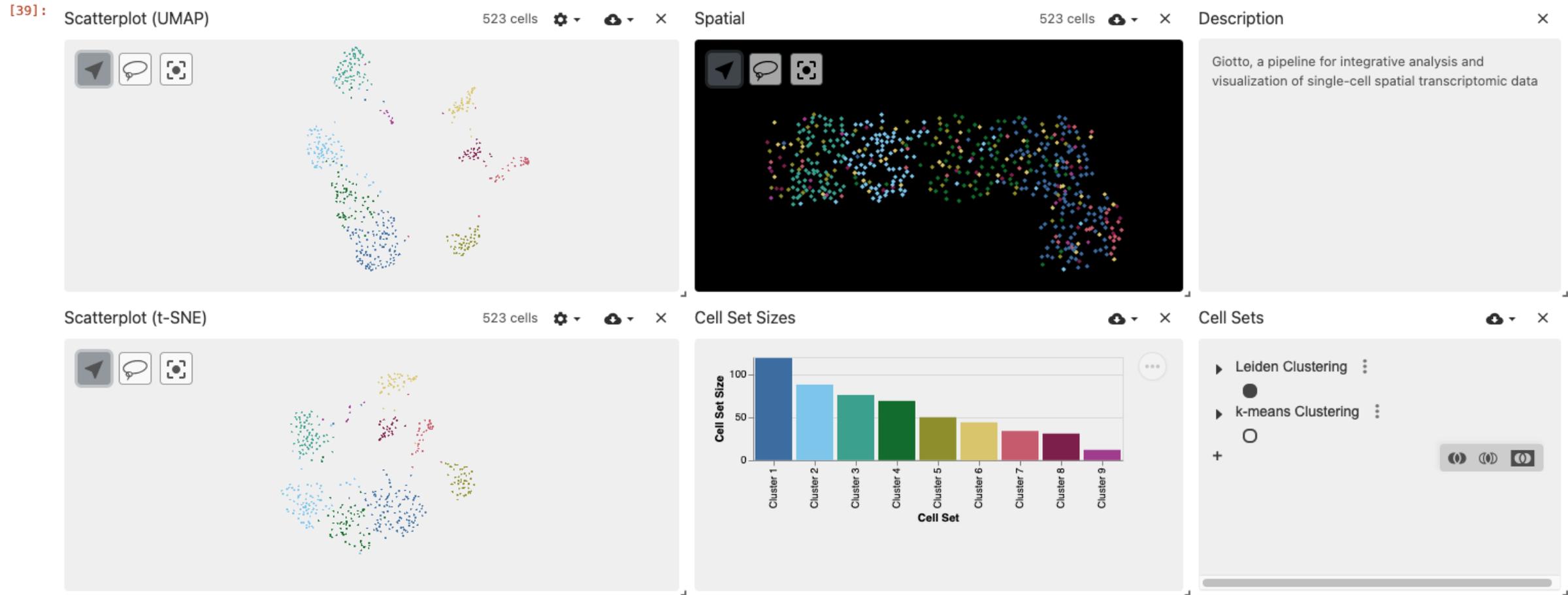
dsRed - autofluorescence ⋮

+ ADD CHANNEL

REMOVE IMAGE LAYER

Status ×

Spatial Transcriptomics



Want to learn more?

Explore our Public Projects

Request access to our MCMICRO on SB-CGC Walkthrough

Contact us:

- **Email:** cgc-SB@velsera.com
- **Website:** CancerGenomicsCloud.org