A clinical data mesh for quality improvement and research in healthcare

Mike Hogarth
MD, Clinical Research Information Officer, UC San Diego Health

Tom Covington
CEO of Tag.bio

Mark Mooney
VP Customer Success, Tag.bio

Campus LISA

Agenda (50 minutes)

UCSD and Tag.bio history	15 min
What is a Data Mesh?	10 min
Nightingale demo	20 min
Q & A	5 min

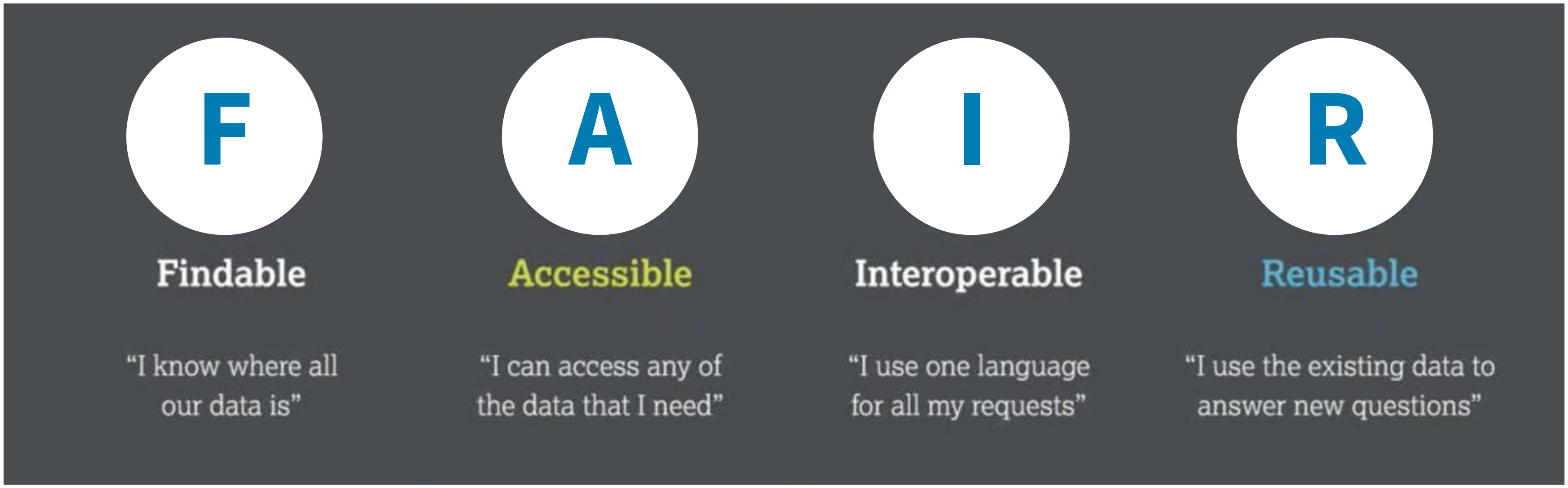
UCSD

Mike Hogarth, MD, Clinical Research Information Officer, UC San Diego Health

The problem (use case) we are trying to solve

- Provide clinical data (protected health information - PHI) to UCSD biomedical researchers:
    - Securely
    - Swiftly
    - Simply
    - Standardized
    - Semantically consistent

Data (and analyses) should be FAIR.



F — **Findable**

"I know where all our data is"

A — **Accessible**

"I can access any of the data that I need"

I — **Interoperable**

"I use one language for all my requests"

R — **Reusable**

"I use the existing data to answer new questions"

Campus
L›SA

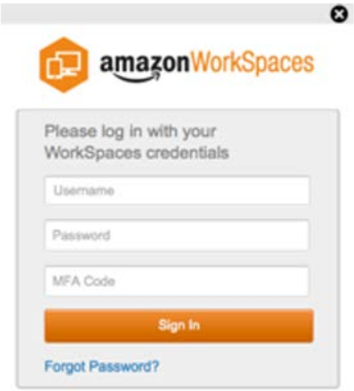
UCSD Health Secure Research Cloud in AWS

iDASH

- What is it?
  - Integrating **D**ata for **A**nalysis, anonymization, and **SH**aring (iDASH) - v2.0
  - A secure computing environment for sensitive data
  - HIPAA compliant
  - Health system CSO approved for use involving protected health information (PHI)
  - Includes virtual research desktops and VMs in a locked-down AWS VPC without the ability for users to connect to the “outside”

- Why we needed it
  - c2017 - EHR data (PHI) for research routinely given to investigators through a download -- some data ended up shared it with external entities without data use agreements, no visibility into the location/use of the data, the ADCS situation along with ~800 other AWS accounts by UCSD Health staff/faculty without visibility or controls

UCSD health secure research cloud with nodes in a VPC



amazonWorkSpaces

Please log in with your WorkSpaces credentials

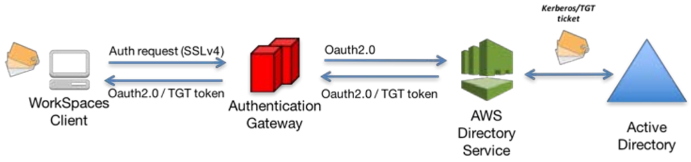
Username

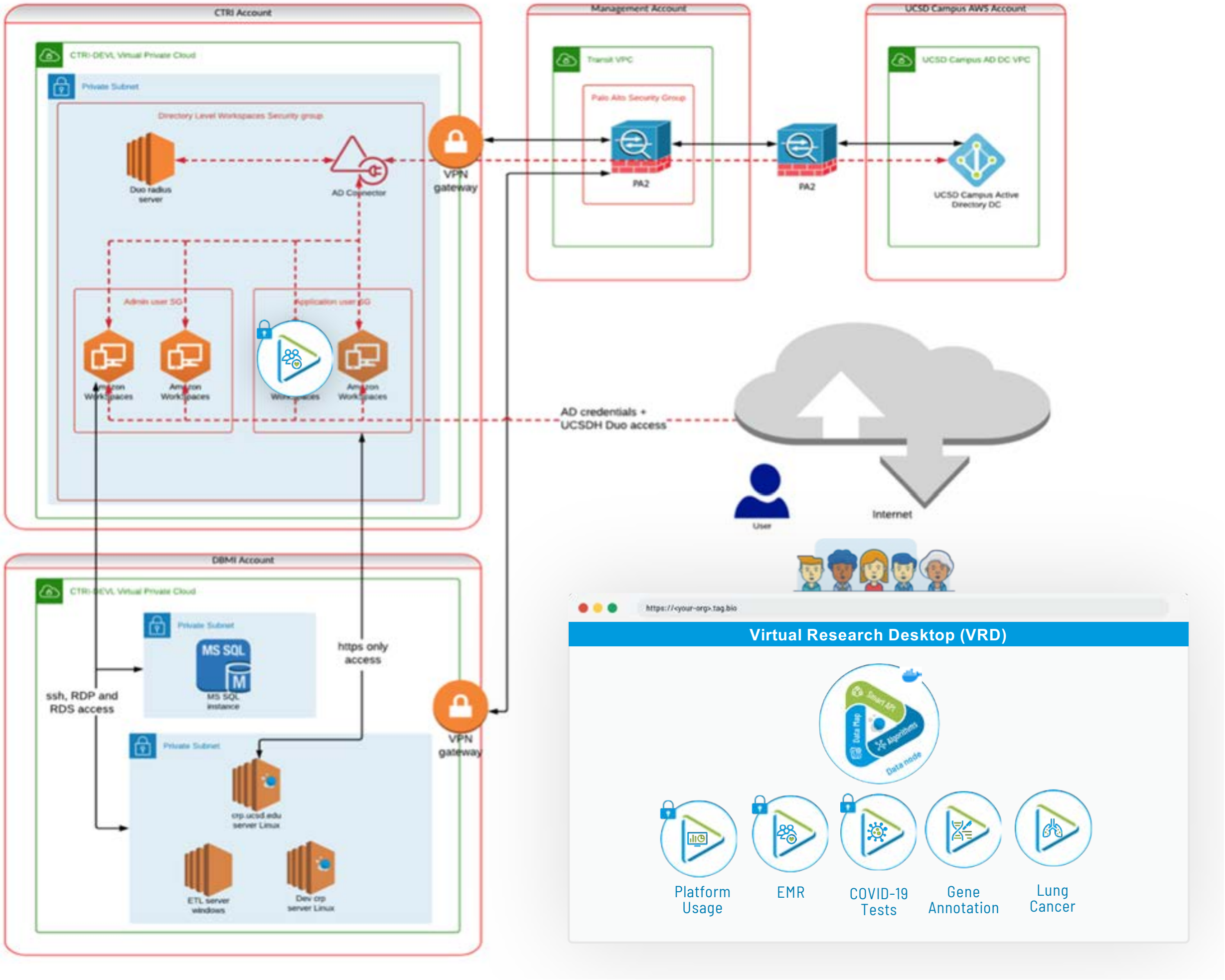
Password

MFA Code

Sign In

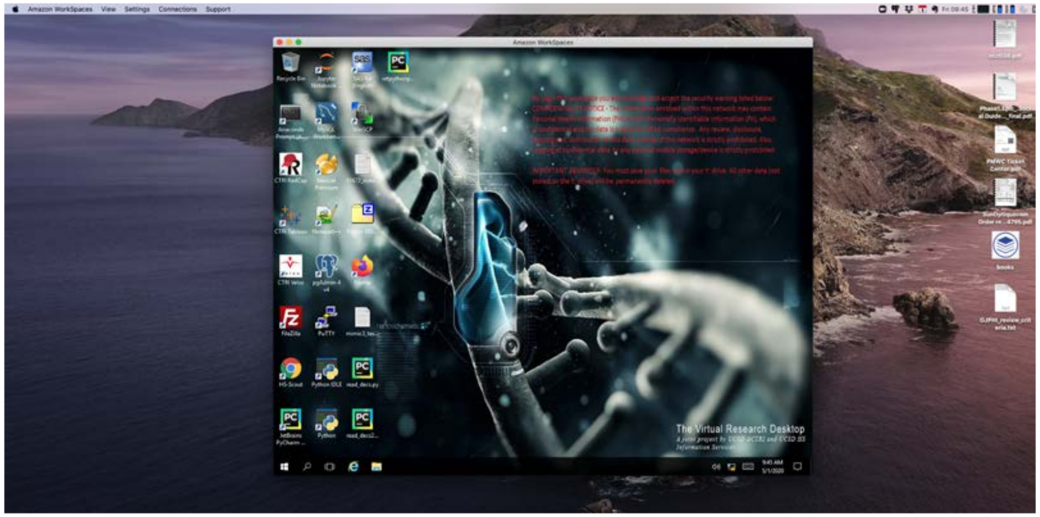
Forgot Password?





THE VIRTUAL RESEARCH DESKTOP (VRD)

- It is a modified version of the Amazon Web Services (AWS) Windows 10 “Workspace” virtual machine
- Runs in the protected UCSDH Secure Cloud in AWS
  - in the AWS HIPAA environment
  - approved by UCSDH CSO for PHI
- Provisioned with:
  - SPSS
  - R/RStudio
  - Python/PyCharm
  - Java 8 JDK
  - Depending on approval, access to internal databases – ie, UC CORDS
  - tag.bio based access to available databases



UCSD Health Virtual Research Desktop

UC San Diego
ALTMAN CLINICAL AND TRANSLATIONAL RESEARCH INSTITUTE

NIGHTINGALE

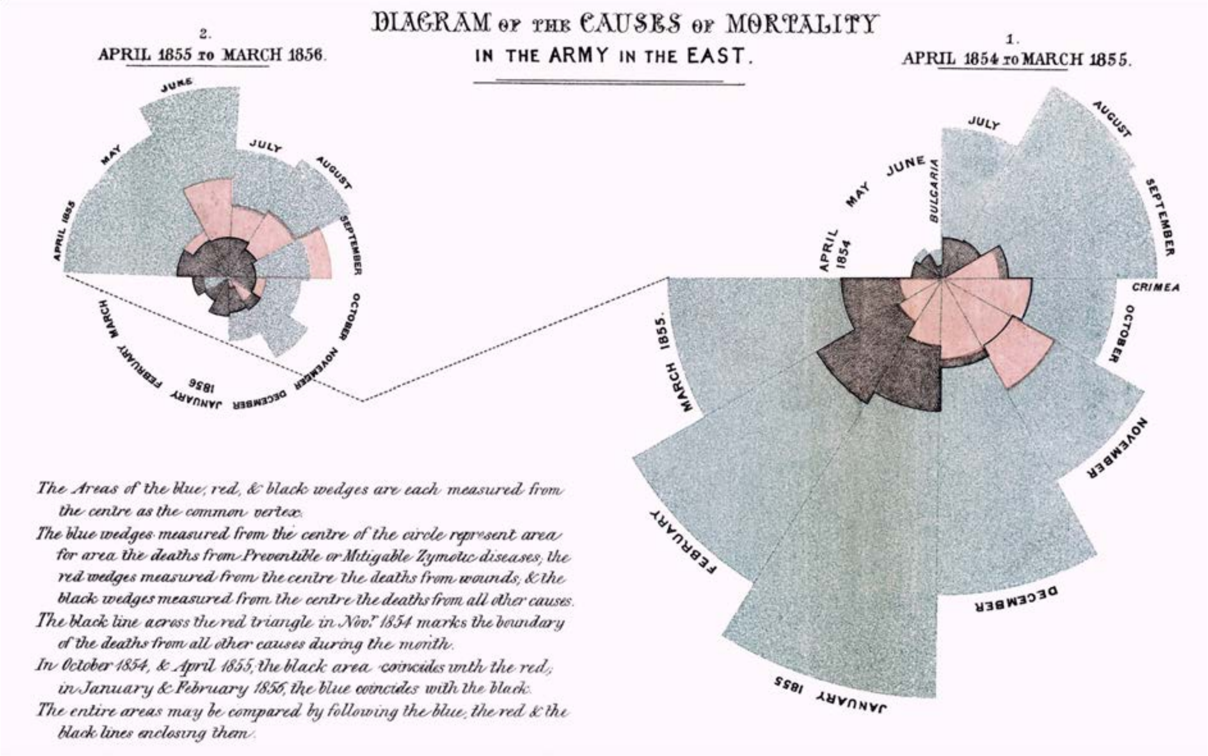


DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Novr 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Florence Nightingale

OM RRC DStJ

Florence Nightingale, c. 1860

Campus L›SA

History

Tom Covington, CEO Tag.bio

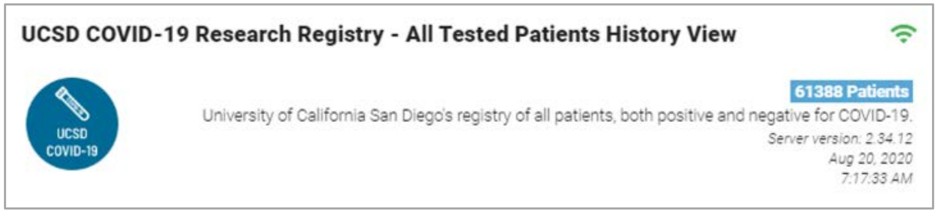
How did the data mesh platform arise?

- Jesse Paquette (CSO Tag.bio) worked at UCSF Helen Diller, Family Comprehensive Cancer Center (2007-2010)
  - Working with Oncology researchers he realized that enabling them to answer their own questions would speed the turnaround time of question to answer
  - Created an initial software called EGAN (Exploratory Gene Association Networks)
  - Formed Tag.bio in 2014 with Tom Covington (CEO) and built first versions of what were then called Flux Capacitors or FC's but became data nodes.
  - Began projects with UCSF Med Center on billing, encounters and claims data in 2018.
  - Realized that the architecture we were working on was an implementation of a "data mesh" after reading Zhamak Dhegani's article: How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh

What’s The History?

- In January 2020 met Tag.bio at Precision Medicine World Conference
  - Initial discussion about work with UCSF on value based healthcare.
  - Set up a visit to UCSD in February
  - Initiated research collaboration with Mike Hogarth in March

- The Pandemic
  - Realized there was an immediate need to make COVID data accessible
  - Built first COVID registries in April



UCSD COVID-19 Research Registry - All Tested Patients History View

61388 Patients

University of California San Diego's registry of all patients, both positive and negative for COVID-19.

Server version: 2.34.12

Aug 20, 2020

7:17:33 AM

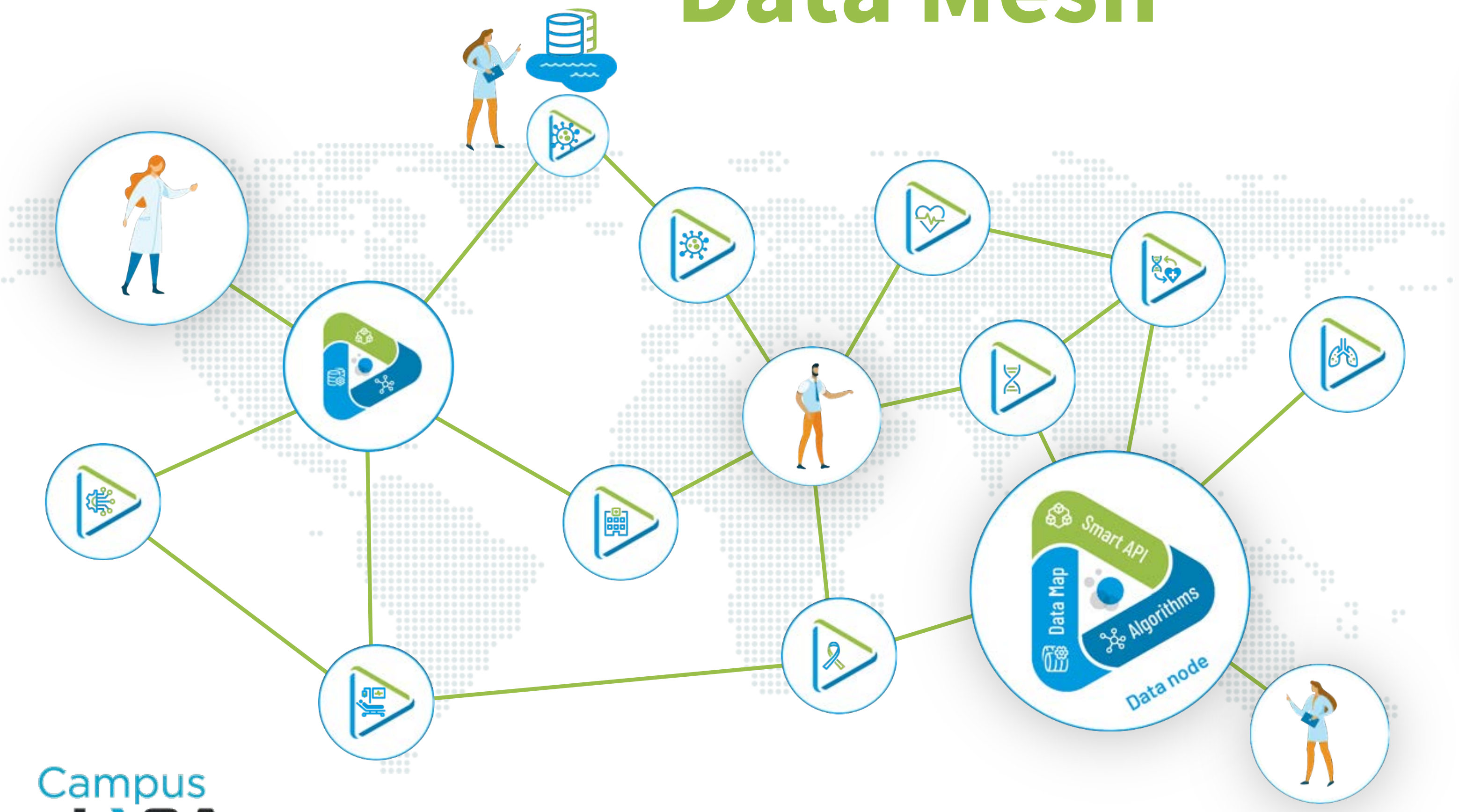
Building a range of patient registries at the present time

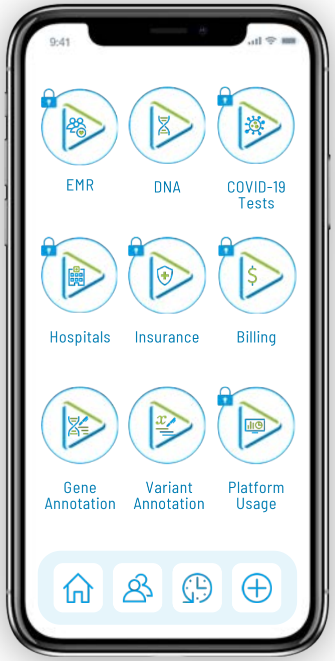
What is a Data mesh?

Mark Mooney, VP of Customer Tag.bio

Data nodes deployed and registered in a decentralized

Data Mesh

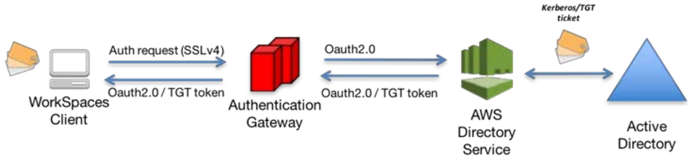


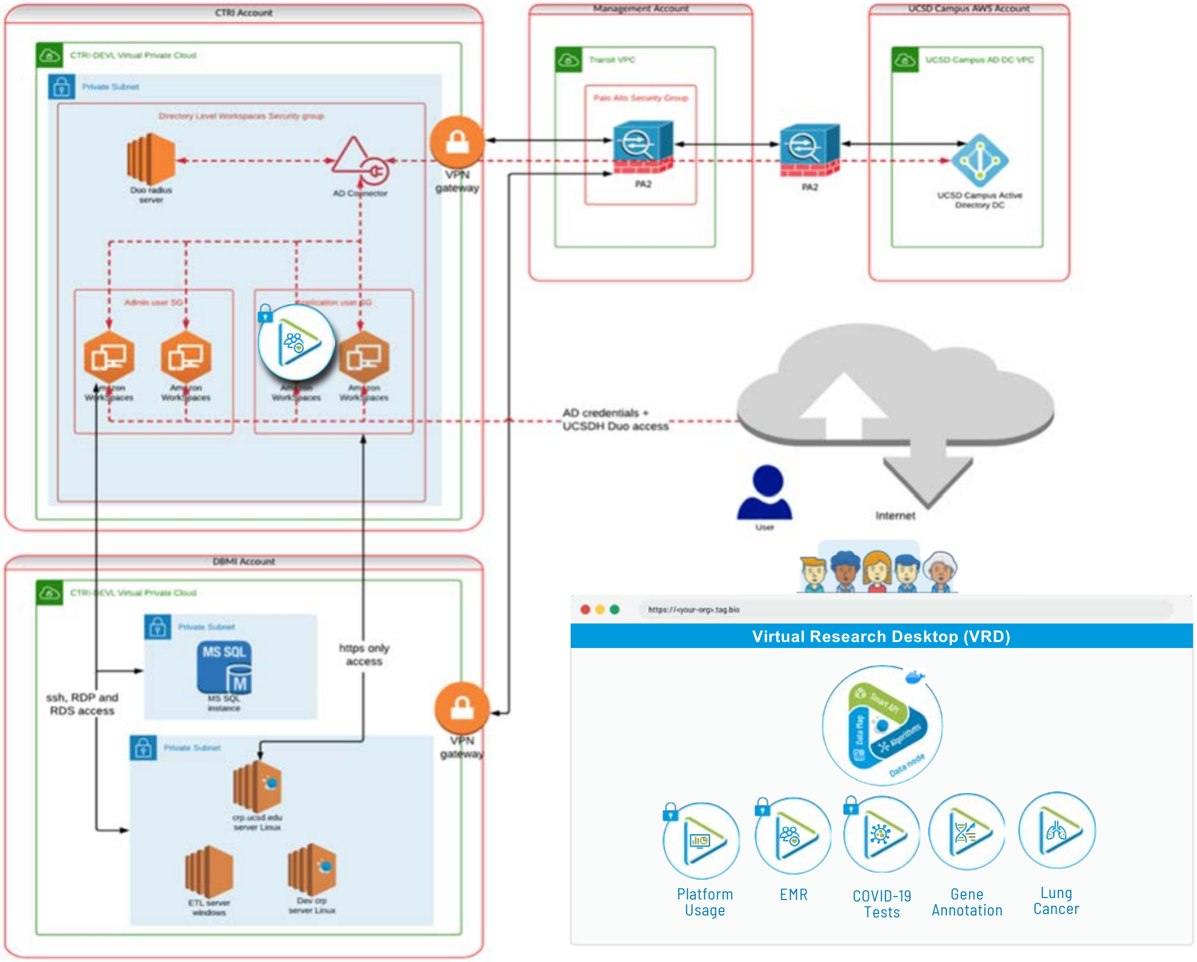


Similar to an app store or a library of data products.

Zhamak Dehghani (Thoughtworks): How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh

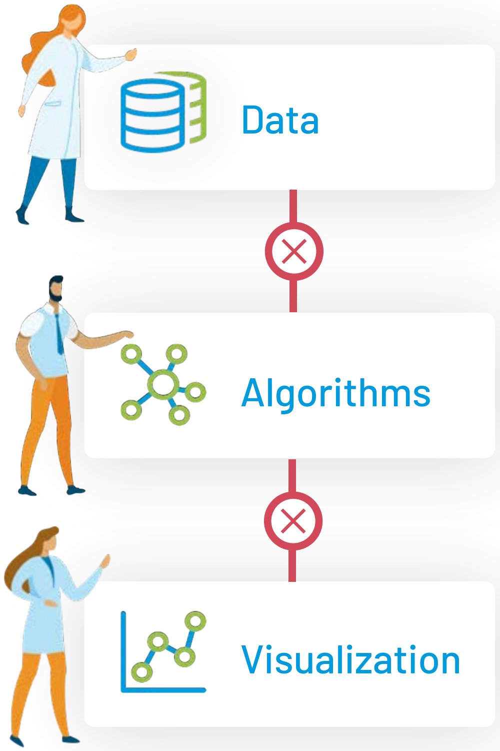
UCSD health secure research cloud with nodes in a VPC



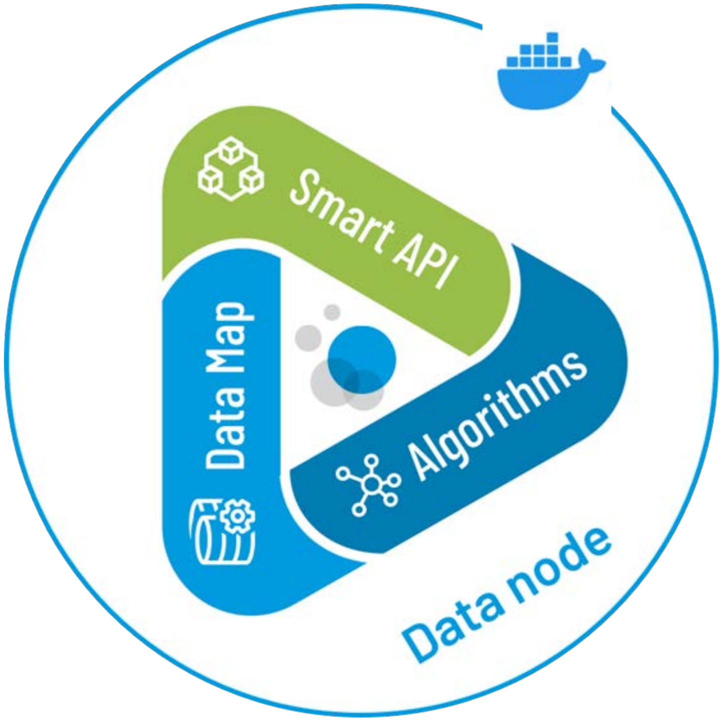


Campus L>SA

Tying 3 technical components...

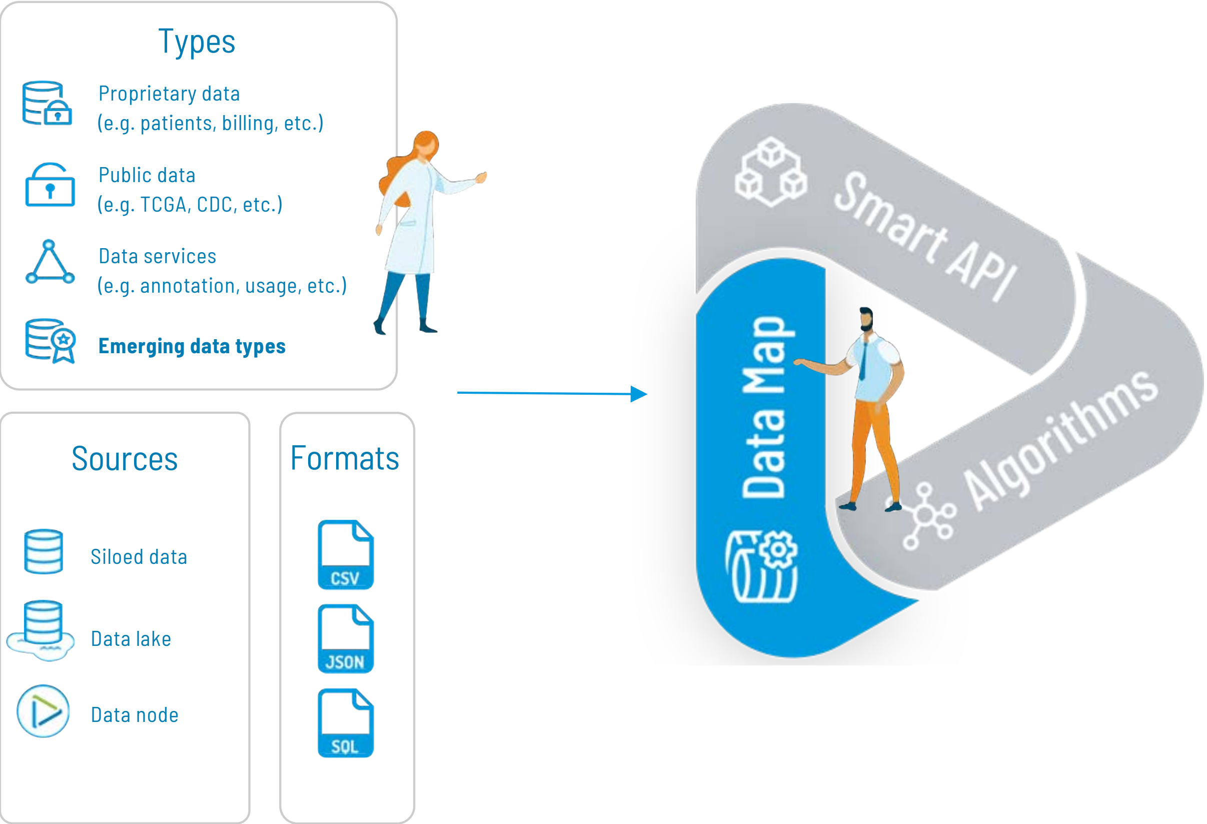


...into a domain-driven data node



Campus LISA

NODE: Data Map



Types

- Proprietary data (e.g. patients, billing, etc.)
- Public data (e.g. TCGA, CDC, etc.)
- Data services (e.g. annotation, usage, etc.)
- **Emerging data types**

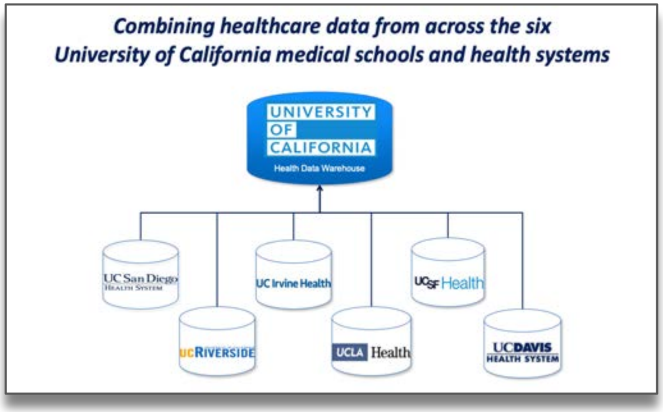
Sources

- Siloed data
- Data lake
- Data node

Formats

- CSV
- JSON
- SQL

UC CORDS COVID-19



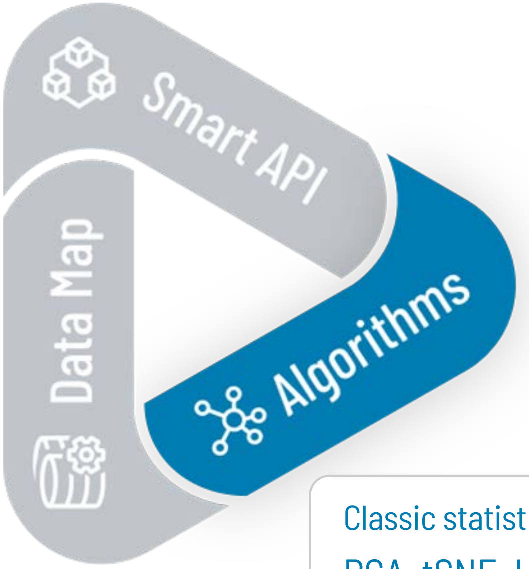
Combining healthcare data from across the six University of California medical schools and health systems

Aug 14 2020

- 175,517 COVID tested patients
- 6,056 COVID+ patients
- all labs, meds, vitals, 29 ICU data elements
- 319,952,837 “data points”

de-identified data
Ingested in 5 hours

NODE: Algorithms



Classic statistical methods

PCA, tSNE, UMAP,
t-test, Mann-Whitney,
Hypergeometric,
Cox regression,
Paired analysis over time,
Pathway analysis, etc.

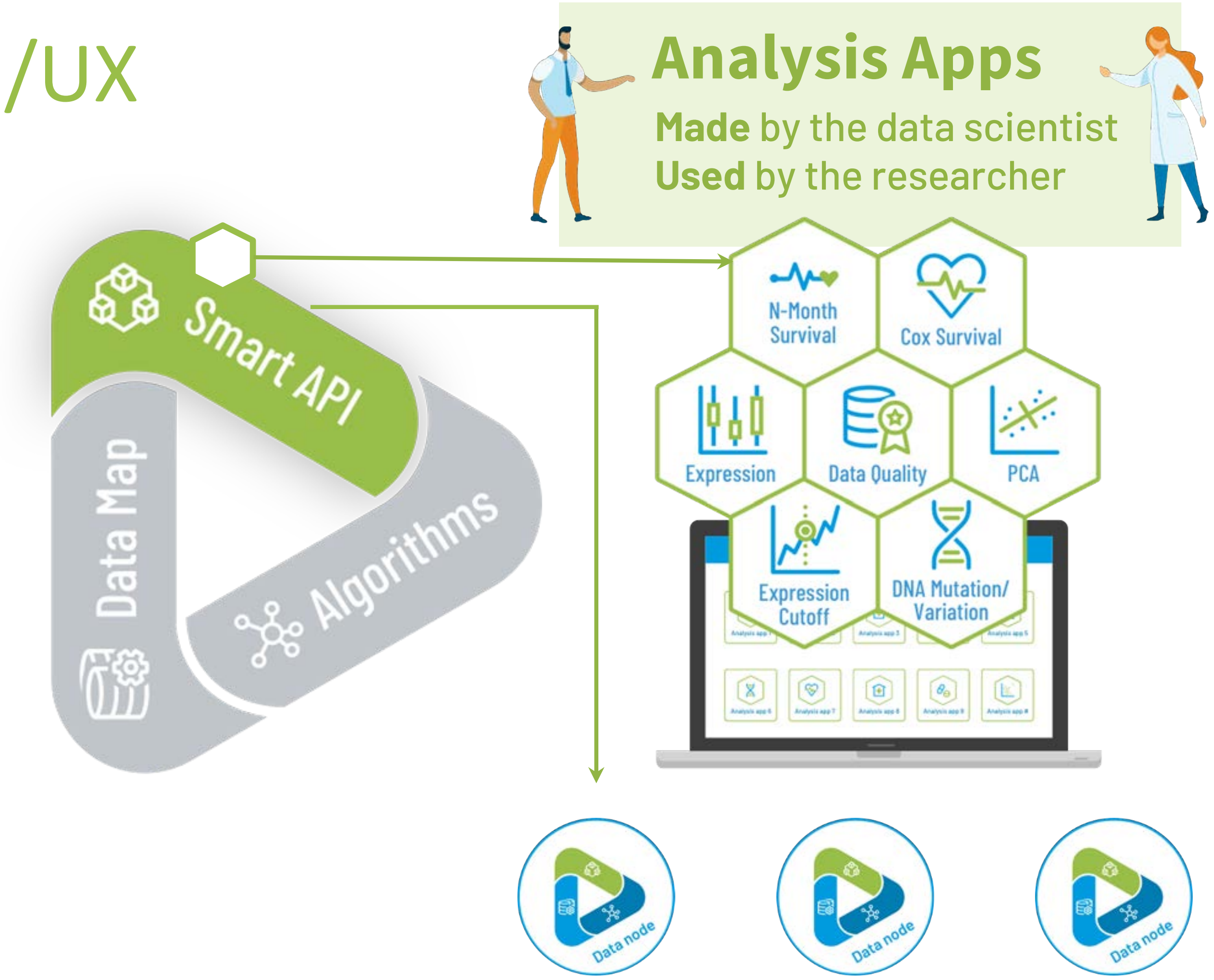
Integration with

R

python

Machine Learning / AI

NODE: Smart API/UX

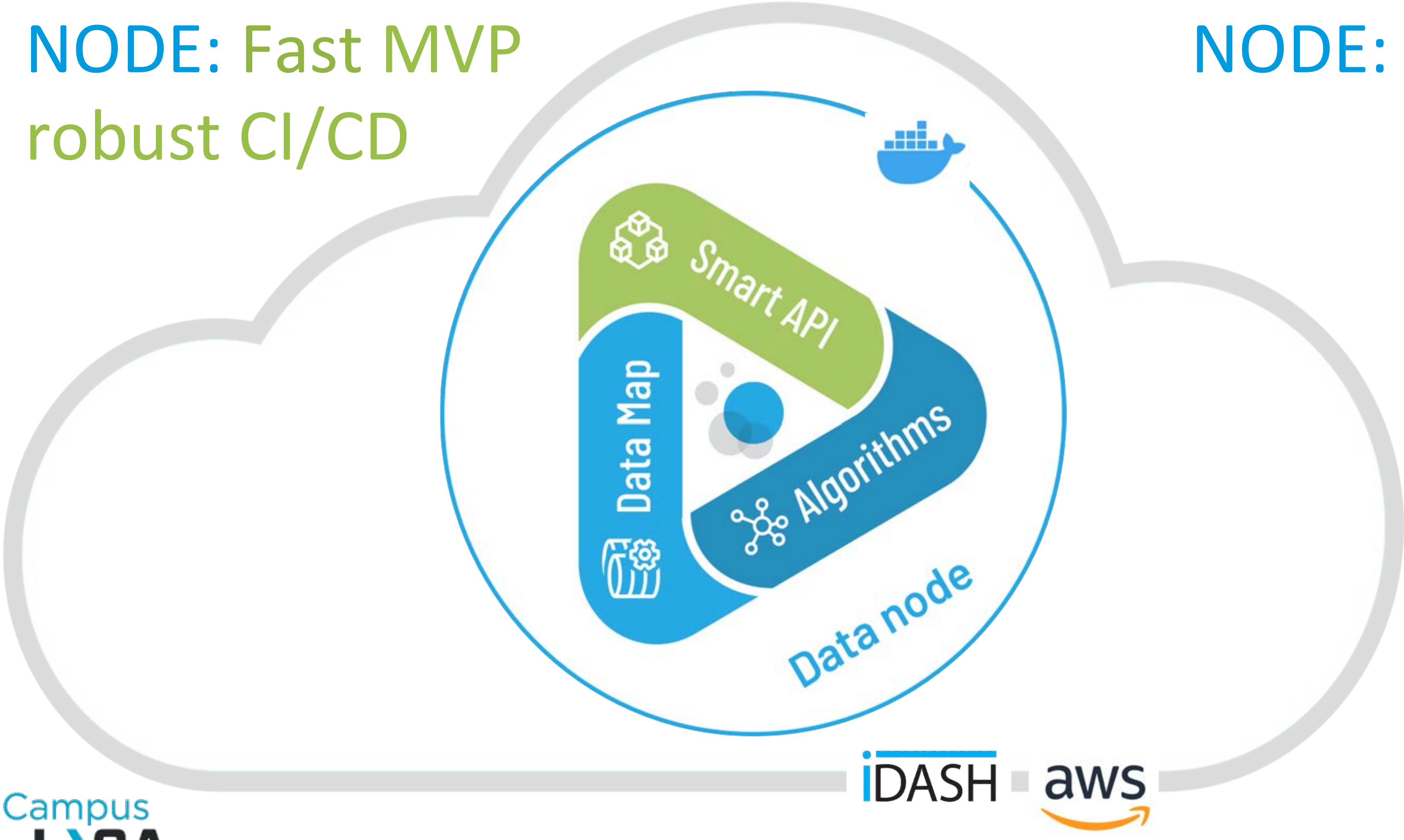


Analysis Apps

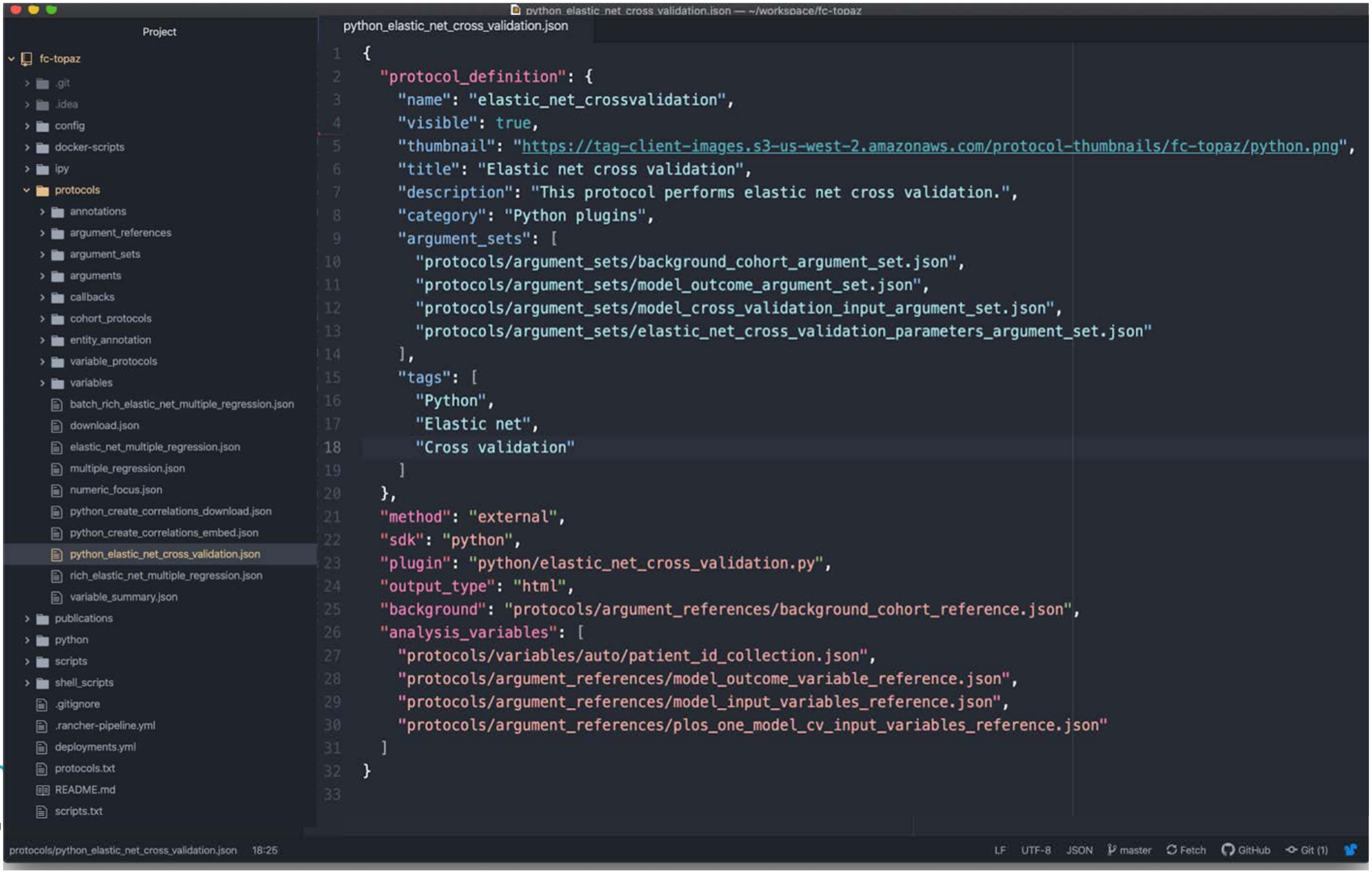
Made by the data scientist
Used by the researcher

NODE: Fast MVP
robust CI/CD

NODE: DIY



Campus
LISA



```json
{
  "protocol_definition": {
    "name": "elastic_net_crossvalidation",
    "visible": true,
    "thumbnail": "https://tag-client-images.s3-us-west-2.amazonaws.com/protocol-thumbnails/fc-topaz/python.png",
    "title": "Elastic net cross validation",
    "description": "This protocol performs elastic net cross validation.",
    "category": "Python plugins",
    "argument_sets": [
      "protocols/argument_sets/background_cohort_argument_set.json",
      "protocols/argument_sets/model_outcome_argument_set.json",
      "protocols/argument_sets/model_cross_validation_input_argument_set.json",
      "protocols/argument_sets/elastic_net_cross_validation_parameters_argument_set.json"
    ],
    "tags": [
      "Python",
      "Elastic net",
      "Cross validation"
    ]
  },
  "method": "external",
  "sdk": "python",
  "plugin": "python/elastic_net_cross_validation.py",
  "output_type": "html",
  "background": "protocols/argument_references/background_cohort_reference.json",
  "analysis_variables": [
    "protocols/variables/auto/patient_id_collection.json",
    "protocols/argument_references/model_outcome_variable_reference.json",
    "protocols/argument_references/model_input_variables_reference.json",
    "protocols/argument_references/plos_one_model_cv_input_variables_reference.json"
  ]
}
```

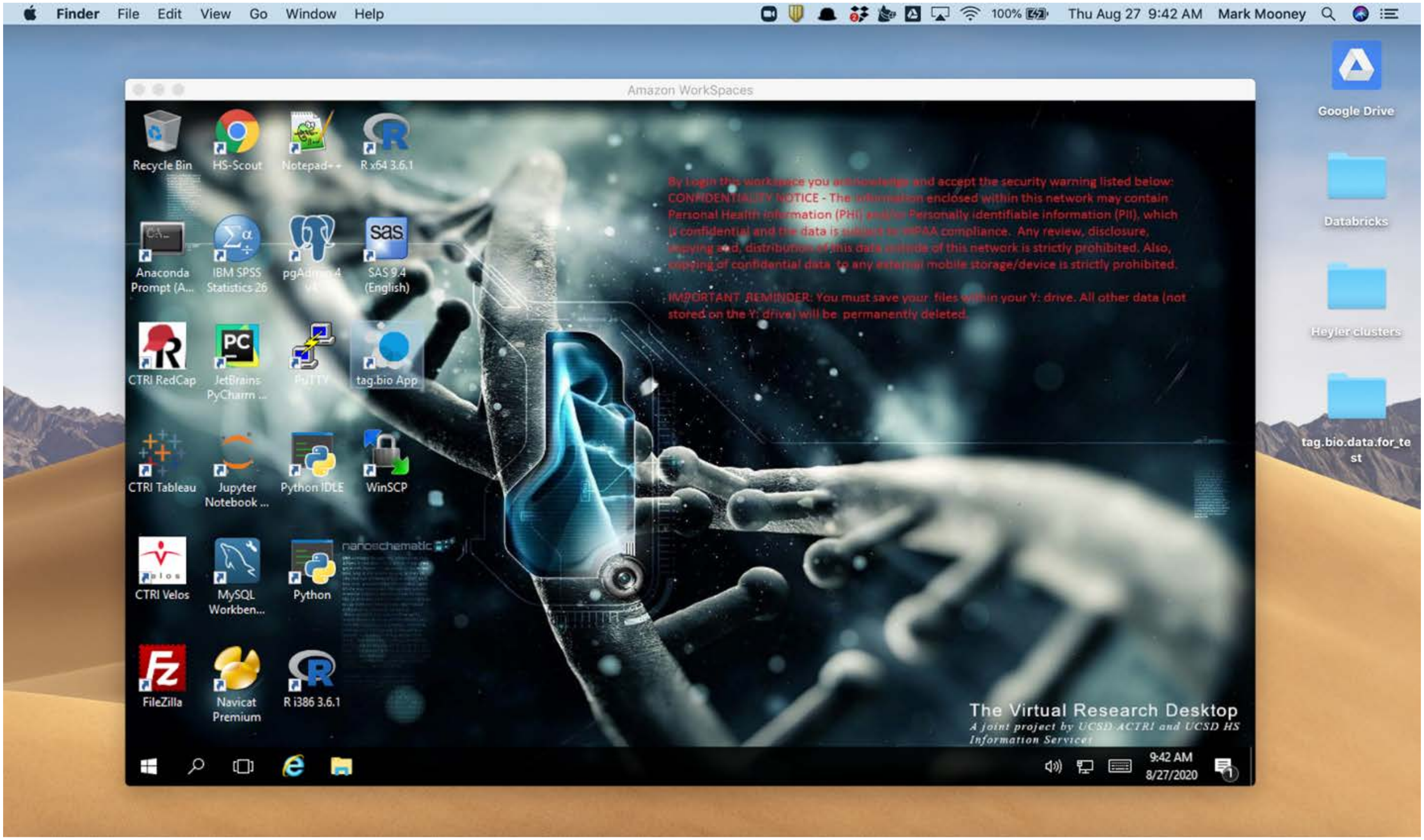
NIGHTINGALE

Nightingale registries

(a de-identified, automated, OMOP data product)

Mark Mooney, VP of Customer Tag.bio

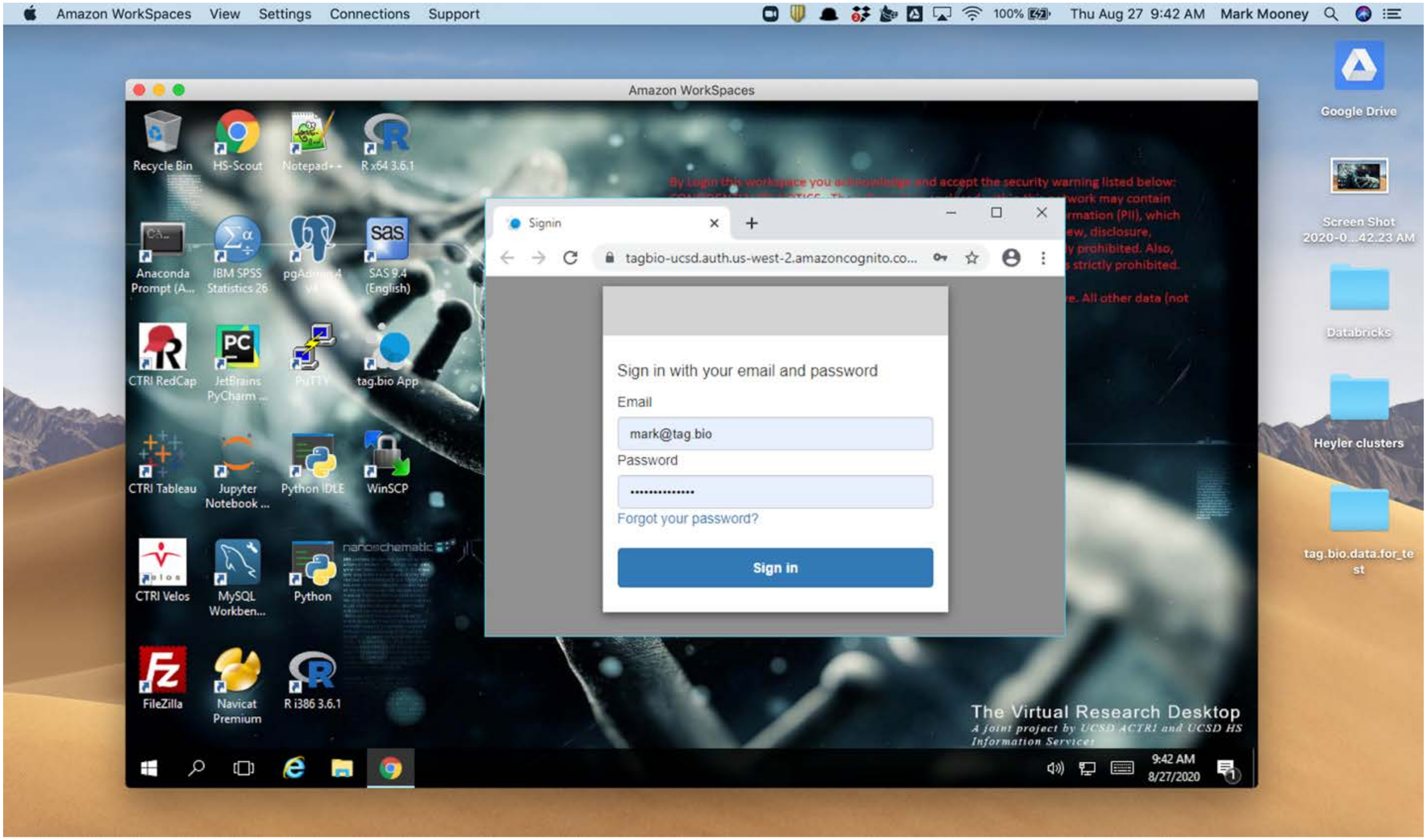
Campus LISA



By Login this workspace you acknowledge and accept the security warning listed below: CONFIDENTIALITY NOTICE - The information enclosed within this network may contain Personal Health information (PHI) and/or Personally identifiable information (PII), which is confidential and the data is subject to HIPAA compliance. Any review, disclosure, copying and, distribution of this data outside of this network is strictly prohibited. Also, copying of confidential data to any external mobile storage/device is strictly prohibited.

IMPORTANT REMINDER: You must save your files within your Y: drive. All other data (not stored on the Y: drive) will be permanently deleted.

The Virtual Research Desktop
A joint project by UCSD ACTRI and UCSD HS Information Services



Sign in with your email and password

Email

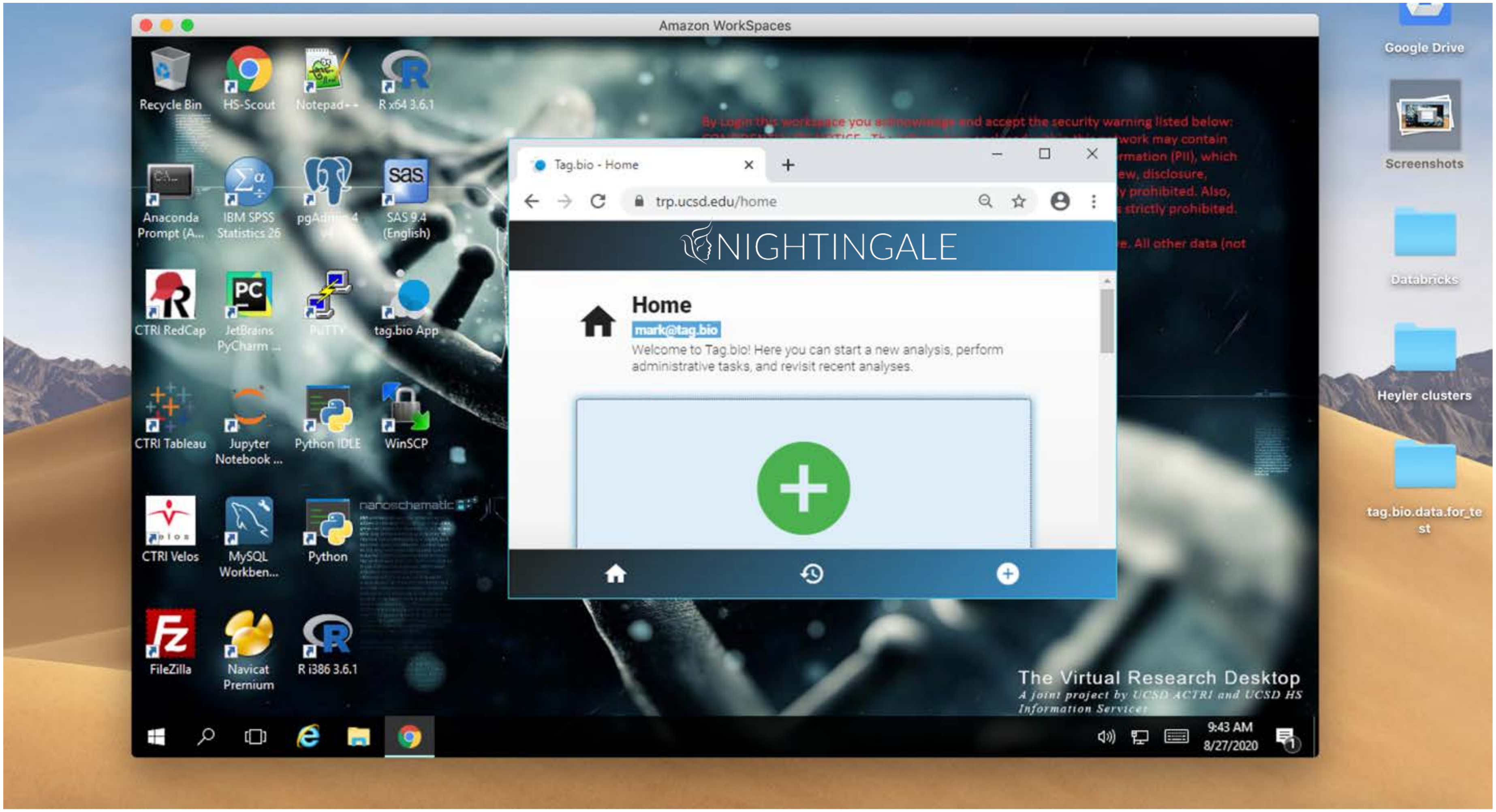
mark@tag.bio

Password

••••••••••••••

Forgot your password?

Sign in



By Login this workspace you acknowledge and accept the security warning listed below: ... work may contain ... rmation (PII), which ... ew, disclosure, ... y prohibited. Also, ... strictly prohibited. ... e. All other data (not

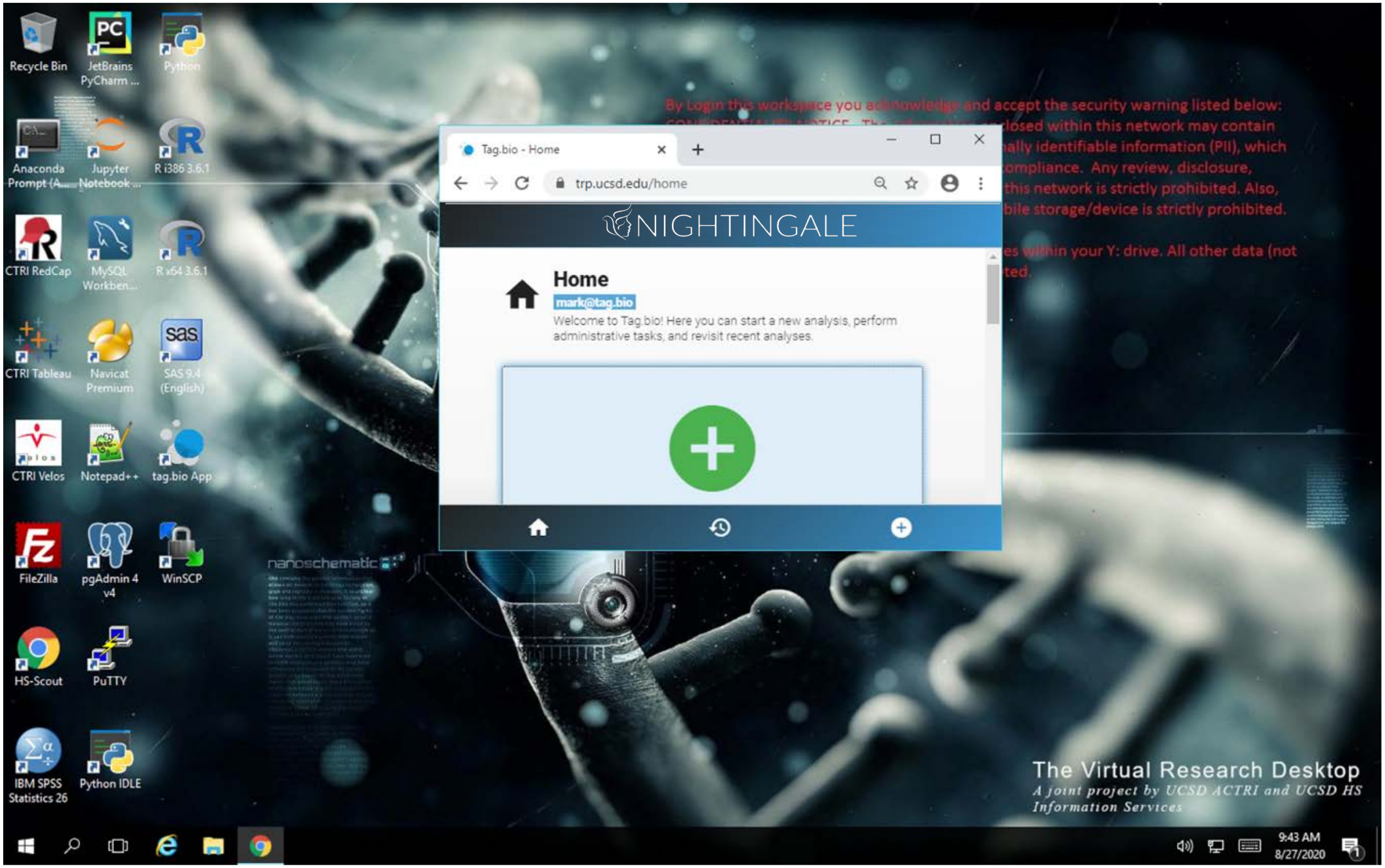
Home

mark@tag.bio

Welcome to Tag.bio! Here you can start a new analysis, perform administrative tasks, and revisit recent analyses.

The Virtual Research Desktop

A joint project by UCSD ACTRI and UCSD HS Information Services



Tag.bio - Home

trp.ucsd.edu/home

NIGHTINGALE

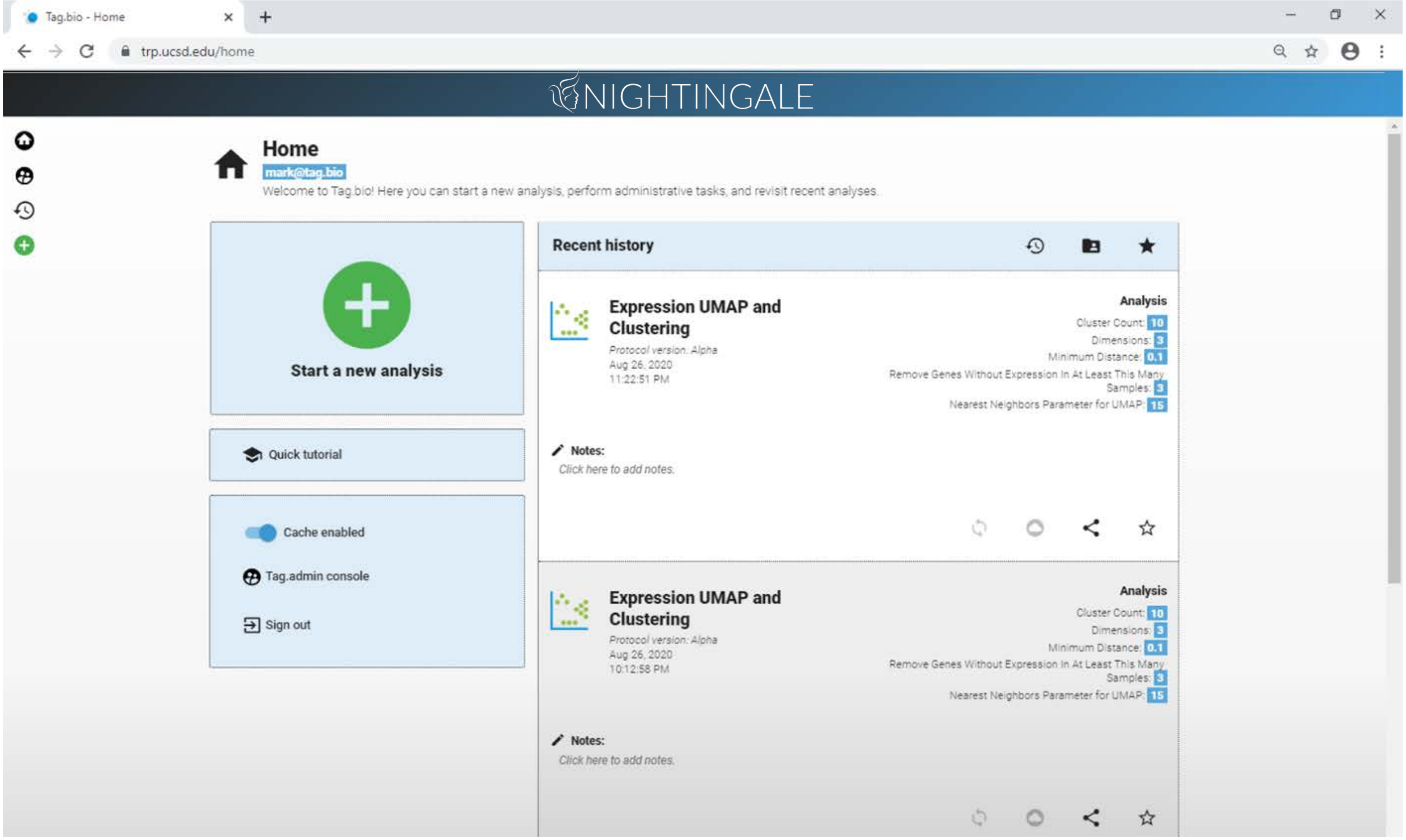
Home

mark@tag.bio

Welcome to Tag.bio! Here you can start a new analysis, perform administrative tasks, and revisit recent analyses.

By Login this workspace you acknowledge and accept the security warning listed below:
CONFIDENTIALITY NOTICE. Th… …closed within this network may contain
…ally identifiable information (PII), which
…ompliance. Any review, disclosure,
…this network is strictly prohibited. Also,
…bile storage/device is strictly prohibited.

…es within your Y: drive. All other data (not
…ted.



Home

mark@tag.bio

Welcome to Tag.bio! Here you can start a new analysis, perform administrative tasks, and revisit recent analyses.

Cache enabled

Expression UMAP and Clustering

Protocol version: Alpha
Aug 26, 2020
11:22:51 PM

Analysis

Cluster Count: 10
Dimensions: 3
Minimum Distance: 0.1
Remove Genes Without Expression In At Least This Many Samples: 3
Nearest Neighbors Parameter for UMAP: 15

Notes:
Click here to add notes.

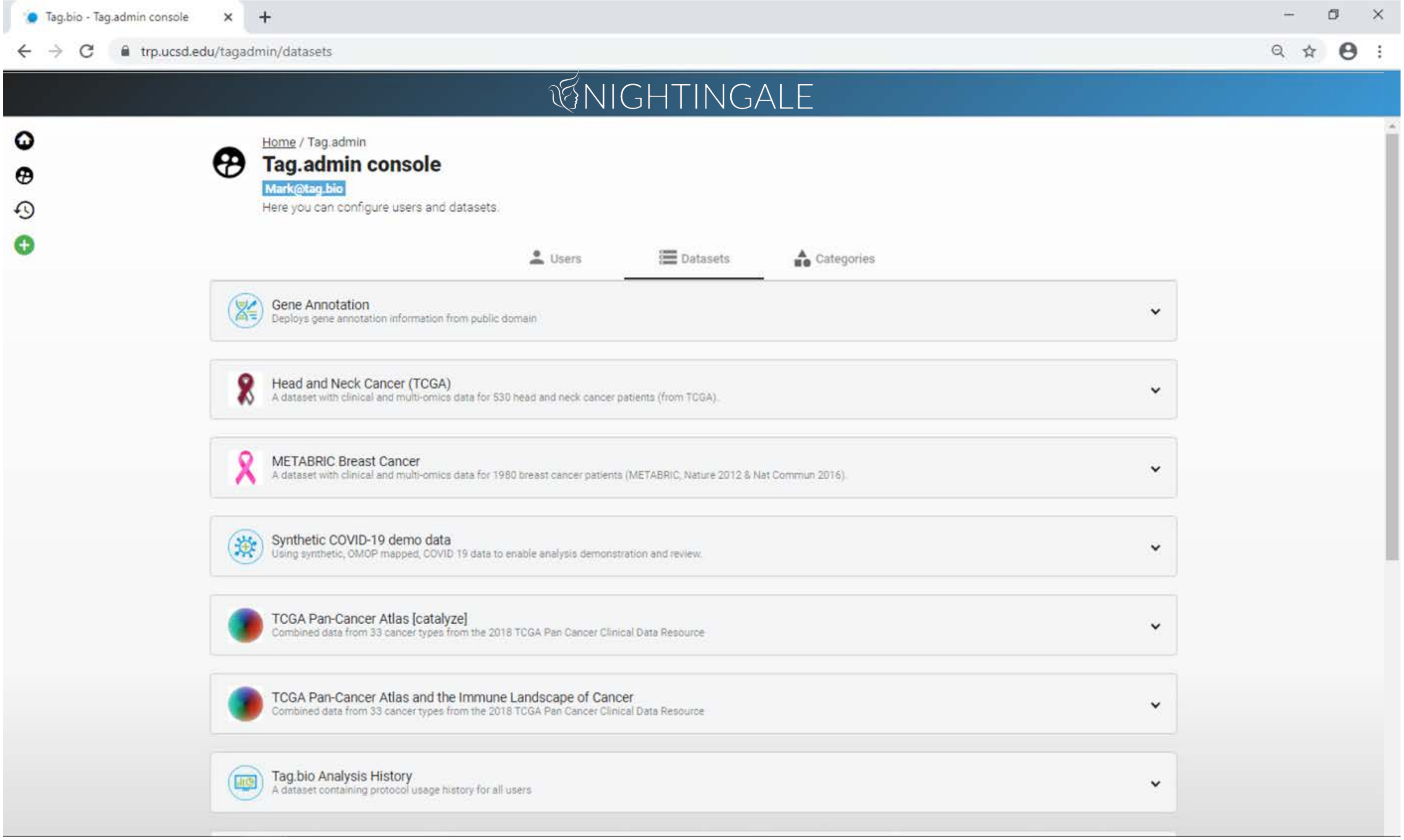
Expression UMAP and Clustering

Protocol version: Alpha
Aug 26, 2020
10:12:58 PM

Analysis

Cluster Count: 10
Dimensions: 3
Minimum Distance: 0.1
Remove Genes Without Expression In At Least This Many Samples: 3
Nearest Neighbors Parameter for UMAP: 15

Notes:
Click here to add notes.



Home / Tag.admin

Tag.admin console

Mark@tag.bio

Here you can configure users and datasets.

Gene Annotation
Deploys gene annotation information from public domain

Head and Neck Cancer (TCGA)
A dataset with clinical and multi-omics data for 530 head and neck cancer patients (from TCGA).

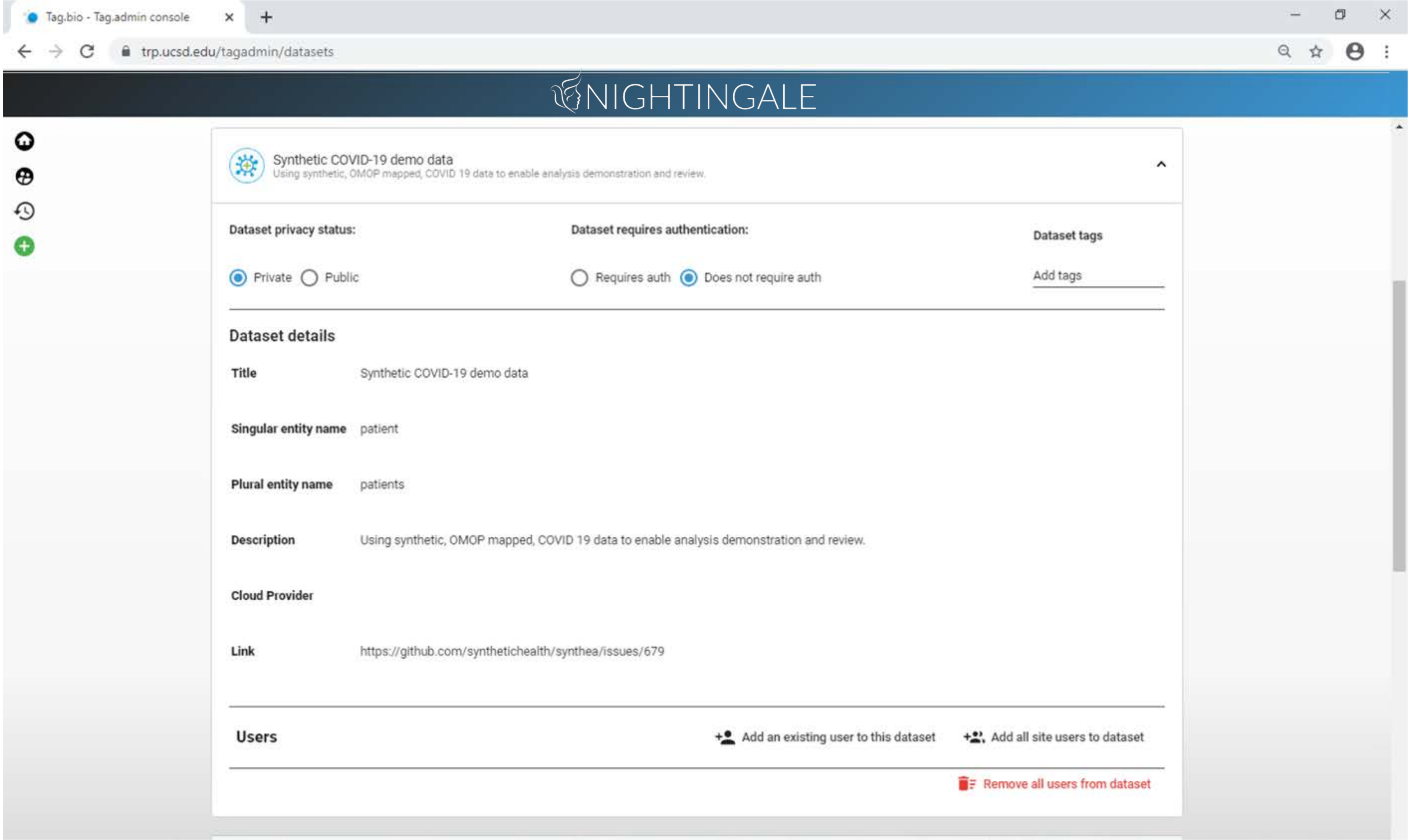
METABRIC Breast Cancer
A dataset with clinical and multi-omics data for 1980 breast cancer patients (METABRIC, Nature 2012 & Nat Commun 2016).

Synthetic COVID-19 demo data
Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

TCGA Pan-Cancer Atlas [catalyze]
Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer
Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

Tag.bio Analysis History
A dataset containing protocol usage history for all users



Synthetic COVID-19 demo data
Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

Dataset privacy status: ● Private ○ Public

Dataset requires authentication: ○ Requires auth ● Does not require auth

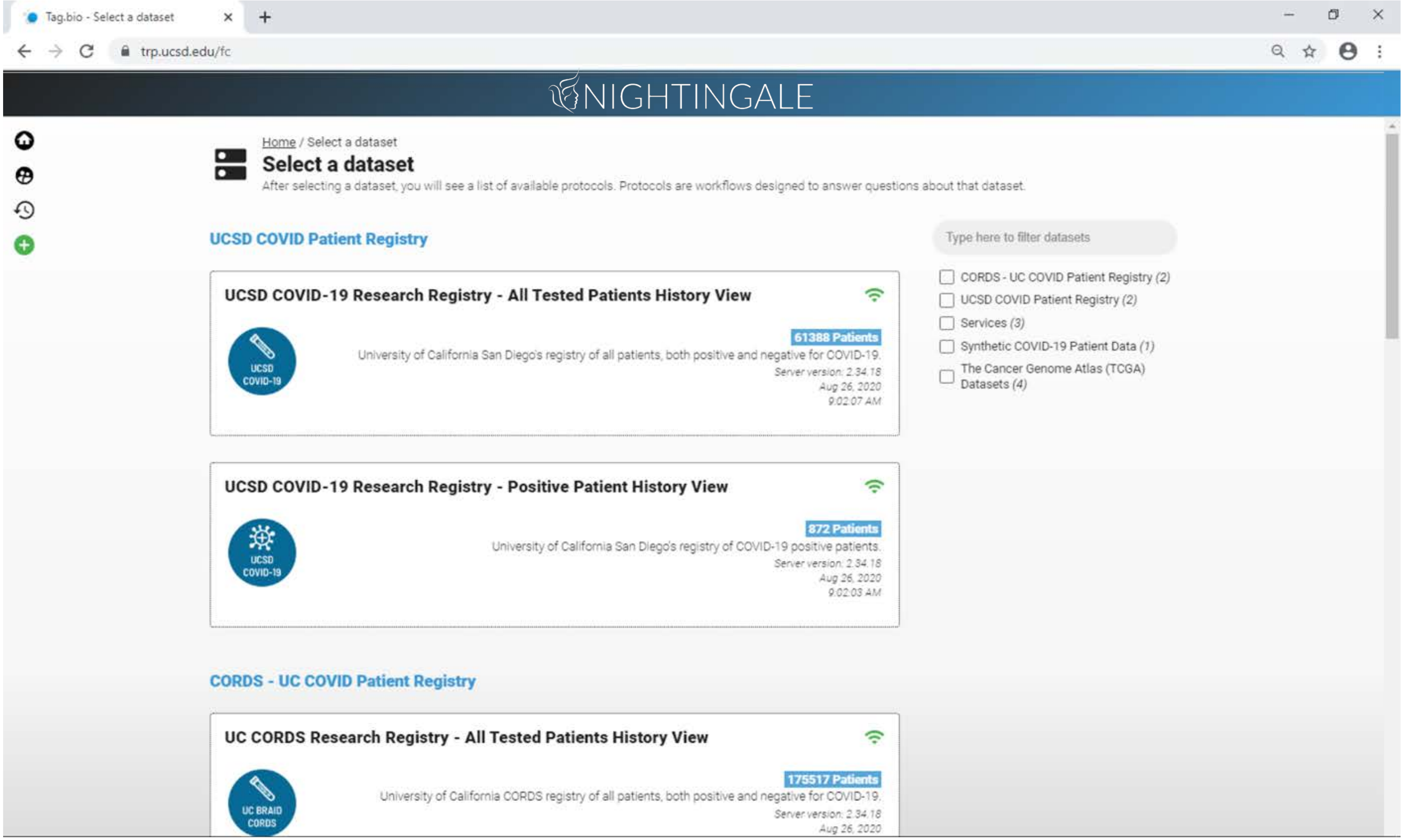
Dataset tags
Add tags

Dataset details

Title	Synthetic COVID-19 demo data
Singular entity name	patient
Plural entity name	patients
Description	Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.
Cloud Provider	
Link	https://github.com/synthetichealth/synthea/issues/679

Users

Add an existing user to this dataset
Add all site users to dataset
Remove all users from dataset



Home / Select a dataset

Select a dataset

After selecting a dataset, you will see a list of available protocols. Protocols are workflows designed to answer questions about that dataset.

Type here to filter datasets

□ CORDS - UC COVID Patient Registry *(2)*
□ UCSD COVID Patient Registry *(2)*
□ Services *(3)*
□ Synthetic COVID-19 Patient Data *(1)*
□ The Cancer Genome Atlas (TCGA) Datasets *(4)*

UCSD COVID Patient Registry

UCSD COVID-19 Research Registry - All Tested Patients History View

UCSD COVID-19

61388 Patients

University of California San Diego's registry of all patients, both positive and negative for COVID-19.

Server version: 2.34.18
Aug 26, 2020
9:02:07 AM

UCSD COVID-19 Research Registry - Positive Patient History View

UCSD COVID-19

872 Patients

University of California San Diego's registry of COVID-19 positive patients.

Server version: 2.34.18
Aug 26, 2020
9:02:03 AM

CORDS - UC COVID Patient Registry

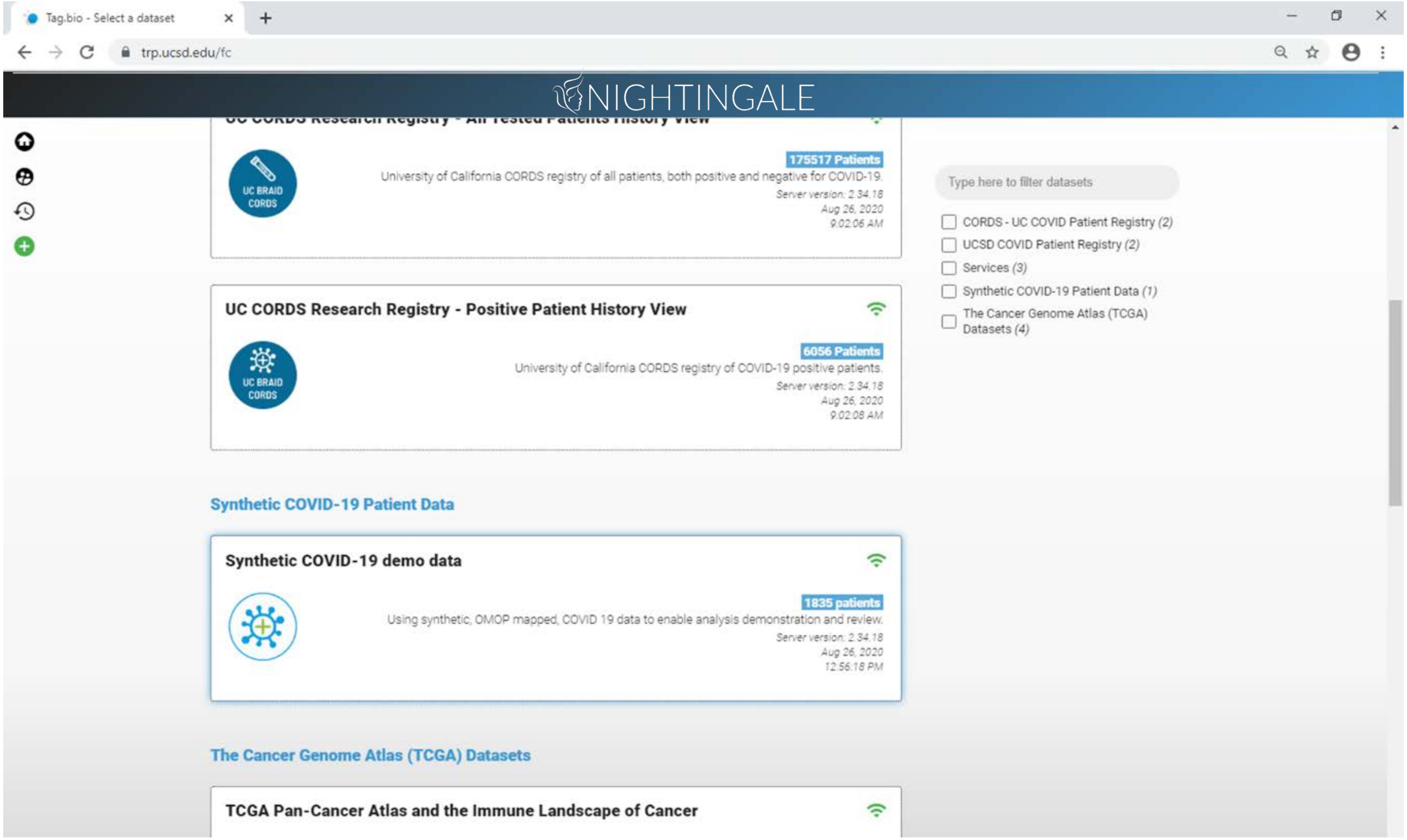
UC CORDS Research Registry - All Tested Patients History View

UC BRAID CORDS

175517 Patients

University of California CORDS registry of all patients, both positive and negative for COVID-19.

Server version: 2.34.18
Aug 26, 2020



UC CORDS Research Registry - All Tested Patients History View

175517 Patients

University of California CORDS registry of all patients, both positive and negative for COVID-19.

Server version: 2.34.18
Aug 26, 2020
9:02:06 AM

UC CORDS Research Registry - Positive Patient History View

6056 Patients

University of California CORDS registry of COVID-19 positive patients.

Server version: 2.34.18
Aug 26, 2020
9:02:08 AM

Synthetic COVID-19 Patient Data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

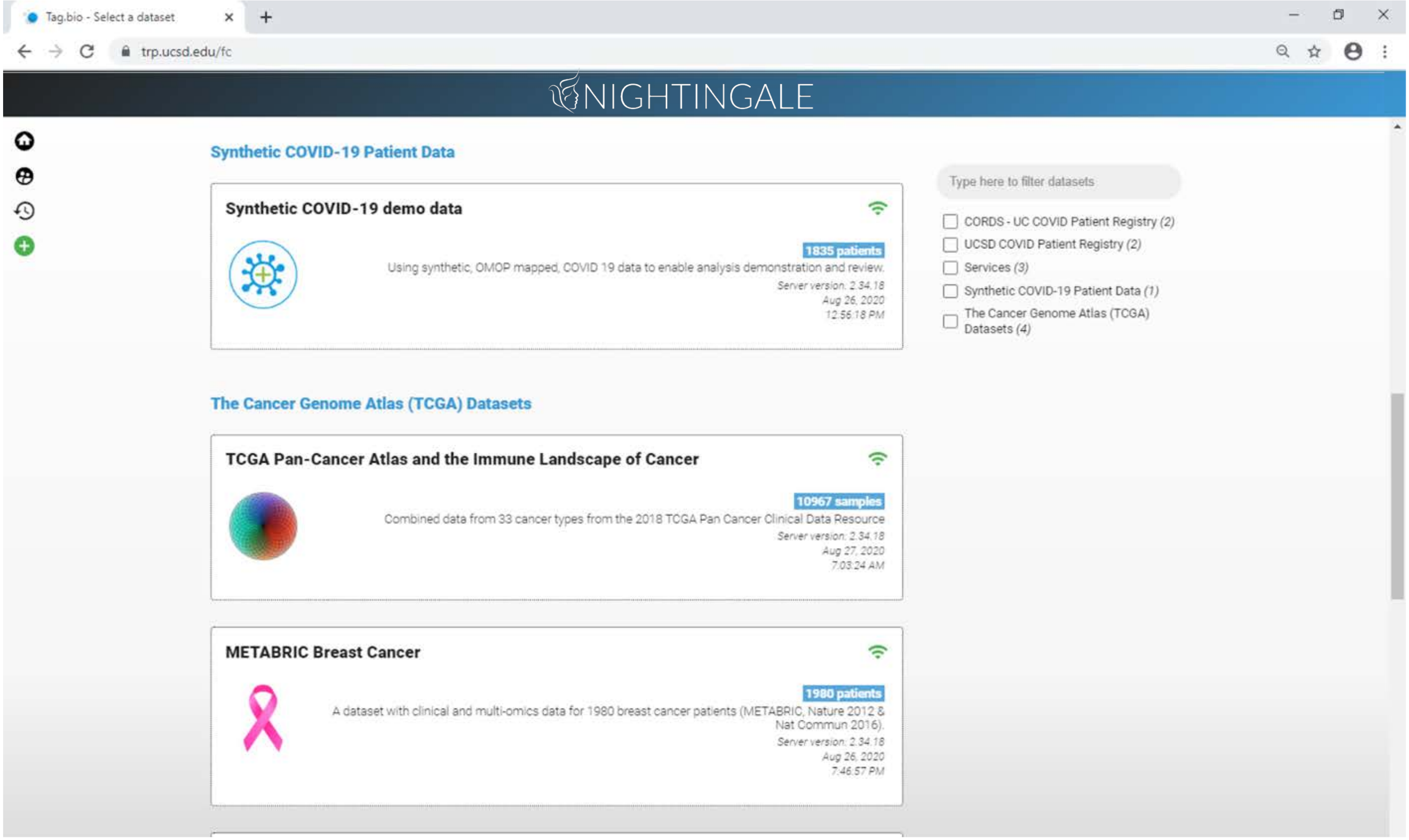
Server version: 2.34.18
Aug 26, 2020
12:56:18 PM

The Cancer Genome Atlas (TCGA) Datasets

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

Type here to filter datasets

- □ CORDS - UC COVID Patient Registry *(2)*
- □ UCSD COVID Patient Registry *(2)*
- □ Services *(3)*
- □ Synthetic COVID-19 Patient Data *(1)*
- □ The Cancer Genome Atlas (TCGA) Datasets *(4)*



Synthetic COVID-19 Patient Data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

Server version: 2.34.18
Aug 26, 2020
12:56:18 PM

The Cancer Genome Atlas (TCGA) Datasets

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

10967 samples

Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

Server version: 2.34.18
Aug 27, 2020
7:03:24 AM

METABRIC Breast Cancer

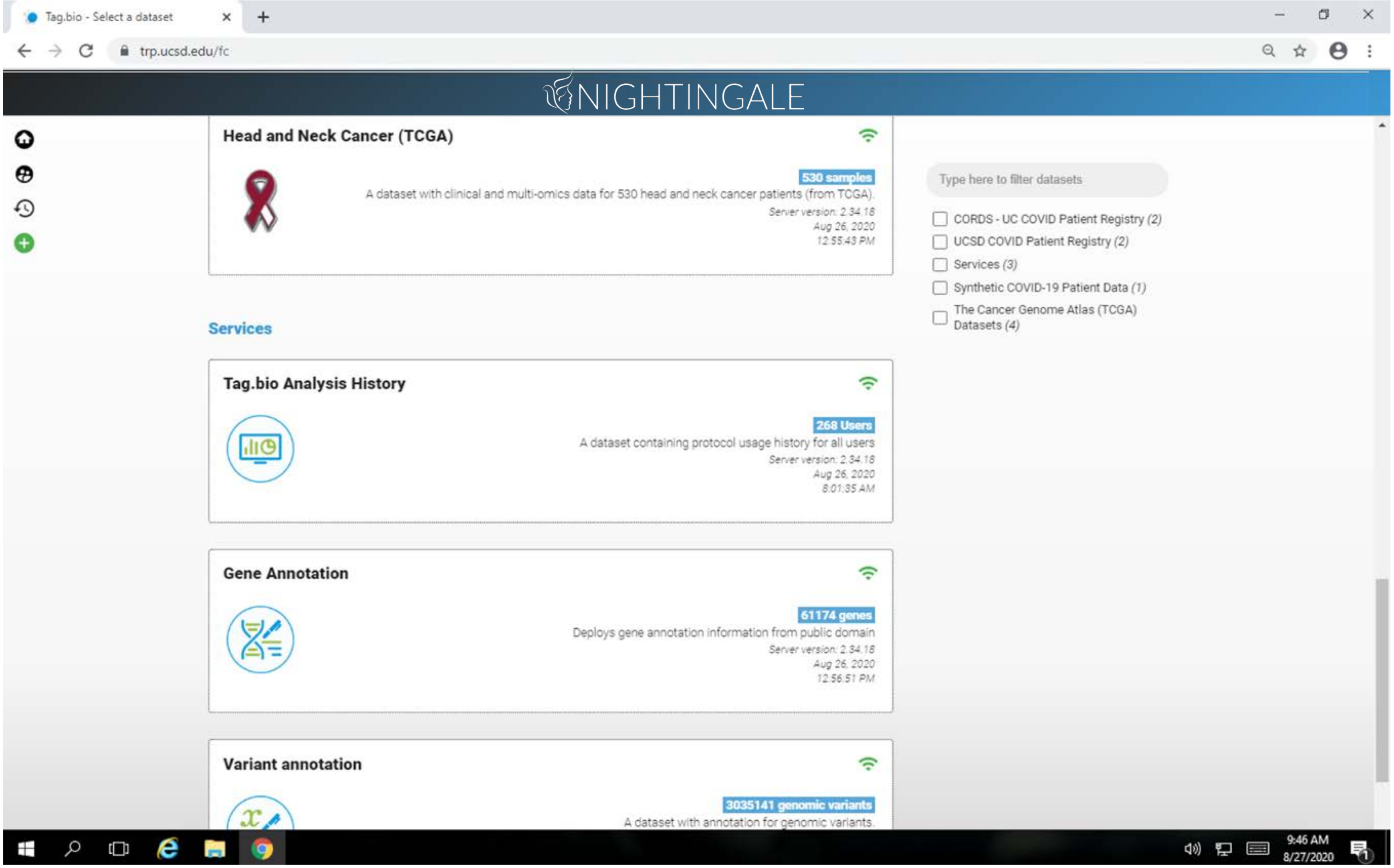
1980 patients

A dataset with clinical and multi-omics data for 1980 breast cancer patients (METABRIC, Nature 2012 & Nat Commun 2016).

Server version: 2.34.18
Aug 26, 2020
7:46:57 PM

Type here to filter datasets

- □ CORDS - UC COVID Patient Registry *(2)*
- □ UCSD COVID Patient Registry *(2)*
- □ Services *(3)*
- □ Synthetic COVID-19 Patient Data *(1)*
- □ The Cancer Genome Atlas (TCGA) Datasets *(4)*



Head and Neck Cancer (TCGA)

530 samples

A dataset with clinical and multi-omics data for 530 head and neck cancer patients (from TCGA).

Server version: 2.34.18
Aug 26, 2020
12:55:43 PM

Type here to filter datasets

- □ CORDS - UC COVID Patient Registry *(2)*
- □ UCSD COVID Patient Registry *(2)*
- □ Services *(3)*
- □ Synthetic COVID-19 Patient Data *(1)*
- □ The Cancer Genome Atlas (TCGA) Datasets *(4)*

Services

Tag.bio Analysis History

268 Users

A dataset containing protocol usage history for all users

Server version: 2.34.18
Aug 26, 2020
8:01:35 AM

Gene Annotation

61174 genes

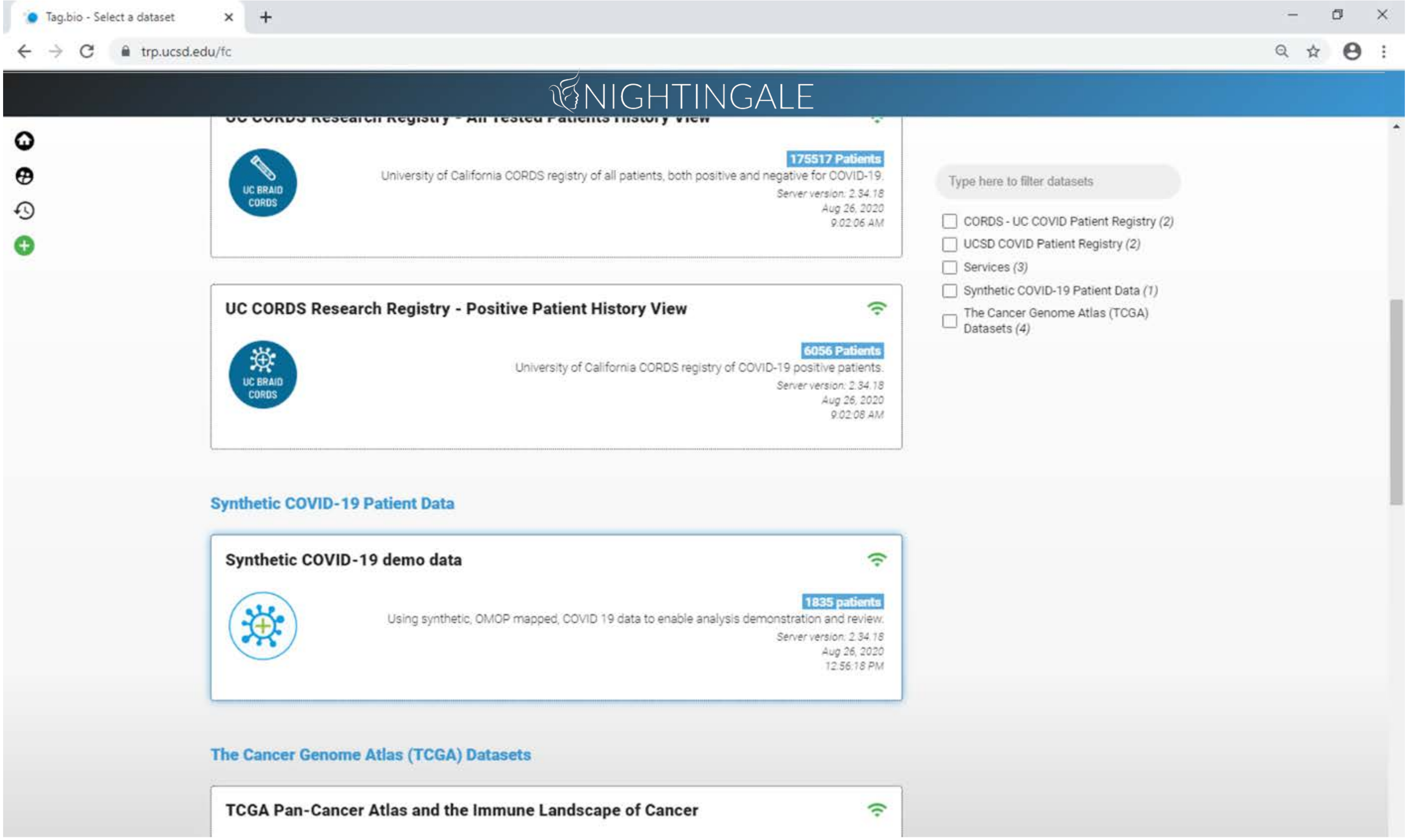
Deploys gene annotation information from public domain

Server version: 2.34.18
Aug 26, 2020
12:56:51 PM

Variant annotation

3035141 genomic variants

A dataset with annotation for genomic variants.



Tag.bio - Select a dataset

trp.ucsd.edu/fc

NIGHTINGALE

UC CORDS Research Registry - All Tested Patients History View

175517 Patients

University of California CORDS registry of all patients, both positive and negative for COVID-19.

Server version: 2.34.18
Aug 26, 2020
9:02:06 AM

UC CORDS Research Registry - Positive Patient History View

6056 Patients

University of California CORDS registry of COVID-19 positive patients.

Server version: 2.34.18
Aug 26, 2020
9:02:08 AM

Synthetic COVID-19 Patient Data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

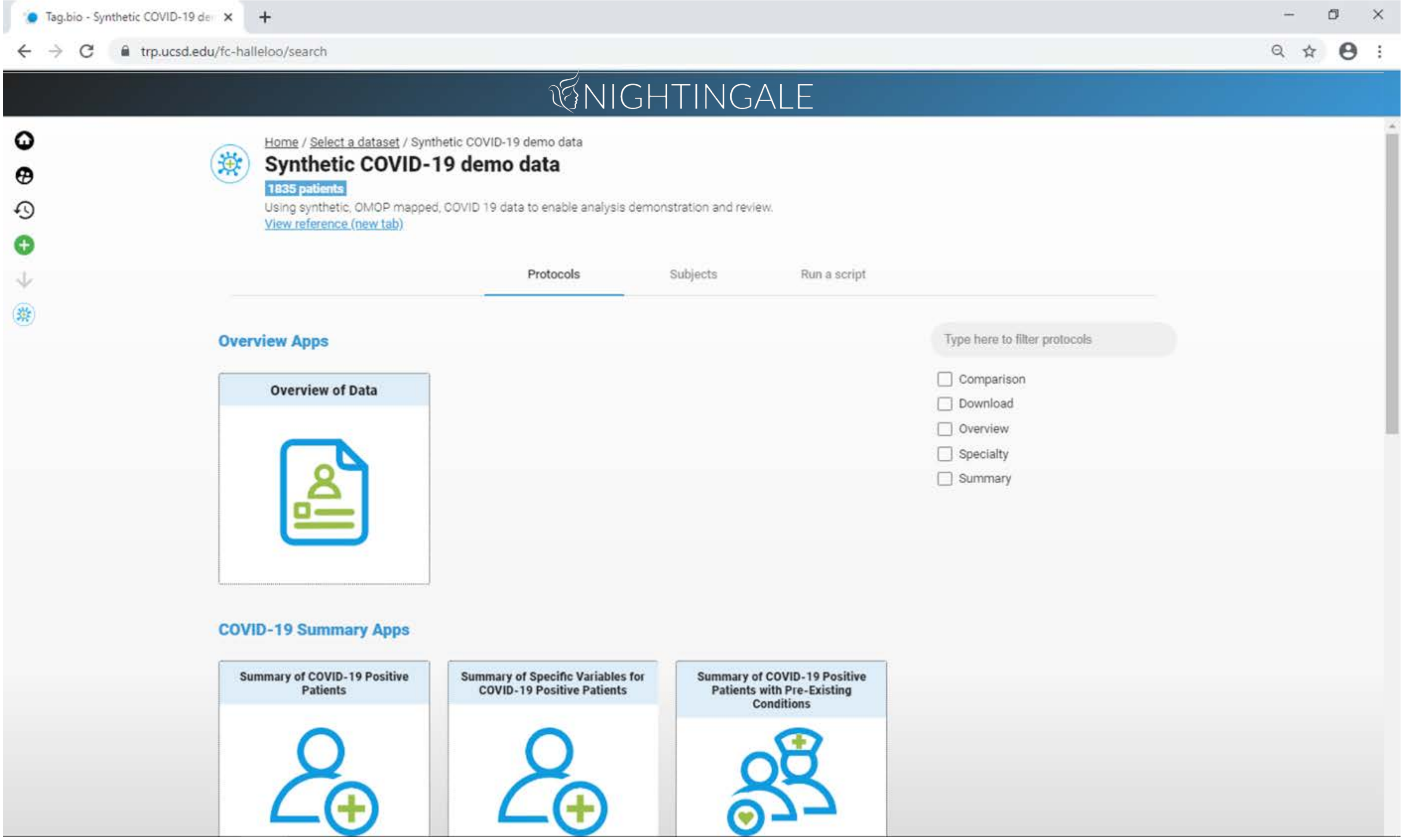
Server version: 2.34.18
Aug 26, 2020
12:56:18 PM

The Cancer Genome Atlas (TCGA) Datasets

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

Type here to filter datasets

□ CORDS - UC COVID Patient Registry *(2)*
□ UCSD COVID Patient Registry *(2)*
□ Services *(3)*
□ Synthetic COVID-19 Patient Data *(1)*
□ The Cancer Genome Atlas (TCGA) Datasets *(4)*



Tag.bio - Synthetic COVID-19 de
trp.ucsd.edu/fc-halleloo/search
NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

View reference (new tab)

Protocols
Subjects
Run a script

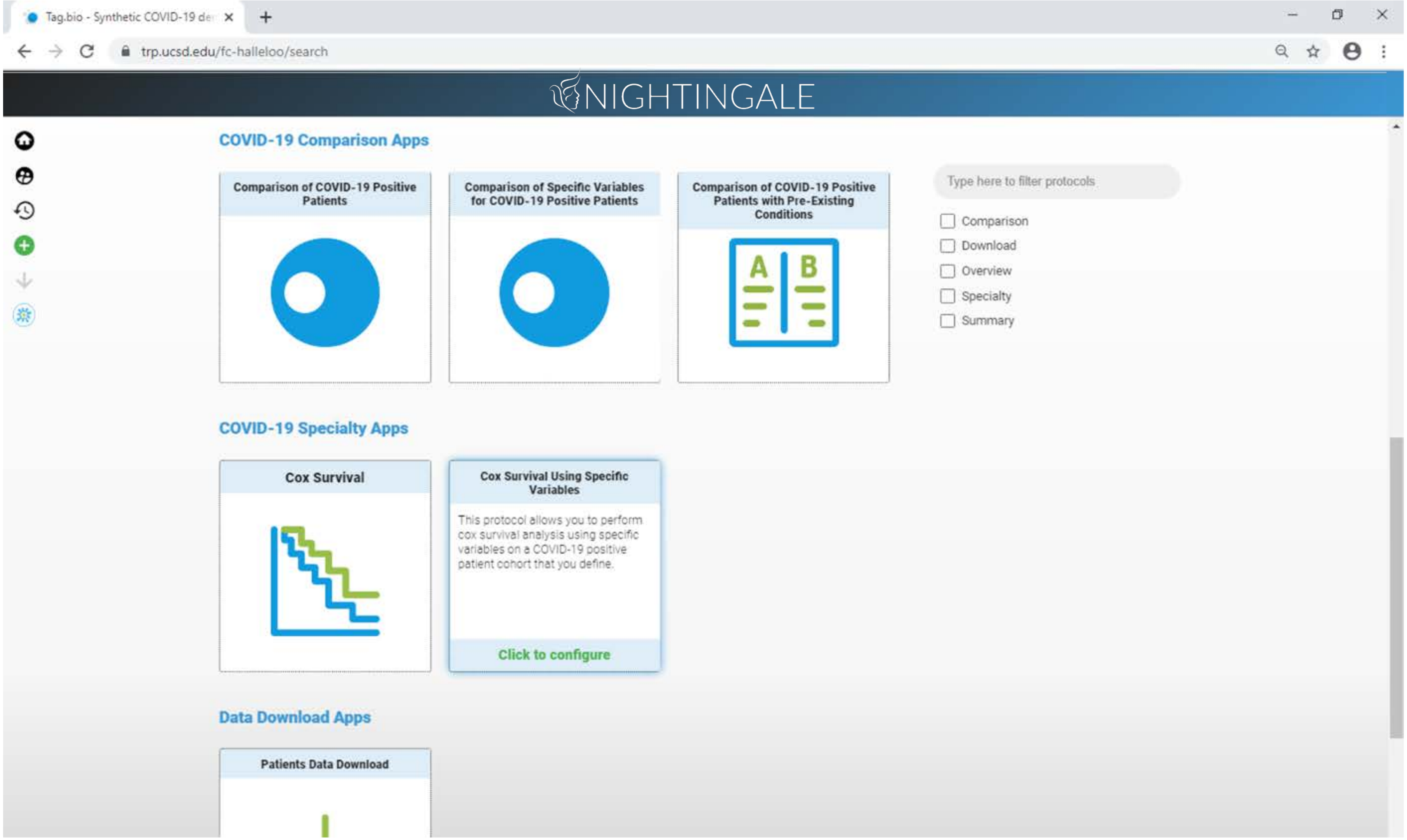
Overview Apps

Type here to filter protocols

□ Comparison
□ Download
□ Overview
□ Specialty
□ Summary

COVID-19 Summary Apps

Summary of COVID-19 Positive Patients
Summary of Specific Variables for COVID-19 Positive Patients
Summary of COVID-19 Positive Patients with Pre-Existing Conditions



Tag.bio - Synthetic COVID-19 de
trp.ucsd.edu/fc-halleloo/search
NIGHTINGALE

COVID-19 Comparison Apps

Comparison of COVID-19 Positive Patients

Comparison of Specific Variables for COVID-19 Positive Patients

Comparison of COVID-19 Positive Patients with Pre-Existing Conditions

Type here to filter protocols

□ Comparison
□ Download
□ Overview
□ Specialty
□ Summary

COVID-19 Specialty Apps

Cox Survival

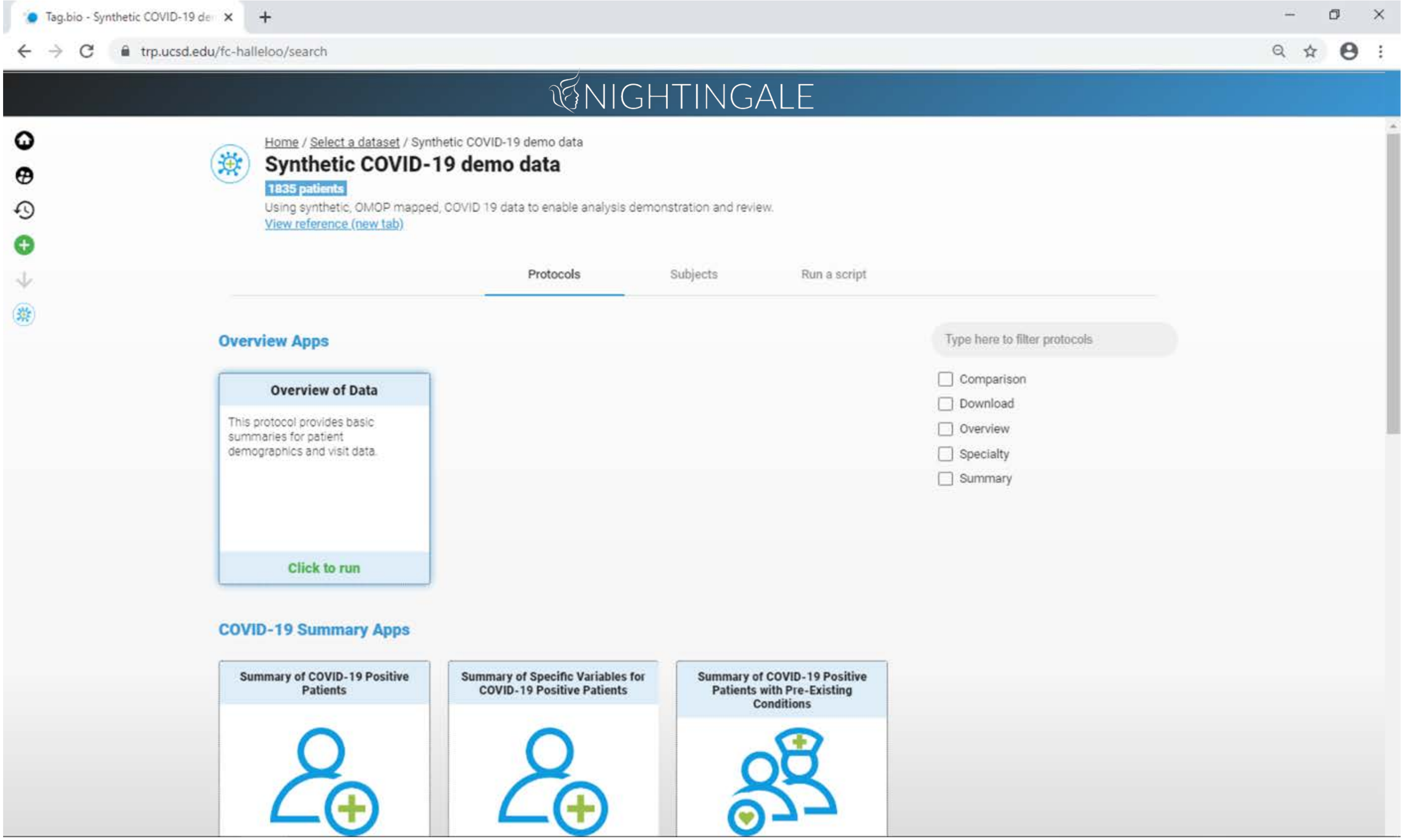
Cox Survival Using Specific Variables

This protocol allows you to perform cox survival analysis using specific variables on a COVID-19 positive patient cohort that you define.

Click to configure

Data Download Apps

Patients Data Download



Tag.bio - Synthetic COVID-19 de

trp.ucsd.edu/fc-halleloo/search

NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

View reference (new tab)

Overview Apps

Overview of Data

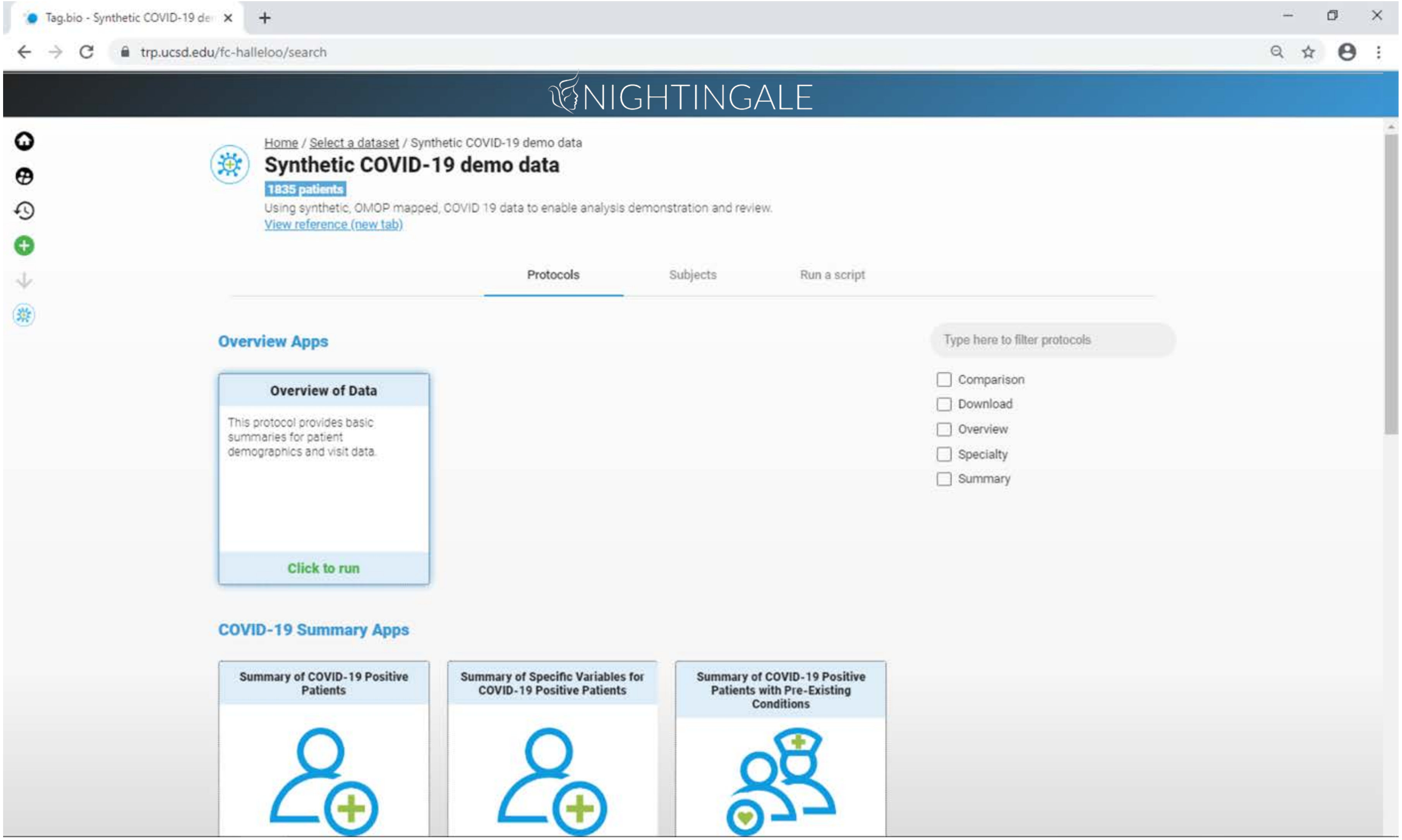
This protocol provides basic summaries for patient demographics and visit data.

COVID-19 Summary Apps

Summary of COVID-19 Positive Patients

Summary of Specific Variables for COVID-19 Positive Patients

Summary of COVID-19 Positive Patients with Pre-Existing Conditions



Tag.bio - Synthetic COVID-19 de

trp.ucsd.edu/fc-halleloo/search

NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

View reference (new tab)

Protocols | Subjects | Run a script

Overview Apps

Overview of Data

This protocol provides basic summaries for patient demographics and visit data.

Click to run

Type here to filter protocols

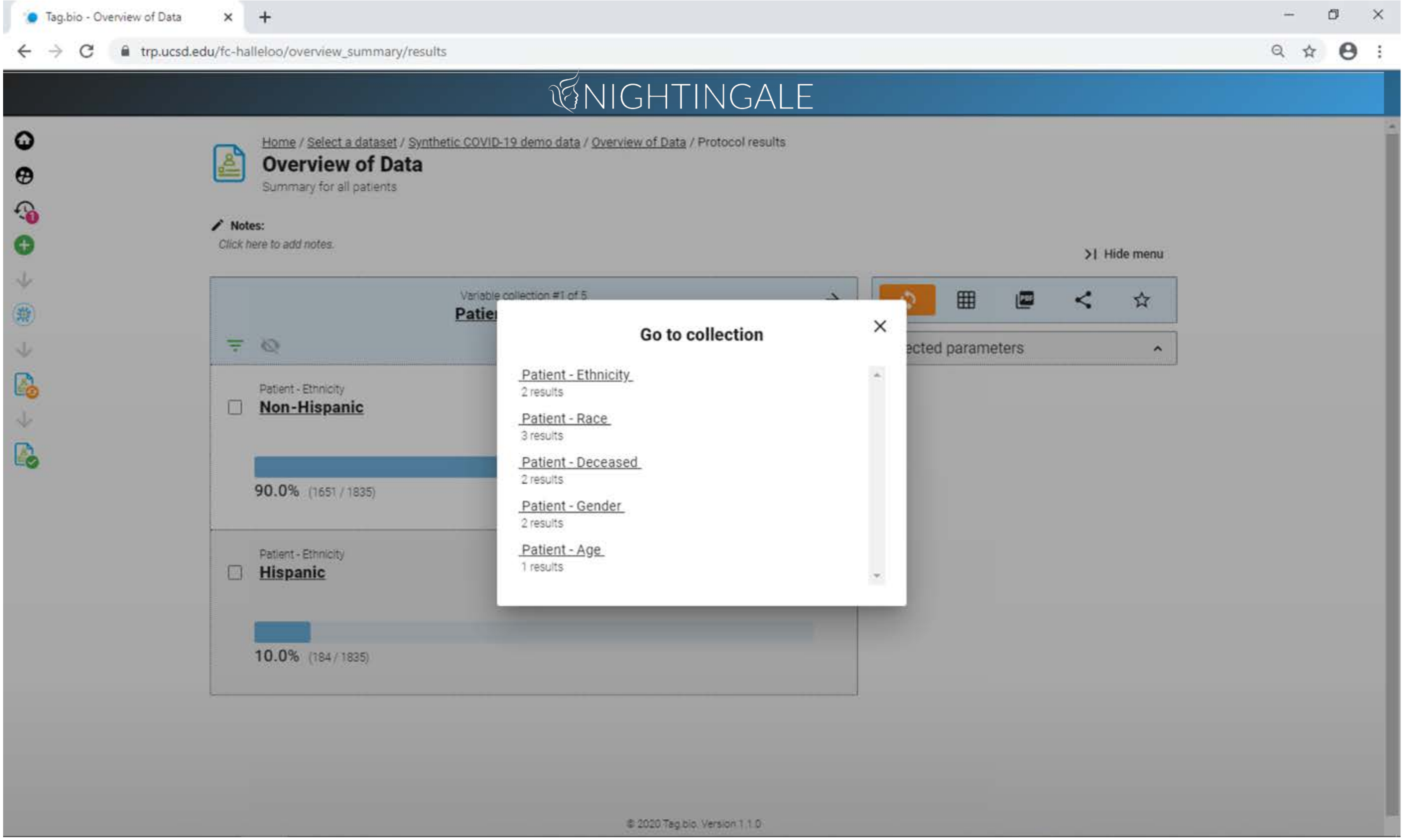
- □ Comparison
- □ Download
- □ Overview
- □ Specialty
- □ Summary

COVID-19 Summary Apps

Summary of COVID-19 Positive Patients

Summary of Specific Variables for COVID-19 Positive Patients

Summary of COVID-19 Positive Patients with Pre-Existing Conditions



Tag.bio - Overview of Data

trp.ucsd.edu/fc-halleloo/overview_summary/results

NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data / Overview of Data / Protocol results

Overview of Data

Summary for all patients

Notes:

Click here to add notes.

>| Hide menu

Variable collection #1 of 5

Patie

...ected parameters

Patient - Ethnicity

□ Non-Hispanic

90.0% (1651 / 1835)

Patient - Ethnicity

□ Hispanic

10.0% (184 / 1835)

Go to collection

Patient - Ethnicity
2 results

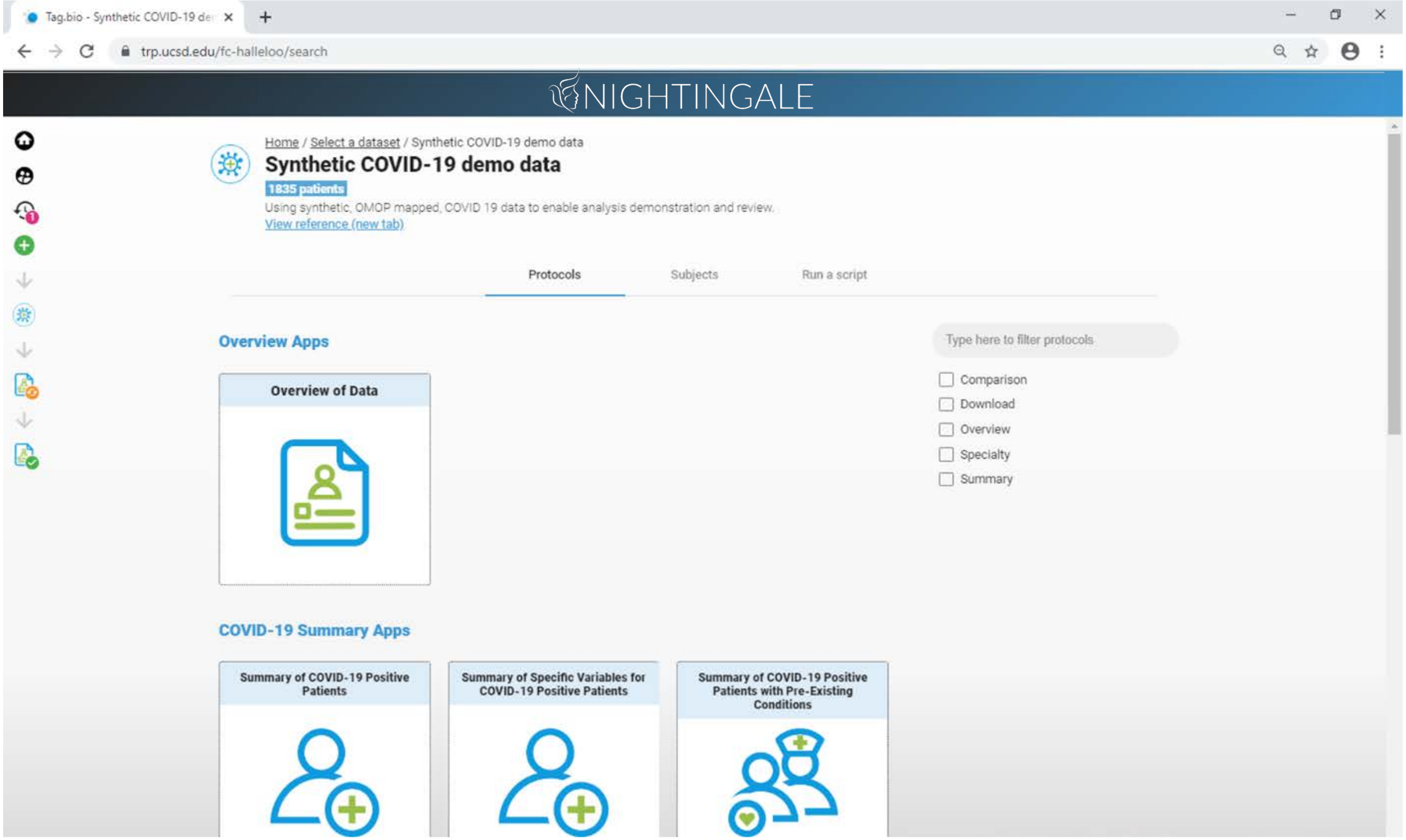
Patient - Race
3 results

Patient - Deceased
2 results

Patient - Gender
2 results

Patient - Age
1 results

© 2020 Tag.bio. Version 1.1.0



Tag.bio - Synthetic COVID-19 de
trp.ucsd.edu/fc-halleloo/search
NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

View reference (new tab)

Protocols
Subjects
Run a script

Overview Apps

Type here to filter protocols

□ Comparison
□ Download
□ Overview
□ Specialty
□ Summary

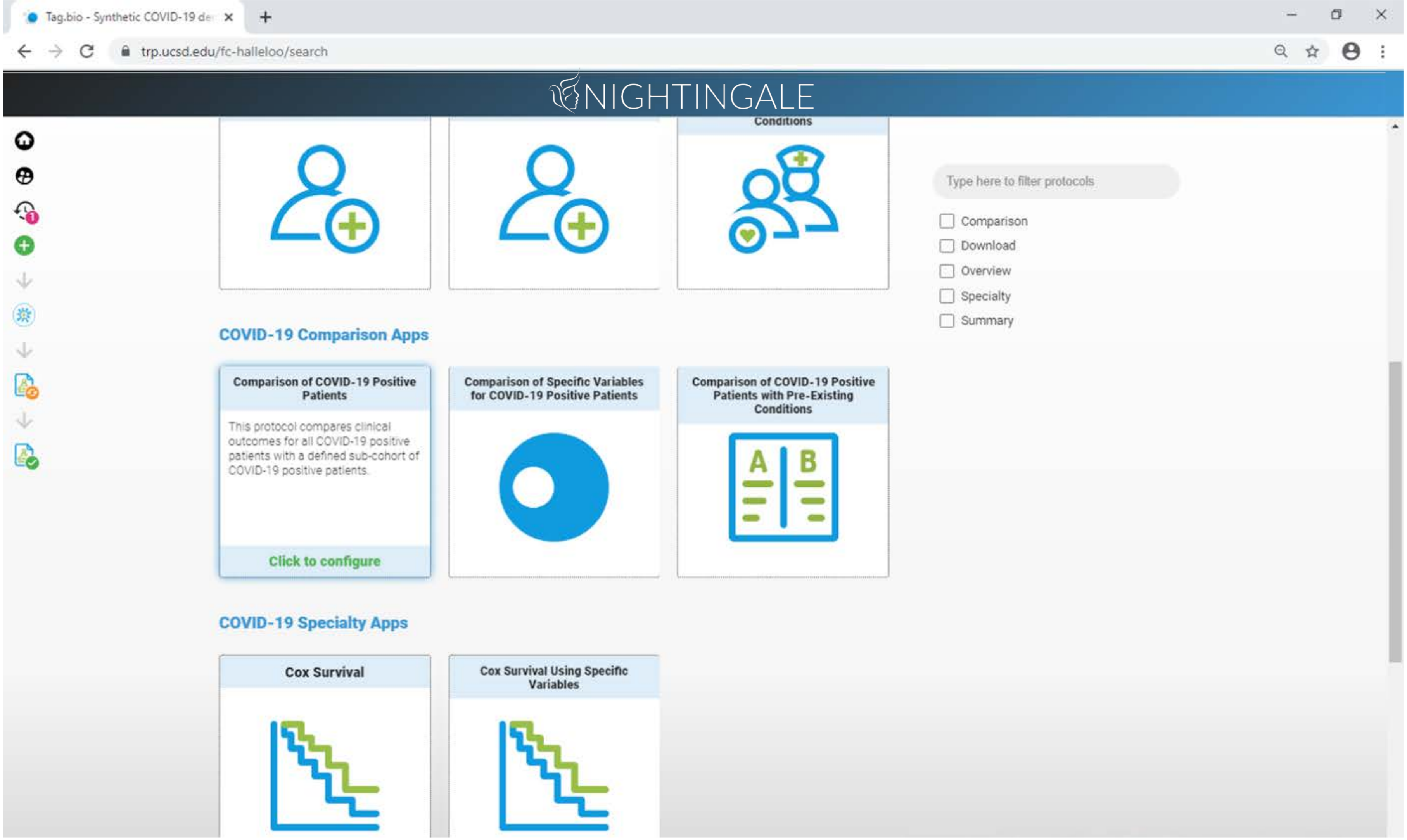
Overview of Data

COVID-19 Summary Apps

Summary of COVID-19 Positive Patients

Summary of Specific Variables for COVID-19 Positive Patients

Summary of COVID-19 Positive Patients with Pre-Existing Conditions


Tag.bio - Synthetic COVID-19 de
trp.ucsd.edu/fc-halleloo/search
NIGHTINGALE
Conditions
Type here to filter protocols

□ Comparison
□ Download
□ Overview
□ Specialty
□ Summary

COVID-19 Comparison Apps

Comparison of COVID-19 Positive Patients

This protocol compares clinical outcomes for all COVID-19 positive patients with a defined sub-cohort of COVID-19 positive patients.

Click to configure

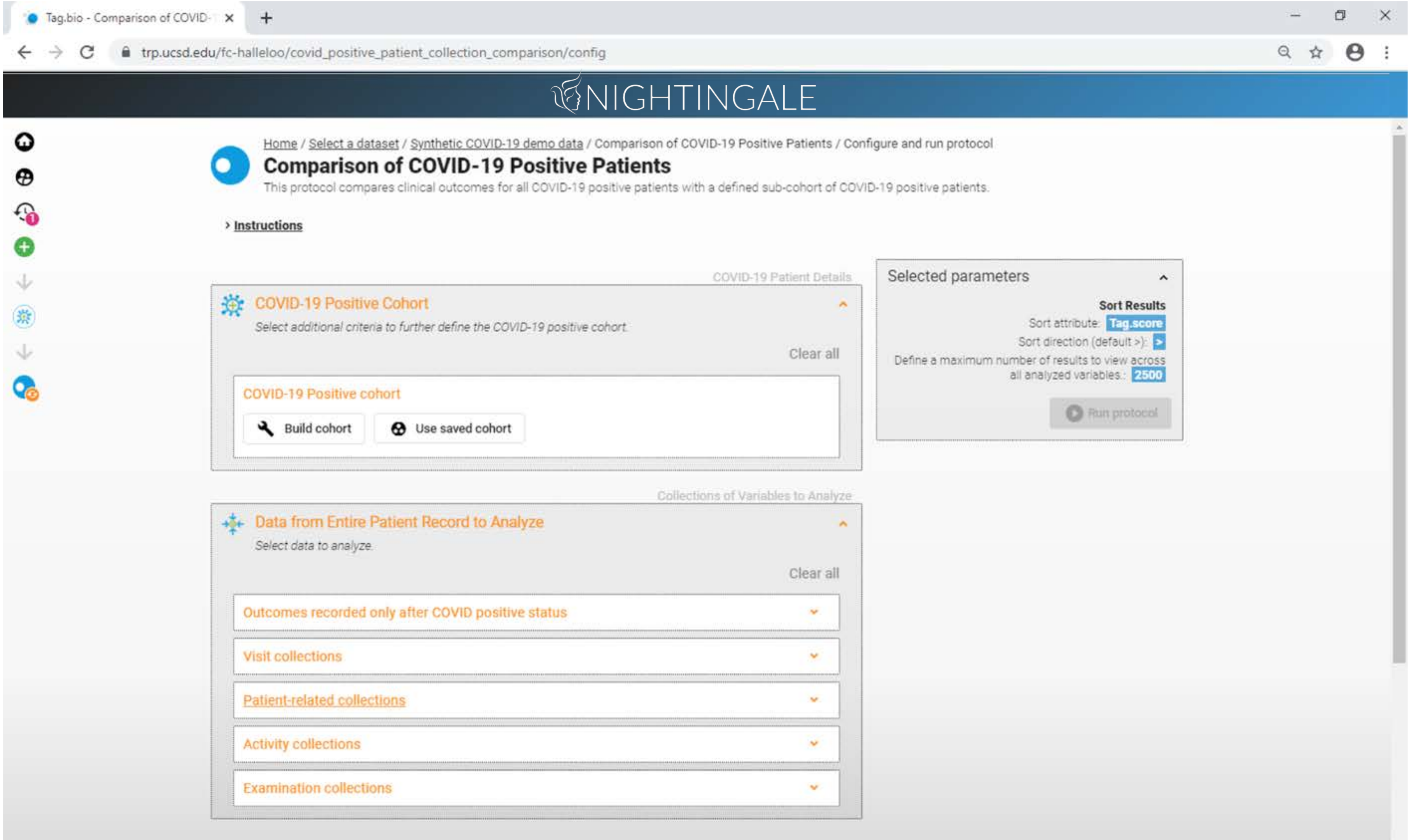
Comparison of Specific Variables for COVID-19 Positive Patients

Comparison of COVID-19 Positive Patients with Pre-Existing Conditions

COVID-19 Specialty Apps

Cox Survival

Cox Survival Using Specific Variables



Tag.bio - Comparison of COVID-

trp.ucsd.edu/fc-halleloo/covid_positive_patient_collection_comparison/config

NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data / Comparison of COVID-19 Positive Patients / Configure and run protocol

Comparison of COVID-19 Positive Patients

This protocol compares clinical outcomes for all COVID-19 positive patients with a defined sub-cohort of COVID-19 positive patients.

› Instructions

COVID-19 Patient Details

COVID-19 Positive Cohort

Select additional criteria to further define the COVID-19 positive cohort.

COVID-19 Positive cohort

Collections of Variables to Analyze

Data from Entire Patient Record to Analyze

Select data to analyze.

Outcomes recorded only after COVID positive status

Visit collections

Patient-related collections

Activity collections

Examination collections

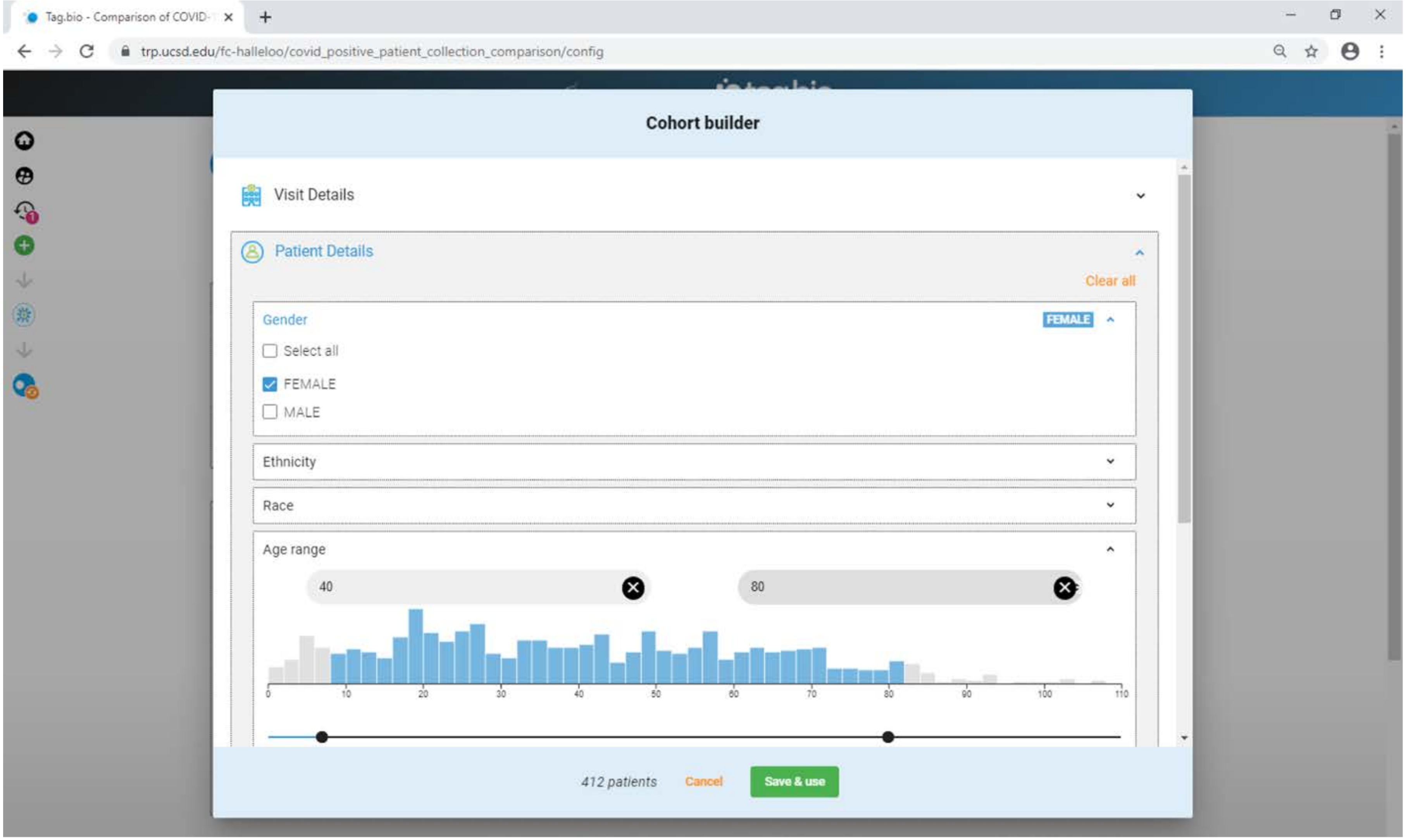
Selected parameters

Sort Results

Sort attribute: Tag.score

Sort direction (default >): >

Define a maximum number of results to view across all analyzed variables.: 2500



Tag.bio - Comparison of COVID-

trp.ucsd.edu/fc-halleloo/covid_positive_patient_collection_comparison/config

Cohort builder

Visit Details

Patient Details

Clear all

Gender FEMALE

☐ Select all

☑ FEMALE

☐ MALE

Ethnicity

Race

Age range

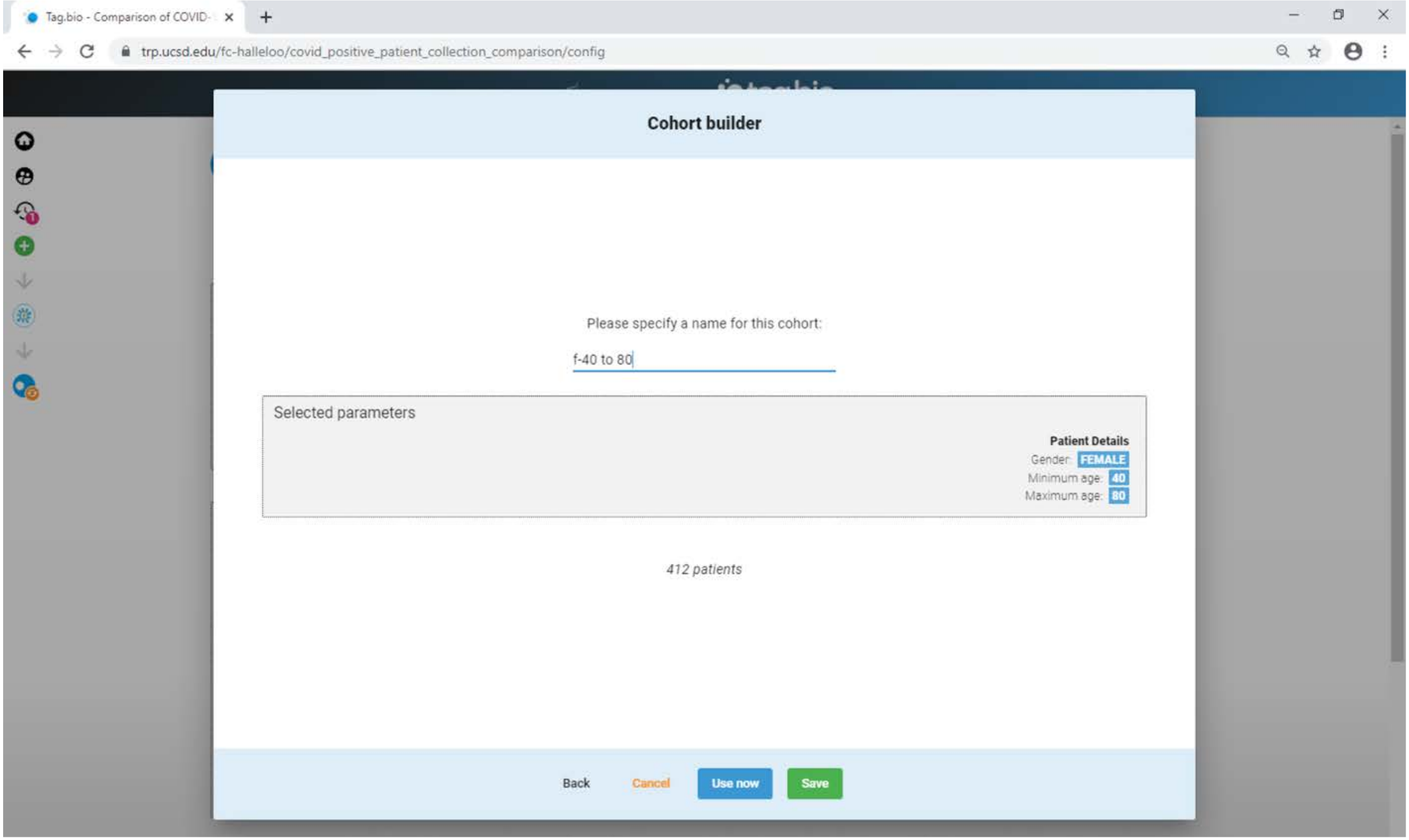
40

80

412 patients

Cancel

Save & use



Cohort builder

Please specify a name for this cohort:

f-40 to 80

Selected parameters

Patient Details

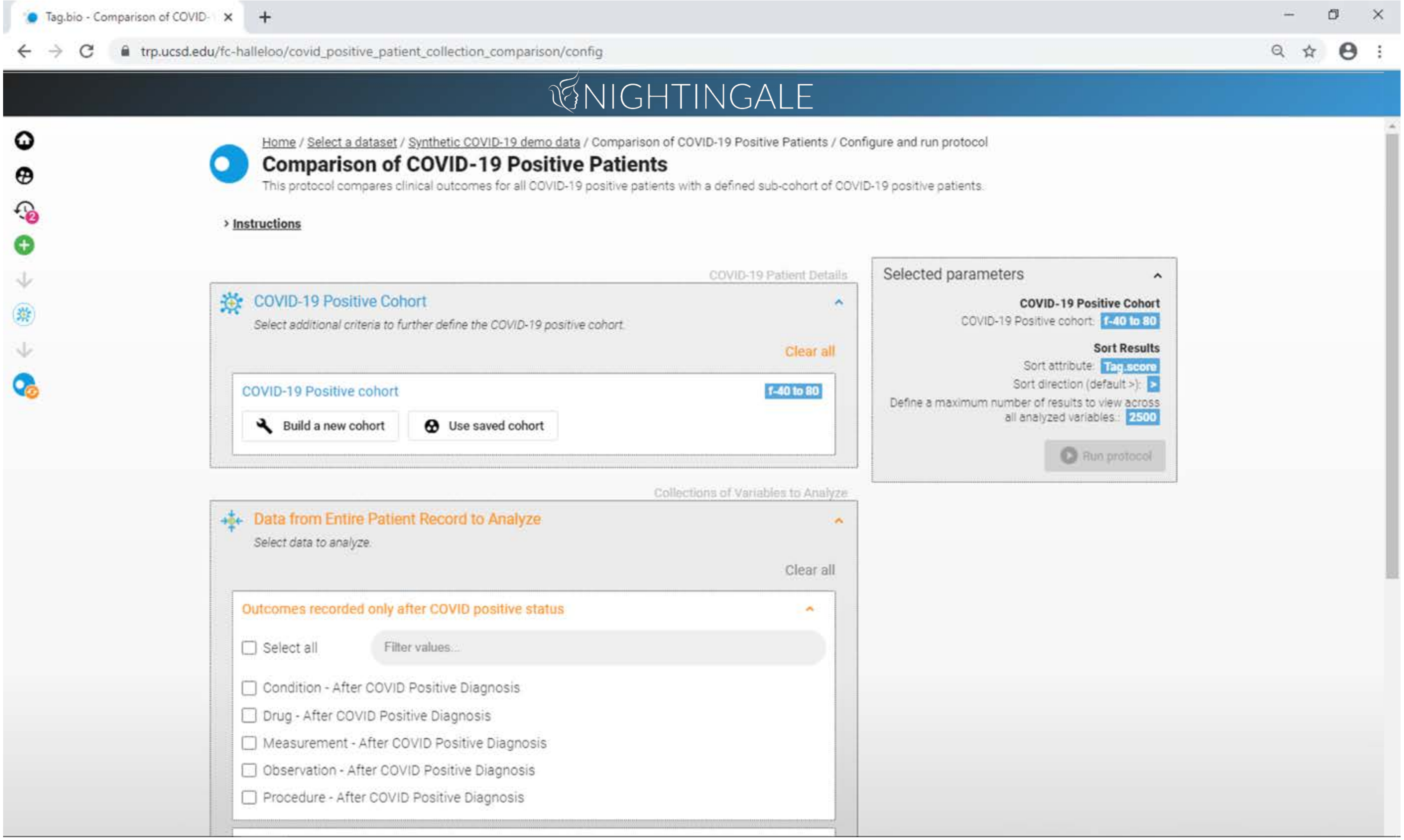
Gender: FEMALE

Minimum age: 40

Maximum age: 80

412 patients

Tag.bio - Comparison of COVID-
trp.ucsd.edu/fc-halleloo/covid_positive_patient_collection_comparison/config
Back
Cancel
Use now
Save



Tag.bio - Comparison of COVID-
trp.ucsd.edu/fc-halleloo/covid_positive_patient_collection_comparison/config
NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data / Comparison of COVID-19 Positive Patients / Configure and run protocol

Comparison of COVID-19 Positive Patients

This protocol compares clinical outcomes for all COVID-19 positive patients with a defined sub-cohort of COVID-19 positive patients.

› Instructions

COVID-19 Patient Details

COVID-19 Positive Cohort

Select additional criteria to further define the COVID-19 positive cohort.

Clear all

COVID-19 Positive cohort: f-40 to 80

Selected parameters

COVID-19 Positive Cohort

COVID-19 Positive cohort: f-40 to 80

Sort Results

Sort attribute: Tag.score

Sort direction (default >): >

Define a maximum number of results to view across all analyzed variables.: 2500

Collections of Variables to Analyze

Data from Entire Patient Record to Analyze

Select data to analyze.

Clear all

Outcomes recorded only after COVID positive status

□ Select all

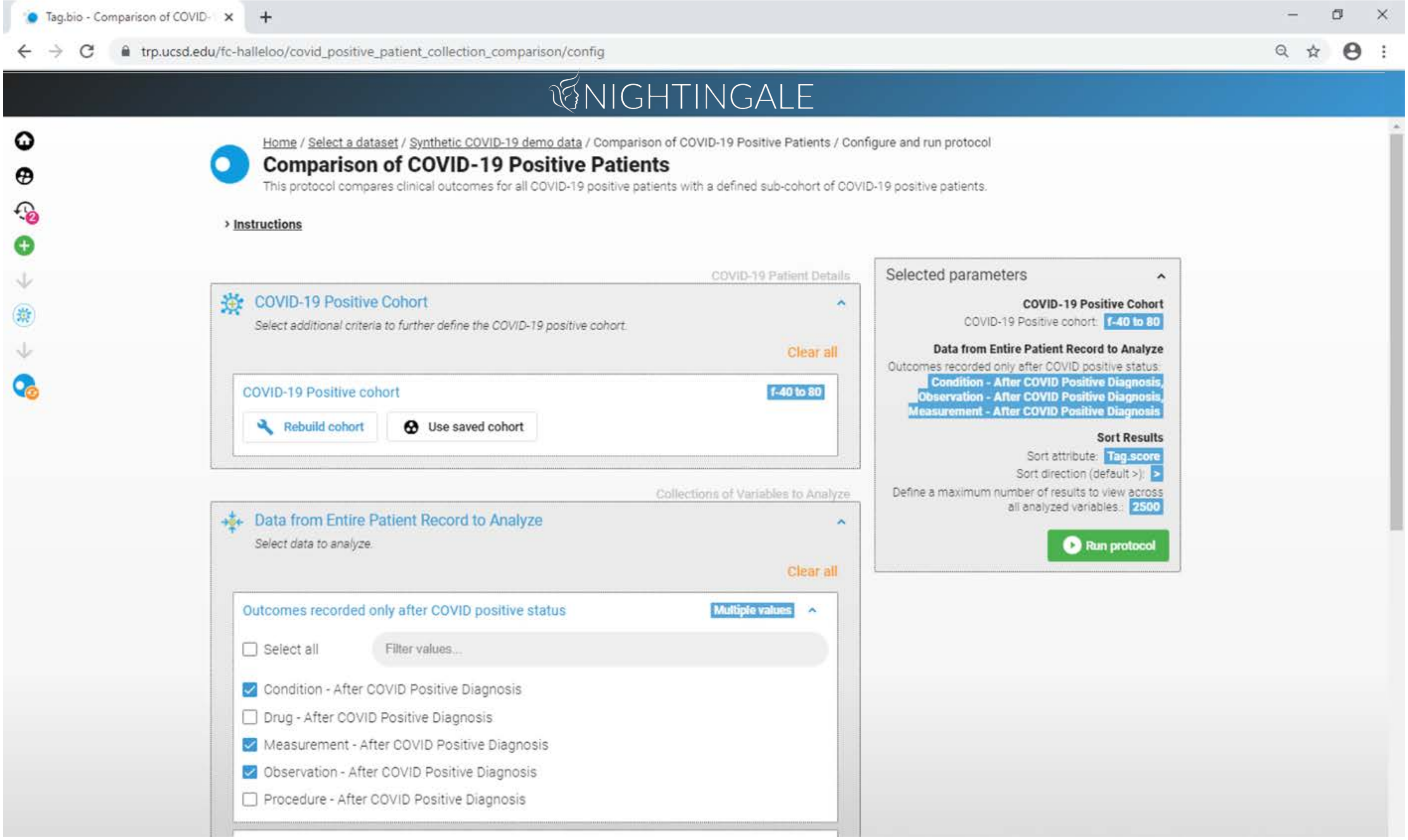
□ Condition - After COVID Positive Diagnosis

□ Drug - After COVID Positive Diagnosis

□ Measurement - After COVID Positive Diagnosis

□ Observation - After COVID Positive Diagnosis

□ Procedure - After COVID Positive Diagnosis


Tag.bio - Comparison of COVID-

trp.ucsd.edu/fc-halleloo/covid_positive_patient_collection_comparison/config

NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data / Comparison of COVID-19 Positive Patients / Configure and run protocol

Comparison of COVID-19 Positive Patients

This protocol compares clinical outcomes for all COVID-19 positive patients with a defined sub-cohort of COVID-19 positive patients.

› Instructions

COVID-19 Patient Details

COVID-19 Positive Cohort

Select additional criteria to further define the COVID-19 positive cohort.

Clear all

COVID-19 Positive cohort f-40 to 80

Rebuild cohort | Use saved cohort

Collections of Variables to Analyze

Data from Entire Patient Record to Analyze

Select data to analyze.

Clear all

Outcomes recorded only after COVID positive status Multiple values

□ Select all Filter values...

☑ Condition - After COVID Positive Diagnosis

□ Drug - After COVID Positive Diagnosis

☑ Measurement - After COVID Positive Diagnosis

☑ Observation - After COVID Positive Diagnosis

□ Procedure - After COVID Positive Diagnosis

Selected parameters

COVID-19 Positive Cohort

COVID-19 Positive cohort: f-40 to 80

Data from Entire Patient Record to Analyze

Outcomes recorded only after COVID positive status: Condition - After COVID Positive Diagnosis, Observation - After COVID Positive Diagnosis, Measurement - After COVID Positive Diagnosis

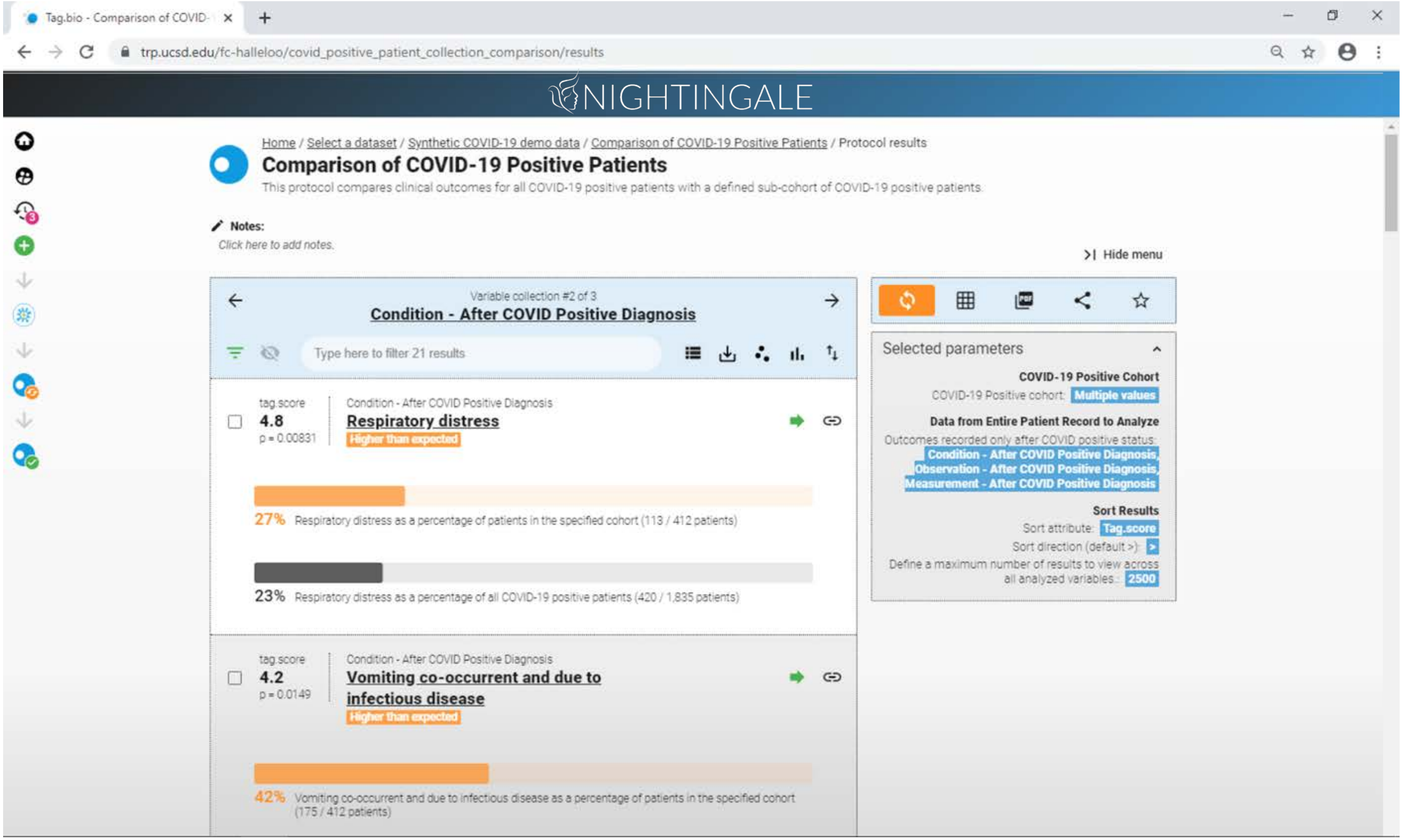
Sort Results

Sort attribute: Tag.score

Sort direction (default >): >

Define a maximum number of results to view across all analyzed variables.: 2500

Run protocol



Tag.bio - Comparison of COVID-
trp.ucsd.edu/fc-halleloo/covid_positive_patient_collection_comparison/results
NIGHTINGALE

Home / Select a dataset / Synthetic COVID-19 demo data / Comparison of COVID-19 Positive Patients / Protocol results

Comparison of COVID-19 Positive Patients

This protocol compares clinical outcomes for all COVID-19 positive patients with a defined sub-cohort of COVID-19 positive patients.

Notes:

Click here to add notes.

Variable collection #2 of 3

Condition - After COVID Positive Diagnosis

□ tag.score **4.8** p = 0.00831

Condition - After COVID Positive Diagnosis

Respiratory distress

Higher than expected

27% Respiratory distress as a percentage of patients in the specified cohort (113 / 412 patients)

23% Respiratory distress as a percentage of all COVID-19 positive patients (420 / 1,835 patients)

□ tag.score **4.2** p = 0.0149

Condition - After COVID Positive Diagnosis

Vomiting co-occurrent and due to infectious disease

Higher than expected

42% Vomiting co-occurrent and due to infectious disease as a percentage of patients in the specified cohort (175 / 412 patients)

Selected parameters

COVID-19 Positive Cohort

COVID-19 Positive cohort: Multiple values

Data from Entire Patient Record to Analyze

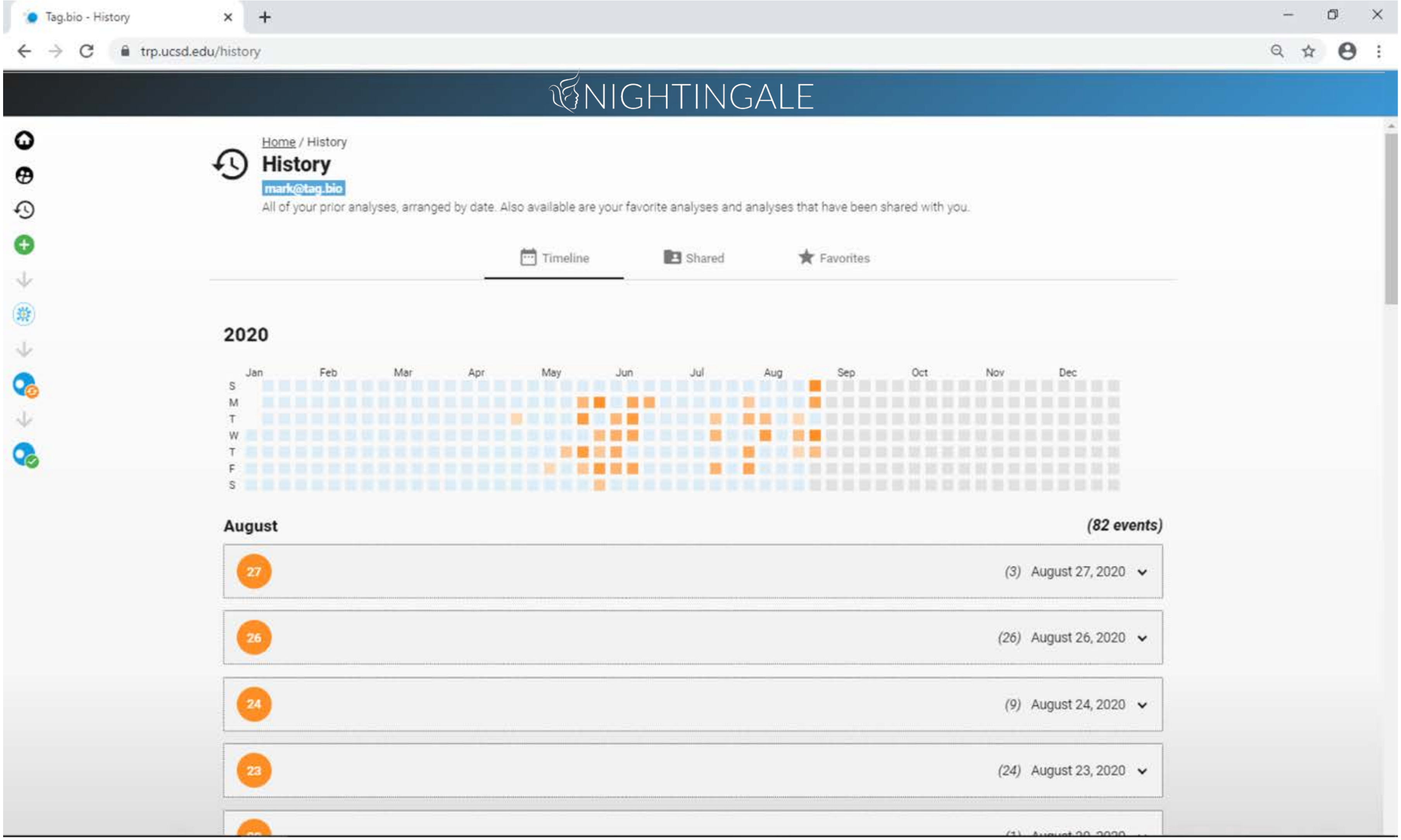
Outcomes recorded only after COVID positive status: Condition - After COVID Positive Diagnosis, Observation - After COVID Positive Diagnosis, Measurement - After COVID Positive Diagnosis

Sort Results

Sort attribute: Tag.score

Sort direction (default >): >

Define a maximum number of results to view across all analyzed variables.: 2500



Home / History

History

mark@tag.bio

All of your prior analyses, arranged by date. Also available are your favorite analyses and analyses that have been shared with you.

2020

August *(82 events)*

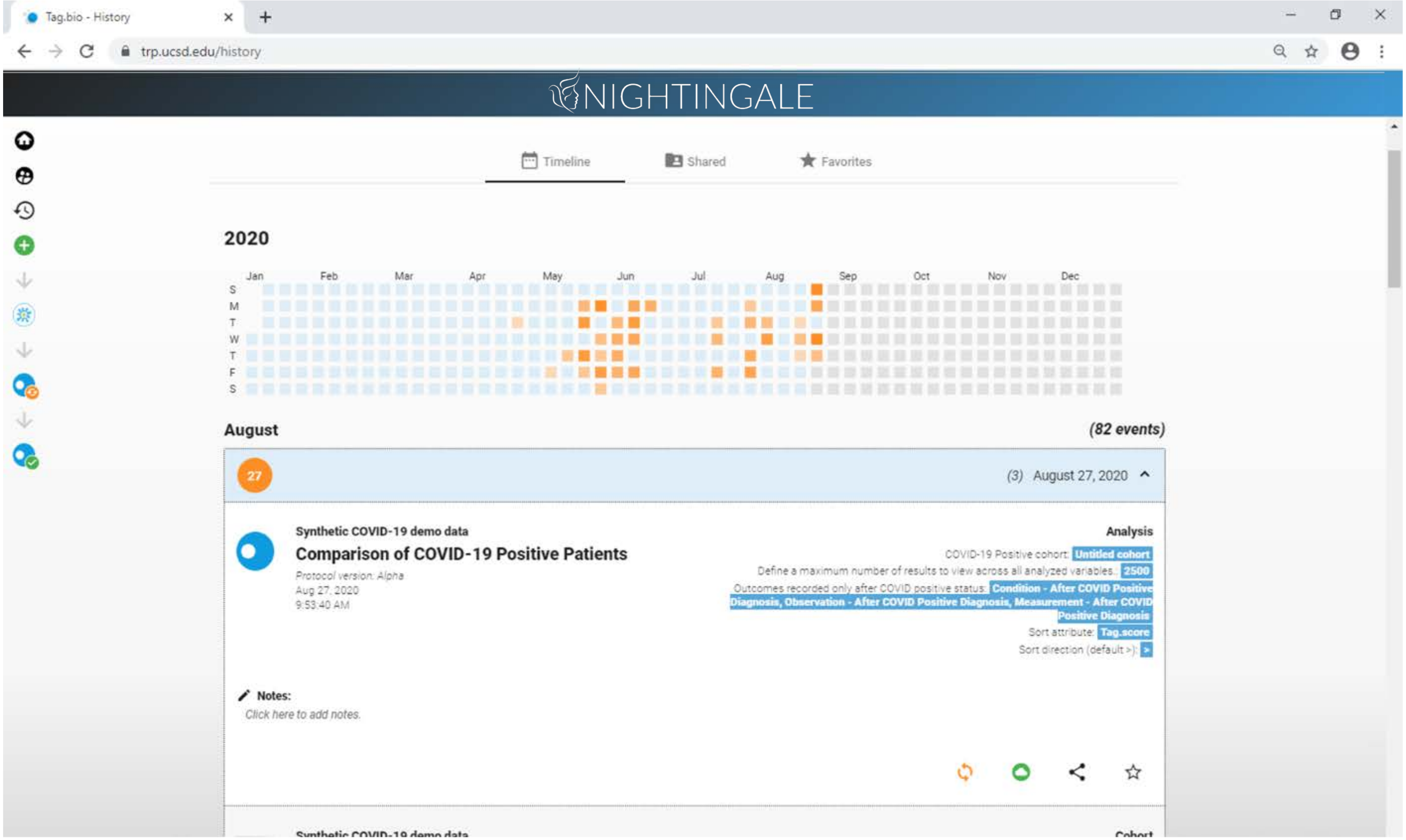
27 *(3)* August 27, 2020

26 *(26)* August 26, 2020

24 *(9)* August 24, 2020

23 *(24)* August 23, 2020

20 *(1)* August 20, 2020



2020

August *(82 events)*

27 *(3)* August 27, 2020

Synthetic COVID-19 demo data

Comparison of COVID-19 Positive Patients

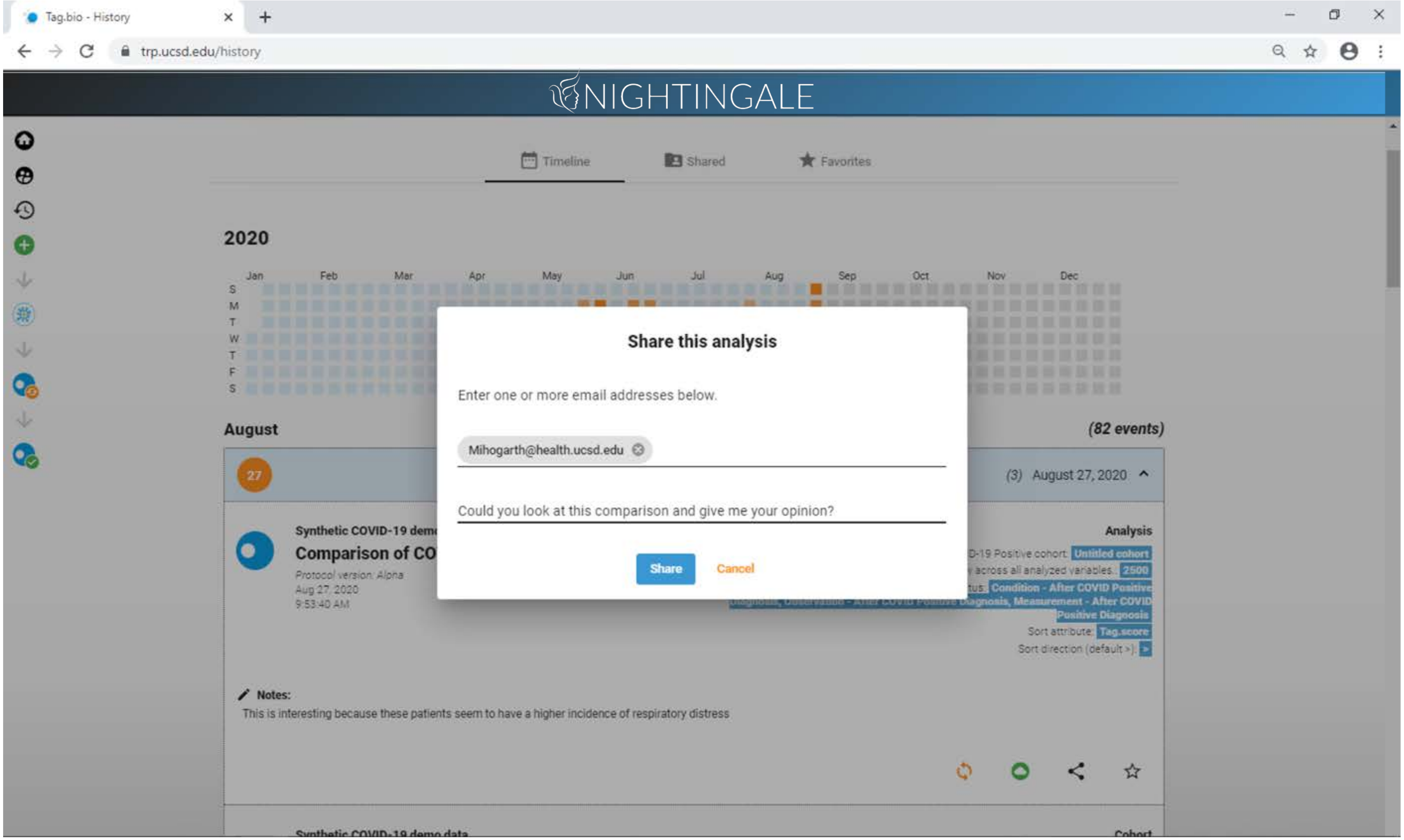
Protocol version: Alpha
Aug 27, 2020
9:53:40 AM

Analysis

COVID-19 Positive cohort: **Untitled cohort**
Define a maximum number of results to view across all analyzed variables.: **2500**
Outcomes recorded only after COVID positive status: **Condition - After COVID Positive Diagnosis, Observation - After COVID Positive Diagnosis, Measurement - After COVID Positive Diagnosis**
Sort attribute: **Tag.score**
Sort direction (default >): **>**

Notes:
Click here to add notes.

Synthetic COVID-19 demo data **Cohort**



August

(82 events)

27

(3) August 27, 2020

Synthetic COVID-19 demo

Comparison of CO

Protocol version: Alpha
Aug 27, 2020
9:53:40 AM

Analysis

D-19 Positive cohort: Untitled cohort
across all analyzed variables.: 2500
tus: Condition - After COVID Positive Diagnosis, Observation - After COVID Positive Diagnosis, Measurement - After COVID Positive Diagnosis
Sort attribute: Tag.score
Sort direction (default >): >

Notes:
This is interesting because these patients seem to have a higher incidence of respiratory distress

Synthetic COVID-19 demo data

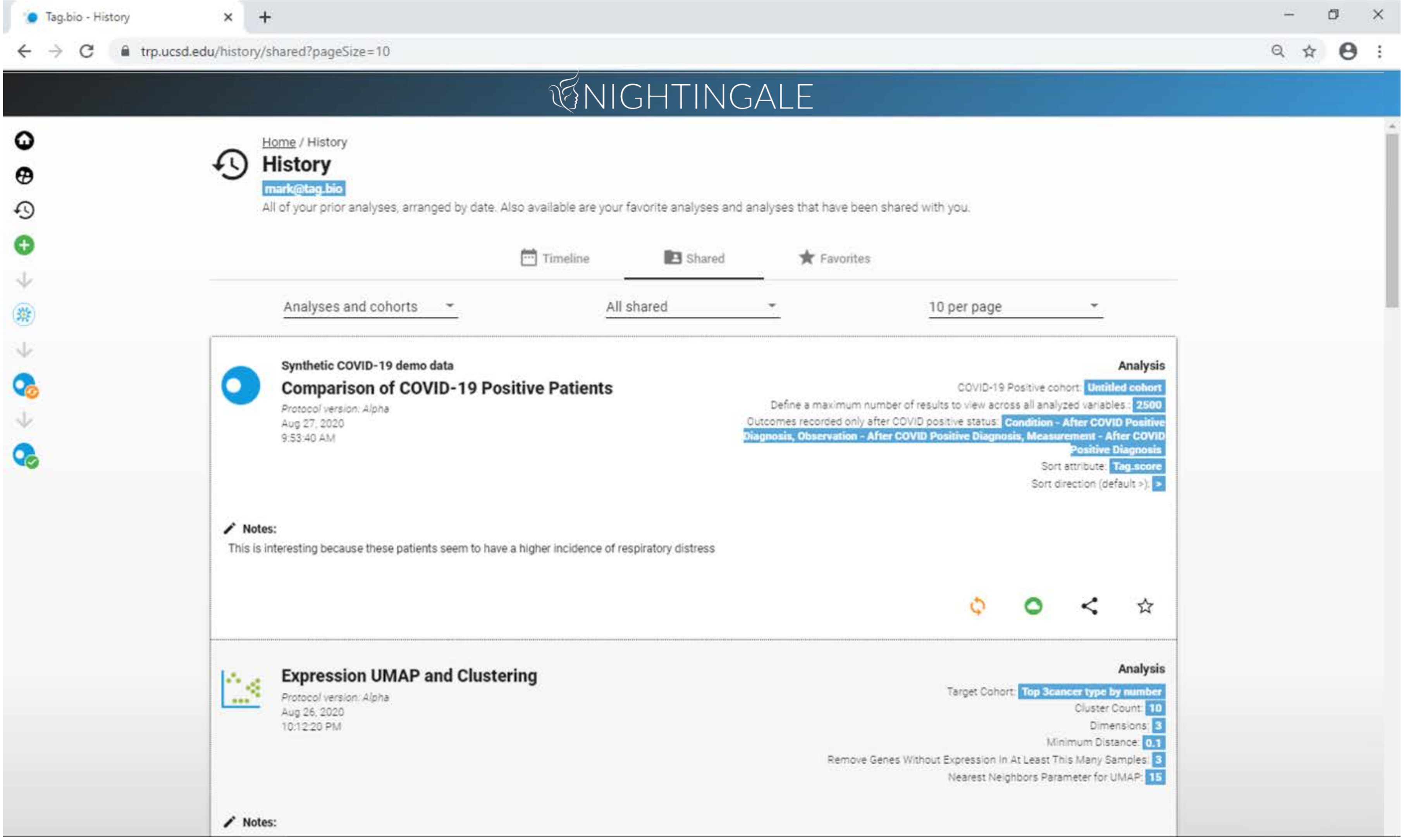
Cohort

Share this analysis

Enter one or more email addresses below.

Mihogarth@health.ucsd.edu ⊗

Could you look at this comparison and give me your opinion?



Home / History

History

mark@tag.bio

All of your prior analyses, arranged by date. Also available are your favorite analyses and analyses that have been shared with you.

Synthetic COVID-19 demo data

Comparison of COVID-19 Positive Patients

Protocol version: Alpha
Aug 27, 2020
9:53:40 AM

Analysis

COVID-19 Positive cohort: **Untitled cohort**
Define a maximum number of results to view across all analyzed variables.: **2500**
Outcomes recorded only after COVID positive status: **Condition - After COVID Positive Diagnosis, Observation - After COVID Positive Diagnosis, Measurement - After COVID Positive Diagnosis**
Sort attribute: **Tag.score**
Sort direction (default >): **>**

Notes:

This is interesting because these patients seem to have a higher incidence of respiratory distress

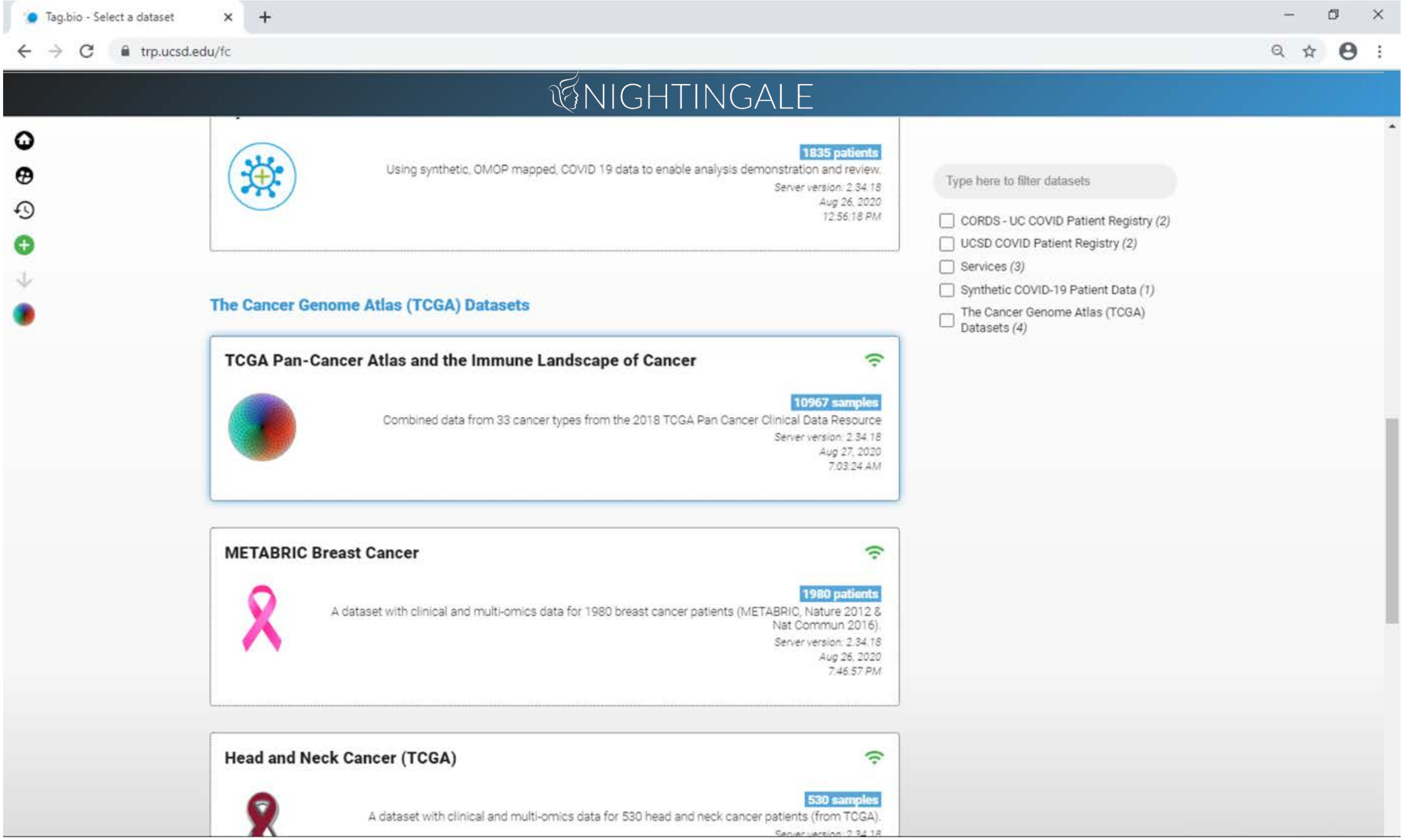
Expression UMAP and Clustering

Protocol version: Alpha
Aug 26, 2020
10:12:20 PM

Analysis

Target Cohort: **Top 3cancer type by number**
Cluster Count: **10**
Dimensions: **3**
Minimum Distance: **0.1**
Remove Genes Without Expression In At Least This Many Samples: **3**
Nearest Neighbors Parameter for UMAP: **15**

Notes:



1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

Server version: 2.34.18
Aug 26, 2020
12:56:18 PM

The Cancer Genome Atlas (TCGA) Datasets

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

10967 samples

Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

Server version: 2.34.18
Aug 27, 2020
7:03:24 AM

METABRIC Breast Cancer

1980 patients

A dataset with clinical and multi-omics data for 1980 breast cancer patients (METABRIC, Nature 2012 & Nat Commun 2016).

Server version: 2.34.18
Aug 26, 2020
7:46:57 PM

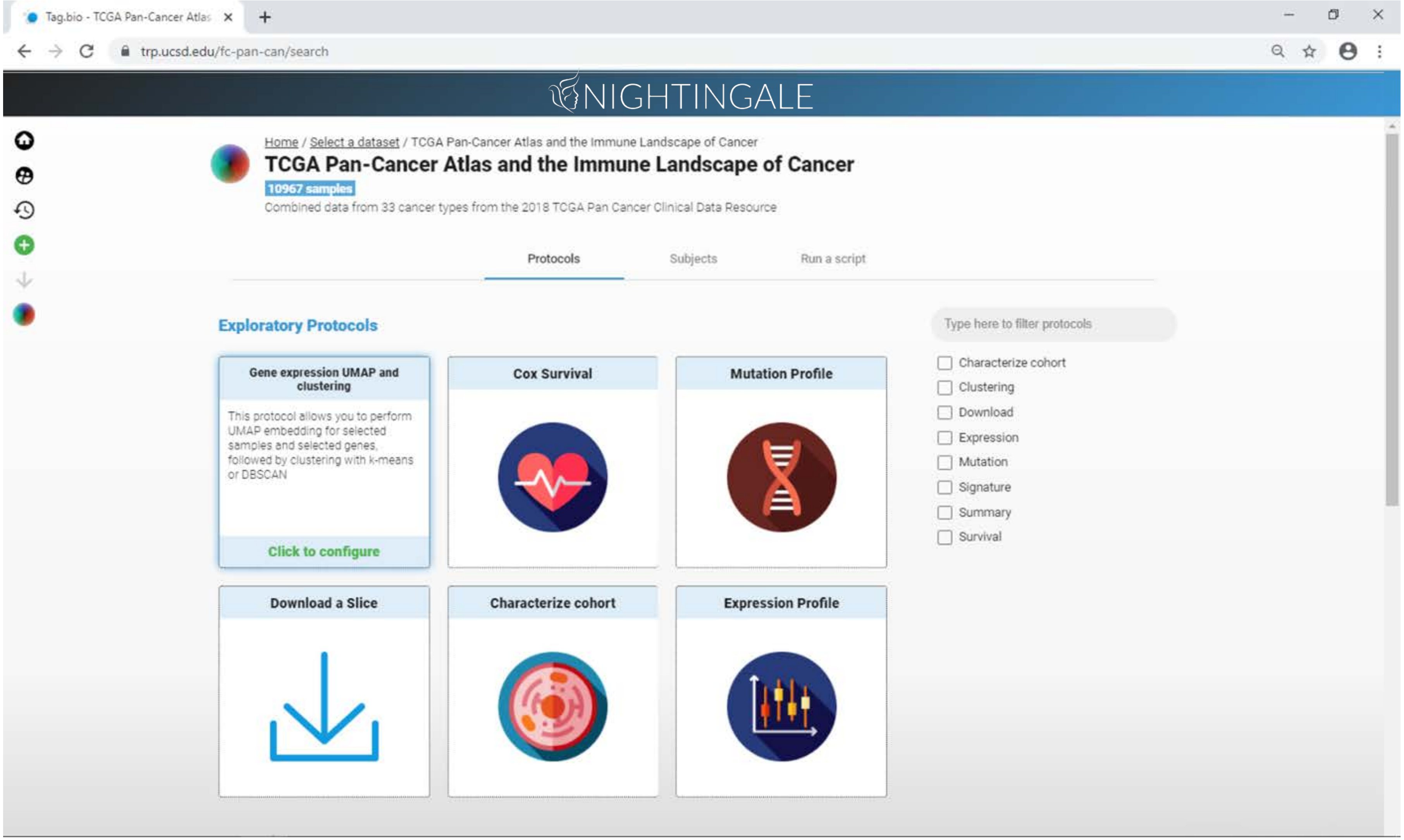
Head and Neck Cancer (TCGA)

530 samples

A dataset with clinical and multi-omics data for 530 head and neck cancer patients (from TCGA).

Server version: 2.34.18

☐ CORDS - UC COVID Patient Registry *(2)*
☐ UCSD COVID Patient Registry *(2)*
☐ Services *(3)*
☐ Synthetic COVID-19 Patient Data *(1)*
☐ The Cancer Genome Atlas (TCGA) Datasets *(4)*



Tag.bio - TCGA Pan-Cancer Atlas
trp.ucsd.edu/fc-pan-can/search
NIGHTINGALE

Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

10967 samples

Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

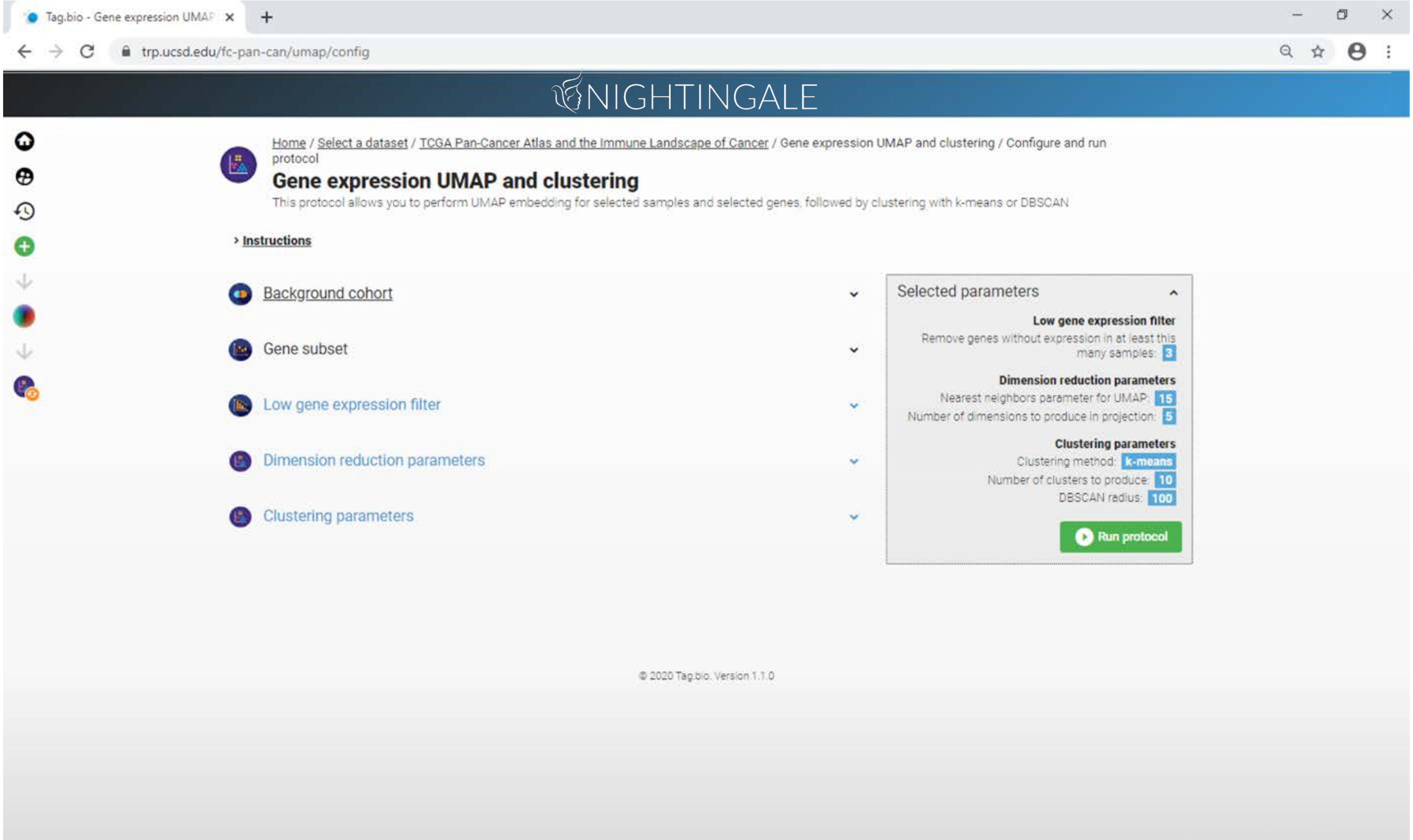
Protocols
Subjects
Run a script

Exploratory Protocols

Gene expression UMAP and clustering

This protocol allows you to perform UMAP embedding for selected samples and selected genes, followed by clustering with k-means or DBSCAN

- □ Characterize cohort
- □ Clustering
- □ Download
- □ Expression
- □ Mutation
- □ Signature
- □ Summary
- □ Survival


Tag.bio - Gene expression UMAP
trp.ucsd.edu/fc-pan-can/umap/config
NIGHTINGALE

Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer / Gene expression UMAP and clustering / Configure and run protocol

Gene expression UMAP and clustering

This protocol allows you to perform UMAP embedding for selected samples and selected genes, followed by clustering with k-means or DBSCAN

› Instructions

Background cohort

Gene subset

Low gene expression filter

Dimension reduction parameters

Clustering parameters

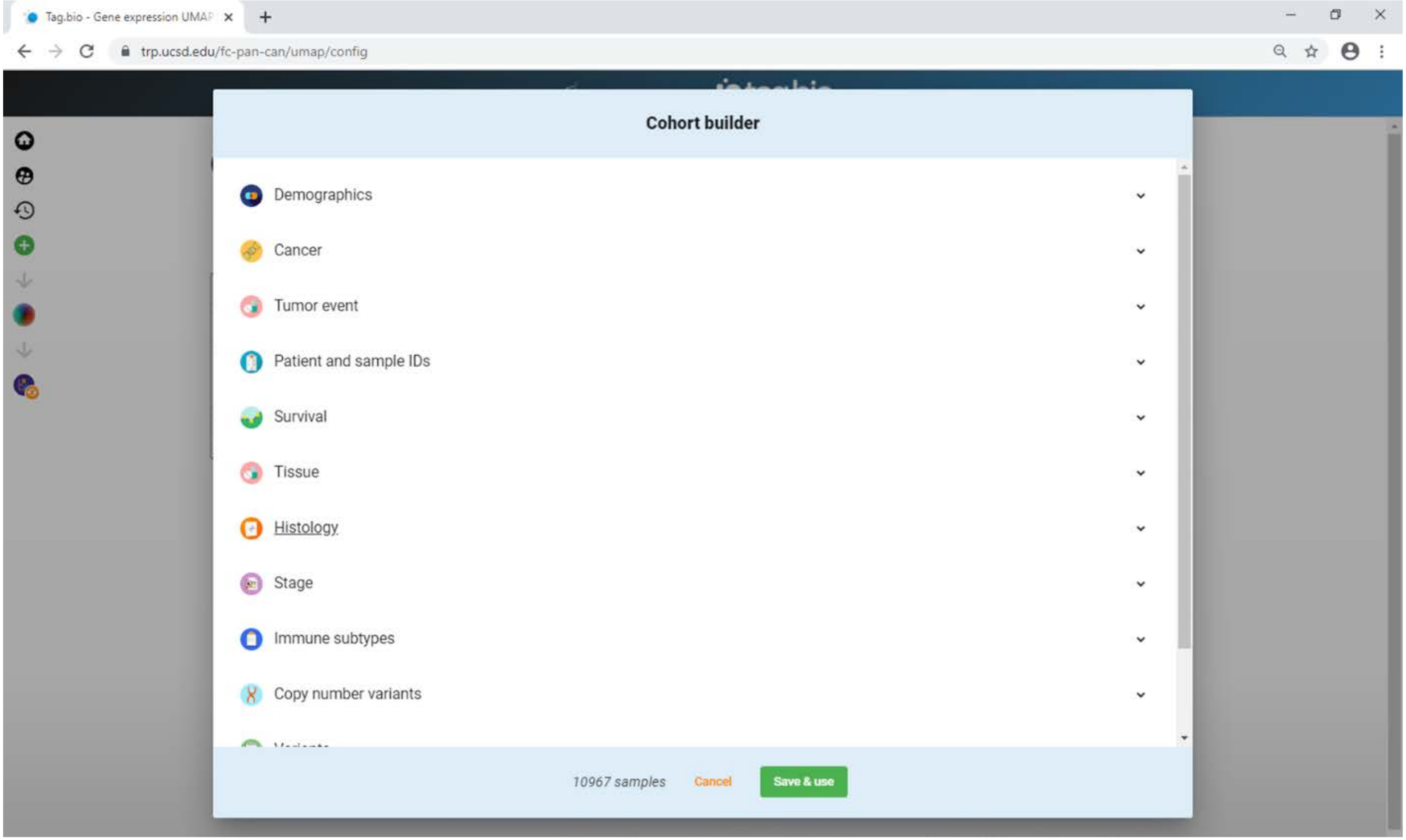
Selected parameters

Low gene expression filter
Remove genes without expression in at least this many samples: 3

Dimension reduction parameters
Nearest neighbors parameter for UMAP: 15
Number of dimensions to produce in projection: 5

Clustering parameters
Clustering method: k-means
Number of clusters to produce: 10
DBSCAN radius: 100

© 2020 Tag.bio. Version 1.1.0

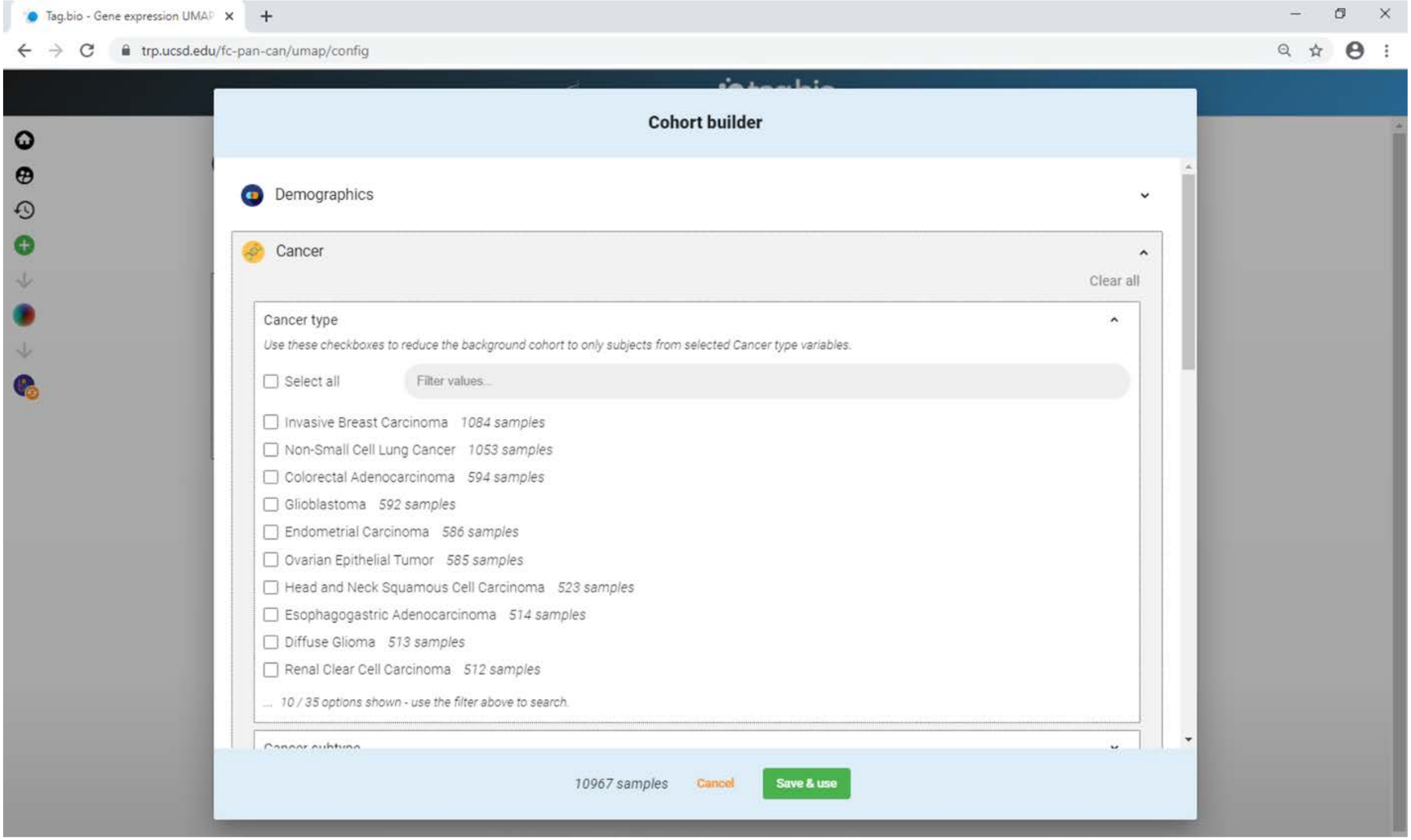


Tag.bio - Gene expression UMAP
trp.ucsd.edu/fc-pan-can/umap/config

Cohort builder

- Demographics
- Cancer
- Tumor event
- Patient and sample IDs
- Survival
- Tissue
- Histology
- Stage
- Immune subtypes
- Copy number variants
- Variants

10967 samples



Tag.bio - Gene expression UMAP
trp.ucsd.edu/fc-pan-can/umap/config

Cohort builder

Demographics

Cancer

Clear all

Cancer type

Use these checkboxes to reduce the background cohort to only subjects from selected Cancer type variables.

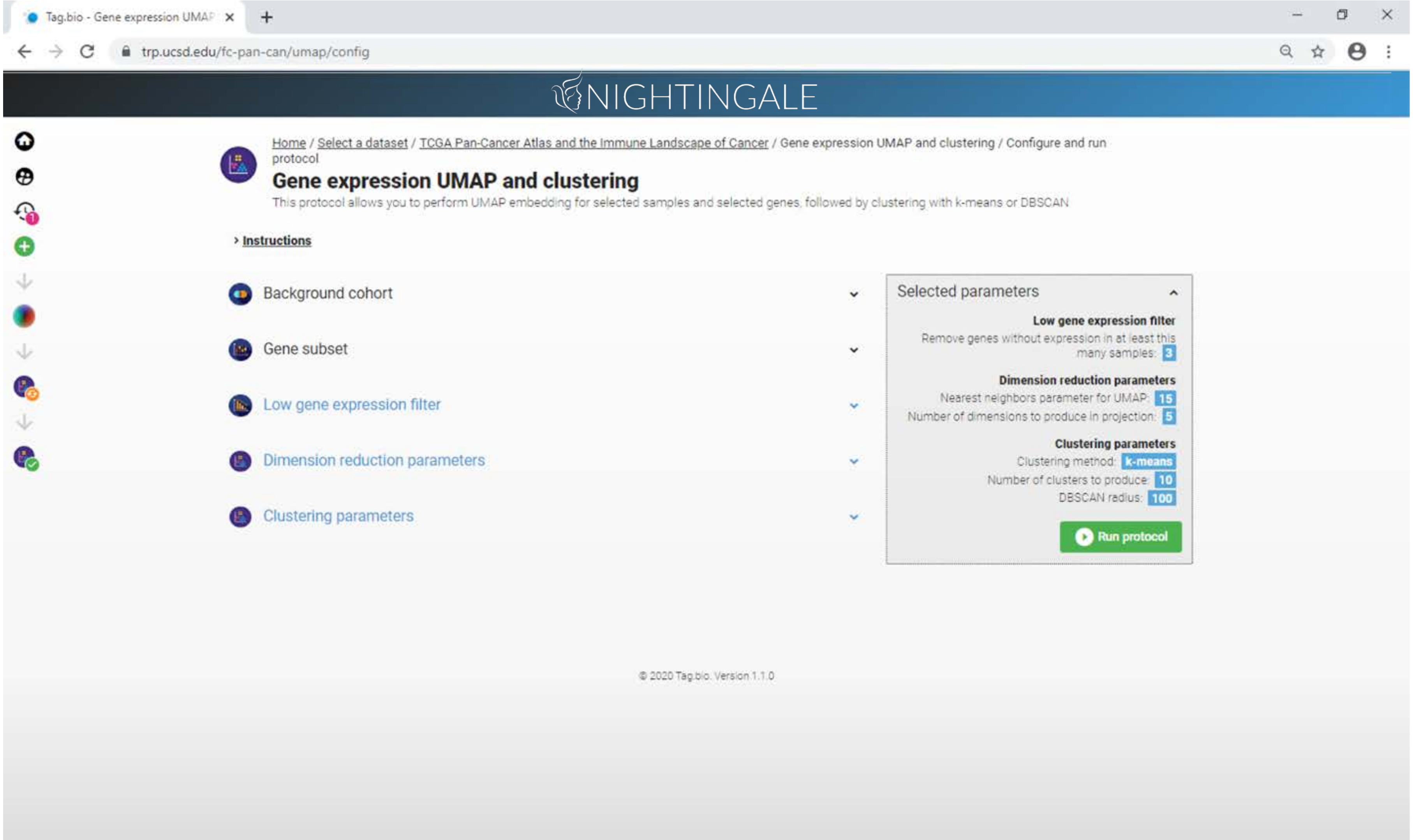
□ Select all   Filter values...

□ Invasive Breast Carcinoma *1084 samples*
□ Non-Small Cell Lung Cancer *1053 samples*
□ Colorectal Adenocarcinoma *594 samples*
□ Glioblastoma *592 samples*
□ Endometrial Carcinoma *586 samples*
□ Ovarian Epithelial Tumor *585 samples*
□ Head and Neck Squamous Cell Carcinoma *523 samples*
□ Esophagogastric Adenocarcinoma *514 samples*
□ Diffuse Glioma *513 samples*
□ Renal Clear Cell Carcinoma *512 samples*

... 10 / 35 options shown - use the filter above to search.

Cancer subtype

10967 samples   Cancel   Save & use


Tag.bio - Gene expression UMAP
trp.ucsd.edu/fc-pan-can/umap/config
NIGHTINGALE

Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer / Gene expression UMAP and clustering / Configure and run protocol

Gene expression UMAP and clustering

This protocol allows you to perform UMAP embedding for selected samples and selected genes, followed by clustering with k-means or DBSCAN

› Instructions

- Background cohort
- Gene subset
- Low gene expression filter
- Dimension reduction parameters
- Clustering parameters

Selected parameters

Low gene expression filter

Remove genes without expression in at least this many samples: 3

Dimension reduction parameters

Nearest neighbors parameter for UMAP: 15

Number of dimensions to produce in projection: 5

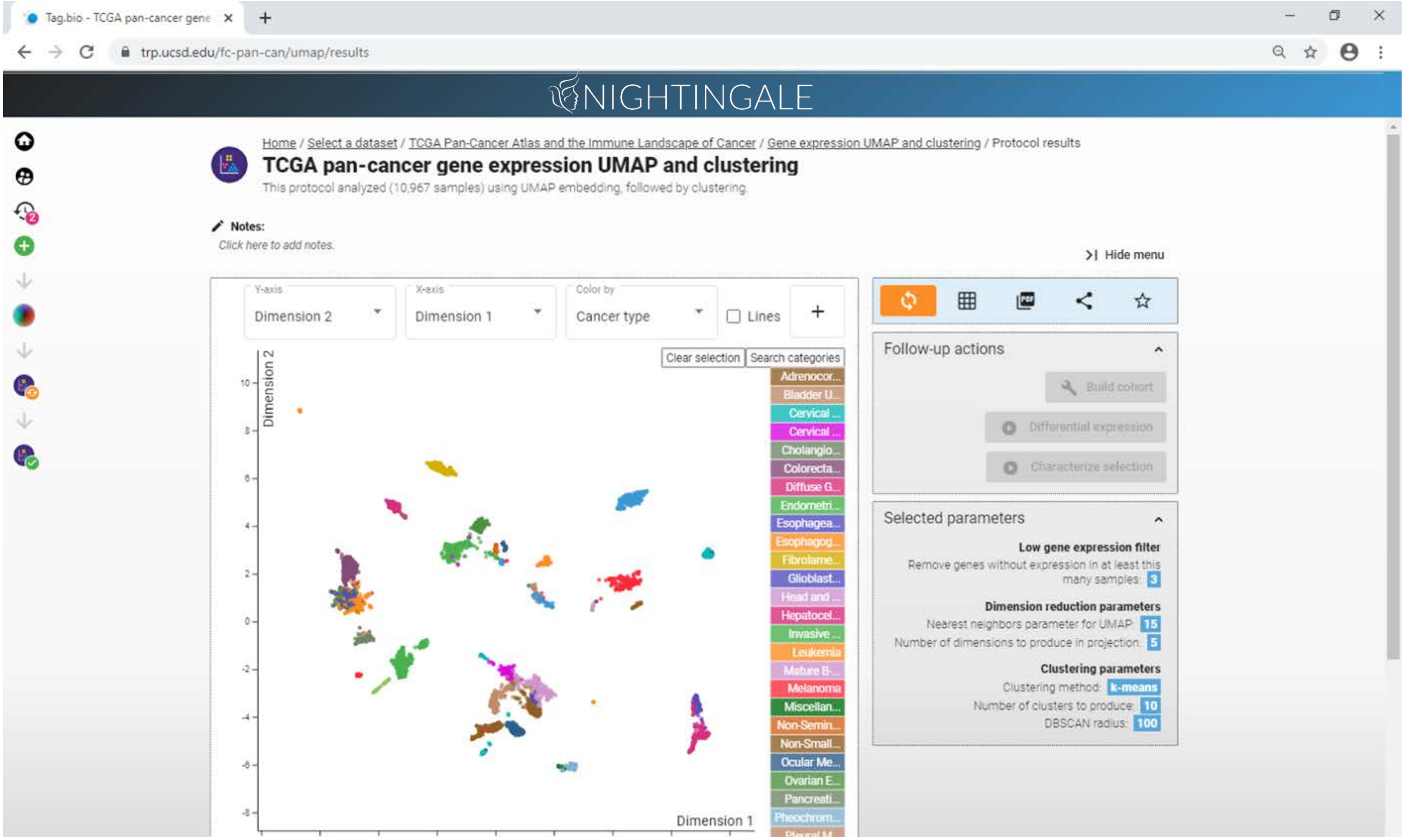
Clustering parameters

Clustering method: k-means

Number of clusters to produce: 10

DBSCAN radius: 100

© 2020 Tag.bio. Version 1.1.0



Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer / Gene expression UMAP and clustering / Protocol results

TCGA pan-cancer gene expression UMAP and clustering

This protocol analyzed (10,967 samples) using UMAP embedding, followed by clustering.

Notes:

Click here to add notes.

Y-axis
Dimension 2

X-axis
Dimension 1

Color by
Cancer type

□ Lines

Follow-up actions

Selected parameters

Low gene expression filter

Remove genes without expression in at least this many samples: 3

Dimension reduction parameters

Nearest neighbors parameter for UMAP: 15

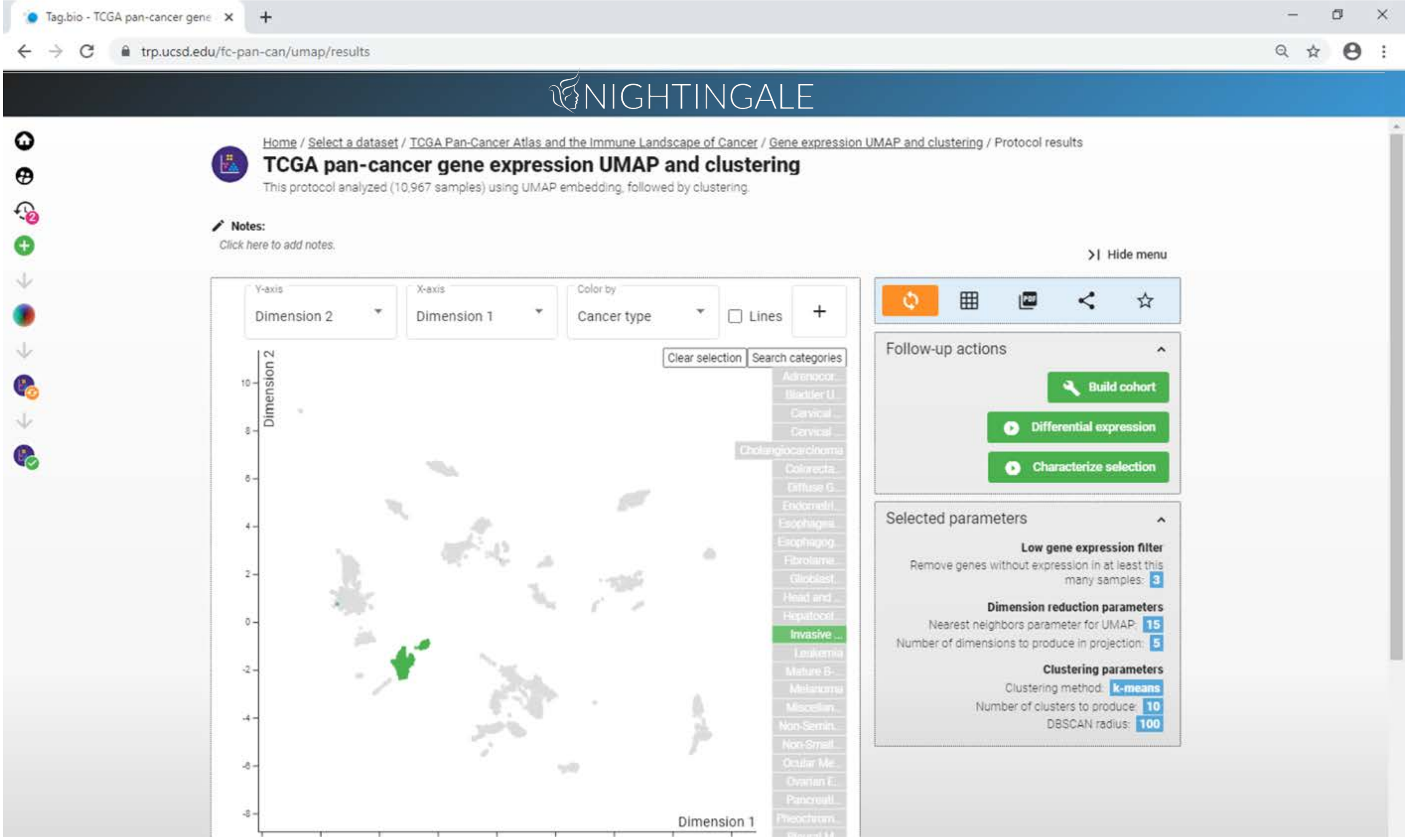
Number of dimensions to produce in projection: 5

Clustering parameters

Clustering method: k-means

Number of clusters to produce: 10

DBSCAN radius: 100



Tag.bio - TCGA pan-cancer gene
trp.ucsd.edu/fc-pan-can/umap/results
NIGHTINGALE
Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer / Gene expression UMAP and clustering / Protocol results

TCGA pan-cancer gene expression UMAP and clustering

This protocol analyzed (10,967 samples) using UMAP embedding, followed by clustering.

Notes:

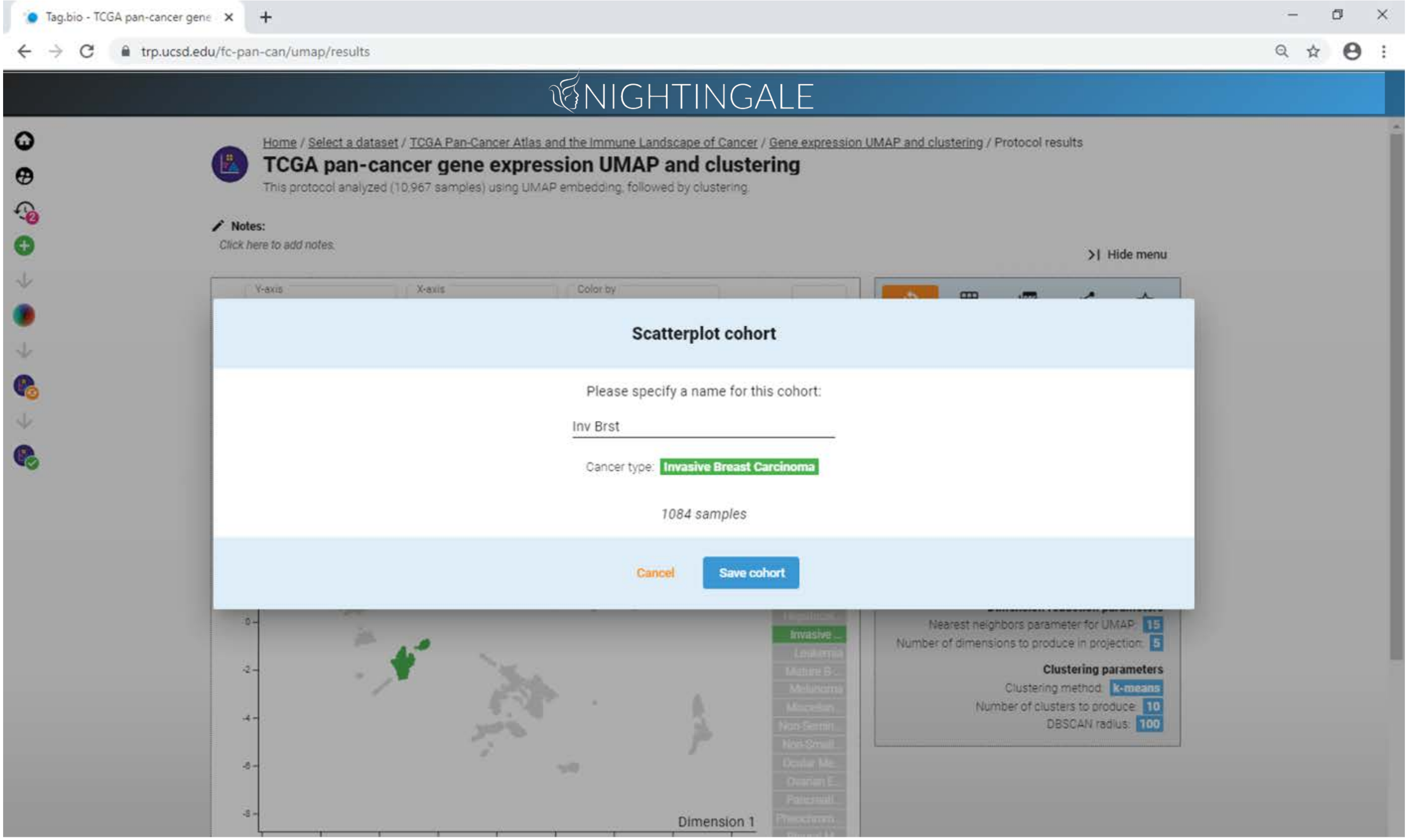
Click here to add notes.

Selected parameters

Low gene expression filter
Remove genes without expression in at least this many samples: 3

Dimension reduction parameters
Nearest neighbors parameter for UMAP: 15
Number of dimensions to produce in projection: 5

Clustering parameters
Clustering method: k-means
Number of clusters to produce: 10
DBSCAN radius: 100



Tag.bio - TCGA pan-cancer gene
trp.ucsd.edu/fc-pan-can/umap/results
NIGHTINGALE

Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer / Gene expression UMAP and clustering / Protocol results

TCGA pan-cancer gene expression UMAP and clustering

This protocol analyzed (10,967 samples) using UMAP embedding, followed by clustering.

Notes:

Click here to add notes.

>| Hide menu
Y-axis
X-axis
Color by

Scatterplot cohort

Please specify a name for this cohort:

Inv Brst

Cancer type: Invasive Breast Carcinoma

1084 samples

Nearest neighbors parameter for UMAP: 15

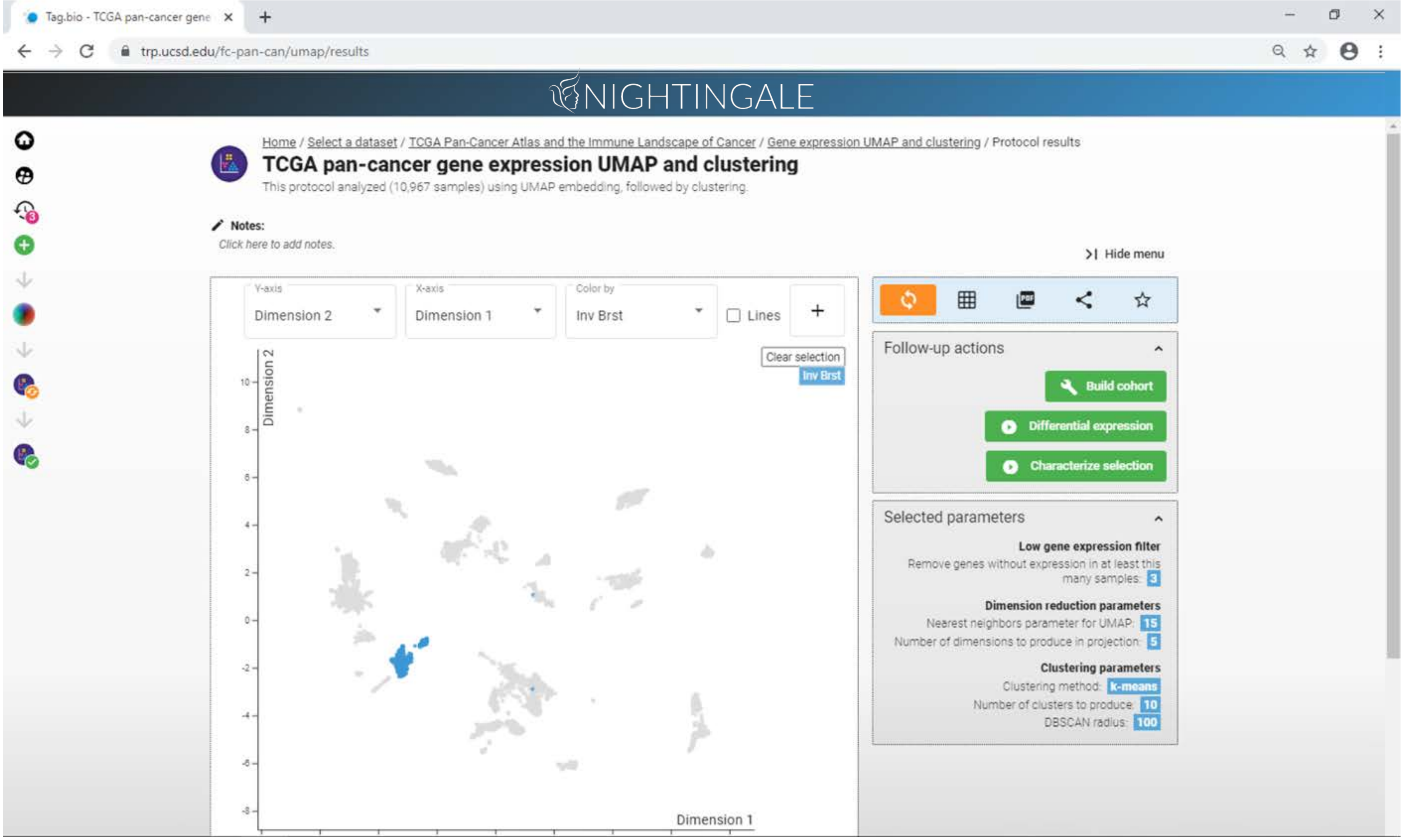
Number of dimensions to produce in projection: 5

Clustering parameters

Clustering method: k-means

Number of clusters to produce: 10

DBSCAN radius: 100


Tag.bio - TCGA pan-cancer gene
trp.ucsd.edu/fc-pan-can/umap/results
NIGHTINGALE

Home / Select a dataset / TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer / Gene expression UMAP and clustering / Protocol results

TCGA pan-cancer gene expression UMAP and clustering

This protocol analyzed (10,967 samples) using UMAP embedding, followed by clustering.

Notes:

Click here to add notes.

Y-axis
Dimension 2

X-axis
Dimension 1

Color by
Inv Brst

□ Lines

Selected parameters

Low gene expression filter

Remove genes without expression in at least this many samples: 3

Dimension reduction parameters

Nearest neighbors parameter for UMAP: 15

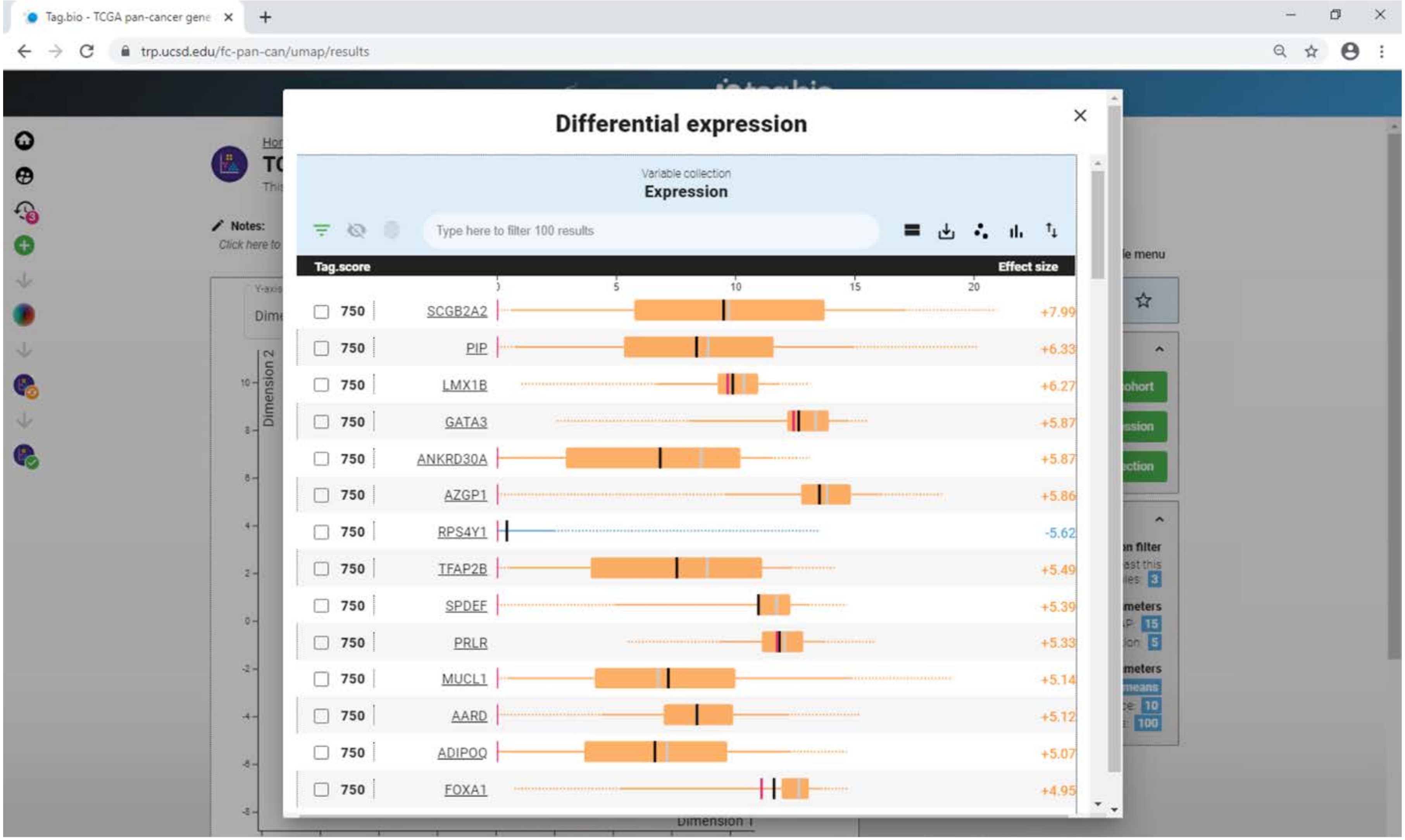
Number of dimensions to produce in projection: 5

Clustering parameters

Clustering method: k-means

Number of clusters to produce: 10

DBSCAN radius: 100

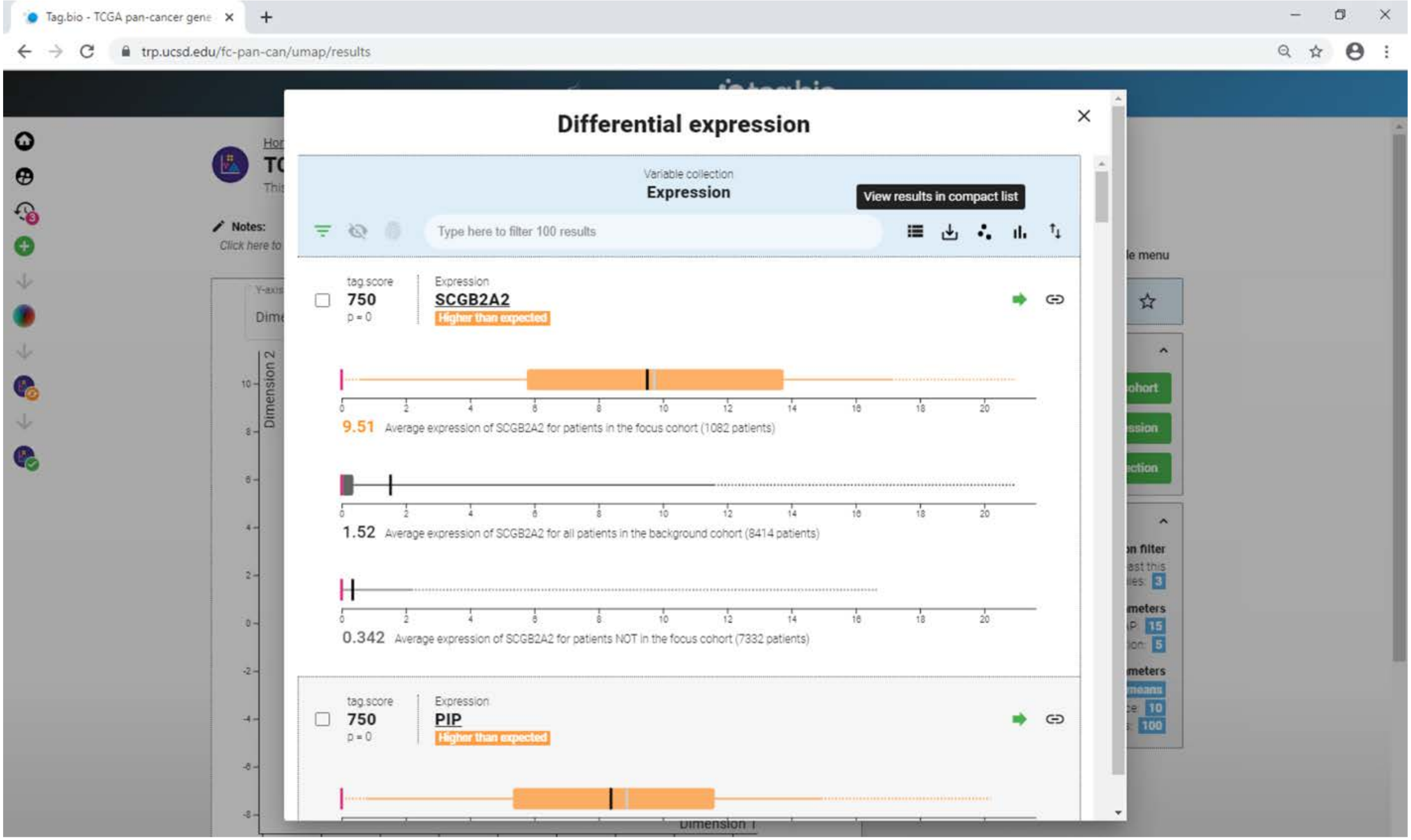


Differential expression

Variable collection
Expression

Type here to filter 100 results

	Tag.score		Effect size
□	750	SCGB2A2	+7.99
□	750	PIP	+6.33
□	750	LMX1B	+6.27
□	750	GATA3	+5.87
□	750	ANKRD30A	+5.87
□	750	AZGP1	+5.86
□	750	RPS4Y1	-5.62
□	750	TFAP2B	+5.49
□	750	SPDEF	+5.39
□	750	PRLR	+5.33
□	750	MUCL1	+5.14
□	750	AARD	+5.12
□	750	ADIPOQ	+5.07
□	750	FOXA1	+4.95



Tag.bio - TCGA pan-cancer gene

trp.ucsd.edu/fc-pan-can/umap/results

Differential expression

Variable collection

Expression

View results in compact list

Type here to filter 100 results

☐ tag.score **750** p = 0

Expression **SCGB2A2** Higher than expected

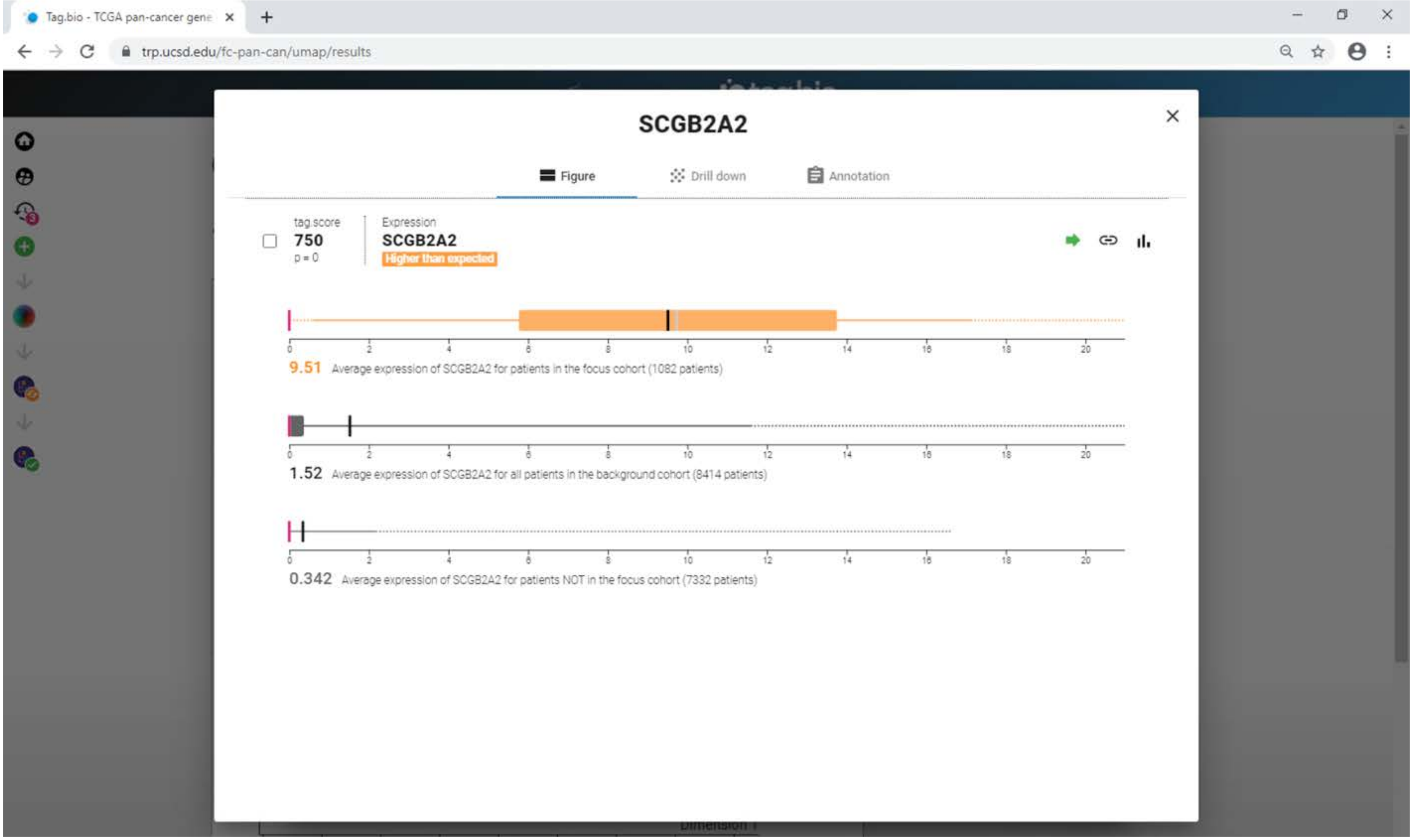
9.51 Average expression of SCGB2A2 for patients in the focus cohort (1082 patients)

1.52 Average expression of SCGB2A2 for all patients in the background cohort (8414 patients)

0.342 Average expression of SCGB2A2 for patients NOT in the focus cohort (7332 patients)

☐ tag.score **750** p = 0

Expression **PIP** Higher than expected


Tag.bio - TCGA pan-cancer gene
trp.ucsd.edu/fc-pan-can/umap/results
Figure
Drill down
Annotation
0
2
4
6
8
10
12
14
16
18
20

SCGB2A2

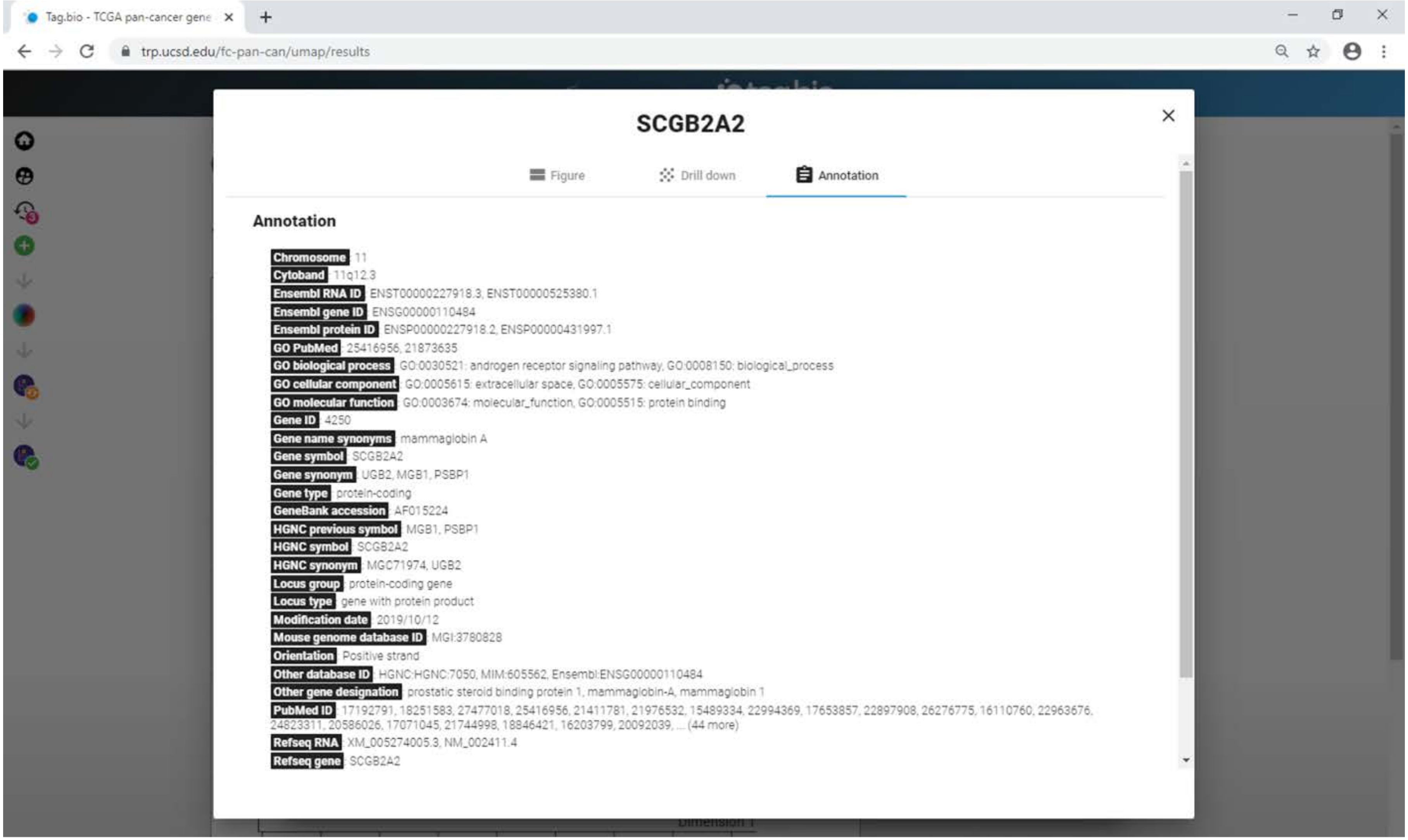
□ tag.score 750 p = 0

Expression SCGB2A2 Higher than expected

9.51 Average expression of SCGB2A2 for patients in the focus cohort (1082 patients)

1.52 Average expression of SCGB2A2 for all patients in the background cohort (8414 patients)

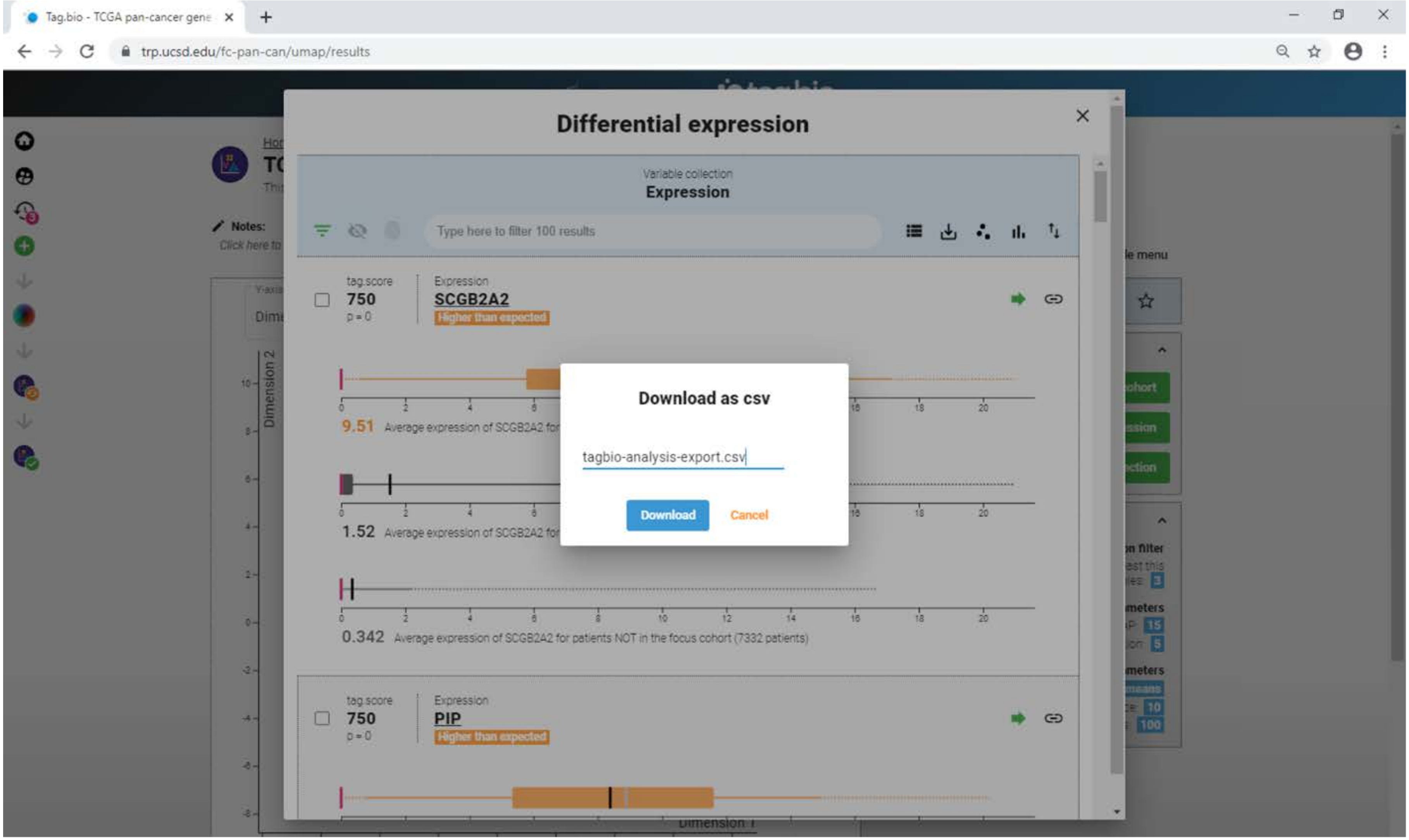
0.342 Average expression of SCGB2A2 for patients NOT in the focus cohort (7332 patients)


Tag.bio - TCGA pan-cancer gene
trp.ucsd.edu/fc-pan-can/umap/results
Figure
Drill down
Annotation
Dimension 1

SCGB2A2

Annotation

Chromosome: 11
Cytoband: 11q12.3
Ensembl RNA ID: ENST00000227918.3, ENST00000525380.1
Ensembl gene ID: ENSG00000110484
Ensembl protein ID: ENSP00000227918.2, ENSP00000431997.1
GO PubMed: 25416956, 21873635
GO biological process: GO:0030521: androgen receptor signaling pathway, GO:0008150: biological_process
GO cellular component: GO:0005615: extracellular space, GO:0005575: cellular_component
GO molecular function: GO:0003674: molecular_function, GO:0005515: protein binding
Gene ID: 4250
Gene name synonyms: mammaglobin A
Gene symbol: SCGB2A2
Gene synonym: UGB2, MGB1, PSBP1
Gene type: protein-coding
GeneBank accession: AF015224
HGNC previous symbol: MGB1, PSBP1
HGNC symbol: SCGB2A2
HGNC synonym: MGC71974, UGB2
Locus group: protein-coding gene
Locus type: gene with protein product
Modification date: 2019/10/12
Mouse genome database ID: MGI:3780828
Orientation: Positive strand
Other database ID: HGNC:HGNC:7050, MIM:605562, Ensembl:ENSG00000110484
Other gene designation: prostatic steroid binding protein 1, mammaglobin-A, mammaglobin 1
PubMed ID: 17192791, 18251583, 27477018, 25416956, 21411781, 21976532, 15489334, 22994369, 17653857, 22897908, 26276775, 16110760, 22963676, 24823311, 20586026, 17071045, 21744998, 18846421, 16203799, 20092039, ... (44 more)
Refseq RNA: XM_005274005.3, NM_002411.4
Refseq gene: SCGB2A2



Tag.bio - TCGA pan-cancer gene

trp.ucsd.edu/fc-pan-can/umap/results

Differential expression

Variable collection

Expression

Type here to filter 100 results

tag.score 750 p = 0

Expression SCGB2A2 Higher than expected

9.51 Average expression of SCGB2A2 for

1.52 Average expression of SCGB2A2 for

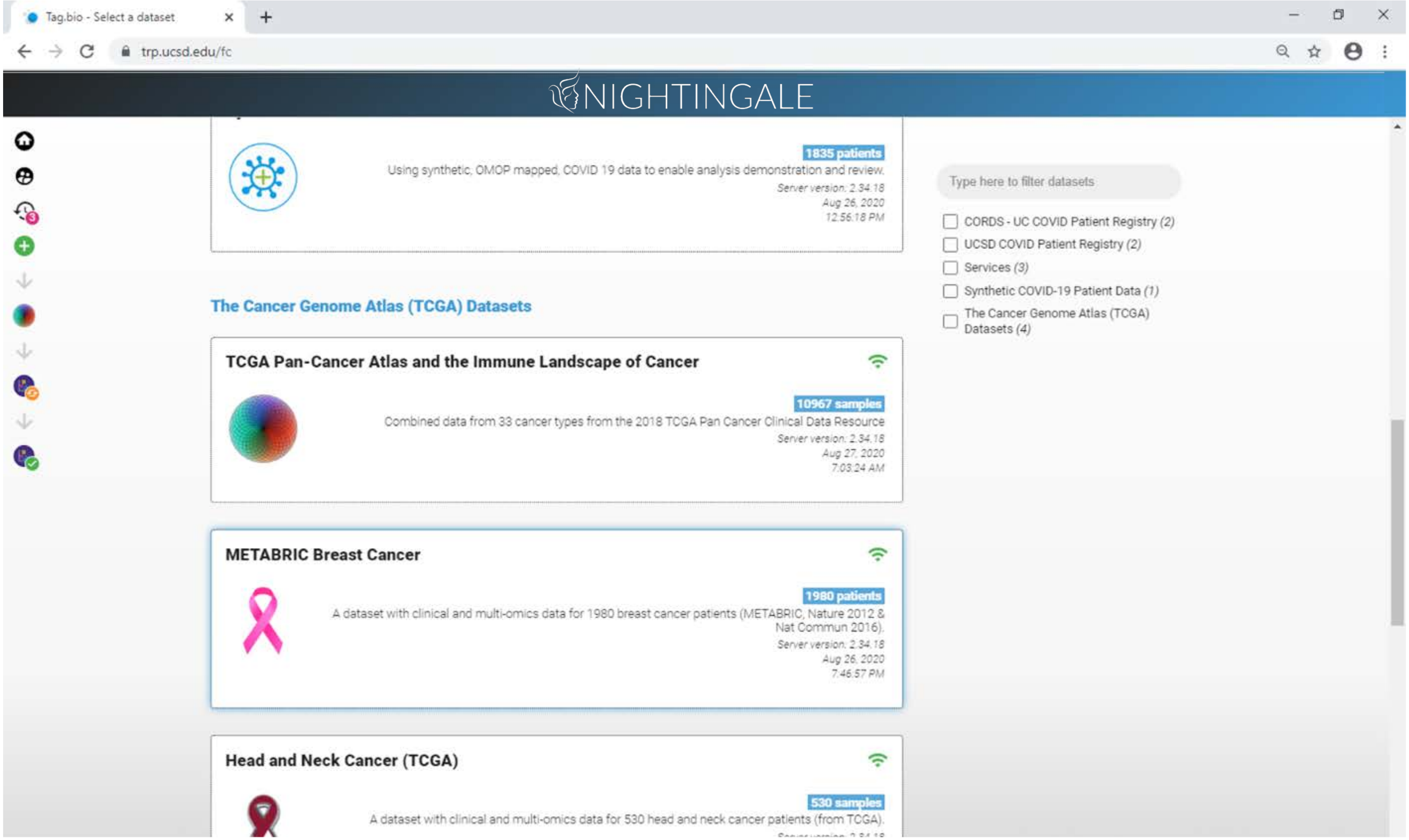
0.342 Average expression of SCGB2A2 for patients NOT in the focus cohort (7332 patients)

tag.score 750 p = 0

Expression PIP Higher than expected

Download as csv

tagbio-analysis-export.csv



1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

Server version: 2.34.18
Aug 26, 2020
12:56:18 PM

The Cancer Genome Atlas (TCGA) Datasets

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

10967 samples

Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

Server version: 2.34.18
Aug 27, 2020
7:03:24 AM

METABRIC Breast Cancer

1980 patients

A dataset with clinical and multi-omics data for 1980 breast cancer patients (METABRIC, Nature 2012 & Nat Commun 2016).

Server version: 2.34.18
Aug 26, 2020
7:46:57 PM

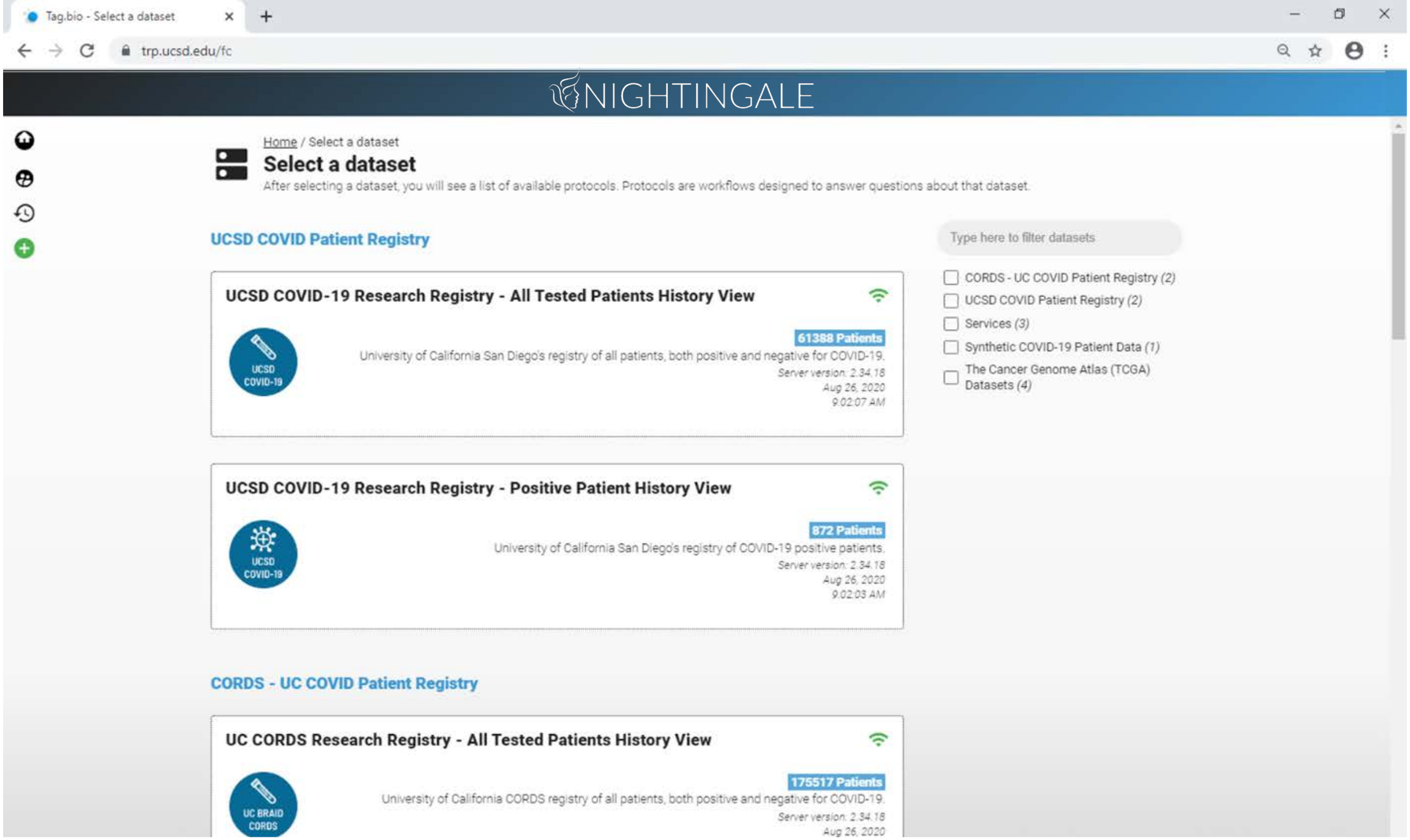
Head and Neck Cancer (TCGA)

530 samples

A dataset with clinical and multi-omics data for 530 head and neck cancer patients (from TCGA).

Type here to filter datasets

- □ CORDS - UC COVID Patient Registry *(2)*
- □ UCSD COVID Patient Registry *(2)*
- □ Services *(3)*
- □ Synthetic COVID-19 Patient Data *(1)*
- □ The Cancer Genome Atlas (TCGA) Datasets *(4)*



Home / Select a dataset

Select a dataset

After selecting a dataset, you will see a list of available protocols. Protocols are workflows designed to answer questions about that dataset.

Type here to filter datasets

□ CORDS - UC COVID Patient Registry *(2)*
□ UCSD COVID Patient Registry *(2)*
□ Services *(3)*
□ Synthetic COVID-19 Patient Data *(1)*
□ The Cancer Genome Atlas (TCGA) Datasets *(4)*

UCSD COVID Patient Registry

UCSD COVID-19 Research Registry - All Tested Patients History View

61388 Patients

University of California San Diego's registry of all patients, both positive and negative for COVID-19.
Server version: 2.34.18
Aug 26, 2020
9:02:07 AM

UCSD COVID-19 Research Registry - Positive Patient History View

872 Patients

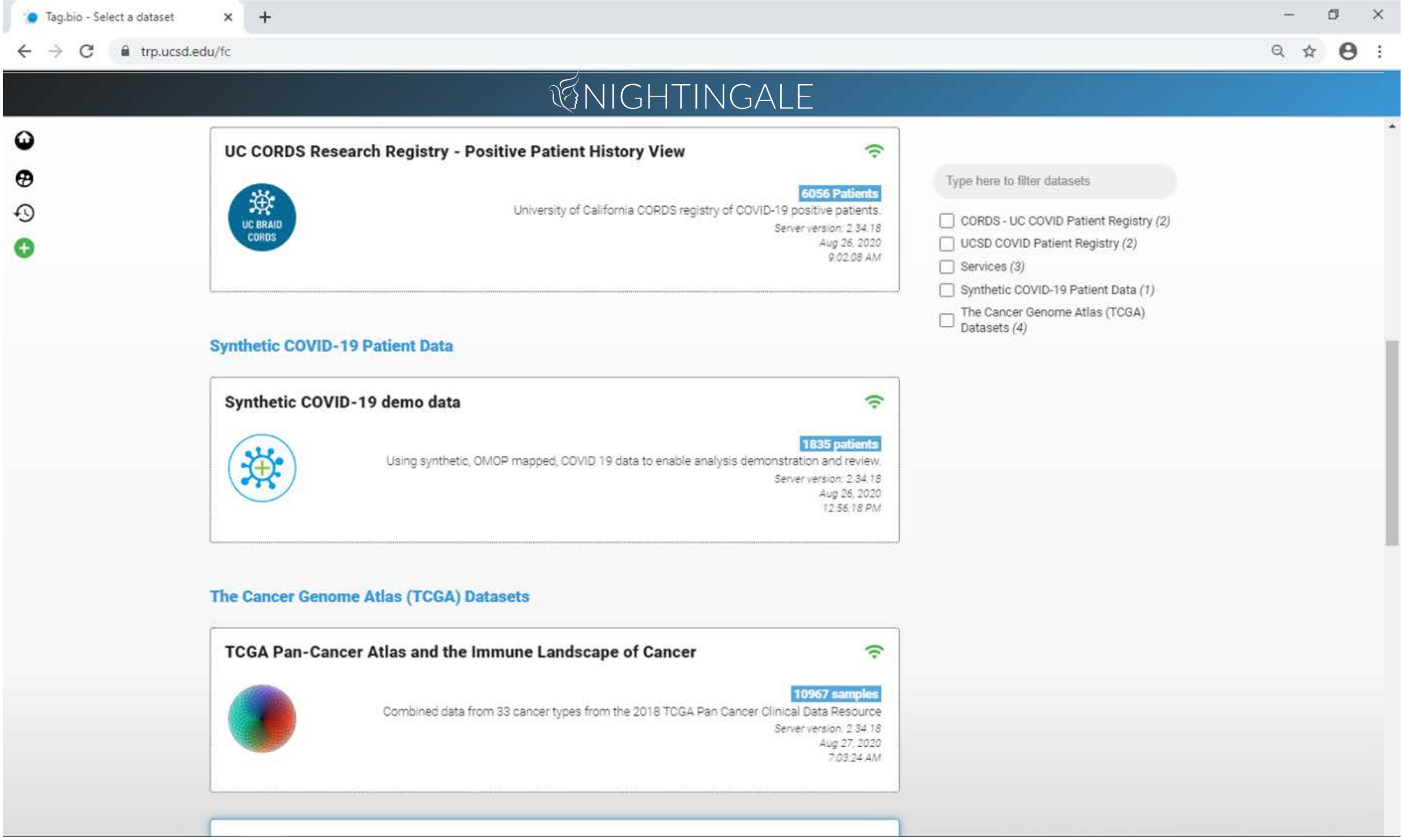
University of California San Diego's registry of COVID-19 positive patients.
Server version: 2.34.18
Aug 26, 2020
9:02:03 AM

CORDS - UC COVID Patient Registry

UC CORDS Research Registry - All Tested Patients History View

175517 Patients

University of California CORDS registry of all patients, both positive and negative for COVID-19.
Server version: 2.34.18
Aug 26, 2020



UC CORDS Research Registry - Positive Patient History View

6056 Patients

University of California CORDS registry of COVID-19 positive patients.

Server version: 2.34.18
Aug 26, 2020
9:02:08 AM

Synthetic COVID-19 Patient Data

Synthetic COVID-19 demo data

1835 patients

Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.

Server version: 2.34.18
Aug 26, 2020
12:56:18 PM

The Cancer Genome Atlas (TCGA) Datasets

TCGA Pan-Cancer Atlas and the Immune Landscape of Cancer

10967 samples

Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource

Server version: 2.34.18
Aug 27, 2020
7:03:24 AM

Type here to filter datasets

- ☐ CORDS - UC COVID Patient Registry *(2)*
- ☐ UCSD COVID Patient Registry *(2)*
- ☐ Services *(3)*
- ☐ Synthetic COVID-19 Patient Data *(1)*
- ☐ The Cancer Genome Atlas (TCGA) Datasets *(4)*

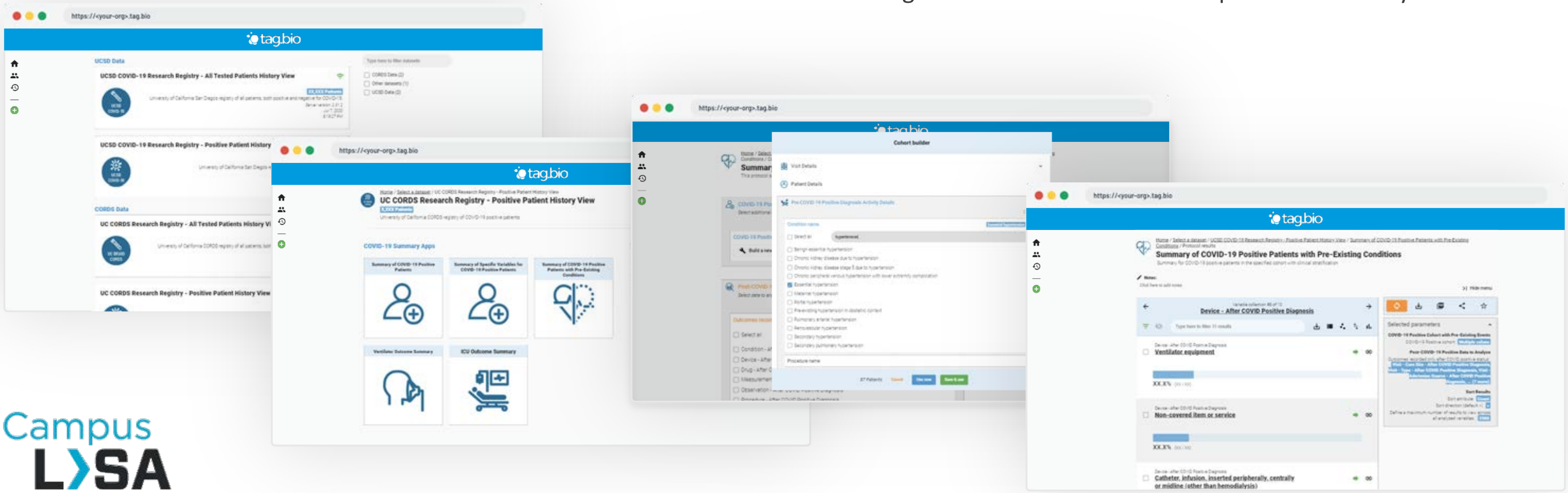
Summary

Mike Hogarth MD, Clinical Research Information Officer, UC San Diego Health

NIGHTINGALE

A tool for data exploration and analysis

- We have installed the tag.bio system in our research cloud and it has access to data sets in our ‘secure data commons database’
- The Nightingale portal provides population level access and ability to perform analysis
- A user can ‘slice’ the cohort and select specific analyses (demographic, survival, comparison between cohorts)
- Planned, pending approval, provide ‘download’ of limited data set (LDS) row-level data from selected data set into the investigator’s virtual research desktop for further analysis


https://<your-org>.tag.bio
tag.bio
UCSD Data
UCSD COVID-19 Research Registry - All Tested Patients History View
UCSD COVID-19 Research Registry - Positive Patient History
CORDS Data
UC CORDS Research Registry - All Tested Patients History View
UC CORDS Research Registry - Positive Patient History View
UC CORDS Research Registry - Positive Patient History View
COVID-19 Summary Apps
Summary of COVID-19 Positive Patients
Summary of Specific Variables for COVID-19 Positive Patients
Summary of COVID-19 Positive Patients with Pre-Existing Conditions
Ventilator Outcome Summary
ICU Outcome Summary
Cohort builder
Visit Details
Patient Details
Summary of COVID-19 Positive Patients with Pre-Existing Conditions
Device - After COVID Positive Diagnosis
Ventilator equipment
Non-covered item or service
Catheter, infusion, inserted peripherally, centrally or midline (other than hemodialysis)
Selected parameters

Campus LISA

Help us evolve the mesh

- What other registries should be available?
- How would you like to query them?
- Are there public data sources you would like to see here?
- Could we use the mesh for other data sources?

Please contact Mike Hogarth at mihogarth@health.ucsd.edu with suggestions or comments.

Thank You!

Questions?

Next presentation

Come see our next UC TECH Presentation 9/03:

Email Overload: Practical Tools for Influencing Email Volume in the Age of Telecommuting

Are you overwhelmed by the number of emails you receive daily? Has email management become a burdensome core task that monopolizes your time? When volume exceeds 200+ emails a day, generic email tips & tricks for email management simply won’t cut it. This session will go beyond email platform use to focus on email management from a behavior modification and process improvement perspective. We will cover practical tools and strategies for actively managing virtual work and interactions with your co-workers more effectively, giving you the ability to actually influence the volume of emails you receive.

Speaker:

Loralyn Cross, Office of Research Affairs, UC San Diego

Campus LISA

Reference slides

Enabling doctors to provide instant answers

9 years of billing and encounter data

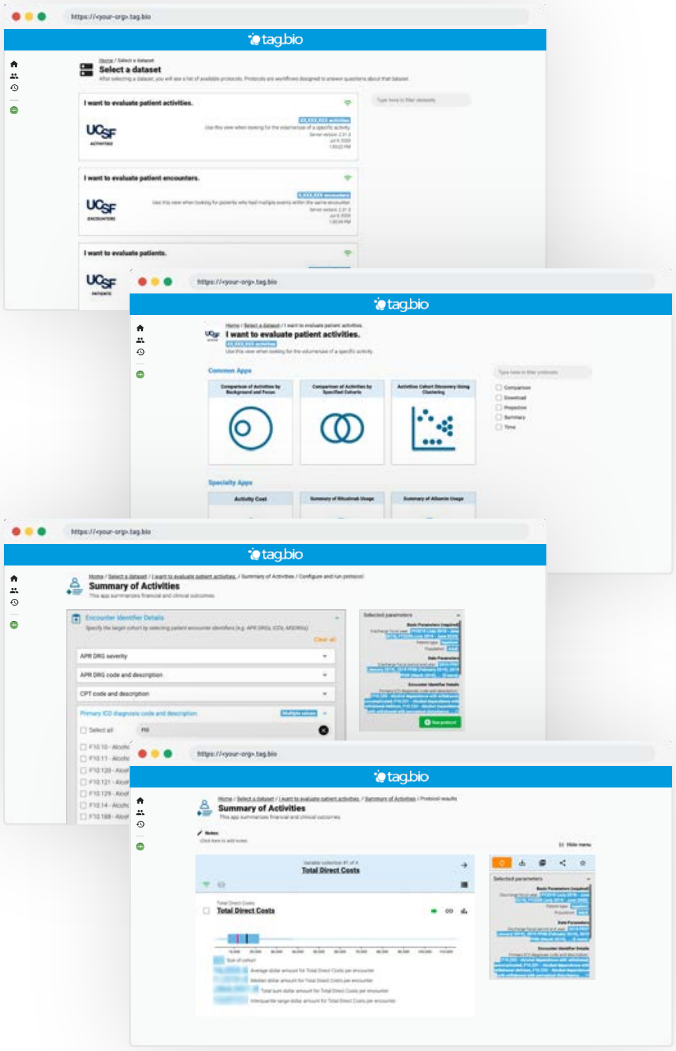
- All inpatient and outpatient data in a combined dataset

➔ Over **2,000 analyses** performed by value improvement physicians in the past year

> "The ability to have this kind of **on-demand information** completely **changes the culture**.
> I can't imagine doing my job without the Tag.bio platform."
>
> *- Jahan Fahimi, Director of Value Improvement at UCSF Health*

Campus LISA

Jahan Fahimi, MD, PhD,
Associate Professor of Emergency Medicine,
Director of Value Improvement at UCSF Health



https://<your-org>.tag.bio

Home / Select a dataset

Select a dataset

I want to evaluate patient activities.

I want to evaluate patient encounters.

I want to evaluate patients.

Home / Select a dataset / I want to evaluate patient activities.

I want to evaluate patient activities.

Common Apps

- Comparison of Activities by Background and Focus
- Comparison of Activities by Specified Cohorts
- Activities Cohort Discovery Using Clustering

□ Comparison
□ Download
□ Projection
□ Summary
□ Time

Specialty Apps

- Activity Cost

Home / Select a dataset / I want to evaluate patient activities. / Summary of Activities / Configure and run protocol

Summary of Activities

Encounter Identifier Details

APR DRG severity

APR DRG code and description

CPT code and description

Primary ICD diagnosis code and description

□ Select all

□ F10.10 - Alcoho
□ F10.11 - Alcoho
□ F10.120 - Alcoh
□ F10.121 - Alcoh
□ F10.129 - Alcoh
□ F10.14 - Alcoho
□ F10.188 - Alcoh

Selected parameters

Home / Select a dataset / I want to evaluate patient activities. / Summary of Activities / Protocol results

Summary of Activities

Notes

Hide menu

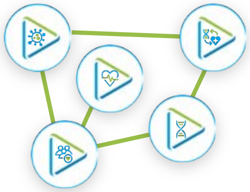
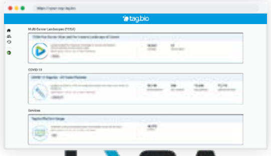
Total Direct Costs

Total Direct Costs

Total Direct Costs

Selected parameters

What does this enable?

	Data	Development	Collaboration
Node 	Is a Data Product Domain-specific analysis Immutable data and transient node Deploy anywhere	Build a rapid Data Product Iterative dev cycle Integrate other functions (R, Python, ML)	SMART API Secure data Secure deploy
Mesh 	Many Nodes, many data types Node functional diversity Distributed analyses Network effect on data value	Transfer apps between nodes Federated functionality No down time An ecosystem of nodes	Distributed querying Centralized analysis Public/Private nodes
Portal  	Publish data with analyses Track analyses through history Create COHORT/UDAT Reference & Annotation nodes	Rapidly populated by an admin Cohorts into Nodes Transactional Nodes Usage allows evolution of mesh	Reproducible, replayable analyses Share analyses Versioned resources De-silo analysis types