

Outreach and Training

29OCT2021



Jeff Leek and Tiffany Miller



Overview



- Mission
- Awareness
- Asynchronous Support
- Synchronous Support
- Tailored Support
- User Research
- Goal

2:30-3:45

Session 2: Breakout rooms

Infrastructure		Outreach and training	
Moderators: Ms. Karen L. Davis (RTI International) and Dr. Carolyn M. Hutter (NHGRI)		Moderators: Dr. Siddharth Pratap (Meharry Medical College) and Mr. Christopher Wellington (NHGRI)	
2:30-2:35	Moderator introduction	2:30-2:35	Moderator introduction
2:35-2:50	AnVIL presentation: <i>Dr. Jeremy Goecks (OHSU) and Dr. Benedict Paten (UCSC)</i>	2:35-2:50	AnVIL presentation: <i>Dr. Jeffrey Leek (JHU) and Ms. Tiffany Miller (Broad)</i>
2:50-3:35	Discussion	2:50-3:35	Discussion
3:35-3:45	Prepare breakout report	3:35-3:45	Prepare breakout report

Breakout room: Outreach and training

Moderators: Dr. Siddharth Pratap and Mr. Christopher Wellington

Dr. Cinnamon Bloss
Dr. C. Titus Brown
Dr. Carol Bult
Dr. John Kwagyan
Dr. Andrew Lee

Dr. Robert Meller
Dr. Peter Robinson
Dr. Sourav Roy
Dr. William (Bill) Southerland

Mission

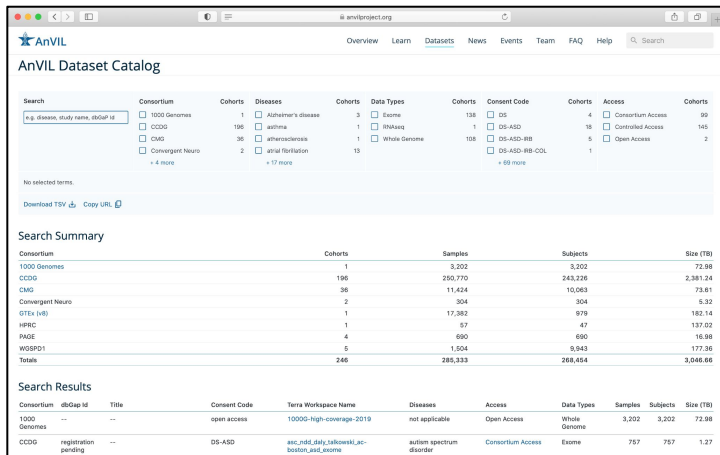
To develop scalable training, support, and incentives to make it easier for the scientific community to adopt and use the AnVIL Cloud Platform to share data, perform genomic and clinical analysis, & manage their computing.





Why AnVIL is awesome

Data!



AnVIL Dataset Catalog

Search:

Download TSV Copy URL

No selected terms

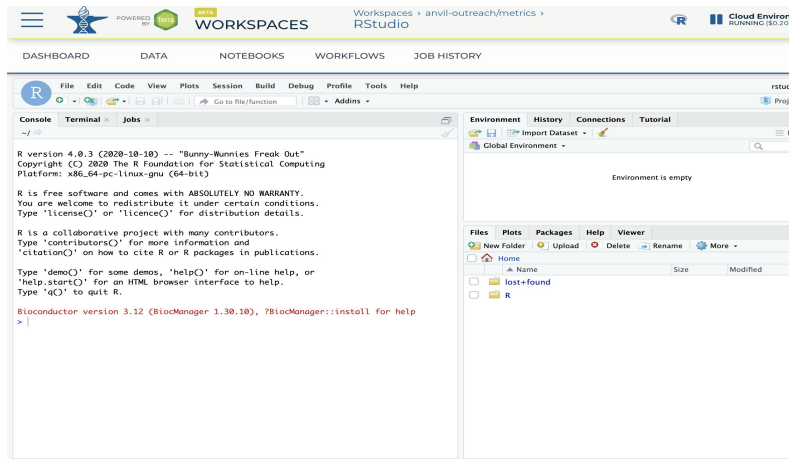
Search Summary

Consortium	Cohorts	Samples	Subjects	Size (TB)
1000 Genomes	1	3,202	3,202	72.98
CCDG	196	250,770	243,226	2,381.24
CMG	36	11,424	10,063	73.61
Convergent Neuro	2	304	304	5.32
GTEx v8	1	17,382	879	182.14
HPRC	1	57	47	137.02
PAGE	4	690	690	16.98
WGSPP1	6	1,504	1,504	177.36
Totals	246	285,333	268,454	3,046.68

Search Results

Consortium	dbGap id	Title	Consent Code	Terra Workspace Name	Diseases	Access	Data Types	Samples	Subjects	Size (TB)
1000 Genomes	--	--	open access	1000G-high-coverage-2019	not applicable	Open Access	Whole Genome	3,202	3,202	72.98
CCDG	registration pending	--	DS-ASD	acc_nhlbi_data_biospecimen_ac-biospecimen_name	autism spectrum disorder	Consortium Access	Exome	757	757	1.27

Computing!



WORKSPACES

Workspace: anvil-outreach/metrics

Cloud Environment: RUNNING (\$0.20)

DASHBOARD DATA NOTEBOOKS WORKFLOWS JOB HISTORY

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Console

```
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Bioconductor version 3.12 (BioManager 1.30.10), 7BioManager::install for help
```

Environment History Connections Tutorial

Global Environment

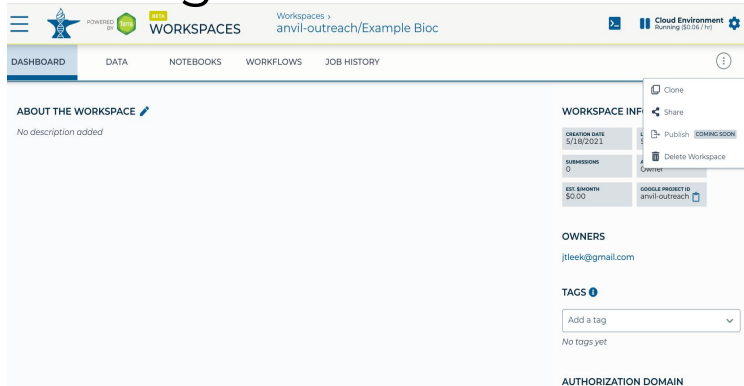
Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Home lost-found R

Sharing!



WORKSPACES

Workspace: anvil-outreach/Example Bioc

Cloud Environment: Running (\$0.10 / hr)

DASHBOARD DATA NOTEBOOKS WORKFLOWS JOB HISTORY

ABOUT THE WORKSPACE

No description added

WORKSPACE INFO

CREATION DATE: 5/18/2021

SUBMISSIONS: 0

EXT. BENCHMARK: \$0.00

OWNERS: jtleek@gmail.com

TAGS: Add a tag

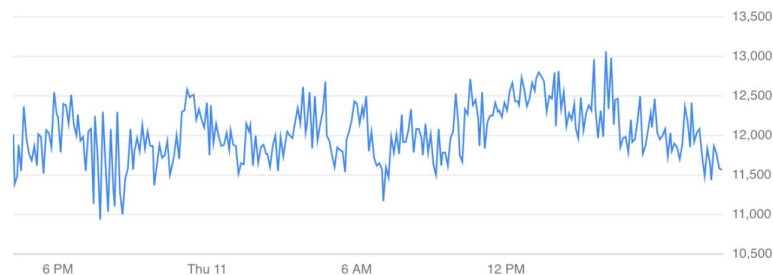
No tags yet

AUTHORIZATION DOMAIN

Scale!

Preview

1 hour 4 hours 1 day



instance/cpu/reserved_cores: 11,552.00



YEEESSS!

Barriers to adoption and use of AnVIL

1. Making the community aware of AnVIL and its capabilities
 - a. Getting AnVIL into users hands
 - b. Keeping them updated about changes
2. Overcoming hurdles to cloud adoption
 - a. Billing
 - b. Security
 - c. New way of thinking
3. Training people in using the cloud/AnVIL
 - a. Using the cloud
 - b. Managing people
 - c. Managing data
 - d. Managing computing
 - e. Managing bills
4. Providing support to diverse communities
 - a. Diverse in background/resources
 - b. Diverse in applications
 - c. Diverse in tools



AnVIL is “renting computers”



Standard Computing

- You buy a laptop one time
- You get that one laptop
- You pay little per use

Introduction

This vignette will walk you through how to examine results from a `DESeq2` analysis. The output data should have been saved to the bucket in the previous vignette [DESeq2 Analysis](#).

Installation

Instructions for installing packages necessary for this notebook are given in [An Overview of AnVIL Bulk RNASeq](#). Refer to that vignette for installation steps.

Load the packages to be used in this notebook:

```
In [1]: # Load packages
suppressPackageStartupMessages({
  library(DESeq2)
  library(ggplot2)
})
```

Load the DESeq results:

```
In [2]: # Move the result saved in the bucket to the compute workspace
AnVIL::avFiles_restore(source = "DESeq_result.RData")

# Load the results
dds <- readRDS("DESeq_result.RData")
dds
```

Copying gs://fc-7be0133b-f0f4-456e-8e0e-834b2053af18/DESeq_result.RData...
/ [0/1 files] 0.0 B/ 22.4 MiB 0% Done
/ [0/1 files] 264.0 KiB/ 22.4 MiB 1% Done
-
- [1/1 files] 22.4 MiB/ 22.4 MiB 100% Done
Resolving metadata from 1 objects (24.4 MiB)

Cloud computing

- You use any web browser
- You rent the computers
- You pay per hour/gigabyte/etc.

Team Design

Centralized

- UX research
- Content Development
- Lightweight funding mechanisms
- Responding to community requests
- Supporting GDSCN faculty

Decentralized

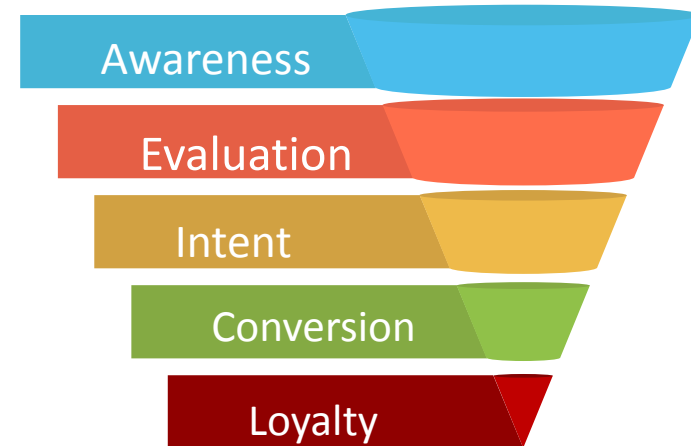
- Leverage distinct user community outreach efforts
- Conference and popup workshops
- Training webinars

Our approach



Overview of Training Activities

- Awareness and advertising
- Asynchronous Support
- Synchronous Support
- Tailored Support
- User Research



Awareness



Blog Posts



Date Published	Article	Views: Publish date - Oct 26
January 22, 2021	Try out RStudio in Terra	905 (one of the top posts)
January 29, 2021	Try out Galaxy in Terra via AnVIL	332
June 9, 2021	A Galaxy of tools at your fingertips	439
June 17, 2021	How Terra fits within the AnVIL ecosystem	339
October 7, 2021	Access Terra/AnVIL resources easily with BioConductor and the AnVIL package	246

Social Media



AnVIL
@useAnVIL

Congrats to AnVIL Cloud Credits awardee
[@tycheleturner](#)!

Tychele Turner @tycheleturner · Oct 12

Our @TNTurnerLab @WashUGenetics ACES manuscript is now out in Bioinformatics! doi.org/10.1093/bioinf... Code available here github.com/TNTurnerLab/AC... and it now also runs in the cloud! #genetics #genomics #bioinformatics

8:45 PM · Oct 12, 2021 · Twitter Web App

2 Retweets 5 Likes



Tweet your reply

Reply



Tychele Turner @tycheleturner · Oct 13

Replying to [@useAnVIL](#)
Thank you for supporting our work!



AnVIL Retweeted



Galaxy Project
@galaxyproject

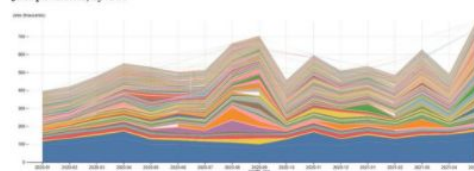


[#UseGalaxy](#) for [@useAnVIL](#) Cloud Cost Analysis, this Thursday at noon US Eastern [galaxyproject.org/events/2021-08...](#) Presented by [@victorW65326694](#) and Bridget Carr. Comparing [@jetstream_cloud](#) [@awscloud](#) & [@googlecloud](#)

Goals

- Determine approximate cost of cloud based jobs based off benchmark results.
- Get a better picture of which Galaxy tools are the most popular, or consume the most resources.
- Visualize and predict which tools consume the most resources or are most popular in volume.

Jobs per month, by tool.



2:01 PM · Aug 17, 2021 · TweetDeck

7 Retweets 8 Likes


















<https://twitter.com/useAnVIL/>

Asynchronous Support



all categories ▸ all tags ▸ **Latest** Top Categories

[+ New Topic](#)

Topic		Replies	Views	Activity
AnVIL Office Hours 18NOV2021 @ 11 am ET		0	11	1d
Running bwa mem on AnVIL	 	4	71	Sep 15
Public dissemination of derived data on AnVIL? terra	 	5	78	Sep 14
  Welcome to Discourse		0	80	Aug 11
Error creating billing project	 	5	108	Jun 10
Embed images in Workspace descriptions		0	83	May 21
Load error when selecting 'Workspaces' tab terra, bioconductor, anvilpublish	 	7	100	May 19
Error Creating Cloud Environment	 	4	101	Mar 19
Workflow execution status in Terra terra	 	9	150	Mar 17

AnVIL powered by Terra - Support

Need Help?

Search our documentation and community forum

AnVIL



Terra is a cloud-native platform for biomedical researchers to access data, run analysis tools, and collaborate. Terra powers important scientific projects like FireCloud, AnVIL, and BioData Catalyst. [Learn more.](#)

Terra Support > Search results

66 results for "AnVIL"

Knowledge base

AnVIL: Linking Gen3 and AnVIL on Terra

Terra Support > Documentation > Scientific Partnerships > AnVIL
· Allie Hajian · 1 year ago

Next, go to your Terra Profile page -
<https://anvil.terra.bio/#profile>

Launching seqr through Terra

Terra Support > Documentation > Analysis > Galaxy · Anton
Kovalsky · 5 months ago

To learn more about how to use seqr and how it's
implemented in *AnVIL*

Community

Workspace email notifications (e.g. AnVIL) 0

Terra Support > Community > Known Issues · Tiffany Miller · 1
year ago

Many people are getting email notifications with the subject
"Terra: Data has been added to *AnVIL*...workspace" and would

Add the "AnVIL" package in the cloud environments by default 1

Terra Support > Community > Feature & Documentation Requests
· Eugenio Mattei · 5 months ago

The *AnVIL* package is an essential tool to interface with the

Channels for
support:

- AnVIL Discourse (new this year!)
- Terra Support forum
- Contact Us button via anvil.terra.bio
- Email (for Terra questions)

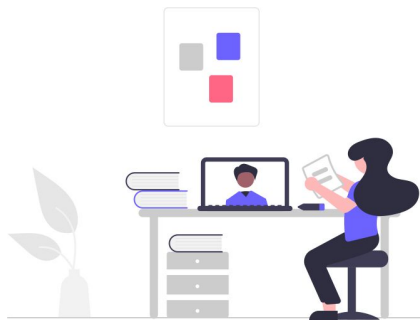
<https://support.terra.bio/hc/en-us>

Persona-oriented approach for training materials

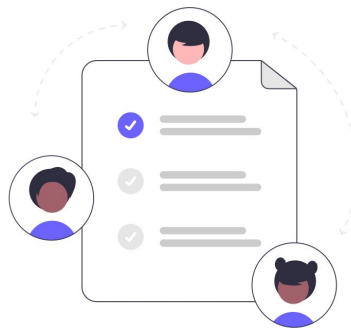
PIs



Analysts



Consortia



Teachers



Budget Templates for AnVIL



Overview Learn Datasets News Events Team FAQ Help

Search

Learn

Introduction Investigators Consortia

Setting up Your Lab in AnVIL

Preparing a Cloud Cost
Budget Justification

Preparing a Budget Justification

To prepare a budget justification, you can use the template Google Doc [AnVIL Budget Justification](#) as a guide to creating a budget justification paragraph for your proposal by including the information highlighted in pink (mostly copying entries from your AnVIL Cost Estimator Google Sheet).

Be sure to check that the prices are up to date by using the links listed above or in the AnVIL Cost Estimator.

Budget Justification Example Text

An example budget justification is given below. For a Google Doc version of the example, see [AnVIL Budget Justification Example](#).

Example

AnVIL Data Storage - We anticipate collecting and storing genotype data on between 100,000 individuals representing a maximum of 10 TB of storage. These data will be stored on the AnVIL system for distribution and analysis through the AnVIL projects consortium data management. These data are hosted on Google Cloud Platform, and their storage will be governed under consortia data storage agreements arranged through the NHGRI and will not be charged to this grant.

AnVIL Temporary Data Storage - We anticipate that up to 4 TB of intermediate

On This Page

Understanding GCP Fees

Estimating your Cloud
Costs


Preparing a Budget
Justification

Budget Justification
Example Text

Completed: AnVIL Getting Started Guide - 1.0

Upcoming:

Teacher Guide, Exercises Book & Galaxy Guide

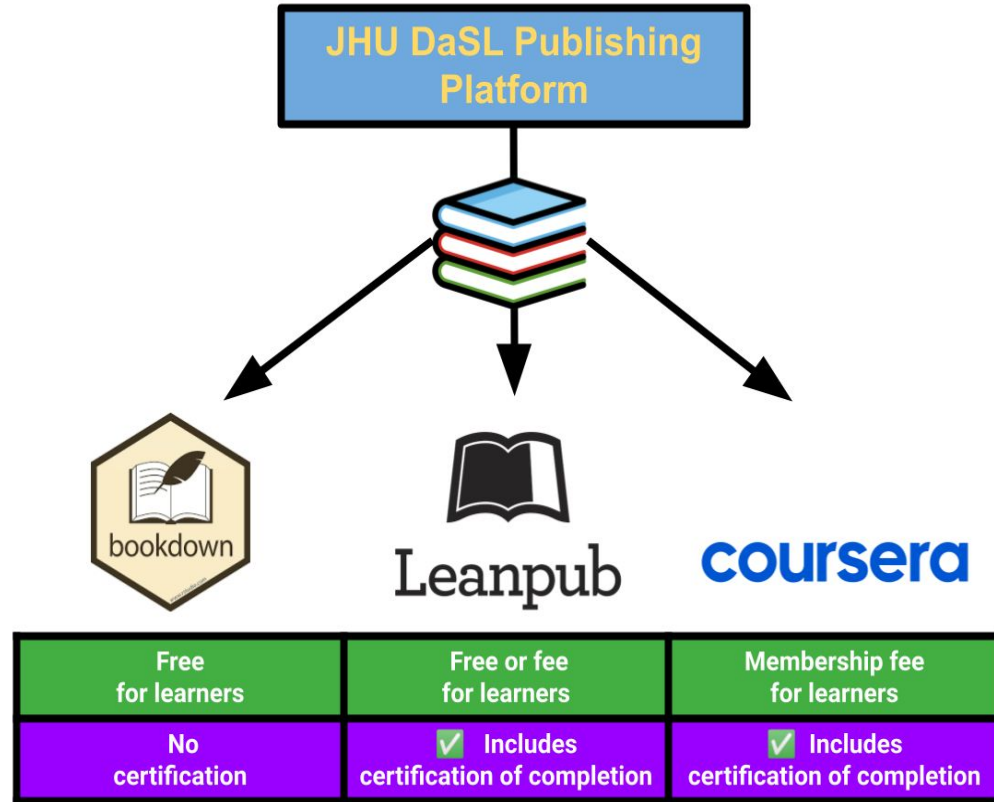


AnVIL_Cost_Estimator						
File Edit View Insert Format Data Tools Add-ons Help Last edit was made 2 days ago by Kai Kammerers						
100% \$ % .00 123 Arial 10 B I U A						
A1 fx Costs for Computing						
1	Costs for Computing	Costs/Hour	Number of hours	Costs/Month	Costs/Year	
2	n1-standard-4 instance consisting of 4 vCPUs and 15 GB of RAM			\$97.09	\$1,165.08 (monthly rates selected)	
3	n1-standard-8 instance consisting of 8 vCPUs and 30 GB of RAM	\$0.379998	174	\$66.12	\$793.44 (hourly rates selected)	
4						
5						
6	Costs for Storage	Costs/Month (1 GB)	Number of GB	Costs/Month	Costs/Year	
7	Local SSD provisioned space	\$0.080	375	\$30.00	\$360.00	
8	Standard Storage, single region storage: iowa (us-central1)	\$0.02	4096	\$81.92	\$983.04	
9						
10						
11	Costs for Network usage (egress)	Cost/GB	Number of GB	Costs/Month	Costs/Year	
12	0-1 TB tier	\$0.12	1024	\$122.88	\$1,474.56	
13	1-10 TB tier	\$0.11	1024	\$112.64	\$1,351.68	
14	10+ TB tier	\$0.08	0	\$0.00	\$0.00	
15	Total of 2 TB egress to Worldwide Destinations				\$2,826.24	
16						
17						
18	Additional Information					
19	Please tailor the numbers highlighted in pink.					
20	Pricing based on rates on 12/01/2020, please check for up-to-dateness by using the links listed below.					
21	Storage and network usage are calculated in binary gigabytes (GB): 1 TB is 1024 GBs.					
22						
23	Costs for Computing is driven by CPU and memory requirements.					
24	https://cloud.google.com/compute/all-pricing#n1_standard_machine_types					

AnVIL Content

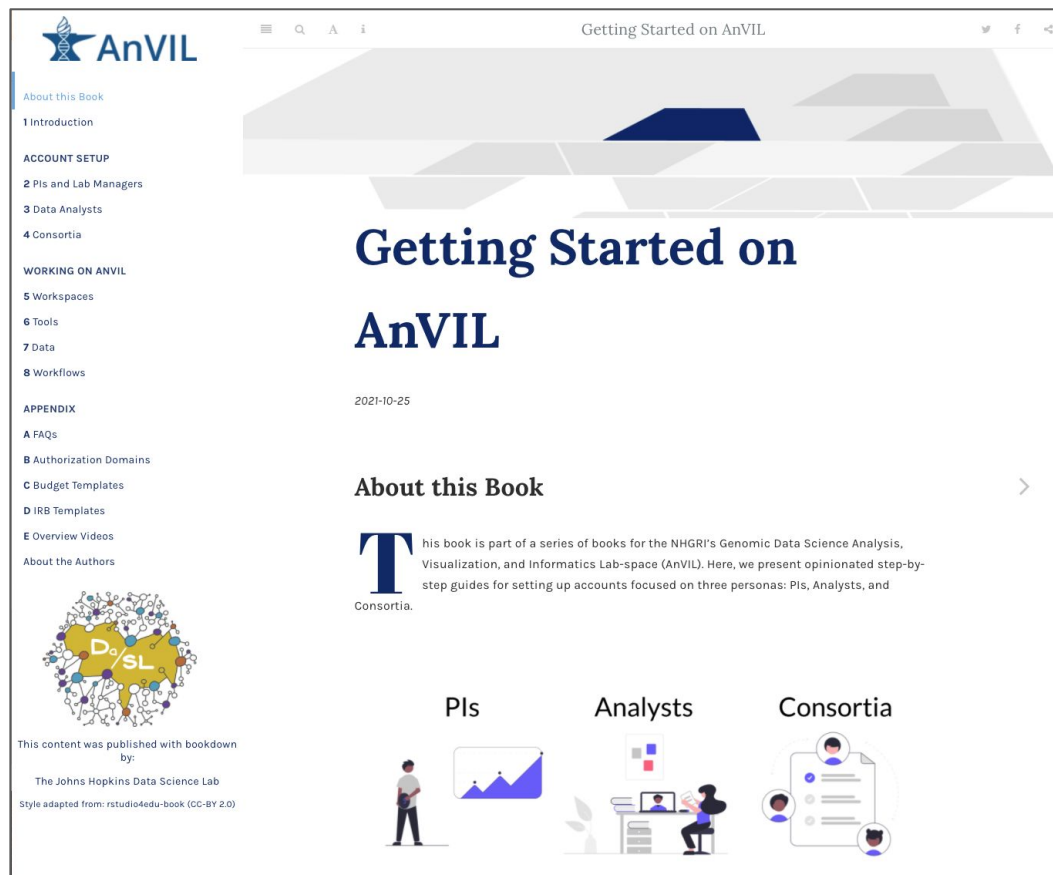


JHU DaSL EdTech Stack



Publish and maintain on all 3 platforms!

AnVIL Getting Started Guide - 1.0



Opinionated guide for
PIs/Lab Managers,
Analysts, and Consortia
Members

Walks through:

- Billing Account Setup
- How to work on AnVIL
 - Workspaces
 - Tools
 - Data
 - Workflows
- Budget and IRB templates

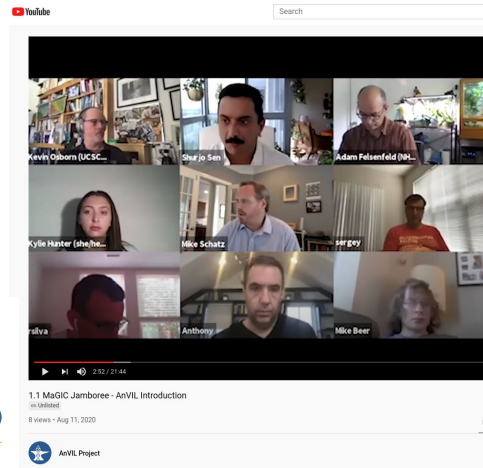
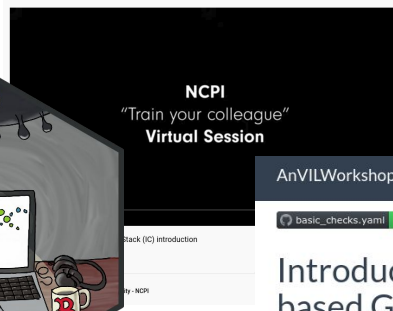
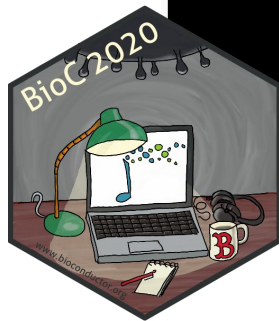
https://jhudatascience.org/AnVIL_Book_Getting_Started/

Synchronous Support



Synchronous Events

- Galaxy Webinars
- Health Disparities Researchers
- ASHG
- T2T/HPRC
- Bioconductor Community Conference
- Bioinformatics Community Conference
- eMERGE Researchers
- ISMB
- MaGIC Jamboree
- NCPI “Train your colleague”
- PAG
- GDSCN
- + many more



AnVILWorkshop 0.1.1 Workshop Reference

basic_checks.yaml passing

Introduction to the Terra/AnVIL Cloud-based Genomics Platform

Galaxy as an educational tool and community resources for undergraduate training

PAG 2020

Office Hours



Attendees:

- Data Coordinating Center member
- Data analyst applying ML techniques
- AC2 awardee

Discussed solutions to roadblocks with:

1. Set up/Billing
2. Data access
3. Analysis
4. Future work and support needed

Highlighted help.anvilproject.org as a resource to get help for new issues and questions in the future.

Rich Green to **Everyone**

11:44 AM

Thanks! Sorry I have to jump off. This was super helpful

Next Office Hours: 18NOV



AnVIL Office Hours 18NOV2021 @ 11 am ET



nakucher

6m

The AnVIL Outreach Working Group is hosting virtual AnVIL Office Hours on Thursday, November 18, 2021 at 11:00 - 11:50 am ET. These Office Hours are a new opportunity for you to get your questions about working on AnVIL answered in person – whether you're trying to set up a billing account, launch Galaxy or RStudio, looking for methods and featured workspaces, and more. Members of the AnVIL team will be available to help users including PIs, analysts, and data submitters get unstuck, troubleshoot issues, and discover online resources that provide further information.

Please post your questions in this thread ahead of the session!

Register here to receive the meeting link: <https://forms.gle/ghgHXYmvV3b23Vgc6>.

    Reply

<https://help.anvilproject.org/t/anvil-office-hours-18nov2021-11-am-et/56>

Tailored Support



AC2 Awardees

Alex Greiner | The University of Iowa | Graduate Student

- “Burden analysis of inherited cardiac arrhythmia genes in epilepsy”

Melissa Suzanne Cline | UC Santa Cruz Genomics Institute |
Principal Investigator

- “Leveraging AnVIL and Terra for secure collaboration on genetic variant interpretation”

Andrew Davidson | University of California | Graduate Student

- “Comprehensive characterization of transposable element expression across human tissues”

Anahita Khojandi | University of Tennessee-Knoxville |
Associate Professor

- “Deep Learning for Accurate Tissue-Specific Prediction of Gene Expression in Large Deeply-Phenotyped Population”

Anshul Kundaje | Stanford University | Principal Investigator

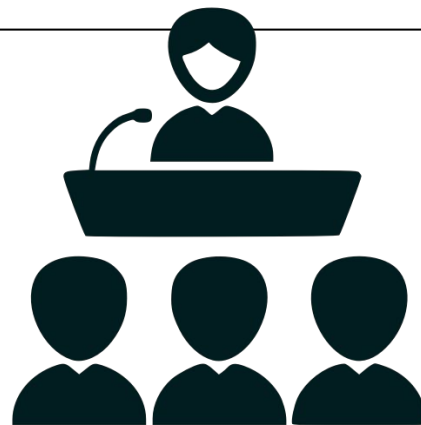
- “Deciphering cis-regulatory syntax of a transcription factor binding atlas with interpretable deep learning models”

Tychele N. Turner | Washington University in St. Louis |
Principal Investigator

- “A k-mer based approach to assess copy number in PacBio HiFi data”

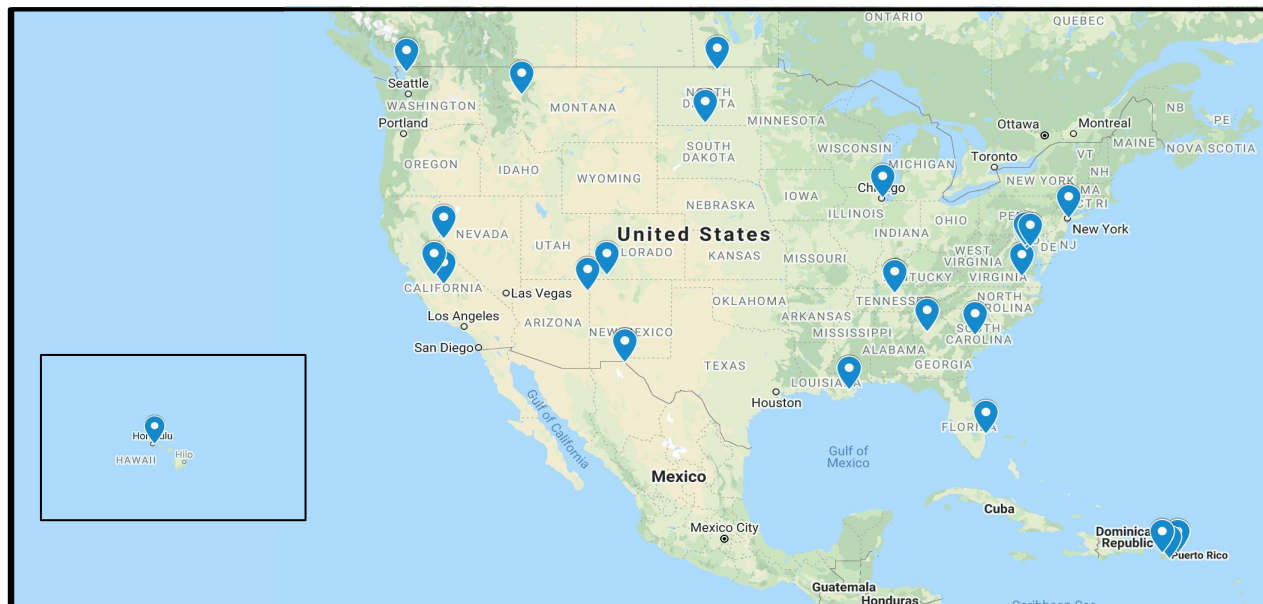
Meet and Greets

Hosted 2 sessions for awardees to meet, present their work and progress, and share their experiences using AnVIL



Genomic Data Science Community Network (GDSCN)

To broaden the spectrum of diverse institutions active in bioinformatics and genomic data science, we seek partnerships with educators and researchers at Historically Black Colleges and Universities (HBCUs), Minority Serving Institutions (MSIs), Tribal Colleges and Universities (TCUs), and Community Colleges (CCs).



<http://www.gdscn.org/>

Events

1 Kickoff Meeting

March 2021

2 Symposium #2

April 2021

3 Symposium #3

November 2021

Developing Lectures and Exercises



Genomic
Sequencing and
Variant Detection



SARS-CoV-2
Mutation Detection
with Galaxy

Laboratory Introduction



publicdomainvectors.org



JOHN HENRIKSEN
UNIVERSITY OF YORK
GENOMICS

GDSCN

User Research



D. Workspace Clone - Total

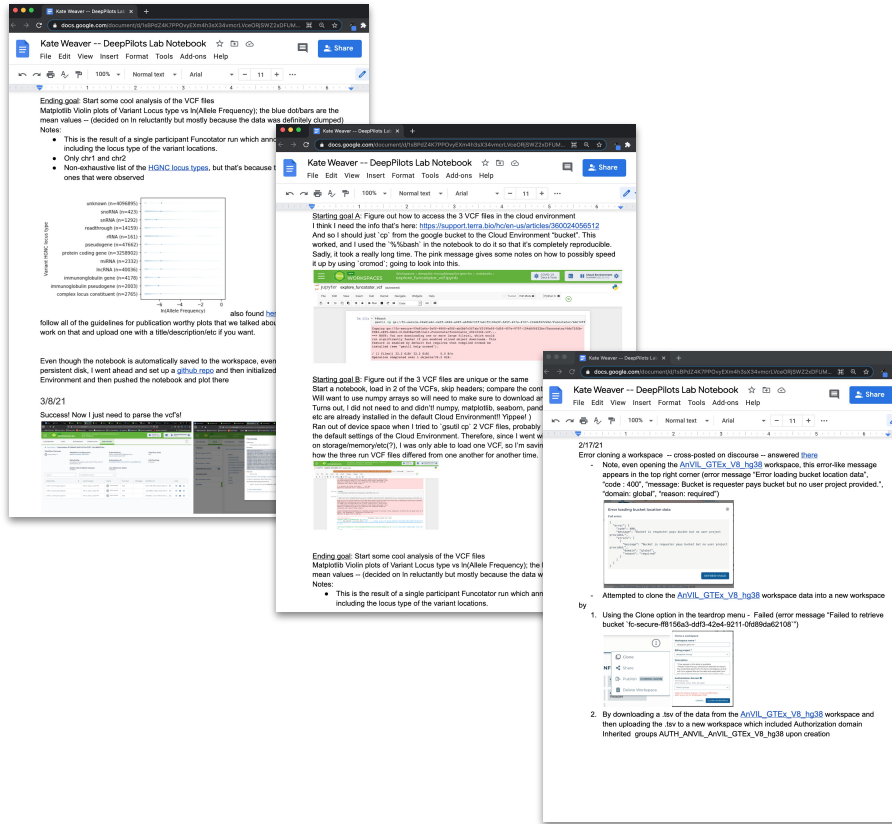
Jupyter Launch

BCC2020

AnVIL analyzed on AnVIL :)

MaGIC Jamboree

DeepPilots - harnessing researcher experience



By tracking the progress of researchers performing novel research on the cloud we can:

Identify our largest roadblocks

Chart a course to circumvent these roadblocks

Build appropriate tutorials and help documents to streamline transition to research on AnVIL

Goal

To build a happy, big, diverse, vibrant AnVIL user community doing teaching & research.



Vision for the Future

We built it, now we want tons of users to come to AnVIL!

1. Scale up our asynchronous (scalable) and synchronous (deep) training efforts
2. Continue to develop content and tools that can support platform changes
3. Build the AnVIL community (GDSCN, clinical, AnVIL Discourse)
4. Leverage our user research to improve user experience
 - a. *Simplified/scalable financial support through a free tier*
 - b. Easier access points to tools through the website
 - c. Simplified billing/data management/user management interfaces