



Workshop Report: Future Directions of the NHGRI Analysis, Visualization, and Informatics Lab-space (AnVIL)

Marylyn D. Ritchie, Ph.D.

Department of Genetics

Director, Institute for Biomedical Informatics

University of Pennsylvania

Presentation Overview

- AnVIL Analysis Workspaces Components
 - Overview of Future Directions Of The NHGRI AnVIL Workshop
 - AnVIL SWOT Analysis
 - Next Steps for AnVIL
- 



AnVIL Analysis Workspaces Components

Faceted search

Established pipelines

Exploratory Analysis

Integrated development environment

WORKSPACES Home / Workspaces / Evaluation protocol for predicting cancer driver genes

DASHBOARD DATA ANALYSIS TOOLS HISTORY

Jupyter Notebook Jupyter Lab RStudio Terminal

NOTEBOOKS +

Name	Created by	Last changed	
[Notebook Name] sometimes this can be very long	JChen	12:15 PM	Recently Updated
[Notebook Name] that is shorter	mdlantrey	Jan 5, 2019	
[Notebook Name] sometimes this can be very long	mdlantrey	Jan 1, 2019	
[Notebook Name] sometimes this can be very long	JChen	Dec 23, 2018	
[Notebook Name]	pamratu	Dec 20, 2018	
[Notebook Name] sometimes this can be very long	pamratu	Dec 17, 2018	
[Notebook Name]	JChen	Dec 15, 2018	
[Notebook Name] sometimes this can be very long	JChen	Dec 3, 2018	
[Notebook Name]	pamratu	Nov 4, 2018	



Dockstore
Create, Share, Use

AnVIL / Dockstore: sharing containerized tools and workflows



Genome Browser

Overview of Future Directions Of The NHGRI AnVIL Workshop

Virtual Meeting
Oct 29, 2021; 12pm-5pm ET

Goal	Identify gaps, challenges, and future opportunities related to NHGRI's investments in the AnVIL's cloud-based infrastructure, tools, and services.
Discuss	The current state of AnVIL
Identify	Activities needed to expand, diversify, and support the AnVIL user community

Workshop Breakout Rooms



Data submission and consortia
engagement



Analysis tools



Infrastructure



Outreach and training

Cross Cutting Themes

How cloud-based platforms can better serve the needs of genomic researchers

What tools and services would better support clinical genomic researchers

How to improve interoperability with other cloud-based resources in a federated genomic data ecosystem

Workshop Attendees

- Invited participants and assigned discussants
- AnVIL key personnel
- NHGRI AnVIL staff
- AnVIL External Consultants Committee
- Other NHGRI staff as listeners
- 105 people attended



Strengths - Where does AnVIL excel?

Strong commitment to data security

A versatile platform for training students and consortia on the use of the cloud for genomic analyses

NHGRI's leadership of the NCPI efforts and AnVIL's extensive use of APIs and GA4GH standards facilitates interoperability

Extensive documentation, which enables users to help themselves and help each other

Ability by third party groups to build on the platform while still prioritizing security

NHGRI's robust engagement with other NIH Institutes and Centers to encourage the adoption of the GA4GH Data Use Ontology (DUO) and Data Use Oversight System (DUOS) to streamline the data access review process

Weaknesses: Where is AnVIL at a disadvantage?

Curation of tools and workflows to facilitate searches by users could be improved

Lack of phenotypic harmonization across programs

The relationships between AnVIL and data coordination centers of NHGRI funded consortia could be better defined

Significant hurdles required to access AnVIL just to test the platform

Lack of video documentation

Users Cannot log in anonymously

AnVIL lacks embedded personnel (e.g., key personnel and developers) with clinical research priorities

Opportunities: Where AnVIL can grow and improve?

AnVIL could create a safe space for groups hesitant to host diverse controlled access datasets in public repositories

AnVIL could transition over to tools development and analyses for discovery, in addition to acting as a data repository

The AnVIL team could demonstrate that AnVIL can work for the clinical community

AnVIL could add additional data standards and data models to improve the interoperability model for Terra

The outreach team could conduct robust research into how AnVIL is being used

AnVIL has the opportunity to introduce cloud computing for the next generation of scientists

AnVIL could integrate a diversity of human genetic datasets and reference genomes and make them backwards compatible

Threats: Which factors jeopardize AnVIL?

Cloud costs are a major barrier for many users.

Challenges in making AnVIL interoperable with other platforms.

Difficulties in shifting scientific culture to the cloud.

Challenges in making tools and resources in a manner that meets users where they are

AnVIL faces data security threats, both by outsiders and by people with authorized access.

Potential institutional fear of investing in a dead-end utility if NHGRI's long term commitment to the resource is unclear.

Potential lack of users' skill transferability between AnVIL and other cloud platforms.

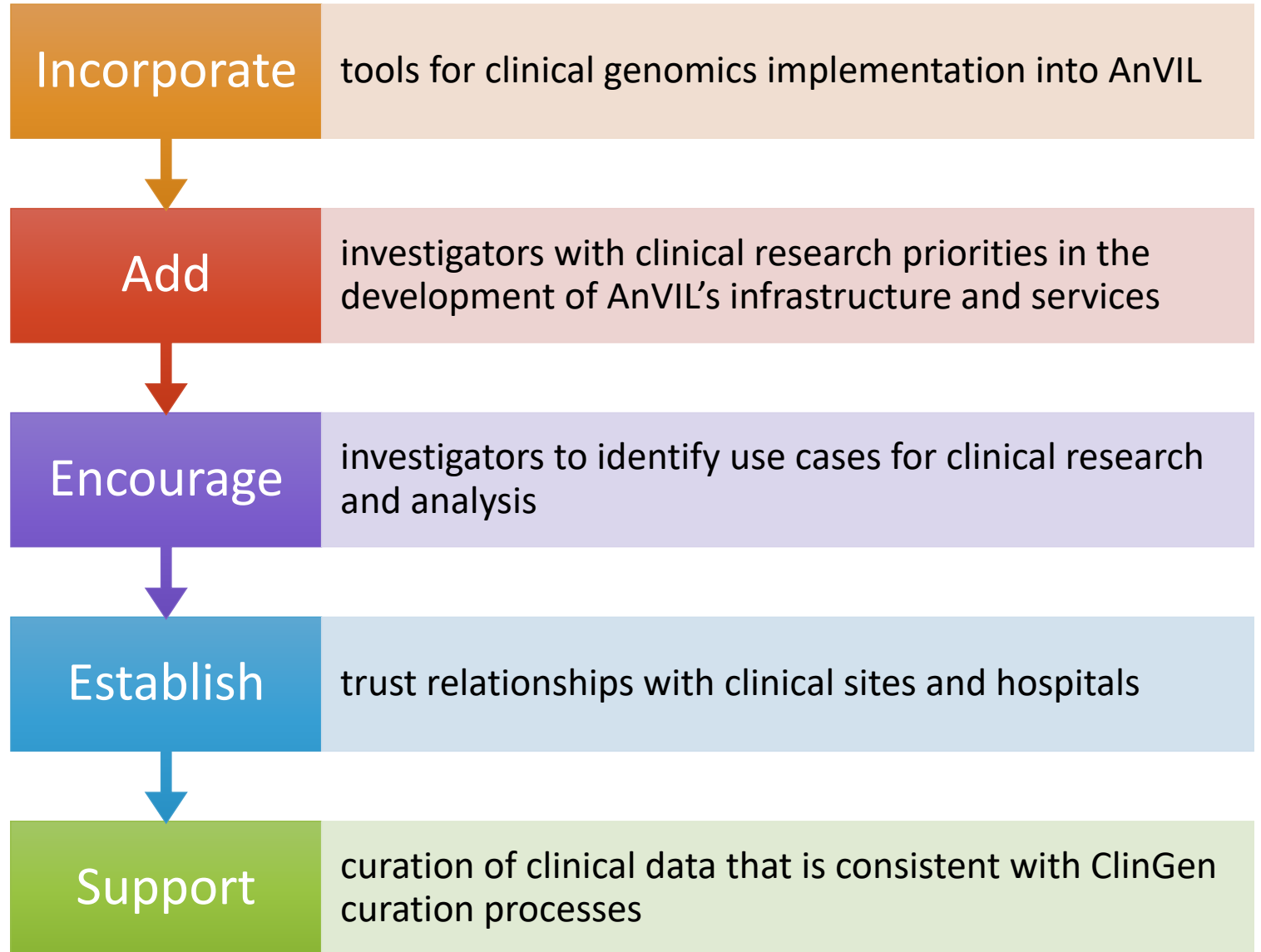
Challenges Related To The Use Of The Cloud

Culture change for most investigators who have exclusively worked using on-premises infrastructure

Onboarding users who are inexperienced with cloud resources

Cloud costs are both a practical and psychological barrier

Support For The Clinical Research Community



Outreach And Training

Make AnVIL accessible to naïve users by:

- generating and sharing popular analysis workflows
- developing and advertising an extensive video guide to its features
- providing streamlined billing and payment for cloud services
- providing easier to access data in the interactive workspaces

Make onboarding easier for all users, in particular for users from low-resource institutions

Leverage the research communities funded by NHGRI to build a sense of community around the platform

Identify use cases to demonstrate AnVIL's capabilities and promote them

Interoperability, Analysis Tools, And Other Considerations



Continue to engage more broadly with other NIH Institutes and Centers



Increase phenotypic data harmonization efforts across NHGRI-funded consortia and initiatives



Support more data standards and common data models to improve semantic interoperability with other platforms



Support the development and availability of more analysis tools for both basic and clinical genomics researchers



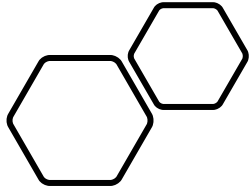
Increase the search capabilities for datasets, tools, and workflows that are available on the platform



Rebrand AnVIL from a data submission site to a multi-functional discovery platform



Clarify relationship between AnVIL and data coordination centers



Next Steps

NHGRI to prioritize AnVIL's future activities and prepare for renewal

The screenshot shows the NHGRI website with a dark blue background and a network diagram. The page title is "Future Directions of the NHGRI Analysis, Visualization, and Informatics Lab-space (AnVIL)". Below the title, there is a section for "Event Details" which describes a workshop aimed at identifying gaps and future opportunities related to NHGRI's investments in the AnVIL's cloud-based infrastructure, tools, and services. The workshop will include discussions on the current status of the infrastructure, tools, and services provided by AnVIL to serve both the basic and clinical genomics research communities. These discussions will also help NHGRI identify activities needed to expand, diversify, and support the AnVIL user community.

<https://www.genome.gov/event-calendar/Future-Directions-of-the-NHGRI-Analysis-Visualization-and-Informatics-Lab-Space-AnVIL>



Thank You
Workshop Participants and Organizers

Questions?