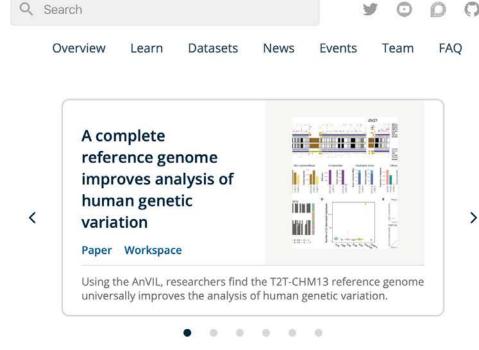
Migrate Your **Genomic Research** to the Cloud

Secure, cost-effective genomic analysis at scale.

Get Started

Learn More





Terra

Collaborate in Terra, AnVIL's secure, scalable, cloud compute environment.

Launch Learn More



Gen3

Manage, harmonize, and share large datasets.

Launch Learn More



Dockstore

Create and share Docker-based workflows.

Launch Learn More



NCPI

Interoperate with other NIH data commons.

Help

Learn More



Bioconductor

Analyze genomic data in the R statistical language.

Learn More



Galaxy

Run batch analysis workflows and interactive visualizations.

Learn More



Jupyter

Run interactive analysis with python or R.

Learn More



Segr

Identify disease-causing variants.

Launch Learn More

AnVIL's Capabilities and Services





Cloud-based infrastructure and software platform



Shared analysis and computing environment



Data access and data security



Genomic datasets, phenotypes and metadata



Cost Control for Cloud services



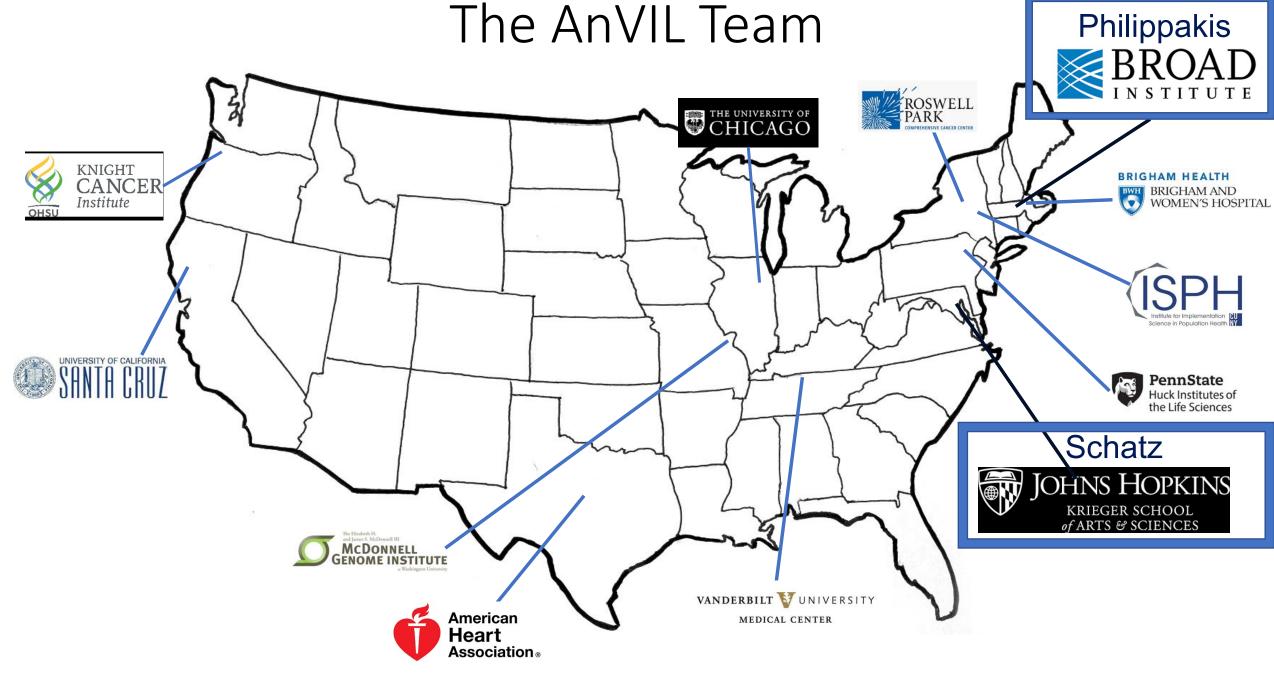
User training and outreach



Participation in a federated genomic data commons ecosystem



Incorporation of scientific and technology advances



AnVIL External Consultant Committee (ECC)

Nadav Ahituv, Ph.D.

University of California, San Francisco

Cinnamon Bloss, Ph.D.

University of California, San Diego

Carol Bult, Ph.D.

Jackson Labs

Karen Davis, M.S.

RTI International

Aleksandar
Milosavljevic, Ph.D.
Baylor College of
Medicine

George Hripcsak, M.D.

Columbia University

Siddharth Pratap, Ph.D.

Meharry Medical College

Adam Resnick, Ph.D.

Children's Hospital of Philadelphia

Marylyn Ritchie, Ph.D.

University of Pennsylvania

NHGRI AnVIL Staff



Valentina
Di Francesco, M.S.
AnVIL Co-Lead



Anneliese
Galczynski, M.B.A.
Grants Management
Specialist



Elena M.
Ghanaim, M.A.
Policy Analyst



Carolyn M.
Hutter, Ph.D.
Director of the Division of Genome Sciences



Teri Manolio,
M.D., Ph.D.
Director of the Division
of Genomic Medicine



Shurjo K. Sen, Ph.D.
Program Director



Ana Pilar Stevens
Scientific Program
Analyst



Helen Thompson
Scientific Program
Analyst



Christopher
Wellington, B.S.
Program Director



Ken L. Wiley, Jr. Ph.D.
AnVIL Co-Lead



AnVIL / Terra: analysis workspaces and batch workflows

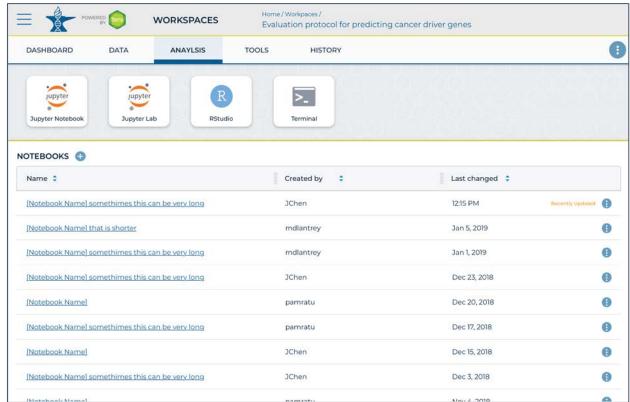
Faceted search

Established pipelines

Exploratory Analysis

Integrated development environment





wdl



AnVIL / Dockstore: sharing containerized tools and workflows









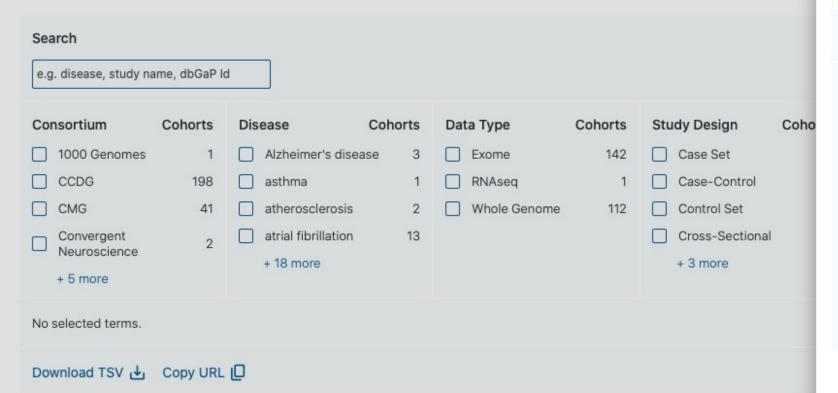
Genome Browser

NHGRI Analysis Visualization and Informatics Lab-space

Q Search

Overview Learn

AnVIL Dataset Catalog



Search Summary

Consortium	Cohorts	Sampl
1000 Genomes	1	3,2
CCDG	198	272,3

Consortium

Current selection: 9 Consortia 254 Cohorts 324,235 Samples 290,929 Participants 3,971.91 Size (TB)

No selected terms.

	Cohorts
1000 Genomes	1
CCDG	198
CMG	41
Convergent Neuroscience	2
GTEx (v8)	1
HPRC	1
PAGE	4
T2T	1
WGSPD1	5

Q Search

Overview

iew Learn

AnVIL Dataset Catalog

Search	Consortium	Cohorts	Disease Co	horts	Data Type	Cohorts	Study Design	Cohorts
e.g. disease, study name,	1000 Genomes	1	Alzheimer's disease	3	Exome	142	Case Set	13
	CCDG	198	asthma	1	RNAseq	1	Case-Control	88
	☐ CMG	41	atherosclerosis	2	☐ Whole Genome	112	Control Set	1
	Convergent	2	atrial fibrillation	13			Cross-Sectional	1
	Neuroscience + 5 more		+ 18 more				+ 3 more	
No selected terms.								
Download TSV ₺ Copy URL [0							

Search Summary

Consortium	Cohorts	Samples
1000 Genomes	1	3,202
CCDG	198	272,306
смб	41	20,706
Convergent Neuroscience	2	304
GTEx (v8)	1	17,382
HPRC	1	57
PAGE	4	690
Т2Т	1	0
WGSPD1	5	9,588
Totals	254	324,235

Disease

Current selection: 9 Consortia 254 Cohorts 324,235 Samples 290,929 Participants 3,971.91 Size (TB)

Cohorts

13 43

10

16 58

30

No :	selected terms.
	Alzheimer's disease
	asthma
	atherosclerosis
	atrial fibrillation
	autism spectrum disorder
	cardiovascular disease
	cardiovascular disease and risk factors
	cardiovascular system disease
	control
	coronary artery disease
	epilepsy
	heart and blood vessel disease
	hemorrhagic Stroke
	Hemorrhagic Stroke
	inflammatory bowel disease
	Mendelian
	myocardial infarction
	population
	Schizophrenia and Bipolar Disorder cases and contro
	Schizophrenia cases and controls

Spinal Muscular Atrophytype 1 diabetes mellitus

Sparch Posults

AnVIL Work Groups







PORTAL

DATA PROCESSING

TECHNICAL







DATA ACCESS

PHENOTYPE

OUTREACH

Outreach Activities



Develop AnVIL training materials, projects, evaluations, and videos



Release training and education materials as open source so they can be used across multiple courses/projects



Work with STRIDES and NHGRI to evaluate credit provision models



AnVIL Data Access





Public Access

There are two types

- Open
- Controlled

Open Access

Accessible to the public without prior approval

Controlled Access

- Apply through dbGaP
- Reviewed by a NIH/NHGRI Data Access Committee (DAC)
- Approved requestors added to a list managed by dbGaP which will be provided to the AnVIL staff for data access

Consortium Access



Consortium members will have access to their own data



Access will be managed by a POC agreed upon by the Consortium members (e.g. a Coordinating Center)



The approved POC will develop and maintain a whitelist of approved members



The authorization list will be sent to the AnVIL team to inform them who should have access to consortium data

Common Questions

Will datasets hosted by the AnVIL also have to be uploaded to dbGap as part of the dbGaP registration process?

NO

- data hosted by the AnVIL will not be required to be uploaded also to dbGaP
- The forms for registering datasets with dbGaP will still be required to be sent dbGaP

Will all Consortium Data hosted by the AnVIL automatically became available to the public (either by open or controlled access)

• NO

- Consortium data that is registered through dbGaP will be made available through controlled access
- Data approved by the data generator(s) can be made accessible for open access



Interoperability



Authentication & Authorization



Phenotypic Data Harmonization

Challenges



Overview

Participating Platforms

Guiding Principles

Working Groups

Training Materials

Progress Updates



