

**National Advisory Council for Human Genome Research (NACHGR)
May 16, 2022
Concept Clearance for FOAs**

**NHGRI Analysis, Visualization, and Informatics Lab-space (AnVIL)
Program Renewal**

Purpose

This concept proposes two Requests for Applications (RFAs) for the renewal of the AnVIL program: 1) a Limited Competition RFA to enhance and expand AnVIL's platform, tools, services, and management processes; and 2) an Open Competition RFA for the AnVIL Clinical Resource (ACR), a suite of genomic-based clinical tools, workflows, and services that will be incorporated into and leverage the existing AnVIL ecosystem to foster genomic-based clinical research. The awardees of both RFAs will work together to ensure the AnVIL ecosystem serves the basic and clinical genomics research communities.

Background

NHGRI established the [AnVIL platform](#) in 2018 as a federated, cloud-based environment for researchers to securely access, integrate, analyze, and share large genomic datasets from NHGRI initiatives. The AnVIL environment supports collaboration through workspaces for sharing data, analyses, and workflows. The platform supports both users with limited computational expertise and sophisticated data scientists. Detailed information about AnVIL is available in a 2022 [marker paper](#), in a [workshop booklet](#), and in a [list of accomplishments](#) previously shared with Council.

NHGRI supports AnVIL through two cooperative agreement awards, one to the Broad Institute and one to Johns Hopkins University, both including several partners. The two groups have complementary responsibilities and collaborate to achieve the overall program goals. Program oversight is provided by NHGRI staff and the multidisciplinary AnVIL [External Consultants Committee \(ECC\)](#).

NHGRI recently convened the "Future Directions of AnVIL" [workshop](#) to identify gaps, challenges, and future opportunities for AnVIL for basic and clinical genomics research communities. ECC member Dr. Marylyn Ritchie (UPenn) [reported on the workshop](#) at the February 2022 Council meeting. Workshop participants commended the AnVIL team's progress over the past 3.5 years, highlighting impressive accomplishments in data ingestion, strict data security standards, the extensive catalog of available analysis tools, and AnVIL's involvement in nascent NIH efforts to promote interoperability among NIH cloud-based data resources.

The workshop also identified opportunities to expand and diversify AnVIL activities supporting the genomic research community, specifically: 1) increasing availability of analysis tools and workflows; 2) increasing outreach and educational offerings; 3) addressing challenges related to the use of cloud-based services; 4) improving interoperability with other cloud-based data resources; and 5) increasing support for the clinical research community. NHGRI's vision for this renewal period is for

AnVIL to evolve from being primarily a data submission and sharing platform to being a multi-functional discovery platform for basic genomics and clinical researchers and a resource for genomic data science educational efforts.

This concept proposes two Requests for Applications (RFAs) for this renewal: a Limited Competition RFA for AnVIL; and an Open Competition RFA for the AnVIL Clinical Resource (ACR), which will be incorporated into the overall AnVIL project.

AnVIL Limited Competition RFA (U24 funding mechanism):

To become a multi-functional discovery platform for genomics, AnVIL awardees will continue to build and improve on the current user interfaces for data submission, access, secure sharing, and analysis, while also addressing opportunities 1-4 above. They will also work in close collaboration with the ACR awardee on opportunity 5. Specifically, they will:

(1) Increase availability of analysis tools and workflows: The AnVIL has integrated and made available genomic analysis workflows such as those in Galaxy, Bioconductor, and Dockstore. As more datasets are added and more researchers use the platform, it will be important to expand the analysis and workflow offerings. This renewal aims to establish an AnVIL “genomic app store” to easily share and deploy third-party analysis tools and workflows for analysis of both open and controlled-access data in a secure environment.

(2) Increase outreach and educational offerings: The AnVIL only recently reached a critical mass of available data and analysis workflows to make the platform appealing to genomic researchers. This is a key time to ramp up user outreach and engagement activities. During the renewal, AnVIL will encourage wider adoption by adding to online training materials and by running workshops and hackathons at scientific conferences and NHGRI consortium meetings. NHGRI and AnVIL grantees will also aim to streamline data access request processes to make the use of the platform more attractive. Finally, AnVIL is expected to collaborate with the Educational Hub for Enhancing Diversity in Computational Genomics and Data Science ([NOT-HG-22-012](#)) to support genomic data science curriculum development in undergraduate courses and to enhance genomic educational opportunities in limited-resource institutions.

(3) Address challenges related to use of cloud-based services: Transitioning data sharing, access, and analyses to the cloud is a significant change for investigators who have been using on-premise infrastructure. Enhancing educational offerings and awarding cloud credits or service discounts to AnVIL users will help investigators making that transition. NHGRI plans to expand the [AnVIL Cloud Credits Program](#) for investigators using AnVIL while also developing tools for investigators to budget and manage cloud costs, both of which address some of the biggest concerns investigators have about moving to the cloud. Throughout these activities, AnVIL outreach efforts will highlight advantages to using cloud computing for popular analysis workflows.

(4) *Improve interoperability with other cloud-based data resources:* Interoperability between AnVIL and other cloud-based resources is integral to AnVIL. AnVIL is a founding member of the [NIH Cloud Platform Interoperability Effort \(NCPI\)](#), which seeks to establish and implement guidelines and technical standards to enable end-user analyses across cloud platforms and to facilitate the realization of a trans-NIH [Secure and Authorized FAIR Environment \(SAFE\)](#) in which NIH platforms can interoperate. AnVIL grantees are expected to continue to play a crucial role in the trans-NIH and NCPI interoperability activities that are outlined in [OTA-22-084](#) and sponsored by the NIH Office of Data Science Strategy. More broadly, AnVIL should contribute to and adopt interoperability standards emerging from the international community of genomic researchers, such as Health Level Seven's Fast Healthcare Interoperability Resources (FHIR), Global Alliance for Genomics and Health standards (GA4GH), and Application Programming Interface standards.

AnVIL Clinical Resource (ACR) RFA (U24 funding mechanism):

The recent [Genomic Medicine XIII: Developing a Clinical Genomic Informatics Research Agenda](#) meeting and the AnVIL workshop highlighted the genomic-based clinical research community as a specific group that AnVIL should better support. With a separate award, this concept proposes to integrate a suite of genomic-based clinical components within the existing AnVIL ecosystem. These clinical components will include aspects such as: implementing clinical software, tools, and workflows; developing innovative and secure methods for storing, generating, and returning genomic-based clinical results; conducting outreach and educational efforts tailored to the clinical genomic research community; and facilitating data access and security that addresses clinical research needs. The ACR team will also work to leverage and combine diverse datasets (such as clinical, imaging, -omics, and social determinants of health, or SDOH, data) for genomic-based clinical research.

Clinical software, tools, and workflows: ACR will leverage the existing AnVIL ecosystem to adopt and integrate user-friendly, open-source software, tools, and workflows for users to visualize and analyze relationships between and among genomic, imaging, clinical, molecular, SDOH, and related data to develop computational-based predictive modeling of disease states and genomic variant interpretation. ACR will also provide a foundation for users to develop and research the use of digital health applications that can monitor and manage genomic-based health conditions that involve the analysis and reporting of genomic results for clinical use.

Developing innovative and secure methods for storing, generating, and returning clinical results: ACR will provide a foundation for users to develop innovative methods for conducting analysis, storing, and disseminating clinical reports for researchers and clinical laboratories. The ACR team will work with members of the genomic-based clinical research community to identify how AnVIL can serve their needs. In addition, the ACR team will develop and implement ways to leverage FHIR and GA4GH standards and APIs to access and perform analysis in AnVIL.

Outreach and user education: The ACR team will participate in AnVIL outreach activities and leverage educational resources to enable the genomic research

community to utilize and analyze genomic-based clinical data. Such efforts include, but are not limited to, conducting focus groups, developing use cases, providing individualized education sessions to serve specific groups, and soliciting and addressing users' feedback. The ACR team will also work to build trust relationships with clinical sites to support genomic-based clinical research on the AnVIL platform.

Data access and security: Data protection and proper stewardship of clinical, genomic, phenotypic, and other sensitive data through ACR are of the utmost importance. Therefore, ACR is expected to adhere to the data access and security requirements of AnVIL. In addition, the ACR team will work with the existing AnVIL team in making AnVIL a HIPAA-compliant resource, as needed to support the use of the platform by clinical sites.

Relationship to Ongoing Activities: As per [NOT-HG-19-024](#), AnVIL serves as the designated data repository and primary dissemination resource for key unrestricted and controlled-access datasets (including molecular data, phenotypic data, and metadata) generated with NHGRI funds. Numerous NHGRI funded programs [currently use](#) or will use AnVIL as their data sharing and consortium collaborative platform, including Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR), Developmental Genotype-Tissue Expression (dGTEx), Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium, and Electronic Medical Records and Genomics (eMERGE).

The AnVIL is also expected to collaborate with the NHGRI Educational Hub for Enhancing Diversity in Computational Genomics and Data Science ([NOT-HG-22-012](#)). In addition, AnVIL will continue to play a key role in [NCPI](#) efforts and will be involved in future NCPI activities of the NIH Office of Data Science Strategy.

Mechanism of Support: Both the AnVIL and the ACR RFAs will use the U24 Resource-Related Cooperative Agreement funding mechanism.

Eligibility for Limited and Open Competition RFAs:

Only recipient organizations funded under the existing AnVIL program ([RFA-HG-17-011](#)) will be eligible to apply for the proposed AnVIL limited competition RFA. The current primary awardees are expected to apply independently and to continue their collaborative efforts. A limited competition will allow NHGRI to leverage the existing investments made in establishing the platform, tools, services, documentation, and management processes while minimizing disruptions for AnVIL users. The AnVIL ACR RFA will be an open competition, but the current AnVIL primary awardees will not be eligible to apply.

Funds Anticipated: NHGRI will commit ~\$6.5M in total costs per year for the AnVIL RFA (flat budget relative to the current funding cycle) and an additional ~\$1.5M in total costs per year for the ACR RFA. Both will be funded for five years, FY23-FY27, for a total cost of approximately \$40M.