National Advisory Council for Human Genome Research
September 9, 2024
Concept Clearance: Impact of Genomic Variation on Function (IGVF) Renewal

Purpose:
The National Human Genome Research Institute (NHGRI) proposes a renewal of the Impact of Genomic Variation on Function Program (IGVF). The goal of the initial phase of IGVF was to develop a framework for understanding the effects of genomic variation on genome function and how these effects shape phenotypes. The proposed IGVF renewal will build on this initial activity with a focused set of collaborative research projects designed to enhance development and application of this framework and to create an expanded, user-friendly community resource to support future studies. The objectives of the renewal include: (1) systematically perturbing the genome to assess and catalog the impact of individual genomic variants on genome function and phenotype; (2) studying the effects of genetic or environmental interactions on genomic variant impact; (3) developing and applying predictive models and integrative analyses of genomic variation impact on genome function; and (4) expanding and further developing a broadly accessible set of resources, including a data portal and a comprehensive and searchable catalog of measured and predicted variant impacts.

Background:
A current challenge in genomics is defining the role of genomic variation in influencing phenotypes in human biology and diseases. As outlined in NHGRI's 2020 Strategic Vision, studies characterizing the functional consequences of variants have not kept pace with studies of variant discovery or association, and this lack of functional genomic information remains a barrier that impedes genomic progress. "Systematic approaches, including new tactics that connect high-throughput molecular readouts of functional genomic assays to organismal phenotypes, are required for establishing the phenotypic consequences of all genomic variants". Further, it is not feasible to experimentally probe all genomic variants of interest in all contexts, so novel data collection strategies and analytical approaches are required to predict the relationships between variation, function, and phenotypes.

IGVF was established in 2021 to address the challenge of understanding how genomic variation affects genome function to influence phenotype. Using a team-science, coordinated approach and emerging experimental and computational methods, the consortium brought together >100 labs across five different components to model different approaches to this challenge, and to work together to develop a framework and infrastructure to map, perturb, and predict variant impact on function.

Over the past three years, IGVF has made advances in several areas that will serve as foundations upon which a renewal phase will build:

- A series of collaborative projects have been organized to explore non-coding and coding variants both individually, and in the context of networks and exemplar diseases. A renewal phase will leverage this collaborative framework, connecting data generation to analysis to modeling and to resource building.
- An initial resource has been developed including datasets, predictive models, standards, and uniform processing pipelines, with data files made available through the IGVF Data Portal. IGVF is developing a searchable catalog that will enable the community to find information about measured and predicted effects for specific variants, genomic regions, and genes, as outlined in the recent IGVF update presentation to Council. A central goal of a renewal phase will be further development and refinement of this user-focused community resource.

- IGVF is partnering with other consortia and programs including the Atlas of Variant Effects (AVE) Alliance, the Clinical Genome (ClinGen) Resource, the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR), and the Human Pangenome Reference Consortium (HPRC) to enhance communication and opportunities to exchange data and resources. Interactions with these consortia have resulted in a cross-consortia meeting and joint projects that are currently underway (see meeting report). A renewal phase will continue coordination with related projects.
- Within IGVF, graduate students and postdoctoral fellows participate in a career development series and as active participants and leads of scientific focus groups and collaborative projects. Leadership and participation of early-career scientists will be further emphasized in a renewal phase.

In the spring of 2024, NHGRI hosted a series of listening sessions to obtain community input about future directions and opportunities for understanding the effects of variants on function. Session participants provided short term and long term recommendations for the field outlined in a meeting report. High level recommendations relevant to IGVF include:

- Continue testing the function of individual variants at scale.
- Increase testing of variants in biologically-relevant and disease-relevant contexts at scale, emphasizing cellular, environmental, and dynamic contexts.
- Prioritize diversity in genetic ancestries when choosing model systems and selecting variants for testing.
- Improve synergy between modelers and data generators through iterative data collection, AI/ML based modeling, and model testing with targeted experiments.
- Develop and promote standards for assays, protocols, metadata, analyses, and predictions that would enable data sharing and re-use.

Proposed Scope and Objectives:
This concept proposes four RFAs to renew the IGVF consortium for five years. The consortium will be comprised of four components that will address the following:

1) Individual Variant Characterization Centers (up to 5 centers; $15M total cost (TC)/year)
Systematically perturb variants or elements to assess the effects of individual variants on genome function and phenotype. The results will be integrated into a resource to enable community searches of individual variant impacts and to power variant impact predictors.

- Systematically apply 1-3 existing, state-of-the-art high-throughput genomic perturbation methods (for example, reporter assays, genome editing, epigenome editing). Assays should be appropriate for the variant types proposed for study.
- Assay impact of genome variation on molecular, cellular, or organismal phenotypes.
- Test variants in DNA, RNA, or protein-coding elements in one or more cell types and cell states of high value to the research community.
- Develop robust, reproducible, and portable data processing pipelines, as well as standardized models for data and metadata sharing, in collaboration with consortium.
- If needed, centers may include limited generation of single-cell, multi-omic (transcription, accessibility) maps of genes and regulatory elements for the specific cell types being studied.

2) Variant Interaction Centers (up to 5 centers; $10M TC/year)
Study the effects of genetic or environmental interactions on genomic variant impact. Centers will propose novel research frameworks to investigate whether interactions with effects beyond additive are important modifiers of the impact of variants on human

phenotypes. Results will be integrated into the IGVF catalog. The rationale for this combination of research and resource work is if there is structure within the interaction landscape, these centers might identify the subspace of interactions to test, leading to more effective variant testing, better understanding of GWAS, and improved polygenic scores.

- Propose a research framework (variants to study, context, and rationale) for studying the frequency and magnitude of interactions.
- Develop and apply the proposed framework to study interactions.
- Contribute results of tested combinations and models for predicting the most important combinations to test as part of the IGVF catalog.
- Examples of systems for studying the interaction of variants:
    - Genetic interactions, e.g. combinations of variants, combinations of variant types (coding and non-coding variants), range of genetic backgrounds
    - Environment interactions, e.g. models such as inflammation, stress, or altered metabolism

3) Modeling and Analysis Centers (up to 5 centers; $5M TC/year)

Develop predictive models and perform integrative analyses to determine the impact of genomic variation on genome function. Centers are expected to develop and/or refine computational approaches that are generalizable to systems beyond those that are studied and should demonstrate flexibility to accommodate new and emerging data.

- Develop, apply, and experimentally test predictive models. Examples include but are not limited to: a) the impact of genomic variation on function and phenotype; b) the location and function of variants in specific cell and spatial contexts; c) interactions of genomic variants; d) systems or network-level effects.
- Guide and optimize IGVF experimental design.
- Analyze existing IGVF data and other data to aid in model development and testing.
- Provide analytical expertise and support for consortium-wide efforts, including integrative analyses across center projects from all four components.
- Work with other consortia and international groups to facilitate data integration and joint analyses.
- Create tools (e.g., visualization tools) to enable inferences about variant impact and genome function.

4) Data and Coordinating Center (1-2 centers; $6M TC/year)

Expand and further develop a comprehensive catalog of the effects of genomic variation to enable others in the community to search for measured and predicted variant impacts.

- Build on existing IGVF efforts to develop a catalog and knowledge graph of genomic function and the molecular, cellular, and organismal effects of genomic variants.
- Enable search for information about specific variants, genomic loci, or genes derived from IGVF data, analyses, and computational predictions, and other data sources.
- Make data FAIR (Findable, Accessible, Interoperable, and Reusable) and available in a form ready for advanced machine-learning, artificial intelligence, and other computational approaches.
- Seek and incorporate user community feedback in development of catalog.
- Provide community access to data, hardened pipelines, software, standards, and resources from all components of this program as a resource of high utility and value to the research community.
- Organize and facilitate consortium activities, including convening working groups and outreach activities.

The data and coordinating center will actively seek input on the design of the catalog from within and outside the consortium. This may include seeking opportunities to work with other consortia to enable searches across resources. The center will work with all consortium members to maximize the consortium's overall scientific impact and contribute to accelerating understanding of genomic variation.

Across all components, centers are expected to coordinate assays, variants, and cell types and to develop shared analysis strategies to meet consortium goals, rather than function as independent projects with independent goals. When appropriate, consideration should be given to variants and cell types that were previously characterized in the initial phase of the IGVF consortium. Research is expected to be done in mammalian systems with a preference for human studies with inclusion of samples derived from individuals of diverse genetic ancestry. Variants selected for testing should account for allele frequencies and disease associations within and across human populations. Proposed cell types and states should be derived from tissues important in development, differentiation, or human diseases and include consideration of diseases disproportionately affecting under-represented populations. Centers will be expected to generate pipelines, tools, and educational materials that facilitate community uptake of their data and methods, and to work together to ensure all consortium resources are accessible to a wide variety of potential users. Results from variant characterization assays and prediction models should be shared early and organized in ways that enable a variety of uses by the broader scientific community from searches of individual variant impacts to large scale integrative or machine learning approaches. Applicants to all components will be expected to develop and include a plan that describes how they will foster diversity and inclusion to advance the science within their proposed projects and the IGVF consortium.

Relationship to Ongoing Activities:
NHGRI will continue to support investigator-initiated research and technology development in functional genomics and genomic variation. The IGVF consortium will continue to offer affiliate membership, which is the opportunity for researchers without IGVF funding to apply to join the consortium and contribute to consortium goals. IGVF will also continue to collaborate with other projects with a shared interest in variant interpretation. For example, IGVF will work with AVE and the Molecular Phenotypes of Null Alleles in Cells (MorPhiC) Program on variant and gene perturbation analysis and on how the resources of these efforts can be better coordinated with one another. IGVF will work with ClinGen and GREGoR to identify, prioritize, and benchmark disease relevant variants, and to understand how to best communicate and share functional genomic data so it can be used in these more clinically focused settings. Additionally, IGVF will work with the HPRC on how to incorporate state-of-the-art pangenome builds into the analysis of IGVF data.

Mechanism of Support:
All awards will be cooperative agreements. The Individual Variant Characterization Centers, Variant Interaction Centers, and Modeling and Analysis Centers will use the UM1 activity code (Research Project with Complex Structure Cooperative Agreement). The Data and Coordinating Center will use the U24 activity code (Resource-Related Research Projects--Cooperative Agreements).

Funds Anticipated:
NHGRI intends to commit approximately $36M total costs per year for 5 years for fiscal years 2026-2030 for this program.