National Advisory Council for Human Genome Research

September 9-10, 2024

Concept Clearance for RFA

GREGoRi: Catalyzing Innovation in Rare Disease Diagnosis

Purpose:

The National Human Genome Research Institute (NHGRI) proposes to renew the Genomics Research to Elucidate the Genetics of Rare Disease (GREGoR) program. The overarching goal of GREGoR is to develop novel strategies to identify the causal variant(s) in individuals with genetic disorders. This phase of the program builds upon this effort by further catalyzing the development and application of innovative methods or approaches that have the potential to make transformative advances in how we elucidate the genetics of rare disease. The objectives of this renewal, called GREGoR:Innovation or GREGoRi are to 1) develop and systematically evaluate alterative frameworks for identifying causal genes or variants in individuals with rare genetic disease, with an emphasis on the use of new and emerging genomic and molecular technologies, 2) develop novel computational methods and tools that enable a more comprehensive assessment of genomic and molecular data from individuals with rare diseases, and 3) foster connection, collaboration and data sharing among GREGoRi investigators and with the broader research community.

Background:

Over the last decade, rapid advancements in the development of high throughput and cost effective genome sequencing technologies have made it possible to genetically diagnose (e.g. identify the causal variant) a large proportion of individuals with an undiagnosed but likely genetic disease. Using methods like whole exome sequencing (WES), efforts such as the NHGRI Centers for Mendelian Genomics were able to identify the genes associated with more than 1600 rare genetic disorders. Still, a significant proportion of individuals with suspected genetic disease remain undiagnosed, even after sequencing. In 2021, NHGRI launched the GREGoR (Genomics Research to Elucidate the Genetics of Rare Disease), with the goal of developing new approaches to identify causal variants in individuals that remain undiagnosed after undergoing standard clinical genetic screening. In its first three years, GREGoR has generated a robust dataset of over 1000 families that is widely available to the research community through AnVIL, and currently includes a range of molecular data types on different subsets of families, such as WES, short read or long-read whole genome sequencing (WGS), and RNA sequencing (RNA-seq).

Recently, NHGRI hosted a community feedback workshop to identify the major gaps, challenges, and opportunities for improving the proportion of individuals that obtain a precise molecular diagnosis following genetic testing. Next generation DNA sequencing approaches, particularly panel testing, WES and short read WGS, are the current state of the art for identifying causal variants underlying Mendelian diseases. The workshop participants noted a number of areas where progress has been hampered. First, participants noted that although a number of molecular technologies that allow researchers to comprehensively evaluate DNA and its modifications, RNA, protein, and metabolites are available at scale, they are not widely in the clinical genetics setting. Participants also felt that there was a need for continued innovation in this space, including developing and/or pilot testing newer, emerging molecular technologies in the context of rare disease gene/variant discovery. It was noted that it may now be possible to consider a new paradigm for identifying the gene(s) underlying genetic disease that does not rely on sequencing as the first step. Second, participants observed that, although it is easier than

ever before to obtain WGS, it remains challenging to fully analyze these data to identify and prioritize variants that may contribute to rare disease, especially more complex variants or those located outside of coding regions. It was noted that there is a need for new computational methods and user-friendly tools that enable researchers to get the most out of whole genome sequence data, and effectively integrate other molecular data (such as RNA-seq, DNA methylation, or metabolomic data) to identify the gene(s) and variant(s) underlying rare disease. The group also noted a need for tools that better account for the potential of more complex forms of inheritance. Finally, participants discussed the current landscape of data sharing in this space and the need for continued progress developing federated solutions to sharing sequence level data.

Proposed Scope and Objectives:
This concept proposes a second phase of GREGoR that is intended to spur the development and uptake of innovative molecular and computational approaches for identifying the causal gene(s) and variant(s) underlying rare human genetic disease, that move beyond the current WES/WGS-based standard approach. This phase, called GREGoR:Innovation (GREGoRi) will be made up of four coordinated initiatives:

Technology Integration Center (1 award, $3M TC/year)
The primary goal of the Technology Integration Center is enable the development of standards and best practices for applying new and emerging molecular methods in the identification of the gene(s) or variant(s) underlying rare genetic disorders. The Technology Integration Center will be charged with generating a multidimensional dataset that can be used for the development and benchmarking of novel tools and strategies that facilitate rare disease diagnosis. The center will identify and obtain samples from at least 500 individuals with rare genetic disorders and appropriate family members, generate a core set of molecular data from all individuals, and work with the Coordination and Outreach center to make the data available to researchers. The Center will work with the Technology Innovation and Computational Innovation Projects to incorporate emerging technologies into this dataset, and to apply novel analysis methods and tools to the data, as appropriate.

Technology Innovation Projects (up to 5 awards, $3.75M TC/year)
This initiative will support proof of concept studies that explore novel strategies for identifying the gene(s) or variant(s) underlying rare genetic disorders, which take advantage of new and emerging molecular technologies. A primary goal of this initiative is to move beyond the current state of the art approach that uses DNA sequencing-based methods as an initial step. Examples of work that would be within scope include pilot studies using emerging technology like multi-omics, or that evaluate the effectiveness of using molecular methods like RNA-seq or methylation as an initial step. Technologies that show potential in these proof of concept studies will be integrated into the activities of the Technology Integration Center in later years. It is anticipated that this RFA will have two receipt dates.

Computational Innovation Projects (up to 5 awards, $3.75M TC/year)
This initiative seeks to develop user-friendly tools and/or novel methods that facilitate the identification of the gene(s) or variant(s) that contribute to rare genetic diseases. Examples of work that would be in scope of this initiative include the development of accessible tools that enable seamless, comprehensive analysis of whole genome sequencing data, or that leverage AI to integrate multiple diverse data types to inform analyses and variant interpretation, and that have entry points that serve many different kinds of users. In addition to more polished, end-stage tools, applicants may also propose innovative, early stage work, such as the development of algorithms that effectively address the potential for more complex inheritance or the combinatorial impact of multiple variants. Applicants

should leverage the GREGoR dataset, along with any other relevant data generated by other programs as described below (see: Relationship to Ongoing Activities). It is anticipated that this RFA will have two receipt dates.

Data Coordination and Outreach Center (1 award, $2M TC/year)
This Center will be responsible for logistical coordination, data dissemination, and outreach, including: a) receiving data, phenotype data, metadata, protocols and tools from the Technology Integration Center and Innovation Projects, b) ensuring that high quality data is submitted to AnVIL and other resources, and c) sharing these data and resources more broadly by participating in and enhancing federated data sharing platforms, such as Matchmaker Exchange. The Coordination Center will foster the continued development of innovative computational tools and analysis strategies to enable rare disease diagnosis by organizing community analysis events or similar workshops.

Across all components
GREGoRi will be a collaborative environment where all funded members will be expected to fully engage in consortium activities, including annual meetings, steering committee meetings and working group calls. Funded members will be expected to share pre-publication data and research results with other consortium members, and to be flexible in adapting to consortium data standards, pipelines, or other policies that promote data harmonization and interoperability. 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 GREGoRi consortium.

Relationship to Ongoing Activities:
NHGRI supports investigator-initiated research in broad scientific areas relevant to the goals of GREGoRi, including genomic technology development, and computational genomics.

A number of existing research consortia generate and make available data that could be leveraged by applicants to GREGoRi. This includes programs focused on rare disease diagnosis and discovery such as the first phase of GREGoR, the Undiagnosed Diseases Network (UDN), and the Gabriella Miller Kids First program. This also includes programs in which relevant functional data is available, such as Impact of Genomic Variation on Function (IGVF), the Encyclopedia of DNA Elements (ENCODE), Molecular Phenotypes of Null Alleles in Cells (MorPhiC), and Multi-Omics for Health and Disease (MOHD). GREGoRi will work to share standards, methods and approaches with this programs, as appropriate. Most notably, MOHD seeks to develop approaches for using multi-omic data to better understand healthy and disease states in common diseases. GREGoRi will focus on the development of innovative methods and tools that leverage similar datatypes as used in MOHD, but with a primary goal of identifying the genes and/or variants that underly rare genetic diseases.

Some of the goals of GREGoRi relate to enhancing data sharing, and developing data standards. There is an opportunity for synergy with these efforts of the Global Alliance for Genomics and Health (GA4GH). In addition, GREGoR data will continue to be shared via AnVIL, the NHGRI supported cloud-based data sharing and analysis platform.

Mechanism of Support:
Technology and Computational Innovation Projects will use the U01 activity code, with companion NOFOs for small business grants. The Technology Integration Center will use the U01 activity code. The Coordination and Outreach Center will use the U24 activity code.

Funds Anticipated:
NHGRI intends to commit $12.5M in total costs per year from FY26-FY30.