

**National Advisory Council for Human Genome Research (NACHGR)**  
**February 13, 2023**  
**Concept Clearance**  
**Human Genome Reference Program**

**Purpose**

The National Human Genome Research Institute (NHGRI) proposes to renew the Human Genome Reference Program (HGRP). The high-level objective of the HGRP is to provide an improved human genome reference to the research community that represents the diversity of human populations.

The current phase of HGRP funded the establishment of the Human Pangenome Reference Consortium (HPRC), which added high-quality genome sequence assemblies with diverse genetic ancestries to the reference and built computational representations of this multi-genome reference (aka the “pangenome”).

The goals for the renewal include: 1) continuing to maintain and improve the pangenome reference, including generation of additional reference quality sequence assemblies; 2) facilitating adoption of the pangenome reference by the research and clinical genomics communities; 3) fostering the development and deployment of informatics tools for the pangenome; 4) embedding ethical, legal and social implications (ELSI) research; and 5) forming international partnerships.

**Background**

A human genome reference sequence is an accepted representation of the human genome used as a standard to align and assemble genome sequence data. Since the completion of the Human Genome Project, the genome reference has been steadily improved by resolving errors and adding information from 15 new assemblies. New versions of the reference have been released by the Genome Reference Consortium (GRC), a collaboration between The National Center for Biotechnology Information (NCBI) and the European Bioinformatics Institute (EBI), with the current version being GRCh38. The existing reference, however, has limitations including assembly gaps, lack of haplotype resolution, and — most importantly — inadequate representation of genetic diversity that perpetuates reference bias and risks exacerbating health disparities.

The HGRP was established in 2019 to improve the human genome reference by using then-newly available methods for more complete assemblies, and by increasing genetic diversity. In the ensuing three years, funded investigators formed the HPRC, which built the foundation for a pangenome reference. Specifically, the HPRC released the first pangenome draft assembled from diverse genomes and represented as a graph, established an embedded ELSI team which guided the HPRC on issues such as consent and participant community engagement, and initiated collaborations with international partners to develop a resource representative of global populations.

Overall, this new human pangenome reference is poised to substantially increase our ability to discover, interpret, and translate genetic variants that impact health and disease. By resolving formerly hidden genetic variation in the most intractable regions of the genome, and by being more inclusive of populations across the world, the new reference will help solve rare medical conditions and allow for more powerful genetic association studies of complex traits in diverse cohorts. Also, by improving the representation of structural variation and other complex genomic regions, the new reference will provide a comprehensive foundation for functional annotations, as well as population genetics and epigenomic analyses. Thus, continuing to develop and adopt a pangenome reference will open new possibilities and greatly facilitate research and medical genomics from the lab to the clinic.

In October 2022, NHGRI hosted a workshop to identify future directions for the HGRP. The report from this meeting can be found [here](#). Meeting participants strongly encouraged the continuation of the program, with key recommendations including:

- Prioritize utility over quantity and focus on sample selection and data generation of highest benefit to the broader genomics research community.
- Emphasize adoption of the new pangenome by identifying use cases, building user-friendly tools, and minimizing disruption to existing workflows.
- Establish partnerships with international organizations and global participant communities and strive to ensure equitable benefit of the pangenome.
- Integrate ethical, legal and social considerations at all stages of research including project design, recruitment, adoption, dissemination, and access.

### **Proposed Scope and Objectives**

We propose issuing RFAs for Cooperative Agreements for three main HGRP components.

***Human Pangenome Reference Center*** (renewal; one award, \$3M/year; five years).

This component will serve as the logistic and scientific coordinating center for the HGRP and will create, improve, release, and maintain new pangenome reference versions. Key activities include:

- Explore, select, and implement state-of-the-art reference representations, building on the progress made on graph representations.
- Develop a basic tool infrastructure (i.e., for applications needed by all users including downstream tool developers) and aggregate key pangenome tools.
- Coordinate with other groups (e.g., NCBI, EBI, GRC) to do annotation, QC and other activities needed to release a useful reference to the community.
- Emphasize engagement with different broad sectors of the research and clinical communities to accelerate adoption of the reference.
- Identify and oversee “adopter projects” in partnership with existing genomics consortia to use the pangenome, identify and address challenges, and collect data on the scientific and clinical value of adopting the pangenome reference.
- Coordinate outreach to international partners, research collaborators, and participant communities with the larger goal of creating an inclusive and representative resource.

***Center for High-Quality Reference Genomes*** (renewal, one award, up to \$3M/year; five years).

This component will provide high-quality human genome sequence data and assemblies to fill gaps in representation, including:

- Establish criteria and metrics for prioritizing and adding new assemblies to the pangenome in a way that balances scientific and social priorities, considers both quantity and quality, and emphasizes utility for basic and clinical studies.
- Identify, recruit, and collect new or existing samples and datasets that will fulfill these criteria and metrics. This in turn will require working with both researcher and participant communities.
- Generate reference quality sequence data for at least 200 diploid genomes. An updated evaluation of sequencing and funding needs will be made based on state of the science and sequencing costs at the time of the award.
- Support an embedded ELSI component focused on issues such as diversity and equity of the pangenome resource, open access, data privacy and sovereignty, regulatory aspects of clinical use, and participant community engagement.

**Informatics Tools for the Pangenome Reference** (new; 4-6 awards, \$2M total year 1; \$4M total year 2, possible re-issue after evaluation; multiple receipt dates; 3 years).

This new component addresses the high priority need to facilitate uptake and use of the pangenome reference. Emphasis will be on tools for common use cases that are relevant to different broad sectors of the genomics community, e.g., clinical, population, and functional genomics. Possible examples include selecting the best subset of linear genomes or paths along the graph for a given set of samples, visualizing complex variation, and annotating regulatory elements and disease associations. These tools will build on those developed by the Human Pangenome Reference Center, with the latter focusing on general computational infrastructure for pangenome use.

Applicants to all three FOAs will be expected to include a plan to advance the scientific and technical merit of the proposed project through diversity and inclusion. This is particularly important given the program focus on a resource representative of worldwide populations.

### **Relationship to Ongoing Activities**

HGRP awardees will work directly with the GRC which has a role in coordinating with major genomics databases, liaising with groups annotating the reference, providing views of the reference to the community, and receiving community feedback.

The HGRP will coordinate with GA4GH regarding reference standards and establishing international partnerships. The HGRP will continue to pursue relationships with other international genomics organizations, such as H3Africa and the Australian National Centre for Indigenous Genomics.

HGRP data will continue to be publicly available through multiple genomic resources. Cell lines developed from pangenome samples will be available through Coriell.

NHGRI encourages related investigator-initiated research outside of the HGRP itself, including development of improved computational tools for representing and use of pangenome references, technology development for high quality genome sequencing and assemblies, and independent ELSI research. HPRC will encourage associate membership for independently funded investigators.

### **Mechanism of Support**

RFAs for cooperative agreements (UM1) will be published for the Human Pangenome Reference Center and the Center for High-Quality Human Reference Genomes. An RFA for U01s will be issued for Informatics Tools, with a companion FOA for Small business grants. Notices or Program Announcements (without dedicated funds) may be made for complementary investigator-initiated research as described above. Adopter projects will be supported through the Human Pangenome Reference Center with additional funds available to "adopter consortia" through supplements or co-funding.

### **Limited Competition**

We propose issuing the Human Pangenome Reference Center and the Center for High Quality Reference Genomes components as limited competitions. These integrated centers have established infrastructure and critical collaborations that would be very difficult and time consuming to re-establish. This includes the efforts in generating a pangenome reference representation, the embedded ELSI component, interactions with associate consortium members, and existing and burgeoning international partnerships.

### **Funds Anticipated**

NHGRI will commit up to \$10M in total costs per year over 5 years (FY24-FY28) for a total of \$50M. Co-funding will be sought especially for outreach and adoption.