Human Genome Reference Program (HGRP) Pre-Application Webinar for RFA-HG-23-026 (Informatics Tools for the Pangenome)

Adam Felsenfeld, Alexander Arguello, Carolyn Hutter, Jyoti Dayal, Nicole Lockhart, and Sara Currin

National Human Genome Research Institute, Division of Genome Sciences

Thursday, July 20th, 2023 at 12:00PM Eastern

The Forefront of Genomics®

Format

Overview: The HGRP (Adam Felsenfeld – 10 min)

RFA-HG-23-026: Informatics Tools for the Pangenome (Adam Felsenfeld – 15 min)

Questions (All - 30 min)

Key Takeaways

- Please look at the Review Criteria specific to this NOFO (Notice of Funding Opportunity)
- Submission deadlines:
  - **Wednesday, November 1st, 2023**
  - Monday, March 3rd, 2025

NIH NHGRI

To Note

- This call will be recorded and posted to the NHGRI HGRP webpages
- Your questions may be rendered into FAQs with answers linked to the HGRP webpages
- You do not have to identify yourself to ask a question
- Please use the Q&A feature on Zoom to ask questions

Part 1: Overview

The Human Genome Reference Program (HGRP)

HGRP Goals 2019-2024

- Adequately represent human haplotype variation – add 350 new genomes to build and improve the human reference resource
- Easy and effective computational representations of reference data

Later added:

- Embedded ethical, legal, and social implications (ELSI) scholars
- International partnerships

Overall:

- Better representation of human variation in the human reference resource

NIH NHGRI

- Includes sequences from 47 people
- Four papers recently published

NIH NHGRI

HGRP Structure 2019-2024

Human Pangenome Sequencing Center

UNIVERSITY OF CALIFORNIA SANTA CRUZ

Genome Reference Representations

USC University of Southern California

Dana-Farber Cancer Institute

Stanford University

Human Pangenome Reference Center

The Elizabeth H. and James S. McDonnell III McDonnell Genome Institute at Washington University

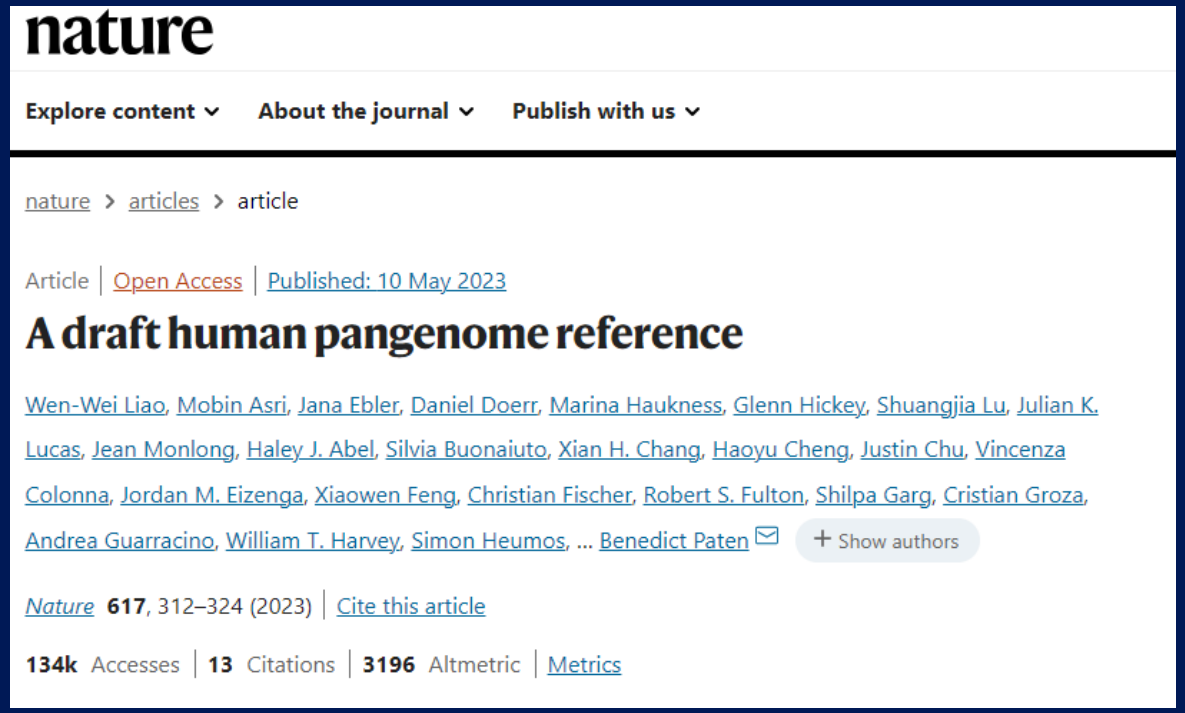
Comprehensive Human Genome Sequencing

UNIVERSITY OF CALIFORNIA SANTA CRUZ

ELSI work across all 3 components

NIH NHGRI

HGRP Products 2019-2023



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Open Access | Published: 10 May 2023

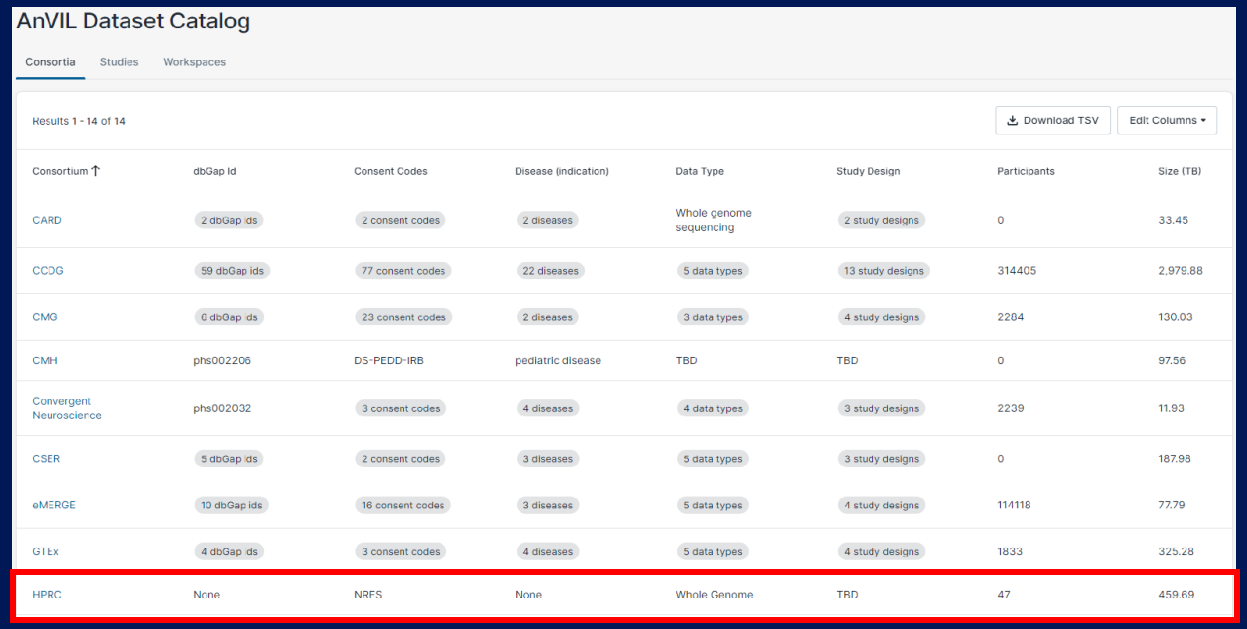
A draft human pangenome reference

Wen-Wei Liao, Mobin Asri, Jana Ebler, Daniel Doerr, Marina Haukness, Glenn Hickey, Shuangjia Lu, Julian K. Lucas, Jean Monlong, Haley J. Abel, Silvia Buonaiuto, Xian H. Chang, Haoyu Cheng, Justin Chu, Vincenza Colonna, Jordan M. Eizenga, Xiaowen Feng, Christian Fischer, Robert S. Fulton, Shilpa Garg, Cristian Groza, Andrea Guarracino, William T. Harvey, Simon Heumos, ... Benedict Paten + Show authors

Nature **617**, 312–324 (2023) | Cite this article

134k Accesses | **13** Citations | **3196** Altmetric | Metrics

Initial draft pangenome consisting of nearly 50 human genome assemblies



AnVIL Dataset Catalog

Consortia | Studies | Workspaces

Results 1 - 14 of 14 | Download TSV | Edit Columns

Consortium	dbGap Id	Consent Codes	Disease (indication)	Data Type	Study Design	Participants	Size (TB)
CARD	2 dbGap ids	2 consent codes	2 diseases	Whole genome sequencing	2 study designs	0	33.45
CCDG	59 dbGap ids	77 consent codes	22 diseases	5 data types	13 study designs	314405	2,979.88
CMG	6 dbGap ids	23 consent codes	2 diseases	3 data types	4 study designs	2284	130.03
CMH	phs002206	DS-PEDD-IRB	pediatric disease	TBD	TBD	0	97.56
Convergent Neuroscience	phs002032	3 consent codes	4 diseases	4 data types	3 study designs	2239	11.93
CSER	5 dbGap ids	2 consent codes	3 diseases	5 data types	3 study designs	0	187.93
eMERGE	10 dbGap ids	16 consent codes	3 diseases	5 data types	4 study designs	114118	77.79
GTEx	4 dbGap ids	3 consent codes	4 diseases	5 data types	4 study designs	1833	325.28
HPRC	None	NRES	None	Whole Genome	TBD	47	459.69

Sequencing data, assemblies, and pangenomes in AnVIL

Current Sampling Strategy

- 1000G samples so far, BioMe going forward
- Multiple approaches to prioritizing samples
  - Maximize genetic variants – choose samples within populations that increase the number of common variants (MAF>1%) represented
  - Maximize genetic distance – choose samples, agnostic to population, that increase the genetic distance between any two samples

HGRP Renewal 2024 Planning

- 2022 HGRP Planning Meeting Report
- Prioritize utility over quantity; high benefit to genomics research community
- Adoption of the pangenome: use cases, user-friendly tools, minimize disruption to workflows
- Establish partnerships with international organizations and global communities. Strive to ensure equitable benefit
- Integrate ethical, legal, and social implications (ELSI) at all stages of research including project design, recruitment, adoption, dissemination, and access

HGRP Renewal Goals (2024-2029)

To produce a human pangenome reference that optimizes both the population genetic diversity represented, and also the utility for, and adoption by, the genomics research community.

- **Maintain and improve pangenome reference**
- **Facilitate adoption by research and clinical genomics communities**
- **Foster development, deployment of informatics tools for the pangenome**
- **Embed ELSI research**
- **Foster international partnerships, collaboration for global adoption**

NIH NHGRI

HGRP Renewal Structure (2024-2029)

High Quality Reference Genomes	Human Pangenome Coordinating Center	Informatics Tools for the Pangenome
Limited Competition	Limited Competition	**NEW!** Open Competition

ELSI work across all 3 components

NIH NHGRI

Part 2: RFA-HG-23-026

Informatics Tools for the Pangenome

NIH NHGRI

Informatics Tools for the Pangenome

Objectives:

- To facilitate uptake and use of the human pangenome reference
- To advance compelling use cases relevant to broad sectors of the genomics community (clinical, population, functional)
- Useful to investigators/clinicians with a range of genomics informatics expertise and resources

Application Suggestions

- Can be:
  - New tools to take advantage of information in pangenome reference and graph representation
  - Modifications of existing tools to enable important use cases of the pangenome
- Should:
  - Have the greatest utility to the genomics research community and follow best practices in software development
  - Break the path for other scientists to benefit from the pangenome reference

Application Examples

- Selecting the best subset of linear genomes or paths along the pangenome graph for use in genome alignments of short read sequence data
- Creating a data visualization tool to aid in identifying structural variation in the pangenome
- Using genetic variation in the pangenome to inform population genetic or comparative genomic analysis
- Integrating the pangenome with multi-omic or functional genomic data to improve variant characterization in genomically diverse populations

These examples are intended to convey the intent and are not the only responsive topics!

More Application Examples

- Visualizing regulatory elements, or other functional genomic features within the pangenome
- Using the pangenome reference to improve fine mapping and polygenic risk scores
- Using the pangenome to annotate disease associations or clinically relevant variants
- Integrating controlled access genomic data with open access pangenome data in ways that enable secure and private use of protected data

These examples are intended to convey the intent and are not the only responsive topics!

NIH NHGRI

Function and Coordination

- All three components (Coordinating Center, Genomes Center, Tools Centers) will work collaboratively towards production and community adoption of the human pangenome reference
- Tools NOFO seeks development of user applications; Coordinating Center will support infrastructure and delivery of a human pangenome reference
- Both Tools & Coordinating Center will encourage and support the use and adoption of the pangenome by the broader research community
- Existing assembly representation and tools available here

NIH NHGRI

A Word About Reference Representations

Tools development may need to anticipate HPRC plans for reference representations moving forward

Language from NOFO:

- Current representation is graph-based; other representations (e.g. k-mer) may evolve to accommodate scale and complexity
- Applicants should address how their tool works with current graph-based representation, and how they could work with, and adapt to, other ways in which the pangenome may be represented.

Relevant Links

Current NOFO: RFA-HG-23-026: Informatics Tools for the Pangenome (U01 Clinical Trial Not Allowed)

Companion Funding Opportunities:

- RFA-HG-23-024: Limited Competition: High Quality Reference Genomes (UM1 Clinical Trial Not Allowed)
- RFA-HG-23-025: Limited Competition: Human Pangenome Coordinating Center (U41 Clinical Trial Not Allowed)

Email contact: adam_felsenfeld@nih.gov

Link to this Webinar

Program Webpage

2022 HGRP Planning Meeting Report

2020 NHGRI Strategic Vision

NIH NHGRI

Reminders

- Please look at Review Criteria specific to this NOFO
- Submission deadlines:
  - **Wednesday, November 1st, 2023**
  - Monday, March 3rd, 2025

NIH NHGRI

Questions?

- Please use Q&A
- Email Adam
- FAQ will be posted


The Forefront
of Genomics®