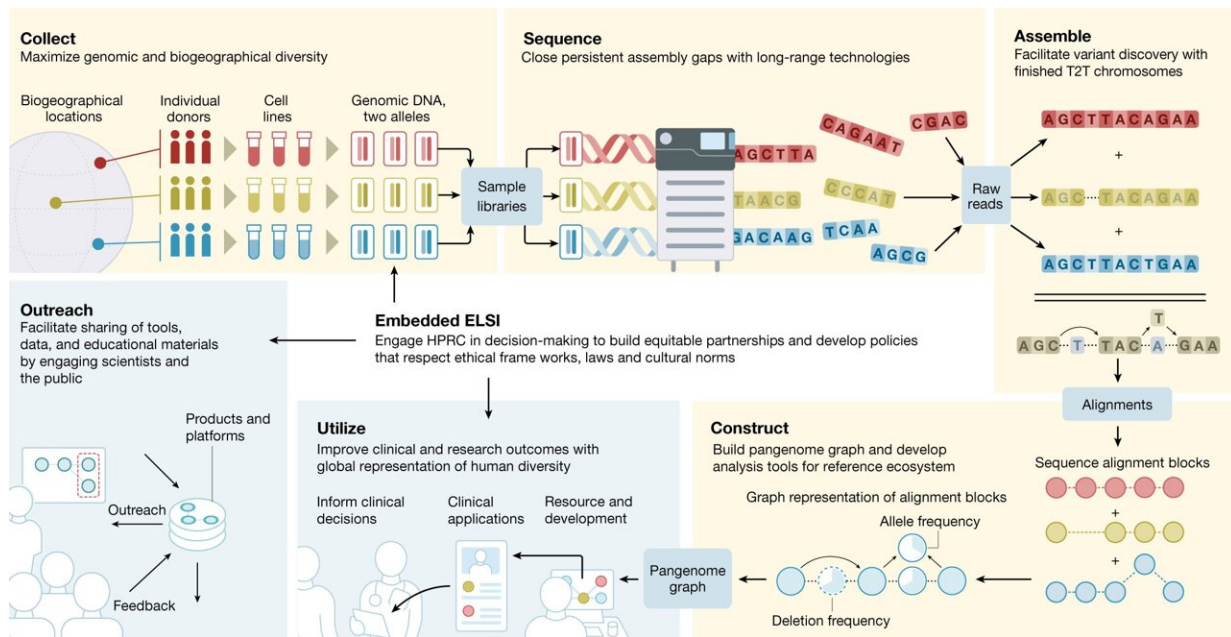


Scientists set out to map the world's genomic diversity

April 20 2022, by Michael McCarthy



An overview of several components of the HPRC. Collect: 1,000 Genomes samples start the project and will be followed by additional samples collected through community engagement and recruitment. Sample selection efforts will ensure that the graph-based reference captures global human genomic diversity. Sequence: long-read and long-range technologies are used to generate genome graphs and bridge gaps in difficult-to-assemble genomic regions. Assemble: T2T finished diploid genomes will foster variant discovery, especially in complex, difficult-to-assemble genomic regions. Construct: scalable bioinformatics approaches assemble, quality control, call variants and benchmark graph assembly accuracy. The graph is annotated with gene descriptions and transcriptome data, making it more accessible and interpretable. Utilize: collaboration across scientific and stakeholder communities will create a new

ecosystem of analysis tools. Clinical applications and research use will involve analysis, validation, interpretation and publication of results. Outreach: members of the HPRC outreach community engage and educate the user community and broadly share all genomic products and informatics platforms. ELSI: ELSI scholars will develop selection processes and policy frameworks that meet investigator needs as well as respect research partner autonomy and cultural norms. Credit: *Nature* (2022). DOI: 10.1038/s41586-022-04601-8

An international consortium of scientists has launched a new effort to create a reference genome that captures the genetic diversity of all the peoples of the world. The researchers describe the initiative, called the Human Pangenome Project, in a paper published Wednesday, April 20, in the journal *Nature*.

"The goal is to collect, organize and make accessible a representation of all the [genetic variation](#) that exist in humans, big, small, common and rare," said Evan Eichler, a professor of [genome](#) sciences at the University of Washington School of Medicine in Seattle and a co-author of the paper.

More than a dozen research institutions are collaborating on the international project. The effort is funded by the U.S. National Human Genome Research Institute and National Institutes of Health.

The human genome consists of sequence of 3.1 billion DNA molecules. Overall, our genomes are remarkably similar. But small differences in the sequences of our genomic DNA play a large role in determining what makes each individual unique, including his or her risk for disease.

A reference genome helps describe those differences by mapping the location of genes and other elements of the genome. Researchers use this map to identify new genes, variants of known genes and other functional

elements, and to share and compare their findings with other scientists.

Earlier this spring, scientists announced that they had finished a two-decades long effort create a genome reference that represented a complete map of a human genome. This [reference genome](#) is a composite created from the DNA sequences of a very small group, about 20 individuals, with most of the sequence coming from a single individual. As a result, the reference does not reflect the diversity seen in the world's population. Indeed, most studies of the [human genome](#) are based on samples from individuals of European origin.

The new initiative seeks to address this shortcoming by creating multiple complete reference genomes representing hundreds of people from around the world. "We don't want to have a (single) reference; we want to have many references that will capture human diversity," Eichler said.

An important part of the initiative will be obtaining the consent and collaboration from groups who might be wary of efforts by Western scientists to sequence their population's DNA, Eichler said. "Some groups have said, 'Thanks, but no thanks.' And we have to respect their autonomy."

Other populations have expressed interest, but want to do their own genome-sequencing and make the results available on their terms. The project is trying to provide these communities with the technology and skills they need to conduct this science.

"We think there are huge advantages for everyone if we had all people represented in the pangenome, but we need to let people go at their own pace and make their own decisions," Eichler said.

The initial goal of the project is to create genome sequences from 350 individuals from diverse populations over the next five years. The

researchers will use a laborious and time-consuming process called "long read" sequencing that allows them to map an entire genome, error-free, from end to end. Previous methods yielded sequences riddled with gaps, often missing large segments of genome. Ultimately, the project organizers hope to sequence thousands of genomes to capture as much human [genetic diversity](#) as possible.

The long-range goal is to someday enable a person anywhere in the world to go into a doctor's clinic and have a DNA sample collected and sequenced. The patient's individual sequence could then be compared with the reference pangenome to characterize the patient's genotype, providing information about their genetic risk for cardiovascular disease, diabetes and other conditions.

It may be years before such tests are available for patient care, Eichler cautioned. "We're just at the beginning. Lots of things need to be worked out. But we really believe this is the future of human genetics."

More information: Ting Wang et al, The Human Pangenome Project: a global resource to map genomic diversity, *Nature* (2022). [DOI: 10.1038/s41586-022-04601-8](#)

Provided by University of Washington School of Medicine

Citation: Scientists set out to map the world's genomic diversity (2022, April 20) retrieved 22 June 2024 from <https://phys.org/news/2022-04-scientists-world-genomic-diversity.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.