The Human Genome Reference Program

A Concept for Program Renewal

NACHGR

February 13, 2023

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

NIH National Human Genome Research Institute

The **Forefront** of **Genomics**®

Thanks To

Xander Arguello

Sara Currin

Adam Felsenfeld

Carolyn Hutter

Nicole Lockhart

Heidi Sofia

Mike Smith

Kris Wetterstrand

Current human reference sequence:

Better human reference sequence(s):

And more. But not enough…

Need to Represent Many Genomes at Once- One Way:

Example from sequenceTubeMap
https://github.com/vgteam/sequenceTubeMap/blob/master/images/header.png

Current Program

- Adequately represent human variation – Goal: 350 new genomes
- Develop better ways to represent a multi-genome reference
- Build and maintain a useful pangenome resource
- Further develop methods to sequence and assemble complete genomes

Later added:

- Embed ethical, legal, and social issues (ELSI) scholars
- Foster international partnerships

HGRP 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

GA4GH

GRC (NCBI, EBI)

Human Pangenome Reference Center

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

T2T

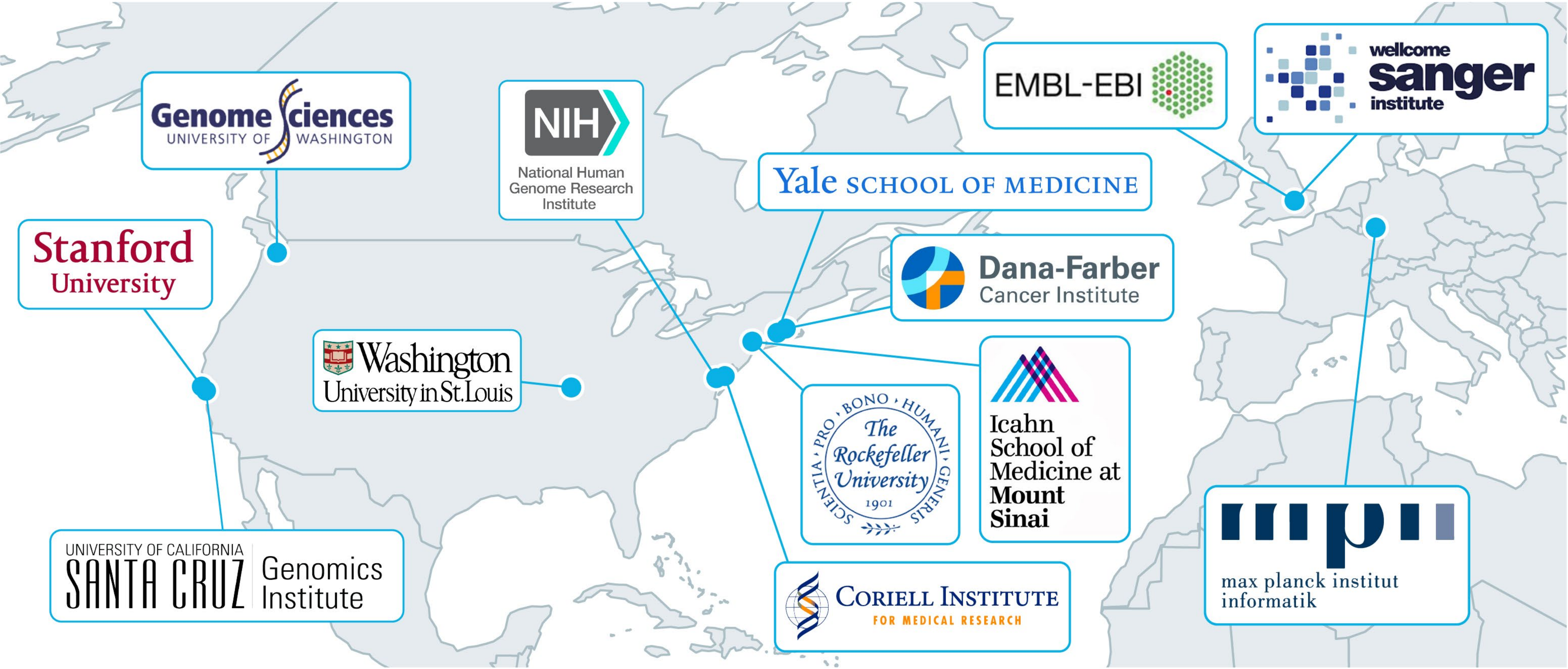
Tech Dev for Comprehensive Human Genome Sequencing

University of California Santa Cruz

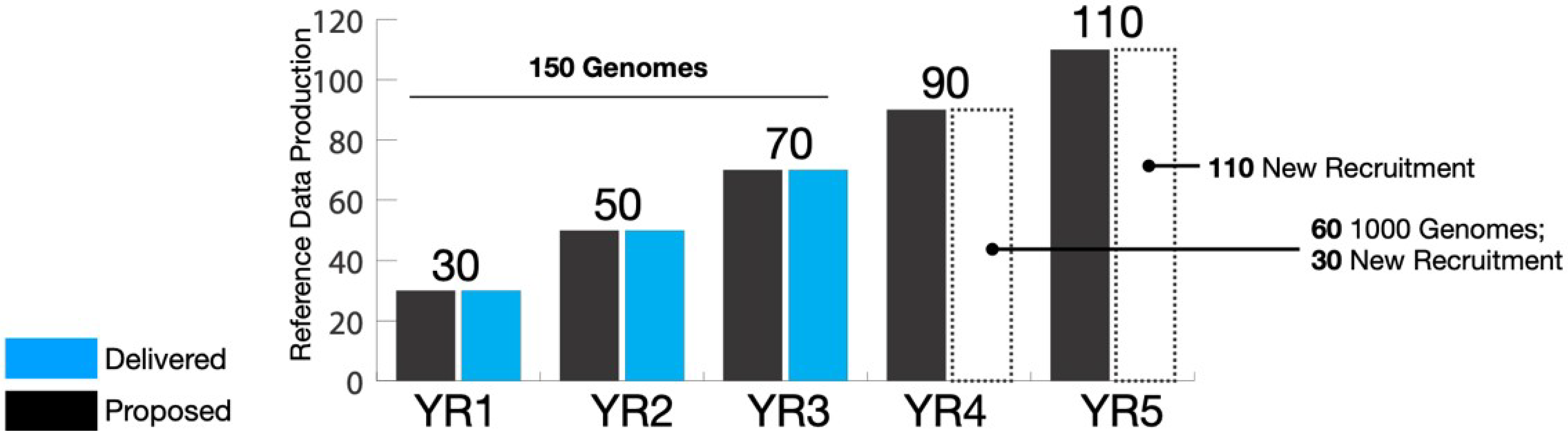
ELSI

Human Pangenome Reference Consortium: HPRC

https://humanpangenome.org



HGRP Progress– High Quality Genome Assemblies

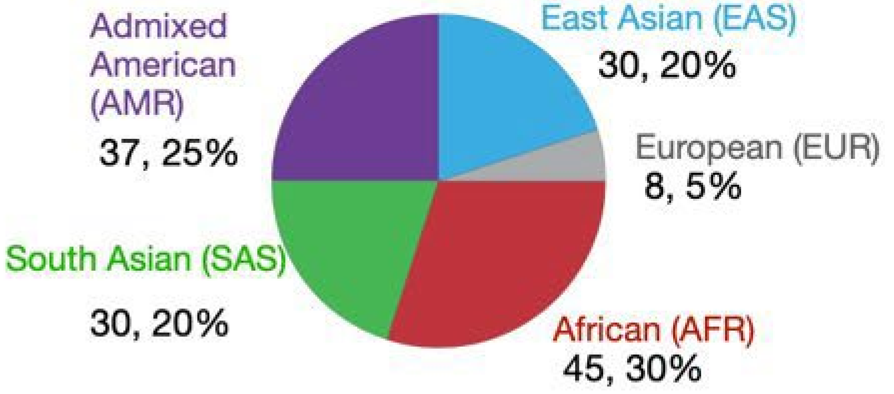


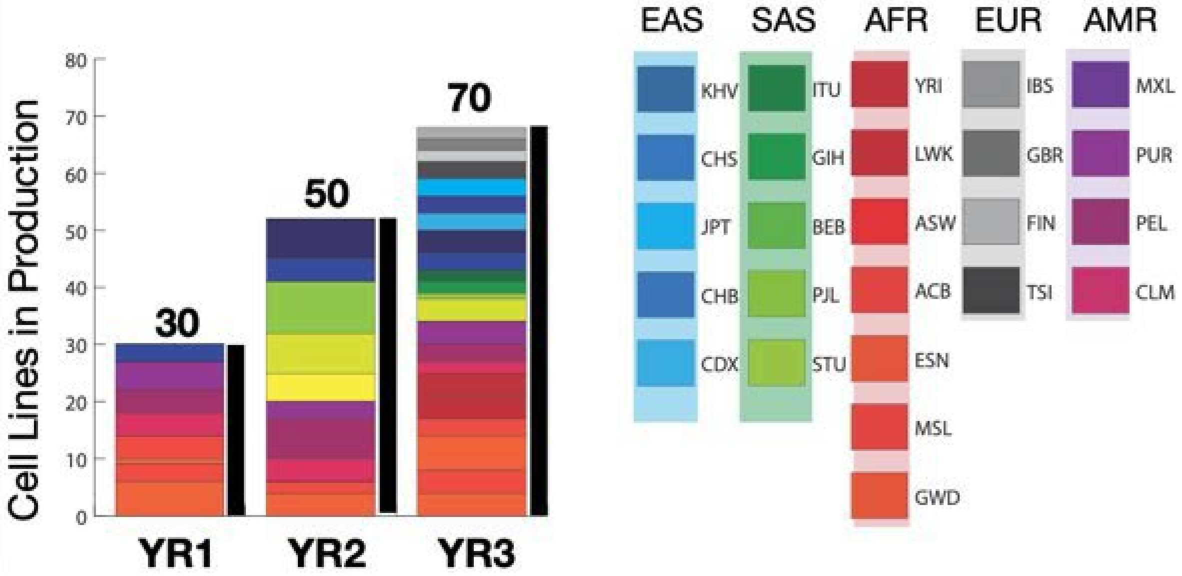
Credit: HPRC and Deanna Church

Phase 1 HPRC Sample Selection

- ✓ Cover genetic and geographic diversity
- ✓ Availability of low passage cell lines
- ✓ Availability of trios/parental data (YR1-2)

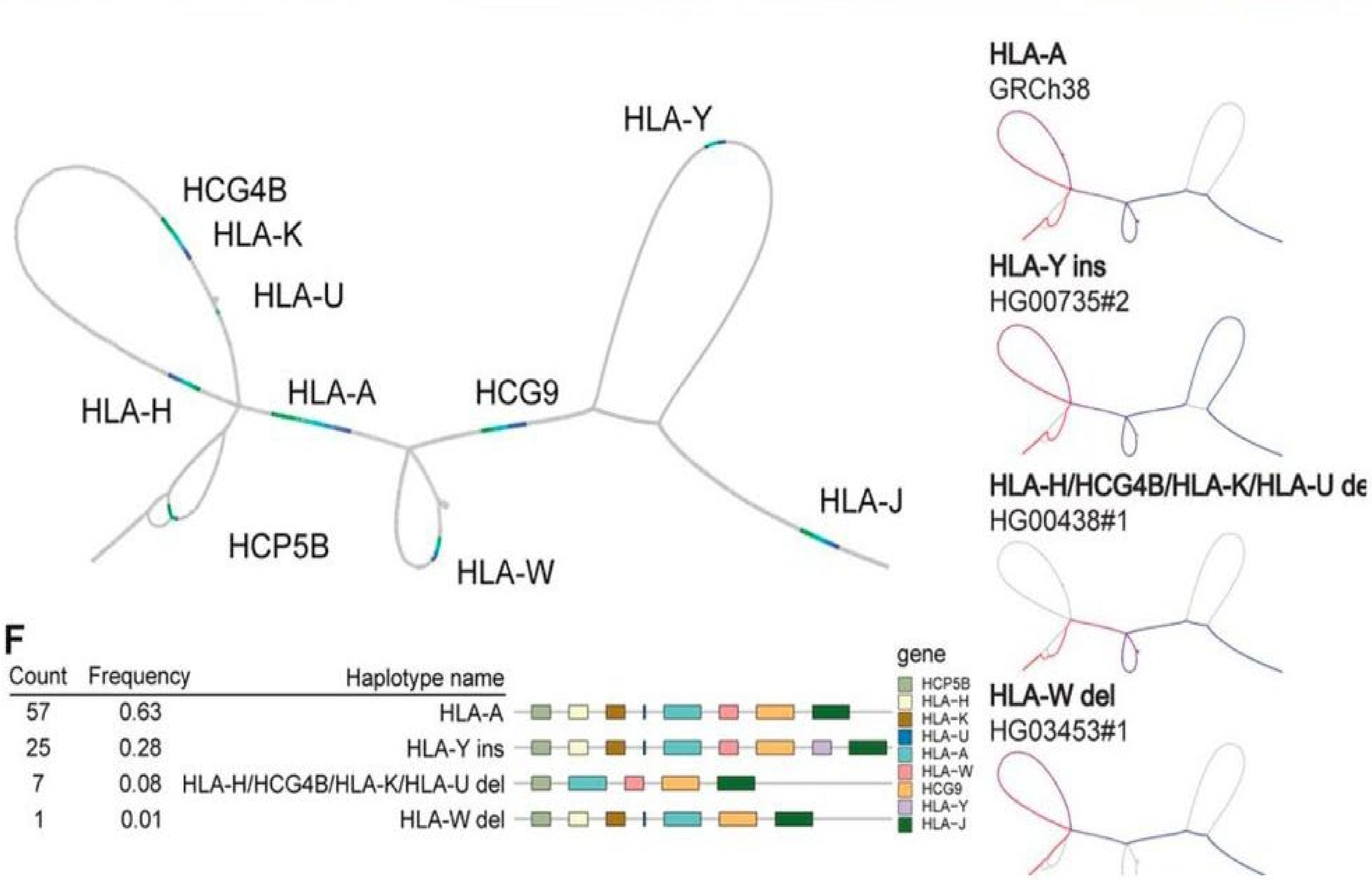
Phase 1 Production: 150 total lines





Credit: HPRC and Deanna Church

Progress: Pangenome Representation



Count	Frequency	Haplotype name
57	0.63	HLA-A
25	0.28	HLA-Y ins
7	0.08	HLA-H/HCG4B/HLA-K/HLA-U del
1	0.01	HLA-W del

Credit: HPRC and Deanna Church

Background: Planning Workshop October 2022

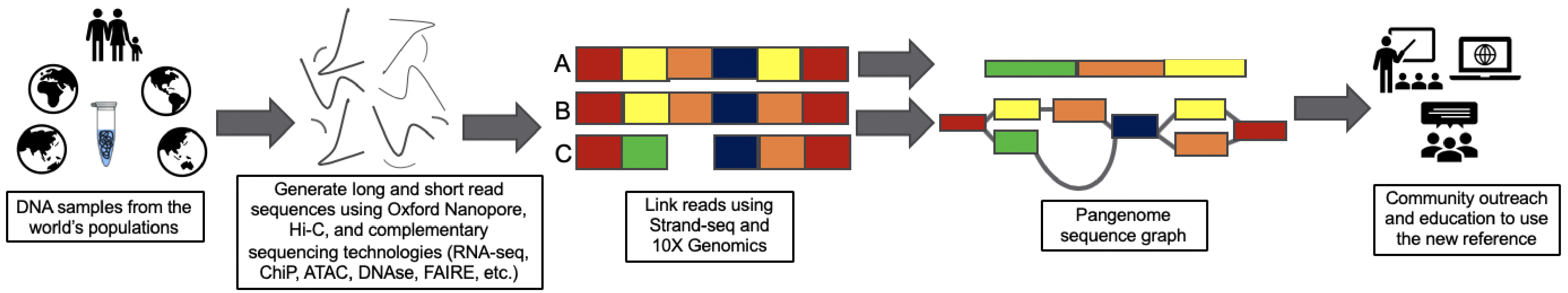
Topics:

1. **Samples and sequencing**
2. **Representation and implementation**
3. **Dissemination**
4. **Engaging worldwide partners**

Key Recommendations:

- **Focus on demonstrating benefit to a broad basic and clinical user-base.**
- **Emphasize adoption and practical implementation. Identify exemplar "adopter" projects. Build user-friendly tools.**
- **Establish partnerships with international organizations to ensure equitable benefit.**
- **Integrate ethical and social considerations at all stages of research.**

HGRP 2019-2024



DNA samples from the world's populations

Generate long and short read sequences using Oxford Nanopore, Hi-C, and complementary sequencing technologies (RNA-seq, ChiP, ATAC, DNAse, FAIRE, etc.)

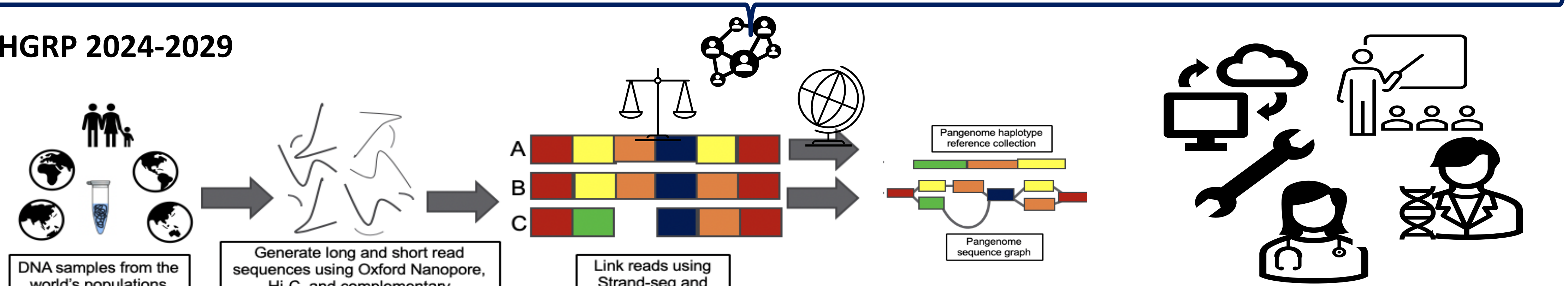
Link reads using Strand-seq and 10X Genomics

Pangenome sequence graph

Community outreach and education to use the new reference

350

HGRP 2024-2029



DNA samples from the world's populations

Generate long and short read sequences using Oxford Nanopore, Hi-C, and complementary sequencing technologies (RNA-seq, ChiP, ATAC, DNAse, FAIRE, etc.)

Link reads using Strand-seq and 10X Genomics

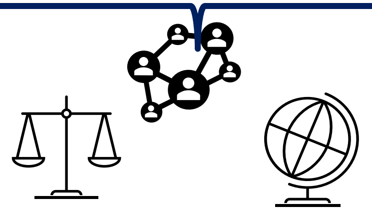
Pangenome haplotype reference collection

Pangenome sequence graph

R&D -> Invest. Init.
Implement -> HGRP

Emphasize computational tools, use-cases, outreach, adoption

200



Proposed Scope and Objectives

FOA 1: Human Pangenome Reference Center

- **Construct and release new pangenome reference versions**
- **Implement state-of-the-art reference representations**
- ***New emphasis*: Community adoption will be emphasized in this RFA and others. Use an “adopter projects” model and other means**
- **Develop/aggregate basic tools/infrastructure for applications needed by all for use of the pangenome reference**
- **Coordinate outreach and collaborations, including international.**
- **Liaise with other genomic resources**
- **Logistic and scientific coordinating center for the program.**
- **Cooperative agreement; one award, $3M/year; five years**

FOA 2: Center for High-Quality Reference Genomes

- **Identify, recruit, and collect new samples (or existing data sets). This includes developing criteria for prioritizing that considers quantity, quality, cost, and added utility for basic and clinical studies.**

- **Generate reference quality sequence data for at least 200 diploid genomes for addition to the pangenome reference.**

- **Support embedded ELSI research into the creation and use of a human pangenome.**

- **Cooperative agreement; one award, $3M/year; five years**

FOA 3: *New:* Informatics Tools for the Pangenome Reference

- **Address high-priority need to facilitate uptake and use of the pangenome reference.**
- **Emphasize tools for common use cases relevant to different broad sectors of the genomics community, e.g. clinical genomics, population genetics, functional genomics.**

  Examples:

  - **Selecting the best subset of linear genomes or paths along the graph for analysis of a given set of samples/variants**
  - **Visualizing complex variation**
  - **Annotating functional elements and disease associations**
- **These tools will complement those developed by the Human Pangenome Reference Center, which will focus on general infrastructure for pangenome use.**
- **Cooperative agreements; 4-6 awards, $2M total year 1; $4M total available in Y2-5; multiple receipt dates**

Relationship to Ongoing Activities

- Continue to pursue relationships with international genomics organizations with common interests.
- Liaise with other large genomic resources.
- Encourage associate membership.
- Leverage existing investigator-initiated and SBIR programs that could mesh. Consider ways to stimulate.

Limited Competition

- For the “Human Pangenome Reference Center” and the “Center for High Quality Reference Genomes” components.

- Investigators have laid a solid foundation for a Human Pangenome Community resource.

- Created an open community of investigators building the resource. Established many key relationships: ELSI embed, associate consortium members.

- Leading the organization of an international alliance around the idea of a human pangenome, together with GA4GH .

- Established key policies/standards for the pangenome, including consent, sample prioritization for diversity, assembly quality metrics, and others.

Re-building these relationships and policies would cause long delays.

Open Competition:

- Tools FOA

New Ideas:

- Investigator initiated research (incl. SBIR)- e.g. pangenome representation; structural variation; tools; tech dev
- Adoption projects
- Associate consortium membership

Summary of Proposed Annual Budget

	Mechanism	FY24	FY25	FY26	FY27	FY28
Pangenome Reference Center	**U41**	**3**	**3**	**3**	**3**	**3**
HQ Reference Genomes	**U41**	**3**	**3**	**3**	**3**	**3**
Pangenome Tools Y1	**U01**	**2**	**2**	**2**		
Pangenome Tools Y2	**U01**		**2**	**2**	**2**	
Tools Y4*	**U01**				**2**	**2**
Total		**$8M**	**$10M**	**$10M**	**$10M**	**$8M**

* If funds are available in Y3, this would take place then. Year 4 awards would extend to FY 29.

END