Human Genome Reference Program Webinar

Adam Felsenfeld, Heidi Sofia, Michael Smith

March 8, 2019

NIH National Human Genome Research Institute

The **Forefront** of **Genomics**®

WebEx

- You will automatically be **muted** when joining the conference.
- You can connect to audio under the “Quick Start” tab. You can have WebEx call you by inputting your number, or by calling **650-479-3208** (a toll number) and using the access code **624 755 444**.
  - Additionally, you can connect to audio through your computer with headphones or a Bluetooth device

Questions

- Email questions to **gsp@mail.nih.gov** or type them in the WebEx chat box
  - To anonymously send questions via WebEx chat, message them privately to Taylorlyn Stephan's account
- To speak, please "raise your hand" to be unmuted:

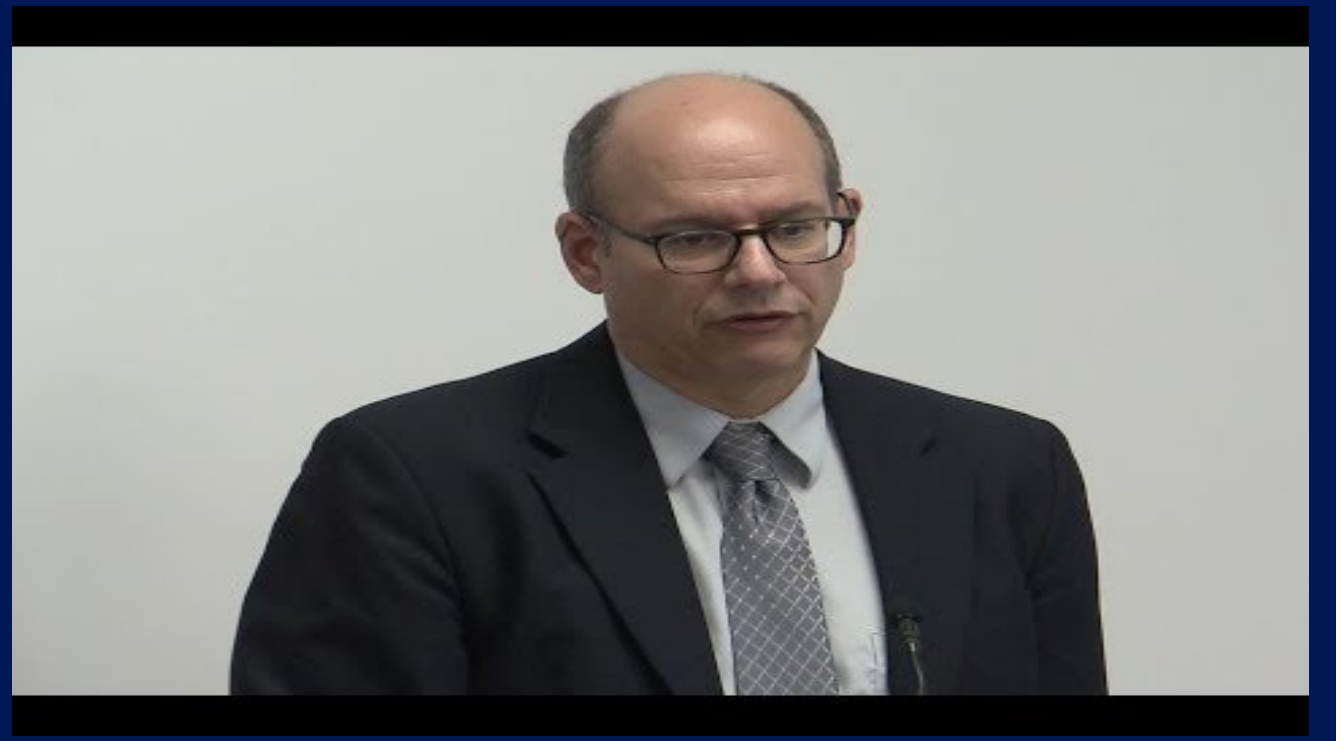
Participant

If you want to speak and the host has muted your microphone, you can ask the host to unmute your microphone so that you can speak. You can cancel a request to speak at any time.

To ask to speak	Windows	result
Select **Raise Hand** on the Participants panel		The **Raised Hand** indicator appears on the participant list for the host and presenter.

To ask to speak	Mac	result
Select the Raise Hand icon at the bottom of the Participants panel		The **Raised Hand** indicator appears on the participant list for the host and presenter.

Background Documents



Rapporteur: Taylorlyn Stephan
NHGRI Program Analyst

Human Genome Reference Sequence Opportunities Webinar – March 1, 2018
Executive Summary

Background
On March 1, 2018, over 65 basic research, clinical, and bioinformatic scientists in the genomics community convened for a four-hour web meeting on NHGRI-funded components of the Genome Reference Consortium (GRC). The GRC is an essential resource for the biomedical community and due to recent technological advances, NHGRI staff determined it was time to evaluate current GRC activities and discuss its possible future scope. The web meeting addressed these topics: key research and resource opportunities for improving the human reference; activities necessary to keep the reference relevant and useful; clinical and research community needs (including education); related resources; and collaborations. Several major themes and recommendations emerged during this web meeting.

A Pan-Genome
Many meeting participants expressed the need for a "pan-genome" – a reference genome that represents all human variation. The pan-genome will facilitate alignment for every sequence of interest, with a primary goal that no haplotype will go unaligned. The pan-genome would replace the current reference build, GRCh38, and its multiple alternative paths, as they do not fully represent the diversity of variation and haplotypes in humans. Additional sample collection and sequencing of diverse genomic ancestries were strongly advised for inclusion in the pan-genome. 1000 Genomes Project samples, or samples with the same consent categories and diverse provenance, were regarded as the highest sequencing priority. A corollary request to these additional samples were ones with original blood available for use in genome characterization and benchmarking. While there was discussion on incorporating existing genomes from large-scale sequencing projects, participants agreed that new samples would be easier to incorporate for several reasons: 1) samples from populations missing in GRCh38 could be obtained, 2) they could all be sequenced at the same depth and quality at one or a few institutions in a narrow timeframe, 3) their consents would be standard. For sequencing efforts, a pilot of 50-300 new genomes with haplotype resolution was suggested. There was not a detailed conversation or consensus as to what sequencing technologies or metrics (e.g., 30X sequencing, PacBio long reads, Hi-C) should be used in the pilot. Participants noted that how the pan-genome is displayed (in a graph, linear coordinates, etc.) should not limit its development. If most users can understand and access the data, then the reference format should be a back-end concern. Community members also argued that a pan-genome effort will motivate tool development and use.

Bioinformatic Tools
Several participants argued that bioinformatic tool development for use with the reference has been severely underfunded. In addition, existing tools are difficult to use or find, so future tool development should emphasize ease-of-use for the average biologist and medical scientist. There was an argument that tool development will follow naturally from a pan-genome effort, but there still is a critical lack of bioinformatic tools for current users of builds GRCh37 and GRCh38.

Human Genome Reference Center (HGRC)

Due: April 2, 2019 by 5PM local time

Program Contact: Adam Felsenfeld (felsenfa@mail.nih.gov)

High Quality Human Reference Genomes

Due: April 2, 2019 by 5PM local time

Program Contact: Adam Felsenfeld (felsenfa@mail.nih.gov)

Research and Development for Genome Reference Representations (GRR)

Due: April 2, 2019 by 5PM local time

Program Contact: Heidi Sofia (heidi.sofia@nih.gov)

Developing Comprehensive Human Genome Sequencing Methodologies

Due: June 27, 2019 by 5PM local time

Program Contact: Michael Smith (smithmw@mail.nih.gov)