NHGRI Planning Meeting for the Human Genome Reference Program
October 13, 2022
Executive Summary

On October 13, 2022, the National Human Genome Research Institute (NHGRI) sponsored *The Human Genome Reference Program (HGRP) Planning Meeting*, at the Bethesda Marriott hotel in Bethesda, Maryland. The objective was to gather input on the next steps for the HGRP, and to anticipate future needs for the human genome reference resource. The current round of funding for the HGRP will end in 2024. NHGRI convened about 100 members of the genomics community. Four broad topics were discussed: samples and sequencing; representation and implementation of a pangenome resource; dissemination --- versioning, user outreach/education, and user tools; and engaging worldwide partners. For the full meeting agenda see Appendix A.

Meeting participants emphasized several high-level points:

- Continuing the HGRP should be a high priority for NHGRI.

- The next stages of the HGRP should provide concrete deliverables that focus on utility. A pangenome reference resource will never be "finished" if the goal is to represent all of human genetic diversity. Therefore the program should also consider goals in terms of use cases, (e.g., improving ability to diagnose diverse individuals) and building a common framework for genetic variation.

- The pangenome reference will have utility not only for NHGRI but for other NIH institutes, and the broader international community. World health organizations have a stake in this. This goes beyond considerations of co-funding: without buy-in from the larger community the resource won't be useful.

- There is a need to involve, enable and promote equity for diverse research communities, including international communities, throughout the project from design to implementation.

- ELSI should be included from the start in to navigate critical challenges, including:
    - Regulatory challenges in clinical use of a continuously changing reference.
    - Defining what is meant by "diversity" in this context, and how that should impact project priorities and activities.
    - The future of privacy: we can't predict the implications of a full genome online in 10 years.
    - Engagement of indigenous communities.
    - Limitations of consent as a mechanism for human participation.

Participants raised additional specific points for each meeting topic:

1. Samples and sequencing

    - Address the scale of production that is needed for a program renewal. This includes consideration of whether all constituent genome assemblies really need to be telomere-to-telomere, and whether technical improvements or changes in cost structures will allow for increased production.

- Define "diversity" in the context of the science and also the number of genomes we can afford.
- Be specific about interim goals/achievements and benchmarks.
- Leverage international relationships for addition of more samples/genomes.
- Make available an end-to-end non-proprietary pipeline to be used by all. This includes data production protocols, processing pipelines, and tools.

2. Pangenome reference representation

- Assess the advantages, timelines and requirements for moving away from using GRCh38 as a bridge.
- Clarify the right level of granularity for considerations of what features need to be represented (all variants, common variants).
- Clarify who is responsible for annotation of the updated version of the pangenome. It is important that existing annotations be carried over.
- Consider whether the best approach is to have a single authoritative version of the reference or whether multiple references, or multiple ways of presenting reference genome information, need to be tailored for specific applications.

3. Dissemination: versioning, user outreach, and user tools

Tools

- Focus on tool development and give priority to stable funding for the development and maintenance of tools.
- Encourage tools that will enable researchers to transition existing activities to a new reference (e.g., lift-over and community-specific interfaces), as well as tools that enable novel activities that can only be done using the new pangenome reference (e.g., new types of variant caller or association study analyses based on the pangenome).
- Promote open-source tools that are easy to use, and that enable affordable computing costs.
- Address the existing gap between developing the tools and building and making available a tool ecosystem.

Users

- Support user communities to transition from GRCh38 to the pangenome; use a "driver project" model.
- Define who the users are and work with user groups from the start to determine their needs. Choose and develop specific key use cases. Use cases that demonstrate utility (e.g., improve diagnostic rate, improve variant-to-function insight, etc.) will be important to encourage larger adoption of the pangenome reference.
- Provide education around the resource and its importance that addresses barriers such as understandability and consider community needs.
- Work to position NHGRI's leadership role in order to enable buy-in from the larger genomics community.

Versioning

- Consider tradeoffs between continuous delivery vs. freezing the reference.
- Be clear in naming and versioning. The format must make it clear to the to the user what they are actually using.

- Address the 'coordinate' problem, provide clarity on how to communicate coordinates in a pangenome world.

4. Engage diverse partners and the international genomics community

    - Approach this as an international effort; work with international organizations, e.g., WHO, GA4GH, H3Africa, DSI Africa, H3Bionet. Build trust by establishing relationships and collaborations.
    - Maintain commitment to open data science, but balance that with engagement to communities to explore and understand other pathways for data sharing.
    - Prioritize communication and the need to develop true engagement and trust with diverse communities.
    - Consider secure, federated methods that keep data locally.
    - Lower barriers of access in ways that allow a broad and diverse community to make use of the pangenome reference.
    - Recognize equity is more than inclusion; inclusion is important, but insufficient.
    - Proactively consider how a pangenome can lower disparities.
    - Provide the funding, support, specific interaction, etc. to engage partner communities.

Full Meeting Report

BACKGROUND

The human genome reference is a foundational resource used by essentially all researchers who need to align experimental or patient genome sequence data. The human genome reference also serves as a consensus coordinate system for reporting results. In 2019, the NHGRI launched the Human Genome Reference program (HGRP) in order to re-organize and re-focus its contribution to genome reference activities. A key goal of the HGRP was to move from a linear composite view of the human reference to a pangenome view that uses alternative representations (such as graph-based representations) of the genetic diversity of genome sequences from multiple individuals. The intent is to improve genomic analyses broadly, to ensure that the genome reference is useful for analyses of worldwide populations, and to reduce the risk that the use of the reference will lead to disparities.

The HGRP funds multiple elements, including: 1. addition of high-quality, phased, genome assemblies chosen to ensure that human haplotype variation is adequately represented; 2. development of computational methods to provide a useful representation for the community of the growing pangenome reference; 3. development of technologies to routinely sequence and assemble very high-quality genomes; 4. outreach to the user community to ensure adoption of the pangenome reference.

The grantees initially funded by the HGRP formed the Human Pangenome Reference Consortium (HPRC; humanpangenome.org). The consortium has made excellent progress on its goals (DOI 10.1038/s41586-022-04601-8.). ~150 high-quality genomes have been sequenced and assembled so far, representing >99% of SNVs of ≥1% frequency in the 1000 Genomes samples. The HPRC has demonstrated the utility of graph genome representations. In the last five years the HPRC has also highlighted the importance of having an embedded Ethical, Legal and Social Issues (ELSI) component.

Even with this progress, more work on the human reference remains to be done. Most notably, the tools for use of the pangenome reference remain underdeveloped. Although the community is becoming aware of the pangenome reference, the vast majority still use GRCh38, which remains an unphased linear representation of a mosaic of about 15 individuals. This points to a need for increased tool building and outreach.

In addition, in order to have a global impact, a human pangenome resource will need to increasingly engage the international scientific community. This engagement is needed to ensure that the participants who are sequenced and included in the reference are diverse, and also to ensure that international scientists are engaged and can participate in both the development and use of a pangenome resource. The current phase of the HGRP ends in 2023.

OPENING SESSIONS

Welcome: Dr. Eric Green welcomed participants to the meeting and put the HGRP within the larger context of the NHGRI's 2020 Strategic Vision. The HGRP is a key element in providing a **Robust Foundation for Genomics** and its goal to make the reference more representative of human genetic variation and diversity globally across populations aligns with the NHGRI's **Guiding Principles and Values**. Finally,

Dr. Green emphasized the importance of continuing the collaborative nature of the program, especially including international scientists in the effort.

- **Introduction:** Drs. Adam Felsenfeld and Alexander Arguello, Program Directors in the Division of Genome Sciences (DGSci) at the NHGRI, provided an overview of the meeting's purpose, goals, themes and deliverables. Dr. Felsenfeld presented an overview of the history and structure, and introduced the "Big Questions" that the meeting sought to answer, including:
    - Should the NHGRI have a coordinated program focused on the human genome reference in ~2024-2029?

  If yes:
    - What high-level goals should it have?
    - What major program elements should it include? How should they be related, integrated, and prioritized?
    - How should it balance clinical, population, and functional genomics along with other user community needs?
    - How can it creatively incorporate diversity, equity, and inclusion in all of its aspects (including participants, investigators, and users/the community)?

  To guide the conversation, a list of key considerations to keep in mind were introduced ahead of the panel discussions:
    - What are the opportunities and gaps that we experience and can be anticipated? What about the challenges and barriers?
    - What strategies are going to be the most efficient to address these challenges?
    - What does success look like and how do we measure that?

PANEL SESSION #1: SAMPLES AND SEQUENCING

Moderator: Charles Rotimi
Panelists: Carlos Bustamante, Ben Neale, Alice Popejoy, Genevieve Wojcik

This session focused on the sample and sequencing needs of a human pangenome program. The panelists were provided several questions to spark conversation:

- The current program has a goal of ~350 high-quality genomes from diverse populations by 2024; what should the goal of a renewed program be? Why?
- What does a "finished" human pangenome look like (i.e., when are we done)?
- Do we need more "pangenome quality" assemblies to achieve this?
- How do we prioritize sample selection from different perspectives (population genetics, clinical utility, diversity and inclusion, others)?
- What consent and population naming standards should be used and how should they be encouraged and integrated?

Discussants noted that a pangenome will never truly be "finished" since you could always capture more variation by adding additional individuals. Multiple participants agreed that the idea of a "finished" pangenome should be abolished, and the conversation shifted more towards how the NHGRI can mark progress and success within the program. Several participants indicated that the purpose of the pangenome must be identified to evaluate success. Participants suggested that the purpose of the pangenome is to support a common framework that will eventually be representative across the diversity of a species. Ultimately, the group suggested that defining specific interim goals, achievements, and benchmarks will be necessary for defining success for the pangenome efforts.

Participants also discussed the advantages of making the pangenome a dynamic resource that changes with time. This was followed by some discussion about whether annotations will be able to immediately follow a dynamic graph representation and if freezing the pangenome at certain points for release might be the more widely accepted way to move forward. Others visualized this freezing and releasing as somewhat similar to a phone software update, where several improvements are released periodically, and it does not disrupt the user workflow. Though there was support for a dynamic pangenome, there was concern that clinical labs might not jump aboard the idea and a suggestion that it would be critical to have some early demonstration projects that showed that clinical labs can feasibly implement a dynamic pangenome.

Discussion around the "pangenome quality" assemblies included the idea that the Telomere-to-Telomere (T2T) quality assemblies are the norm, and that more reference assemblies need to be generated to this level of quality to achieve metrics of success with the pangenome. But several other participants felt that perhaps lower quality genome assemblies would be adequate to fulfill the same purpose.

In terms of prioritizing sample selection, many ideas and opinions were put forward by the meeting participants. One suggestion was to use the graph-based structures as a tangible way to determine how diverse a sample is compared to the other genomes that are already included in the pangenome reference. This would allow for the prioritization of samples that are quantifiably different to enhance the overall genetic diversity that is included in the pangenome. There was also discussion about equity and how our social constraints have impacted the pangenome and the overall field of genomics in the past. One challenge that was identified is that the social/political categories of differences among and between humans have historically been conflated with genomic background. Arguably, by prioritizing the genomes that have the widest amount of variance according to the graph, genome variance is prioritized and less focus on social categories is supported. By leveraging international relationships, the addition of more diverse samples will be available for the pangenome in the future.

Participants discussed the complications around population naming. One idea put forward was that harmonization of population naming will be more successful than standardization of naming in the long-term. Standardization is not an effective strategy since it requires guidelines that are set in stone that may not be applicable globally. The idea that a future program should demonstrate stewardship instead of control in terms of consent was also brought up. Although it might be impossible to know now what the risks of being part of a reference will be in the future, the program is responsible for ensuring that leadership and stewardship are demonstrated to the communities included.

Throughout the discussion questions, themes of equity and inclusion were raised. Instead of focusing on what is missing in terms of DNA variation in the pangenome, some argued that the bigger question is which populations and even individual people are not represented in the resource. Being able to minimize the differences in impact to different groups of people was considered heavily, as well as who is and is not being mapped, and who will have access to the pipeline in the future. The suggestion to define what "diversity" means in the context of science and the number of genomes a renewed program can afford was also discussed during this session and should be part of ongoing conversations. Ultimately, the

group felt that to address equity and inclusion, it is imperative that an end-to-end non-proprietary pipeline including data production protocols, processing pipelines, and tools need to be openly available and widely accessible.

Several questions were also raised to complement this discussion, including:

- What would the timeline for freezing and releasing data on the pangenome look like?
- What are the active harms to the priorities of sample selection right now?
- How does the community know what is missing in the pangenome? What about who this effort is leaving behind?
- What scale of production is needed for a program renewal? Do all genomes really need to be T2T? Will technical improvements allow more genomes?
- Is it worth doing additional sequencing to fully capture copy number variants?
- Can minimum amounts of variant frequency or other measures be used to define the goalposts and metrics for success in the program?

PANEL SESSION #2: REPRESENTATION AND IMPLEMENTATION OF A PANGENOME RESOURCE

Moderator: Deanna Church
Panelists: Erik Garrison, Gabor Marth, Pavel Pevzner, Adam Phillippy

This session focused on the success and shortcomings of the current representations for the pangenome, and how these representations can be better supported in the future. The guiding questions included:

- Are the current graph and other data structure representations able to usefully represent the number of genome assemblies in the pangenome, now or in the near-to-mid future?
- How much R&D on graph or other representations is still needed? Or should we concentrate on implementation of available representations?
- In the next phase of the program, what should be the relative emphasis on R&D for pangenome representations vs. aggressive roll-out of the pangenome reference to the community?
- What is the best way to integrate this element into a larger problem – how closely does this activity need to be linked to other components of the HGRP as well as other NIH programs?

Several points were raised in support of the current data structure being able to usefully represent the number of genome assemblies needed for the pangenome, as well as some suggestions for future advances needed to improve the data structure. Now is the time to start looking beyond what works for people building the pangenome, but to determine what the user community will need. While graph representations offer a dynamic way to look at the genome, stable sequence models allow for further rare disease studies and observation of de novo mutations that might be difficult to detect in the dynamic graph representations. One participant indicated that a good metric for representations will be whether users of the pangenome can accurately call variation for every individual. A desire for increased focus on haplotypes was also discussed during this session. One participant stated plainly that bioinformatic software development for the pangenome model needs more funding.

During the discussion about increased research and development on pangenome representations, several participants shared the idea that communication between the HPRC

and other consortia studying the genomic basis of disease might be helpful for figuring out community needs. Several participants also shared the idea that these collaborations may be an iterative process where we learn what is needed along the way. There was also a call for more research and development of representations that can be used to better understand certain more complex, or understudied, regions of the genome. One participant shared that while the graph representation may be helpful in developing a deeper understanding of genome structure, the current use-cases are limited, and the community needs more tool development in order to move forward in this space.

This session also included discussion about the coordinate system used in the reference, and the need for innovation in this area The coordinate system as available now will have to be redesigned to function for the pangenome. This also led to conversation about how to lift over annotations if there is a change in the coordinate system, and the expectation that the community needs to ensure that existing tools continue to function, even when the underlying reference might be changing.

Several questions were raised by participants, including:

- How can the community and a renewed program work together to ensure that the pangenome will have rapid benefit?
- Are current mechanisms for tool and software development sufficient for the tools and software that are needed going forward?
- Does the community need just one reference or are multiple references (focused on specific applications) needed?
- How does the current representation scale to a future version of the pangenome that includes genomes from a much larger group of individuals?
- How will privacy concerns be addressed as more rare variants are being put into the pangenome?
- How will a renewed program ensure that genomics does not move forward without its complementary understanding of other -omics?
- How will the community move away from using GRCh38 as a bridge with the pangenome and how can a renewed program help navigate this?
- How will the community maintain the annotation and ensure that it is carried over to the pangenome?
- How will the community address the 'coordinate' problem? How will coordinates be communicated with the use of the pangenome?
- What is the right level of granularity to include (all variants, common variants)?

PANEL SESSION #3: DISSEMINATION: VERSIONING, USER OUTREACH/EDUCATION, AND USER TOOLS

Moderator: Eric Venner
Panelists: Aravinda Chakravarti, Valerie Schneider, Liz Worthey

This session focused on dissemination of the pangenome, how users will interact with the resource, and how a renewed program can facilitate the transition to make use of the pangenome as quickly as possible. The questions that were used to guide the discussion included:

- What will the user community need/want for the next phase of the program?
- What tools are needed? How best to encourage their generation and use?
- How do we minimize version churn or achieve backward compatibility?
- More generally, what other methods or resources are needed to gain acceptance?

- In two years, what emphasis will this component need?

Several participants had thoughts about how to split up the user community into groups for better understanding of what will be required for the next phase of the program. One suggestion was to look at the community as those interested in three areas: 1) human genetics, 2) model systems, and 3) the clinical community. Some participants agreed this relatively broad division would be helpful for figuring out tools and usage for the pangenome. Whereas others believed that you would need to consider even more specific user communities to accurately depict what will be necessary for groups to utilize the pangenome. Others mentioned that expectation management will be needed, since it is likely that not all communities will be supported at the same level. One of the more unique ideas discussed included a conversation on how to include the lay public community on the discussion of the pangenome through the private DNA testing kit companies and having explanations of the pangenome available to that group. Participants indicated that a ”driver project” model, similar to that used by GA4GH, may be a helpful model for the transition of different use communities from GRCh38 to the pangenome. These driver projects should be complemented by education around the resource and its importance while also addressing the barriers to the pangenome like understandability and community needs.

Further discussion focused on defining what tools would be needed and how to support them. Several participants indicated that support is needed for resources that work with both pangenome and linear versions of the reference.. Other participants thought that key deliverables such as alignments, remapping offunctional data and -omics, pangenome graph, and annotation will have to be considered going forward to make use of the current structures.

Participants agreed that it is a high priority to have stable funding for both the development and maintenance of the tools. They also noted some areas to prioritize for tool development. Specifically, participants felt that new types of variant caller or tools for novel association study analyses based on the pangenome would be more critical than lift-over and community-specific interfaces. These new tools should be open-source and easy to use, with compute being affordable. Several participants also noted that there is a gap between developing the tools and creating an ecosystem where tools are maintained and made available for others to use. It will also be important to users that naming and versioning are easily understood. Naming and versioning will likely be more challenging for the pangenome, and the format must make it clear to the user what they are using.

A discussion on version churn and backwards compatibility yielded a general opinion that backwards compatibility may not be the exact goal of a renewed program. Instead, education and making the tools usable going forward and ensuring that they are forward compatible with updates to the pangenome was preferred. Because the pangenome will be dynamic to some degree, most of the participants agreed that tools had to be able to support both the necessity for stability in some analysis contexts and the fluctuation that is expected with the pangenome. This sparked a conversation on diversity, equity, and inclusion, with the concern raised that certain institutions or global regions might not have the compute needed for more advanced tools. Ensuring that the reference is available equitably will be a responsibility of a renewed program going forward.

Participants spent some time addressing what broad adoption of the pangenome would look like. Several participants said that the balance of the duality of a stable, yet actively evolving resource will be necessary. The user community will need reassurance that the switch to the pangenome as a new reference is going to be meaningful in terms of the ability to make new scientific discoveries that have an impact on society. One participant indicated that funding diverse teams would be helpful for broad adoption, since that approach would better illustrate the different user communities and their needs. Having these diverse groups of individuals build the tools and publish their scientific findings may garner increased interest in the pangenome and more support for its uptake and use.

Additional questions that were sparked during discussion of this panel include:

- Are other NIH institutes interested in supporting or using the pangenome?
    - How can the NHGRI's leadership role be positioned to help ensure buy-in from the larger genomics community? How can the NHGRI demonstrate its utility (improving diagnostic rate, improving variant-to-function insight, etc.)?
- What communities are going to have access to this first? How can a renewed program help with compute in areas that should have access but might not have the resources available?
- How can the community and a renewed program work together to act as a steward for tools that have long since been unsupported?
- Who are the users? How will a renewed program work with them to determine their needs and choose specific key use cases for the pangenome?
- What are the tradeoffs between continuous delivery vs. freezing the reference?

PANEL SESSION #4: ENGAGING WORLDWIDE PARTNERS

Moderator: Erin Riggs
Panelists: Heather Lawson, Nicole Soranzo, Ambroise Wonkam

This session focused on how the HGRP should engage international partners and what could be done to ensure that relationships based on trust and respect are formed within the project and outside of it. The questions used to guide this conversation included:

- Who are key long-term partners both for the implementation of the pangenome reference as the standard across studies and data types as well as achieving an international resource representing humanity?
    - What is needed to attain this, and what should the NHGRI's specific role be?
- How can the NHGRI best establish this international reference with limited resources? ?
- Are there particular barriers that the NHGRI needs to anticipate (political, cultural, ethical)?
- Are there things the NHGRI should ***not*** do, even if it could?

Several attendees agreed that GA4GH, H3Africa, and other NIH institutes could be helpful international partners. There were specific suggestions of groups to work with in Africa, including professional organizations (e.g., African Society of Human Genetics (AfSHG)), consortia interested in the pangenome (e.g., H3Africa, DS-I Africa, and H3ABioNet), and government organizations (e.g., CDC Africa, WHO Africa) to provide structure for the African contributors. Collaborators will be needed in other regions as well, including Japan and Australia which have a well-developed genomics infrastructure. Meeting participants stressed the need to develop relationships and build trust as a global pangenome effort becomes more engaged with international partners.

The follow-up question on how to attain this and the NHGRI's role in an international effort sparked deep conversation. Most participants agreed that the NHGRI's role was to facilitate partnerships. By convening groups and empowering them, the NHGRI can enable further conversations to happen. Several people noted the need to include support for ELSI, policy, training, and education in order to strengthen the connections being made. Another suggestion was made to include community engagement from the start rather than focus on how the outputs will affect those communities. Participants pointed out that stable funding for pangenome reference activities is important for building meaningful international relationships, as it is difficult to build trust if long-term funding prospects are uncertain.

The group also anticipated some cultural barriers that this effort may encounter going forward. Meeting participants noted the need to emphasize building true international partnerships, and not having the development of this resource be based solely in the U.S. Other participants shared that prior malpractice of science would need to be considered, particularly when engaging with communities that have traditionally been underrepresented in genomics. Within this scope, it is important to lower barriers of access and consider how a pangenome can lower disparities. In order to do this, it is also important to recognize that equity is more than inclusion. Inclusion of diverse communities will be important, but it is insufficient for the reference that the community seeks to build.

Different views were expressed about data sharing in the context of international partnerships. The current consortium practice is to only include data from participants who have consented to open, unrestricted availability of the sequence data. Some participants felt strongly that the NHGRI should not insist upon open data sharing for all data in this effort, so as not to lose contributions from areas or groups that have other stances on data sharing. Others believed that the best way forward was for the NHGRI to commit to open data sharing, but to engage in conversation to ensure that potential partners are engaged in the discourse.

At the conclusion of the session, several questions still remained to be answered including:

- How can federated models be moved forward?
- Right now, the HPRC is focused in the U.S. and quasi-governmental. How would a renewed program shift the power structure so it will be more supportive of the international and collaborative framework?
- How will the community plan to respect local traditions and norms?
- How will the community deal with data sovereignty?
- How will a renewed program achieve a group that everyone feels that they are an equal member in?
- What communities will a renewed program be engaging with? Who is the resource for?
- Can the pangenome be multi-level (private/public, one country/multiple countries) to help with data sharing concerns? Maybe the pangenome will not be a single reference but rather multiple.
- How will a renewed program reach out to the smaller communities?
- How will a renwed program develop frameworks with local and frequently revisited oversight?
- How can the NHGRI's leadership role be positioned to ensure buy-in from the international genomics community?

CLOSING DISCUSSION

Discussants: Xander Arguello, Adam Felsenfeld, Deanna Church, Martin Hirst, Matt Lebo, Pavel Pevzner, Bob Waterston

This session started with a summary of key topics and suggestions that came up throughout the meeting:

Overarching Suggestions

- Abolish the concept of "finished" from our thinking.
- Develop definitions of aggressive deliverables/goals for all phases of the project.
- Plan for the sustainability of resources and tools (and international relationships) after the HGRP funding ends.
- Support user communities to transition from GRCh38 to the pangenome.
- Support driver projects to demonstrate utility.
- Involve, enable, and promote equity for diverse research communities throughout the project from design to implementation.
- Have a robust focus on tool development, with a focus on open-source tools that are easy to use, and with compute being affordable.

Samples and Sequencing

- Understand the tradeoff between fewer samples at higher quality vs. more samples at lower quality.
- Carefully assess equity issues.
- Be specific about interim goals/achievements and benchmarks.
- Consider equity and use case metrics, rather than only considering technical metrics.

Representation and Implementation of a Pangenome Resource

- Develop an approach to address the 'coordinate' problem, including a framework to communicate coordinates in a pangenome world.
- Move away from using GRCh38 as a bridge after 2024.
- Determine the right level of granularity for assessing how well the pangenome captures human diversity (all variants, common variants).
- Consider whether the genomics community needs just one reference, versus developing multiple references focused on specific applications.

Dissemination: Versioning, User Outreach/Education, and User Tools

- Consider the trade-offs of continuous delivery vs. data freezes for delivering the reference.
- Determine the appropriate naming and versioning in such a way that it is readily known to the user.
- Define who the users are and work with user groups from the start to determine their needs, including considerations of driver projects and specific use cases.
- Develop ways to educate user communities about the reference and its importance in addressing barriers such as understandability and consideration of community needs.

Engaging Worldwide Partners

- Provide the funding, support, specific interaction, etc. needed to appropriately and beneficially engage diverse partner communities.
- Maintain commitment to open data science, but balance with other pathways to engagement and the need to develop trust.

- Consider secure, federated methods that keep data locally.
- Lower barriers of access.
- Recognize equity is more than inclusion; inclusion will be important but insufficient.
- Proactively consider how a pangenome can lower disparities.
- Build global infrastructure.

Participants raised some additional topics for consideration. First, it was noted that it will be important that the tools and structures needed to use the pangenome are widely available to other institutes and not just to the NHGRI. Similarly, it will be important to gain credibility with international bodies such as the World Health Organization in order to have international use of the pangenome.

Participants revisited the topic of annotation and the challenge that genome annotations are still evolving. Similarly, our understanding of the non-coding regions of the genome are still changing, and support will have to continue for the pangenome to have updated annotations of these non-coding regions. This led to a discussion of how closely genome annotation activity should be associated with the production of pangenomes and the need for a framework that allows for evolving annotations to be made to an evolving pangenome reference representation.

Participants also had additional discussion on the importance of integrating the concept of equity into the HGRP. Several challenges related to equity and to ELSI issues can be forecasted, including:

- Regulatory challenges of a continuously changing reference that is constantly updated.
- The need to define what is meant by diversity in the context of a pangenome.
- Privacy challenges, and the way that privacy challenges are changing over time.
- Ongoing issues with indigenous communities and community engagements.
- Limitations of consent as a mechanism for human participation.

Similar to other sessions, this discussion ended with questions that will be important for the NHGRI to consider going forward, including:

- What infrastructure is needed to support a pangenome reference? Is AnVIL the suitable place for this effort?
- How should organizations seeking to get involved with the HGRP communicate their interest in being involved with the pangenome?
- What are small, incremental goals that a renewed program and the community should be aiming for over the next few years, so that they can ensure things are happening differently in 2029? What is the overall roadmap for this project?
- How will the community and a renewed program bring genes that were not previously represented in other references into the pangenome ecosystem?

And finally, several overarching conclusions were drawn from the final session:

- Continuing the HGRP should be a high priority for the NHGRI.
- Concrete deliverables and milestones must be defined. The HGRP will never be "finished" if the goal is to represent all of human genetic diversity. Rather the project should have a specific focus on utility and consider metrics such as improving the ability to diagnose individuals or building a common framework for genetic variation.
- Engagement with international partners is necessary. This engagement should go beyond considerations of co-funding. Without buy-in from the international community, the resource will not be useful.

- Involving, enabling, and promoting equity for diverse research communities throughout the project from design to implementation is a priority. ELSI should be embedded from the start to navigate critical challenges.

Appendix A
HGRP Planning Meeting Agenda

Start Time	Duration	Session Description	Moderator/Panelists
8:30AM	10′	Welcome	Eric Green
8:40AM	10′	Introduction: purpose, goals, themes, deliverables, logistics	Alexander Arguello, Adam Felsenfeld
8:50AM	60′	Samples and sequencing	**Moderator:** Charles Rotimi **Panelists:** Carlos Bustamante (v), Ben Neale, Alice Popejoy, Genevieve Wojcik (v)
9:55AM	20′	*Break*	
10:15AM	60′	Representation and implementation of a pangenome resource	**Moderator:** Deanna Church **Panelists:** Erik Garrison, Gabor Marth (v), Pavel Pevzner (v), Adam Phillippy
11:15AM	60′	Dissemination: versioning, user outreach/education, and user tools	**Moderator:** Eric Venner **Panelists:** Aravinda Chakravarti (v), Valerie Schneider, Liz Worthey
12:15PM	60′	*Lunch on your own*	
1:15PM	60′	Engaging worldwide partners	**Moderator:** Erin Riggs **Panelists:** Heather Lawson, Nicole Soranzo (v), Ambroise Wonkam
2:15PM	15′	*Break*	
2:30PM	45′	Discussion	**Discussants:** Alexander Arguello, Adam Felsenfeld, Deanna Church, Martin Hirst (v), Matt Lebo, Pavel Pevzner (v), Bob Waterston (v)
3:15PM		*Close*	

Appendix B
Participant List

Rajeev K. Agarwal
Office of Research on Women's Health, NIH
agarwalraj@mail.nih.gov

Lucinda Antonacci-Fulton
The McDonnell Genome Institute at Washington University
lfulton@wustl.edu

Alexander Arguello
National Human Genome Research Institute, NIH
alexander.arguello@nih.gov

Katie Bardsley
National Human Genome Research Institute, NIH
katie.bardsley@nih.gov

Ewan Birney
EMBL-European Bioinformatics Institute
birney@ebi.ac.uk

Alex Brown
Telethon Kids Institute
Australian National University
alex.brown@anu.edu.au

Carol J. Bult
The Jackson Laboratory
carol.bult@jax.org

Carlos D. Bustamante
Stanford University
cdbustam@stanford.edu

Mark Chaisson
University of Southern California
mchaisso@usc.edu

Aravinda Chakravarti
New York University
aravinda.chakravarti@nyulangone.org

Deanna Church
Independent Contractor
deanna.church@gmail.com

Sara Grace Currin
National Human Genome Research Institute, NIH
sara.currin@nih.gov

Valentina di Francesco
National Human Genome Research Institute, NIH
valentina.difrancesco@nih.gov)

Evan Eugene Eichler
University of Washington
ee3@uw.edu

Adam Felsenfeld
National Human Genome Research Institute, NIH
adam_felsenfeld@nih.gov

Adam Frankish
EMBL-European Bioinformatics Institute
frankish@ebi.ac.uk

Robert S. Fulton
The McDonnell Genome Institute at Washington University
rfulton22@wustl.edu

Weiniu Gan
National Heart, Lung, and Blood Institute, NIH
ganw2@nhlbi.nih.gov

Nanibaa' Garrison
University of California, Los Angeles
nanibaa@socgen.ucla.edu

Erik Garrison
University of Tennessee Health Science Center
egarris5@uthsc.edu

Mark Gerstein
Yale University
mark.gerstein@yale.edu

Dan Gilchrist
National Human Genome Research Institute, NIH
daniel.gilchrist@nih.gov

Thomas Raymond Gingeras
Cold Spring Harbor Laboratory
gingeras@cshl.edu

Eric D. Green
National Human Genome Research Institute, NIH
egreen@nhgri.nih.gov

Roderic Guigo
Centre for Genomic Regulation
roderic.guigo@crg.cat

Jonathan Haines
Case Western Reserve University
jonathan.haines@case.edu

Ira Hall
Yale University
ira.hall@yale.edu

Martin Hirst
University of British Columbia
hirstm@mail.ubc.ca

Kerstin Howe
Wellcome Sanger Institute
kj2@sanegr.ac.uk

Carolyn Mary Hutter
National Human Genome Research Institute, NIH
carolyn.hutter@nih.gov

Erich Jarvis
The Rockefeller University and Howard Hughes Medical Institute
ejarvis@rockefeller.edu

Hanlee Ji
Stanford University
genomics_ji@stanford.edu

Eimear E. Kenny
Ichan School of Medicine at Mount Sinai
eimear.kenny@mssm.edu

Barbara Ann Koenig
University of California, San Francisco
barbara.koenig@ucsf.edu

Jan Korbel
European Molecular Biology Laboratory
elena.andreeva@embl.de

Pui-Yan Kwok
University of California, San Francisco
pui.kwok@ucsf.edu

Heather A. Lawson
Washington University in Saint Louis
lawson@wustl.edu

Matt Lebo
Mass General Brigham
mlebo@bwh.harvard.edu

Sandra Soo-Jin Lee
Columbia University
sandra.lee@columbia.edu

Charles Lee
The Jackson Laboratory for Genomic Medicine
charles.lee@jax.org

Heng Li
Dana-Farber Cancer Institute
hli@ds.dfci.harvard.edu

Xihong Lin
Harvard T.H. Chan School of Public Health
xlin@hsph.harvard.edu

Tina Lindsay
The McDonnell Genome Institute at Washington University
tgraves@wustl.edu

Nicole Lockhart
National Human Genome Research Institute, NIH
lockhani@mail.nih.gov

Amy C. Lossie
National Institute on Drug Abuse, NIH
amy.lossie@nih.gov

Aime Lumaka Zola
University of Kinshasa
aime.lumaka@unikin.ac.cd

Tobias Marschall
Heinrich Heine University
tobias.marschall@hhu.de

Gabor T. Marth
University of Utah
gmarth@genetics.utah.edu

Daphne Oluwaseun Martschenko
Stanford Center for Biomedical Ethics
daphnemartschenko@gmail.com

Gilean McVean
Genomics plc
gilean.mcvean@genomicsplc.com

Karen Hayden Miga
University of California, Santa Cruz
khmiga@ucsc.edu

Marilyn Miller
National Institute on Aging, NIH
millerm@nia.nih.gov

Stephanie Morris
National Human Genome Research Institute, NIH
morriss2@mail.nih.gov)

Jennifer Moser
Department of Veterans Affairs
jennifer.moser@va.gov

Anjene Musick
National Institutes of Health
anjene.musick@nih.gov

Benjamin Michael Neale
Massachusetts General Hospital and Broad Institute
bneale@broadinstitute.org

Anne O'Donnell Luria
Broad Institute and Boston Children's Hospital
odonnell@broadinstitute.org

Hardip Patel
The Australian National University
hardip.patel@anu.edu.au

Benedict Paten
UC Santa Cruz Genomics Institute
bpaten@ucsc.edu

Pavel Pevzner
University of California, San Diego
ppevzner@ucsd.edu

Adam Phillippy
National Human Genome Research Institute, NIH
adam.phillippy@nih.gov

Sharon E. Plon
Baylor College of Medicine
splon@bcm.edu

Alice B. Popejoy
University of California, Davis
abpopejoy@ucdavis.edu

Jennifer Ellen Posey
Baylor College of Medicine
jennifer.posey@bcm.edu

Pankaj Qasba
National Heart, Lung, and Blood Institute, NIH
pq5h@nih.gov

Nishadi Rajapakse
National Library of Medicine, NIH
chandima.rajapakse@nih.gov

Heidi L. Rehm
Mass General Brigham and Broad Institute
hrehm@mgh.harvard.edu

Bing Ren
University of California, San Diego School of Medicine
biren@ucsd.edu

Erin Rooney Riggs
Geisinger
eriggs@geisinger.edu

Anna Jollyette Rogers
National Human Genome Research Institute, NIH
anna.rogers@nih.gov

Kristen Ross
National Human Genome Research Institute, NIH
kristen.ross@nih.gov

Charles Nohuoma Rotimi
National Institutes of Health
rotimic@nih.gov

Melissa Rotunno
National Cancer Institute, NIH
rotunnom@mail.nih.gov

Valerie Schneider
National Library of Medicine, NIH
schneiva@ncbi.nlm.nih.gov

Adam C. Siepel
Cold Spring Harbor Laboratory
asiepel@cshl.edu

Michael William Smith
National Human Genome Research Institute, NIH
smithmw@nih.gov

Heidi Sofia
National Human Genome Research Institute, NIH
heidi.sofia@nih.gov

Nicole Soranzo
Human Technopole
nicole.soranzo@fht.org

Nathan Stitziel
Washington University
nstitziel@wustl.edu

Jennifer Troyer
National Human Genome Research Institute, NIH
jennifer.troyer@nih.gov

Maya Vanzanten
National Human Genome Research Institute, NIH
maya.vanzanten@nih.gov

Eric Venner
Baylor College of Medicine
venner@bcm.edu

Ting Wang
Washington University
twang@wustl.edu

Robert H. Waterston
University of Washington
watersto@uw.edu

Kris Wetterstrand
National Human Genome Research Institute, NIH
wettersk@mail.nih.gov

Matthew Thomas Wheeler
Stanford University
wheelerm@stanford.edu

Genevieve L. Wojcik
Johns Hopkins Bloomberg School of Public Health
gwojcik1@jhu.edu

Ambroise Wonkam
Johns Hopkins University
ambroise.wonkam@uct.ac.za

Jonathan Wood
Wellcome Sanger Institute
jmdw@sanger.ac.uk

Elizabeth Anabel Worthey
University of Alabama at Birmingham
eaworthey@uabmc.edu

Andrew David Yates
EMBL-European Bioinformatics Institute
ayates@ebi.ac.uk

Jane Ye
National Institutes of Health
jane.ye@nih.gov

Joseph M. Yracheta
Native Bio-Data Consortium
joseph@nativebio.org

Michael C. Zody
New York Genome Center
mczody@nygenome.org

Justin M. Zook
National Institute of Standards and Technology
jzook@nist.gov