Shaping the Future of Genetic Diversity in Humans: A Reflection on Past Ethical Challenges and Lessons Learned

Keerthana Balaji[1], Robert Cook-Deegan[2]
[1]Arizona State University, Tempe, AZ, USA
[2]Arizona State University, School for the Future of Innovation in Society, Tempe, AZ, USA
https://doi.org/10.38126/JSPG270101
Corresponding author: kbalaji4@asu.edu
Keywords: genome sequencing; population genetics; bioethics; diversity; global health

Executive Summary: Genetic diversity research has long faced ethical challenges, especially regarding community engagement, informed consent, and data ownership. Early projects like the Human Genome Diversity Project (HGDP) had challenges with proper community involvement, leading to significant ethical controversies. Later initiatives, such as the International HapMap Project, made strides by incorporating stronger ethical frameworks and better community engagement practices. Despite these improvements, ongoing issues remain, including the politicization of genetic findings and the oversimplification of population labels that can perpetuate harmful stereotypes. To move forward responsibly, a more inclusive and transparent approach to genetic research is proposed, ensuring that diverse communities are equitably represented and respected. By fostering dialogue between scientists and communities, integrating ethical considerations from the start, and avoiding reductionist population descriptors, we can prevent past mistakes from being repeated. This shift in research practices will ensure that future genetic diversity projects benefit global health while building trust with the communities involved.

I. Introduction

Genetic diversity projects are coordinated efforts to collect and study DNA from many different populations to build comprehensive databases of human genetic variation for research and clinical interpretation. Genetic diversity projects reflect humanity's quest to unravel our genetic makeup. A series of initiatives dating back to the Human Genome Project have been designed to catalog and analyze genetic variations among human populations, with the overarching goal of mapping the genetic diversity of our species. By probing deeply into the genetic landscapes of the world's diverse populations, these projects aim not only to enhance our understanding of human evolution and migration but also to shed light on the genetic basis of various diseases and traits. Uncovering human genetic diversity makes way for advances in science and medicine, offering insights into disease susceptibility, drug responses, and public health (Lu et al. 2014). The data from these projects hold the potential to improve healthcare delivery, inform personalized medicine approaches, and promote a more inclusive and equitable healthcare. However, it is crucial to acknowledge that these projects have exposed flaws in research practices. Each initiative has built upon its predecessors with similar protocols, encountering new challenges, ethical issues, and controversies, while hardy perennial issues persist. Concerns about informed consent, privacy, global inequities, commercialization, and how to classify and label populations have only intensified. Debates surrounding the ownership and use of genetic information have never been fully resolved. These serve as important challenges that demand reflection and scrutiny to ensure that

history does not repeat itself. Efforts to sample human genetic diversity can learn from past mistakes and build on successful models.

Several projects have advanced knowledge of human genetic variation, starting with the Human Genome Diversity Project (HGDP 1992-2003),the International HapMap Project (2002-2016), and the 1000 Genomes Project (2008-2015). These projects span three decades, from their planning and execution to their impact. The HGDP, initiated in the 1990s, aimed to characterize genetic diversity across different populations worldwide. The HapMap project in the early 2000s focused on identifying common genetic variants and their associations with diseases. The 1000 Genomes Project, launched in 2008, marked a significant milestone by sequencing the genomes of thousands of individuals from diverse populations, providing a comprehensive catalog of human genetic variation. Below are brief summaries of the projects.

i. The Human Genome Diversity Project

The Human Genome Project (HGP), not to be confused with HGDP, was launched in 1990 as one of the first efforts to sequence the human genome and several other non-human organisms. It was known to be "one of the most ambitious and important scientific endeavors in human history" (National Human Genome Research Institute n.d). Before high-throughput genome sequencing, researchers assessed genetic diversity using pedigrees and karyotyping, electrophoretic protein polymorphisms, Southern blot–based restriction fragment length polymorphism, microsatellites, targeted Sanger sequencing of candidate loci, and linkage or association mapping with marker panels. These marker-based approaches paved the way for the global cataloging of human variation through whole-genome sequencing (Sunnucks 2000). While HGP was well underway, the Human Genome Diversity Project (HGDP) was then proposed in 1991, in the journal of *Genomics* (Cavalli-Sforza 1991), to explore genetic variation of humans. The Project aimed to analyze individual genetic variation from around 500 "populations" around the world (Cavalli-Sforza 1994). This goal set HGDP apart from preceding projects, which did not explore the diversity of the genomic data that were being collected. By obtaining blood samples from individuals from all inhabited continents, they planned to sequence DNA and create cell lines in order to study human migration and evolution. While the Project was concluded in 2003, the HGDP-CEPH DNA samples and data subsets, in total around 1,064 cell lines, are still being used as of 2024 (Cavalli-Sforza 2005; Koenig et al. 2024). Bergström, et al., sequenced 929 of the samples reported in a 2020 publication (Bergström et al. 2020).

ii. The HapMap Project

The International HapMap Project, whose goal was to develop a haplotype map of the genome, had its initial meeting in 2002. A haplotype is a selection of genomic variations (SNPs) that are often inherited together (National Human Genome Research Institute n.d). By creating a "map" of these common variations amongst humans, disease susceptibility and manifestation could be further studied. This project found "sequence variants, their frequencies, and correlations between them" from populations from Africa, Asia, and Europe (The International HapMap Project n.d). The HapMap had three main phases, with Phase I assembling the first haplotype map, Phase II mapping over 2.1 million SNPs, and Phase III increasing the number of DNA samples to 1,301 from different populations (The International HapMap Project n.d; Wellcome Sanger Institute n.d). For the HapMap, the community engagement process for new sample collection pioneered methods to address community and cultural concerns, and incorporated into the project's informed consent. The project was concluded in 2016 and the sequence data are available in the public domain, allowing future projects and interested scientists to make use of this seminal work.

iii. The 1000 Genomes Project

Launched in 2008, the 1000 Genomes project was designed to catalog human genetic variation that would support studies investigating genetic correlations with diseases (1000 Genomes n.d). The data extended beyond the International HapMap Project and became a comprehensive resource that had genomic information on 2,504 individuals from 26 populations around the world. Unlike the

HapMap, 1000 Genomes aimed to detect variants "with frequencies as low as 1%" (IGSR: The International Genome Sample Resource n.d). Ultimately, the final data set was published in 2015 (The 1000 Genomes Project Consortium 2015).

While these projects and studies were groundbreaking, there were ethical flaws that should be explored to not repeat them in the future. This paper reviews past efforts to sample genetic diversity by identifying the challenges encountered and reflecting on the lessons learned to avoid missteps for the Human Pangenome Reference Consortium, Human Pangenome Project, and other current and future initiatives.

Six interviews were conducted with individuals who were part of these projects' formation and execution to collect perspectives from critics and leaders of the projects. Among the interviewees, 2 were project leaders, 2 were critics, and 2 were historians. These interviewees were selected based on their role in the project development to understand rationale behind decisions made, and the critics (bioethicists in this case) and historians were selected based on their understanding and previous work in outlining ethical guidelines for genomic research. While only six interviews were completed, this paper serves as a starting point for better understanding the ethical issues encountered in early genome projects. In addition, a literature review of project protocols, proposals, policy documents, and peer-reviewed ethical analyses was completed. Using the feedback and reflections made on the projects from the interviewees and analysis of literature on these projects, major themes regarding challenges and lessons emerged. The historians (Interviews E and F) contributed to clarifying project timelines, while the perspectives of the other interviewees are incorporated as appropriate throughout the discussion.

II. The ethical basis and politicization of genetic research

Genetic research raises profound ethical, legal, and social implications. The cautionary tales about the history of eugenics gave context to biological determinism and scientific racism, and how it might lead to contemporary issues in research (Largent 2014). As the initial genome project ideas were being conceived, noteworthy scholarship encompassing genetics, eugenics, and race was emerging. Daniel J. Kelves' analysis of the history of eugenics, serialized in The New Yorker, as "Annals of Eugenics" (Kevles 1984) and then published as a book, *In the Name of Eugenics* (Kelves 1985), gained prominence even as the HGP was getting launched. His work brought recognition to eugenics and how it, in Kelves' words, "casts a shadow over all contemporary discourse concerning human genetic manipulation" (Kelves 1985; Largent 2014). It catalyzed a renewed awareness of how genetic research might have unforeseen or even negative social consequences. Benno Müller-Hill's *Murderous Science*, published in 1988, uncovered the role of scientists and physicians in the Nazi holocaust. Kevles's, Müller-Hill's and a spate of other scholarship, uncovered the dark history of eugenics and drew attention to the need for Ethical, Legal, Social Implications (ELSI) research in the scientific community (Proctor 1988). As the HGP was taking flight and the HGDP was being proposed, the National Institutes of Health (NIH) and the Department of Energy (DOE) began implementing ELSI in research (McEwen 2014; Dolan 2022).

The National Academies of Sciences, Engineering, and Medicine (NASEM) proposed a framework that comprehensively addressed profound ethical, legal, and social implications (NASEM 2019). Ethical considerations such as informed consent, privacy protection, and the potential for discrimination are paramount. Legal issues encompass intellectual property rights, regulatory frameworks for genetic testing, and the forensic use of genetic data in legal proceedings and for population surveillance. Genetic research additionally intersects with questions of identity, familial relationships, and societal attitudes toward diversity and inclusion.

Moreover, the NASEM report framework outlines the importance of promoting societal value, minimizing negative impacts, protecting research participants' interests, advancing patient welfare, ensuring scientific rigor, engaging relevant communities, establishing oversight and accountability mechanisms, and acknowledging governmental and

policy roles to navigate the ethical domain of genetic research responsibly (NASEM 2019). Some of these eight considerations are used to understand where certain projects could have had setbacks and challenges as they progressed.

Genetic research has a penchant for becoming politicized because findings about genetic differences are often misinterpreted or weaponized to justify social hierarchies and discriminatory policies. This emphasizes the importance of vigilance in preventing misunderstandings and misuse of scientific findings. Transparency and clear communication are crucial when navigating the complexities of genetic research, particularly in discussions surrounding sensitive topics such as race. While acknowledging the role of genetic factors in shaping human traits, it is essential to strike a balance in discussing whether race even comes in as a genetic determinant without reinforcing harmful stereotypes.

The NASEM report on *Using Population Descriptors in Genetic and Genomics* provides a historical overview of how race and other population labels have been used in scientific research and further recommendations on how to improve genomic studies (NASEM 2023). The authors state that "race is a socially constructed designation, a misleading and harmful surrogate for population genetic differences, and has a long history of being incorrectly identified as the major genetic reason for phenotypic differences between groups" (NASEM 2023). Yudell et al. "Taking Race out of Human Genetics," make a clear distinction between "race" and "ancestry", where ancestry is about an individual's personal relationships with their genealogical heritage, and race is just a conclusion made about hierarchical organization of people from similar geographies or cultures (Yudell et al. 2016). By doing this, scientists can promote clarity, define variables clearly, and mitigate confusion and inconsistency in understanding human diversity and its implications. Ultimately, genetic research should improve lives while being mindful of ethical considerations and societal impacts.

III. Challenges and lessons

i. Consequences of "science first, ethics second"

Privileging scientific advancement over ethical considerations can have significant consequences. When science is pursued without sufficient attention to ethical implications, it can exacerbate inequalities and injustices. The approach of "science first" may reinforce societal biases and discrimination, particularly in research areas such as genetics and race. Scientific projects featuring grand ideas, sizable grants, and large teams risk inadvertently crashing against the shoals of controversy if the ethical, legal, and social factors are not anticipated and addressed. Placing science first and ethics second risks neglecting the broader societal impacts and ethical complexities. In line with this, the Collingridge Dilemma posits that it is difficult to predict and control the consequences of a technology—such as genomic sequencing in this case—early in its development, but by the time unwanted effects are apparent, it may be too late, expensive, or difficult to change the technology effectively (Collingridge 1980). Integrating ethical reflection and considerations into scientific practices from the outset can ensure that research outcomes align with ethical principles and contribute positively to social justice and human well-being.

For example, the HGDP aimed to collect genetic samples from diverse populations worldwide to study human genetic diversity. However, the project faced significant criticism and controversy, particularly from Indigenous communities. HGDP was labeled "helicopter science", which is the practice when "scientists from wealthy nations visiting lower-income countries, collecting samples, and publishing results with little or no involvement from local scientists, and providing no benefit for the local community" (Adame 2021; Minasny et al. 2020). Indigenous peoples were identified as groups to be sequenced for genetic variation through HGDP. Due to the "helicopter science," "Indigenous peoples felt they were being treated as examples of human fossils, from whom samples had to be collected before they died out" (Dodson and Williamson 1999). The World Congress of Indigenous Peoples dubbed the project the "Vampire Project" in 1993, along with many other organizations actively

opposing the research. They called for a halt as the project planned to draw blood from Indigenous communities amid fears of exploitation and patenting of samples (Central Australian Aboriginal Congress 1994). Critics argued that the project failed to adequately address the ethical implications of genetic research on Indigenous populations, including issues of informed consent, ownership of genetic data, and potential exploitation (Reardon 2005). There were concerns about profits to be made off of the samples obtained, lack of transparency, absence of reciprocity for the populations being sampled, and more. Project organizers did respond to these claims with updated ethical guidelines, reassurance, and answers to the questions raised, but controversy continued (Cavalli-Sforza 1994; Reardon 2005; Greely 1999).

Although the project ultimately went uncompleted, a takeaway from this is that overlooking ethical principles can lead to significant backlash and mistrust, ultimately undermining research goals. For instance, once statements were released answering questions and clarifying the ethics behind HDGP, even more concerns arose about group consent and patenting Indigenous DNA. As a result, "this episode brought into relief Project organizers' difficulty in recognizing the potential political and economic content of their Project, a difficulty that would continue to trouble their attempts to create a social and moral order that could support their endeavor", as Reardon writes (Reardon 2005). Jenny Reardon is a professor of Sociology and the Founding Director of the Science and Justice Research Center at the University of California, Santa Cruz, and author of *Race to the Finish: Identity and Governance in an Age of Genomics*. One lesson to learn here is that by involving ethics in the fabric of scientific inquiry from the beginning, researchers can foster greater trust and collaboration. As stated by one of the HGDP organizers, the project did develop a Model Ethical Protocol and intended to secure local Institutional Review Board (IRB) authorization and national/regional approvals for any new sampling [Interview A] (Weiss et al. 1997). The Model Ethical Protocol also called for benefit-sharing with donors and their communities. But critics, especially Indigenous advocates, argued that formal approvals alone were insufficient without prior community-level consent, clear rules on ownership and benefit-sharing, and protections against patenting [Interview B]. The HapMap, for example, brought consideration of the ethical questions into the first meetings, to address and create a plan about how the genetic variations they identified would not perpetuate harmful stereotypes. This is explored further below.

ii. Overlooking anthropological and historical contexts

Scientists should be equipped with knowledge of history, ethics, and anthropology while formulating their projects. Scientific literacy encompasses more than just an understanding of scientific principles and methodologies; it also involves an appreciation for the historical context and ethical considerations that shape scientific inquiry. When developing the research question or premise of the project, this interdisciplinary approach can help bridge the knowledge gap between the present phenomenon and its historical and cultural influences. For instance, HGDP initially proposed sampling to detect genetic variations without fully considering the complex social structures and historical events that shaped these populations, or how they might react to such a project. By incorporating anthropological insights during inception, researchers could have framed questions that considered past migrations, cultural practices, and environmental pressures, leading to a more nuanced understanding of the collected data and potentially faster discoveries. HGDP involved anthropological expertise later, but this created new concerns about reviving outdated ideas about race.

Reardon argues that gaining anthropological expertise is not as easy as including anthropologists for their advice and expertise, as anthropology itself is divided on questions about methods and goals, which are closely tied to issues of power and knowledge (Reardon 2005). Especially when a project is dealing with race and human differences, it is important to acknowledge how the power structure even came about, who is allowed to define it, and the implications of studying it. This relates to the idea of "science first, ethics second," how the absence of anthropological expertise on diversity and evolution of humans in the early stages ultimately raised concerns about the scientific

legitimacy of the project (Reardon 2005). Further, Reardon observes that a hierarchy was in practice as a result, where anthropological input was viewed as being useful during sampling but not throughout the design process. When conducting research involving diverse people, particularly marginalized communities, scientists need to approach sampling with more than just including an anthropologist during project formation. Instead, prioritizing cultural competence and understanding historical context is essential. This entails extensive research of the culture and history of the people they would engage with before the project design, rather than solely relying on external resources, ensuring a respectful and informed approach to science.

In another aspect, science education at the high school, undergraduate, and graduate levels can also explicitly integrate historical case studies, ethics, and anthropological perspectives into core curricula. Practically, this can mean ELSI modules embedded in introductory biology and genetics, co-taught courses with historians and anthropologists, and reflective assignments that surface power dynamics and positionality. Accreditation standards, grant requirements, and departmental learning outcomes can further normalize these expectations by assessing culturing competence and ethical reasoning in addition to methodological competence. By aligning pedagogy and policy, training programs can produce scientists who are as fluent in the human context of their work as they are in its technical methods, strengthening trust with the communities their research affects and reducing the risk of repeating past harms.

iii. Outdated community engagement

Relying solely on pre-existing foundations and relationships with communities from the past without prioritizing ongoing community engagement can lead to several drawbacks. It perpetuates assumptions of continuity, overlooking the evolving needs and concerns of communities over time. Additionally, not having active engagement risks eroding trust and undermining the integrity of the research process, as there is a potential for communities to feel disregarded. This approach also misses opportunities for collaboration and co-creation of research projects aligned with community interests. Ethical concerns arise regarding informed consent, respect for autonomy, and equitable distribution of benefits and burdens when meaningful community involvement is lacking. While initial gains may be achieved, sustained effort and investment in ongoing engagement are essential for building sustainable partnerships that contribute to long-term research success and positive community outcomes.

Striving to not make the same mistakes of prior projects, the HapMap engaged in community consultation to ensure that they were not doing "helicopter science." In addition to community engagement with the participants involved, the HapMap also adopted a more nuanced approach to population descriptors. According to a scientist and ethicist who was consistently involved in the oversight of HapMap, community engagement reached a new high with the HapMap, which included informed consent forms adapted to the communities at hand. Focus groups with participants informed the selection of population labels for HapMap. After a conversation with the leaders of the participating communities and the members, the informed consent was updated for each community, which "provided a stronger partnership with the communities so their data was accounted for on their terms" [Interview C].

Respecting local communities and enhancing community engagement can contribute to scientific advancement and ethical integrity. By involving local communities in the research process, projects gain valuable insights into cultural practices, beliefs, and values that may be more reflective of a population. Additionally, community engagement ensures trust and builds rapport between researchers and community members, laying the foundation for collaborative partnerships that benefit both parties. Through meaningful engagement and conversations with the communities being sampled, researchers can address community concerns, ensure informed consent, and tailor research approaches to meet the needs and preferences of local populations. Ultimately, respect for local communities is prioritized with an increase in community involvement.

With that said, the projects following the HapMap seemed to have relied on the already-built foundations with communities due to financial constraints and time limitations. In my interview with a bioethicist who had followed the 1000 Genomes project from start to finish, they stated that this may have been seen with the 1000 Genomes, "where sampling could have not been as intentional although pragmatic" [Interview D]. To avoid these pitfalls when funding and timelines are tight, consortia can standardize and underwrite community engagement as a required, budgeted workstream, drawing on established guidance and reusable assets: early, continuous engagement playbooks with clear milestones and line items; data governance frameworks that foreground community rights and benefit-sharing [e.g., CARE Principles for Indigenous Data Governance' (Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE) Principles]; proportionate, interoperable consent and data-sharing policies (e.g., GA4GH Framework and Consent Policy); and scalable consent mechanisms that maintain participant control without bespoke redesign for every project (e.g., the GA4GH Toolkit) (Carroll et al. 2020). Frameworks such as Ownership, Control, Access, and Possession Principles (OCAP), which affirm Indigenous ownership and control over data, and the Transparency, Responsibility, User focus, Sustainability and Technology (TRUST) Principles, which emphasize transparency, accountability, and reciprocity in data use, further exemplify how ethical guardrails can be operationalized (Lin et al. 2020; Mecredy et al. 2018). These policy levers can streamline approvals and make community-centered design feasible within constrained budgets (Carroll et al. 2020; Knoppers 2014).

iv. Generalization of population labels

Population descriptors are terms or labels used to categorize and describe human populations based on factors such as ethnicity, race, geography, culture, religion, language, ancestry, or other features (NASEM 2023). In genetic research, population descriptors are often employed to classify individuals into distinct groups for the purpose of studying genetic variation and understanding the genetic basis of traits or diseases. Discourse surrounding population descriptors and their implications in genetic research involves complex social and historical contexts that shape population categories and the potential consequences of their uncritical use in genetic studies. One conversation surrounding the use of population descriptors is the oversimplification and essentialization of complex human diversity. By reducing populations to broad categories based on race or ethnicity, researchers risk overlooking the cultural and genetic heterogeneity within and across populations. This can reinforce harmful stereotypes, perpetuate scientific racism, and obscure the underlying genetic factors contributing to health disparities. The Coriell Institute provided guidelines on how to refer to populations from scientifically, ethically, and culturally sound standpoints, when using the samples derived from the HapMap and 1000 Genomes (Coriell Institute for Medical Research n.d.). Since there is overlap in the samples from the HapMap and 1000 Genomes, the guidelines are in place when referring to either sample group. These guidelines and recommended descriptors serve as a way to ensure precision when naming communities—not too broad and yet not too specific. The labels applied to HapMap and 1000 Genomes were based on asking groups what labels should be applied to them.

Despite the recommendations given in Coriell, there are still "continental" or "superpopulation" labels applied to the HapMap and 1000 Genomes samples; these labels do not capture the diversity and give an "inflated sense of homogeneity within the continent", which can lead to false positive and false negative findings (HPRC-ELSI Working Group 2023). In the 1000 Genomes metadata the granular, community-engaged "sub-population" labels are present alongside higher-level continental "super-population" labels, and those continental tags are sometimes applied inconsistently by downstream resources. The core problem, therefore, is not merely what identifiers are stored but how continental labels are used in analysis. Grouping or interpreting data at the continental level can obscure fine-scale structure and inflate between-continent differences. A recent inquiry into the use of cell lines labeled with 'Mexican Ancestry from Los Angeles (MXL)' and/or 'African ancestry in the American Southwest (ASW)'—not continental labels—was

brought about by the HPRC-ELSI Working Group (WG), as as project leaders and the ELSI working group did not want to "recruit around" tribal policies governing genomics research, even inadvertently, and there was concern that those cohorts might be interpreted as Indigenous. These were labels determined by the communities themselves, after community consultation. Following discussions with those directly involved in community consultation with those cohorts, and time dedicated to consider and respect Indigenous Data Sovereignty, the HPRC-ELSI WG emphasized the importance of complying with the original recommendations for cohort sample descriptions from 1000 Genomes and with the NASEM's recommendations on population descriptors (HPRC-ELSI Working Group 2023). Ultimately, samples from 'MXL' and 'ASW' populations were approved to be included in HPRC, with recommendations to not use 'super-population' labels, to not use labels as a proxy to measure Indigenous American ancestry, and to not re-identify anyone in any of the HPRC collections (HPRC-ELSI Working Group 2023).

When involving communities of specific cultural needs and historically vulnerable populations, upholding our promises to allow the labels to be what they chose for themselves would be the "best practice" for the use of the labels. However, even when this is implemented, scientists may revert back to the generalizations and inaccurate population descriptors once the genetic database is public. Collecting a high-level continental tag can be useful for coarse sampling summaries, but using continental categories as analytic strata is typically too broad: it collapses heterogeneous groups, invites ecological fallacies, and can bias association tests through population stratification even when more informative, nonracial descriptors (e.g., community self-identification, geography, language, genetic sequence similarity) are available. It can be challenging to achieve consensus and widespread adoption of alternative population categories. To curb this reversion, projects should require community-determined descriptors in metadata; mandate explicit justification and sensitivity analyses when continental groupings are used; provide standardized usage notes and automated flagging of continental-based analyses; and, in education, integrate core training in research ethics, history of science, anthropology, and community-engaged methods across science curricula from high school through graduate school, with practice in study design, consent, data governance, benefit sharing, and power dynamics.

IV. Current projects to utilize lessons learned

Global public good projects like the Human Pangenome Project (HPP) or Human Pangenome Reference Consortium (HPRC) will need to pioneer new approaches. The Human Pangenome Reference Consortium is an NIH-funded program, launched in 2019, that is building a high-quality human pangenome reference to better represent global genomic diversity while developing new assembly methods and tools. Its first public draft in 2023 assembled 47 phased, diploid genomes from a genetically diverse cohort, and the consortium has stated a plan to scale toward about 350 individuals. (Liao et al. 2023). Data for Release 2 have been made public, and a series of publications presenting and using the data are in progress. The Human Pangenome Project is a broader international alliance that coordinates access to high-quality genomes representing worldwide diversity and aims to maintain an integrated pangenome resource, which improves variant discovery compared with a single linear reference (Human Pangenome Project (HPP)" n.d).

The traditional exhortations to "engagement" and "community involvement" pose unique challenges because it is impractical to consult with all possible global constituencies, and yet, the goal is a globally represented public resource. Achieving universal representation requires the inclusion of individual countries, communities, and cultures. However, navigating this complex landscape raises profound questions. How can such projects authentically engage with diverse communities worldwide? How can they ensure equitable participation and representation while addressing the interests and concerns of *each* group?

V. Future directions

ELSI's critical role in genetic research can contribute to sound scientific practices by questioning and raising awareness about topics that may hinder the

legitimacy of a project. These considerations are not secondary but integral to the ethical conduct of scientific inquiry. Researchers can proactively address potential ethical dilemmas, safeguard the rights and well-being of research participants, and mitigate the risks of harm or exploitation. ELSI frameworks provide a structured approach for navigating complex ethical issues, such as informed consent, privacy protection, and the equitable distribution of benefits and burdens. Moreover, integrating ELSI considerations into research planning and implementation fosters a culture of responsible science.

Open dialogue, community engagement, and transparent communication are central to the success of genetic diversity projects. Meaningful engagement with diverse communities ensures that research initiatives are culturally sensitive, respectful, and responsive to the needs and concerns of those involved. By involving communities in the research process from inception to dissemination of findings, researchers can build trust and enhance the relevance and impact of their work. Transparent communication further promotes accountability, clarity, and understanding among stakeholders, enabling informed decision-making and promoting public confidence in genetic research endeavors. Furthermore, the lessons learned from past projects demonstrate the need for continued respect of autonomy, cultural beliefs, and values of research participants, ensuring that their voices are heard and valued throughout the research process. Safeguarding the rights and well-being of marginalized or vulnerable populations by researchers requires vigilance against exploitation and stigmatization. This involves identifying power imbalances, promoting equity and inclusivity, and upholding ethical standards. To translate this into practice, two concrete steps can be implemented immediately to make such efforts much more successful. First, require an ELSI workplan in every proposal as a scored review criterion with dedicated budget lines and auditable milestones, implemented by agencies during grantmaking processes. This plan should map stakeholders and community partners, specify roles and compensation, define consent models, outline data governance and stewardship, describe benefit sharing and capacity building, set engagement checkpoints from study inception through dissemination, and commit to plain-language communications and result return. Second, integrate education in ethics, history of science, anthropology, and community-engaged methods across all points of science training. This can be implemented through core courses such as case studies, practicums with community partners, and assessments of ethical reasoning, data governance, and reflexivity.

VI. Conclusion

Genetic diversity projects represent a journey through scientific discovery, ethical challenges, and societal implications. These initiatives, spanning decades from the Human Genome Project to the 1000 Genomes Project, have provided insights into human genetic variation, disease susceptibility, and population dynamics. However, they also presented significant ethical challenges and controversies, from issues of informed consent and community engagement to the politicization of genetic research and the oversimplification of population descriptors. Through reflection on past efforts, important lessons came about on the importance of integrating ethical considerations into scientific practices, improving cultural competence in research design, and continued community engagement throughout the research process. Improving ethical integrity, cultural sensitivity, and inclusive practices will ensure that scientific advancements benefit all members of society equitably. By learning from the past, the complexities of human genetic diversity can be discovered with greater awareness, compassion, and respect for the diverse communities that science is meant to serve.

Appendix: Interview Code
Interview A: HGDP organizer, 2022
Interview B: Bioethicist, 2023
Interview C: HapMap organizer, 2022
Interview D: Lawyer and bioethicist, 2023
Interview E: HapMap Historian, 2023
Interview F: Historian, 2023

References

Adame, Fernanda. 2021. "Meaningful Collaborations Can End 'Helicopter Research.'" *Nature*. https://www.nature.com/articles/d41586-021-01795-1.

Bergström, Anders, Shane A. McCarthy, Ruoyun Hui, Mohamed A. Almarri, Qasim Ayub, Petr Danecek, Yuan Chen, et al. 2020. "Insights into Human Genetic Variation and Population History from 929 Diverse Genomes." *Science* 367 (6484): eaay5012. https://doi.org/10.1126/science.aay5012.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. "The CARE Principles for Indigenous Data Governance". *Data Science Journal* 19 (1): 43. https://doi.org/10.5334/dsj-2020-043.

Cavalli-Sforza, Luca., A.C. Wilson, C.R. Cantor, R.M. Cook-Deegan, and M.-C. King. 1991. "Call for a Worldwide Survey of Human Genetic Diversity: A Vanishing Opportunity for the Human Genome Project." *Genomics* 11, no. 2: 490–91. https://doi.org/10.1016/0888-7543(91)90169-f.

Cavalli-Sforza, Luca. 1994. "The Human Genome Diversity Project." *The Human Genome Diversity Project (Conference)*. https://www.osti.gov/servlets/purl/505327.

Cavalli-Sforza, Lucacol. 2005. "The Human Genome Diversity Project: Past, Present and Future." *Nature Reviews Genetics* 6: 333–340. https://doi.org/10.1038/nrg1596.

Central Australian Aboriginal Congress. 1994. "The Vampire Project: An Aboriginal Perspective on Genome Diversity Research." *25(3):* 88-90.

Collingridge, David. 1980. *The Social Control of Technology*. St. Martin's Press.

Coriell Institute for Medical Research. n.d. "Population Descriptors | Guidelines for Samples | Coriell." https://www.coriell.org/1/NHGRI/About/Guidelines-for-Referring-to-Populations

Dodson, M., and R. Williamson. 1999. "Indigenous Peoples and the Morality of the Human Genome Diversity Project." *Journal of Medical Ethics* 25(2): 204-208. https://doi.org/10.1136/jme.25.2.204.

Dolan, D. D., Lee, S. S., & Cho, M. K. 2022. Three decades of ethical, legal, and social implications research: Looking back to chart a path forward. *Cell genomics,* 2(7), 100150. https://doi.org/10.1016/j.xgen.2022.100150.

National Academies of Sciences, Engineering, and Medicine. 2019. "Framework for Addressing Ethical Dimensions of Emerging and Innovative Biomedical Technologies: A Synthesis of Relevant National Academies Reports." Washington, DC: *The National Academies Press*. https://doi.org/10.17226/25491.

National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Health and Medicine Division, Committee on Population, Board on Health Sciences Policy, and Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research. 2023. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. Washington, DC: National Academies Press.

Greely, Henry. T. 1999. "The Overlooked Ethics of the Human Genome Diversity Project." *Politics and the Life Sciences* 18(2): 297–299. https://doi.org/10.1017/s073093840002150x.

"HapMap 3." n.d.*Wellcome Sanger Institute*. https://www.sanger.ac.uk/data/hapmap-3/.

"Haplotype." n.d. *National Human Genome Research Institute*. https://www.genome.gov/genetics-glossary/haplotype.

"HPRC-ELSI Working Group Statement on the Use of 'ASW' and 'MXL' Cell Lines from the 1000 Genomes Project Collection at Coriell Institute." 2023. https://humanpangenome.org/sample-use-asw-mxl/.

Human Genome Project. n.d. *National Human Genome Research Institute*. www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project.

"Human Pangenome Project (HPP)." n.d. https://humanpangenomeproject.org/.

IGSR: The International Genome Sample Resource. n.d. *1000 Genomes Project Summary | 1000 Genomes*. https://www.internationalgenome.org/1000-genomes-summary.

Kevles, Daniel J. 1985. 1939-. In the Name of Eugenics : Genetics and the Uses of Human Heredity. New York :Knopf.

Knoppers B. M. 2014. Framework for responsible sharing of genomic and health-related data. The HUGO journal, 8(1), 3. https://doi.org/10.1186/s11568-014-0003-1

Koenig, Z., M. T. Yohannes, L. L. Nkambule, X. Zhao, J. K. Goodrich, H. A. Kim, M. W. Wilson, et al. "A Harmonized Public Resource of Deeply Sequenced Diverse Human Genomes." *Genome Research* 34, no. 5 (2024): 796-809. https://doi.org/10.1101/gr.278378.123.

Largent, Mark. A. 2014. "On Daniel J. Kevles's In the Name of Eugenics." *Historical Studies in the Natural Sciences* 44(5): 514–520. https://doi.org/10.1525/hsns.2014.44.5.514.

Liao, WW., Asri, M., Ebler, J. et al. 2023. *A draft human pangenome reference*. Nature 617, 312–324 . https://doi.org/10.1038/s41586-023-05896-x.

Lu YF, Goldstein DB, Angrist M, Cavalleri G. 2014. Personalized medicine and human genetic diversity. *Cold Spring Harb Perspect Med* 24;4(9):a008581. https://doi.org/10.1101/cshperspect.a008581.

McEwen, J. E., Boyer, J. T., Sun, K. Y., Rothenberg, K. H., Lockhart, N. C., & Guyer, M. S. 2014. The Ethical, Legal, and Social Implications Program of the National Human Genome Research Institute: reflections on an ongoing experiment. *Annual review of genomics and human genetics* 15, 481–505. https://doi.org/10.1146/annurev-genom-090413-025327.

Mecredy, Graham, Roseanne Sutherland, and Carmen Jones. 2018. "First Nations data governance, privacy, and the importance of the OCAP principles." *International Journal of Population Data Science* 3, no. 4. https://doi.org/10.23889/ijpds.v3i4.911.

Michael Yudell, Roberts, D., DeSalle, R., & Tishkoff, S. 2016. "Taking Race Out of Human Genetics." *Science* 51,564-565. https://doi.org/10.1126/science.aac4951.

Minasny, B., D. Fiantis, B. Mulyanto, Y. Sulaeman, and W. Widyatmanti. 2020. "Global Soil Science Research Collaboration in the 21st Century: Time to End Helicopter Research." *Geoderma* 373: 114299. https://doi.org/10.1016/j.geoderma.2020.114299.

Proctor, Robert N. 1988. "Science and Nazism." *Science* 241 (4866): 730–31. https://doi.org/10.1126/science.241.4866.730.

Reardon, Jenny. 2005. *Race to the Finish: Identity and Governance in an Age of Genomics*. Princeton University Press.

Rotimi, Charles., M. Leppert, I. Matsuda, C. Zeng, H. Zhang, C. Adebamowo, I. Ajayi et al., and the International HapMap Consortium. 2007. "Community Engagement and Informed Consent in the International HapMap Project." *Community Genetics* 10(3): 186–198. https://doi.org/10.1159/000101761.

Sunnucks P. 2000. "Efficient genetic markers for population biology." *Trends in ecology & evolution* 15(5), 199–203. https://doi.org/10.1016/s0169-5347(00)01825-5.

The International HapMap Consortium. 2003. "The International HapMap Project." *Nature* 426: 789–796. https://doi.org/10.1038/nature02168.

The 1000 Genomes Project Consortium. 2015. "A global reference for human genetic variation." *Nature* 526, 68–74. https://doi.org/10.1038/nature15393.

Weiss, K. M., Cavalli-Sforza, L. L., Dunston, G. M., Feldman, M., Greely, H. T., Kidd, K. K., King, M., et al., & the North American Regional Committee of the Human Genome Diversity Project. 1997. "Proposed model ethical protocol for collecting DNA samples." *Houston law review*, 33(5), 1431–1474. https://pubmed.ncbi.nlm.nih.gov/12627556/.

Keerthana Balaji is a recent graduate from Barrett, The Honors College at Arizona State University, with degrees in Biomedical Sciences (BS) and in Global Health (BA). She is passionate about bioethics and medicine. Keerthana has been a research fellow through ASU and the Human Pangenome Reference Consortium for the past few years. She plans to pursue a career in medicine in future.

Robert Cook-Deegan is a professor in the School for the Future of Innovation in Society, and the Consortium for Science, Policy & Outcomes at Arizona State University. He directed Duke's Center for Genome Ethics, Law & Policy 2002-2016. Previous positions were at the National Academies of Science, Engineering and Medicine; National Center for Human Genome Research (NIH); and the congressional Office of Technology Assessment. His MD is from the University of Colorado and BA in chemistry (*magna cum laude*) from Harvard. He is the author of *The Gene Wars: Science, Politics, and the Human Genome* and over 350 other publications.