2022 ASHG presidential address—One human race: Billions of genomes

Charles N. Rotimi[1,*]

This article is based on the address given by the author at the 2022 meeting of The American Society of Human Genetics (ASHG) in Los Angeles, California. The video of the original address can be found at the ASHG website.

Welcome

I am excited to welcome you in person to Los Angeles for our 2022 annual meeting. I look forward to reconnecting with as many colleagues as possible in the coming days and to engaging in discussions on research findings and how we move human genetics forward globally. I want to thank all of you for placing your trust in me to serve as your president this year. Performing my responsibilities as the president was made infinitely easier by the dedication and professionalism of the ASHG staff and my colleagues on the board of directors and on the various committees who volunteer to serve our society. My sincere gratitude to all.

Before I go into my presidential address, I would like to draw your attention to two upcoming events. (1) the Presidential Symposium on exciting dialogue about what African genomics is and is becoming. The symposium will highlight the profoundly dynamic and diverse continent's major advances, new directions and goals, emerging scientific leadership, exciting investment in technology infrastructure, and more; how can and will genomics in Africa "spread its wings," and what areas are most exciting? Join a global community for an exciting dialogue about what African genomics is—and is becoming—this Thursday, October 27, 2022, at 8:30 a.m. (2) The International Congress on Human Genetics is taking place February 22 –26, 2023 in beautiful Cape Town, South Africa; the theme of the conference is "Coming Home," which acknowledges our common origin in the geographical region called Africa today. I do sincerely hope that you will join us in person because it is difficult if not impossible to "Come Home" virtually.

I will begin my address with an acknowledgement of the wonderful progress the global scientific community is making towards fulfilling the promise of how understanding the information coded in human genomes will revolutionize our appreciation of who we are, where our genetic ancestors come from, and why there is differential susceptibility or resistance to disease in diverse environmental contexts. The developing understanding of the evolutionary history and function of the about 4% of the modern human genome that came from interbreeding with archaic humans including Neanderthals and Denisovans is a clear demonstration of the power of genomics. I use this opportunity to congratulate our colleague, Svante Pääbo, who was awarded this year's Nobel Prize in Physiology or Medicine for his discoveries concerning the genomes of extinct hominins and human evolution. Collectively, progress in human genetics and related fields is moving us closer to the full integration of genomic understanding of biology into the day-to-day practice of medicine and the development of new therapeutics for previously intractable human diseases, including several types of cancers, and gene-editing to cure sickle cell disease.

However, genomics-driven scientific and medical innovations are currently not shared equitably by all human populations with the resulting well-documented global challenge of lack of diversity in both the participants and scientists that are engaged in genomic sciences. As has been acknowledged by me and others, if not urgently and systematically addressed, these challenges will likely compromise our goal and vision that "people everywhere realize the benefits of human genetics and genomics research." While more work is needed, I am encouraged by efforts underway in Africa, the Middle East, Asia, and North and South America and the call to action by several organizations, including the World Health Organization, the National Human Genome Research Institute at NIH, the International Common Disease Alliance, the Global Alliance for Genomics and Health, and our society.

ASHG is contributing to achieving the promise of genomics in a variety of ways, including (1) serving the research community by providing a forum where scientists present and share transformative approaches and technologies in

[1]Woodstock, Maryland, USA
*Correspondence: charles.rotimi@gmail.com
https://doi.org/10.1016/j.ajhg.2023.01.011.

human genetics and medicine; (2) providing opportunities for trainees and scientists from diverse backgrounds to attend annual meetings; (3) increasing diversity of its board of directors, as illustrated by the change in the number of board members self-identifying as non-European ancestry from 5% as recently as 2017 to 40% today; (4) acknowledging past harms in human genetics, such as eugenics and reinforcement of racial stereotypes; (5) highlighting the evolving use of ancestry, ethnicity, and race in human genetics in the society's journals; and (6) proudly supporting ASHG members who are helping to lead and shape the work of a National Academies committee (https://www.nationalacademies.org/our-work/use-of-race-ethnicity-and-ancestry-as-population-descriptors-in-genomics-research) to assess the existing methodologies, benefits, and challenges in the use of race, ethnicity, and other population descriptors in genomics research.

Very similar, but different

What I would like to talk about in the remaining few minutes of my presidential address is the question that I am often asked when I give talks: why is it important to have diversity in research participants? This question is usually asked after I make the statement that "human genomes are greater than 99% similar" irrespective of where the DNA samples come from globally. Of course, the moment I invoke the non-random distribution of the less than 1% of human genetic variations that differ by geography and between individuals and groups, I am also queried about another statement that I and others make—i.e., genomic data do not support the existence of distinct human racial groups and "race" is not an objective genomic classifier.

So how do we begin to unpack these statements that seem orthogonal to each other so we can advance our science and allow all of us to simultaneously appreciate our individuality and our common humanity? I would like to challenge us to ask why we continue to tolerate the use of imprecise labels that we know are hindering our understanding and interpretation of how genetic and non-genetic factors influence human health and identity. I ask this question because most geneticists, epidemiologists, anthropologists, and clinicians that I interact with acknowledge that descriptors such as Black, White, Hispanic, and Asian are at best imprecise proxies to the causal factors that underly health inequities globally. However, some scientists are reluctant to eliminate these racial labels for different reasons, including how they were trained and their observations that the labels capture some information that is important to scientific and clinical practices. My response is usually, yes that there is some correlation between labels like Black or Hispanic with some health indicators, but this is because such labels are mega-variables that capture several risk factors at once. We need to understand what component of these mega-variables matters for particular outcomes; otherwise, we will continue to just identify correlation with these labels without changing the underlying factors that drive these inequalities, such as income, education, and living in economically neglected neighborhoods. I will mention two recent examples. First, the clean water crises in Flint, Michigan and Jackson, Mississippi. Most individuals affected by this crisis are Black, but this correlation is not necessarily helpful for solving the problem because Blacks that live in rich, mostly White, neighborhoods are not affected and Whites living in poor, mostly Black, neighborhoods are affected. So, in this example, living in certain neighborhoods is more important than whether you are Black or White. However, it would be negligent not to acknowledge why these neighborhoods became predominantly Black and neglected. In the case of Jackson, MS, it is clearly the combined effects of racism and political ideology—residents of Jackson are mostly Democrats, while the state's leadership, including the governors, is mostly Republican. The second example is the well-established causal relationship between *APOL1* and kidney diseases. The prevalence of the *APOL1* renal risk variants is more common in individuals from geographical regions of sub-Saharan Africa affected by trypanosomiasis (the so-called African sleeping sickness) because these variants conferred evolutionary advantage against this disease. It is important to note that it is not a defining characteristic of being African or Black and that because of admixture and gene flow, these risk variants have spread across the world such that using labels like African or Black seriously distorts our understanding of individuals at risk of kidney disease due to these variants.[1] This misunderstanding can lead to missed diagnoses with potentially deadly consequences among persons who look phenotypically European or Hispanic.

Looking back to move forward

To move forward, we must appreciate the pre-genetic history of human racial classifications from as far back as the 15th century. It is important to note that throughout this long history, there has never been agreement regarding the number of races nor their defining characteristics. Taxonomically, François Bernier in his 1684 work, "A new division of the Earth by the different species or races which inhabit it," divided humans into four geographic races: one covering Europe, North Africa, Western Asia, and parts of Southeast Asia; one covering the rest of Africa; one covering the rest of Asia, including Central, East, and North Asia; and one including only the Lapps. He was uncertain if Native Americans constituted a fifth race but was inclined to include them with Europeans. Although Bernier's primary consideration was geographic, he described generalization of physical characteristics including skin color and facial features. Today, the world still has no clear definition of race and racial classifications continue to be fluid.

For example, race in the United States of America is now based on self-identification, with individuals allowed to report more than one race group with the U.S. Office of Management and Budget (OMB) requiring a minimum of

five categories (White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander). Importantly, OMB clearly states that racial categories included in the census questionnaire generally reflect a social definition of race recognized in this country and not an attempt to define race biologically, anthropologically, or genetically (https://www.census.gov/topics/population/race/about.html). In addition, it is recognized that the categories of the race item include racial and national origin or sociocultural groups. People may choose to report more than one race to indicate their racial mixture, such as "American Indian" and "White." People who identify their origin as Hispanic, Latino, or Spanish may be of any race. If these centuries of inconsistencies in the meaning of racial categories do not challenge our professional and personal conscience about the use of race in research, I am at a loss as to what will do it. At a minimum, we may adopt the International HapMap (https://www.genome.gov/10001688/international-hapmap-project) standard for population descriptors, such as Yoruba from Ibadan, Nigeria instead of Blacks or Africans; Utah residents of northern and western European ancestry instead of Whites or Caucasians; and Han Chinese individuals from Beijing, China instead of Asians or East Asians.

Ancestry and identity

To help us move beyond the use of pre-genetically defined racial categories in research and clinical practice, it is important to acknowledge that the level of genetic variation that geneticists see across human populations does not rise to the level of sub-speciation. This is obvious because within anatomically modern humans, there is no biological evidence for any reproductive barriers such as anatomical or physiological incompatibilities. However, the small (<1%) differences in genetic variations between individuals and populations are not random and may be important in understanding our ancestral backgrounds and risks of developing certain diseases and responses to therapeutics. For example, population geneticists can use this non-random genetic variation to define complex patterns of genetic admixture and ancestries across human populations. A recent study[2] that investigated genetic ancestry in a sample of about 6,000 individuals from 30 primary language families from around the world identified 21 genetic ancestries with over half of these ancestries identified among present-day Africans, emphasizing the huge genetic diversity across Africa. Notably, individuals in this global dataset had, on average, DNA from four ancestries, the majority (97.3%) displayed mixed ancestry, and multiple ancestries mapped to each continent. It is essential to point out that because genetic ancestry is determined solely by genomic data, it is not subjectively self-identified and is subject to evolutionary changes. This means that ancestries, in the long term, are subject to birth-death cycles and ancestry-specific allele frequencies can change over time. Also, the number of ancestral groups can change as the diversity of genomic reference databases improve. In all, this study showed that ancestry cross-classifies ethno-linguistic group as well as continent and racial categories. For example, Western Asian ancestry, with highest frequency in peoples from the Caucasus Mountains and the Levant, is the major ancestry in Abkhazian, Georgian, and Druze samples and is present in samples with origins ranging from Morocco to Mongolia and from England to Ethiopia. That is, Western Asian ancestry simultaneously exists in Africa, Asia, and Europe, as well as in the US racial categories Black or African American, Asian, and White. In contrast to race categories, genetic ancestry is a valid genomic classifier.

Ancestry and health

Ancestral backgrounds have significant implications for disease prevalence, severity, and variable response to therapeutics. Examples include Tay-Sachs diseases among persons of Ashkenazi Jewish ancestry, hemoglobinopathies (sickle cell disease, thalassemia, and G6PD deficiency) in people with African or Middle Eastern ancestry, *APOL1*-associated kidney disease and hypersensitivity syndrome (AHS) in response to abacavir for HIV treatment in some African-ancestry populations. However, because the genetic background that underly these biological observations may be present in other ancestries (e.g., Tay-Sachs disease is seen in non-Jewish populations, hemoglobinopathies and AHS in non-African populations), racial or ethnic labeling of these health outcomes can render them invisible in some ancestral groups, leading to many years of misdiagnoses, incorrect public health screening decisions, and pain and suffering. Illustratively, the frequency of the HLA-B*5701 allele that prevents AHS ranges from 0% to ~20% globally. Among individuals who may be labeled as "Black" or "African," frequency varies from ~14% among the Kenyan Maasai group to 1% among African Americans and 0% among Yoruba from Nigeria.[3] Thus, the use of these and similar social labels ("White," "Asian," and Hispanic") can render biomedically relevant variation invisible in some populations and over diagnosed in others.

The use of racial categories in public health and sociopolitical policymaking in countries like the US is where I see perhaps the biggest challenge in moving beyond racial categories in science. I say this because, despite its flaws, some scientists and policymakers are concerned that eliminating them will make it difficult to collect necessary data to address racism in society. My suggestion, perhaps naive, is to keep and use them as social labels without calling them racial categories. For example, when we say "African Americans" in the US, we are referring to a cultural group with complex patterns of admixture and ancestry. It is not a genetically defined distinct group. Personally, and in my research, the only way I can consistently define African Americans are those who are the descendants of the Middle Passage who live in the US. Also, while I commend all persons that are working hard to root out racism in every society where it exists, I am beginning to

accept that racism is a societal infection that cannot be scientifically and medically cured. Therefore, I believe the slow progress in eliminating racism is not because of lack of data or evidence. It is because the source of the social problem of racism lies with racists and not in the intrinsic biology of their victims, as stated in the 2004 *Nature Genetics* editorial entitled "The unexamined population."[4] Hence, eliminating racism and dismantling the sociopolitical structures that give it oxygen requires the awakening in the consciousness of humanity to the devasting health, economic, and safety consequences of the practice of racism. We have all felt the impact of recent examples, such as the public murder of George Floyd by a White police officer in Minneapolis; the shooting death of Ahmaud Arbery while jogging in the streets of Brunswick, Georgia by a White father and son; and the labeling of the devastating coronavirus using disparaging terms because of its origins in China, leading to backlash among those of East Asian ancestry. Racism needs to be investigated and fully prosecuted. Given the unique role of our society in human genetics, we must use a single unapologetic voice to confront lingering eugenics thinking, racism, and systemic inequities and their devastating consequences on humanity.

I would like to end by calling on the ASHG board and membership to continue to take important steps to document and untie itself from the inequities that have been part of the society's past. If we are indeed committed to the goal of building an equitable future for genomics research and medicine, our society and funding agencies need to pay attention to ongoing global genomic efforts specially in low- and middle-income countries so we can continue to expand on the success of programs such as H3Africa (https://h3africa.org/) and GenomeAsia 100K (https://browser.genomeasia100k.org/). Scientists in high income countries, including the US and Europe, must elevate their collaborative strategies to be more fair and supportive of the development of research capacity with full appreciation for the challenges faced by their colleagues in resource-challenged environments. I would like to reinforce some key points. (1) Genetic differences between individuals and by geography are important in understanding human migration history and ancestral background and may have biological meaning that can help us to fully realize the promise of genomic medicine. While self-identification is the current desired approach in genomics and medicine, it is important to note that most individuals have mixed genetic ancestry that may not be fully recognized but is key in the context of providing individualized care. (2) Analysis of data from the global sequencing of hundreds of thousands of human genomes has now made it abundantly clear that we only have one human race but billions of genomes. The individual is the ultimate variant and collectively, human genome variation is a continuum and genetic differences among human populations are overwhelmingly the result of gradations in allele frequencies, not uniqueness. Hence, the Editorial Board of *Nature Biotechnology* in 2002 likened the attempt to use genetics to define distinct racial categories as trying to "slice soup"; you can cut all you want, but the soup stays mixed.[5] I thank you again for giving me the opportunity to serve as this year's president of this great society. I wish you the very best in life and please help to spread the word about our common humanity. We need it now more than ever. Please be safe and enjoy the conference and what the city of Los Angeles has to offer.

Acknowledgments

The author thanks Jamie E. Rotimi (daughter), Shawneequa Callier, Amy Bentley, Daniel Shriner, Mateus Gouveia, and Adebowale Adeyemo for their constructive inputs and feedback on the development of the draft manuscript. This manuscript represents the author's opinions and not the policies and views of any organization with which C.N.R. is affiliated.

References

1. Nadkarni, G.N., Gignoux, C.R., Sorokin, E.P., Daya, M., Rahman, R., Barnes, K.C., Wassel, C.L., and Kenny, E.E. (2018). Worldwide Frequencies of APOL1 Renal Risk Variants. N. Engl. J. Med. *379*, 2571–2572. https://doi.org/10.1056/NEJMc1800748.
2. Baker, J.L., Rotimi, C.N., and Shriner, D. (2017). Human ancestry correlates with language and reveals that race is not an objective genomic classifier. Sci. Rep. *7*, 1572. https://doi.org/10.1038/s41598-017-01837-7.
3. Rotimi, C.N., and Jorde, L.B. (2010). Ancestry and disease in the age of genomic medicine. N. Engl. J. Med. *363* (16), 1551–1558. https://doi.org/10.1056/NEJMra0911564.
4. (2004). The unexamined population. Nat. Genet. *36* (*Suppl 11*). S3. https://doi.org/10.1038/ng2151.
5. (2002). Slicing soup. Nat. Biotechnol. *20*, 637. https://doi.org/10.1038/nbt0702-637.