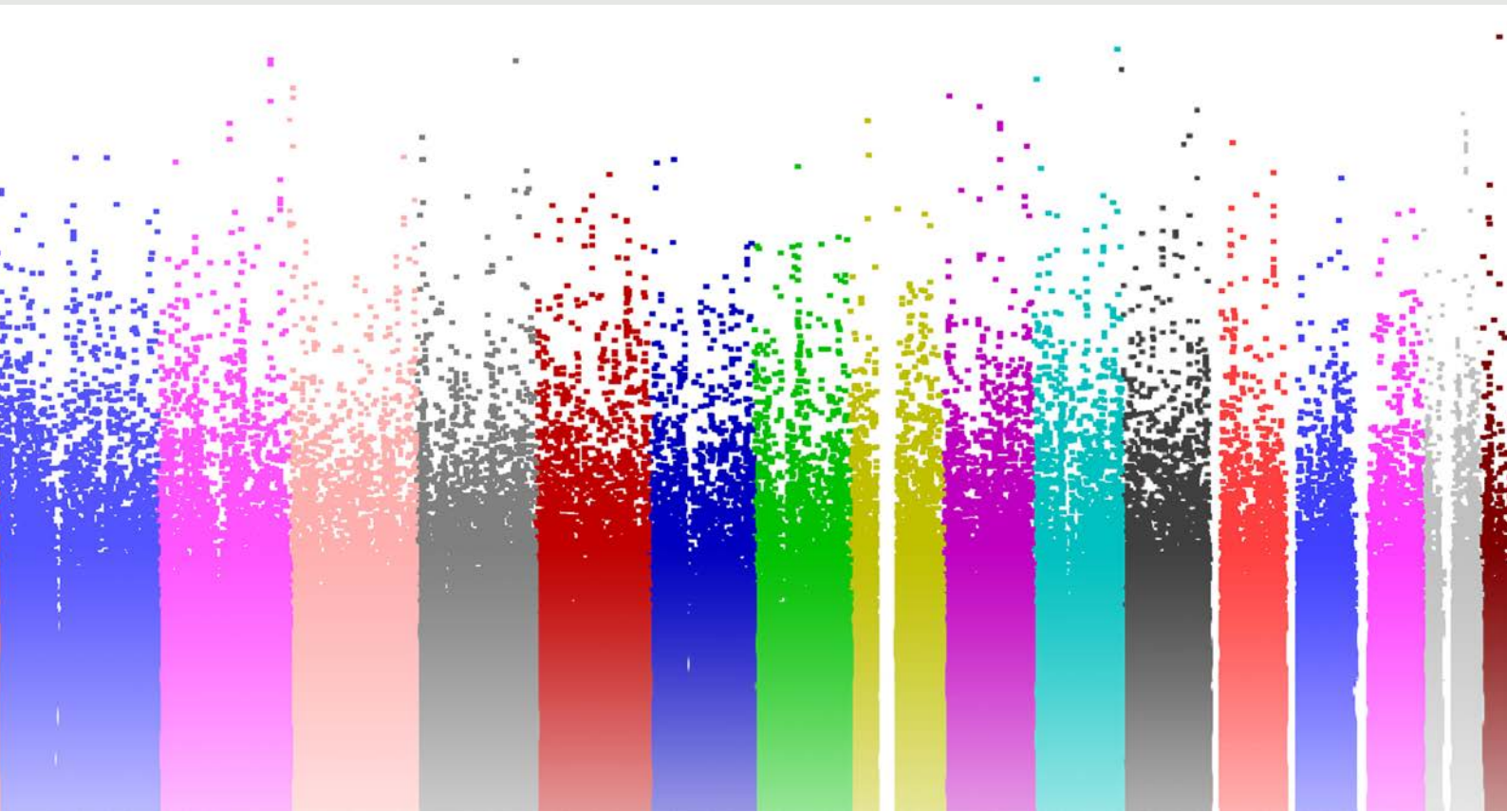


# DIVERSITY AND INCLUSION IN GENOME-WIDE ASSOCIATION STUDIES

How to make racial representation in GWAS more proportionate?  
Thought leaders from Africa, South Korea, and New Zealand, discuss  
the causes, challenges, and action being taken to close the gap.



REPRINTED FROM KIM S-H ET AL. PLOS ONE  
9, E111220 (2014), VIA CC-BY-4.0

For **illumina**<sup>®</sup>

by **nature** research  
custom media

# DIVERSITY AND INCLUSION IN GENOME-WIDE ASSOCIATION STUDIES

How to make **RACIAL REPRESENTATION IN GWAS** more proportionate? Thought leaders from Africa, South Korea, and New Zealand, discuss the causes, challenges, and action being taken to close the gap.

In 2020, an extensive study of African genomic data uncovered more than three million undescribed genetic variants, many of which from ethnic groups whose DNA were sampled for the first time.<sup>1</sup> This landmark genome wide association study (GWAS) was a leap towards addressing underrepresentation of many populations in the world's DNA databases, currently dominated by European data.

Knowledge of diverse genetic variants across populations increase the chances of discovering disease mechanisms and drug targets, as well as applying those findings to clinical practice. With cystic fibrosis, for example, a single genetic mutation accounts for more than 70% of cases in Europe. The disease was thought to be largely absent in Africa until researchers discovered unique variants in people with some African ancestry, which account for 40% of cystic fibrosis cases on the continent.<sup>2</sup> "In cases like this, diagnostic kits developed based on European datasets yield inaccurate results for patients in Africa," says Michèle Ramsay, a professor in the Division of Human Genetics at University of the Witwatersrand in South Africa. "Data from local people are

necessary to validate whether a particular approach will be clinically appropriate in that population."

## **PARTICIPATION IN CONSORTIA**

While Africa is the world's most genetically diverse continent, only 2% of GWAS represent those with African ancestry,<sup>3</sup> most of whom are African Americans. "If Africa is going to benefit from precision medicine or genomic medicine, it will only be possible if we generate data locally and share it," says Ramsay. "It's encouraging that scientists today are forming international consortia to collaborate on genetic studies. Africa has largely been left out of those collaborations, because to participate, data and resources are necessary, and these are scarce in Africa," says Ramsay.

While logistical constraints severely limit data collection, as Samuel Agyei Wiafe, the founder of the Rare Diseases Ghana Initiative observes, the Human Heredity and Health in Africa (H3A) consortium has helped address some of these obstacles. H3A — a collaborative effort between the African Society of Human Genetics (AfSHG), the National Institutes of Health (NIH) in the United States, and the UK's Wellcome Trust — was a game-changer that created the

foundations for coordinated, large-scale genomics in Africa.

"In addition to bringing major funding to Africa for the first time, and creating a 'large grant' culture, it also facilitated a pan-Africa approach to collaboration which will help us interpret data in meaningful ways in the future," says Ramsay. A major milestone was the creation of the H3ABioNet bioinformatics network, which facilitates local analysis of African data. H3ABioNet provides a communication framework across more than 15 countries, as well as computational infrastructure and hardware. Furthermore, it trains the next generation of researchers in data management, analysis and interpretation. By bringing together expertise and building future research capital, H3ABioNet has built a critical mass in bioinformatics that has enabled large-scale genomics in Africa.

**"IT IS NOT ENOUGH TO SIMPLY GIVE A PRESENTATION AND ASK FOR PERMISSION TO COLLECT DATA AND SAMPLES."**

The research cohorts under H3A include more than 100,000 participants in 30 African countries.<sup>4</sup> Ramsay says that

while the consortium's 10-year term ends in June 2022, it leaves a legacy for the next generation of researchers. Some projects and data from H3A will be included in the Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) programme, which was recently launched by the NIH Common Fund.

## **SCALING STUDIES WITH GOVERNMENT SUPPORT**

Major genome projects such as the UK Biobank, or the Singapore 10K Genome project, have typically run with significant government support, adds Ramsay. "GWAS for non-communicable diseases (NCDs) require large cohorts, preferably with longitudinal data, of tens of thousands of research participants," she explains. "Understandably, this has not been a priority for African governments, given the large infectious disease burden. However, the prevalence of NCDs is rising in African communities." In addition to governments, pharmaceutical companies are increasingly coming together to form precompetitive consortia that fund very large projects, such as the whole exome sequencing of the UK Biobank. Conversations have now begun about similar endeavours in Africa.

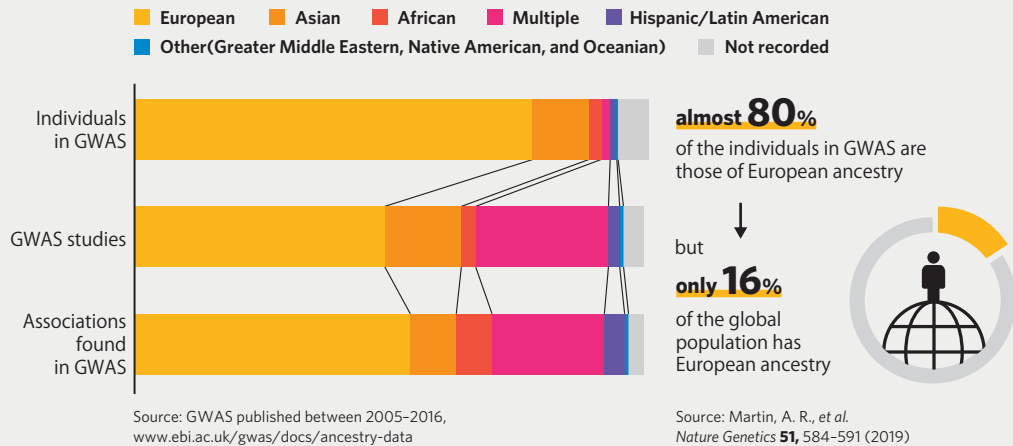
Meanwhile in Asia, the

## DIVERSITY IN GENOMICS

Underrepresentation of different ancestries in genome-wide association studies (GWAS) could create a disparity in healthcare and personalized medicine.

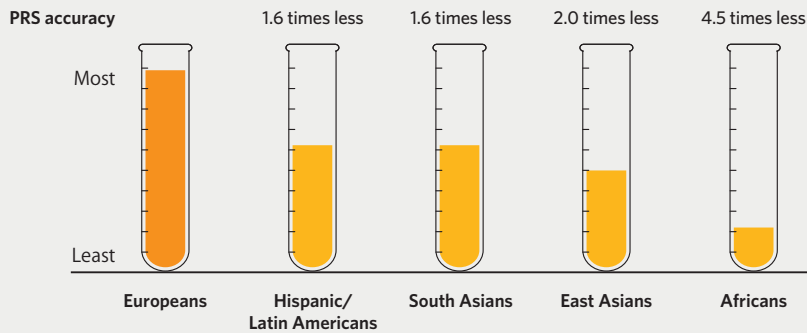
### Who participates in GWAS?

Individuals of European ancestry are over-represented in GWAS. This bias could lead to important associations from other groups, such as those of African descent, not being identified.



### Why does this matter?

Lack of genomic data from a range of ancestries creates a disparity in the level of medical care people receive. For example, polygenic risk scores (PRS) help predict risk based on genetic data, but due to the European bias in GWAS, are less accurate for those of other ancestries.



government of South Korea has advanced plans to increase genome research capability. In a multi-ministerial project involving the nation's top hospitals, the country will spend 1 trillion won (US\$ 830 million) over six years to build a national digital library on health and genome data. The National Project of Bio Big Data, was powered by Illumina's high-throughput DNA sequencing platforms and completed its first pilot establishing a database of 10,000 genomes. The second stage continues to employ Illumina's technology and will examine 12,500 genomes of patients with rare diseases.

"The strength of this

genome data is the medical information associated with it," explains Woong Yang Park, the Director of the Samsung Genome Institute in Seoul. "In South Korea, biannual health checkups are mandatory to keep your insurance. As universal healthcare began in 1988, the government has collected medical data for more than 30 years. With consent, we can use that information. The genome data itself may not be competitive with other countries, but the combined medical data makes it powerful for drug and target discovery, as well as for precision medicine."

In addition to collecting data,

however, researchers must make more of an effort to utilize Asian genome data. "While individuals of Asian ancestry account for 10% of individuals studied in GWAS, this is still not well utilized, even though recent exome analyses in populations show that Asia is much less homogenous than previously thought," says Park.<sup>5</sup>

Park believes this is partly due to the lack of structured cooperation. "The EU has organizations for integrating research across countries," he observes, "in contrast, in Asia, we have no kind of pan-continental organization, neither politically nor scientifically. Individual

countries do not have sufficient funds to finance large-scale GWAS projects". In contrast, in other non-GWAS projects, Park has seen early hints of success through collaborations with other researchers. Park is a co-principal investigator for the Asian Immune Diversity Atlas, which addresses the underrepresentation of Asian donors in the Human Cell Atlas. This brings together researchers from China, Japan, South Korea, India, Thailand, and Singapore. "Our individual datasets may be small and have limited interpretation, but coordinated efforts to use data will help derive meaningful results, which we could then

use to emphasize the value of Asian data. The same principles also apply to GWAS.”

### ENGAGING WITH DIVERSE ETHNIC GROUPS

Including data from diverse ethnic groups within continents is also paramount in improving representation. Challenges with defining ethnolinguistic groups, as well as language differences in raw data, complicate the logistics of studies in Africa. “There are more than 3,000 ethnolinguistic groups in Africa according to Ethnologue, but this can vary widely by how one defines a group. There is no standardized, pan-Africa categorization to date,” says Ramsay.

Widespread distrust of western science from many indigenous groups, given a history of exploitation, is also a challenge. In 2017, the San, from South Africa, published a code of research ethics outlining the requirements for a respectful partnership. “It is difficult for many groups to identify a leadership team who can speak on their behalf. Even for the San, there is a question of whether the group that established the code of ethics could fully represent all San people in sub-

Saharan Africa,” says Ramsay. “In community engagement, it is extremely important to identify who the community is, and to engage with them to reach common understanding. It is not enough to simply give them a presentation and ask for permission to collect data and samples. I would like to see funding agencies provide more funding for engagement, as these activities are time intensive, and can be very costly depending on where the populations are located and how complex their leadership structures are.”

In New Zealand, the nature of engagement with Māori has become more inclusive and the relationship has improved, says Māui Hudson, a research ethicist at the University of Waikato in New Zealand. “Funding agencies and ethics committees now have higher standards for the depths of engagement,” he says. “More recently, the tension is about Western science being positioned as the only truth. There are other ways of understanding the world, which govern how prepared people are to accept the results of a scientific project, as well as what appropriate use of the results looks like.”

Genomics Aotearoa, New Zealand’s strategic platform for building genomic capability, has supported projects to enhance engagement with indigenous communities. In 2016, Hudson and colleagues crafted a guideline on engaging with Māori for health research. In GWAS conducted worldwide, however, the representation of indigenous peoples – including Native Americans, Indigenous Australians and Pacific Islanders – has fallen from 0.06% in 2009 to 0.02% in 2019, despite an increase in participants.<sup>6,7</sup> Moves toward open data complicate the picture even further.

“In a truly open data environment, no one will have to contact indigenous communities to use that information, meaning that there is no consultation after data collection. It’s when researchers have to contact them to use the information, that we begin discussing the types of questions that should have been asked at the beginning of the process,” says Hudson. “If indigenous communities are more comfortable with the way in which they participate, see that they have more control over decision making, and feel that these are being done in a way that their integrity is maintained, then they will participate. But we are not there yet. For a substantial improvement to diversity and inclusion, the scientific community must be more open to having other people making decisions.”

“When there is a precedent of certain groups being described as inferior in the scientific literature, of course people have anxiety around what will happen to the data when it is shared. Some researchers also publish data without acknowledging the people they obtained it from,” adds Wiafe. “However, that does not mean that people are

unwilling to participate.”

Polygenic risk scores (PRS) help predict disease risk based on genetic data. However, due to the eurocentric bias in GWAS studies, it has been shown that PRS for those of African, South Asian, East Asian, and Hispanic/Latin American ancestries are significantly less accurate than for those of European backgrounds.<sup>8</sup> This can lead to a disparity in the level of medical care provided.

For GWAS to bring equitable benefits, genomics studies conducted locally by local researchers, as well as collaboration with local communities, are key. “People across the world are influenced by a complex diaspora from Africa and other parts of the world, and this affects the relationship between genetic diversity and our phenotype,” says Ramsay. “While many genetic variations are universal, we need to know when there are unusual population-specific variants, and then we need to understand their origins and potential impact on disease susceptibility.”

### REFERENCES

1. Choudhury, A. et al., *Nature* **586**, 741-748 (2020).
2. Stewart, C. & Pepper, M. S., *Genetics in Medicine* **18**, 653-662 (2016).
3. Wonkam, A., *Nature* **590**, 209-211 (2021).
4. <https://h3africa.org/>
5. Lee, S. et al., *Scientific Reports* **7**, 4287 (2017).
6. Garrison, N. A. et al., *Annual Review of Genomics and Human Genetics* **20**, 495-517 (2019).
7. Mills, M. C. & Rahal, C., *Communications Biology* **2**, 9 (2019).
8. Martin, A. R. et al., *Nature Genetics* **51**, 584-591 (2019).



For personalized medicine to be fully realized, genomics must be more equitable.

illumina®

[www.illumina.com](http://www.illumina.com)

ANDREW BROOKES/IMAGE SOURCE/GETTY