

Increasing diversity of functional genetics studies to advance biological discovery and human health

Sophia H.L. George,^{1,2,*} Alejandra Medina-Rivera,³ Youssef Idaghdour,^{4,5,6} Tuuli Lappalainen,^{7,8,*} and Irene Gallego Romero^{9,10,11}

Summary

In this perspective we discuss the current lack of genetic and environmental diversity in functional genomics datasets. There is a well-described Eurocentric bias in genetic and functional genomic research that has a clear impact on the benefit this research can bring to underrepresented populations. Current research focused on genetic variant-to-function experiments aims to identify molecular QTLs, but the lack of data from genetically diverse individuals has limited analyses to mostly populations of European ancestry. Although some efforts have been established to increase diversity in functional genomic studies, much remains to be done to consistently generate data for underrepresented populations from now on. We discuss the major barriers for this continuity and suggest actionable insights, aiming to empower research and researchers from underserved populations.

Background

After decades of human genetics research focusing largely on populations of European ancestry, the field has recently begun emphasizing and investing in research that encompasses a broader spectrum of environmental and human ancestral diversity, with the twin aims of diversifying existing human genetics resources and developing a global workforce. Projects such as H3Africa¹ and major biobank studies in several Asian countries^{2–4} have already provided crucial insights into genetic diversity and its links to human traits and diseases via genome-wide association studies (GWASs) carried out in cohorts more representative of global genetic diversity. However, much work remains to be done, as the Eurocentric bias in genetic research still prevails according to the GWAS Diversity Monitor: 95.2% European, 3.1% Asian, 0.2% African, 0.5% Afro-Caribbean/African American, 0.3% Hispanic/Latine, 0.7% other/mixed.⁵ Moreover, with research moving forward from GWASs to population functional genomic studies, representation bias remains a challenge. Addressing this issue is on the agenda of major research programs, funders, and many research groups.

While the lack of diversity in genetic and clinical studies has been widely discussed, there is an even more profound lack of genetic and environmental diversity in functional genomics datasets (Figure 1), which is the focus of this perspective and an issue only slowly starting to receive attention.⁶ Functional genomics data are used to understand the molecular processes that underlie human traits

and bridge the gap between genotype and trait. Indeed, functional interpretation of GWASs has recently emerged as perhaps the biggest contemporary challenge in human genetics, as identifying molecular mediators of GWAS effects is necessary to translate GWAS discoveries into biological insights and medical benefits.

In contrast to well-established practices for metadata curation in GWAS repositories, functional genomic studies do not systematically identify ethnicity, ancestry, or environmental metadata of samples, hampering further analysis aiming to integrate this information. Despite efforts such as TOPMed and All of Us investing in more balanced multi-omic studies of minority populations in the United States as well as multiple national population biobanks, this alone will not fix gaps in global diversity in functional genomics research.^{11,12} While some of the underlying reasons are partially shared with those limiting the diversity of GWASs, such as the long legacy of research funding and capacity building in high-income countries (HICs),¹³ functional studies have additional unique challenges that we discuss in this perspective.

We argue for decisive efforts to address these bottlenecks to diversify functional genomics studies and realize the benefits of human genetics both for basic biological discovery and for human health worldwide. The imperative for remedying this bias is not simply moral. A focus on individuals leading predominantly sedentary lives in urbanized settings in HICs provides only a narrow representation of both extant human genetic diversity and the range of social, cultural, and physical environments occupied by

¹Department of Obstetrics, Gynecology and Reproductive Sciences, Miller School of Medicine, University of Miami, Miami, FL, USA; ²Sylvester Comprehensive Cancer Center, Miami, FL, USA; ³Laboratorio Internacional de Investigación Sobre El Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Querétaro, México; ⁴Program in Biology, Division of Science and Mathematics, New York University Abu Dhabi, Abu Dhabi, UAE; ⁵Public Health Research Center, New York University Abu Dhabi, Abu Dhabi, UAE; ⁶Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, UAE; ⁷Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden; ⁸New York Genome Center, New York, NY, USA; ⁹Melbourne Integrative Genomics and School of BioSciences, University of Melbourne, Parkville, VIC, Australia; ¹⁰Center for Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu, Tartu, Estonia

¹¹Present address: Human Genomics and Evolution Group, St Vincent's Institute of Medical Research, Fitzroy, VIC, Australia

*Correspondence: sophia.george@med.miami.edu (S.H.L.G.), tuuli.lappalainen@scilifelab.se (T.L.)

<https://doi.org/10.1016/j.ajhg.2023.10.012>

© 2023 American Society of Human Genetics.

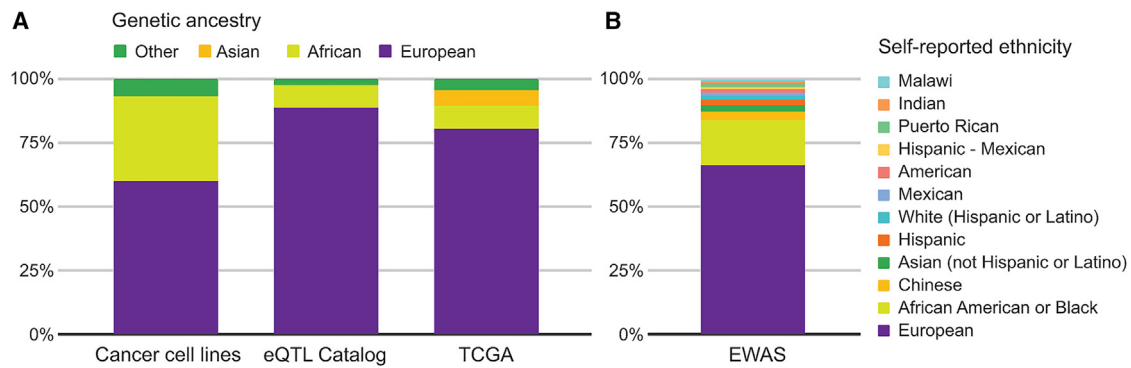


Figure 1. Donor diversity in different functional genomics datasets and resources (A and B) Genetic ancestry in a set of commonly used cancer cell lines,⁷ the eQTL Catalog,⁸ and TCGA⁹ are shown in (A), with (B) showing self-identified ethnicity across a variety of epigenome-wide association studies (EWASs).¹⁰ In addition to overrepresentation of samples of European ancestry and ethnicity, these data highlight the complexity of capturing the multiple dimensions of genetic ancestries, self-identities, and donor origins—information that is typically not collected and/or poorly reported in functional genomics data.

humans. This significantly limits our collective ability to move beyond fundamental discovery to equitable clinical outcomes, as discussed further below.^{5,14}

Importance of diversity in molecular studies

Two of the National Human Genome Research Institute’s ten “bold predictions” for 2030 for improving human health at the forefront of genomics are centered on diversity (genomic and workforce). The bold predictions also have a goal to go beyond historic social constructs such as race.¹⁵ However, lack of ancestral and environmental diversity in studies of molecular variation creates multiple roadblocks and challenges for GWAS discovery, interpretation, follow-up, and applications in precision medicine and predictive health (Figure 2). Firstly, most studies of molecular variation are performed in individuals living in urban or suburban environments that represent a narrow range of the environmental conditions in which humans live. Consequently, these studies may miss disease-relevant physiological variation and its molecular underpinnings, including host-pathogen interactions, gene-environment interactions in more extreme environments, societal stressors, and other factors.^{16–19} These molecular processes can include those also involved in disease processes, including chronic diseases that are not directly and immediately linked to environmental exposure, such as the role of infections in later autoimmune disease. With increasingly compelling evidence of the importance of such environmental exposures to human disease and their interactions with genetic risk,^{20–22} this represents a major gap in molecular studies and a missed opportunity in understanding diverse molecular triggers and risk factors of disease. This is not in disagreement with recent research on large cohorts in HICs suggesting that GWAS effect sizes are typically robust across ancestries,²³ as environmental effects may be quite specific.

Secondly, one of the primary tools for characterizing potential molecular effects of noncoding GWAS loci is mapping their associations to molecular phenotypes (molecular quantitative trait locus [molQTL] mapping).²⁴ While major molQTL resources have been assembled especially for expression and splicing QTLs (eQTLs and sQTLs, respectively), these are dominated by European ancestries and largely lack variants and haplotype patterns that are specific to other ancestries.⁸ Thus, a simple *in silico* molQTL query is less fruitful for non-European GWAS loci, and the variant-to-function gap is thus wider for non-European GWAS loci due to the lack of data. This hampers our ability to use the growing GWAS datasets to learn about disease biology and to identify potential targets for drugs and other interventions. Furthermore, increased diversity of molQTL studies helps to break down linkage disequilibrium blocks, leading to better fine-mapping of likely causal regulatory variants.^{25,26} This is key to experimental follow-up, as discussed below, and may also help to increase trans-ancestry portability of polygenic risk scores.²⁷

Studies of severe disease caused by rare variants—often with immediate clinical genomics applications—rely on large genetic databases of variant frequencies, and biases caused by lack of diversity are well recognized.²⁸ Emerging multi-omic approaches for clinical genomics aim to use not only genomic data but also transcriptome, epigenome, and other data to support identification of disease-causing rare variants.²⁹ These approaches are typically based on identifying population outliers in molecular profiles, such as situations where an individual has an aberrant splicing pattern in a gene where they carry a rare intronic variant. Lack of genetic variation data and molecular patterns that are an appropriate reference for individuals can give rise to false-positive diagnoses in individuals of non-European ancestries that may erroneously appear as being outside the general population spectrum. However, effects of population ancestry on molecular states and functions of cells should not be overstated: most of

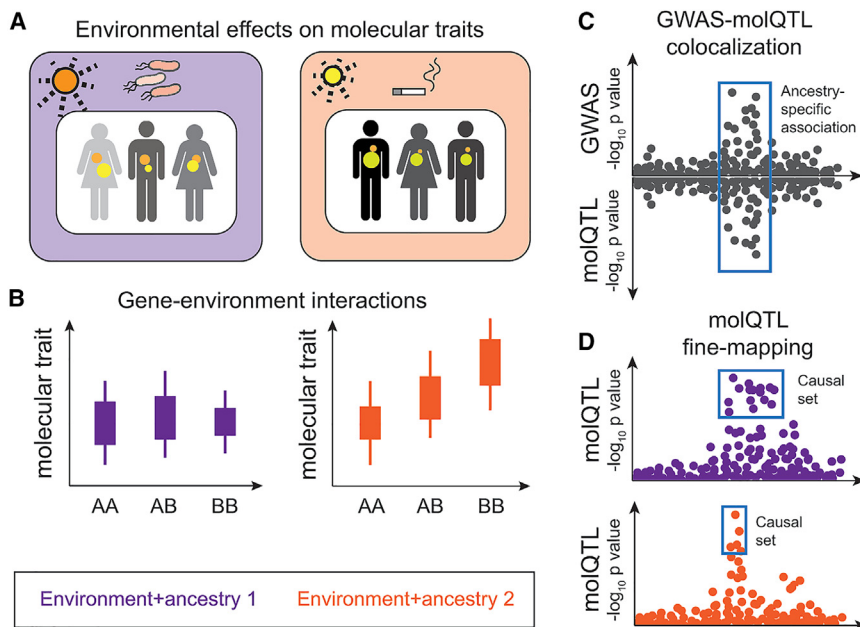


Figure 2. Illustration of some of the advantages of diversity in functional genomics studies

(A) Capturing molecular traits and processes in tissues and cells (colored dots) that are associated to specific environmental exposures.

(B) Detecting gene-environment or gene-gene interactions.

(C) Enabling molecular QTL-GWAS colocalization including in loci that are present in only specific ancestries or environmental contexts.

(D) Improving fine-mapping by leveraging different linkage disequilibrium (LD) patterns in different ancestries, with the lower population having less LD in this locus and thus a smaller set of putative causal variants.

Barriers and ways to fix them

Unlike DNA samples, which can be readily collected from blood or saliva and withstand substantial processing

delays before nucleic acid isolation without impacting data quality, generation of high-quality functional genomics data is not just costly but also requires sample collection in a controlled environment, rapid initial processing, and special sample shipping conditions.³⁵ This presents significant logistical challenges when working in low- and middle-income countries (LMICs), both in urban and non-urban environments, where facilities can be limited or under-resourced.^{36–38} Establishing local capacity for cost-effective functional genomic research and developing cohort studies and biobanks that actively engage local research communities, with the support of government, is critical in this regard.

Obtaining informed donor consent for biospecimen sampling without prior community engagement and buy-in is challenging and can be further complicated by a lack of trust in scientists. Varying cultural attitudes surrounding the handling of bodies after death, and toward invasive or postmortem biospecimen sampling introduce additional layers of complexity.^{39–41} It is important to have meaningful community involvement and engagement with ethical, legal, and social issues (ELSI) facilitators, which are fundamental to success in the long term (<https://www.inegi.org.mx/programas/enpecyt/2017/>). Ideally, this should be viewed not as an imposition, but as an opportunity for sustained inclusion and participation of community members at all stages of the project, including collaboration with local researchers and capacity building. Further opportunities for inclusion can be gained from innovation in less invasive biospecimen collection and study designs that accommodate different cultural contexts.⁴²

Once consent has been secured and samples collected, the labile nature of RNA and other biomolecules of interest and their large dynamic range introduces additional

the biology at the molecular, cellular, tissue, and physiological levels is shared among all humans (and often also among many species). Furthermore, similarly to the partitioning of genetic variation,^{30,31} the majority of inter-individual variation in molecular traits is not explained by population or continental groupings.³² Thus, the lack of diversity and/or representativeness does not invalidate basic molecular biology discoveries, and the reasonable default assumption for most molecular processes is that they are broadly shared. Yet, in studies that aim to characterize basic molecular biology and typically analyze up to a handful of individuals as biological replicates, diverse sampling not only prevents biased assignments of what constitutes an “archetype” of humans but can also lead to more robust discoveries. When our attention is focused on *variation* between individuals, sample selection that includes genetic and environmental variation is key for the reasons described above, and the study of diverse populations can provide valuable insights into population-specific genetic risk factors. While the extent of variation in GWAS effect sizes due to the individual’s genetic and environmental background, and driven by epistatic and gene-environment interactions, is debated,^{23,33,34} it appears that effect-size heterogeneity is the exception rather than the rule. This offers promise for integration and portability of results across populations with careful statistical methods and empirical data from diverse populations, in a manner that does not necessarily require extremely large datasets from all populations, which would be a daunting challenge. Nevertheless, this recommendation should not be taken as justification to limit resources to increase sample sizes in studies focused on underrepresented populations, as samples from HICs do continue to be extensively analyzed.

difficulties. Small differences in sample processing across sites or handlers can introduce unwanted variability in datasets, making downstream integration and harmonization complex and time intensive. However, it is often impossible to process all samples in a central location. Beyond logistics, some countries have strict export bans on human biological samples because of past bio-colonialist, extractive research carried out by scientists based in HICs. Rather than a reason for exclusion of participants from LMICs, this should be seen as an opportunity to increase equity not only in sample collections but in the scientific teams generating them. In the short term, some consortia are making use of cross-site controls to aid in correcting batch effects, although even with these in place the analytical challenges are not to be underestimated. Complexity increases even more when the aim is to generate multiple modalities of molecular data, such as simultaneous assaying of gene expression and chromatin structure, or to integrate molecular data with other orthogonal types such as health records, which are not always interoperable across countries or health systems and, in any case, are much less likely to be digitized outside of HICs. Some groups have developed standards and then harmonized survey instruments across diverse LMICs and HICs to allow further metadata integration.⁴³

Data generation and infrastructure building in LMICs is also hindered by higher costs for sequencing services and reagents, which due to lower demand are sometimes artificially inflated, a vicious cycle that hinders global scientific advancement.^{44,45} While international collaboration can be key for overcoming these challenges, these collaborations can be easily unbalanced in power. It is not uncommon for scientists in HICs to establish collaborations to access samples in LMICs without including as equal partners the scientists that collect the samples and that have been working with the local communities, excluding LMIC scientists from data analysis and discovery research stages. This is noticeable in the authorship positions of the articles resulting from these collaborations.⁴⁶ Not including the perspective of LMIC scientists along the research process also hampers scientific development, as these perspectives could open new avenues for discovery. In this context, it is also important to highlight the tension between current discipline expectations for fully open and unrestricted data access and the growing recognition of Indigenous and national data sovereignty, e.g., as exemplified by the CARE principles.^{47,48} Here, the possibility of federating both data storage and analyses⁴⁹ such that participants retain ongoing ownership of their data emerges as a solution worth exploring further. Additionally, peer review bias toward research developed in HICs^{50–53} slows down publication of work produced in LMICs and affects the availability of novel genomic data generated in these studies. Some journals, such as *eLife* and *F1000*, have proposed changes in the peer review process to increase accountability on the reviewers' side to reduce unconscious bias and improve scientific input, but analyses

remain to be done to probe if this strategy is yielding the desired results.

Governments and funders, too, have significant roles to play in bringing about change. Science spending in LMICs is generally low (https://www.oecd-ilibrary.org/science-and-technology/main-science-and-technology-indicators_2304277x), which hampers national capability to generate functional genomic data as grant budgets are limited. Fundamental science is not always a funding priority, especially considering more urgent medical needs.^{54–56} Few international funding agencies offer funding for research groups based in LMICs, especially for discovery-based research, and too often there is limited support for the preliminary outreach, community engagement, and logistic pilot studies that by necessity precede full-scale studies, because their small sample sizes are seen as uninformative. As exceptions to these trends, we highlight the Chan Zuckerberg Initiative's Ancestry Networks for the Human Cell Atlas, which is supporting the generation of large single-cell-resolution datasets from countries across Africa, Asia, the Caribbean, and Latin America to incorporate into the Human Cell Atlas, as well as local outreach and community engagement, alongside the Wellcome Trust's long-standing record of directly supporting fundamental research in LMICs. Nevertheless, these exceptional examples are not enough to shift the current tendency, and systematic funding opportunities in HIC funding agencies (private and governmental) should be made open to LMIC scientists. It is understandable that HIC funding agencies prioritize studies focused on their own populations; however, resource limitations in LMICs are, in many cases, a consequence of a colonial history. Making funding opportunities available could be a way of restorative justice to support research development.

We need funders in HICs to recognize the colonial debt with LMICs and allocate funds to support research development. Because this work can take many years to lead to results, it is also worth taking advantage of complementary approaches that can bear more immediate fruit. Amongst these we highlight the nascent potential of high-throughput multiplex assays of variant effect (MAVEs) or CRISPR-based perturbation screens to test the functional potential of tens of thousands of variants simultaneously.^{57,58} Adopting such approaches can aid in fine-mapping molQTLs or GWAS loci and in dissecting their molecular mechanisms regardless of the population where the variants and/or associations were originally detected. It is also possible to expose multiple cell lines to a defined set of external stimuli (e.g., pathogens, chemicals) and identify molQTLs associated with inter-individual response differences.⁵⁹ However, except for lymphoblastoid cell lines from the 1000 Genomes Project,⁶⁰ most public cell line collections are also highly biased toward European ancestries, perpetuating systematic exclusions of other populations.

Conclusions

Global challenges require global support. To increase our ability to improve and map genetic variation to health

and disease, it is essential that we diversify across multiple variables, including genuine inclusion of LMICs. Beyond genomic ancestry, variables such as age, sex, environment, lifestyle, ethnicity, and country of origin, and the interactions between them, are important factors to incorporate into study designs to understand the influence of and interplay between sociology, geography, and biology. Inclusion of participants and scientists from diverse communities and the Global South will increase feasibility and improve studies targeting these populations to prevent helicopter science.⁶¹ In women, racial and ethnic minorities, and other underrepresented groups, we can disentangle and measure the influence of societal factors like social status (environment) on disease etiology and pathogenesis. The contribution of genomic ancestry is a modifier of the experiential, cultural factors that increase risk to disease and likely influence health. While there is much to be done to genuinely have an impact in diversifying functional genomics data generation, there is a major opportunity on multiple fronts to advance scientific discovery and global equity.

Acknowledgments

We would like to thank Charles Breeze for providing EWAS statistics of Figure 1B for re-plotting purposes. S.G. has grants from Chan Zuckerberg Initiative Ancestry Network (DAF2021-240624) and CDMRP Ovarian Cancer program (W81XWH-18-1-0072). Y.I. is supported by Chan Zuckerberg Initiative Ancestry Network grant DAF2022-239720. A.M.-R. is supported by Chan Zuckerberg Initiative Ancestry Network (2021-240438), CONACYT-FORDECYT-PRONACES grant numbers 11311 and 6390, and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica–Universidad Nacional Autónoma de México (PAPIIT-UNAM) grant numbers IA203021 and IN218023. I.G.R. is supported by a Chan Zuckerberg Initiative Ancestry Network grant, Australian National Health and Medical Research Council Ideas grant 2020501, the European Union through Horizon 2020 Research and Innovation Program under grant number 810645, and the European Regional Development Fund project number MOBEC008.

Declaration of interests

S.G. has grants from Pfizer Inc. T.L. is a paid advisor or consultant to GSK, Pfizer, and Goldfinch Bio and has equity in Variant Bio.

References

1. Fatumo, S., and Inouye, M. (2023). African genomes hold the key to accurate genetic risk prediction. *Nat. Human Behav.* 7, 295–296.
2. Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F., Li, L.; and China Kadoorie Biobank CKB collaborative group (2011). China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* 40, 1652–1666.
3. Lee, S., Jung, P.E., and Lee, Y. (2016). Publicly-funded biobanks and networks in East Asia. *SpringerPlus* 5, 1080.
4. Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., et al. (2017). Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* 27, S2–s8.
5. Mills, M.C., and Rahal, C. (2020). The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat. Genet.* 52, 242–243.
6. Breeze, C.E., Beck, S., Berndt, S.I., and Franceschini, N. (2022). The missing diversity in human epigenomic studies. *Nat. Genet.* 54, 737–739.
7. Hooker, S.E., Jr., Woods-Burnham, L., Bathina, M., Lloyd, S., Gorjala, P., Mitra, R., Nonn, L., Kimbro, K.S., and Kittles, R.A. (2019). Genetic Ancestry Analysis Reveals Misclassification of Commonly Used Cancer Cell Lines. *Cancer Epidemiol. Biomarkers Prev.* 28, 1003–1009.
8. Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J., et al. (2021). A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* 53, 1290–1299.
9. Yuan, J., Hu, Z., Mahal, B.A., Zhao, S.D., Kensler, K.H., Pi, J., Hu, X., Zhang, Y., Wang, Y., Jiang, J., et al. (2018). Integrated Analysis of Genetic Ancestry and Genomic Alterations across Cancers. *Cancer Cell* 34, 549–560.e9.
10. Breeze, C.E., Wong, J.Y.Y., Beck, S., Berndt, S.I., and Franceschini, N. (2022). Diversity in EWAS: current state, challenges, and solutions. *Genome Med.* 14, 71.
11. Stark, Z., Dolman, L., Manolio, T.A., Ozenberger, B., Hill, S.L., Caulfield, M.J., Levy, Y., Glazer, D., Wilson, J., Lawler, M., et al. (2019). Integrating Genomics into Healthcare: A Global Responsibility. *Am. J. Hum. Genet.* 104, 13–20.
12. Lazareva, T.E., Barbitoff, Y.A., Changelidis, A.I., Tkachenko, A.A., Maksiutenko, E.M., Nasykhova, Y.A., and Glotov, A.S. (2022). Biobanking as a Tool for Genomic Research: From Allele Frequencies to Cross-Ancestry Association Studies. *J. Personalized Med.* 12, 2040.
13. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177, 1080–1131.
14. Ruan, Y., Lin, Y.F., Feng, Y.C.A., Chen, C.Y., Lam, M., Guo, Z., Stanley Global Asia Initiatives, He, L., Sawa, A., Martin, A.R., et al. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* 54, 573–580.
15. Green, E.D., Gunter, C., Biesecker, L.G., Di Francesco, V., Easter, C.L., Feingold, E.A., Felsenfeld, A.L., Kaufman, D.J., Ostrander, E.A., Pavan, W.J., et al. (2020). Strategic vision for improving human health at The Forefront of Genomics. *Nature* 586, 683–692.
16. Idaghdour, Y., Czika, W., Shianna, K.V., Lee, S.H., Visscher, P.M., Martin, H.C., Miclaus, K., Jadallah, S.J., Goldstein, D.B., Wolfinger, R.D., and Gibson, G. (2010). Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat. Genet.* 42, 62–67.
17. Favé, M.J., Lamaze, F.C., Soave, D., Hodgkinson, A., Gauvin, H., Bruat, V., Grenier, J.-C., Gbeha, E., Skead, K., Smargiassi, A., et al. (2018). Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat. Commun.* 9, 827.
18. Lea, A., Subramaniam, M., Ko, A., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Mononen, N., Raitakari, O.T., Ala-Korpela, M., et al. (2019). Genetic and environmental perturbations lead to regulatory decoherence. *Elife* 8, e40538.
19. Temba, G.S., Kullaya, V., Pecht, T., Mmbaga, B.T., Aschenbrenner, A.C., Ulas, T., Kibiki, G., Lyamuya, F., Boahen, C.K., Kumar, V., et al. (2021). Urban living in healthy Tanzanians is

- associated with an inflammatory status driven by dietary and metabolic changes. *Nat. Immunol.* 22, 287–300.
20. Laville, V., Majarian, T., Sung, Y.J., Schwander, K., Feitosa, M.F., Chasman, D.I., Bentley, A.R., Rotimi, C.N., Cupples, L.A., de Vries, P.S., et al. (2022). Gene-lifestyle interactions in the genomics of human complex traits. *Eur. J. Hum. Genet.* 30, 730–739.
 21. Virolainen, S.J., VonHandorf, A., Viel, K.C.M.F., Weirauch, M.T., and Kottyan, L.C. (2023). Gene–environment interactions and their impact on human health. *Gene Immun.* 24, 1–11.
 22. Wu, H., Eckhardt, C.M., and Baccarelli, A.A. (2023). Molecular mechanisms of environmental exposures and human disease. *Nat. Rev. Genet.* 24, 332–344.
 23. Hu, S., Ferreira, L.A.F., Shi, S., Hellenthal, G., Marchini, J., Lawson, D.J., and Myers, S.R. (2023). Leveraging fine-scale population structure reveals conservation in genetic effect sizes between human populations across a range of human phenotypes. Preprint at bioRxiv.
 24. Aguet, F., Alasoo, K., Li, Y.I., Battle, A., Im, H.K., Montgomery, S.B., and Lappalainen, T. (2023). Molecular quantitative trait loci. *Nat. Rev. Methods Primers* 3, 4.
 25. Vandiedonck, C. (2018). Genetic association of molecular traits: A help to identify causative variants in complex diseases. *Clin. Genet.* 93, 520–532.
 26. Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* 1, 59.
 27. Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., Koido, M., Dey, K.K., Matsuda, K., Murakami, Y., Price, A.L., Kawakami, E., et al. (2020). Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat. Genet.* 52, 1346–1354.
 28. George, S.H.L., Donenberg, T., Alexis, C., DeGennaro, V., Jr., Dyer, H., Yin, S., Ali, J., Butler, R., Chin, S.N., Curling, D., et al. (2021). Gene Sequencing for Pathogenic Variants Among Adults With Breast and Ovarian Cancer in the Caribbean. *JAMA Netw. Open* 4, e210307.
 29. Wang, R.-S., Maron, B.A., and Loscalzo, J. (2023). Multiomics Network Medicine Approaches to Precision Medicine and Therapeutics in Cardiovascular Diseases. *Arterioscler. Thromb. Vasc. Biol.* 43, 493–503.
 30. Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J.K., and Feldman, M.W. (2005). Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure. *PLoS Genet.* 1, e70.
 31. Rosenberg, N.A. (2006). Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann. Hum. Genet.* 70, 841–847.
 32. Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M., et al. (2012). Patterns of Cis Regulatory Variation in Diverse Human Populations. *PLoS Genet.* 8, e1002639.
 33. Abdellaoui, A., Dolan, C.V., Verweij, K.J.H., and Nivard, M.G. (2022). Gene–environment correlations across geographic regions affect genome-wide association studies. *Nat. Genet.* 54, 1345–1354.
 34. Patel, R.A., Musharoff, S.A., Spence, J.P., Pimentel, H., Tcheandjieu, C., Mostafavi, H., Sinnott-Armstrong, N., Clarke, S.L., Smith, C.J., et al.; VA Million Veteran Program (2022). Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. *Am. J. Hum. Genet.* 109, 1286–1297.
 35. Gallego Romero, I., Pai, A.A., Tung, J., and Gilad, Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* 12, 42.
 36. Abdrabou, W., Dieng, M.M., Diawara, A., Sermé, S.S., Almojil, D., Sombié, S., Henry, N.B., Kargougou, D., Manikandan, V., Soulama, I., and Idaghdour, Y. (2021). Metabolome modulation of the host adaptive immunity in human malaria. *Nat. Metab.* 3, 1001–1016.
 37. Suwalowska, H. (2022). The invisible body work of 'last responders' - ethical and social issues faced by the pathologists in the Global South. *Global Publ. Health* 17, 4183–4194.
 38. Lea, A.J., Garcia, A., Arevalo, J., Ayroles, J.F., Buetow, K., Cole, S.W., Eid Rodriguez, D., Gutierrez, M., Highland, H.M., Hooper, P.L., et al. (2023). Natural selection of immune and metabolic genes associated with health in two lowland Bolivian populations. *Proc. Natl. Acad. Sci. USA* 120, e2207544120.
 39. Ben Taher, M., Pearson, J., Cohen, M., and Offiah, A.C. (2018). Acceptability of post-mortem imaging among Muslim and non-Muslim communities. *Br. J. Radiol.* 91, 20180295.
 40. Das, M.K., Arora, N.K., Debata, P., Chellani, H., Rasaily, R., Gaikwad, H., Meena, K.R., Kaur, G., Malik, P., Joshi, S., and Kumari, M. (2021). Why parents agree or disagree for minimally invasive tissue sampling (MITS) to identify causes of death in under-five children and stillbirth in North India: a qualitative study. *BMC Pediatr.* 21, 513.
 41. Selket, K., Glover, M., and Palmer, S. (2015). Normalising post-mortems – whose cultural imperative? An indigenous view on New Zealand post-mortem policy. *Kotuitui* 10, 1–9.
 42. Martorella, M., Kasela, S., Garcia-Flores, R., Gokden, A., Castel, S.E., and Lappalainen, T. (2023). Evaluation of noninvasive biospecimens for transcriptome studies. Preprint at bioRxiv.
 43. Odedina, F.T., Ragin, C., Martin, D., Moser, R., J., O., McDonald, A., Rice, E.L., Nguyen, J., Chingwundoh, F., Morisson Blidgen, B., et al. (2019). Standardized Global Behavioral and Epidemiological Measures for Prostate Cancer Studies in Black Men (Cancer Health Disparities), pp. e1–e16.
 44. Ciocca, D.R., and Delgado, G. (2017). The reality of scientific research in Latin America; an insider's perspective. *Cell Stress Chaperones* 22, 847–852.
 45. Olufadewa, I.I., Adesina, M.A., and Ayorinde, T. (2020). From Africa to the World: Reimagining Africa's research capacity and culture in the global knowledge economy. *J. Glob. Health* 10, 010321.
 46. Odeny, B., and Bosurgi, R. (2022). Time to end parachute science. *PLoS Med.* 19, e1004099.
 47. Jennings, L., Anderson, T., Martinez, A., Sterling, R., Chavez, D.D., Garba, I., Hudson, M., Garrison, N.A., and Carroll, S.R. (2023). Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. *Nat. Ecol. Evol.* 7, 1547–1551.
 48. Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., et al. (2020). The CARE Principles for Indigenous Data Governance. *Data Sci. J.* 19, 12.
 49. Boscarino, N., Cartwright, R.A., Fox, K., and Tsosie, K.S. (2022). Federated learning and Indigenous genomic data sovereignty. *Nat. Mach. Intell.* 4, 909–911.

50. Harris, M., Macinko, J., Jimenez, G., and Mullachery, P. (2017). Measuring the bias against low-income country research: an Implicit Association Test. *Glob. Health* 13, 80.
51. Harris, M., Marti, J., Watt, H., Bhatti, Y., Macinko, J., and Darzi, A.W. (2017). Explicit Bias Toward High-Income-Country Research: A Randomized, Blinded, Crossover Experiment Of English Clinicians. *Health Aff.* 36, 1997–2004.
52. Fox, C.W., Meyer, J., and Aimé, E. (2023). Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology* 37, 1144–1157.
53. Liu, F., Rahwan, T., and AlShebli, B. (2023). Non-White scientists appear on fewer editorial boards, spend more time under review, and receive fewer citations. *Proc. Natl. Acad. Sci. USA* 120, e2215324120.
54. Pramesh, C.S., Badwe, R.A., Bhoo-Pathy, N., Booth, C.M., Chinnaswamy, G., Dare, A.J., de Andrade, V.P., Hunter, D.J., Gopal, S., Gospodarowicz, M., et al. (2022). Priorities for cancer research in low- and middle-income countries: a global perspective. *Nat. Med.* 28, 649–657.
55. Pramesh, C.S., Venkataramanan, R., Suvarna, V., Goel, N.S., Lakshman, S., Venkatesh, V., Gupta, V., and Badwe, R. (2016). Involvement of general public in biomedical research. *Perspect. Clin. Res.* 7, 152–155.
56. Charani, E., Abimbola, S., Pai, M., Adeyi, O., Mendelson, M., Laxminarayan, R., and Rasheed, M.A. (2022). Funders: The missing link in equitable global health research? *PLOS Glob. Public Health* 2, e0000583.
57. Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., and Fowler, D.M. (2017). Variant Interpretation: Functional Assays to the Rescue. *Am. J. Hum. Genet.* 101, 315–325.
58. Morris, J.A., Caragine, C., Daniloski, Z., Domingo, J., Barry, T., Lu, L., Davis, K., Ziosi, M., Glinos, D.A., Hao, S., et al. (2023). Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science* 380, eadh7699.
59. Findley, A.S., Monziani, A., Richards, A.L., Rhodes, K., Ward, M.C., Kalita, C.A., Alazizi, A., Pazokitoroudi, A., Sankararaman, S., Wen, X., et al. (2021). Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *Elife* 10, e67077.
60. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science* 352, 600–604.
61. Atutornu, J., Milne, R., Costa, A., Patch, C., and Middleton, A. (2022). Towards equitable and trustworthy genomics research. *EBioMedicine* 76, 103879.