

<https://doi.org/10.1038/s43856-024-00647-z>

Inclusiveness of the All of Us Research Program improves polygenic risk scores and fosters genomic medicine for all

Benson R. Kidenya & Gerald Mboowa

Check for updates

Polygenic risk scores (PRS) show promise but have accuracy disparities across ancestries due to underrepresentation in the existing genomic databases. Here, we outline the initiative of All of Us Research Program in refining PRS and advancing genomic medicine for all.

The nature of genomic data and its use in medicine

For years, clinicians and researchers have diligently pursued precise predictions of genetic diseases and traits through the analysis of genetic data. Early identification of genetic predisposition for complex diseases allows for timely interventions, potentially averting adverse outcomes. While the genetics of many Mendelian diseases is well understood, predicting genetic susceptibility for complex diseases such as diabetes mellitus, stroke, schizophrenia, cancers, cardiovascular diseases, and others remains challenging. This is in part due to the cumulative effects of numerous genetic variants, each with minor individual impacts on disease development¹.

Over the past two decades, studies to investigate the multiple genetic variations in humans such as genome-wide association studies (GWAS) have revealed polygenic variants linked to common complex disorders and certain traits². While individual genetic variants' contributions to disease risk or traits may seem modest, this insight has facilitated the development of polygenic risk scores (PRS). These scores aggregate a weighted lifetime sum of independent and significantly associated genetic variants or single nucleotide polymorphisms³. They aim to predict an individual's susceptibility to acquiring a genetic disease or propensity towards a specific trait. PRS have found a widespread application in research, particularly in elucidating associations between scores, and disease status or traits. However, their clinical utility remains largely unestablished and constrained⁴. This is attributed in part to the insufficient inclusiveness of the existing genomic databases, leading to gaps in data representation across different populations.

PRS are developed to forecast, and thus improve, health outcomes through genomic medicine. This encompasses predicting disease risk, traits, treatment outcomes, and disease prognosis. Comprising hundreds to thousands of genetic variants, PRS are constructed from a compilation of independent risk genetic variants linked with a disease. These individual variants are derived from the latest evidence from the most expansive or informative GWAS. This compilation yields a single score representing each individual's genetic predisposition for a disease or continuous trait. However, a primary ethical and scientific obstacle in the clinical integration of PRS is the significant discrepancy in accuracy across ancestries. Presently available PRS are notably more precise in individuals of European descent

compared to other ancestries. This disparity stems from inherent Eurocentric biases in existing GWAS as the existing GWAS are mostly from white European populations, underscoring that the current clinical utilization of PRS predominantly benefits populations of European descent. Analyses indicate markedly lower accuracy of PRS among non-European populations such as African and Hispanic populations, posing a substantial challenge to equitable genomic medicine⁵.

This underrepresentation in genomic data, coupled with the high diversity of genomes and short blocks of genetic variants that are non randomly inherited together (linkage disequilibrium), particularly among African populations^{6,7}, contributes to the challenge. In essence, using Eurocentric genomic datasets for PRS training and development leads to less accurate PRS for underrepresented populations. To fully realize the equitable potential of PRS, prioritizing greater diversity in genetic studies is essential. To bring this to realization, there should be a concerted effort among researchers, funders, and hosts of genomic databases. Additionally, public dissemination of summary statistics from all ancestries is crucial among authors/researchers and journals to prevent exacerbating health disparities among the most underserved individuals. This will facilitate the construction of PRS with improved accuracy in prediction of complex human diseases and traits.

Making genomic data more equitable by addressing existing disparities

In an effort to enhance healthcare by prioritizing the genetic and health data of historically marginalized populations, the National Institute of Health (NIH) in the United States recently established the All of Us Research Program^{8,9}. With more than \$3.1 billion in funding from the NIH, this initiative aims to compile detailed health profiles for one million diverse individuals within the US by 2026, thereby bridging existing gaps. From its inception in 2018¹⁰ to April 2023, the program has enrolled 413,000 participants, with 46% belonging to minority racial or ethnic groups. Impressively, it has shared nearly 250,000 genomes, comprising the most extensive assembly of African American, Hispanic, and Latin-American genomes to date.

Data collected in the All of Us Research Program consists of whole-genome sequences, health records, and surveys, with intention to not only compile GWAS data, but also to provide insight into health across diverse ancestries, and levels of access to healthcare. This ambitious endeavor currently ranks as one of the largest and most accessible biomedical research repositories worldwide. In addition to the genomic data and participant surveys, electronic health records and data from wearable devices, such as Fitbits, have also been included to enhance its utility as a comprehensive genomic resource. Meaningful contributions to our understanding of genetic risk are already being realized from this database. Primary analyses of up to 245,000 diverse genomes from the All of Us Research Program have

revealed over 275 million new genetic variants linked to a range of complex diseases, including nearly 150 potentially linked to type 2 diabetes mellitus⁸. These results serve as a demonstration of existing disparities in genetics research regarding non-white populations, as novel pathogenic variants are discovered in these diverse populations that have not already been identified in European populations. Additionally, new genetic information gathered in the All of Us Research Program shows that there are fewer people with pathogenic genetic variants and more with previously unknown variants¹¹.

The dataset is freely accessible upon reasonable request, facilitating its sharing and enabling the recalibration and development of new PRS to enhance accuracy. In essence, the All of Us Research Program represents a pivotal step towards leveraging genomic diversity to foster inclusive genomic medicine and improve PRS accuracy, thereby advancing healthcare for all⁸.

The All of Us Program's diverse genomic dataset is poised to revolutionize the development of PRS with heightened accuracy, while also facilitating the refinement of existing PRS initially constructed using Eurocentric genomic data. Consequently, the collection and utilization of additional genomic and health data from varied populations will be essential for generating precise PRS that offer an accurate assessment of an individual's genetic susceptibility to developing a disease⁴. The diverse genomic dataset from the All of Us Research Program has been utilized to create and validate PRS customized for enhanced performance in clinical settings⁴. Past research has revealed that these scores, soon to be integrated into clinical practice for personalized healthcare, are often less accurate for minority populations than for majority populations. However, recent studies have already leveraged the inclusive All of Us Research Program data to enhance and validate scores for various conditions, including coronary heart disease and diabetes mellitus^{12,13}.

This underscores the significance of the diverse genome dataset in updating and refining these PRS to enhance their accuracy for use in clinical practice. Nonetheless, a challenge persists in precisely interpreting these scores among clinicians. Upcoming research should concentrate on understanding how healthcare professionals interpret these scores, and how to apply them in treatment decisions. Currently, there is no specific African country widely recognized for actively using PRS in clinical settings as seen in the United States or United Kingdom. However, there are ongoing efforts and research initiatives across Africa to develop and utilize PRS, especially within the context of enhancing genomic data and precision medicine for African populations. One notable effort is the establishment of biobanks and genomic datasets in several African countries. For instance, the H3Africa initiative (Human Heredity and Health in Africa) is working to increase the understanding of how genomic and environmental factors influence disease in African populations, which could pave the way for future use of PRS¹. Additionally, there is growing recognition of the need for more inclusive genomic research that represents the genetic diversity of African populations to improve the accuracy and applicability of PRS in these regions¹⁴. Countries like South Africa and Nigeria are also part of international collaborations aiming to gather more comprehensive genomic data. These efforts are essential steps toward potentially implementing PRS in healthcare systems across the continent in the future¹. While the direct clinical use of PRS in Africa may still be in the developmental stages, these foundational efforts indicate a promising direction for the integration of genetic risk prediction in the continent's healthcare landscape.

Conclusion

Emerging data from the All of Us Research Program shows that there are fewer people with deleterious genetic variants and more with genetic variants we do not fully understand in groups who have been less well studied in the past. This stresses the need for more genetic research in these groups. Furthermore, the All of Us Research Program's diverse genomic data represents a pivotal platform for addressing genetic disparities and unlocking the full potential of optimized and accurate PRS. This, in turn, will advance the equitable application of genomic medicine and its tools, including PRS, heralding a new era of personalized healthcare for all.

Benson R. Kidenya ^{1,2} & **Gerald Mboowa** ^{3,4,5}

¹Department of Biochemistry and Molecular Biology, Weill Bugando School of Medicine, Catholic University of Health and Allied Sciences, Mwanza, Tanzania. ²Train-The-Trainers for Bioinformatics Community, Human Heredity and Health for Africa Bioinformatics Network (H3ABioNet), Cape Town, South Africa. ³Department of Immunology and Molecular Biology, College of Health Sciences, School of Biomedical Sciences, Makerere University, Kampala, Uganda. ⁴The African Center of Excellence in Bioinformatics and Data-Intensive Sciences, the Infectious Diseases Institute, College of Health Sciences, Makerere University, Kampala, Uganda. ⁵Africa Centres for Disease Control and Prevention, African Union Commission, Addis Ababa, Ethiopia.

✉ e-mail: benkidenya@gmail.com

Received: 21 March 2024; Accepted: 15 October 2024;

Published online: 08 November 2024

References

1. Fatumo, S. et al. Polygenic risk scores for disease risk prediction in Africa: current challenges and future directions. *Genome Med.* **15**, 87 (2023).
2. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Prim.* **1**, 1–21 (2021).
3. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
4. Lennon, N. J. et al. Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat. Med.* **30**, 480–487 (2024).
5. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* **12**, 44 (2020).
6. Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genom. Hum. Genet.* **9**, 403–433 (2008).
7. Lonjou, C. et al. Linkage disequilibrium in human populations. *Proc. Natl. Acad. Sci. USA* **100**, 6069–6074 (2003).
8. Bick, A. G. et al. Genomic data in the All of Us research program. *Nature* 1–7 <https://doi.org/10.1038/s41586-023-06957-x> (2024).
9. Kozlov, M. Ambitious survey of human diversity yields millions of undiscovered genetic variants. *Nature*. <https://doi.org/10.1038/d41586-024-00502-0> (2024).
10. All of Us Research Program Investigators. The “All of Us” research program. *N. Engl. J. Med.* **381**, 668–676 (2019).
11. Venner, E. et al. The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities. *Commun. Biol.* **7**, 1–11 (2024).
12. Mapes, B. M. et al. Diversity and inclusion for the All of Us research program: a scoping review. *PLoS ONE* **15**, e0234962 (2020).
13. Venner, E. et al. Whole-genome sequencing as an investigational device for return of hereditary disease risk and pharmacogenomic results as part of the All of Us research program. *Genome Med.* **14**, 34 (2022).
14. Slunecka, J. L. et al. Implementation and implications for polygenic risk scores in healthcare. *Hum. Genom.* **15**, 46 (2021).

Acknowledgements

We acknowledge Human Heredity and Health Africa Bioinformatics Network (H3ABioNet) for the training that helped us to conceptualize and execute this work.

Author contributions

B.R.K. conceptualized the study while discussing it with G.M. B.R.K. and G.M. gathered the literature and summarized the main findings. B.R.K. drafted the manuscript and B.R.K. and G.M. revised the manuscript critically.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Benson R. Kidenya.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024