Polygenic risk scores for disease risk prediction in Africa: current challenges and future directions

- Segun Fatumo,
- Dassen Sathan,
- Chaimae Samtal,
- Itunuoluwa Isewon,
- Tsaone Tamuhla,
- Chisom Soremekun,
- James Jafali,
- Sumir Panji,
- Nicki Tiffin &
- Yasmina Jaufeerally Fakim

-
- Explore all metrics

Abstract

Early identification of genetic risk factors for complex diseases can enable timely interventions and prevent serious outcomes, including mortality. While the genetics underlying many Mendelian diseases have been elucidated, it is harder to predict risk for complex diseases arising from the combined effects of many genetic variants with smaller individual effects on disease aetiology. Polygenic risk scores (PRS), which combine multiple contributing variants to predict disease risk, have the potential to influence the implementation for precision medicine. However, the majority of existing PRS were developed from European data with limited transferability to African populations. Notably, African populations have diverse genetic backgrounds, and a genomic architecture with smaller haplotype blocks compared to European genomes. Subsequently, growing evidence shows that using large-scale African ancestry cohorts as discovery for PRS development may generate more generalizable findings. Here, we (1) discuss the factors contributing to the poor transferability of PRS in African populations, (2) showcase the novel Africa genomic datasets for PRS development, (3) explore the potential clinical utility of PRS in African populations, and (4) provide insight into the future of PRS in Africa.

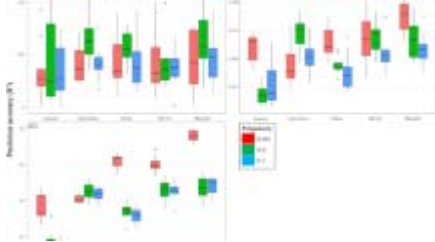**Similar content being viewed by others**
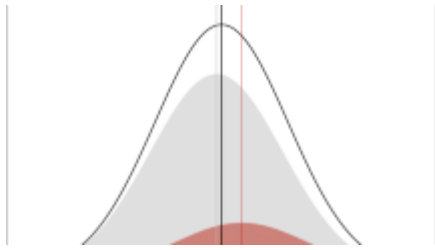
Polygenic risk scores: from research tools to clinical instruments

Article Open access 18 May 2020
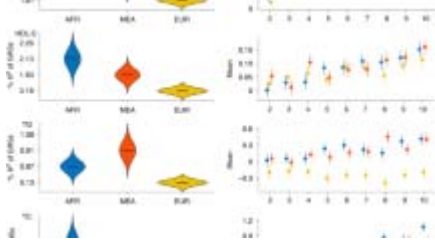
Transferability of genetic risk scores in African populations

Article Open access 02 June 2022

Evaluation of optimal methods and ancestries for calculating polygenic risk scores in East Asian population
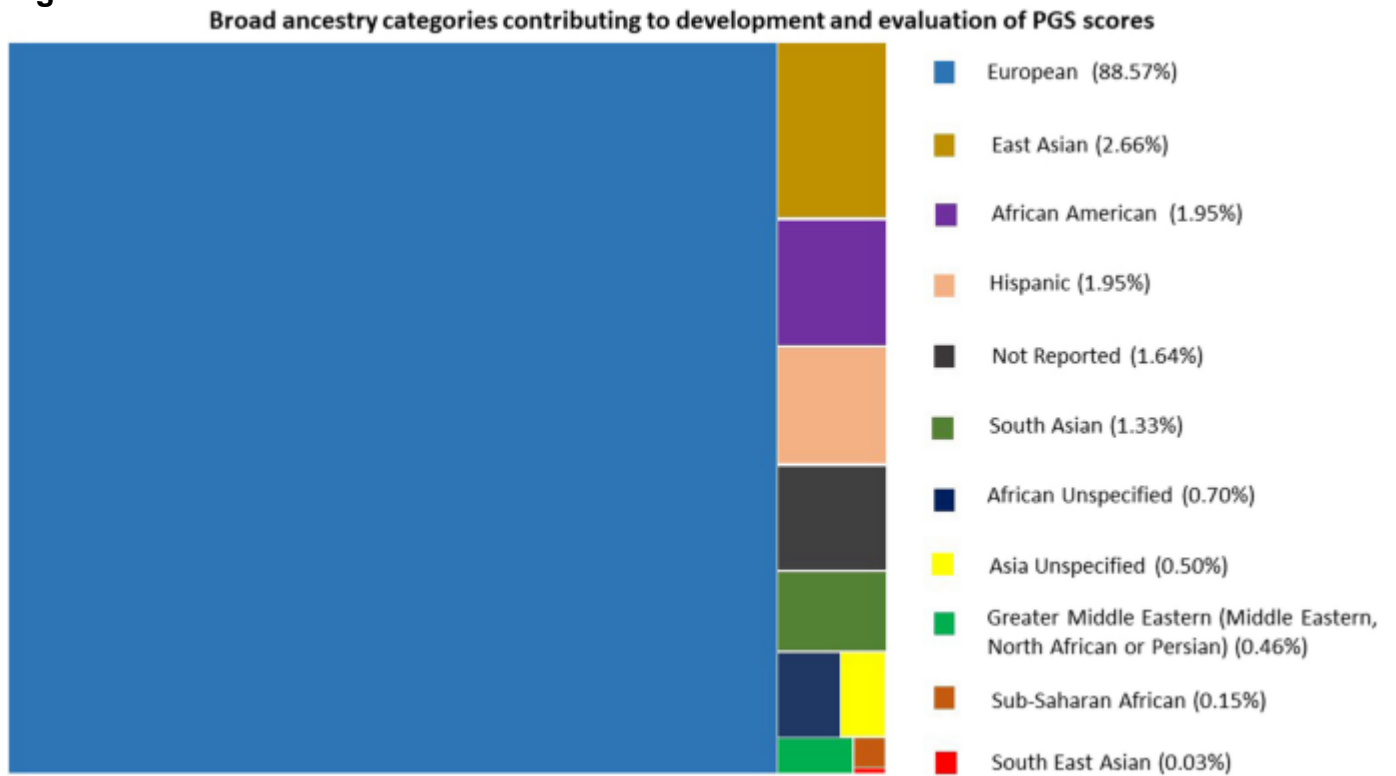
Article Open access 06 November 2023

Background

Disease prevalence varies largely across the world, and some diseases are often specific to certain geographic locations. Lifestyle, diet, and environmental determinants, as well as genetic factors, explain pathological conditions in diverse settings and are likely to impact on the severity in different individuals and populations (https://www.who.int/data/gho,18/10/2022). Clinical risk can be evaluated from the analysis of blood biomarkers, symptoms, and prevailing family history. However, recent work has suggested that risk prediction for common chronic diseases can be improved using genetic data [1].

Genome-wide association studies (GWAS) have significantly contributed to identifying a huge number of loci associated to a variety of complex diseases and traits. However, most genetic association discoveries have been made in European ancestry individuals [2,3,4,5]. The strength of the genetic association with phenotypes is enhanced when phenotypic data is available from

large-scale studies linked with relevant phenotypic data and electronic health records. Recently, polygenic risk scores (PRSs), which weigh the genetic effect of numerous common variations associated to disease or traits, have gained popularity to quantify an individual's genetic risk for a disease or trait. The pace of research in this area has recently improved, and PRS scores are now available for a variety of traits and diseases, mostly in the European population (Figs. 1 and 2). As a result, PRS is quickly becoming a common tool for estimating genetic liability in predicting disease risks, which is essential for early disease identification, prevention, and intervention.

Fig. 1

Broad ancestry categories contributing to development and evaluation of PGS scores

- European (88.57%)
- East Asian (2.66%)
- African American (1.95%)
- Hispanic (1.95%)
- Not Reported (1.64%)
- South Asian (1.33%)
- African Unspecified (0.70%)
- Asia Unspecified (0.50%)
- Greater Middle Eastern (Middle Eastern, North African or Persian) (0.46%)
- Sub-Saharan African (0.15%)
- South East Asian (0.03%)

Proportion of broad ancestry populations that contributed to the development and evaluation of 2555 PGS scores within the PGS Catalogue (version 2023–06-23). African unspecified: African that could not be classified as African American, Afro-Caribbean, or Sub-Saharan African. Asian unspecified: Asians that could not be classified as East Asian, Central Asian, South Asian, or South-East Asian. Multiple studies and subsets of data can contribute towards the development and evaluation of a PGS scores within the PGS Catalogue

Full size image

Fig. 2
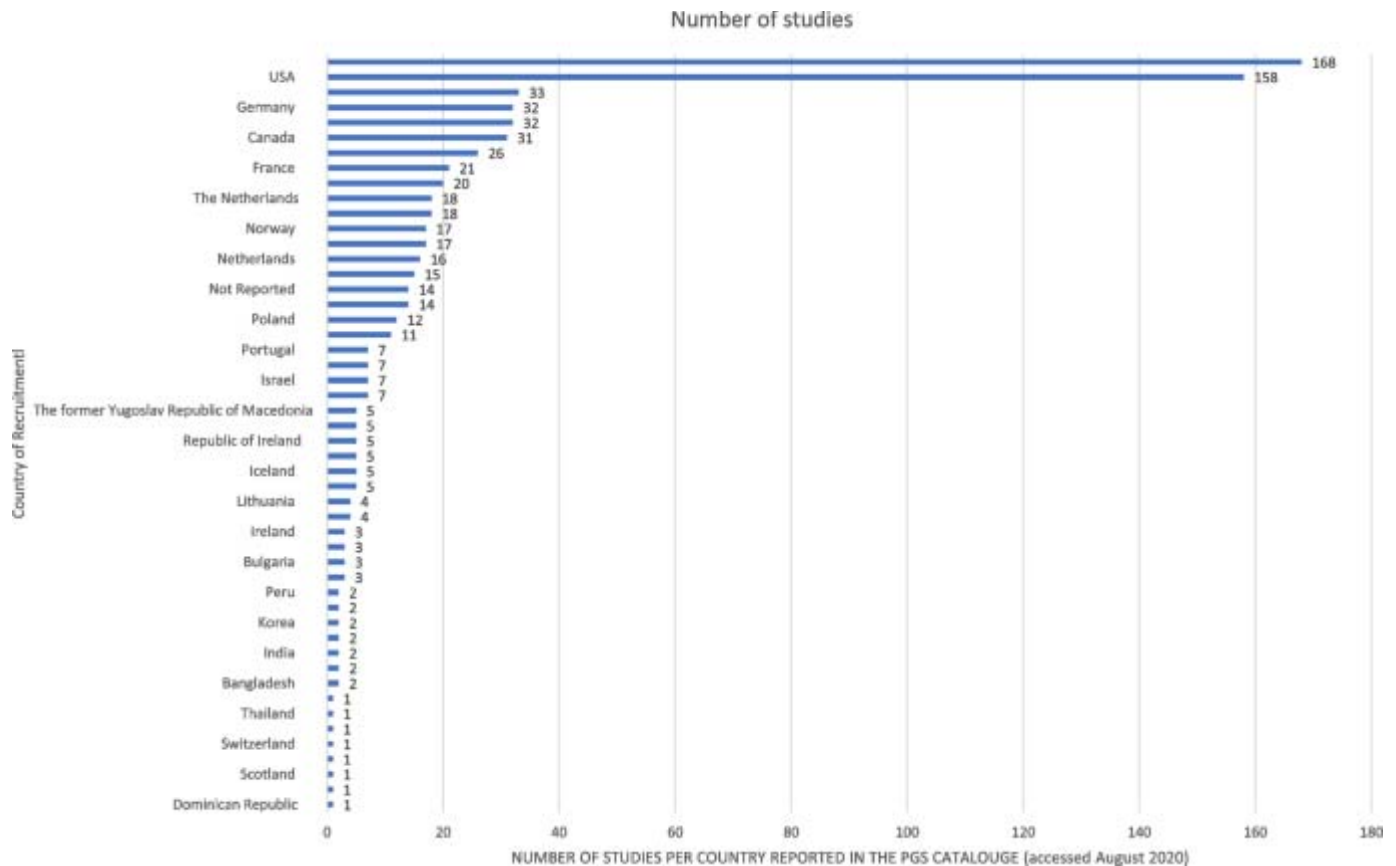
Count of reported countries of recruitment for samples included in the PGS Catalogue (accessed August 2020)

Full size image

The poor transferability of PRS derived from European ancestry dataset to diverse African populations is a cause of concern. This is likely to be due to unique differences in genetic architecture and environmental exposures of the different populations [4]. The lack of accurate PRS in African ancestry individuals may cause barrier to achieve precise risk stratification which is critical for precision medicine. Given that the human genetic diversity is greater in Africa, and when large-scale African ancestry cohorts are available for the development of PRS, this may generate more generalizable findings [6]. This is high importance, not only for Africa but for the entire global medical and research community. For example, the identification of *PCSK9* missense mutations and their impact on plasma low-density lipoprotein cholesterol levels across diverse ancestries. This breakthrough discovery exemplifies how African ancestry individuals have contributed to advancing medical knowledge, thereby benefiting the entire human race. The rich genetic variation in African populations provides so many opportunities that extends well beyond the scope of PRS. In this review article, the aim is to (1) review factors contributing to poor transferability of PRS in African populations, (2) showcase the novel genomic datasets that could enhance PRS transferability in continental Africa, and (3) explore the potential clinical utility of PRS in African populations.

Factors contributing to poor transferability of PRS in African populations

There are many factors contributing to poor transferability of PRS in African populations. This includes genetic factors such as minor allele frequencies, difference in linkage disequilibrium patterns, and their interactions with environmental considerations like diet, exercise, age, gender, and variability in phenotype measurement.

PRS are calculated by aggregating the effect of many common variants that are associated with the diseases of interest. Given that Africa is the continent where all humans originated, it has the highest genetic diversity in the whole world. However, the current lack of diversity in genomic studies have implications on the predictive power of the methods that are trained and developed on euro-centric datasets. PRS constructed with such method may differ primarily on how the weight of the effect size is generated and how the number of single nucleotide polymorphism (SNPs) to be included in the PRS calculation is determined. For example, the interrogation of high-risk variants may involve inclusion of a causal variant from a population, whereas PRS estimates may incorporate variants that are not perfectly correlated with the causal genetic factor [7]. The implication of this is that the method may incorporate a variant with uncertain effect size in the PRS which invariably may reduce the generalizability of PRS risk estimates in the target population.

A linkage disequilibrium (LD) reference panel and data on allele frequencies are prerequisites for application of PRS methods in a heterogeneous background. These factors are important for PRS development. For example, allele frequency differences may cause predicted risks of a disease to vary across populations. Given that LD blocks are shorter in African populations, the SNPs which are in high LD can be removed as they inflate the score. Several studies have shown lower levels of LD in African populations compared to other populations [8] which may imply that the power to detect untyped causal loci is reduced. This LD and distance between the causal variants and the GWAS tagging SNPs can explain lower accuracy and limited transferability. The relative accuracy of polygenic scores is enhanced when LD and minor allele frequencies are integrated into the model [9] assuming that causal variants are shared between populations. Invariably, LD pattern differences between discovery and target populations may impact the effect size calculation and determination of causal variant. Therefore, the transferability of risk score across populations is a major challenge when they do not share the same genetic architecture for each disease [10].

The phenomenon of pleiotropy is an indication of the complexity of how mutations in one locus may influence several pathways or functions. One gene may be responsible for different unrelated traits. SNP markers that are selected for PRS calculations may well impact on another phenotype(s) than the one for which the risk is being calculated. Graff et al. found patterns of pleiotropy when investigating PRS in Europeans for 16 different types of cancers [11]. Positive associations were reported between several forms of cancer. Some variants may be associated

with a disease while being protective for another pathology [11]. Pleiotropic effects of certain variants may well confound risk estimations when used for PRS calculations [11].

Many PRS methods have attempted to solve the problem of different LD pattern using LD clumping or/and penalised regression. Ge and colleagues [12] and Baker and colleagues [13] have extensively reviewed such different PRS methods, and, as such, this paper is not intended to duplicate this effort. However, we provide a summary of how some popular PRS tools account for LD (Additional file 1: Table S1). The main problem with the clumping and thresholding approach is that it does not take environmental factors such diet and exercise into consideration which might confound the predictive accuracy of these measures.

In addition to genetic contributions to lower PRS accuracy in African population, environmental exposures can also play a major role in contributing to poor transferability of PRS across populations [5]. As most GWAS may have already been subject to ascertainment bias [7], such study recruitment mainly from rural, urban, healthier, poorer, or educated participants only may introduce collider bias [14]. In a recently published paper describing PRS in African populations, Kamiza and colleagues[4] show that environmental factors such as diet, exercise, age, gender, and living in rural or urban community can influence PRS portability [15]. The results from this paper suggest that poor transferability of PRS between South African Zulu and Ugandan populations is due to differences in environmental and genetic factors between the two African populations [15]. They also showed that lipid predictability was lower in East Africa Uganda population than in South Africa Zulu population which was attributed to non-fasting of participants before blood collection for lipid analysis. Similarly, a type 2 diabetes PRS paper [16] shows a varied predictability between Kenya and Ghana and Nigeria, where predictability was much higher. Although these two studies show that PRS derived from data of African American individuals enhance polygenic prediction in sub-Sahara Africa compared to European and multi-ancestry scores, it is important to note that the studies further show that PRS prediction varied greatly within SSA, implying that African American-derived PRS may not be generalizable across populations in Africa. This reason may be that only certain geographies and genetic variation are represented in African American [17].

In the study by Reisberg and colleagues [18], SNPs from the European cohorts were used for calculating the risk for type 2 diabetes, and the 1000 Genomes dataset of African populations had the highest scores compared to Europeans. The report from Hugues and colleagues[15] verifies that when population-specific SNPs are included in the calculations, the risk calculation is improved. Considerable efforts to understand GxE interaction effects is key for transferability of PRS as the effects of genetic variants on phenotype can be different between populations as demonstrated by Chikowore and colleagues [16].

In addition, research strategies and medical procedures are not always consistent across all countries in Africa. This is critical for diseases such as psychiatric disorders, where phenotype reporting requires intricate and complicated procedures. As an alternative, minimal phenotyping, which has recourse to hospital records, self-reporting symptoms, or prescription of medications, is used for identification of cases. This approach consists of sampling based on heterogeneous

self-reported symptoms and not on the recommended criteria for diagnosis. GWAS based on minimal phenotyping produce a large number of associated loci which are however of lower heritability and have non-specific effects. Cai et al. show that when minimal phenotyping is used, for major depressive disorder (MDD), the genetic architecture is different from when the strictly defined MDD is used [19].

Collectively, genetic factors such as differences in effect sizes, allele frequencies, LD patterns, phenomenon of pleiotropy, and phenotyping in addition to environmental exposures are limiting the generalizability of genetic predictions of diseases and traits to African populations. Pereira, et al. discussed in detail these factors that influence PRSs and limit transferability including highlighting the complex scenarios of the importance of using genomic data from multiple populations to develop appropriate population-specific applications [20].

Growing collection of continental African genomic datasets for more accurate PRS

To do genomic research, biobanks are essential. As a result, national biobanks have been established by several governments globally to support scientific research and advance precision medicine. One notable example is UKBioBank [21], a well-known biobank that gathers health and genetic information from 500,000 people in the UK. (2) The All of Us Research Programme (USA) seeks to recruit one million or more individuals from a variety of backgrounds to provide a resource for precision medicine [22]. (3) The Estonian Biobank—a nationwide biobank effort with the aim of enhancing genetic research and healthcare in the nation—collects genomic and health-related data from over 200,000 members [23]. Some other genomic medicine initiatives include those from Canada, Qatar, Turkey, Japan, Finland, Denmark, Australia, Saudi Arabia, Switzerland, China, and Brazil [24], but such national biobank is lacking in Africa.

Growing evidence shows that using large-scale African ancestry cohorts as discovery for PRS development may generate more generalizable findings. Data from GWAS are fundamental as they are used for developing PRS. To date, GWAS has increasingly identified a large number of genetic variants which are associated with a range of complex traits [5, 7, 24, 25]. However, the majority of GWAS has been conducted with data from individual of primarily European and Asian descents [5, 25]. PRS can help to estimate individual's genetic risk to a disease or condition by aggregating the effect of many common variants associated with the condition, but studies have shown that well-powered large-scale-based data are required to derive PRS which are currently lacking in continental Africa. This calls for the need to initiate a step-change in the scale of such studies in African populations to enhance PRS prediction or aggregate emerging genomic datasets comparable with European and Asian genomic initiatives. African genetic data have revealed highly relevant African-enriched variants in genes such as *APOL1*, PCSK9, and *G6PD* for kidney diseases, lipid traits, and diabetes respectively [26]. In Table 1, we show a growing collection of rich continental African genomic datasets linked to mostly non-communicable disease phenotypes becoming available for generating PRS for African populations.

Table 1 Description of a range of genomic datasets in Africa for potential polygenic risk scores for disease risk in Continental Africa

One key factor to determine the accuracy and predictive power of PRS is the power of the discovery GWAS data to avoid reaching misleading conclusions [9]. To improve cross-population polygenic risk prediction, specifically, Weissbrod and colleagues [59] recommended that base GWAS should have at least 100 K individuals to observe relate prediction and accuracy of PRS. Unfortunately, the most current GWAS data from continental Africa are under-powered with sample sizes ranging from 150 to 12,000 individuals representing only 1.1% of genomic studies from all African ancestry individuals [5].

In order to improve the representation of African genomic data in the global context for discovery and genetic risk prediction in the last decade, some initiatives have been initiated in Africa including the Africa America Diabetes Mellitus (AADM) [30], the Uganda genome project [27, 28], the Human Heredity and Health in Africa consortium [60, 61], the Nigerian 100 K genome project [50], and many others with smaller sample sizes and limited potential to get published. Aggregating all the datasets including many emerging ones (Table 1) will (1) improve discovery power for GWAS and PRS, (2) improve representation of African genomics in the global context, (3) provide a unique framework to examine a wide range of health indices in African populations, and (4) aid insights into the biological mechanisms and aetiology underlying disease risk in African populations, informing the wider application of potential preventative or therapeutic strategies.

Barriers and potential clinical utility of PRS in African populations

Hundreds of PRS studies have been carried out including those on its clinical utility mainly in European populations [62,63,64,65]. A recent systematic review by Kumuthini and colleagues [65] shows a conflict claim for and against utility of PRS. This analysis did not discover published evidence of a PRS's clear clinical usefulness, though they show numerous examples of near evidence of clinical utility and ample demonstration of clinical validity [65]. Conversely, there is also a growing number of investigations suggesting that PRS are not more predictive than standard of care [66], for example, two retrospective studies that integrated coronary disease PRS and found no and a modest statistically significant improvement in accuracy compared to use of the same models without the score [67, 68]. In the analysis of two US cohorts, Mosley and colleagues [68] show that the PRS was associated with incident of coronary heart disease events but did not significantly improve discrimination, calibration, or risk reclassification compared with conventional predictors. However, a few PRS-based genetic risk estimates from continental Africa [15, 16] have shown promises in the ability of PRS to identify subgroups of individuals who may benefit from the prioritisation of preventive actions.

The potential utility of PRS in African populations is limited by many factors. First, the current PRS methods limit the general utility of PRS as they have mostly been developed and optimised in European populations. Unless sufficient research is also undertaken to optimise the application

of PRS in African populations, there is a risk of inequitable distribution of health benefits from future clinical utility of PRS. PRS calculations cannot, for now, capture the full spectrum of disease risk because of allele types, their frequencies, and their effect sizes. For precise estimates to be possible, a complete representation of all contributing loci is desirable [69]. Current PRS methods can be improved with non-genetic parameters included in the models. More dynamic methods to estimate the effects of specific genetic variations given the genetic, demographic, and clinical risk factor backgrounds of the individual are anticipated to be developed as representation from Africa and other underrepresented populations increases. It is reassuring to see a coordinated efforts such as the Polygenic RIsk MEthods in Diverse Populations (PRIMED) that promises to deliver new methods for risk prediction in diverse ancestry and specifically a pan-Africa initiative—CARdiometabolic Disorders IN African-ancestry PopuLations (CARDINAL) project which aim to test PRS performance on African individuals with phenotype and genotype data available from H3Africa projects [70].

Lack of infrastructure and difficulties with accurate phenotyping are major barriers for conducting genomic research in resource limit settings like Africa. However, to ensure collection of more accurate phenotype, a standardised data collection instrument known as the H3Africa Standard Case Report Form (CRF) was developed by H3Africa [71], which enables efficient and complete data collection, processing, analysis, and reporting. While the issue of heterogenous phenotyping remains, there exists some commonly statistical method for analysing a collection of studies for which the effect sizes are expected to vary. Random-effects model for GWAS meta-analysis is designed specifically for the case in which there is heterogeneity [72]. The other commonly used fixed-effects meta-analysis will only increase power if effects are homogeneous across studies.

For a sustainable solution to some lack of infrastructure in the continent for genomic research, H3Africa [60] and other initiatives in Africa are partnering with biotechnology company such as Illumina. Africa can now boast of large genomics facility with the latest cutting-edge technology in Nigeria, South Africa, Kenya, Uganda, and other places. Notably are African Center of Excellence for Genomics in Infectious Diseases (ACEGID) Nigeria, KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Centre for Epidemic Response and Innovation (CERI), and Centre for Proteomic and Genomic Research (CPGR) in South Africa. These new facilities enable African researchers to avoid major delays in cross-border shipping of biological samples and to ensure the ability to reuse these valuable datasets. In the past, infrastructure for sample processing, biobanking, genotyping, or sequencing and computational analysis are often outsourced, but these are now gradually changing. To optimally benefit from these technologies that foster implementation of PRS, Africa countries must first address the collapsing primary health care mostly in the rural communities, although, in the urban context, Africans have choice to access advanced medical technologies despite the current shortcomings of the healthcare systems including an increased in demand for genetic tests for preventive purposes (for example: cancer panels such as MammaPrint in South Africa).

Even with the most accurate PRS, addition of conventional risk factors to PRS would be central to potential clinical utility in Africa [16, 73]. Such clinical utility of PRS in Africa will require an

extensive awareness and education for both the physicians, patients, and the public regarding the importance and interpretation. In particular, methods that integrate uncertainty deriving from measured as well as unmeasured factors would be valuable for communicating the uncertainty associated with genetic risk estimations at an individual level [74, 75] including approaches to mitigate incidental findings, genetic discrimination, and what the role of counsellors and expert mediators would be in such cases. Such clinical utility of PRS would need to be fully supported by a robust ethical framework and effective regulatory system. For example, ethically, the genetic study of cognitive ability remains controversial both scientifically and ethically, and as such, the utility of PRS would need to be regulated within certain traits and phenotypes.

Ultimately, as shown by a few African PRS studies, PRS may have clinical utility in Africa when combined with traditional risk factors for some diseases, such as cardiometabolic traits, but first, right healthcare system and genomic infrastructure must be in place, and large-scale African genomic studies are required to demonstrate the utility of polygenic risk estimation. This might require the development of multiple models for every disease given the broad genetic diversity within Africa.

Future directions and conclusion

PRS currently have limited transferability, caused mainly lack diversity in genomic studies. To improve the prediction accuracy of PRS in African ancestry individuals, it is most important to include ethnically diverse individuals from continental Africa in genomic studies. Wonkam [76] recommended a rough estimate of about three million African genomes (3MAG) to capture the full scope of Africa's genetic variation and a representative human reference genome. This is mostly hindered by the lack of accurate population descriptions of African populations. Participants are mostly defined as per their geographic region or country, while it is well established that most countries are not homogeneous and can have profound genetic differences. Botswana, for example, hosts populations that are descendants of Bantu from West Africa and people of South African ancestry [77]. Similarly, Bantu speakers of Uganda contrast with non-Bantu speakers from the same country. Substantial genetic variations across regions of Africa must be carefully addressed for the integration of genomics data in health care. In a personal communication with Wonkam [76], he explained that 3MAG was a very rough estimate base on two simple assumptions: (1) the Human Genomes Project (HGP) estimated that between two unrelated individuals, there is a SNV every 1300 bp, therefore about 3,000,000 SNV difference, considering that each genome has 3 billion nucleotides. (2) owing to the great diversity in Africa, if we assume that most African has at least one uncaptured SNV, we need a minimum of 3 million African to capture, at least, the SNVs in our genome, although we see the potential of bias in this estimation [78] and several logistical and financial challenges to consider. Nevertheless, we agree with the proposition that a comprehensive and extensive genome sequencing programme in Africa is of utmost importance. This undertaking is essential for the comprehensive representation of the continent's extensive genetic diversity. New initiatives in Africa, such as the ambitious plan to establish eight Genomics Centres of Excellence (GenCoE) across the continent, seek to revolutionise access to cutting-edge genomics technologies and reshape the continent's response to some of its most pressing health challenges (https://www.nature.com/articles/d44148-023-00052-z). The initiative, which carries a

significant price tag of US$200 million, is built from the 3MAG programme and seeks to obtain financial support from many sources globally. These sources include African governments, industrial partners, the US government, and other funding agencies. It is imperative that Africa actively participates in the genomic medicine revolution, ensuring that it does not lag behind in harnessing the transformative potential it offers.

Such large-scale African genomic studies like 3MAG can reveal novel genes including causal genetic variants not found in previous Eurocentric studies. In addition, it would offer the opportunity to develop regional PRS within Africa to cater for genetic differences within Africans which is even larger than between Africans and Eurasians. Invariably, this would largely solve many barriers poised by difference in allele frequencies, effect sizes, and LD patterns when developing PRS. Leveraging the greater genetic diversity in Africa, within representative genomic data from Africa, PRS derived from African population may be more predictive to all global populations [6].

Currently, PRS statistical models are trained with Eurocentric datasets. While representation of African genomics is being improved which might take decades, statistical model could be trained to estimate the projected effect sizes and allele frequency of those unknown African GWAS loci for genetic risk prediction. With current advances in machine learning and artificial intelligence, expanding the PRS models is a more practical solution to addressing the effect of genetics and its interaction with environmental exposure [79]. This method will require to be trained with different datasets. Such datasets for PRS across diverse ethnic groups in African populations have been highlighted in Table 1. Eventually, more dynamic methods for estimating effects associated with individual genetic variants given the individual's genetic, demographic, and clinical risk factor background should be developed [79]. We think the future of PRS in diverse Africa population lies in the development of multiple PRS models per disease from African discovery datasets.

Considering the current poor state of many healthcare settings in Africa, even with best models and perfect PRS transferability, the prospect of clinical utility of PRS is slim in resource-limited medical settings. It is most likely that PRS would first be accessible across Africa via direct-to-consumer (DTC) company and specialist private hospitals for only those who could afford it, but there are concerns about ethical legal and social issues (ELSI) and how PRS will be regulated. Regulatory bodies should consider limiting power in the hands of PRS service providers to use their discretion to test and report any conditions or traits; otherwise, the easy access to PRS may also lead to inappropriate use and abuse. For example, the use of PRS for embryo selection, intelligence, and other psychiatric and socio-behavioural traits is strongly recommended to be restricted.

Collectively, in the future, with increased representation of Africans in genomics, sophisticated predictive PRS models which account for both genetic and non-genetic factors, it may well be possible for PRS to be utilised in the medical practice for some diseases with multiple polygenic scores generated for different diseases or traits in combination with conventional risk factors.

This would need to be guided with robust ethical framework, but more translational research is needed.

References

1. Wang Y, Guo J, Ni G, Yang J, Visscher PM, Yengo L. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. Nat Commun. 2020;11(1):1–9. https://doi.org/10.1038/s41467-020-17719-y.

   Article **CAS** **Google Scholar**

2. Fatumo S. The opportunity in African genome resource for precision medicine. EBioMedicine. 2020;54:102721. https://doi.org/10.1016/j.ebiom.2020.102721.

   Article **PubMed** **PubMed Central** **Google Scholar**

3. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161–4.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

4. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. Cell. 2019;177(1):26–31.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

5. Fatumo S, Chikowore T, Choudhury A, et al. A roadmap to increase diversity in genomic studies. Nat Med. 2022;28(2):243–50. https://doi.org/10.1038/s41591-021-01672-4.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

6. Fatumo S, Inouye M. African genomes hold the key to accurate genetic risk prediction. Nat Hum Behav. 2023;7:295–6. https://doi.org/10.1038/s41562-023-01549-1.

   Article **PubMed** **Google Scholar**

7. Uffelmann E, Huang QQ, Munung, NS,Vries J, Okada Y, Martin AR et al. Genome-wide association studies. Nat Rev Methods Primers. 2021;59. https://doi.org/10.1038/s43586-021-00056-9.

8. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9:403–33. https://doi.org/10.1146/annurev.genom.9.081307.164258.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

9. Choi SW, Mak TSH, Porsch RM, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genet Epidemiology. 2017;41(6):469–80.

   Article **PubMed** **Google Scholar**

10. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020;12(1):1–11.

    Article **Google Scholar**

11. Graff RE, Cavazos TB, Thai KK, Kachuri L, Rashkin SR, Hoffman JD, Sakoda LC. Cross-cancer evaluation of polygenic risk scores for 16 cancer types in two large cohorts. Nat Commun. 2021;12(1):970.

    Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

12. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10(1):1–10. https://doi.org/10.1038/s41467-019-09718-5.

    Article **CAS** **Google Scholar**

13. Baker E, Escott-Price V. Polygenic risk scores in Alzheimer's disease: current applications and future directions. Front Digit Heal. 2020;2(August):1–7.

    Google Scholar

14. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet. 2019;104(1):21–34.

    Article **CAS** **PubMed** **Google Scholar**

15. Kamiza AB, Toure SM, Vujkovic M, Machipisa T, Soremekun OS, Kintu C, et al. Transferability of genetic risk scores in African populations. Nat Med. 2022;28(6):1163–6. https://doi.org/10.1038/s41591-022-01835-x.

    Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

16. Chikowore T, Kamiza AB, Oduaran OH, Machipisa T, Fatumo S. Non-communicable diseases pandemic and precision medicine: is Africa ready? EBioMedicine. 2021;65:103260. https://doi.org/10.1016/j.ebiom.2021.103260.

    Article **PubMed** **PubMed Central** **Google Scholar**

17. Fatumo S, Choudhury A. African American genomes don't capture Africa's genetic diversity. Nature. 2023;617(7959):35–35.

    Article **CAS** **PubMed** **Google Scholar**

18. Reisberg S, Iljasenko T, Läll K, Fischer K, Vilo J. Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. PLoS One. 2017;12(7):1–9.

Article **Google Scholar**

19. Cai NA, Revez JA, Adams MJ, Andlauer TF, Breen G, Byrne EM, Clarke TK, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet. 2020;52(4):437–47.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

20. Pereira L, Mutesa L, Tindana P, et al. African genetic diversity and adaptation inform a precision medicine agenda. Nat Rev Genet. 2021;22:284–306. https://doi.org/10.1038/s41576-020-00306-8.

Article **CAS** **PubMed** **Google Scholar**

21. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9. https://doi.org/10.1038/s41586-018-0579-z.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

22. All of Us Research Program Investigators. The "All of Us" research program. N Engl J Med. 2019;381(7):668–76.

Article **Google Scholar**

23. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, Metspalu A. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol. 2015;44(4):1137–47.

Article **PubMed** **Google Scholar**

24. Wei Zhou, Kanai Masahiro, Wu Kuan-Han H, Rasheed Humaira, Tsuo Kristin, Hirbo Jibril B, Wang Ying, et al. Global Biobank meta-analysis initiative: powering genetic discovery across human disease. Cell Genomics. 2022;2(10):100192.

Article **Google Scholar**

25. Peterson RE, Kuchenbaecker K, Walters RK, Popejoy AB, Periyasamy S, Lam M, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell. 2020;179(3):589–603.

Article **Google Scholar**

26. Rotimi CN, Dunston GM, Berg K, Akinsete O, Amoah A, Owusu S, et al. In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. Ann Epidemiol. 2001;11(1):51–8.

Article **CAS** **PubMed** **Google Scholar**

27. Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, Prado-Martinez J, et al. Uganda genome resource enables insights into population history and genomic discovery in Africa. Cell. 2020;179(4):984–1002.

Article **Google Scholar**

28. Fatumo S, Mugisha J, Soremekun OS, Kalungi A, Mayanja R, Kintu C, Kaleebu P. Uganda Genome Resource: a rich research database for genomic studies of communicable and non-communicable diseases in Africa. Cell Genomics. 2022;2(11):100209.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

29. Hird TR, Young EH, Pirie FJ, Riha J, Esterhuizen TM, O'leary B, et al. Study profile: The Durban Diabetes Study (DDS): a platform for chronic disease research. Glob Heal Epidemiol Genomics. 2016;1:e2.

Article **CAS** **Google Scholar**

30. Tekola-Ayele F, Adeyemo AA, Rotimi CN. Genetic epidemiology of type 2 diabetes and cardiovascular diseases in Africa. Prog Cardiovasc Dis. 2013;56(3):251–60. https://doi.org/10.1016/j.pcad.2013.09.013.

Article **PubMed** **Google Scholar**

31. Agyemang C, Beune E, Meeks K, Owusu-Dabo E, Agyei-Baffour P, De-Graft Aikins A, et al. Rationale and cross-sectional study design of the research on obesity and type 2 diabetes among African migrants: the RODAM study. BMJ Open. 2014;4(3):e004877.

Article **PubMed** **PubMed Central** **Google Scholar**

32. Mtatiro SN, Singh T, Rooks H, Mgaya J, Mariki H, Soka D, et al. Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. PLoS One. 2014;9(11):7–14.

Article **Google Scholar**

33. Ramsay M, Crowther N, Tambo E, Agongo G, Baloyi V, Dikotope S, Sankoh O. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. Glob Health Epidemiol Genomics. 2016;1:e20.

Article CAS Google Scholar

34. Choudhury A, Brandenburg JT, Chikowore T, Sengupta D, Boua PR, Crowther NJ, et al. Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. Nat Commun. 2022;13(1):2578. https://doi.org/10.1038/s41467-022-30098-w.

Article CAS PubMed PubMed Central Google Scholar

35. Ekoru K, Young EH, Adebamowo C, Balde N, Hennig BJ, Kaleebu P, et al. H3Africa multi-centre study of the prevalence and environmental and genetic determinants of type 2 diabetes in sub-Saharan Africa: study protocol. Glob Heal Epidemiol Genomics. 2016;1:e5.

Article CAS Google Scholar

36. Machipisa T, Chong M, Muhamed B, Chishala C, Shaboodien G, Pandie S, et al. Association of novel locus with rheumatic heart disease in Black African individuals: findings from the RHDGen study. JAMA Cardiol. 2021;6(9):1000–11. https://doi.org/10.1001/jamacardio.2021.1627.

Article PubMed PubMed Central Google Scholar

37. Adebamowo SN, Dareng EO, Famooto AO, Offiong R, Olaniyan O, Obende K, et al. Cohort Profile: African Collaborative Center for Microbiome and Genomics Research's (ACCME's) Human Papillomavirus (HPV) and Cervical Cancer Study. Int J Epidemiol. 2017;46(6):1–11.

Article Google Scholar

38. Kaplan MH, Contreras-Galindo R, Jiagge E, Merajver SD, Newman L, Bigman G, et al. Is the HERV-K HML-2 Xq21.33, an endogenous retrovirus mutated by gene conversion of chromosome X in a subset of African populations, associated with human breast cancer? Infect Agent Cancer. 2020;15(1):1–15.

Article Google Scholar

39. Mentzer AJ, Dilthey AT, Pollard M, Gurdasani DT, Karakoc E, Carstensen T, Muhwezi A, et al. High-resolution African HLA resource uncovers HLA-DRB1 expression effects underlying vaccine response. medRxiv. 2022;2022:11.

Google Scholar

40. Muriuki JM, Mentzer AJ, Mitchell R, Webb EL, Etyang AO, Kyobutungi C, Atkinson SH. Malaria is a cause of iron deficiency in African children. Nat Med. 2021;27(4):653–8.

Article CAS PubMed PubMed Central Google Scholar

41. Crampin AC, Kayuni N, Amberbir A, et al. Hypertension and diabetes in Africa: design and implementation of a large population-based study of burden and risk factors in rural and urban Malawi. Emerg Themes Epidemiol. 2016;13:3. https://doi.org/10.1186/s12982-015-0039-2.

   Article **PubMed** **PubMed Central** **Google Scholar**

42. Sarfo FS, Ovbiagele B, Gebregziabher M, Wahab K, Akinyemi R, Akpalu A, et al. Stroke among young West Africans: Evidence from the SIREN (stroke investigative research and educational network) large multisite case-control study. Stroke. 2018;49(5):1116–20.

   Article **PubMed** **PubMed Central** **Google Scholar**

43. Musanabaganwa C, Jansen S, Wani A, Rugamba A, Mutabaruka J, Rutembesa E, Mutesa L. Community engagement in epigenomic and neurocognitive research on post-traumatic stress disorder in Rwandans exposed to the 1994 genocide against the Tutsi: lessons learned. Epigenomics. 2022;14(15):887–95.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

44. Rudahindwa S, Mutesa L, Rutembesa E, Mutabaruka J, Qu A, Wildman DE, et al. Transgenerational effects of the genocide against the Tutsi in Rwanda: a post-traumatic stress disorder symptom domain analysis. AAS Open Res. 2020;1:10.

   Article **Google Scholar**

45. Hennig BJ, Unger SA, Dondeh BL, Hassan J, Hawkesworth S, Jarjou L, et al. Cohort profile: The Kiang West Longitudinal Population Study (KWLPS)-a platform for integrated research and health care provision in rural Gambia. Int J Epidemiol. 2017;46(2):1–12.

   Google Scholar

46. Anie KA, Olayemi E, Paintsil V, Owusu-Dabo E, Adeyemo TA, Sani MU, et al. Sickle Cell Disease Genomics of Africa (SickleGenAfrica) Network: ethical framework and initial qualitative findings from community engagement in Ghana. Nigeria and Tanzania BMJ Open. 2021;11(7):1–10.

   Google Scholar

47. Parekh RS, Rasooly RS, Kimmel P. Genomic approaches to the burden of kidney disease in Sub-Saharan Africa: the Human Heredity and Health in Africa (H3Africa) Kidney Disease Research Network. Kidney Int. 2016;90(1):2–5. https://doi.org/10.1016/j.kint.2015.12.059.

   Article **PubMed** **PubMed Central** **Google Scholar**

48. Mboowa G, Mwesigwa S, Katagirya E, Retshabile G, Mlotshwa BC, Williams L, et al. The Collaborative African Genomics Network (CAfGEN): applying genomic technologies to probe host factors important to the progression of HIV and HIV-tuberculosis infection in sub-Saharan Africa. AAS Open Res. 2018;1:3.

49. Walker R, Whiting D, Unwin N, Mugusi F, Swai M, Aris E, et al. Stroke incidence in rural and urban Tanzania: a prospective, community-based study. Lancet Neurol. 2010;9(8):786–92. https://doi.org/10.1016/S1474-4422(10)70144-7.

Article **PubMed** **Google Scholar**

50. Fatumo S, Yakubu A, Oyedele O. et al. Promoting the genomic revolution in Africa through the Nigerian 100K Genome Project. Nat Genet 2022:531–536. https://doi.org/10.1038/s41588-022-01071-6.

51. Ilboudo H, Noyes H, Mulindwa J, Kimuda MP, Koffi M, Kabore JW, TrypanoGEN Research Group as members of The H3Africa Consortium. Introducing the TrypanoGEN biobank: a valuable resource for the elimination of human African trypanosomiasis; 2017.

52. Happi C. Genomic analysis of Lassa virus during an increase in cases in Nigeria in 2018.

53. Stevenson A, Akena D, Stroud RE, Atwoli L, Campbell MM, Chibnik LB, Koenen KC. Neuropsychiatric genetics of African populations-psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya South Africa and Uganda. BMJ Open. 2019;9(2):e025469.

Article **PubMed** **PubMed Central** **Google Scholar**

54. Kipkemoi, Patricia, Heesu Ally Kim, Bjorn Christ, Emily O'Heir, Jake Allen, Christina Austin-Tse, Samantha Baxter et al. Phenotype and genetic analysis of data collected within the first year of NeuroDev. medRxiv. 2022:2022-08.

55. Matimba A, Oluka MN, Ebeshi BU, Sayi J, Bolaji OO, Guantai AN, Masimirembwa CM. Establishment of a biobank and pharmacogenetics database of African populations. Eur J Hum Genet. 2008;16(7):780–3.

Article **CAS** **PubMed** **Google Scholar**

56. Tindana P, Bull S, Amenga-Etego L, de Vries J, Aborigo R, Koram K, Parker M. Seeking consent to genetic and genomic research in a rural Ghanaian setting: a qualitative study of the MalariaGEN experience. BMC Med Ethics. 2012;13(1):1–12.

Article **Google Scholar**

57. Sirugo G, Loeff MS, Sam O, Nyan O, Pinder M, Hill AV, Kwiatkowski D, et al. A national DNA bank in The Gambia, West Africa, and genomic research in developing countries. Nat Genet. 2004;36(8):785–6.

Article **CAS** **PubMed** **Google Scholar**

58. Sgaier SK, Jha P, Mony P, Kurpad A, Lakshmi V, Kumar R, Ganguly NK. Biobanks in developing countries: needs and feasibility. Science. 2007;318(5853):1074–5.

Article **CAS** **PubMed** **Google Scholar**

59. Weissbrod O, Kanai M, Shi H, Gazal S, Peyrot WJ, Khera AV et al. Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. Nat Genet. 2022:450–458. https://doi.org/10.1038/s41588-022-01036-9.

60. H3Africa Consortium. Enabling the genomic revolution in Africa: H3Africa is developing capacity for health-related genomics research in Africa. Science (New York, NY). 2014;344(6190):1346.

Article **Google Scholar**

61. Adoga MP, Fatumo SA, Agwale SM. H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa. Source Code Biol Med. 2014;9:1–3.

Article **Google Scholar**

62. Padilla-Martínez F, Collin F, Kwasniewski M, Kretowski A. Systematic review of polygenic risk scores for type 1 and type 2 diabetes. Int J Mol Sci. 2020;21(5):1703.

Article **PubMed** **PubMed Central** **Google Scholar**

63. Polygenic Risk Score Task Force of the International Common Disease Alliance. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nature Medicine. 2021;27(11):1876–84. https://doi.org/10.1038/s41591-021-01549-6.

64. Wang Y, Tsuo K, Kanai M, Neale BM, Martin AR. Challenges and opportunities for developing more generalizable polygenic risk scores. Ann Rev Biomed Data Sci. 2022;5:293–320.

Article **Google Scholar**

65. Kumuthini J, Zick B, Balasopoulou A, Chalikiopoulou C, Dandara C, El-Kamah G, et al. The clinical utility of polygenic risk scores in genomic medicine practices: a systematic review. Hum Genet. 2022;141(11):1697–704. https://doi.org/10.1007/s00439-022-02452-x.

Article **PubMed** **PubMed Central** **Google Scholar**

66. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet. 2017;100(4):635–49. https://doi.org/10.1016/j.ajhg.2017.03.004.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

67. Elliott J, Bodinier B, Bond TA, Chadeau-Hyam M, Evangelou E, Moons KGM, Dehghan A, Muller DC, Elliott P, Tzoulaki I. Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. JAMA. 2020;323(7):636–45. https://doi.org/10.1001/jama.2019.22241.

Article **PubMed** **PubMed Central** **Google Scholar**

68. Mosley JD, Gupta DK, Tan J, Yao J, Wells QS, Shaffer CM, et al. Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease. JAMA. 2020;323(7):627–35. https://doi.org/10.1001/jama.2019.21782.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

69. Igo RP Jr, Kinzy TG, Cooke Bailey JN. Genetic risk scores. Curr Protoc Hum Genet. 2019;104(1):e95.

Article **PubMed** **PubMed Central** **Google Scholar**

70. Adebamowo CA, Adeyemo A, Ashaye A, Akpa OM, Chikowore T, Choudhury A, Adebamowo SN. Polygenic risk scores for CARDINAL study. Nat Genet. 2022;54(5):527–30.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

71. H3ABioNet Phenotype Standardisation: Project Documentation. https://www.h3abionet.org/images/DataAndStandards/DataStandards/h3abionetphenstddoc_v1.1.pdf

72. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet. 2011;88(5):586–98.

Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

73. Ekoru K, Adeyemo AA, Chen G, Doumatey AP, Zhou J, Bentley AR, et al. Genetic risk scores for cardiometabolic traits in sub-Saharan African populations. Int J Epidemiol. 2021;50(4):1283–96. https://doi.org/10.1093/ije/dyab046.

Article **PubMed** **PubMed Central** **Google Scholar**

74. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581–90.

Article **CAS** **PubMed** **Google Scholar**

75. Sheehan NA, Didelez V. Epidemiology, genetic epidemiology and Mendelian randomisation: more need than ever to attend to detail. Hum Genet. 2020;139(1):121–36. https://doi.org/10.1007/s00439-019-02027-3.

Article **PubMed** **Google Scholar**

76. Wonkam A. Sequence three million genomes across Africa. Nature. 2021;590(7845):209–11. Available from: https://www.nature.com/articles/d41586-021-00313-7.

77. Choudhury A, Aron S, Botigué LR, et al. High-depth African genomes inform human migration and health. Nature. 2020;586:741–8. https://doi.org/10.1038/s41586-020-2859-7.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

78. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, Corvelo A, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell. 2022;185(18):3426–40.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

79. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584–91.

   Article **CAS** **PubMed** **PubMed Central** **Google Scholar**

Download references

Acknowledgements

Funding

Author information

Authors and Affiliations

1. **The African Computational Genomics (TACG) Research Group, MRC/UVRI and LSHTM, Entebbe, Uganda**

Segun Fatumo & Chisom Soremekun

2. **H3Africa Bioinformatics Network (H3ABioNet) Node, Centre for Genomics Research and Innovation, NABDA/FMST, Abuja, Nigeria**
Segun Fatumo & Chisom Soremekun

3. **Department of Non-Communicable Disease Epidemiology (NCDE), London School of Hygiene and Tropical Medicine, Keppel St, London, WC1E 7HT, UK**
Segun Fatumo

4. **H3Africa Bioinformatics Network (H3ABioNet) Node, University of Mauritius, Reduit, Mauritius**
Dassen Sathan & Yasmina Jaufeerally Fakim

5. **Laboratory of Biotechnology, Environment, Agri-Food and Health, Faculty of Sciences Dhar El Mahraz–Sidi Mohammed Ben Abdellah University, 30000, Fez, Morocco**
Chaimae Samtal

6. **Department of Computer and Information Sciences, Covenant University, P. M. B. 1023, Ota, Ogun State, Nigeria**
Itunuoluwa Isewon

7. **Covenant University Bioinformatics Research (CUBRe), Covenant University, Km 10 Idiroko Road, P.M.B. 1023, Ota, Ogun State, Nigeria**
Itunuoluwa Isewon

8. **Covenant Applied Informatics and Communication African Centre of Excellence (CApIC-ACE), Covenant University, P.M.B. 1023, Ota, Ogun State, Nigeria**
Itunuoluwa Isewon

9. **Division of Computational Biology, Integrative Biomedical Sciences Department, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa**
Tsaone Tamuhla

10. **South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa**
Tsaone Tamuhla & Nicki Tiffin

11. **Department of Immunology and Molecular Biology, College of Health Science, Makerere University, Kampala, Uganda**
Chisom Soremekun

12. **Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi**
James Jafali

13. **Clinical Infection, Microbiology & Immunology, The University of Liverpool, Liverpool, UK**
James Jafali

14. **Computational Biology Group, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, 7925, South Africa**
Sumir Panji

Contributions

SF, YJF, and NT conceptualised and designed the manuscript. DS, YJF, and SF wrote the original manuscript. CS, II, TT, JJ, SP, and NT commented and elaborated on the contents of the

manuscript. SF reworked and revised the manuscript. SF and YJF finalised the manuscript. All authors read and approved the final version of the manuscript.

Corresponding author

Correspondence to Segun Fatumo.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Additional information

Publisher's Note

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

Supplementary Information

Additional file 1: Table S1.

Different methods for the calculation of PRS and their relevant parameters.

Rights and permissions

Reprints and permissions

About this article

Cite this article

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link
Provided by the Springer Nature SharedIt content-sharing initiative

Keywords

Associated Content
Part of a collection:
Advancing human genetics in underrepresented populations

- Sections
- Figures
- References

- Abstract