



An African perspective on genetically diverse human induced pluripotent stem cell lines

Tracey Hurrell, Jerolen Naidoo, Tiro Ntlhafa & Janine Scholefield



Human induced pluripotent stem cell-derived models are a well-established preclinical tool, with the ability to retain the genetics of the individual from which they are derived. Here we comment on the global representation and accessibility of such cellular tools from African population groups.

The equitable representation of global genetic diversity in preclinical cellular models such as human induced pluripotent stem cells (hiPSCs) has been advocated for by many in recent years^{1–3}. This has been driven by the recognition of extraordinary genetic diversity across the globe, especially within historically underrepresented population groups, including those on the African continent. This genetic diversity must be translated into physiologically relevant models to realise equitable benefits from advances in preclinical research for all population groups. Since the (a) efficacy of treatments from vaccines to gene therapies (both *ex vivo* and *in vivo*), (b) preclinical testing of small molecules, (c) diagnostic systems, and (d) deconvolution of the molecular basis of disease, could all be dependent on individual genetic background, the impact for healthcare is bound to be expansive.

Whilst not yet market-approved, over 100 clinical trials for hiPSC-based cellular therapies are ongoing⁴. In addition, as a cellular and disease modelling resource, hiPSC technologies are being utilized to advance preclinical research across myriad fields within drug discovery and regenerative medicine, allowing for the functional investigation of individual donor genetics as they relate to specific diseases and treatment outcomes. Africa's extensive genetic diversity is associated with a unique disease burden which could be repositioned to instead offer a rich resource for the development of therapeutics related to human leucocyte antigen (HLA) matching. Allogeneic strategies represent a significant proportion of current hiPSC clinical trials, yet few studies are actively taking this into consideration (as an example here by Japan⁵) to expand broader population applications. Given the increasing evidence of HLA diversity^{6,7} inclusion of a wider representation of population samples will be required to ensure equitable benefit from advances in cellular therapeutic intervention for historically underrepresented and underserved population groups.

The scale of African genetic diversity

Africa harbours the greatest genetic diversity in the world - which contextualized from within Africa relates to the extraordinary diversity across the >2000 ethnolinguistic groups (constituting the four larger population structures i.e. Afroasiatic, Khoisan, Niger-Congo, and Nilo-Saharan) across the continent which are not adequately represented in research (the extent of which is shown to some degree in Fig. 1).

The initial 1000 genome study dataset revealed that typical African genomes harbour ~25% more genetic variant sites per genome than any other population, as well as the highest number of common variants (>5%) defined as globally rare (<0.5%)⁸. Several initiatives such as the Human Heredity and Health in Africa (H3Africa) program have made great strides towards increasing the representation of African genomes to more closely represent Africa's 17.5% of the global population⁹. In addition, version four of the genome aggregation database (gnomAD V4.0¹⁰) currently includes a three-fold increase in deposited African genomes sequenced since the release of V2.0, yet an overall reduction in the total proportion of deposited African genomes (from 8.8% to 4.6%) is reflected due to an increase in the number of genomes of European ancestry being deposited.

It is therefore critical to ensure that equitable representation of African genetic diversity is addressed in hiPSC collections to mitigate the stark resource gap in the global research landscape. Ghosh and colleagues analysed the status of global ancestry reported for hiPSC collections in 2022², where lines of non-African ancestry were significantly more common than those from African populations. The authors eloquently outlined the importance of greater inclusivity and the subsequent challenges of accessing diverse lines for both academia and industry. For this commentary, we focus on evaluating the representation of iPSCs of African ancestry derived from ethnolinguistic groups (outlined above) from the African continent, based on data accessible in January 2024.

African hiPSC representation within global collections

Several organisations host hiPSC lines denoting population descriptors including 'African', 'African American', or 'Black'. Of the larger global repositories, WiCell¹¹ (United States of America) has reported the highest proportion, comprising approximately 15% of reported lines (Fig. 2A). Importantly the majority (191) of the 203 available African American hiPSC lines hosted at the WiCell repository are each derived from individual donors, serving as a true reflection of genetic diversity present within this repository. The HipSci repository¹² based in Europe, hosts ten lines characterised as Black or African, derived from seven donors with population definitions including donor ethnicity and predicted population. The human pluripotent stem cell registry (hPSCreg¹³) lists 68 hiPSC registered lines denoted to be of Black or African ethnicity, which are however derived from only 12 donors. Of the approximately 6000 lines registered on hPSCreg, 62% lack reported population descriptors referring to genetic ancestry which perhaps reflects the lack of importance historically assigned to genetic diversity by the research community at large. Finally, the Coriell repository¹⁴ (United States of America) holds 17 Black or African American lines which are each derived from individual donors.

Whilst several WiCell lines are derived from individuals with reported clinical phenotypes linked to a specific disease area of research, the majority (74%) are assigned "no disease reported" and

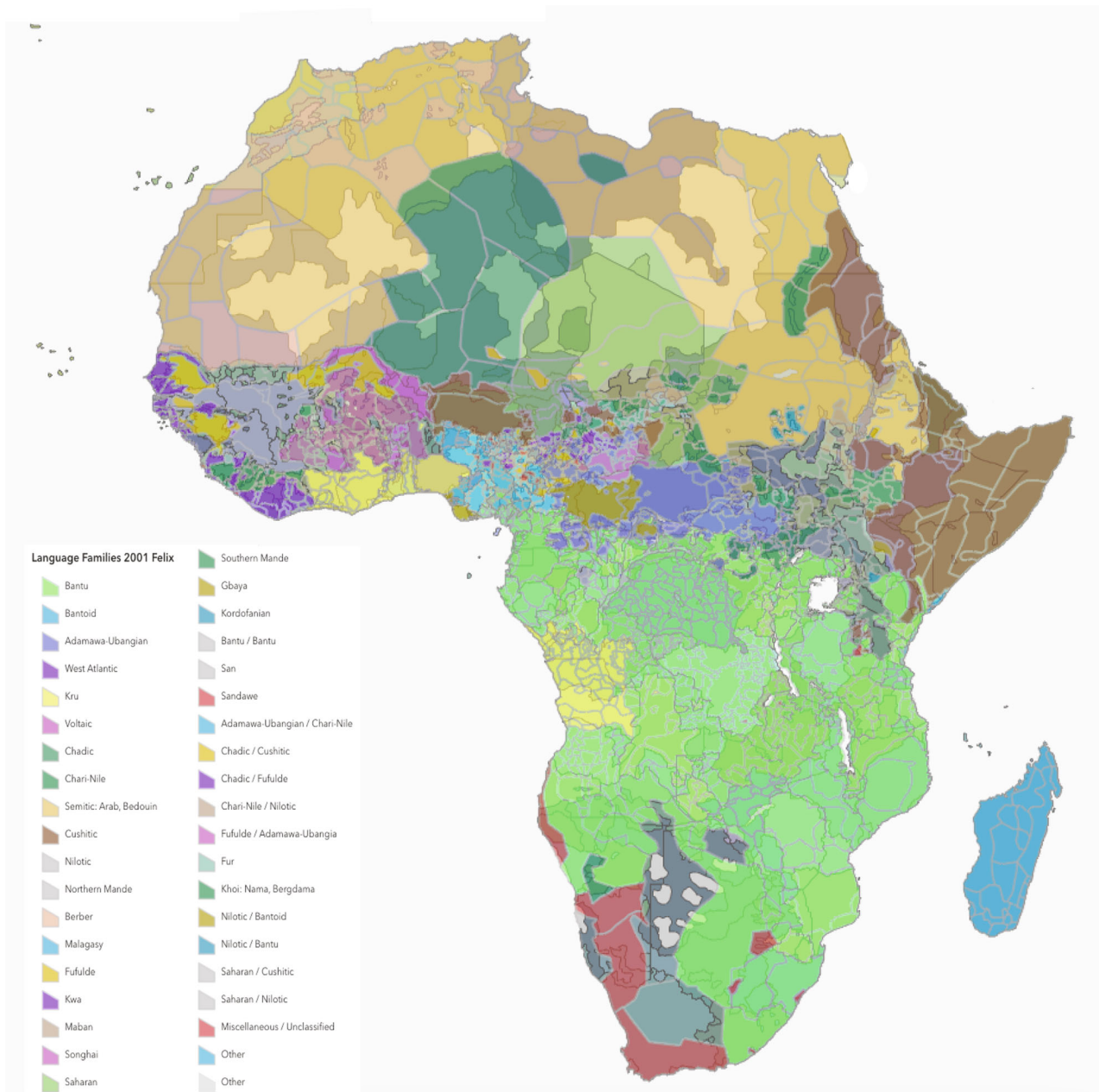


Fig. 1 | Graphic representation of ethnolinguistic diversity in Africa. Overlaid Geo-referencing of Ethnic Groups (GREG) and Language families (Felix 2001) layers using Harvard WorldMap²⁴ are visible. The latter are shown within the key and

reproduced with permission from Marc Leo Felix²⁵, whilst the GREG layers are noted here in ref. 26, derived from ref. 27.

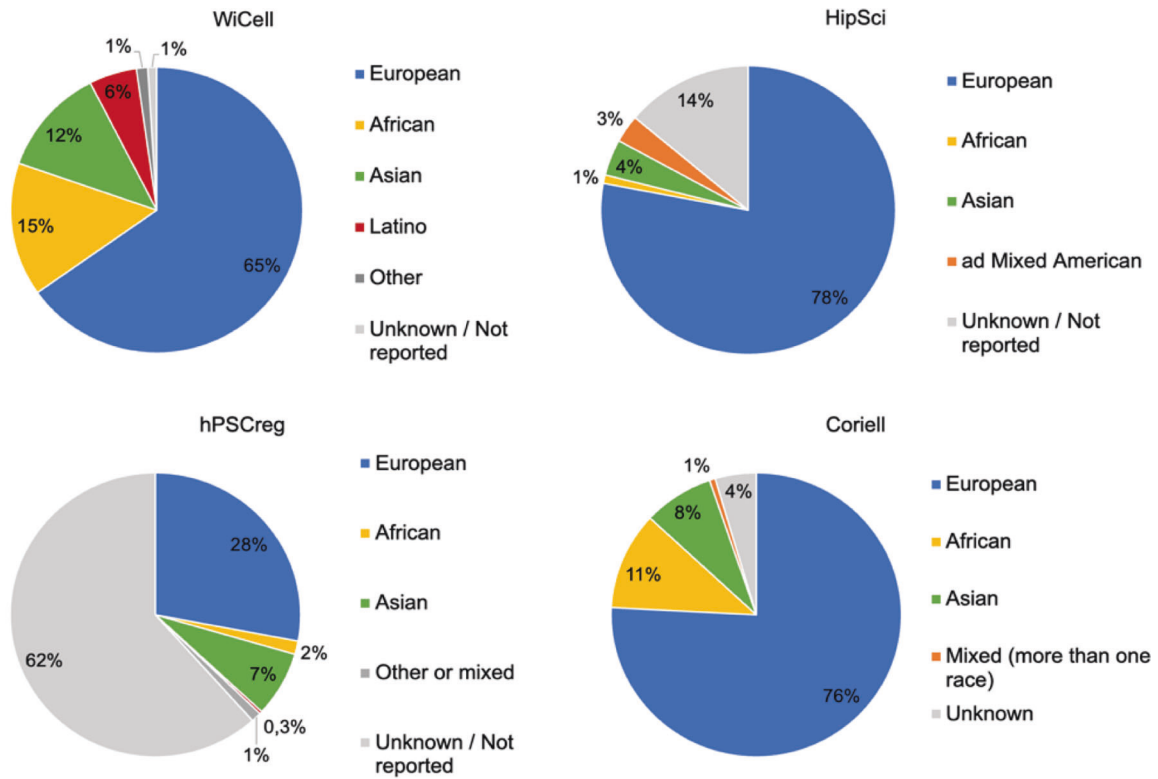
could be used as background population control lines for disease modelling studies. At the time of writing, 57% (91) of these potential control lines are also not restricted *for use* to a specific disease area (Fig. 2B). It's important that such lines should adhere to the International Society for Stem Cell Research recommendations of iPSC characterisation, to ensure quality control standards and confidence in their utility in scientific research studies¹⁵. To the best of our knowledge, only 12 of the 215 African American WiCell lines are annotated as lacking the necessary karyotyping standards and are usefully marked

as such. The Coriell repository also hosts an additional 11 Black/African American hiPSC lines from 11 donors with no reported disease, representing an important resource adding to the global numbers of potential control samples of diverse African representation.

WiCell hiPSC lines associated with a specific disease area are largely derived from a combination of inherent blood and cardiac disease patients. With sickle cell disease¹⁶ and cardiovascular diseases¹⁷ being predominant within African populations, this represents an important potential resource for identifying novel gene therapies and small

Comment

A



B

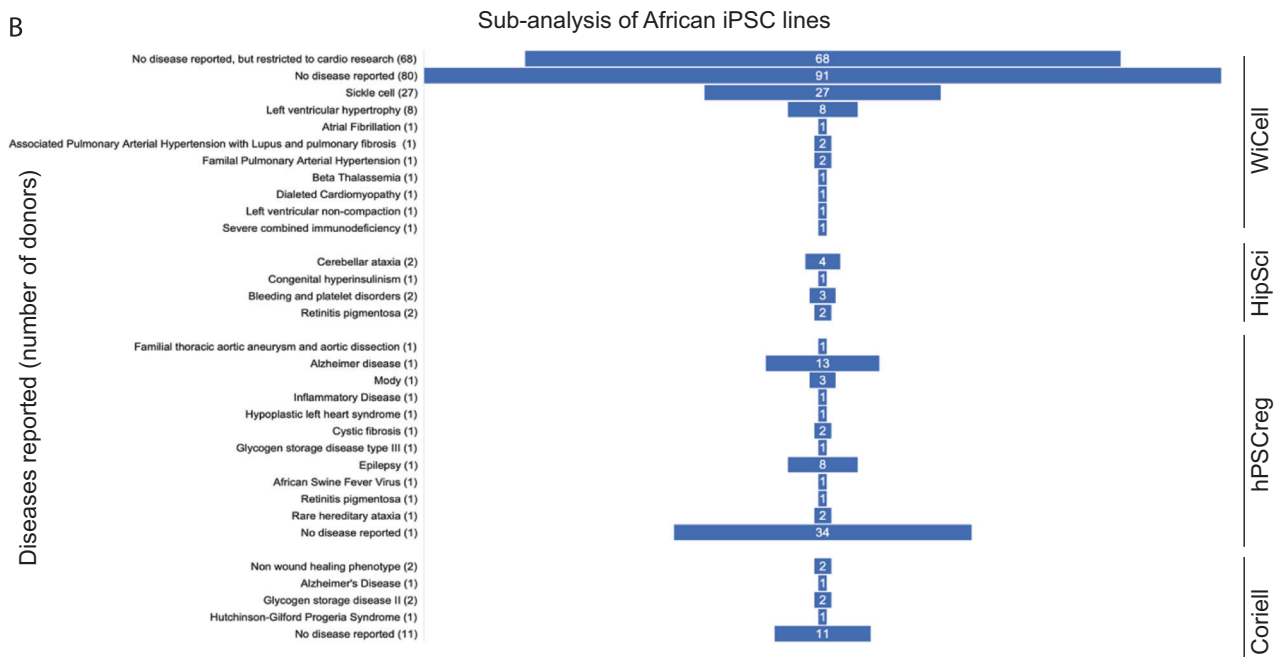


Fig. 2 | Global analysis of iPSC collections. A Relative representation by identified genetic population groups within each repository and the hPSC registry. **B** Detailed sub-analysis of available African iPSC lines outlining clinical pathologies of each repository and associated donor representation.

molecule drugs which are relevant to population-specific disease aetiologies. HipSci, hPSCreg and Coriell add to this global resource by including hiPSCs representing diseases with varying degrees of genetic predisposition including retinitis pigmentosa, Alzheimer's disease, cystic fibrosis (CF), glycogen storage disorder, diabetes and epilepsy (Fig. 2B).

It is clear from this data that ancestral African genetics is present in global repositories. However, improved representation is needed to fully recapitulate Africa's contributions to global genetic diversity. Increased international migration, the need to replace historically genetically homogeneous cell line experimental pipelines with more diverse genetics inherent within the African population, and Africa's reliance on pharmaceuticals, collectively necessitate increased diversity in accessible iPSC lines.

Yet, a concerning narrative still holds across the scientific community; that African hiPSCs from outside the continent itself e.g. African Americans, would sufficiently assuage concerns of diverse representation. The genetic diversity present in many African American populations is predominantly derived from West African populations¹⁸ leading to alteration of allele frequencies through genetic drift, subsequent founder effects and admixture (created by recently linked, but previously distinct geographically distant ancestors within a few generations¹⁹). These effects collectively reinforce that neither this population group nor any other single minority population group outside the African continent, can solely serve as a proxy for the entirety of genetic diversity encompassed within the diverse ethnolinguistic groups on the African continent.

For the purposes of comprehending the implications of ethnolinguistic genetic diversity, we need only look at a few examples within the continent itself. A variant responsible for the metabolism of the antimalarial, chloroquine, has vastly differing allele frequencies across two South African population groups, at 16% in Tsonga-speakers vs 0.8% in Xhosa-speakers²⁰. Other examples include the breadth of disease-causing mutations in spinal muscular atrophy, Duchenne muscular dystrophy, and CF where rare or previously unknown mutations are increasingly recognized as being predominantly responsible for the disease phenotype²¹ offering a valuable resource for disease modelling. Indeed, the globally predominant CF disease-causing *delf508* mutation is present in less than 50% of patients on the African continent with an additional 21 uniquely African mutations littered across the continent.

In addition, the equitable representation of patient-derived samples of idiopathic and/or complex disease, underpinned by diverse genetic variation across African ancestral groups, would serve as a critical turning point for the future impact of hiPSC technologies on the African continent. This would provide preclinical resources that not only facilitate multiple avenues of research related to disease pathophysiology and therapy but are also relevant to the continent's diversity. These are some of the many demonstrations that justify the need to diversify hiPSC representation from within and across the African continent.

To the best of our knowledge, most of the samples in global repositories represent individuals from only a few African countries. However, Africa's regional diversity is broader than a simple designation of 'African descent' where a breakdown of ethnolinguistic sub-populations highlights the diversity even within a single African country²². This is critical to reinforce given the recent evidence of extraordinary regional diversity between 111 sub-Saharan ethnolinguistic and geographical groups across the continent²³.

A local solution for a global problem

A requirement for the inclusion of African ancestry within hiPSC studies would represent a welcome step towards addressing existing disparities within the literature and global hiPSC resources, but there is still a clear need for a Pan-African approach to address equitable representation for the continent's diverse population groups. At the same time, the potential for symbolic representation of only a few 'African ancestry' hiPSC lines should be cautioned against to avoid a common misconception which is underpinned by the idea that this would adequately represent an entire continent which harbours more genetic diversity than the rest of the world combined.

Whilst the number of lines derived from ancestral African population groups is increasing, more still needs to be done to ensure that cellular models such as hiPSCs that are representative of African genetic diversity are locally generated, regulated and made globally accessible. At the time of publication, only five lines generated in Africa have been submitted to hPSCreg which highlights the necessity for investing resources within the continent. Building on converting this vision into further scalable action, initiatives that seek to grow the cellular representation of the African diaspora should be strongly advocated for, and accelerated to implementation ensuring that it is rooted in a collaborative network of governmental, clinical, and scientific stakeholders.

A lack of resources (skills and funding) has led to the under-representation of samples spanning African ethnolinguistic groups in global collections. Our goal is to ensure that representation of the African genetic diaspora is incorporated into global R&D pipelines to ensure a future of equitable healthcare for the continent. Whilst this requires consultation with established global programmes to include diverse hiPSCs, we, as contributors of African research, reinforce that community consultation (research bodies and the general population) through public lectures is an absolute requirement to ensure Africa is not harvested of her diversity, without due benefit to the local population.

Partnerships that support the financially sustainable localisation of iPSC generation in under-represented regions with associated training of local scientists are needed. We envisage that with a public-private partnership supported by the government, a mutually beneficial structure could be financially sustainable in the long-term, with industry players supporting the generation and expansion of diverse cellular lines encompassing both bespoke clinical cohorts as well as much-needed background controls, accessible by academia. Providing accessibility of African-derived iPSCs must be balanced by the need to prevent exploitation of this resource. Commercial licencing of the African lines to industry is an essential first step towards improving accessibility to these resources for African researchers. Another strategy worth considering would be to ensure African research interests are equitably represented within the IP positions defined for industry use. The most optimal structure may involve a variation on such strategies but must, at its inception, ensure that African genetic diversity is not exploited. In addition, considerations for equitable IP could provide an opportunity for additional long-term sustainability, and benefit sharing. To do this, we need an established iPSC centre of excellence in Africa which can serve as a launchpad for the diffusion of skills and technologies to promote the uptake of hiPSC technologies across the continent. Through training workshops and visiting researcher programmes this could be expanded across the continent to establish strategically positioned regional nodes creating an African-led hiPSC network to

drive the adequate representation of African genetic diversity in global stem cell repositories.

We acknowledge that the lack of adequate representation within global hiPSC resources available to the academic and industry community is not an African problem alone—investment is required to ensure that global population diversity is fairly represented within current repositories as this is critical to assuring equitable benefit and impact from advances within the hiPSC field for all population groups.

With the right partners, the localisation of such endeavours on the continent would strive to build capability and capacity, whilst ensuring that Africa's extraordinary genetic diversity is harnessed for global accessibility and consequently, impact.

Tracey Hurrell ^{1,4}, Jerolen Naidoo ^{1,2,4}, Tiro Ntlhahfu¹ & Janine Scholefield ^{1,2,3} 

¹Bioengineering and Integrated Genomics Group, Council for Scientific and Industrial Research, Pretoria, South Africa. ²Department of Human Biology, University of Cape Town, Cape Town, South Africa. ³Division of Human Genetics, University of the Witwatersrand, Johannesburg, South Africa. ⁴These authors contributed equally: Tracey Hurrell, Jerolen Naidoo.  e-mail: jscholefield@csir.co.za

Received: 24 January 2024; Accepted: 16 September 2024;
Published online: 03 October 2024

References

- Zaaijer, S. & Capes-Davis, A. Ancestry matters: building inclusivity into preclinical study design. *Cell* **184**, 2525–2531 (2021).
- Ghosh, S., Nehme, R. & Barrett, L. E. Greater genetic diversity is needed in human pluripotent stem cell models. *Nat. Commun.* **13**, 7301 (2022).
- Horwitz, R., Riley, E. A. U., Millan, M. T. & Gunawardane, R. N. It's time to incorporate diversity into our basic science and disease models. *Nat. Cell Biol.* **23**, 1213–1214 (2021).
- Guhr, A. et al. Recent trends in research with human pluripotent stem cells: impact of research and use of cell lines in experimental research and clinical trials. *hPSCreg*. https://hpscereg.eu/clinical_studies (2018).
- Yoshida, S. et al. A clinical-grade HLA haplobank of human induced pluripotent stem cells matching approximately 40% of the Japanese population. *Med* **4**, 51–66.e10 (2023).
- Arrieta-Bolaños, E., Hernández-Zaragoza, D. I. & Barquera, R. An HLA map of the world: a comparison of HLA frequencies in 200 worldwide populations reveals diverse patterns for class I and class II. *Front. Genet.* **14**, 866407 (2023).
- Tshabalala, M., Mellet, J. & Pepper, M. S. Human leukocyte antigen diversity: a Southern African perspective. *J. Immunol. Res.* **2015**, 746151 (2015).
- Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Choudhury, A. et al. High-depth African genomes inform human migration and health. *Nature* **586**, 741–748 (2020).
- Stats | gnomAD. <https://gnomad.broadinstitute.org/stats>.
- Advanced Search. *WiCell* <https://www.wicell.org/advanced-search.cmsx>.
- HipSci | Human Induced Pluripotent Stem Cell Initiative. <https://www.hipsci.org/>.
- Home · hPSCreg. <https://hpscereg.eu/map>.
- Coriell Institute for Medical Research. https://www.coriell.org/Search?grid=0&q=ipsc&csld=&f_2=iPSC.
- Ludwig, T. E. et al. ISSCR standards for the use of human stem cells in basic research. *Stem Cell Rep.* **18**, 1744–1752 (2023).
- Adigwe, O. P., Onoja, S. O. & Onavbavba, G. A critical review of sickle cell disease burden and challenges in sub-Saharan Africa. *J. Blood Med.* **14**, 367–376 (2023).
- Minja, N. W. et al. Cardiovascular diseases in Africa in the twenty-first century: gaps and priorities going forward. *Front. Cardiovasc. Med.* **9**, 1008335 (2022).
- Tishkoff, S. A. et al. The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
- Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field* (National Academies Press, Washington, 2023).
- da Rocha, J. et al. G6PD distribution in sub-Saharan Africa and potential risks of using chloroquine/hydroxychloroquine based treatments for COVID-19. *Pharmacogenomics J.* **21**, 649–656 (2021).
- Stewart, C. & Pepper, M. S. Cystic fibrosis on the African continent. *Genet. Med.* **18**, 653–662 (2016).
- Sengupta, D. et al. Genetic substructure and complex demographic history of South African Bantu speakers. *Nat Commun.* **12**, 2080 (2021).
- Fortes-Lima, C. A. et al. The genetic legacy of the expansion of Bantu-speaking peoples in Africa. *Nature* **625**, 540–547 (2024).
- Harvard WorldMap. <https://worldmap.maps.arcgis.com/home/index.html>.
- Language Families Felix—overview. <https://www.arcgis.com/home/item.html?id=053aa06015fd45abb9893263d41e3fe4>.
- GREG (Geo-referencing of Ethnic Groups)—overview. <https://www.arcgis.com/home/item.html?id=27814e8c824248a784fbf57e303cf4c5>.
- Weidmann, N. B., Rød, J. K. & Cederman, L.-E. Representing ethnic groups in space: a new dataset. *J. Peace Res.* **47**, 491–499 (2010).

Acknowledgements

The authors acknowledge the following funding, Department of Science and Innovation, V6YBG65, J.S. We also thank Marc Leo Felix for permission to reproduce the “Language Families Felix” dataset using WorldMap.

Author contributions

Conception and writing: J.S., T.H., and J.N. Data analysis: J.S. and T.N.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Janine Scholefield.

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