Polygenic prediction of type 2 diabetes in Africa

Tinashe Chikowore[1,2], Kenneth Ekoru[3], Marijana Vujkovic[4,] Dipender Gill[5,12], Fraser Pirie[6], Elizabeth Young[7], Manjinder S Sandhu[8], Mark McCarthy[9], Charles Rotimi[3], Adebowale Adeyemo[3], Ayesha Motala[6] and Segun Fatumo[10,11]

[1] MRC/Wits Developmental Pathways for Health Research Unit. Department of Pediatrics. Faculty of Health Sciences. University of the Witwatersrand. Johannesburg. South Africa. [2]Sydney Brenner Institute for Molecular Bioscience. Faculty of Health Sciences. University of the Witwatersrand. Johannesburg. South Africa. [3]Center for Research on Genomics and Global Health, National Institute of Health, Bethesda, MD, USA [4]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA. [5] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom [6]Department of Diabetes and Endocrinology, University of KwaZulu-Natal, Durban, 4013 South Africa. [7]Omnigen Biodata Ltd, Cambridge. [8]Department of Epidemiology & Biostatistics, Imperial College, London. *[9]Wellcome Trust Centre for Human Genetics, University of Oxford, UK. [10]London School of Hygiene and Tropical Medicine, London, UK.[11]The African Computational Genomics (TACG) Research Group, MRC/UVRI and LSHTM (Uganda Research Unit), Entebbe, Uganda. [12]Clinical Pharmacology and Therapeutics Section, Institute of Medical and Biomedical Education and Institute for Infection and Immunity, St George's, University of London, London, UK.*

Correspondence:

Segun Fatumo: segun.fatumo@lshtm.ac.uk

Tinashe Chikowore: tinashe.chikowore@wits.ac.za

Abstract

Objective. Polygenic prediction of type 2 diabetes in continental Africans is adversely affected by the limited number of genome-wide association studies (GWAS) of type 2 diabetes from Africa and the poor transferability of European derived polygenic risk scores (PRS) in diverse ethnicities. We set out to evaluate if African American, European or multi-ethnic derived PRSs would improve polygenic prediction in continental Africans.

Research Design and Methods. Using the PRSice software, ethnic-specific PRSs were computed with weights from the type 2 diabetes GWAS multi-ancestry meta-analysis of 228,499 cases and 1,178,783 controls. The South African Zulu study (1602 cases and 981 controls) was used as the target data set. Validation and assessment of the best predictive PRS association with age at diagnosis was done in the Africa America Diabetes Mellitus (AADM) study (2148 cases and 2161 controls).

Results. The discriminatory ability of the African American and Multi-ethnic PRS were similar. However**,** the African American derived PRS was more transferable in all the countries represented in the AADM cohort, and predictive of type 2 diabetes in the country combined analysis compared to the European and multi-ethnic derived scores. Notably, participants in the 10^{th} decile of this PRS had a 3.63-fold greater risk (OR 3.63; 95%CI (2.19 - 4.03), $p = 2.79 \times 10^{-17}$) per risk allele of developing diabetes and were diagnosed 2.6 years earlier compared to those in the first decile.

Conclusions African American derived PRS enhances polygenic prediction of type 2 diabetes in continental Africans. Improved representation of non-European populations (including Africans) in GWAS promises to provide better tools for precision medicine interventions in type 2 diabetes.

Keywords: Africans, PRS, type 2 diabetes

Introduction

The global prevalence of diabetes mellitus in 2019 was estimated to be 463 million individuals(1), of which 19.4 million were from Africa. Type 2 diabetes is the most common form of diabetes in Africa, accounting for 90% of the cases. African countries are adversely affected by limited resources to manage this burden. Nonetheless, by 2045 it is projected that Africa will experience the largest increase in diabetes prevalence in the world of 143% (1; 2). In addition, the highest proportion of undiagnosed (59.7%) people living with diabetes in the world reside in Africa(1). Therefore, urgent strategies and resources for improving screening and early identification interventions are required to help curb this pandemic in Africa.

Type 2 diabetes is a multifactorial disease that is hypothesised to be increasing in prevalence due to the interaction of genetic and environmental factors(3). Although the genetic factors are stable over time, the surge in diabetes prevalence over the past decades is thought to be caused by urbanization and the adoption of westernized lifestyles characterised by consumption of energy-dense foods and physical inactivity(3; 4). However, diabetes has been noted to be preventable, and its onset delayed for 15 years by diet and exercise interventions in the Diabetes Prevention Program(5). Since diet and exercise strategies are readily accessible and relatively low-cost, coupling these lifestyle interventions with approaches that identify people more susceptible to developing diabetes earlier might effectively lower the diabetes burden. The use of polygenic risk scores for early identification of people that are more genetically susceptible to developing type 2 diabetes is such an approach(6). Recent studies conducted in Europeans have indicated that individuals in the 10^{th} decile have a 5.21-fold higher risk (OR=5.21; 95% CI 4.94–5.49) of developing diabetes compared to those in the first decile(7). However, evidence exists of the poor transferability of European derived polygenic scores in diverse populations. For example, Martin et al. 2019 reported that European PRSs had a 4.9-fold

reduced predictive in African Americans across 17 traits. There is now a concern that African ancestry and other similarly under-studied population groups may not benefit from the clinical translation efforts of these polygenic risk scores and thereby further exacerbate existing health disparities (8; 9).

Large multi-ethnic cohorts such as the Million Veteran Program improve the representation of African Americans in GWAS and offer a promise of enhanced polygenic prediction in this group (10). However, the representation of continental Africans in GWAS is still very low, both in the number of studies and the total number of study participants. For example, Type 2 diabetes GWAS with over a million European participants are being reported, while the sample sizes of continental Africans remain under 10,000 (7; 11). Therefore, continental Africans face a much worse threat than African Americans of under-representation in precision medicine efforts for type 2 diabetes(9). It has been reported that multi-ethnic PRS (compared to European only PRS) might enhance prediction in diverse populations(12; 13). However, the predictive ability of the multi-ethnic derived PRS and that of African Americans who originated mainly from the Western part of Africa and have ~80% Africa admixture is yet to be evaluated in continental Africans (12; 13). We set up this study to assess the predictive ability of European, African-American and multi-ethnic derived polygenic risk scores for type 2 diabetes in continental Africans.

Methods

Study participants

Black South African participants from the Durban Case-Control (DCC) study (1602 cases) that were attending a diabetes clinic in the same location in Durban with the 981 controls from the cross-sectional study Durban Diabetes Study (DDS) were aggregated and collectively

regarded as the South African Zulu study, as indicated elsewhere (11; 14). These individuals were above 18 years, not pregnant, and from urban black African communities in Durban, South Africa(14). The WHO criteria was used to define type 2 diabetes status. The validation study participants were from the AADM study, which has been described in detail elsewhere(15-17). The 2148 cases and 2161 controls from this study were enrolled at university medical centers in Nigeria (1325 cases and 1363 controls), Ghana (449 cases and 435 controls) and Kenya (374cases and 363controls) (17). In this study, diabetes was defined based on an oral glucose tolerance test or pharmacological treatment of diabetes(17) . Written informed consent was completed by the study participants. The respective studies were approved by relevant ethics committees under the following references DCC (BF078/08), DDS (BF030/12) and AADM (14/WM/1061).

Genotyping and Imputation

Participants in the South African Zulu study (Supplementary Table 1) were genotyped using the Illumina Multi-Ethnic Genotyping Array (Illumina, Illumina Way, San Diego, CA, USA). The Affymetrix Axiom PANAFR SNP array or Illumina Multi-Ethnic Genotyping Array was used to genotype participants in the AADM study. Detailed quality control and imputation for these studies was done using African whole genomes from the Uganda 2000 Genomes (UG2G) and the 1000 Genomes as reference panels, as has been described elsewhere (11; 18). A minimum MAF threshold of 0.5% and imputation information score > 0.4 was applied(11).

Statistical Analysis

PRSice 2 software was used to implement the clumping and threshold approach for developing PRS. After sensitivity analysis, a clumping distance of 500kb and r2 of 0.5 were parameters used for computing PRS. GWAS summary statistics from the multi-ancestry GWAS of type 2

diabetes by Vujkovic *et al*., 2020, comprising of participants representative of European, African Americans, Hispanics and Asians(7) were used as the base (discovery), while genotype data from the South African Zulu study and AADM was used as the target data and validation datasets respectively as illustrated in Table 1.

In the discovery analysis, multiple PRS were computed at p-value thresholds from 1 to $5x10^{-8}$ of the base dataset and LD clumping from the target data set. The predictivity of these PRSs was then evaluated through linear models that adjusted for age, sex and population stratification (five principal components). The p-values of these PRS and the Nagelkerke R2 were evaluated to assess transferability and predictability, respectively (Supplement Figure 2-4). The best predictive Multi-ethnic, African American and European PRSs were then validated in the AADM study as shown in Table 1 and Supplementary Table 2.

During the validation stage, the best predictive PRSs were assessed for transferability and predictivity through the p-values and Nagelkerke R2 in linear models implemented in PRSice, which corrected for age, sex, BMI and population stratification (five principal components) as shown in Table 1. This was first done for the whole of the AADM study and then at the country level, as shown in Figure 1B.

The best predictive PRS from the three discovery datasets was then further used to assess its risk stratification and diagnostic utility. Logistic regression models for the PRS deciles as a predictor variable were computed while correcting for age, sex, body mass index (BMI) and residual population structure using principal components (five principal components). A shape plot was computed to show the differences in risk of the PRS deciles from the first, as shown

in Figure 1A. Finally, a linear regression model was used to evaluate whether the age of diagnosis in patients with diabetes (n=1031) is affected by PRS in the AADM study .

Results

Polygenic score development and validation

From the linear models of the multiple PRSs generated using the PRSice software (Supplementary Figure 2-4), the best predictive PRS from the Europeans, Multiethnic, and African Americans was significant and had the highest variance as indicated by Nagelkerke R2 of 0.69% ($p = 5.09x10^{-6}$), 0.69% ($p = 3.90x10^{-9}$) and 1.11% ($p = 4.62x10^{-6}$) respectively (Table 1). The best PRSs were validated in the AADM study and noted to be all significant in a similar trend. The African American PRS had the highest predictability indicated by Nagelkerke R2 of 2.92% ($9.38 x10^{-24}$) in the combined analysis of the countries, as illustrated in Table 1.

Polygenic risk score stratification and transferability in African countries

The participants in the 10^{th} decile of the African American derived PRS had a more than 3-fold higher risk for developing type 2 diabetes per risk allele, compared to those in the first decile in the AADM study OR 3.63 (95%CI (2.19 - 4.03), ;$p = 2.79 x 10^{-17}$) (Figure 1A). On average, participants in the 10^{th} decile of the African American PRS in the AADM study were diagnosed with type 2 diabetes 2.6 years earlier (Beta = -2.61; p = 0.046) than participants in the first decile (Figure 2B). The African American PRS was transferable in all countries compared to the multi-ethnic that was not in Kenya. The predictability (indicated by Nagelkerke R2) varied greatly between the East African country of Kenya and the two West African countries Ghana

and Nigeria, where it was much higher for both the African-American and the multi-ethnic PRSs.

Discriminatory ability of the polygenic risk score

The model with the conventional risk factors of age, BMI, five PCs and sex had an area under the curve (AUC)/C -statistic of 67.9% while that of the African American PRS, five PCs, age, BMI and sex was 69.8% (Figure 2) almost similar to the multi-ethnic PRS of multi-ethnic of 69.9%. There was therefore improved discriminatory ability by 1.9%, with the addition of the African American PRS to the conventional risk factors.

Conclusions

Our study set out to assess the predictive value of type 2 diabetes PRS in continental Africans. In this study, we set out to compare the polygenic prediction of African American, European and multi-ethnic PRSs for type 2 diabetes in continental Africans. The PRS with the best prediction was derived from an African American restricted GWAS(7). Participants in the 10^{th} decile of this PRS had a more than 3-fold increased risk of developing type 2 diabetes and were diagnosed 2.6 years earlier on average than those in the first decile.

Limited studies of candidate SNP PRS have been performed in continental Africans. Previously we reported a genetic risk score with weights from Europeans that was associated with OR = 1.21, 95%CI (1.02–1.43) for type 2 diabetes in black South Africans(19). This GRS had an AUC of 0.665 together with conventional risk factors for type 2 diabetes (19). However, this study was limited due to the small sample size (n = 356), the availability of only genotyped SNPs, and the use of weights that were derived from European-only studies. In our current

study, we have substantially expanded the sample size (n = 2383), enhanced genome coverage by imputing to 1000 Genomes and local African Ancestry whole genomes(18), and used a multi-ethnic discovery dataset GWAS that included 1.4 million individuals, which had people of African American ancestry. We performed a country-level analysis which showed less variable predictability within regional countries in West Africa, Ghana and Nigeria and greater variability when comparing with other countries from other regions, such as Kenya in East Africa. This phenomenon is suggestive of the usefulness of regional PRS in Africa. However, this will need to be validated by additional studies.

Nonetheless, polygenic predictions of European derived PRS in Europeans are still higher than that of the African Americans in continental Africans(7). Notably, participants in the top decile of a European derived PRS have recently been reported to have a greater than 5-fold risk for developing type 2 diabetes than those in the first decile in Europeans(7). Failure to reach predictions denoted in Europeans might be due to that in our study, the African American derived PRS are from an admixed population group that is not representative of the genetic diversity and linkage disequilibrium patterns of continental Africans(13; 20). In addition, vast improvements in sizes of the European cohorts that are now over a million individuals is indicative of substantial power compared to African diabetes cohorts that are still below the 10 thousand mark (21). More investments are thus required to increase the representation of continental Africans in GWAS of type 2 diabetes.

Recently, it was reported that the multi-ancestry PRS outperforms the population-specific ones from Europeans and East Asians (22). However, this phenomenon is yet to be validated in continental Africans. Considering that 80% of GWAS have been done in Europeans, most multi-ancestry GWAS meta-analyses are biased towards this population group (8). Another

paper by Marquez-Luna *et al.*, 2017 combined the training and the target dataset summary statistics to compute the PRS and then showed that the multi-ethnic PRS improve prediction in diverse populations(12). However, since this approach is not widely accepted and more research is still required to validate if the multi-ethnic PRS outperforms the population-specific PRS for all the ancestries(23; 24). In our study, the African American and Multiethnic PRS had similar discriminatory abilities. However, the African American PRS was slightly more predictive than the multi-ancestry for the combined AADM study and with improved representations of Africans, these predictions might increase in the future. In addition, the country stratified analyses also indicated that the multi-ancestry PRS was not transferable to participants from Kenya. The failure to tag the causal variant due to differences in allele frequencies, LD patterns, and heterogeneity of effect sizes is a potential reason for the limited predictivity of multi-ancestry meta-analysis in continental Africans that have greater genetic diversity(25-27).

The utility of polygenic risk scores is an issue of paramount importance for clinical translation(6). The African American PRS, though it was predictive for type 2 diabetes in continental Africans, only improved the AUC of conventional risk factors by 1.9%, and when combined with PCs, its AUC was 69.8%, while that of the conventional risk factors was 67.9%. Similarly, in a Swedish type 2 diabetes study, the European derived PRS increased the AUC by 1% compared to conventional risk factors (28). However, the use of AUC as a measure to evaluate the clinical utility of polygenic prediction is being debated, as it is regarded as a less sensitive metric(29). There are ongoing efforts to develop better metrics (30). Nonetheless, findings from this study that people with type 2 diabetes and a high PRS are typically diagnosed at an earlier age and have a 3.6-fold risk of developing diabetes are of clinical importance. They may be useful in the prevention and treatment of diabetes.

Our study was limited by the limited number of GWAS of type 2 diabetes of continental Africans. Nonetheless, the African American derived PRS improved disease classification in this population. The clumping and thresholding approach used to compute the genome-wide PRS did not account for environmental factors such as diet and exercise that might confound the predictive accuracy of these measures. The strengths of our study include validation of the African American PRS in the AADM study and the fact that we used GWAS summary statistics of varied ethnicities from the same study, which minimized bias due to genotyping and GWAS designs.

In summary, an African American derived PRS seems to be the best predictor for type 2 diabetes in continental Africans compared to a European and multi-ethnic PRS. More studies are required to determine whether using continental African GWAS might further enhance these predictions and reach a similar accuracy as in Europeans. Although the PRS prediction of diabetes had low specificity and sensitivity, patient stratification by PRS may prove clinically useful.

Acknowledgements

TC and SF conceptulised the study, performed the main analyses and wrote the first draft. KE and AA performed the validation analysis. All the authors read and provided critical feedback on the paper. TC is an international training fellow supported by the Wellcome Trust grant (214205/Z/18/Z). SF is an international Intermediate Fellow funded by the Wellcome Trust grant (220740/Z/20/Z) at the MRC/UVRI and LSHTM. DG was supported by the British Heart Foundation Centre of Research Excellence (RE/18/4/34215) at Imperial College, and a National Institute for Health Research Clinical Lectureship (CL-2020-16-001) at St. George's,

University of London. The AADM study was supported in part by the Intramural Research Program of the National Institutes of Health in the Centre for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). Support for participant recruitment and initial genetic studies of the AADM study was provided by NIH grant No. 3T37TW00041-03S2 from the Office of Research on Minority Health.

Dr. Segun Fatumo is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis

Conflicts of interest

DG is employed part-time by Novo Nordisk and has received consultancy fees from Policy Wisdom.

No potential conflicts of interest relevant to this article were reported by all other authors.

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes research and clinical practice 2019;157:107843
2. Ekoru K, Doumatey A, Bentley AR, Chen G, Zhou J, Shriner D, Fasanmade O, Okafor G, Eghan B, Agyenim-Boateng K, Adeleye J, Balogun W, Amoah A, Acheampong J, Johnson T, Oli J, Adebamowo C, Collins F, Dunston G, Adeyemo A, Rotimi C. Type 2 diabetes complications and comorbidity in Sub-Saharan Africans. EClinicalMedicine 2019;16:30-41
3. Langenberg C, Lotta LA. Genomic insights into the causes of type 2 diabetes. Lancet 2018;391:2463-2474
4. Hansen T. Type 2 diabetes mellitus--a multifactorial disease. Ann Univ Mariae Curie Sklodowska Med 2002;57:544-549
5. Diabetes Prevention Program Research G. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. The lancet Diabetes & endocrinology 2015;3:866-875
6. McCarthy MI, Mahajan A. The value of genetic risk scores in precision medicine for diabetes. Expert Review of Precision Medicine and Drug Development 2018;3:279-281
7. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, Huffman JE, Assimes TL, Lorenz K, Zhu X, Hilliard AT, Judy RL, Huang J, Lee KM, Klarin D, Pyarajan S, Danesh J, Melander O, Rasheed A, Mallick NH, Hameed S, Qureshi IH, Afzal MN, Malik U, Jalal A, Abbas S, Sheng X, Gao L, Kaestner KH, Susztak K, Sun YV, DuVall SL, Cho K, Lee JS, Gaziano JM, Phillips LS, Meigs JB, Reaven PD, Wilson PW, Edwards TL, Rader DJ, Damrauer SM, O'Donnell CJ, Tsao PS, Atkinson MA, Powers AC, Naji A, Kaestner KH, Abecasis GR, Baras A, Cantor MN, Coppola G, Economides AN, Lotta LA, Overton JD, Reid JG, Shuldiner AR, Beechert C, Forsythe C, Fuller ED, Gu Z, Lattari M, Lopez AE, Schleicher TD, Padilla MS, Toledo K, Widom L, Wolf SE, Pradhan M, Manoochehri K, Ulloa RH, Bai X, Balasubramanian S, Barnard L, Blumenfeld AL, Eom G, Habegger L, Hawes A, Khalid S, Maxwell EK, Salerno WJ, Staples JC, Yadav A, Jones MB, Mitnaul LJ, Aguayo SM, Ahuja SK, Ballas ZK, Bhushan S, Boyko EJ, Cohen DM, Concato J, Constans JI, Dellitalia LJ, Fayad JM, Fernando RS, Florez HJ, Gaddy MA, Gappy SS, Gibson G, Godschalk M, Greco JA, Gupta S, Gutierrez S, Hammer KD, Hamner MB, Harley JB, Hung AM, Huq M, Hurley RA, Iruvanti PR, Ivins DJ, Jacono FJ, Jhala DN, Kaminsky LS, Kinlay S, Klein JB, Liangpunsakul S, Lichy JH, Mastorides SM, Mathew RO, Mattocks KM, McArdle R, Meyer PN, Meyer LJ, Moorman JP, Morgan TR, Murdoch M, Nguyen X-MT, Okusaga OO, Oursler K-AK, Ratcliffe NR, Rauchman MI, Robey RB, Ross GW, Servatius RJ, Sharma SC, Sherman SE, Sonel E, Sriram P, Stapley T, Striker RT, Tandon N, Villareal G, Wallbom AS, Wells JM, Whittle JC, Whooley MA, Xu J, Yeh S-S, Aslan M, Brewer JV, Brophy MT, Connor T, Argyres DP, Do NV, Hauser ER, Humphries DE, Selva LE, Shayan S, Stephens B, Whitbourne SB, Zhao H, Moser J, Beckham JC, Breeling JL, Romero JPC, Huang GD, Ramoni RB, Pyarajan S, Sun YV, Cho K, Wilson PW, O'Donnell CJ, Tsao PS, Chang K-M, Gaziano JM, Muralidhar S, Chang K-M, Voight BF, Saleheen D, The HC, Regeneron Genetics C, Program VAMV. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nature genetics 2020;52:680-691

8. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nature genetics 2019;51:584-591
9. Doumatey AP, Ekoru K, Adeyemo A, Rotimi CN. Genetic Basis of Obesity and Type 2 Diabetes in Africans: Impact on Precision Medicine. Curr Diabetes Rep 2019;19:105
10. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, O'Leary TJ. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. Journal of Clinical Epidemiology 2016;70:214-223
11. Chen J, Sun M, Adeyemo A, Pirie F, Carstensen T, Pomilla C, Doumatey AP, Chen G, Young EH, Sandhu M, Morris AP, Barroso I, McCarthy MI, Mahajan A, Wheeler E, Rotimi CN, Motala AA. Genome-wide association study of type 2 diabetes in Africa. Diabetologia 2019;62:1204-1211
12. Márquez-Luna C, Loh PR, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol 2017;41:811-823
13. Zakharia F, Basu A, Absher D, Assimes TL, Go AS, Hlatky MA, Iribarren C, Knowles JW, Li J, Narasimhan B, Sidney S, Southwick A, Myers RM, Quertermous T, Risch N, Tang H. Characterizing the admixed African ancestry of African Americans. Genome Biology 2009;10:R141
14. Hird TR, Young EH, Pirie FJ, Riha J, Esterhuizen TM, O'Leary B, McCarthy MI, Sandhu MS, Motala AA. Study profile: the Durban Diabetes Study (DDS): a platform for chronic disease research. Global Health, Epidemiology and Genomics 2016;1:e2
15. Rotimi CN, Chen G, Adeyemo AA, Furbert-Harris P, Parish-Gause D, Zhou J, Berg K, Adegoke O, Amoah A, Owusu S, Acheampong J, Agyenim-Boateng K, Eghan BA, Jr., Oli J, Okafor G, Ofoegbu E, Osotimehin B, Abbiyesuku F, Johnson T, Rufus T, Fasanmade O, Kittles R, Daniel H, Chen Y, Dunston G, Collins FS, Africa America Diabetes Mellitus S. A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. Diabetes 2004;53:838-841
16. Rotimi CN, Dunston GM, Berg K, Akinsete O, Amoah A, Owusu S, Acheampong J, Boateng K, Oli J, Okafor G, Onyenekwe B, Osotimehin B, Abbiyesuku F, Johnson T, Fasanmade O, Furbert-Harris P, Kittles R, Vekich M, Adegoke O, Bonney G, Collins F. In Search of Susceptibility Genes for Type 2 Diabetes in West Africa: The Design and Results of the First Phase of the AADM Study. Annals of Epidemiology 2001;11:51-58
17. Adeyemo AA, Tekola-Ayele F, Doumatey AP, Bentley AR, Chen G, Huang H, Zhou J, Shriner D, Fasanmade O, Okafor G, Eghan B, Jr., Agyenim-Boateng K, Adeleye J, Balogun W, Elkahloun A, Chandrasekharappa S, Owusu S, Amoah A, Acheampong J, Johnson T, Oli J, Adebamowo C, Collins F, Dunston G, Rotimi CN. Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans. Front Genet 2015;6:335
18. Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, Prado-Martinez J, Bouman H, Abascal F, Haber M, Tachmazidou I, Mathieson I, Ekoru K, DeGorter MK, Nsubuga RN, Finan C, Wheeler E, Chen L, Cooper DN, Schiffels S, Chen Y, Ritchie GRS, Pollard MO, Fortune MD, Mentzer AJ, Garrison E, Bergström A, Hatzikotoulas K, Adeyemo A, Doumatey A, Elding H, Wain LV, Ehret G, Auer PL, Kooperberg CL, Reiner AP, Franceschini N, Maher D, Montgomery SB, Kadie C, Widmer C, Xue Y, Seeley J, Asiki G, Kamali A, Young EH, Pomilla C, Soranzo N, Zeggini E, Pirie F, Morris AP, Heckerman D, Tyler-Smith C, Motala AA, Rotimi C, Kaleebu P,

Barroso I, Sandhu MS. Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. Cell 2019;179:984-1002.e1036
19. Chikowore T, van Zyl T, Feskens EJ, Conradie KR. Predictive utility of a genetic risk score of common variants associated with type 2 diabetes in a black South African population. Diabetes research and clinical practice 2016;122:1-8
20. Choudhury A, Aron S, Botigué LR, Sengupta D, Botha G, Bensellak T, Wells G, Kumuthini J, Shriner D, Fakim YJ, Ghoorah AW, Dareng E, Odia T, Falola O, Adebiyi E, Hazelhurst S, Mazandu G, Nyangiri OA, Mbiyavanga M, Benkahla A, Kassim SK, Mulder N, Adebamowo SN, Chimusa ER, Muzny D, Metcalf G, Gibbs RA, Matovu E, Bucheton B, Hertz-Fowler C, Koffi M, Macleod A, Mumba-Ngoyi D, Noyes H, Nyangiri OA, Simo G, Simuunza M, Rotimi C, Ramsay M, Choudhury A, Aron S, Botigué L, Sengupta D, Botha G, Bensellak T, Wells G, Kumuthini J, Shriner D, Fakim YJ, Ghoorah AW, Dareng E, Odia T, Falola O, Adebiyi E, Hazelhurst S, Mazandu G, Nyangiri OA, Mbiyavanga M, Benkahla A, Kassim SK, Mulder N, Adebamowo SN, Chimusa ER, Rotimi C, Ramsay M, Adeyemo AA, Lombard Z, Hanchard NA, Adebamowo C, Agongo G, Boua RP, Oduro A, Sorgho H, Landouré G, Cissé L, Diarra S, Samassékou O, Anabwani G, Matshaba M, Joloba M, Kekitiinwa A, Mardon G, Mpoloka SW, Kyobe S, Mlotshwa B, Mwesigwa S, Retshabile G, Williams L, Wonkam A, Moussa A, Adu D, Ojo A, Burke D, Salako BO, Matovu E, Bucheton B, Hertz-Fowler C, Koffi M, Macleod A, Mumba-Ngoyi D, Noyes H, Nyangiri OA, Simo G, Simuunza M, Awadalla P, Bruat V, Gbeha E, Adeyemo AA, Lombard Z, Hanchard NA, Trypano GENRG, Consortium HA. High-depth African genomes inform human migration and health. Nature 2020;586:741-748
21. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature genetics 2018;50:1219-1224
22. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, Matsunaga H, Ieki H, Ozaki K, Onouchi Y, Takahashi A, Nomura S, Morita H, Akazawa H, Kim C, Seo JS, Higasa K, Iwasaki M, Yamaji T, Sawada N, Tsugane S, Koyama T, Ikezaki H, Takashima N, Tanaka K, Arisawa K, Kuriki K, Naito M, Wakai K, Suna S, Sakata Y, Sato H, Hori M, Sakata Y, Matsuda K, Murakami Y, Aburatani H, Kubo M, Matsuda F, Kamatani Y, Komuro I. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nature genetics 2020;52:1169-1177
23. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nature Protocols 2020;15:2759-2772
24. Babb de Villiers C, Kroese M, Moorthie S. Understanding polygenic models, their development and the potential application of polygenic scores in healthcare. Journal of Medical Genetics 2020;57:725
25. Morris AP. Transethnic meta-analysis of genomewide association studies. Genet Epidemiol 2011;35:809-822
26. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B. Analysis of polygenic risk score usage and performance in diverse human populations. Nature communications 2019;10:3328
27. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17:392-406
28. Lyssenko V, Laakso M. Genetic Screening for the Risk of Type 2 Diabetes. Worthless or valuable? 2013;36:S120-S126

29. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928-935
30. Baker SG. Metrics for Evaluating Polygenic Risk Scores. JNCI Cancer Spectrum 2020;5

Table 1 Comparisons of the predictive ability of ethnically derived PRS on type 2 diabetes in continental Africans

	Multi-ethnic	African American	European
Discovery Dataset (Multi-ancestry meta-analysis)			
Cases	228,499	24,646	148,726
Controls	1,178,783	31,446	965,732
PRS Development			
Target Data Set (SA Zulu)			
Cases	1,602	1,602	1,602
Controls	981	981	981
PRS parameters			
P-value threshold	3×10^{-4}	5×10^{-8}	0.0608
Number of SNPs	41,815	65	405,572
Nagelkerke R2 %	0.69	1.11	0.69
P-value	4.62×10^{-6}	3.90×10^{-9}	5.09×10^{-6}
*OR(95%CI)	1.29 (1.16-1.43)	1.58 (1.36-1.84)	1.01 (1.00-1.01)
*P-value	3.52×10^{-6}	4.80×10^{-9}	9.54×10^{-6}
Validation of PRS			
Validation data set (AADM)			
Cases	2148	2148	2148
Controls	2161	2161	2161
PRS parameters			
P-value threshold	3×10^{-4}	5×10^{-8}	0.0608
Number of SNPs	41,553	65	1,408,065
Nagelkerke R2 %	2.62	2.92	0.13
P-value	1.06×10^{-21}	9.38×10^{-24}	2.99×10^{-2}
*OR(95%CI)	1.04 (1.03-1.05)	1.57 (1.47-1.67)	1.004 (1.03-1.05)
*P-value	1.41×10^{-21}	5.91×10^{-23}	3.16×10^{-2}

***models adjusted for ancestry indicated by 5 principal components, age, sex and BMI; OR = odds ratio; CI = confidence interval.**

Figure 1 A. Shape plot for the difference in odds ratio for type 2 diabetes (adjusted for age, sex, BMI and five principal components) in reference to the 1st decile for the African (African American), in the AADM study. **B. Bar** plots showing the transferability of the in African countries represented in the AADM study.

Figure 2. A. Receiver operating curves for the African Americans derived PRS and conventional risk factors for the prediction of type 2 diabetes in the AADM study. Abbreviations; AUC= area under the curve, 5PCs = five principal components ,full model

=Age, sex, BMI, AFR PRS,5 PCs. **B.** Shape plot for the difference of age at diagnosis for type 2 diabetes in the AADM study for the African American derived PRS.