

# Introduction to Polygenic Risk Score

***Segun Fatumo PhD, FHEA***

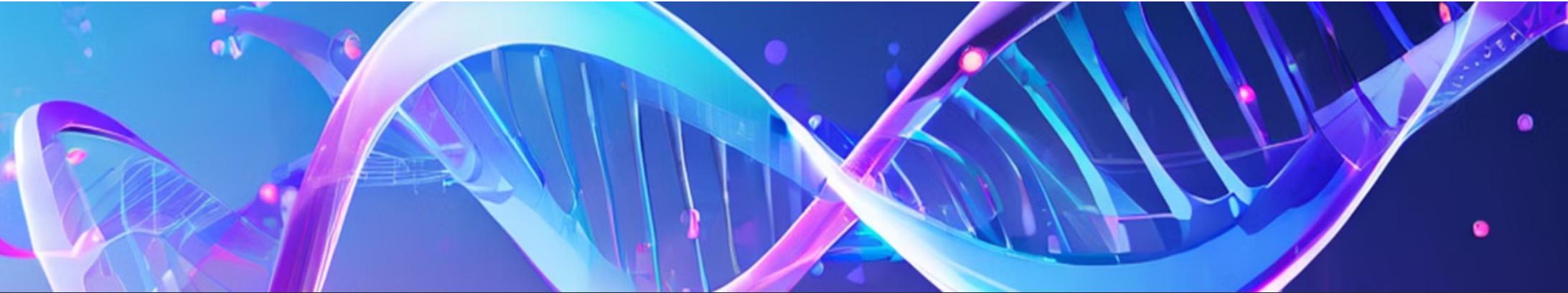
*Professor and Chair of Genomic Diversity*

*Precision Healthcare University Research Institute  
Queen Mary University of London, United Kingdom*

*Head, NCD Genomics  
MRC/UVRI and LSHTM Uganda Research Unit  
Entebbe, Uganda*

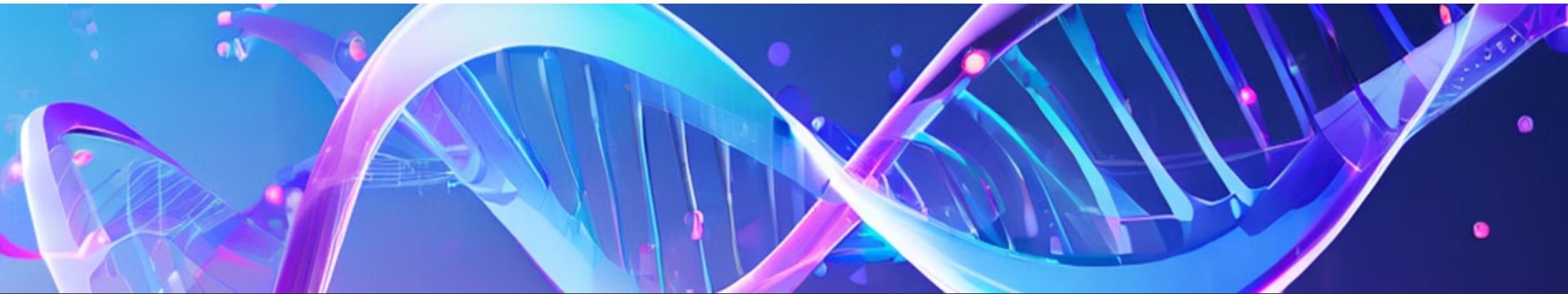
Wellcome Connecting Science Polygenic Risk Score Analysis, 2025, Uganda





## What is a Polygenic Risk Score?

- A polygenic score (PGS) or polygenic risk score (PRS) or genetic risk score (GRS) is an estimate of an individual's genetic liability to a trait or disease, calculated according to their genotype profile
- Polygenic risk scores (PRS) are a powerful tool in the field of personalized genomics, providing insights into an individual's predisposition to complex diseases by analyzing their genetic makeup.
- This presentation will explore the concept of PRS, how they are calculated, and their potential applications in healthcare.



## What is a Polygenic Risk Score?

### 1 Genetic Predisposition

PRS quantifies an individual's genetic risk for a specific disease or trait, taking into account the combined effect of many genetic variants.

### 2 Complex Diseases

PRS are particularly useful for understanding the genetic basis of complex diseases, which are influenced by multiple genes and environmental factors.

### 3 Personalized Medicine

PRS can be used to identify individuals at high risk for certain diseases, enabling earlier intervention and personalized preventive strategies.



# How is a Polygenic Risk Score Calculated?

1

## Genome-Wide Association Studies (No PRS without GWAS)

Genome-Wide Association Studies (GWAS) identify genetic variants associated with specific traits or diseases.

2

## Weighting and Summing

PRS are calculated by summing the effects of multiple genetic variants, weighted by their associated risk estimates from GWAS.

3

## Statistical Modeling

Advanced statistical techniques, such as machine learning, are used to refine the PRS model and improve its predictive accuracy.

# BACKGROUND

**risk allele:** in the context of a disease, this is the allele that confers a risk of developing the disease. Most of the time, risk allele = minor allele, as most people will not carry the risk allele. However, in some case, the risk allele can in fact be the major allele.

- Output from GWAS analysis is a Summary Statistics
- Most important columns are:
  - Chr: Chromosome
  - SNP: SNP Id
  - BP: Basepair
  - EA: Effect Allele
  - NEA: Non Effect Allele or Reference Allele
  - EAF: Effect Allele Frequency
  - Effect: Odd ratio or Beta
  - SE: Standard Er
  - P: Pvalue

Chr	SNP	bp	A1	A2	Freq	b	se	p	K
1	rs2286139	761732	C	T	0.1379	-0.0104056	0.00732416	0.155397	0.5
1	rs12562834	768448	A	G	0.18475	-0.00627592	0.00827054	0.447955	0.5
1	rs4978383	838555	A	C	0.247975	0.00946281	0.00587444	0.107243	0.5
1	rs1806589	853954	C	A	0.3912	0.0152744	0.00523012	0.80349587	0.5
1	rs13382982	861008	A	G	0.018025	-0.0108122	0.0109517	0.341895	0.5
1	rs28576697	870645	C	T	0.29355	0.0116486	0.00556379	0.8362916	0.5
1	rs2348582	882803	A	G	0.05465	0.0119371	0.0111055	0.282426	0.5
1	rs3748594	886384	A	G	0.025975	-0.01244	0.0158797	0.433481	0.5
1	rs28584611	908414	T	C	0.022225	0.00388796	0.0171623	0.828781	0.5
1	rs9777939	929198	A	G	0.03345	-0.00446279	0.0141522	0.752582	0.5
1	rs1091910	932457	A	G	0.2295	-0.00647527	0.00605864	0.285175	0.5
1	rs35940137	940203	A	G	0.050575	-0.10689	0.0115935	2.97533e-20	0.5
1	rs6657048	957640	T	C	0.011825	0.0892934	0.0233322	0.808129691	0.5
1	rs9803031	987200	C	T	0.083875	-0.00434284	0.0091306	0.634334	0.5

# COMPUTING A PRS

- Single variant association analysis has been the primary method in GWAS but requires very large sample sizes to detect more than a handful of SNPs for many complex traits
- In contrast, PRS analysis does not aim to identify individual SNPs but instead aggregates genetic risk across the genome in a single individual polygenic score for a trait of interest (Purcell et al., [2009](#); see Figure [4](#) for a simplified example).
- In this approach, a large discovery sample is required to reliably determine how much each SNP is expected to contribute to the polygenic score (“weights”) of a specific trait.
- Subsequently, in an independent target sample, which can be more modest in size (Dudbridge, [2013](#)), polygenic scores can be calculated based on genetic DNA profiles and these weights

### Discovery GWAS

	Weight*	Risk Allele
SNP1	0.2	A
SNP2	-0.3	C
SNP3	0.1	G

Individual	Alleles SNP1	Alleles SNP2	Alleles SNP3
1	AT	AA	CG
2	AA	CA	GG
3	TT	AC	CG
4	TT	AA	GG
5	TA	CA	GC
6	AT	CA	CG
7	AA	AA	GG
8	AA	CC	CG
9	TA	CC	GC
10	AT	AA	CG

### PRS:

Individual	SNP 1	SNP 2	SNP 3	PRS
1	0.2+0.0	0.0+0.0	0.0+0.1	0.3
2	0.2+0.2	-0.3+0.0	0.1+0.1	0.3
3	0.0+0.0	0.0-0.3	0.0+0.1	-0.2
4	0.0+0.0	0.0+0.0	0.1+0.1	0.2
5	0.0+0.2	-0.3+0.0	0.1+0.0	0.0
6	0.2+0.0	-0.3+0.0	0.0+0.1	0.0
7	0.2+0.2	0.0+0.0	+0.1+0.1	0.6
8	0.2+0.2	-0.3-0.3	0.0+0.1	-0.1
9	0.0+0.2	-0.3-0.3	0.1+0.0	-0.3
10	0.2+0.0	0.0+0.0	0.0+0.1	0.3

## WORKING EXAMPLE OF THREE SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) AGGREGATED INTO A SINGLE INDIVIDUAL POLYGENIC RISK SCORE (PRS).

- ❑ To conduct PRS analysis, trait-specific weights (beta's for continuous traits and the log of the odds ratios for binary traits) are obtained from a discovery GWAS.
- ❑ In the target sample, a PRS is calculated for each individual based on the weighted sum of the number of risk alleles that he or she carries multiplied by the trait-specific weights.
- ❑ For many complex traits, SNP effect sizes are publicly available (e.g., <https://www.med.unc.edu/pgc/downloads>)
- ❑ Although in principle all common SNPs could be used in a PRS analysis, it is customary to first clump (see **clumping**) the GWAS results before computing risk scores
- ❑ p value thresholds are typically used to remove SNPs that show little or no statistical evidence for association

# CONDUCTING POLYGENIC RISK PREDICTION ANALYSES

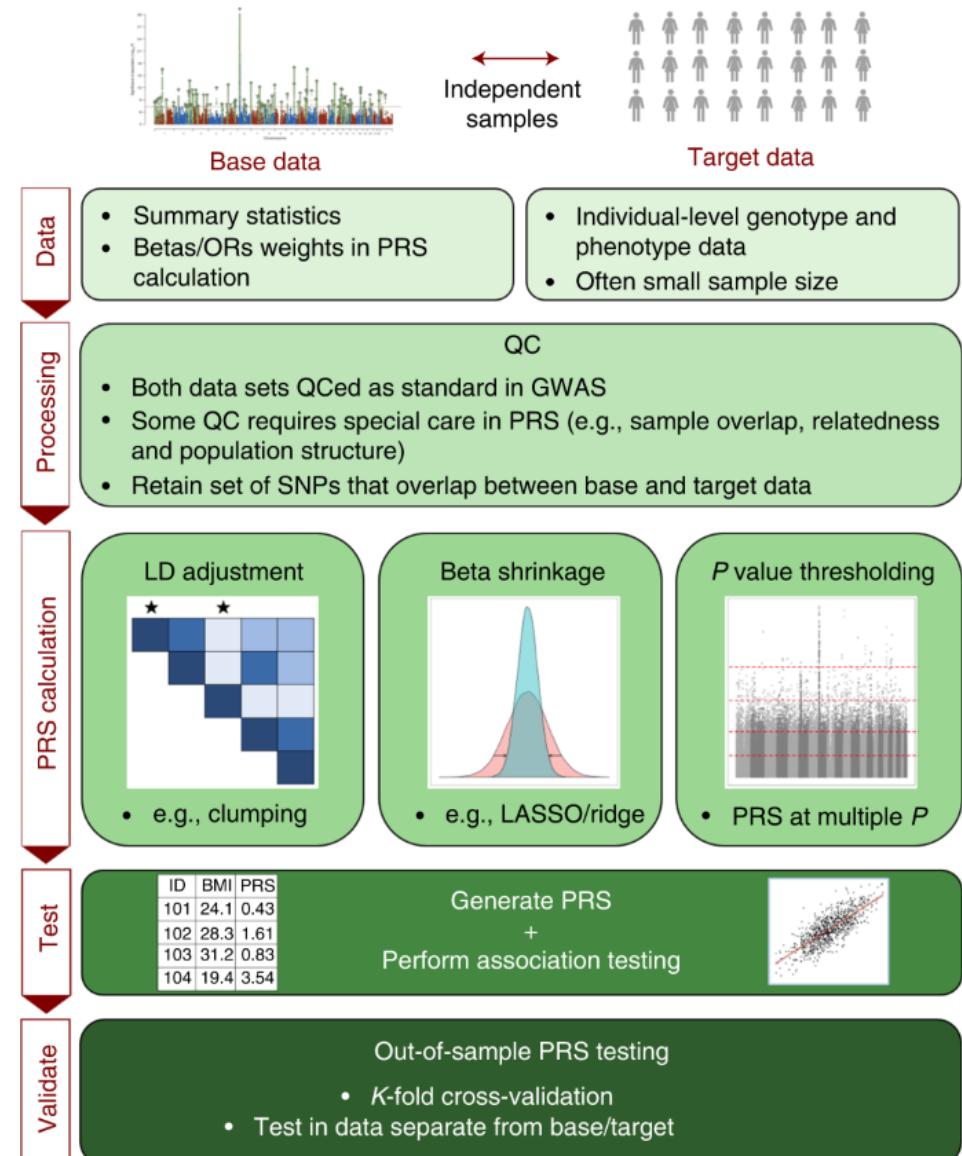
- Once PRS have been calculated for all subjects in the target sample, the scores can be used in a (logistic) regression analysis to predict any trait that is expected to show genetic overlap with the trait of interest.
- The prediction accuracy can be expressed with the (pseudo-) $R^2$  measure of the regression analysis
- It is important to include at least a few PCA or MDS components as covariates in the regression analysis to control for population stratification.
- To estimate how much variation is explained by the PRS, the  $R^2$  of a model that includes only the covariates (e.g., MDS components) and the  $R^2$  of a model that includes covariates + PRS will be compared.
- The increase in  $R^2$  due to the PRS indicates the increase in prediction accuracy explained by genetic risk factors.

# PRS ANALYSIS USING SOFTWARE PRSICE

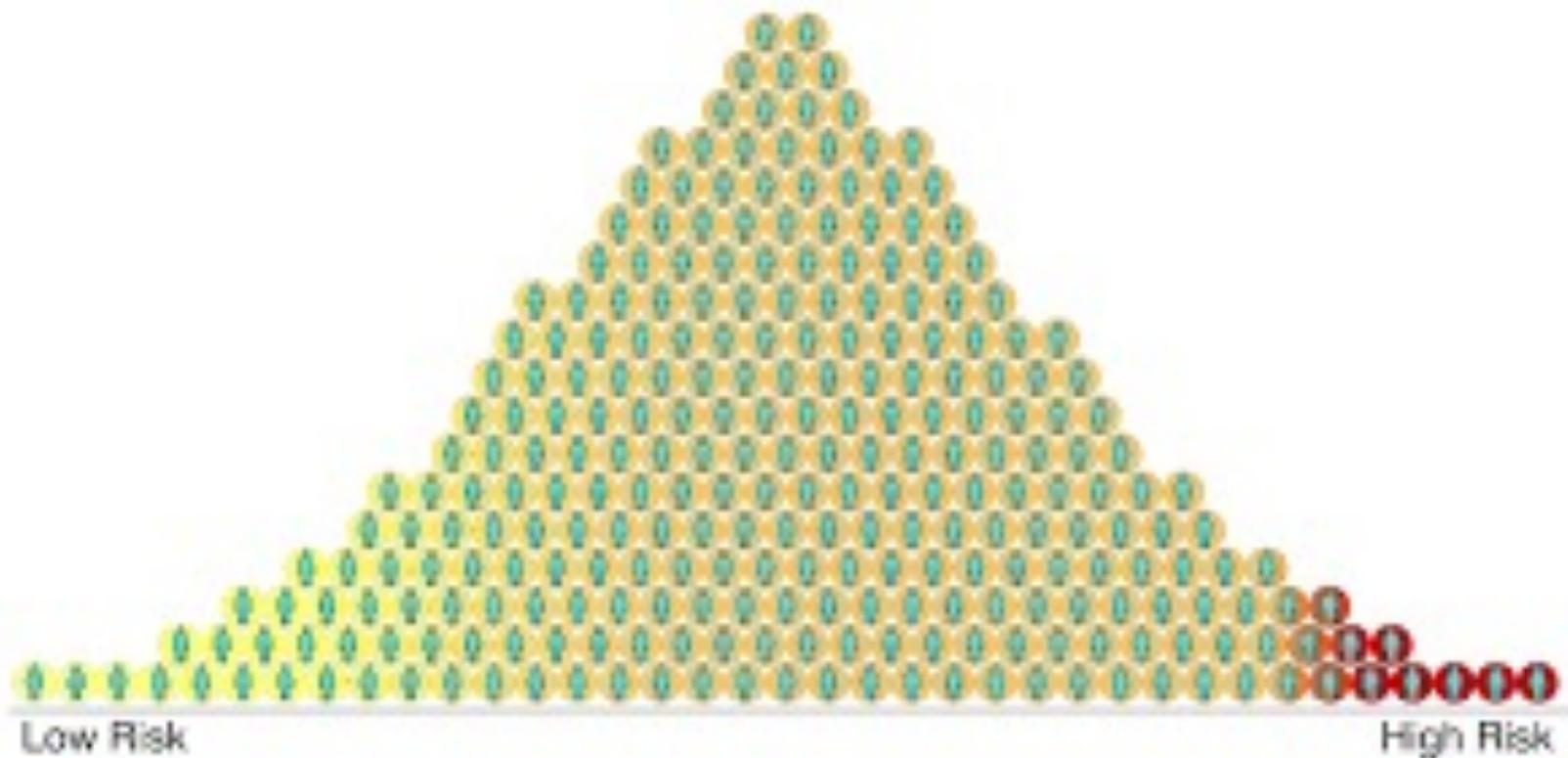
❑ A convenient program to perform PRS analysis is PRSice. It takes care of clumping,  $p$  value thresholds, MDS components, and plots attractive graphs.

❑ There are many other programs for the application of PRS eg,

❑ LDpred, JAMPred, SBLUP, LDpred2-Inf, bigsnpr, LDpred-funct, LDpred2, Lassosum, **BridgePRS**, PRS-CS, **PRS-CSx-auto**, SBayesR, MegaPRS and even PLINK (--score))

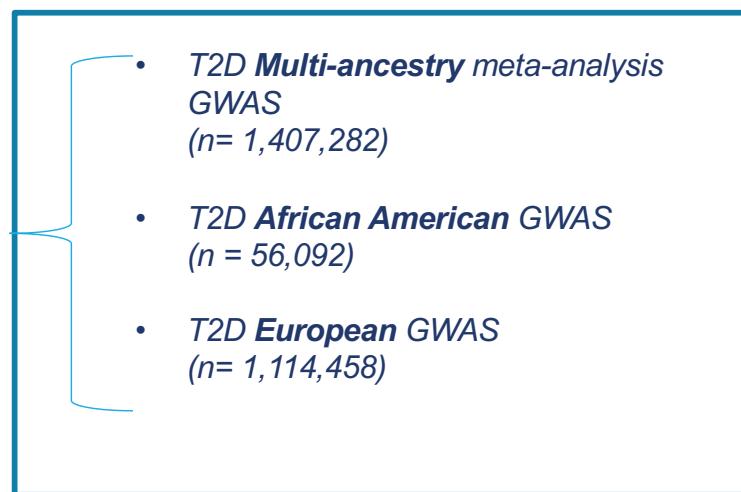


$$PRS_i = \sum_j^M \hat{\beta}_j \times dosage_{ij}$$



# GENETIC PREDICTION OF TYPE 2 DIABETES IN CONTINENTAL AFRICA

- Discovery (Base) Datasets



C+T Approach PRS using PRSice software

We set out to evaluate best PRS to improve polygenic prediction in continental Africans.

## Target Dataset

South Africa – Zulu  
( $n=2,583$ )

## Validation Dataset

AADM  
( $n=4,309$ )  
Kenya – East Africa  
Nigeria – West Africa  
Ghana – West Africa

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

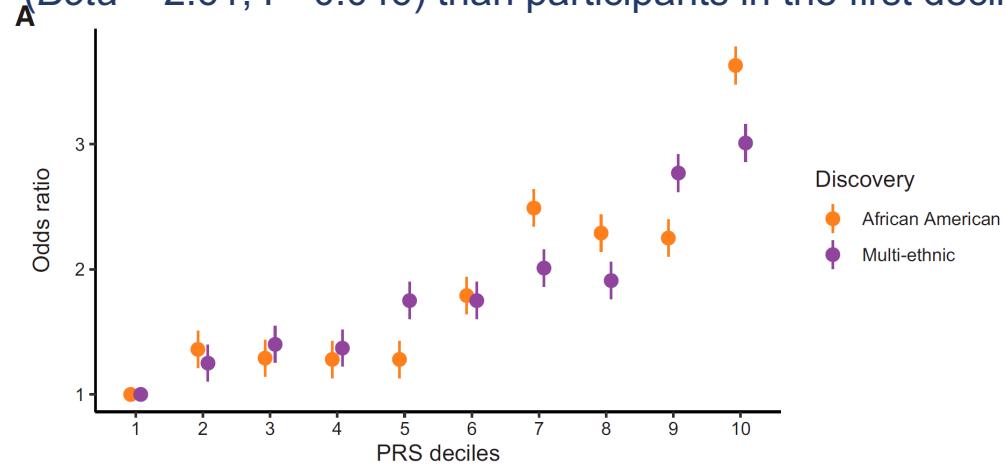
	Multi-ethnic	African American	European
<i>Discovery Dataset (Multi-ancestry meta-analysis)</i>			
Cases	228,499	24,646	148,726
Controls	1,178,783	31,446	965,732
<i>PRS Development</i>			
<i>Target Data Set (SA Zulu)</i>			
Cases	1,602	1,602	1,602
Controls	981	981	981
<i>PRS parameters</i>			
P-value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

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

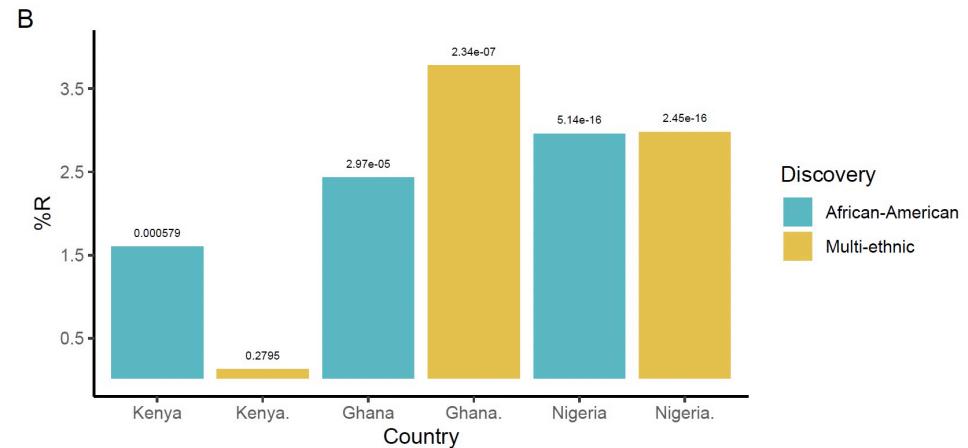
- The best predictive PRS from the data on Europeans, the multi-Ancestry individuals, and African Ancestry was significant
- The highest variance as indicated by Nagelkerke R<sup>2</sup> values of
  - Europeans: 0.69%**
  - multi-Ancestry: 0.69%**
  - African Ancestry: 1.11%**
- The best PRSs were validated in the AADM study and were noted to be all significant in a similar trend
- The African Ancestry PRS had the highest predictability, indicated by a Nagelkerke R<sup>2</sup> of 2.92% ( $9.38 \times 10^{-24}$ ) in the combined analysis of the countries

# PRS STRATIFICATION AND TRANSFERABILITY IN AFRICAN COUNTRIES

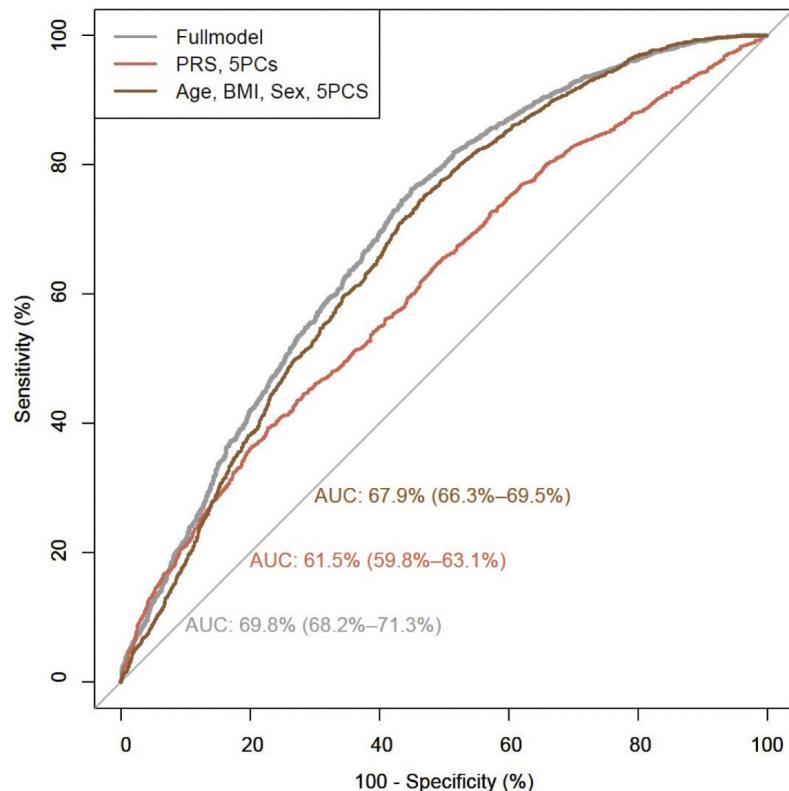
- The participants in the 10th decile of the African Ancestry–derived PRS had **more than threefold higher risk for developing T2D** per risk allele, compared with those in the first decile ( $OR\ 3.63; 95\% CI\ 2.19\text{--}4.03; P = 2.79 \times 10^{-17}$ )
- On average, participants in the 10<sup>th</sup> decile of the African Ancestry PRS were diagnosed with T2D 2.6 years earlier ( $Beta = 2.61; P=0.046$ ) than participants in the first decile
- The African Ancestry PRS was transferable in all countries compared with the multiethnic PRS that was not in Kenya.
- The PRS predictability (indicated by  $R^2$ ) varied greatly between the East Africa and West Africa countries, where predictability was much higher for both the African American and the multi-ancestry PRSs.



Chikowore T et al., ... Fatumo S (2022) Diabetes Care 2022;45(3):717–723



# DISCRIMINATORY ABILITY OF THE POLYGENIC RISK SCORE



- Model with the conventional risk factors of age, BMI, five PCs and sex had an area under the curve (AUC) of **67.9%**
- The African American PRS, five PCs, age, BMI and sex was **69.8%**
- This shows an **improved discriminatory ability by 1.9%**, with the addition of the African American PRS to the conventional risk factors.

Chikowore T et al., ... Fatumo S (2022) Diabetes Care 2022;45(3):717–723



# Diabetes Care

[Diabetes Care](#). 2022 Mar; 45(3): 717–723.

PMCID: PMC8918234

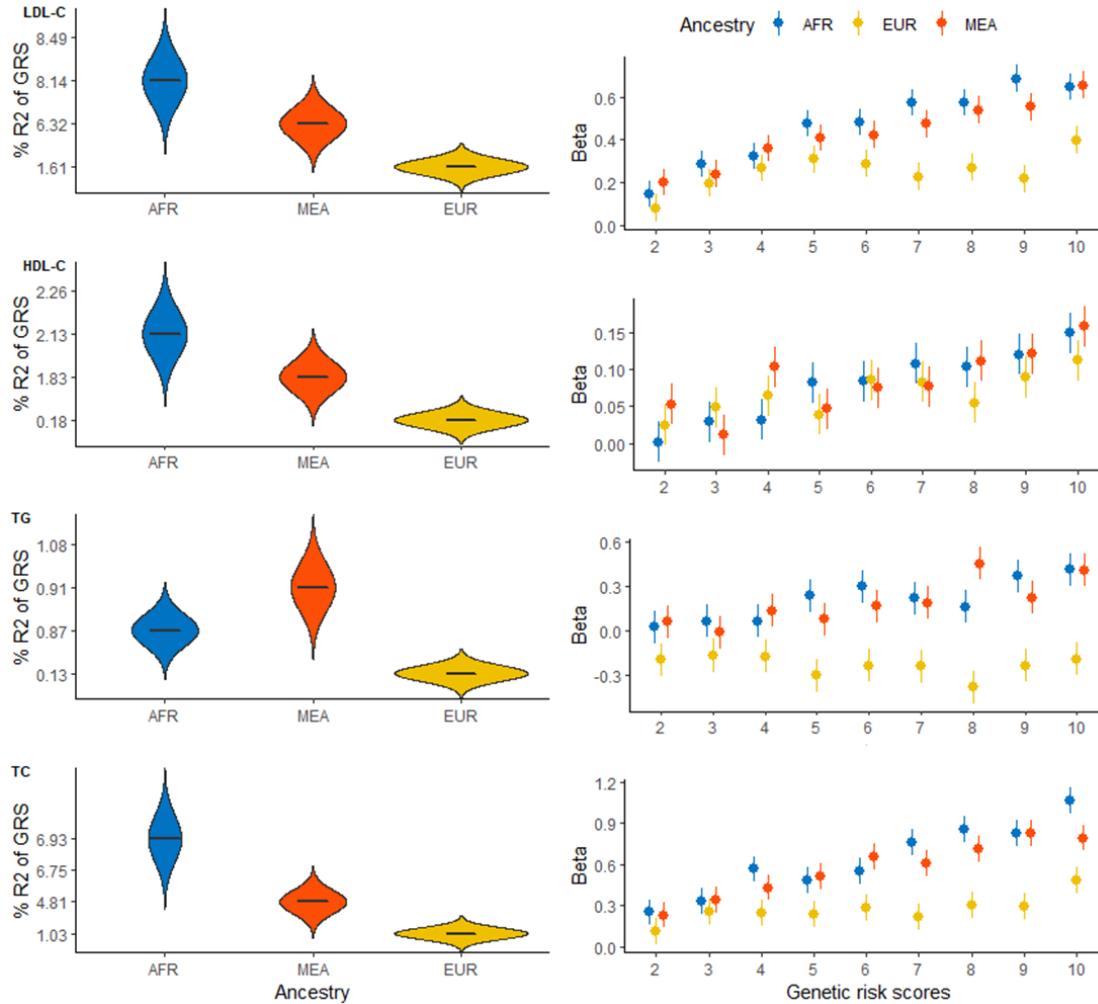
Published online 2022 Jan 11. doi: [10.2337/dc21-0365](https://doi.org/10.2337/dc21-0365)

PMID: [35015074](#)

## Polygenic Prediction of Type 2 Diabetes in Africa

[Tinashe Chikowore](#),<sup>✉</sup> [Kenneth Ekoru](#), [Marijana Vujkovi](#), [Dipender Gill](#), [Fraser Pirie](#), [Elizabeth Young](#),  
[Manjinder S. Sandhu](#), [Mark McCarthy](#), [Charles Rotimi](#), [Adebawale Adeyemo](#), [Ayesha Motala](#), and [Segun Fatumo](#)<sup>✉</sup>

► Author information ► Article notes ► Copyright and License information    [Disclaimer](#)

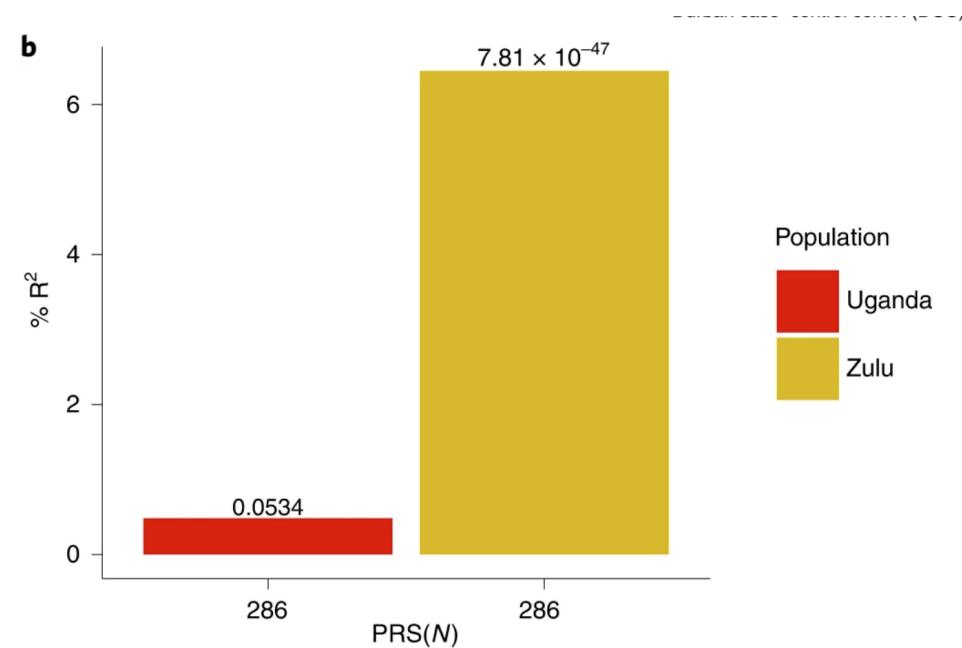
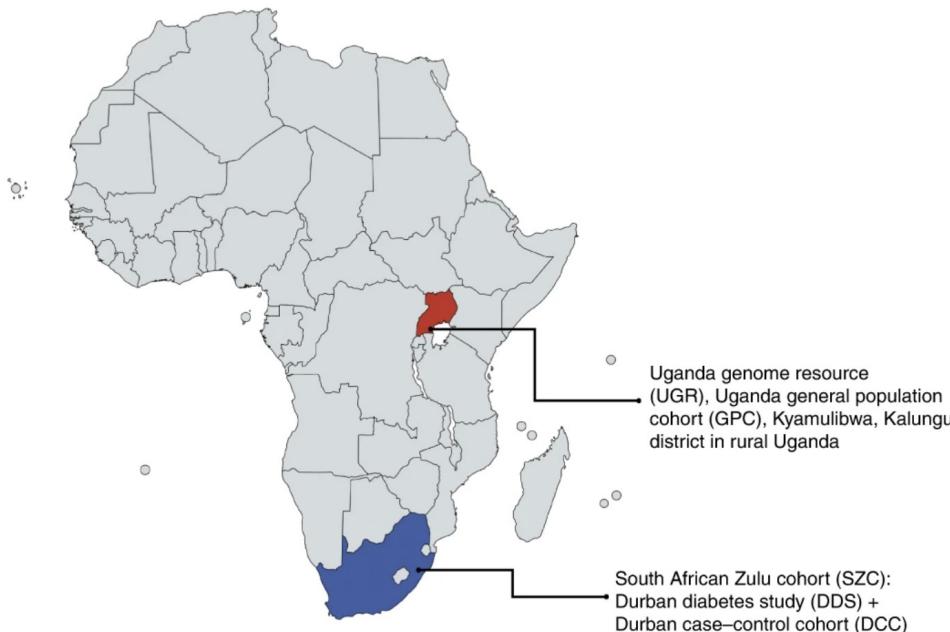


- SIMILARLY, PRSS DERIVED FROM AFRICAN ANCESTRY INDIVIDUALS ENHANCE POLYGENIC PREDICTION OF LIPID TRAITS IN SSA COMPARED TO EUROPEAN AND MULTI-ANCESTRY SCORES.

HOWEVER, OUR PRS PREDICTION VARIED GREATLY WITHIN SSA BETWEEN THE SOUTH AFRICAN ZULU (EG LDL-C,  $R^2 = 8.14\%$  AND UGANDAN COHORTS LDL-C,  $R^2 = 0.026\%$ )

Kamiza A et al., ... Fatumo S (2022) **Nature Medicine** vol 28, pg 1163-1166

Comparative performance of polygenic prediction of Total Cholesterol using the same GRS comprising 286 SNPs, which was developed in Ugandan cohort ( $n = 6,407$ ) and then replicated in the South African Zulu cohort ( $n = 2,598$ ).





OPEN

# Transferability of genetic risk scores in African populations

Abram B. Kamiza<sup>1,2,3</sup>, Sounkou M. Toure<sup>1,4</sup>, Marijana Vujkovic<sup>ID 5</sup>, Tafadzwa Machipisa<sup>ID 6,7,8</sup>, Opeyemi S. Soremekun<sup>1</sup>, Christopher Kintu<sup>1</sup>, Manuel Corpas<sup>ID 9,10,11</sup>, Fraser Pirie<sup>12</sup>, Elizabeth Young<sup>13</sup>, Dipender Gill<sup>ID 14,15</sup>, Manjinder S. Sandhu<sup>14</sup>, Pontiano Kaleebu<sup>16</sup>, Moffat Nyirenda<sup>ID 16</sup>, Ayesha A. Motala<sup>12</sup>, Tinashe Chikowore<sup>ID 3,17,20 ✉</sup> and Segun Fatumo<sup>ID 1,16,18,19,20 ✉</sup>

# Recommended Reading

Review | [Open access](#) | Published: 30 October 2023

## 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](#)

[Genome Medicine](#) **15**, Article number: 87 (2023) | [Cite this article](#)

**2208** Accesses | **3** Citations | **6** Altmetric | [Metrics](#)

<https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-023-01245-9>

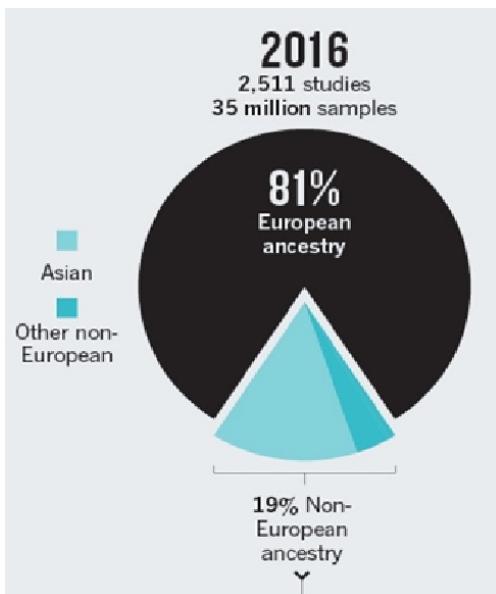
# Factors contributing to poor transferability of PRS in African populations

- **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.**
- Given that Africa has the highest genetic diversity in the whole world. However, 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 (eg different weight for effect size).
- Therefore, the transferability of risk score across populations is a major challenge when they do not share the **same genetic architecture for each disease**

# THERE IS A DRASTIC LACK OF DIVERSITY IN GENOMICS DESPITE REPEATED CALLS AND WARNING

2016

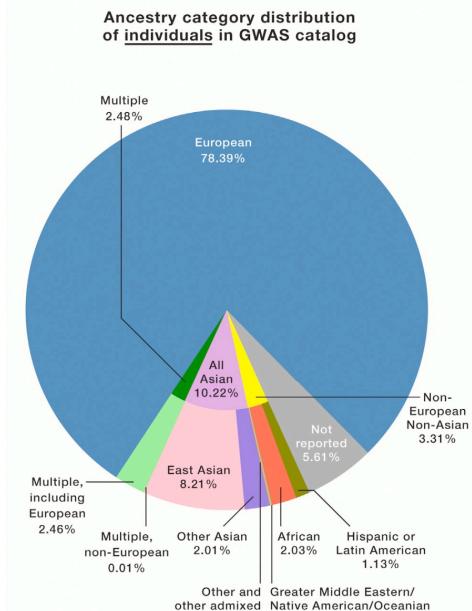
- Europeans: 81%
- Africans: 3%



Popejoy, A., Fullerton, S. 2016 *Nature* 538, 161–164

2019

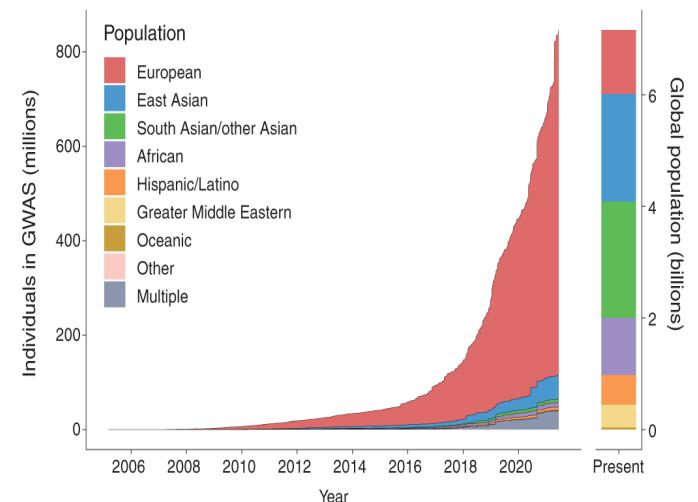
- Europeans: 78.39%
- Africans: 2.03%



Sirugo et al., 2019. *Cell*, 177(1), 26-31.

2022

- Europeans: 86%
- Africans: 1.1%



Fatumo et al., 2022 *Nature Medicine*, 28(2), 243-250.

TO IMPROVE REPRESENTATION OF AFRICAN GENOMIC DATA IN THE GLOBAL CONTEXT FOR DISCOVERY AND GENETIC RISK PREDICTION, MULTIPLE KEY INITIATIVES IN AFRICA HAVE BEEN ESTABLISHED, INCLUDING THE **H3AFRICA, NIGERIAN 100 GENOMES PROJECT,, UGANDA GENOME RESOURCE** AND MANY OTHERS

GREAT VALUE FOR DISCOVER AND GENETIC PREDICTION THAT CAN GENERALIZE MORE BROADLY ACROSS GLOBAL POPULATIONS.

THESE AFRICAN GENOMIC RESOURCES WILL IMPROVE THE POWER AND EQUITABILITY OF PRS AND THEIR UNDERPINNING GWAS.

IT IS VITAL TO DEVELOP ACCURATE PRS ACROSS DIFFERENT POPULATIONS



Fatumo, S., & Inouye, M. (2023). African genomes hold the key to accurate genetic risk prediction. *Nature Human Behaviour*, 7(3), 295-296.

[nature](#) > [nature human behaviour](#) > [correspondence](#) > [article](#)

Correspondence | [Published: 23 February 2023](#)

## African genomes hold the key to accurate genetic risk prediction

[Segun Fatumo](#)  & [Michael Inouye](#)

[Nature Human Behaviour](#) 7, 295–296 (2023) | [Cite this article](#)

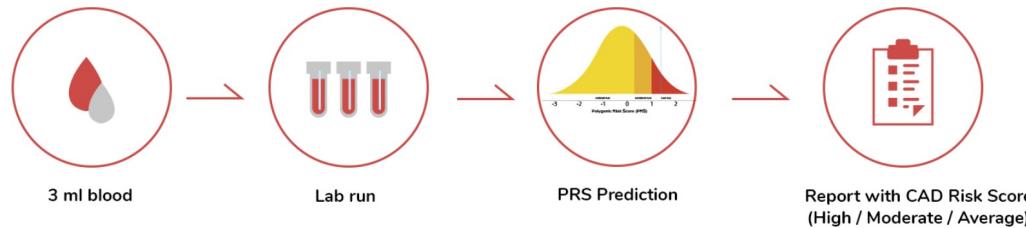
876 Accesses | 1 Citations | 34 Altmetric | [Metrics](#)

To achieve accurate genetic risk prediction and transferability across all populations, large-scale African data are necessary for PRS to achieve their potential.

# KARDIOGEN: PREVENTIVE HEALTHCARE USING PRS TEST FOR CORONARY ARTERY DISEASE

## What is the test process?

Just 3 ml of your blood can give your CAD PRS score



*Results within 12 working days*

## What does one do after receiving your CAD score

CAD Risk Profile	What it means?	Action
High	You are at a higher genetic risk (3 times) of getting CAD than a average risk counterpart.	Take a comprehensive heart check up and consult your doctor for an invasive procedure
Moderate	You are at a higher genetic risk (~1.5 times) towards CAD than a average risk counterpart	Avoid all lifestyle risks such as smoking, stress, etc. Maintain normal blood pressure and blood sugar levels with regular monitoring. Medication, such as blood thinners can be considered.
Average	You are at a average genetic risk towards coronary events	Maintain a healthy lifestyle and get routine health checkups.

# Example: Polygenic Risk Score for Coronary Artery Disease

## Genetic Markers

Hundreds of genetic variants have been associated with an increased risk of coronary artery disease.

## Risk Stratification

PRS can stratify individuals into high, intermediate, and low-risk categories for coronary artery disease.

## Preventive Measures

Individuals with high PRS for coronary artery disease can be targeted for more intensive screening and lifestyle interventions.

## Clinical Utility

PRS for coronary artery disease has shown promise in improving risk prediction and guiding preventive strategies.

# Applications of Polygenic Risk Scores

## Disease Prediction

PRS can be used to identify individuals at high risk for certain diseases, enabling early intervention and preventive strategies.

## Personalized Screening

PRS can guide personalized screening programs, such as targeted cancer screenings, based on an individual's genetic risk profile.

## Drug Response

PRS can help predict an individual's response to certain medications, allowing for more targeted and effective treatments.



# Limitations and Considerations of Polygenic Risk Scores

## 1 Ancestry Bias

PRS models may be biased towards populations with more genetic data available, limiting their applicability across diverse populations.

## 2 Environmental Factors

PRS do not account for the full complexity of disease etiology, as they do not capture the influence of environmental and lifestyle factors.

## 3 Ethical Considerations

The use of PRS raises ethical concerns, such as privacy, discrimination, and the potential for misinterpretation of results.



# Conclusion and Key Takeaways

- 1 Powerful Tool**

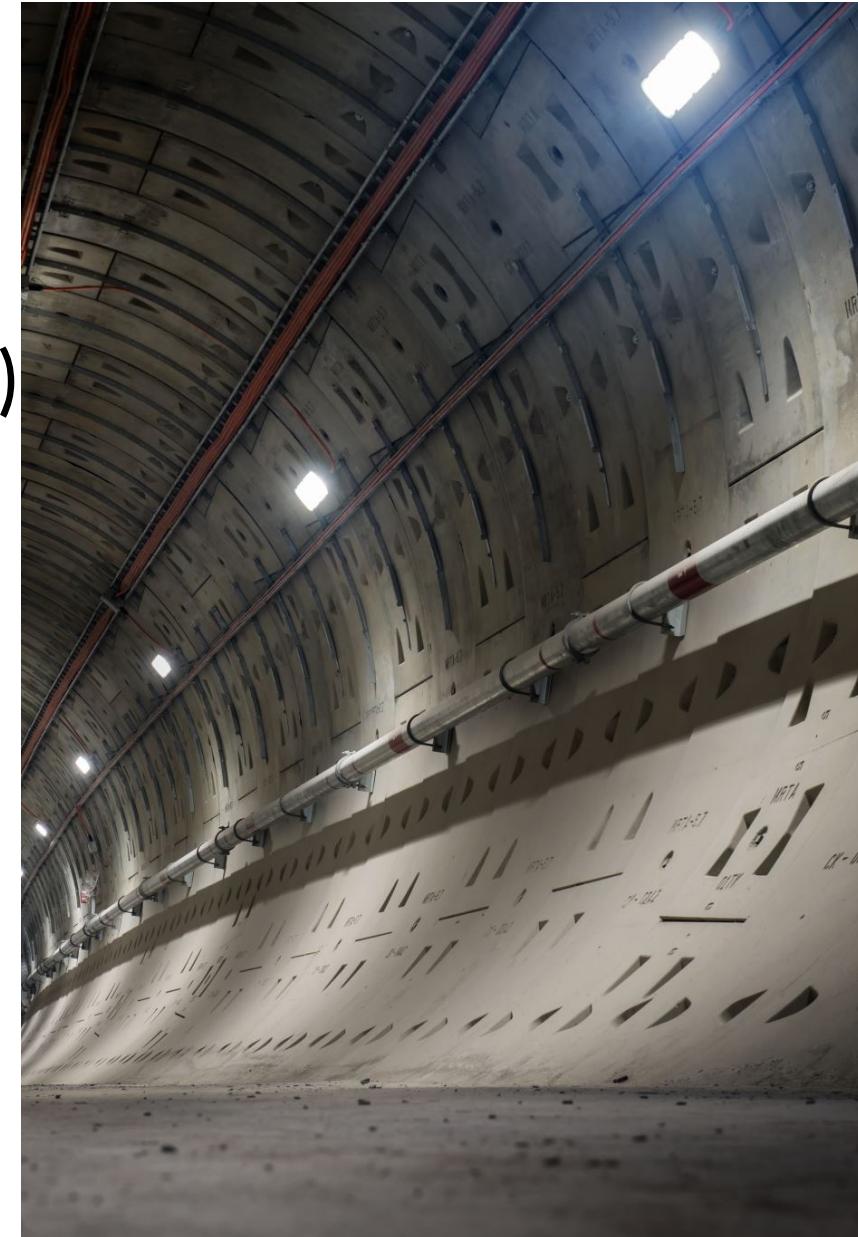
Polygenic risk scores are a powerful tool in personalized medicine, providing insights into an individual's genetic predisposition to complex diseases.
- 2 Limitations and Considerations**

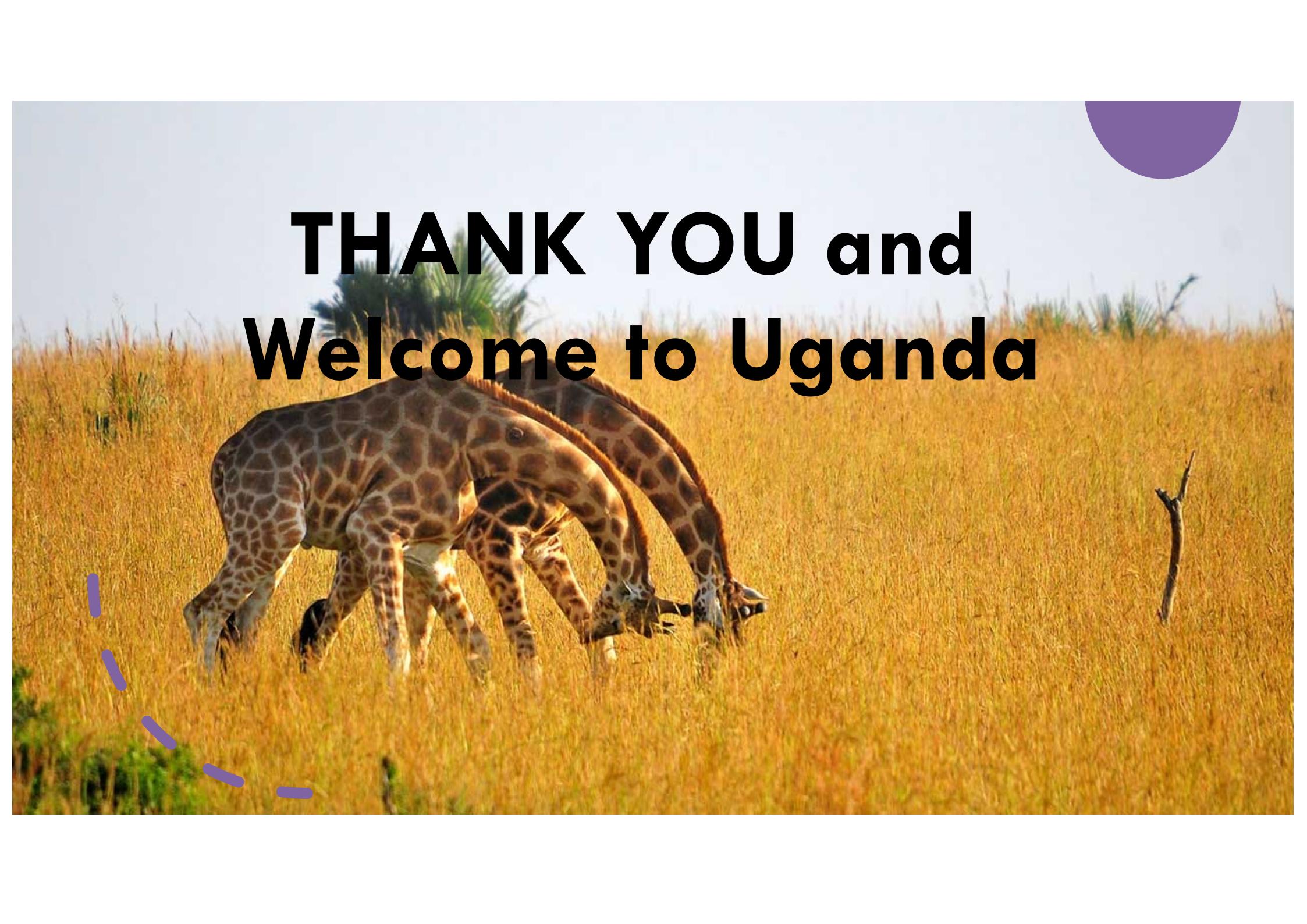
While PRS have great potential, it is essential to consider their limitations, such as ancestry bias and the influence of environmental factors.
- 3 Ongoing Research**

Continued research and development in this field will further enhance the utility of PRS in clinical practice and personalized healthcare.

## EXPECTATIONS AFTER THE COURSE

1. Alumni community (Stay in touch)
2. PRS pipeline Algorithm
3. Biotech Companies
4. PhD/Postdoc using the skills
5. Analyse African Datasets
6. Scientific papers as a group
7. Train others





**THANK YOU and  
Welcome to Uganda**

