Polygenic Risk Scores - Africa

Day 3 – The portability problem

Instructors: Palwende Romuald Boua & Stephie Raveloson

MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES
wellcome connecting science

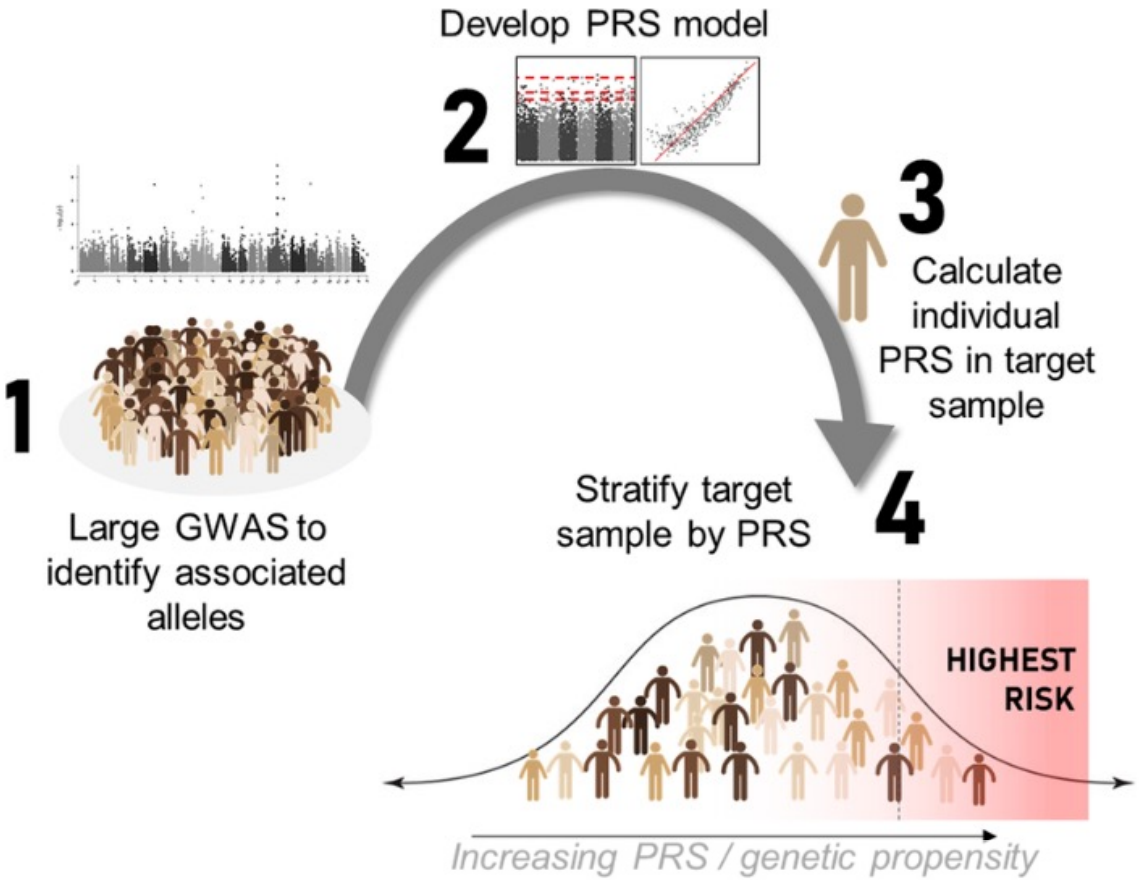
By the end of this lecture, you should be able to…

1. Define what the PRS portability problem is.
2. Describe the likely genetic and non-genetic reasons for poor PRS transferability.
3. Understand these in the context of African populations.

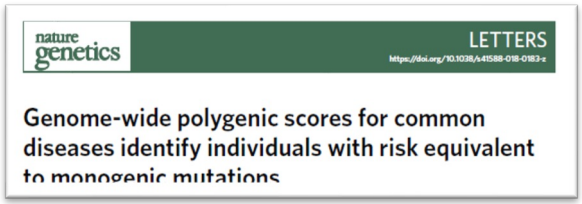
Outline:

- Recap of GWAS and PGS
- Introduction to the Portability Problem
- Theoretical basis for portability issues
  - Genetic factors
  - Non-Genetic factors
- Summary
- Questions and Discussion

Recap: GWAS to PRS



Utility of PRS growing and of increasing public interest


nature genetics

LETTERS

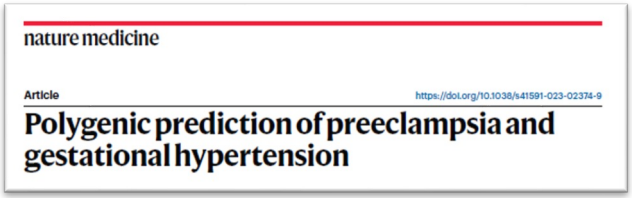
Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations


Comment | Open access | Published: 25 May 2022

The potential of polygenic scores to improve cost and efficiency of clinical trials

Akl C. Fahed, Anthony A. Philippakis & Amit V. Khera

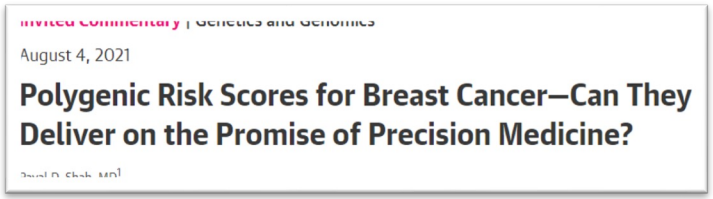
Nature Communications 13, Article number: 2922 (2022) | Cite this article


nature medicine

Article

https://doi.org/10.1038/s41591-023-02374-9

Polygenic prediction of preeclampsia and gestational hypertension


August 4, 2021

Polygenic Risk Scores for Breast Cancer—Can They Deliver on the Promise of Precision Medicine?


Genomics plc and GSK establish precision medicine collaboration to assess polygenic risk scores in clinical trial design

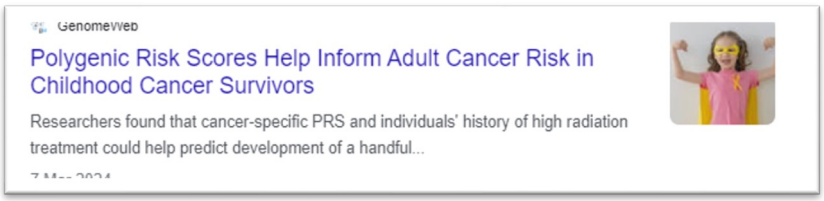
GENOMICS


The New York Times

https://www.nytimes.com/2018/08/13/health/genetic-test-heart-disease.html

Clues to Your Health Are Hidden at 6.6 Million Spots in Your DNA

With a sophisticated new algorithm, scientists have found a way to forecast an individual's risks for five deadly diseases.


GenomeWeb

Polygenic Risk Scores Help Inform Adult Cancer Risk in Childhood Cancer Survivors

Researchers found that cancer-specific PRS and individuals' history of high radiation treatment could help predict development of a handful...

7 May 2024

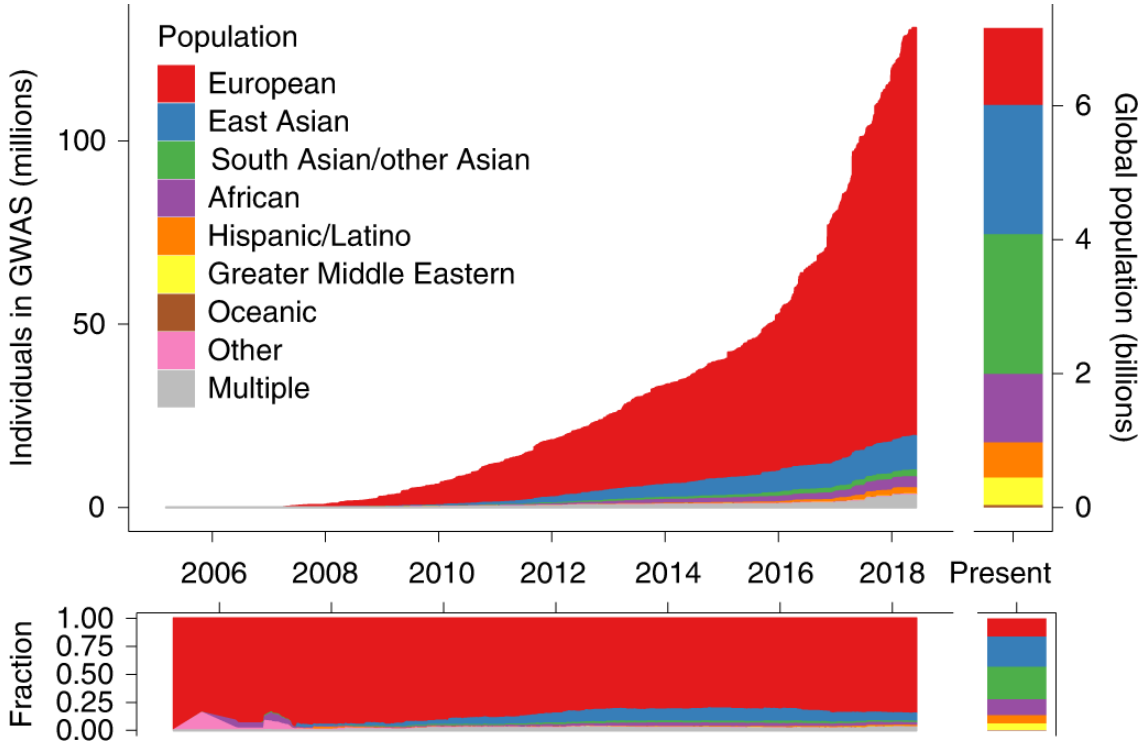
MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

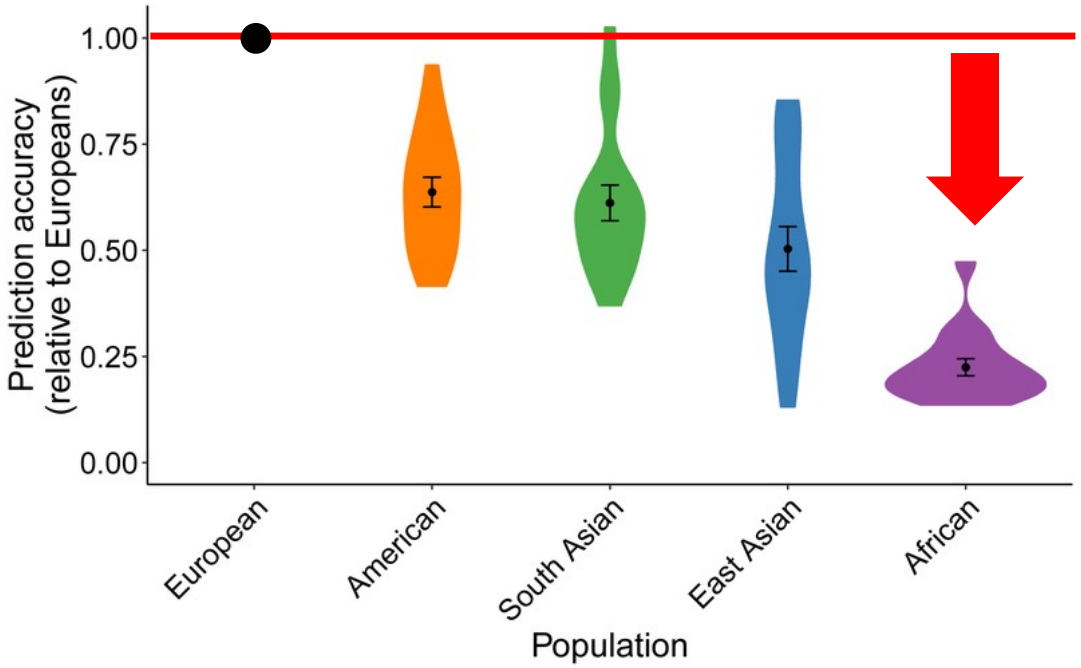
wellcome connecting science

But there is a big DIVERSITY problem…



Martin et al., (2019). Nat. Gen, 51. 584-591

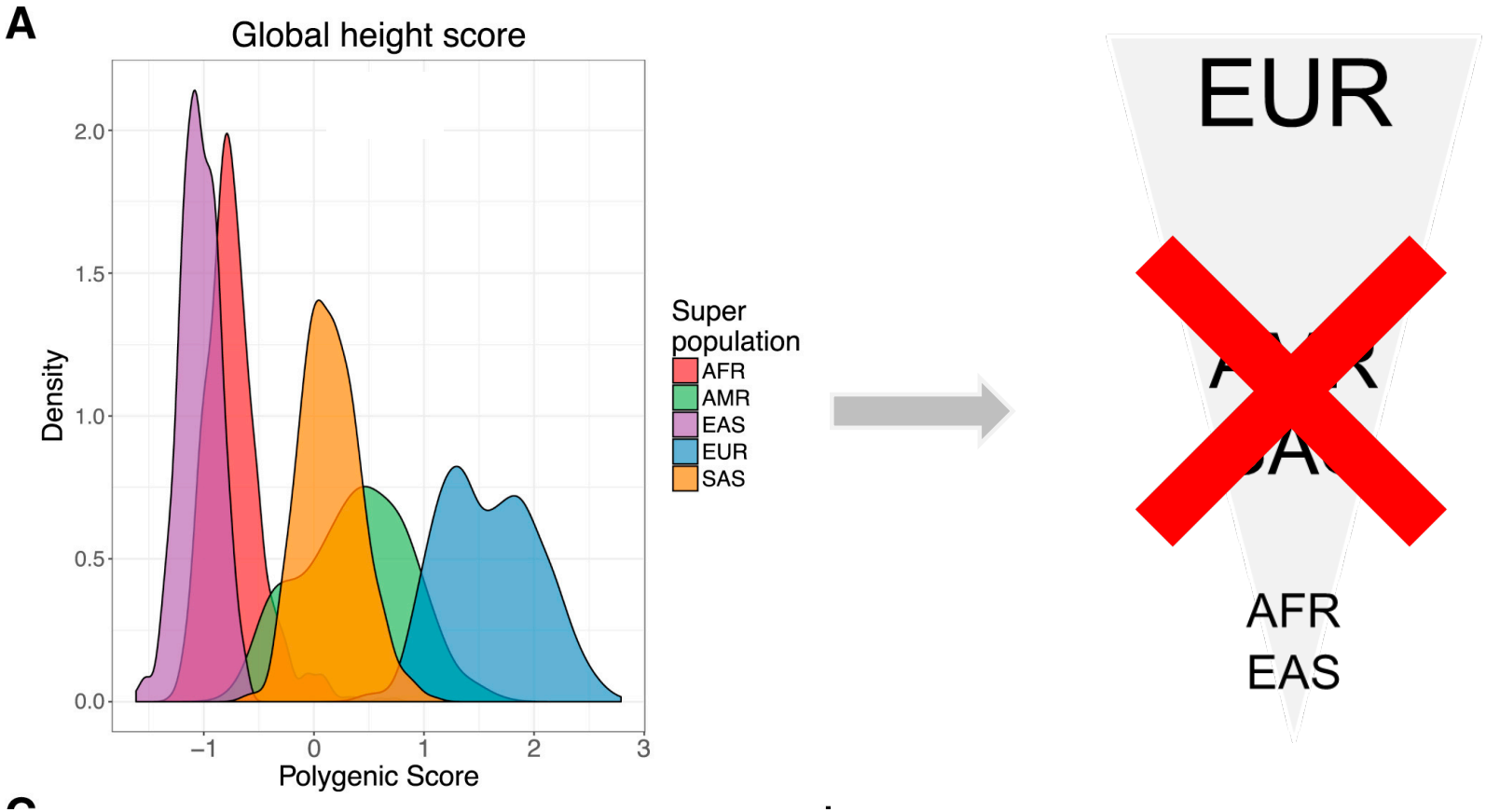
Resulting in PRS that are NOT TRANSFERABLE…



Genetic prediction accuracy decays with increasing genetic distance between the discovery and target populations.

Martin et al., (2019). Nat. Gen, 51. 584-591

Cross-ancestry polygenic risk scores

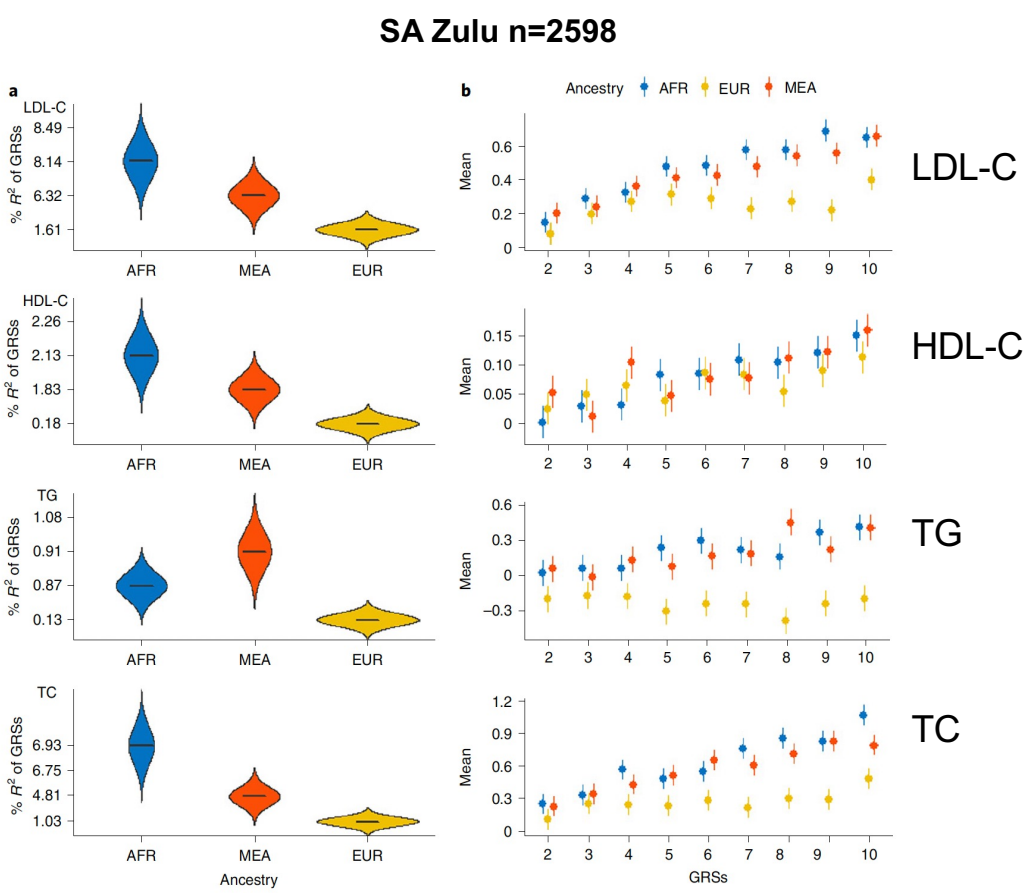


MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

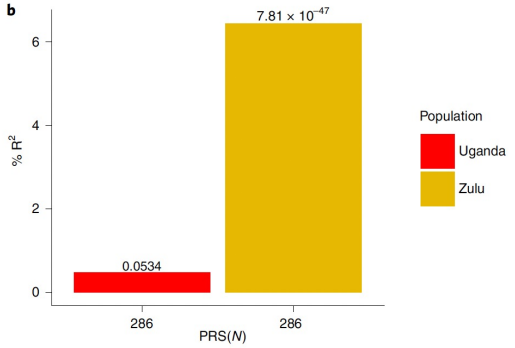
Martin et al., (2017) Am J Hum Genet.100(4): 635–649.

wellcome connecting science

NOT only at superpopulation level but also regionally



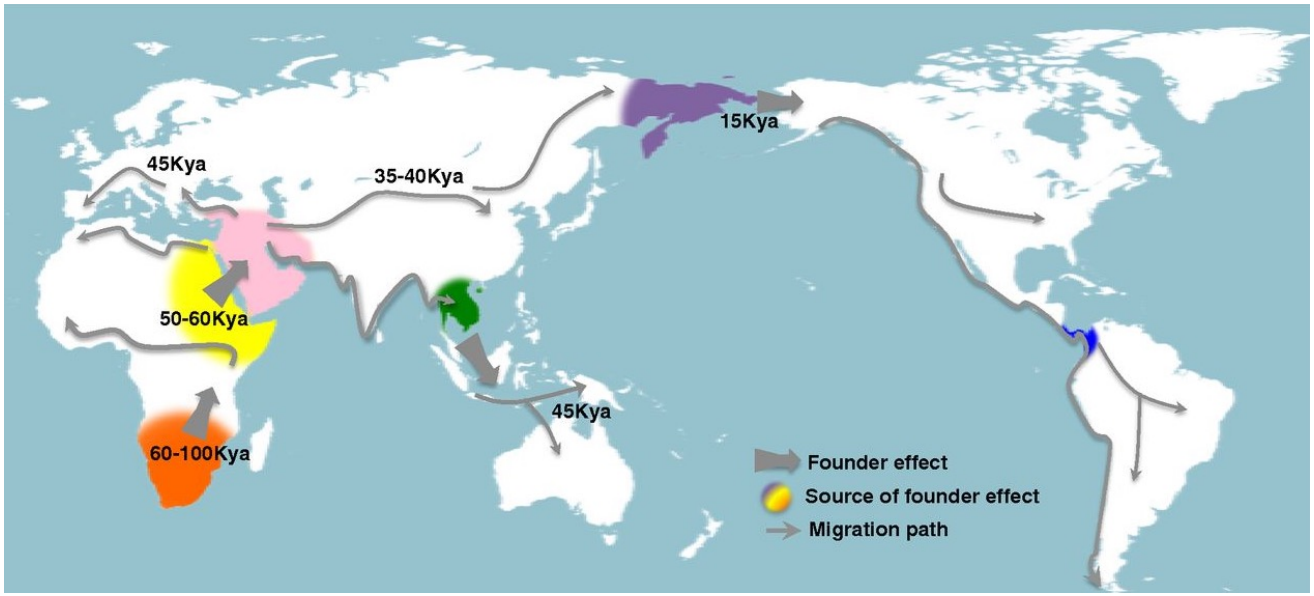
% Variance explained by African American GRS for Total cholesterol

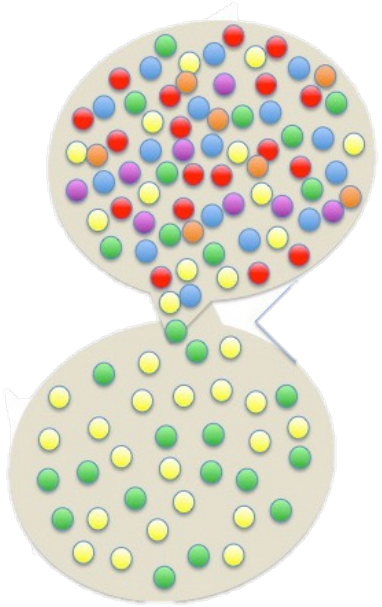


Kamiza et al., (2023). Nat. Med

WHY

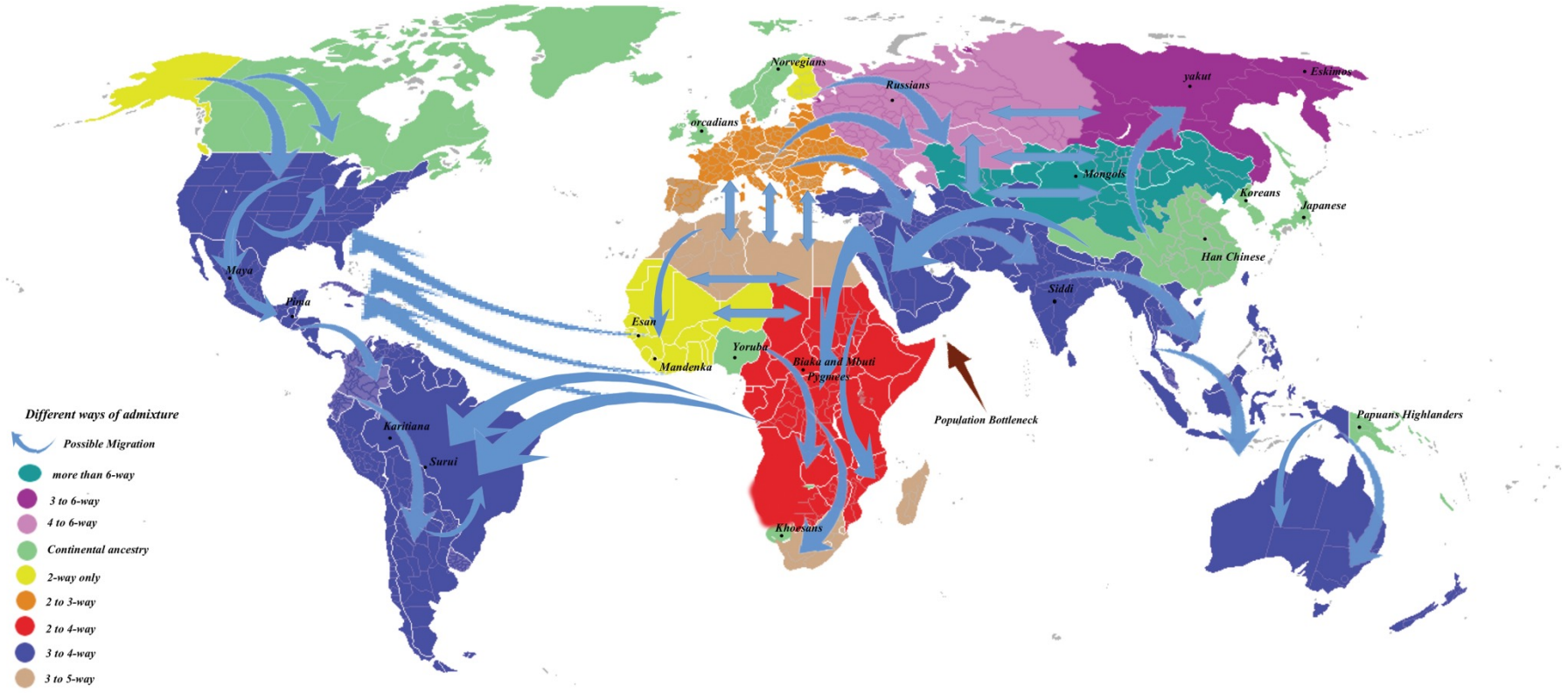
Start at the beginning….





Henn, Cavalli-Sforza, and Feldman (2012). PNAS

RECENT ADMIXTURE



MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

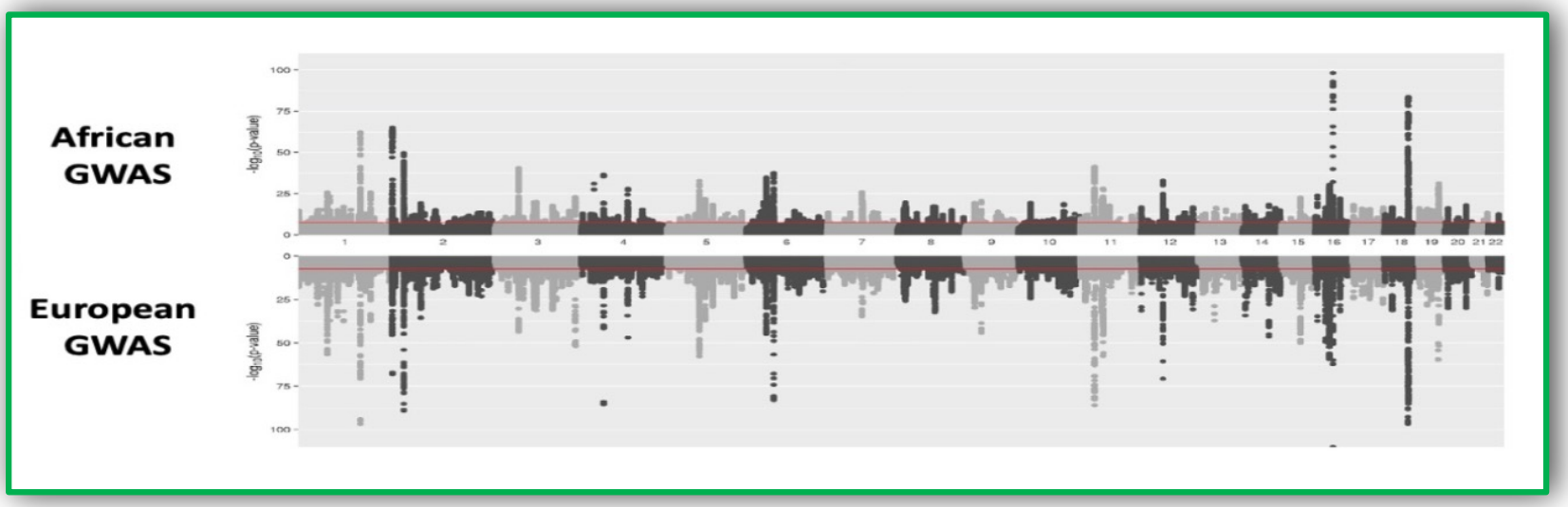
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

Chimusa *et al.*, Briefings in Bioinformatics, 00(00), 2018, 1–12

wellcome connecting science

Genetic similarity >> Genetic variation

Fundamental biology is shared across populations.



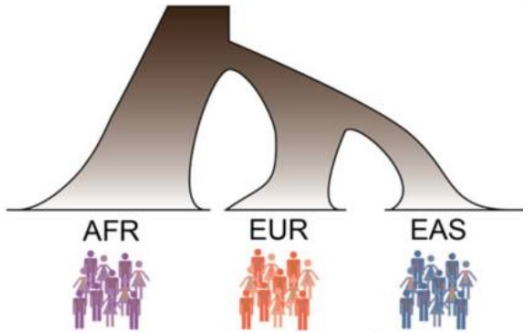
Sharing of genetic architecture provides scope for harnessing GWAS information indiscriminately from all ancestries

MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES
wellcome connecting science

Predictable basis of PRS disparities

Prediction accuracy decays with increasing genetic distance

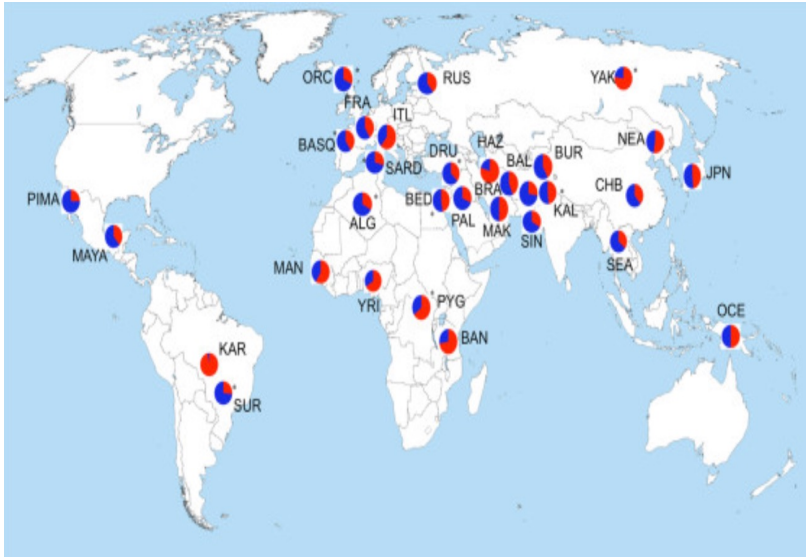
$$PRS_i = \sum_j^M \widehat{B}_j \times dosage_{ij}$$



- GWAS best-powered to discover common variants in the discovery population
- LD difference across populations
- Environment, selection, and other differences

Minor allele frequency differences

- Frequency of the **second most common** allele occurs.
- Variability of MAF across populations leads to different risk allele pools.
- Overestimate or underestimate risk in populations with different allele frequencies.



Discovery

Target

DIFFERENT MAF and EFFECT SIZES

H3Africa
Human Heredity & Health in Africa

Article

High-depth African genomes inform human migration and health

https://doi.org/10.1038/s41586-020-2859-7
Received: 10 May 2019
Accepted: 7 August 2020
Published online: 28 October 2020
Open access
Check for updates

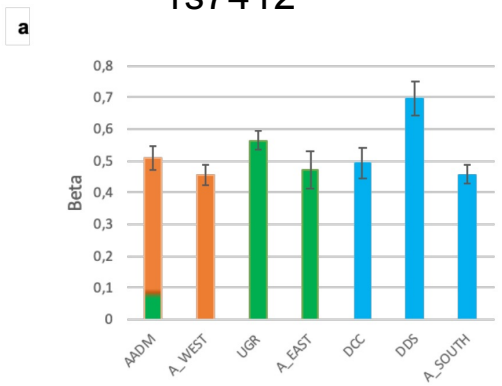
Ananyo Choudhury[1], Shaun Aron[1], Laura R. Botigué[2], Dhriti Sengupta[1], Gerrit Botha[3], Taoufik Bensellak[4], Gordon Wells[5,6], Judit Kumuthini[5,6], Daniel Shriner[7], Yasmina J. Fakim[8,9], Anisah W. Ghoorah[9], Eileen Dareng[10,11], Trust Odia[12], Oluwadamilare Falola[12], Ezekiel Adebiyi[12,13], Scott Hazelhurst[1,14], Gaston Mazandu[3], Oscar A. Nyangiri[15], Mamana Mbiyavanga[3], Alia Benkahla[16], Samar K. Kassim[17], Nicola Mulder[3], Sally N. Adebamowo[18,19], Emile R. Chimusa[20], Donna Muzny[21], Ginger Metcalf[21], Richard A. Gibbs[21,22], TrypanoGEN Research Group*, Charles Rotimi[2], Michèle Ramsay[1,23], H3Africa Consortium*, Adebowale A. Adeyemo[12] ✉, Zané Lombard[23] ✉ & Neil A. Hanchard[22] ✉

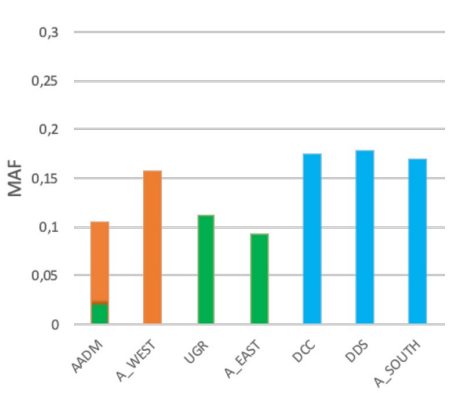
426 whole genomes from 50 ethnolinguistic groups

Effect size

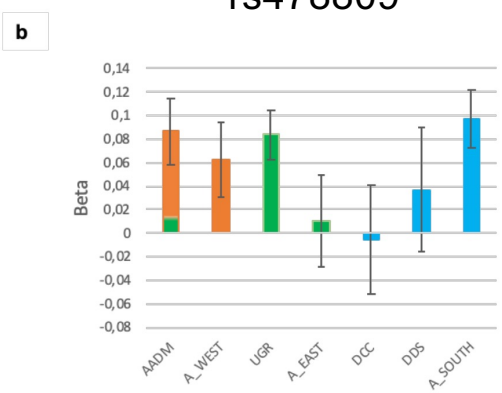
Minor allele frequency

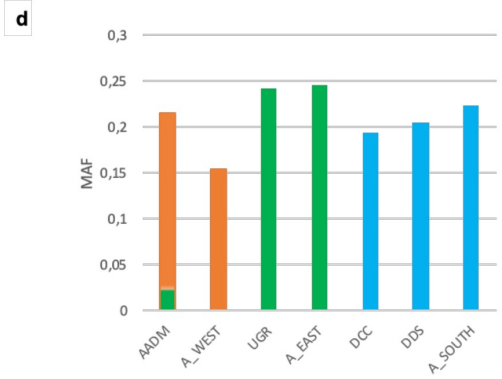
rs7412





rs478809





populations

Choudhury et al. H3Africa Consortium. (2020) Nature

MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES
wellcome connecting science

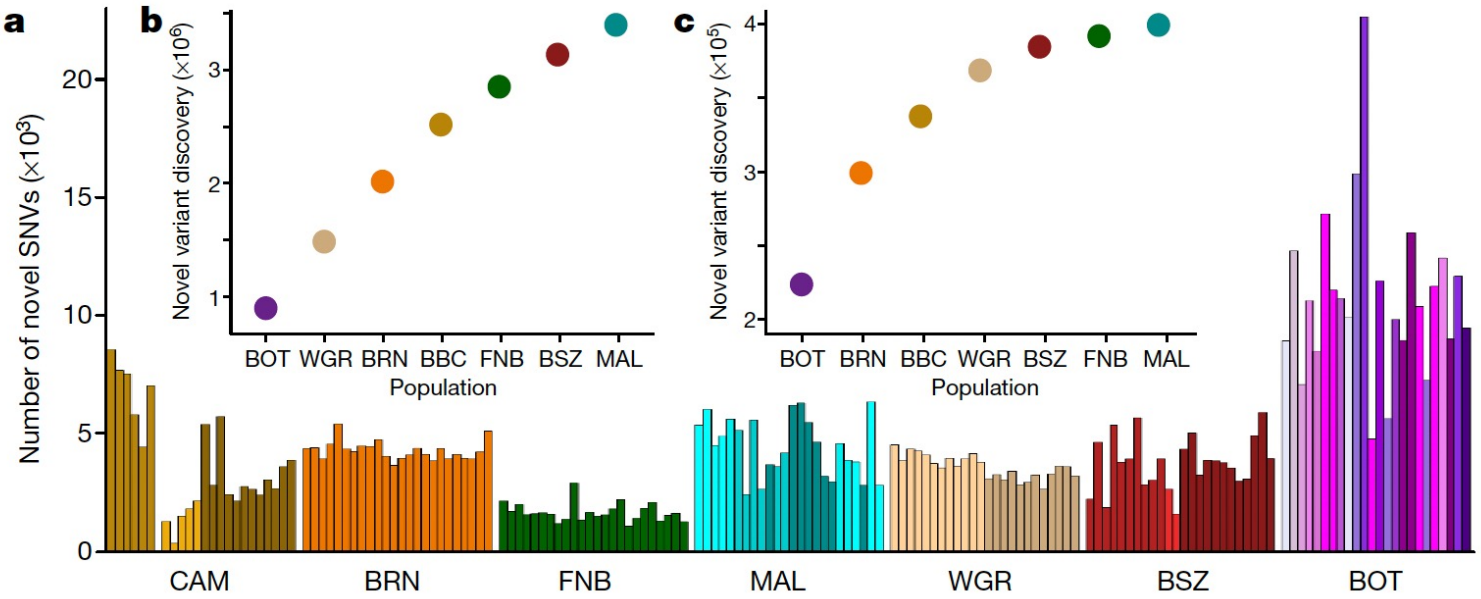
AFRICA: >> Population-specific variants

H3Africa
Human Heredity & Health in Africa

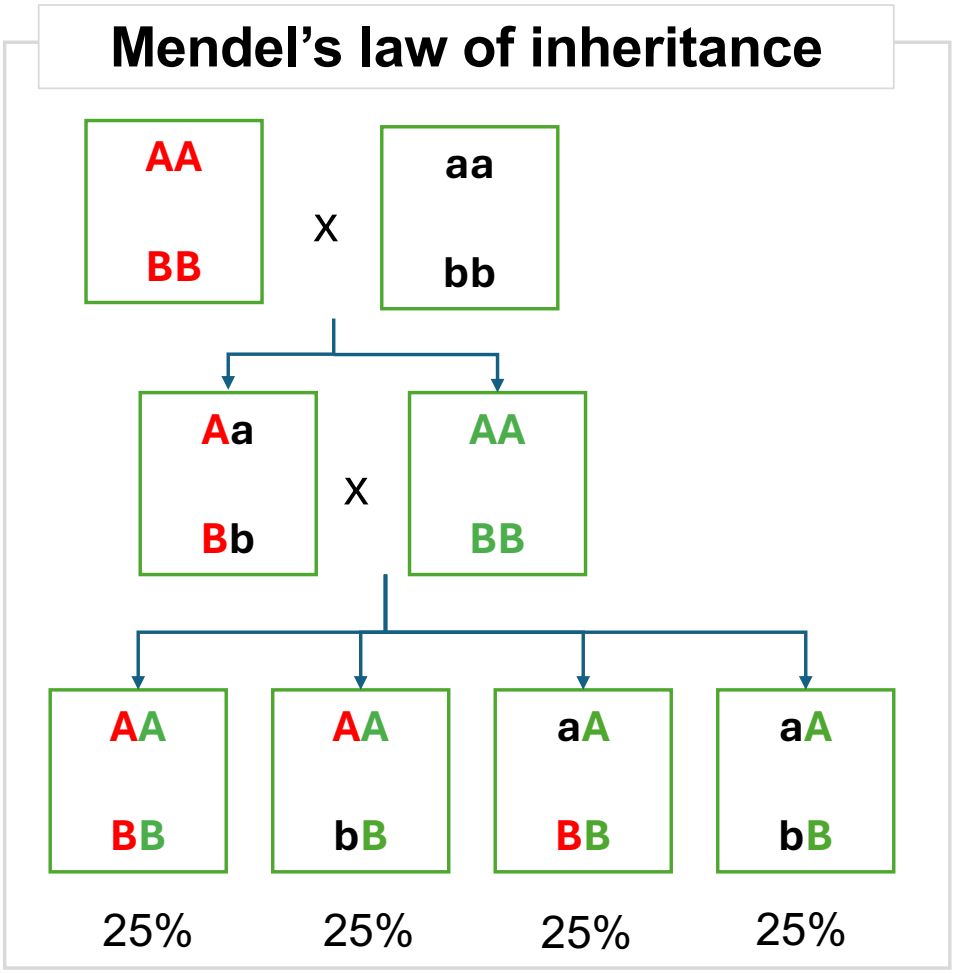
41.6 million variants
~12 to 20 million per population

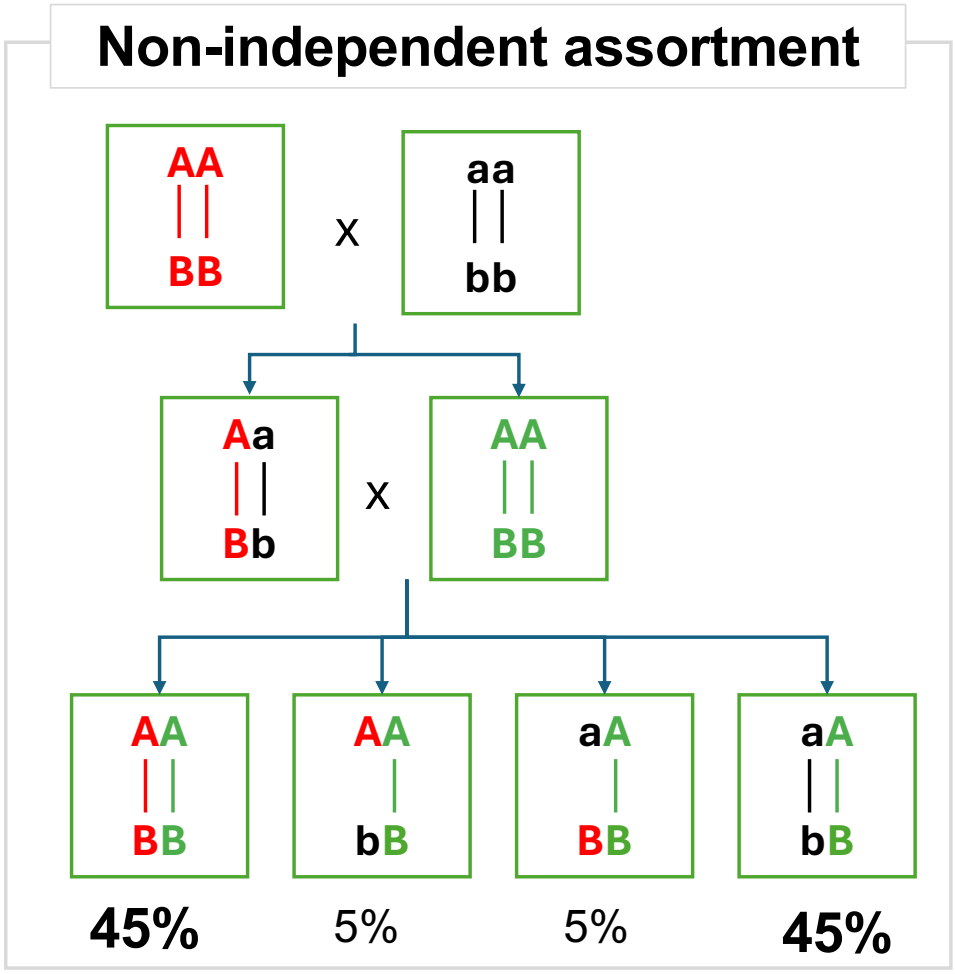
4.3 million novel SNPs

(~2-5% novel SNP in each population)



What is linkage disequilibrium? [1]



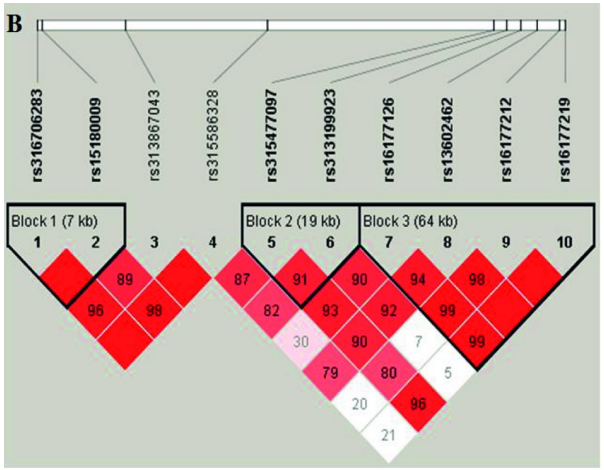


What is linkage disequilibrium? [2]

- Certain alleles on a chromosome **associated more often than expected by chance**.

- The **closer two genetic markers** are, the **less likely recombination** will occur between them.

More likely to be inherited together

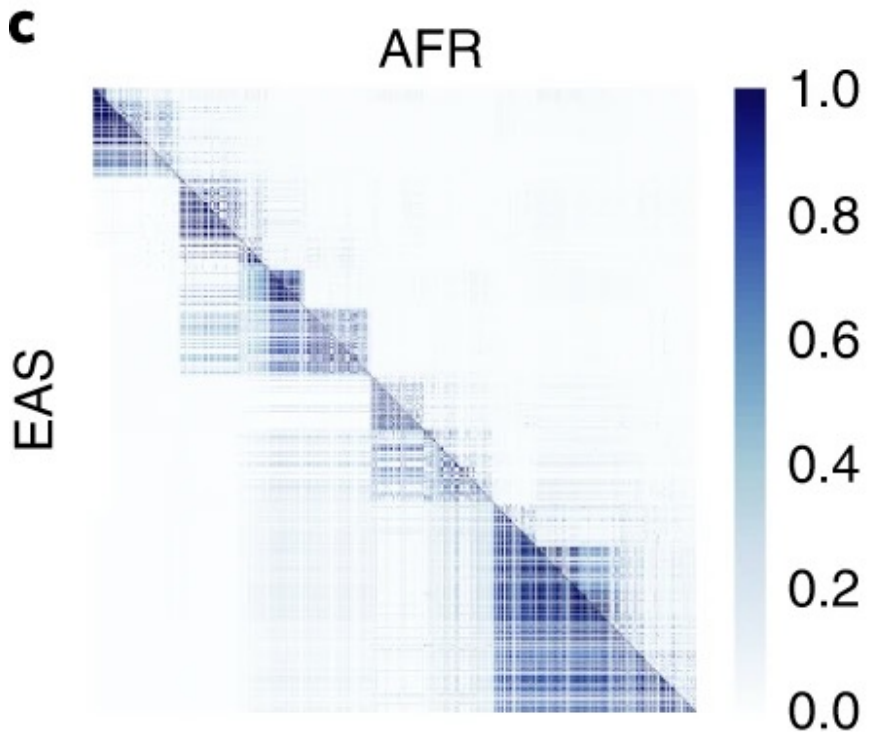


What is linkage disequilibrium? [2]

- Certain alleles on a chromosome **associated more often than expected by chance**.

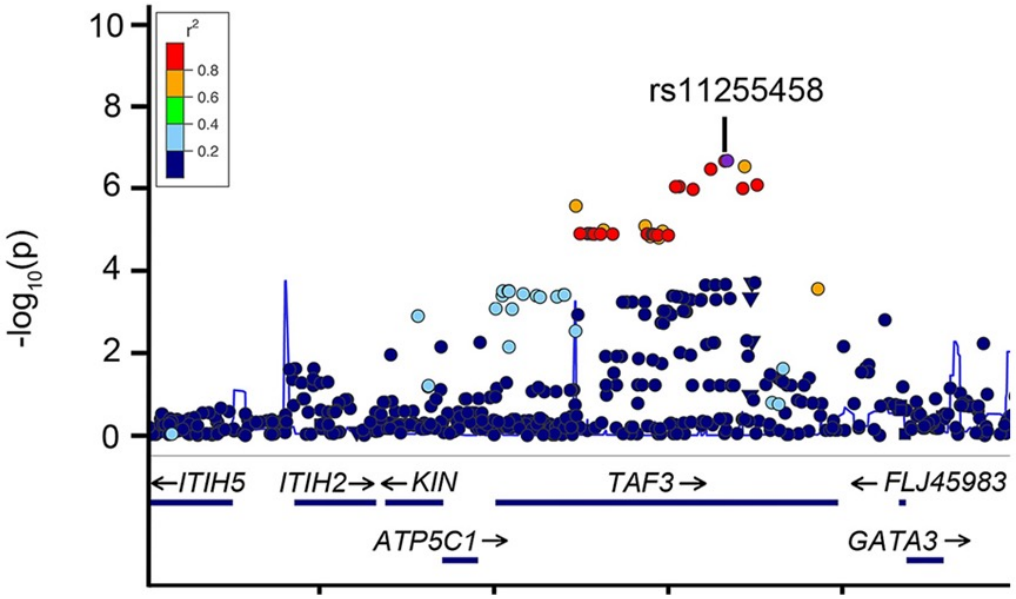
- The **closer two genetic markers** are, the **less likely recombination** will occur between them.

More likely to be inherited together



Why does LD matter?

- GWAS tests **each variant separately**.
- Must account for LD in PGS to **avoid double counting** highly correlated variants.
- LD-based **clumping removes highly correlated variants** within each locus.



MAKERERE UNIVERSITY
IDI INFECTIOUS DISEASES INSTITUTE
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES
wellcome connecting science

Addressing LD: Clumping

1. SNPs sorted by their GWAS *p*-values.

2. Starting from the most significant SNP (the index/lead SNP), any SNPs in high LD (e.g. r^2 > 0.1) with the index SNP are removed.

3. To reduce computational burden, only SNPs that are close to the index SNP are $clumped$.

4. This process is continued until no index SNPs remain.

LD clumping example [1]

P-Value	SNP	A	B	C	D	E	F	G	H	I	J	K	L
0.418	A	■						■					
0.7	B		■					■					
0.0459	C			■			+						
0.0358	D				■		+						
6.35E-63	E					■	+						
2.29E-63	F			+	+	+	+		+				
0.00485	G	■	■					■					
0.00287	H						+		■	■			
0.63	I								■	■		■	■
0.00177	J										■	■	■
0.0026	K									■	■	■	
0.00263	L									■	■		■

Color	Meaning
■	LD Above Threshold
	Index SNP

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

LD Clumping example [2]

P-Value	SNP	A	B	C	D	E	F	G	H	I	J	K	L
0.418	A	■						■					
0.7	B		■					■					
~~0.0459~~	~~C~~			■			■						
~~0.0358~~	~~D~~				■		■						
~~6.35E-63~~	~~E~~					■	■						
2.29E-63	F			■	■	■	■		■				
0.00485	G	■	■					■					
~~0.00287~~	~~H~~						■		■	■			
0.63	I								■	■		■	■
0.00177	J										+	+	+
0.0026	K									■	+	■	
0.00263	L									■	+		■

Color	Meaning
■	LD Above Threshold
■	Index SNP

LD clumping example [3]

P-Value	SNP	A	B	C	D	E	F	G	H	I	J	K	L
0.418	A							+					
0.7	B							+					
~~0.0459~~	~~C~~												
~~0.0358~~	~~D~~												
~~6.35E-63~~	~~E~~												
2.29E-63	F												
0.00485	G	+	+										
~~0.00287~~	~~H~~												
0.63	I												
0.00177	J												
~~0.0026~~	~~K~~												
~~0.00263~~	~~L~~												

Color	Meaning
	LD Above Threshold
	Index SNP

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

LD clumping example [4]

P-Value	SNP	A	B	C	D	E	F	G	H	I	J	K	L
~~0.418~~	~~A~~	■						■					
~~0.7~~	~~B~~		■					■					
~~0.0459~~	~~C~~			■			■						
~~0.0358~~	~~D~~				■		■						
~~6.35E-63~~	~~E~~					■	■						
2.29E-63	F			■	■	■	■		■				
0.00485	G	■	■					■					
~~0.00287~~	~~H~~						■		■	+			
0.63	I								+	+		+	+
0.00177	J										■	■	■
~~0.0026~~	~~K~~									+	■	■	
~~0.00263~~	~~L~~									+	■		■

Color	Meaning
■	LD Above Threshold
■	Index SNP

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

LD Clumping example [5]

P-Value	SNP	A	B	C	D	E	F	G	H	I	J	K	L
~~0.418~~	~~A~~												
~~0.7~~	~~B~~												
~~0.0459~~	~~C~~												
~~0.0358~~	~~D~~												
~~6.35E-63~~	~~E~~												
2.29E-63	F												
0.00485	G												
~~0.00287~~	~~H~~												
0.63	I												
0.00177	J												
~~0.0026~~	~~K~~												
~~0.00263~~	~~L~~												

Color	Meaning
	LD Above Threshold
	Index SNP

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

Table S4: Clumping and LD parameters in the full AWIGen target dataset before splitting into training and validation

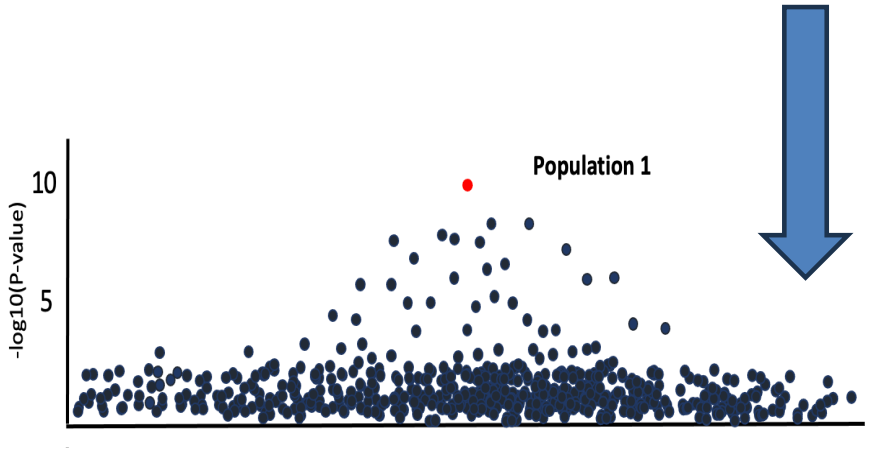
Clumping-kb r2	Best PT	PRS R2	Null Model	Full Model	P -value	Number of SNPs
250kb r2 0.1	0.1161	0.0052	0.2974	0.3026	4.22E-08	248432
250kb r2 0.5	0.0549	0.0082	0.2974	0.3056	1.21E-27	317283
250kb r2 0.7	0.0286	0.0089	0.2974	0.3063	4.41E-30	360599
250kb r2 0.8	0.0471	0.0093	0.2974	0.3067	2.12E-31	494366
500kb r2 0.1	0.0900	0.0047	0.2974	0.3021	1.33E-16	207339
500kb r2 0.3	0.0535	0.0078	0.2974	0.3052	2.12E-08	260888
1000kb r2 0.3	0.0535	0.0077	0.2974	0.3051	4.21E-26	260116

Over-estimated effect size

- LD-based clumping often removes partially independent associations.

- Genome-wide significant effect sizes estimates are **overestimated on average.**

- **Shrink effect sizes**:
  - More realistic
  - Generalisable to other samples



GPS derivation and testing for five common, complex diseases

Disease	Discovery GWAS (*n*)	Prevalence in validation dataset	Prevalence in testing dataset	Polymorphisms in GPS	Tuning parameter	AUC (95% CI) in validation dataset	AUC (95% CI) in testing dataset
CAD	60,801 cases; 123,504 controls16	3,963/120,280 (3.4%)	8,676/288,978 (3.0%)	6,630,150	LDPred ($\rho = 0.001$)	0.81 (0.80–0.81)	0.81 (0.81–0.81)
Atrial fibrillation	17,931 cases; 115,142 controls30	2,024/120,280 (1.7%)	4,576/288,978 (1.6%)	6,730,541	LDPred ($\rho = 0.003$)	0.77 (0.76–0.78)	0.77 (0.76–0.77)
Type 2 diabetes	26,676 cases; 132,532 controls31	2,785/120,280 (2.4%)	5,853/288,978 (2.0%)	6,917,436	LDPred ($\rho = 0.01$)	0.72 (0.72–0.73)	0.73 (0.72–0.73)
Inflammatory bowel disease	12,882 cases; 21,770 controls32	1,360/120,280 (1.1%)	3,102/288,978 (1.1%)	6,907,112	LDPred ($\rho = 0.1$)	0.63 (0.62–0.65)	0.63 (0.62–0.64)
Breast cancer	122,977 cases; 105,974 controls33	2,576/63,347 (4.1%)	6,586/157,895 (4.2%)	5,218	Pruning and thresholding ($r^2 < 0.2$; $P < 5 \times 10^{-4}$)	0.68 (0.67–0.69)	0.69 (0.68–0.69)

AUC was determined using a logistic regression model adjusted for age, sex, genotyping array, and the first four principal components of ancestry. The breast cancer analysis was restricted to female participants. For the LDPred algorithm, the tuning parameter ρ reflects the proportion of polymorphisms assumed to be causal for the disease. For the pruning and thresholding strategy, r^2 reflects the degree of independence from other variants in the linkage disequilibrium reference panel, and P reflects the P value noted for a given variant in the discovery GWAS. CI, confidence interval.

PRSCS, PRSCSx, MegaPRS, and many more…

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

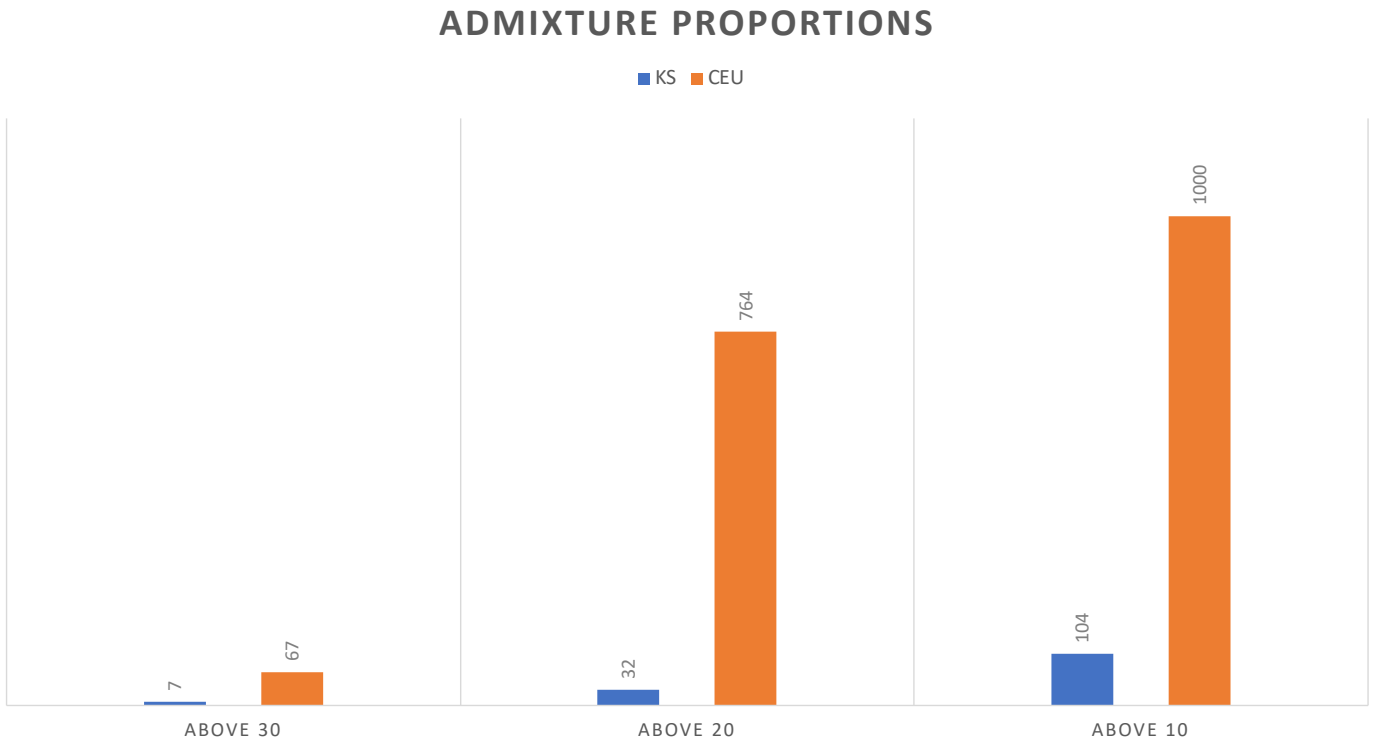
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

CLUMPING AND THRESHOLDING	ALTERNATIVE METHODS
Addresses LD by clumping (removing variants in LD with a variant with a lower p-value)	Addresses LD by modelling its effects on effect sizes (as if all variants were modelled jointly)
Secondary signals are (mostly) lost	Secondary signals are retained
Does not address over-estimation	Addresses over-estimation by shrinking towards zero effect

Admixture for PRS in Africa

Black south African sub-sample from the AWI-Gen Study



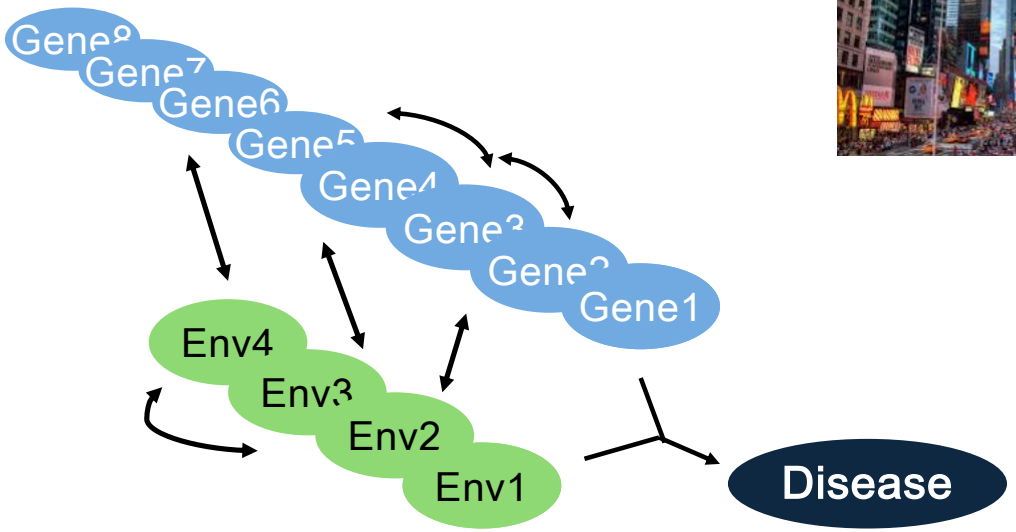
International Journal of Epidemiology, 2025, **54(1)**, dyae173
https://doi.org/10.1093/ije/dyae173
Cohort Profile

Cohort Profile

Cohort Profile: Africa Wits-INDEPTH partnership for Genomic studies (AWI-Gen) in four sub-Saharan African countries

Furahini Tluway[1], Godfred Agongo[2,3], Vukosi Baloyi[4], Palwende Romuald Boua[1,5], Isaac Kisiangani[6], Moussa Lingani[5], Reneilwe Given Mashaba[7], Shukri F Mohamed[6], Engelbert A Nonterah[3], Cairo Bruce Ntimana[7], Toussaint Rouamba[5], Theophilous Mathema[1], Siyanda Madala[1,8], Dylan G Maghini[1,9], Ananyo Choudhury[1], Nigel J Crowther[10], Scott Hazelhurst[1,11], Dhriti Sengupta[1], Patrick Ansah[3], Solomon Simon Rampai Choma[12], Cornelius Debpuur[13], F Xavier Gómez-Olivé[14], Kathleen Kahn[14], Lisa K Micklesfield[4], Shane A Norris[4,15], Abraham R Oduro[13], Hermann Sorgho[5], Paulina Tindana[16], Halidou Tinto[5], Stephen Tollman[14], Alisha Wade[14], Michèle Ramsay[1,*] and as members of AWI-Gen and the H3Africa Consortium

Cultural and Socioeconomic Differences



Africa: Effect of smoking

Novel and Known Gene-Smoking Interactions With cIMT Identified as Potential Drivers for Atherosclerosis Risk in West-African Populations of the AWI-Gen Study

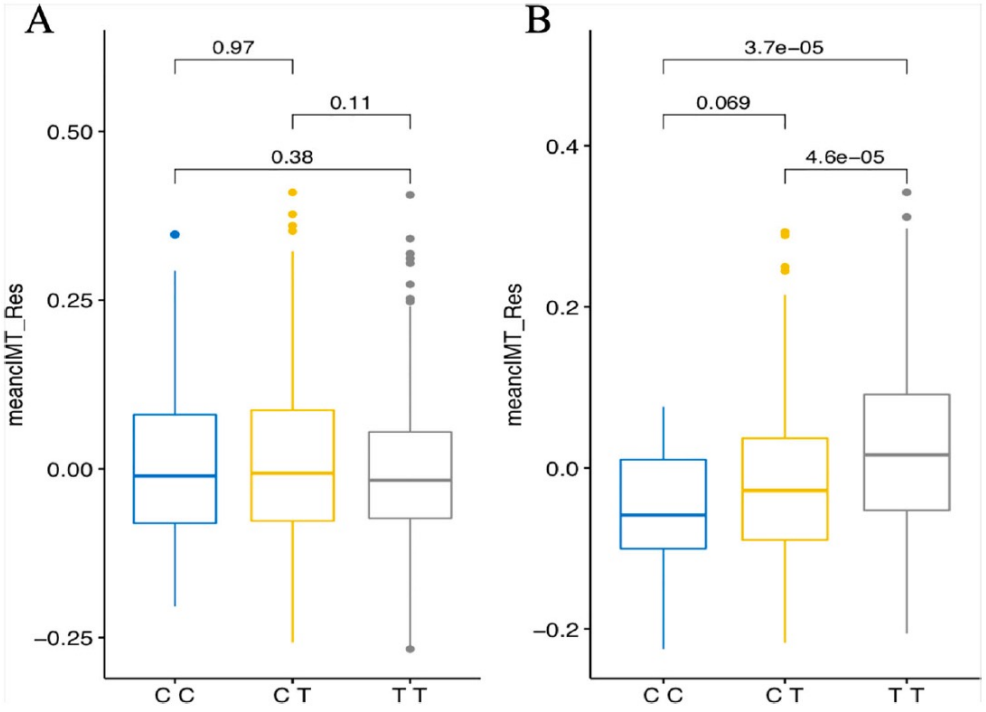
Palwende Romuald Boua[1,2,3], Jean-Tristan Brandenburg[2], Ananyo Choudhury[2], Scott Hazelhurst[2,4], Dhriti Sengupta[2], Godfred Agongo[2,3,5], Engelbert A. Nonterah[5,6], Abraham R. Oduro[5], Halidou Tinto[1], Christopher G. Mathew[2,7], Hermann Sorgho[1] and Michèle Ramsay[2,3]*

Effect of gene – smoking interaction on cIMT

Study restricted to ~2000 West African men

Cultural considerations

Genetic marker:rs1192824
Intergenic region between *TBC1D8* and *CNOT11*



Non-smokers Smokers

Boua et al. Frontiers in Genetics (2020) 10

Gender and menopause

nature COMMUNICATIONS

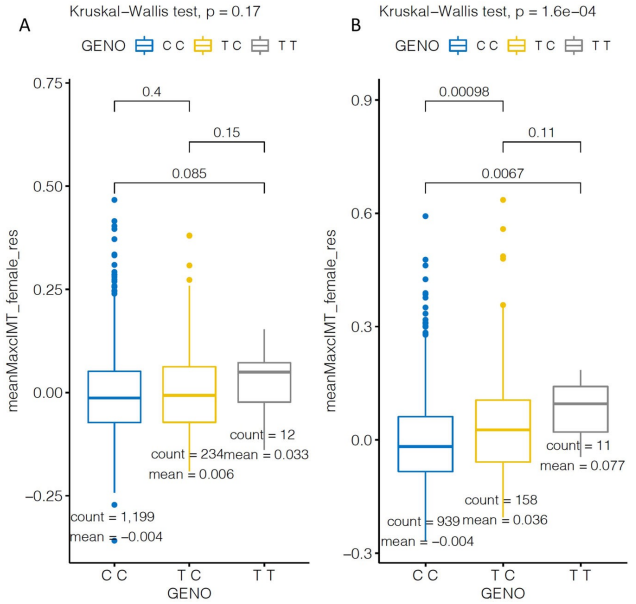
ARTICLE

https://doi.org/10.1038/s41467-022-28276-x OPEN

Check for updates

Genetic associations with carotid intima-media thickness link to atherosclerosis with sex-specific effects in sub-Saharan Africans

Palwende Romuald Boua[1,2,3✉], Jean-Tristan Brandenburg[2], Ananyo Choudhury[2], Hermann Sorgho[1], Engelbert A. Nonterah[4,5], Godfred Agongo[4,6], Gershim Asiki[7], Lisa Micklesfield[8], Solomon Choma[9], Francesc Xavier Gómez-Olivé[10], Scott Hazelhurst[11], Halidou Tinto[1], Nigel J. Crowther[12], Christopher G. Mathew[2,3,13], Michèle Ramsay[2,3✉], AWI-Gen Study* & the H3Africa Consortium*



Combination of modifiers

MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

Social Determinants of Health

JACC

JACC Journals › JACC › Archives › Vol. 84 No. 22

Next

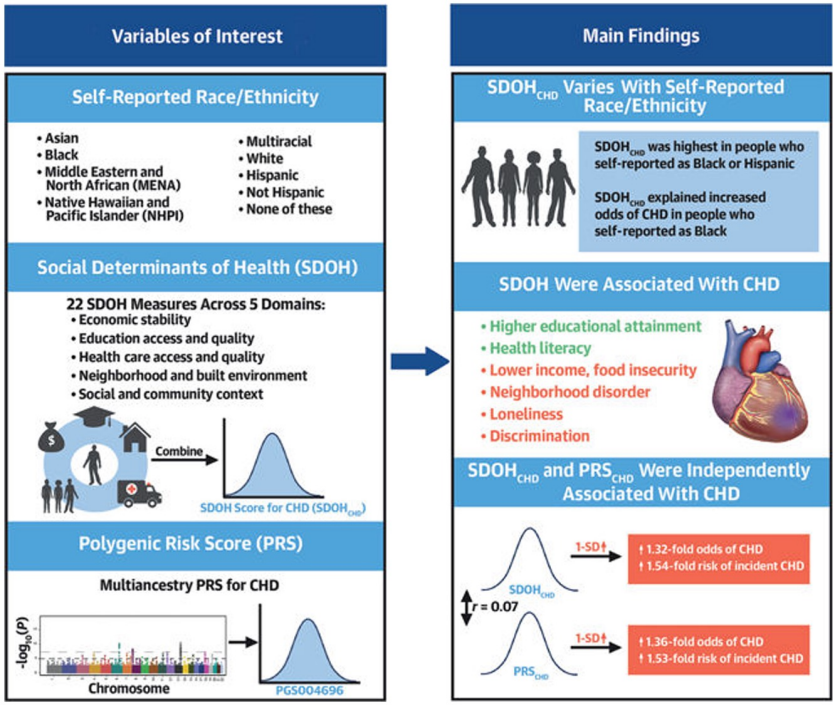
Associations of Self-Reported Race, Social Determinants of Health, and Polygenic Risk With Coronary Heart Disease GET ACCESS

Original Research

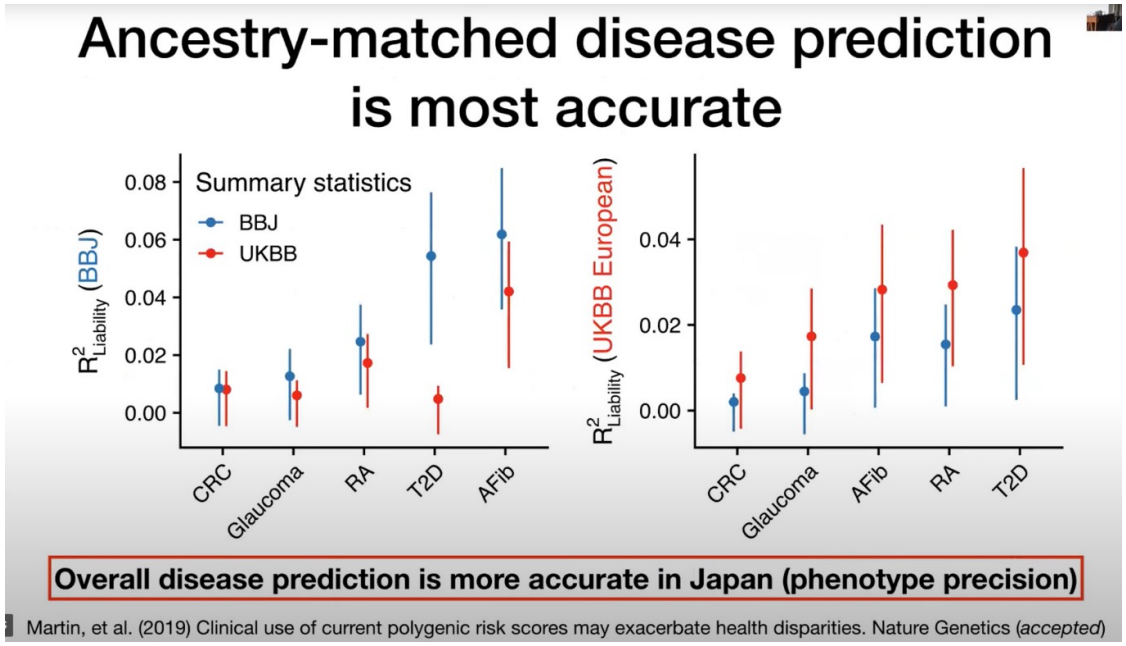
Kristjan Norland, Daniel J. Schaid, Mohammadreza Naderian, Jie Na, and Iftikhar J. Kullo

JACC. 2024 Nov, 84 (22) 2157–2166

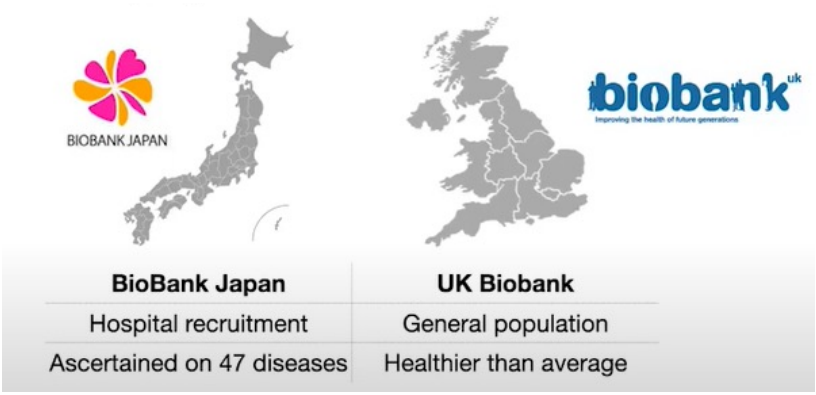
Including both PRS and SDOH in CHD risk models could improve their accuracy.



The impact of what data you use: complexity of comparing across cohorts/populations



Martin, et al. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics (*accepted*)



BioBank Japan	UK Biobank
Hospital recruitment	General population
Ascertained on 47 diseases	Healthier than average

Impact of large African Cohorts

Circulation: Genomic and Precision Medicine

ORIGINAL ARTICLE

Development and Validation of Polygenic Risk Scores for Blood Pressure Traits in Continental African Populations

Ebuka Onyenobi, MS; Michael Zhong, BS; Opeyemi Soremekun, PhD; Abram Kamiza, PhD; Romuald Boua, PhD; Tinashe Chikwore, PhD; ACCME Research Group*; Segun Fatumo, PhD; Ananyo Choudhury, PhD; Scott Hazelhurst, PhD; Clement Adebamowo, MBChB, ScD; Michele Ramsay, PhD; Bamidele Tayo, PhD; Jennifer S. Albrecht, PhD; Timothy D. O'Connor, PhD; Yuji Zhang, PhD; Braxton D. Mitchell, PhD; Sally N. Adebamowo, MBBS, ScD

Circ Genom Precis Med. 2025;18:e005048. DOI: 10.1161/CIRCGEN.124.005048

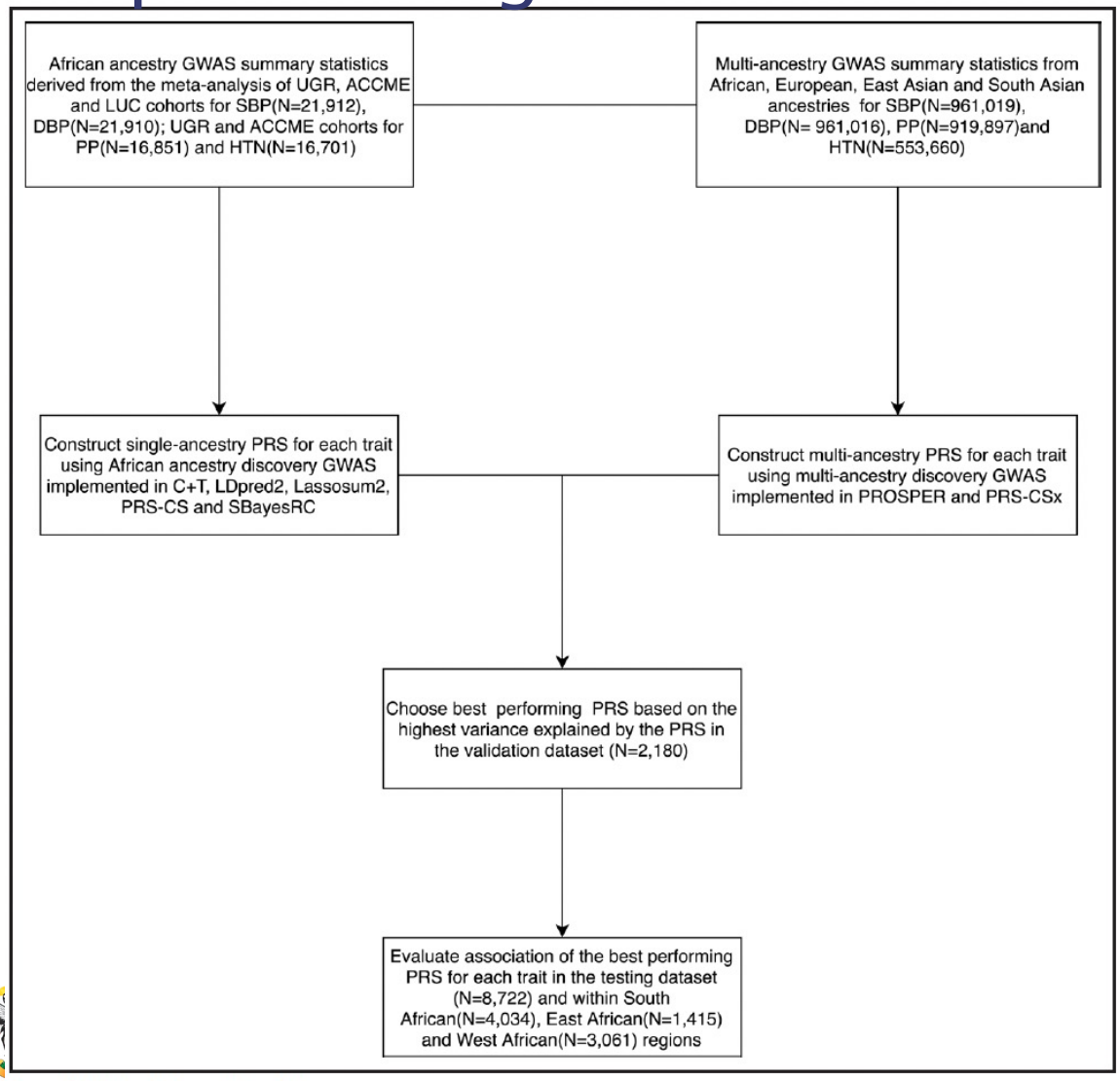
MAKERERE UNIVERSITY

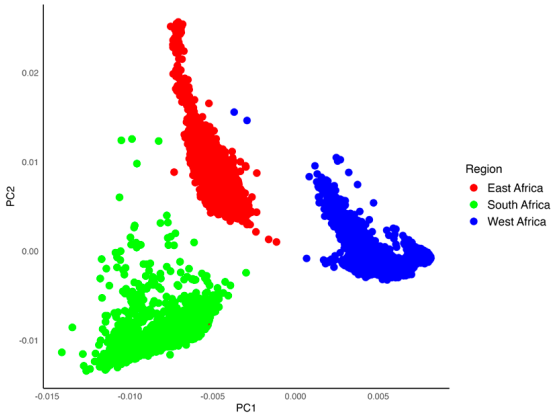
IDI INFECTIOUS DISEASES INSTITUTE

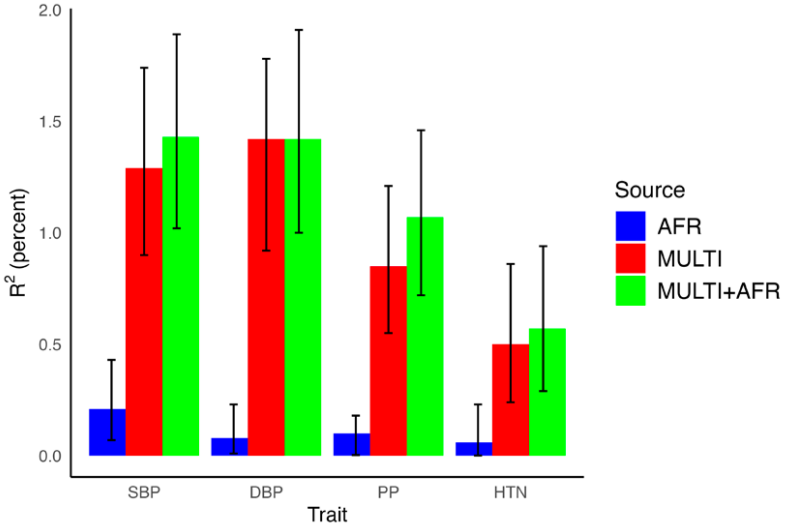
AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

Impact of large African Cohorts

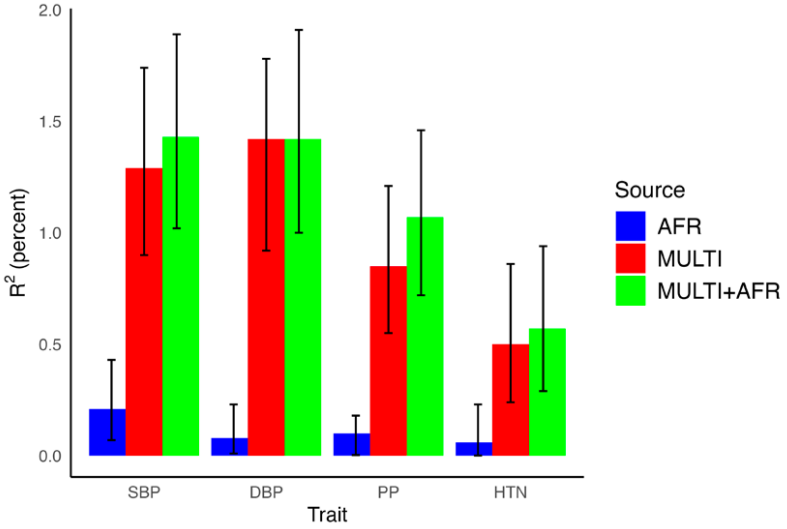


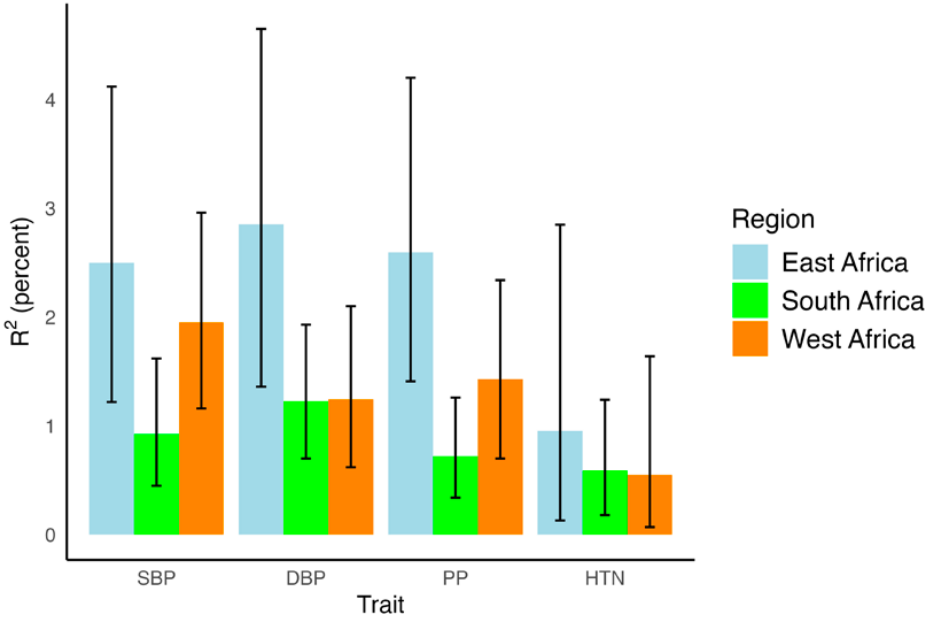




wellcome connecting science

Impact of large African Cohorts





MAKERERE UNIVERSITY

IDI INFECTIOUS DISEASES INSTITUTE

AFRICAN CENTER OF EXCELLENCE IN BIOINFORMATICS & DATA INTENSIVE SCIENCES

wellcome connecting science

Polygenic scores do not generalise well across ancestry groups

- Why?
  - Differences in LD patterns
  - Differences in allele frequencies
  - Different causal variants
  - Differences in environments
  - Vast majority of GWAS are performed in European ancestry individuals.

Perspective | Published: 29 March 2019

Clinical use of current polygenic risk scores may exacerbate health disparities

Alicia R. Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale & Mark J. Daly

Nature Genetics **51**, 584–591(2019) | Cite this article

Considerations for improving PRS portability

Include diverse populations in GWAS:

- More diverse genetic data to uncover a broader spectrum of variants and their associations.
- Encourage international collaborations to include underrepresented populations.

Develop of multi-Ancestry PRS:

- Construct PRS using multiple ancestries to improve predictive accuracy and applicability.
- Use and develop methods that integrate data across populations, accounting for population stratification and heterogeneity in allele frequencies and LD patterns.

Refine statistical models:

- Incorporate novel that can adaptively learn from diverse datasets and account for non-linear interactions between genetic markers.

Integrate environmental and socioeconomic data:

- Better model gene-environment interactions.

Key papers

1. **Martin, A. R., et al. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics.** LINK
   This paper discusses the potential for current PRS to exacerbate health disparities due to their predominant development in European populations. It's an essential read for understanding the ethical and practical implications of PRS in clinical settings.
2. **Duncan, L., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. Nature Communications.** LINK
   This research explores the performance of polygenic risk scores across diverse populations, providing insights into the limitations and challenges of PRS developed using predominantly European genetic data.
3. **Curtis, D. (2018). Polygenic risk score and prediction of treatment outcomes. Molecular Psychiatry.** LINK
   This article reviews the application of polygenic risk scores in predicting treatment outcomes, including the impact of using PRS in diverse populations.
4. **Visscher, P. M., et al. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics.** LINK
   This review covers a decade of GWAS discoveries, discussing the implications for biology, function, and the translation of GWAS findings into clinical practice, including the use of PRS.
5. **Berg, J. J., et al. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. eLife.** LINK
   This study provides insights into the complexity of genetic adaptation and its implications for polygenic scores, including methodological considerations for assessing polygenic adaptation in height and other traits.
6. **Torkamani, A., et al. (2018). The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics.** LINK
   This review discusses the personal and clinical utility of polygenic risk scores, highlighting their potential and limitations in personalized medicine.

Questions

1. What are Polygenic Risk Scores (PRS), and how are they calculated?
2. Why might a PRS developed using data from a European population not perform as well when applied to an African population?
3. What is linkage disequilibrium, and why is it important in the context of PRS portability?
4. What challenges do population-specific genetic variants pose for the development and application of polygenic risk scores?
5. How do gene-environment interactions impact the accuracy of PRS in different populations?
6. Can you name a method that could improve the portability of PRS across diverse populations?

Answers

1. PRS are calculated by summing up the effects of multiple genetic variants across the genome, each weighted by their association with a trait or disease, as determined by GWAS. The formula typically looks like PRS = Σ (effect size * allele count) for selected SNPs.

2. PRS might underperform in non-European populations due to several factors:
   - Allele Frequency Differences: Certain alleles may be more or less common in non-European populations compared to European ones.
   - LD Differences: The patterns of genetic linkage can vary significantly between populations, affecting which genetic variants are tagged by SNPs included in the PRS.
   - Population-Specific Variants: There may be genetic risk factors relevant to African populations that are not present in European populations and vice versa.

3. LD refers to the non-random association of alleles at different loci in the genome. In the context of PRS, LD is important because it influences how well SNPs can tag or represent other nearby genetic variants. Variations in LD structure between different populations can lead to significant differences in the effectiveness of a PRS when transferred from one population to another, as the same set of SNPs may not tag the same causal variants across populations.

4. Population-specific genetic variants pose significant challenges for PRS because these variants may not be present or have been poorly studied in the populations typically included in GWAS, which are often of European descent. When PRS are applied to other populations, crucial genetic factors that influence disease risk may be missing, leading to inaccurate risk predictions. This limitation affects the predictive accuracy and fairness of PRS, as it may lead to biased health outcomes where individuals from underrepresented populations receive less precise or beneficial healthcare interventions based on PRS. Addressing this challenge requires more inclusive genetic studies that capture the full spectrum of human genetic diversity to ensure that PRS are effective and equitable across different populations.

5. One effective method is the development of multi-ancestry PRS, which incorporates genetic data from multiple populations during the GWAS phase. This approach can help to capture a broader spectrum of genetic diversity and reduce bias towards any single population. Additionally, using advanced statistical techniques such as mixed models or Bayesian methods can help to account for population stratification and improve the robustness and accuracy of PRS across different ancestries.