

African perspectives on GWAS for complex disorders and traits

Day 3: KidneyGenAfrica: 1st Training Workshop

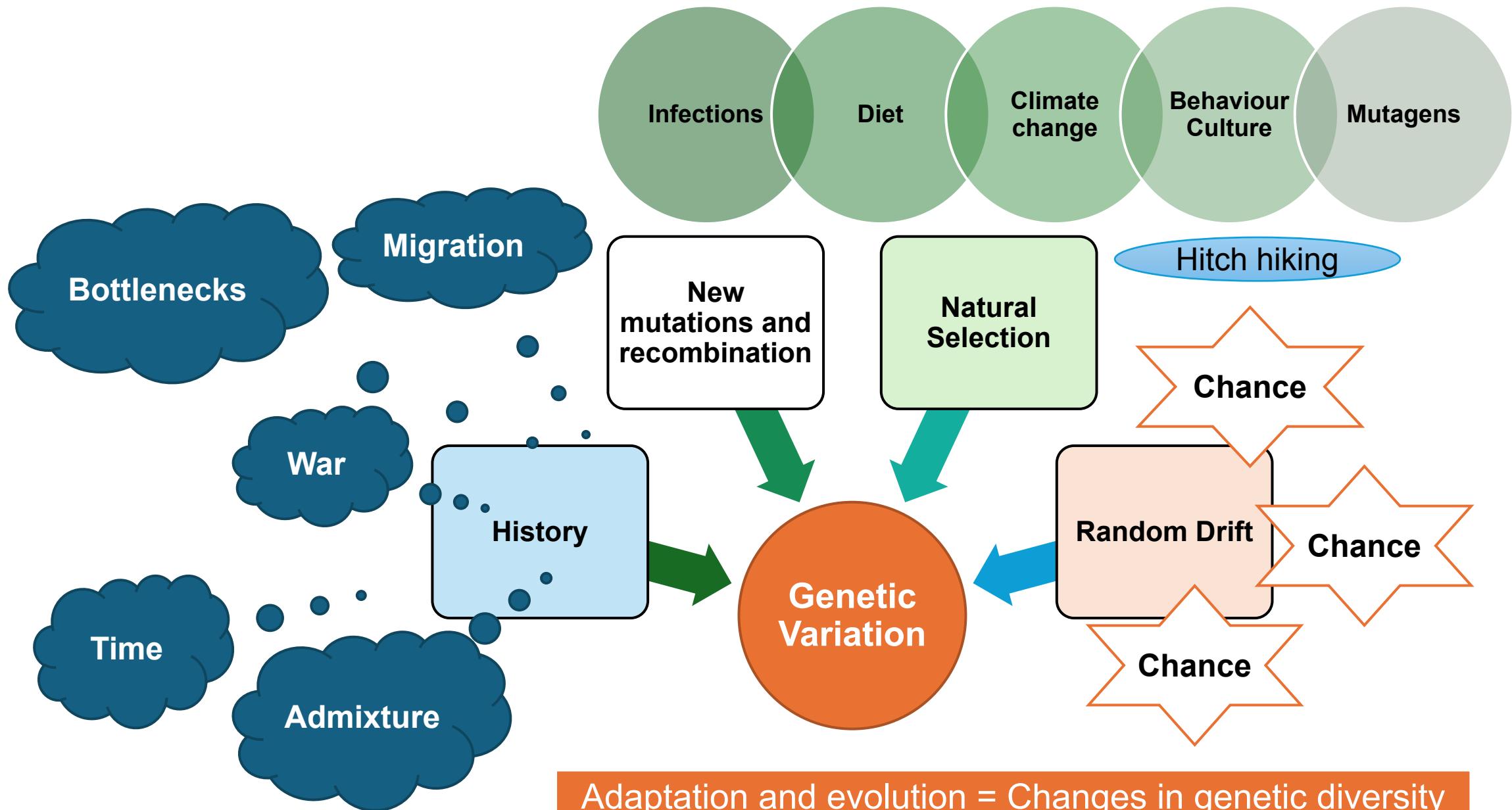
28th January 2026

University of the Witwatersrand, Johannesburg, South Africa

Outline

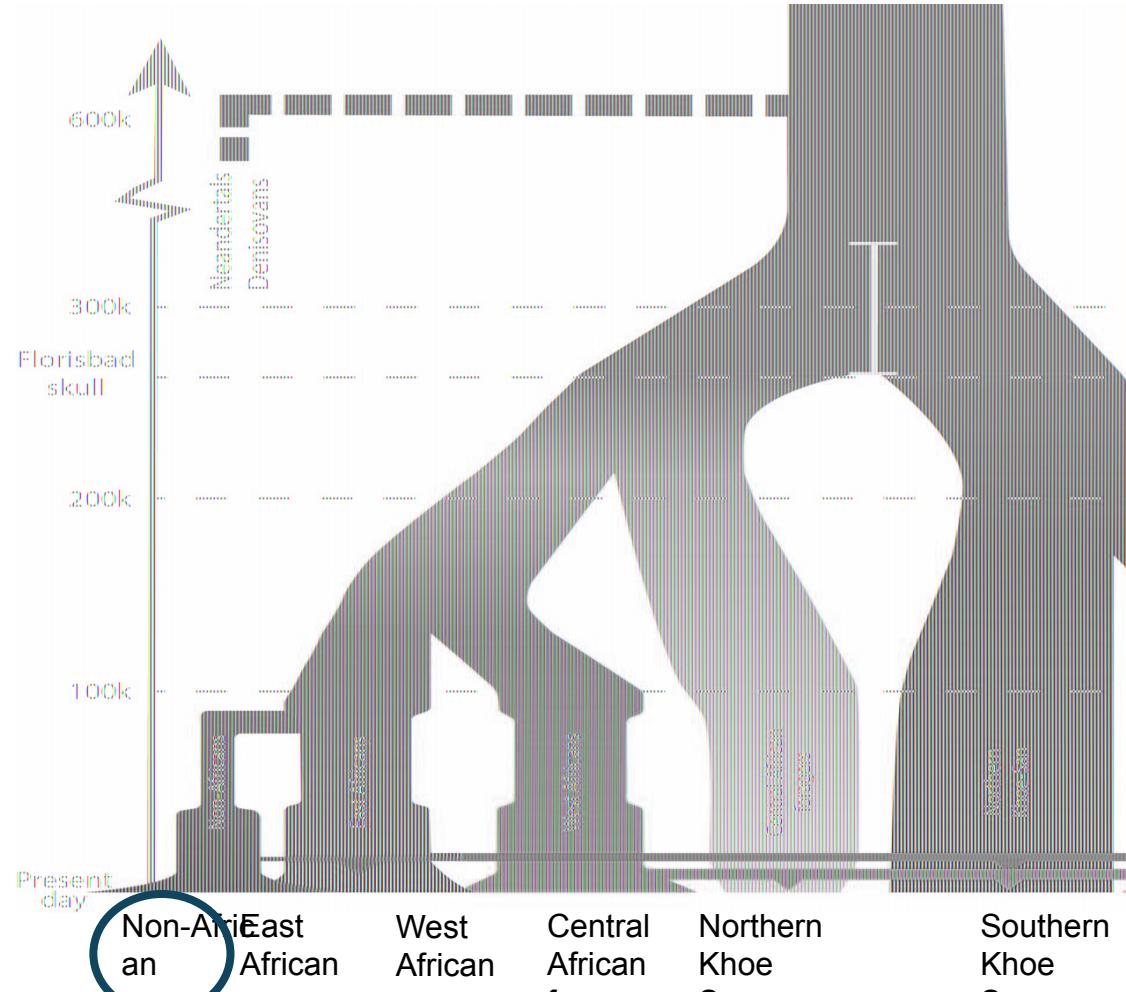
- Increased genetic diversity in African populations
- GWAS considerations in African populations
 - Allele frequency differences and appropriate genotyping arrays
 - Appropriate imputation panels
 - Linkage disequilibrium patterns
 - Replication
 - Sample size
- GWAS examples in African populations (continental and diaspora)
- How far have we come since the start of H3Africa in 2012?
- What to consider when doing GWAS in African populations

Genetic diversity



Adaptation and evolution = Changes in genetic diversity
Presence/absence of variants
Allele frequency differences

Simplified evolutionary model for humans



Schlebusch et al. 2017 Science

GWAS considerations

1. Assumptions about populations made in modeling
2. Data used is often limited and restricted to some populations
3. Effect on interpretations

How do we define African populations?

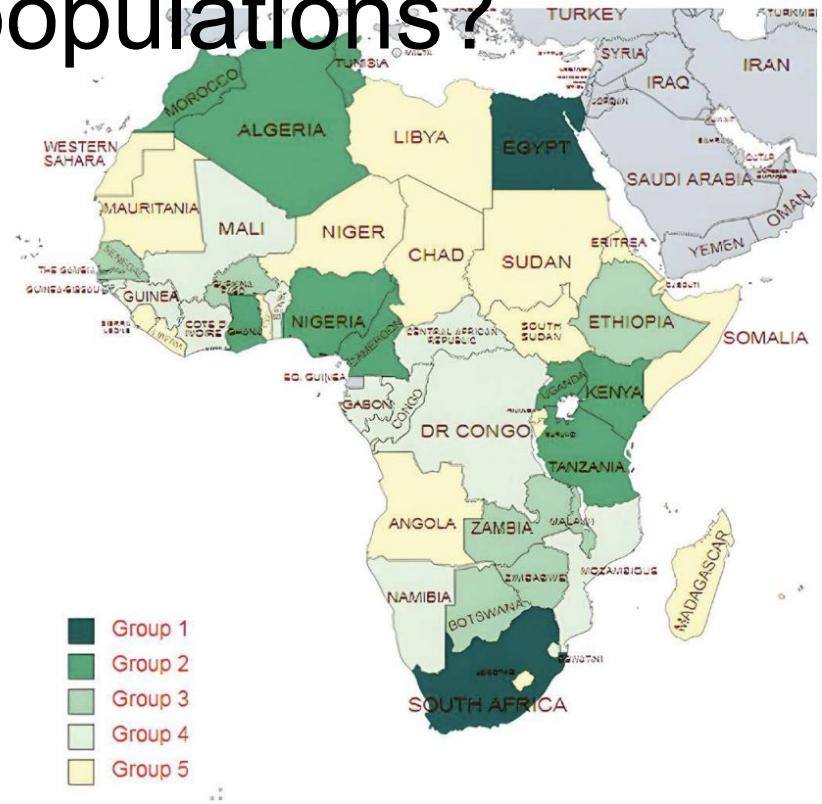
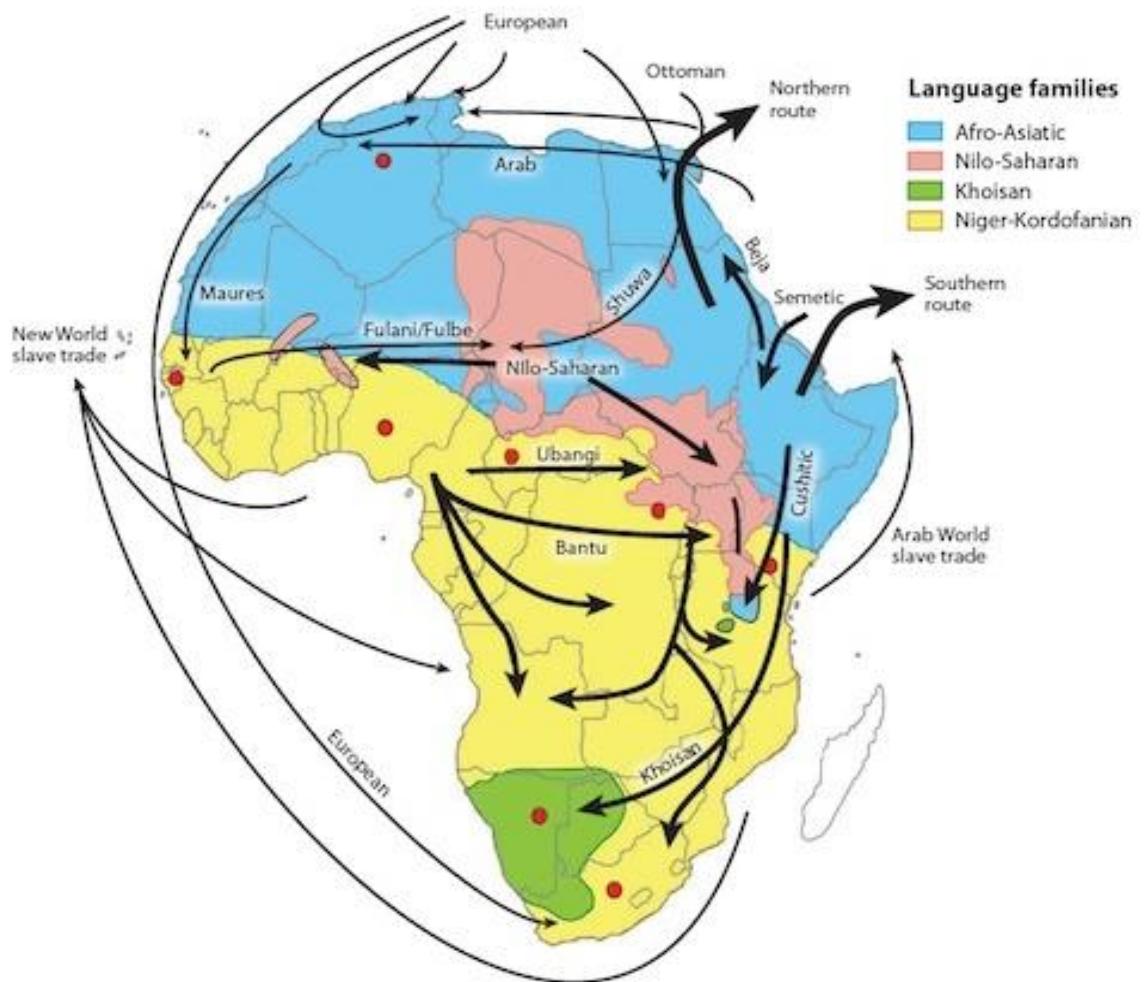


Figure 4: Representation of the PM/genomic capacities in African countries (Source: Sela et al., 2021)

Based on six dimensions of health research

Governance; Financing; Resources; Research Outputs; International collaborations; PM / Genomic research

<https://www.euafrica-permed.eu/>

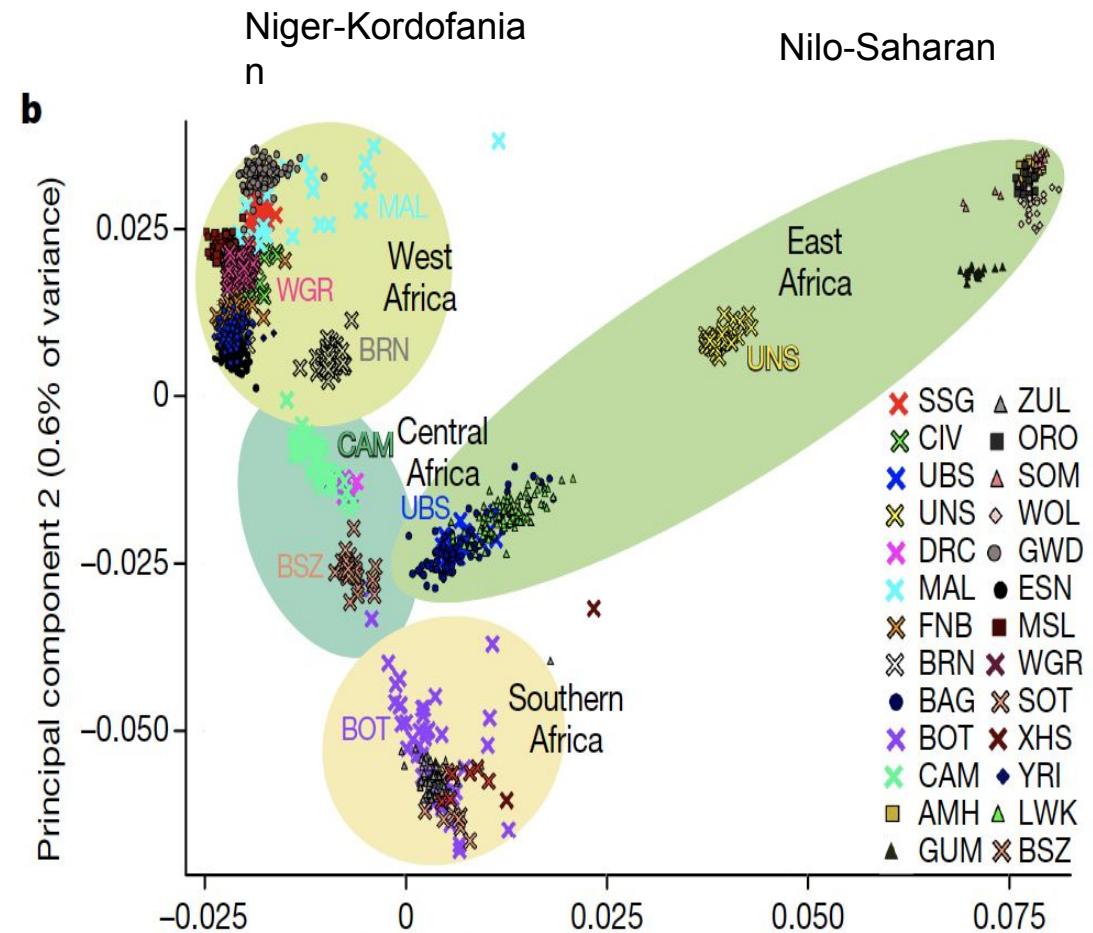
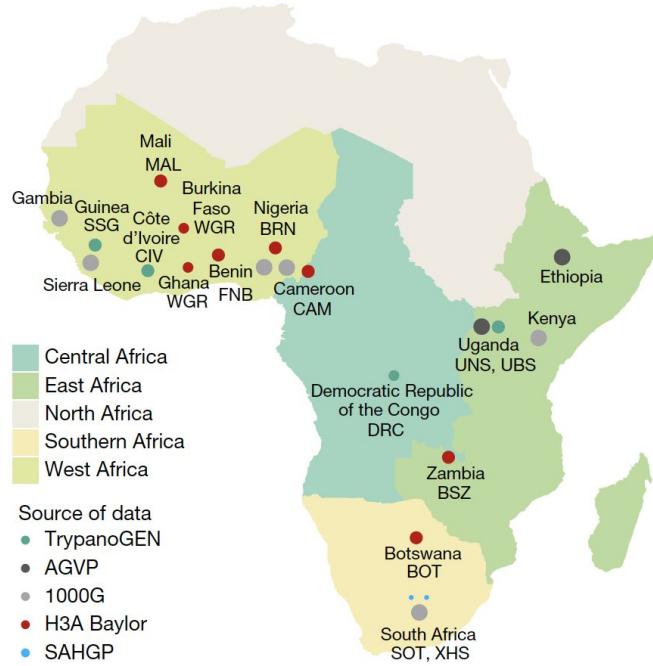


African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping.
Annual Review of Genomics and Human Genetics Vol. 9: 403-433
(2008)

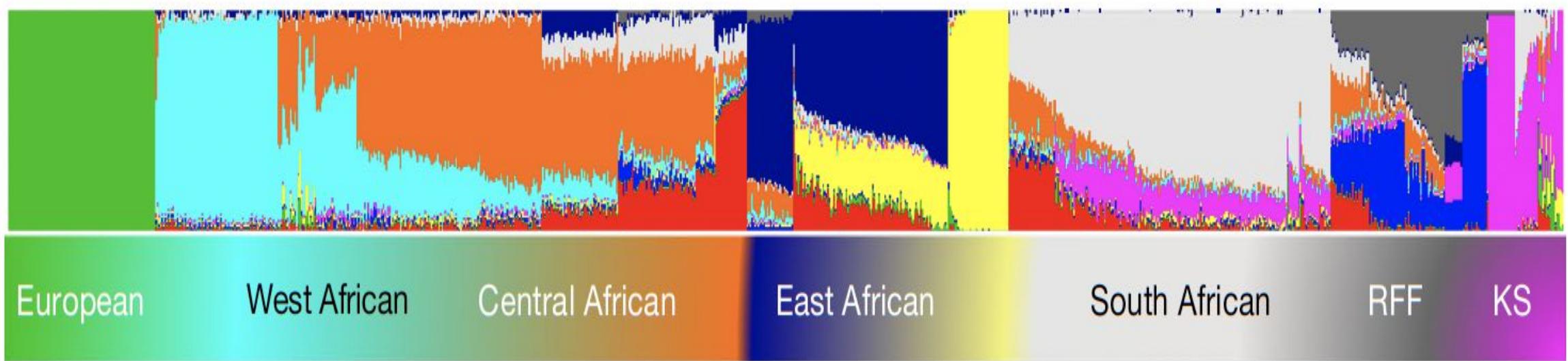
High population structure in Africa



a



High population structure ($k=10$)



Discovery of novel variants in African populations

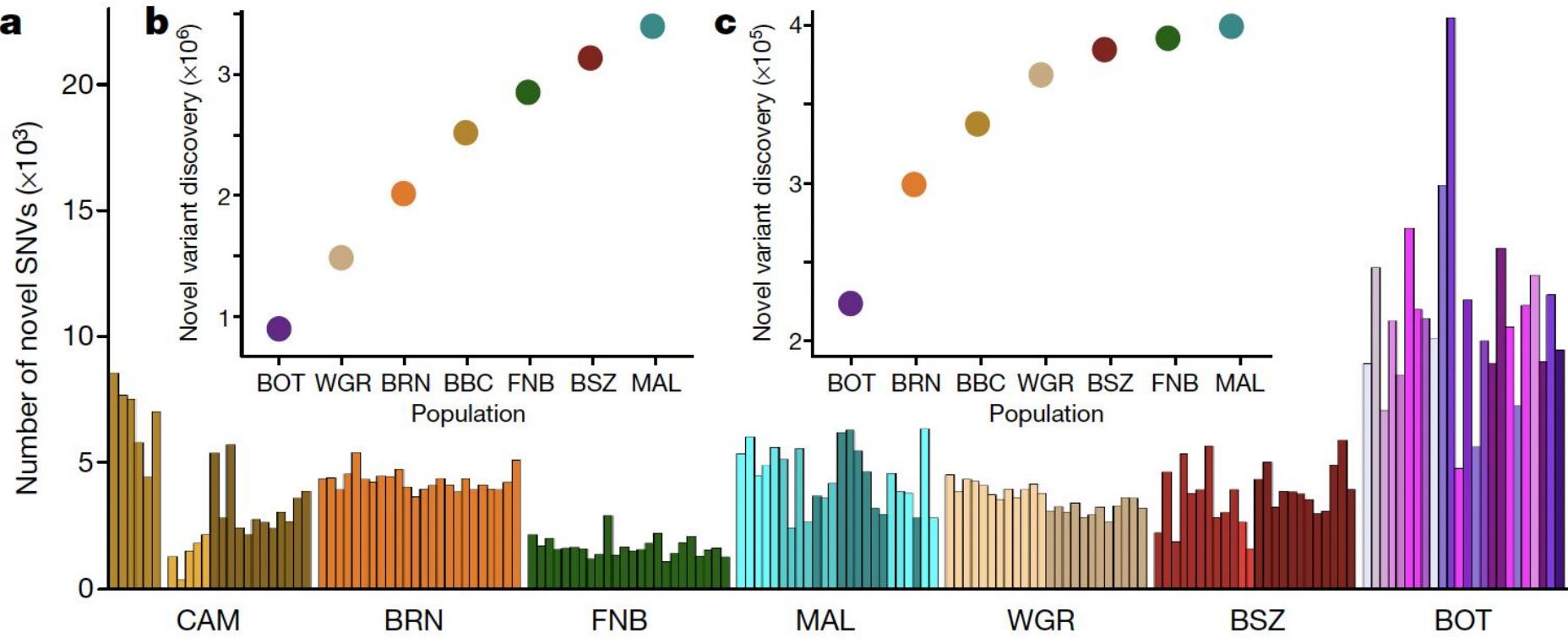
(Much remains unreported)

426 whole genomes from
50 ethnolinguistic groups

41.6 million variants
~12 to 20 million per population

4.3 million novel SNVs
(~88% singletons)
(small sample sizes per population)

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



Not near saturation for novel variant discovery

GWAS considerations in African populations

Read seminal paper - Teo et al. 2010 – explains basics very well and remains conceptually current

Nature Reviews Genetics (2010)



Methodological challenges of genome-wide association analysis in Africa

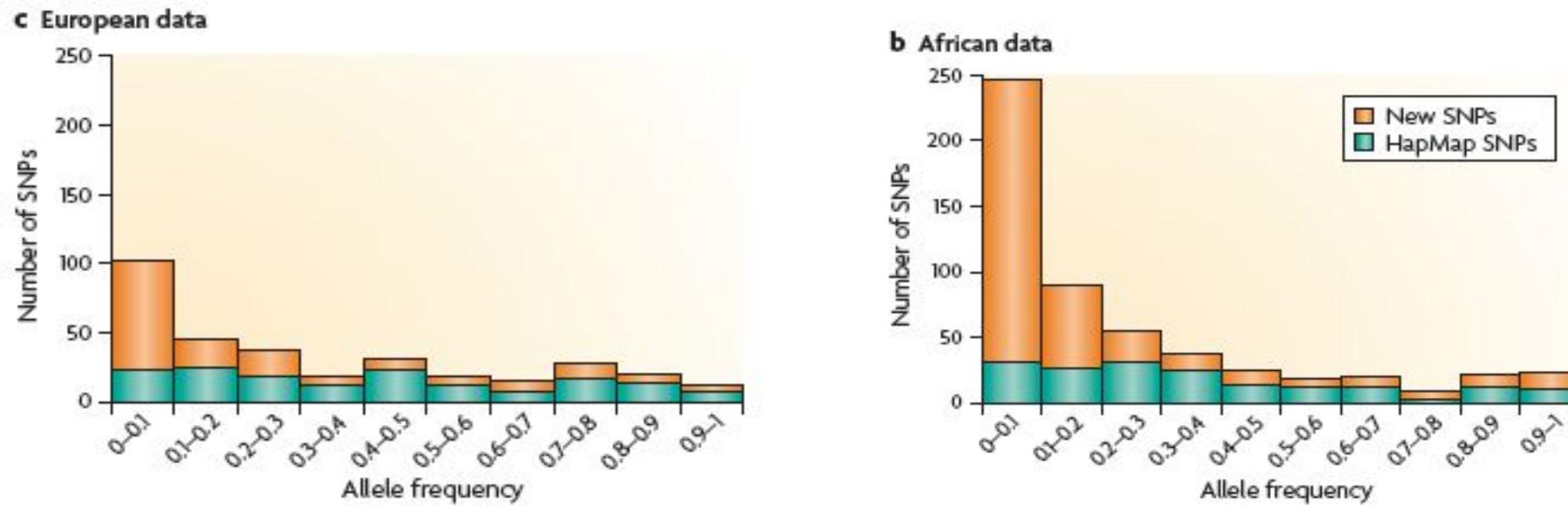
*Yik-Ying Teo *†, Kerrin S. Small *‡ and Dominic P. Kwiatkowski *‡*

Abstract | Medical research in Africa has yet to benefit from the advent of genome-wide association (GWA) analysis, partly because the genotyping tools and statistical methods that have been developed for European and Asian populations struggle to deal with the high levels of genome diversity and population structure in Africa. However, the haplotypic diversity of African populations might help to overcome one of the major roadblocks in GWA research, the fine mapping of causal variants. We review the methodological challenges and consider how GWA studies in Africa will be transformed by new approaches in statistical imputation and large-scale genome sequencing.

GWAS by Linkage Disequilibrium

Stage of analysis	European pops	African pops
Detecting association	High LD increases chance of detecting associations	Low LD reduces likelihood
Replicating association	Good chance of replicating even if causal variant not typed	Reduced likelihood unless causal variant directly typed
Localising causal variant	Can be difficult because of high LD	May be easier because of low LD

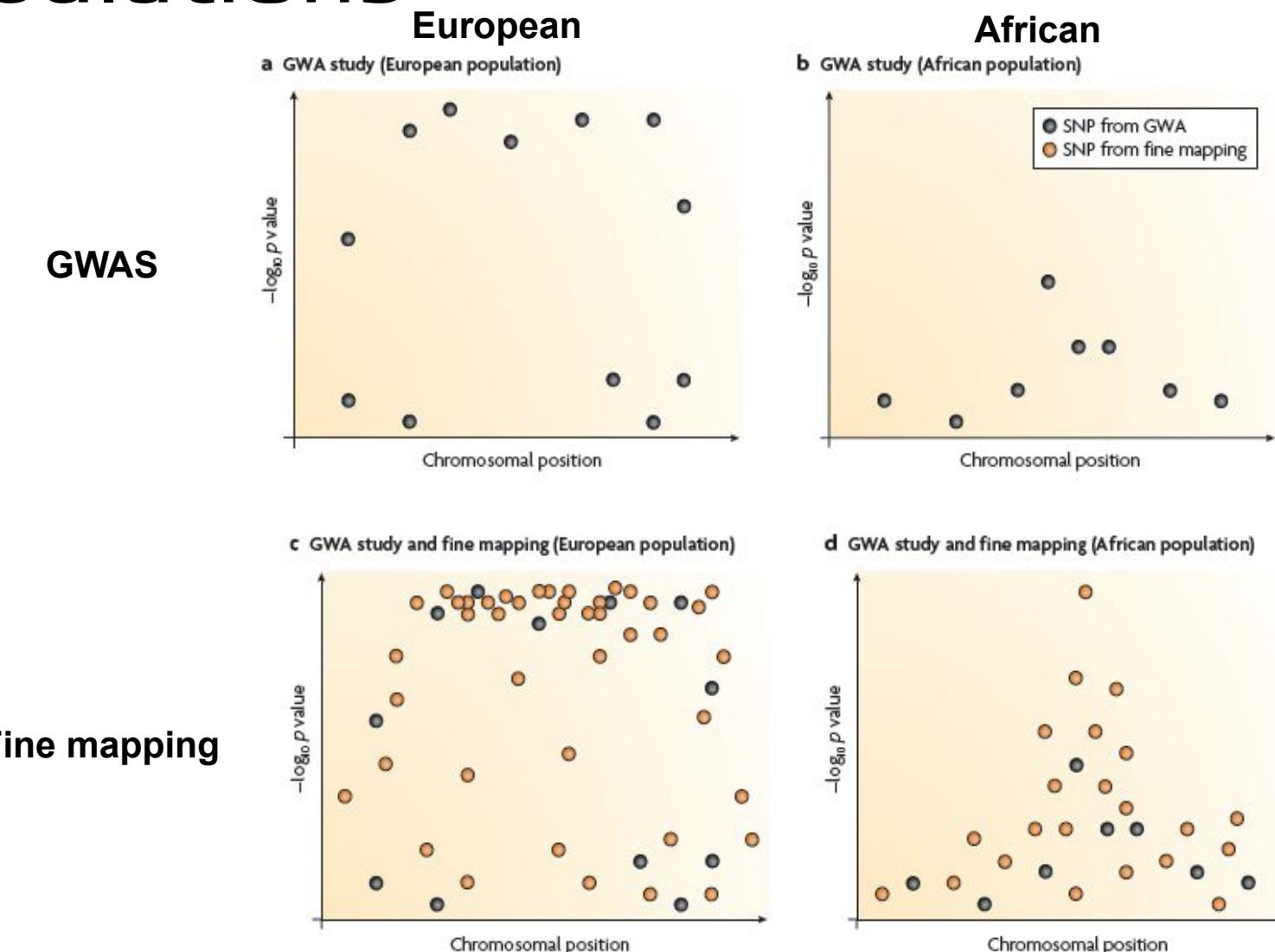
GWAS arrays have more common SNPs for European populations



In African populations you would need about 1.5 million SNPs to give the same coverage as 0.6 million SNPs in European populations

Teo et al. (2010) Nature Reviews, Genetics

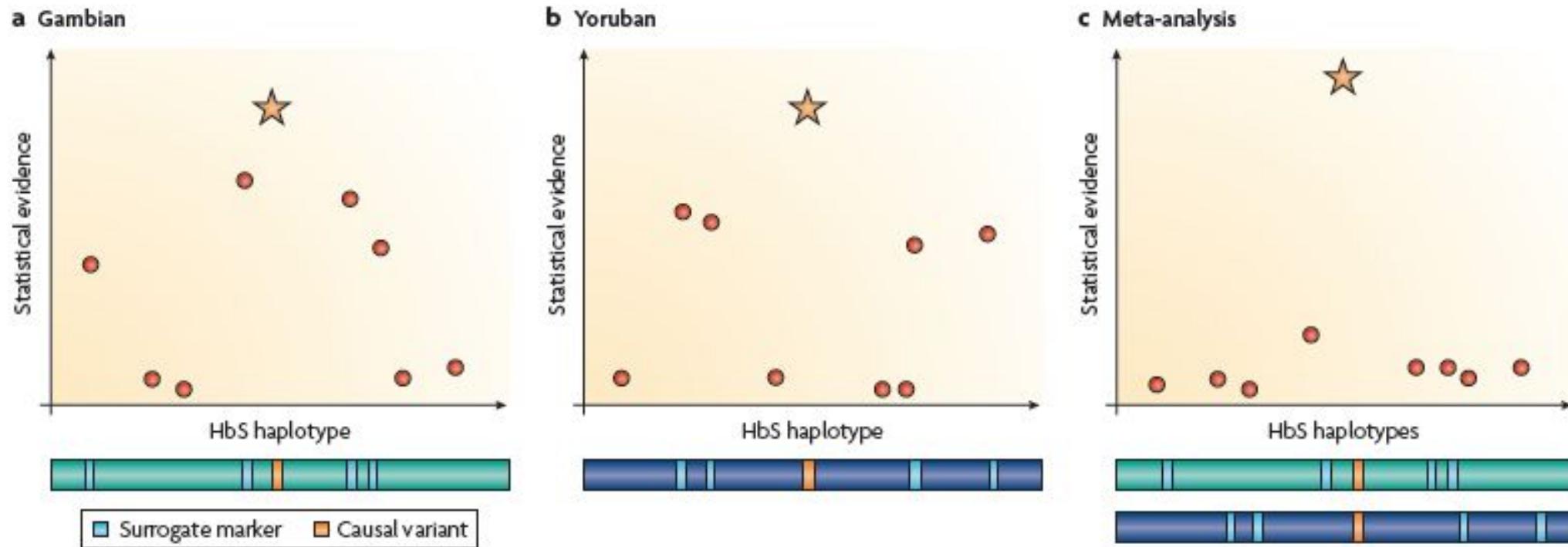
Advantage of studying African populations



African populations have lower linkage disequilibrium (LD) and higher haplotype diversity (more and shorter haplotypes)

Teo et al. (2010) Nature Reviews, Genetics

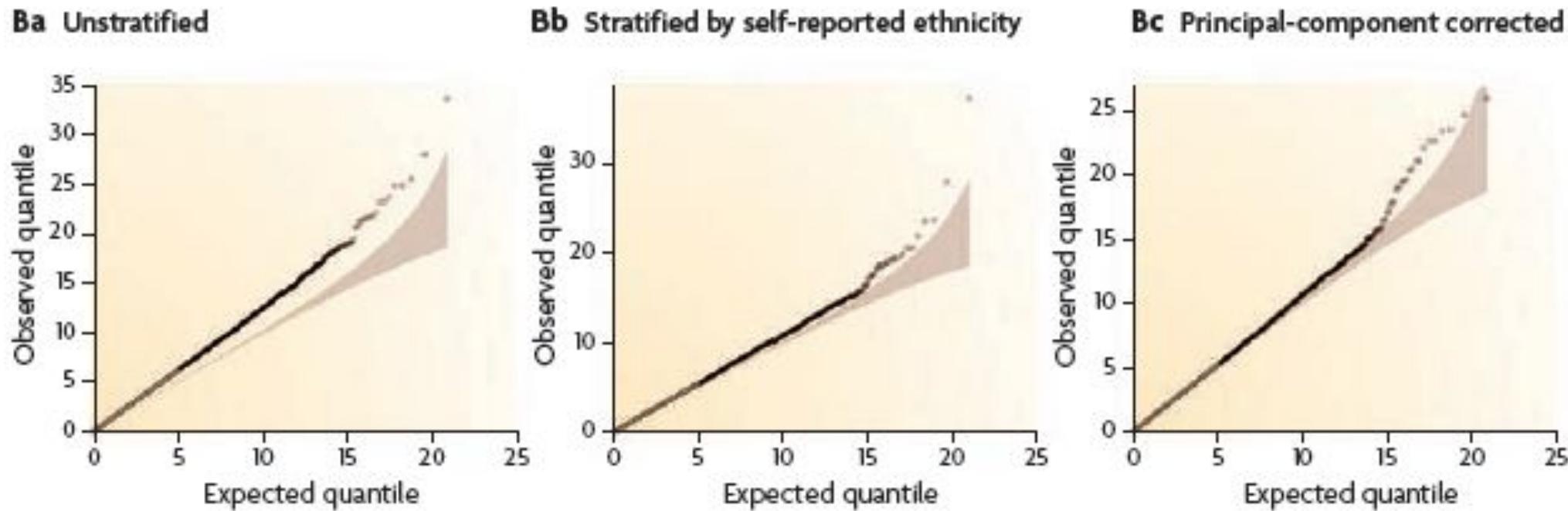
Different chromosomal backgrounds can influence ability to detect associations – Causal variant



Meta-analysis at a site with different associated haplotypes in two populations

Teo et al. (2010) Nature Reviews, Genetics

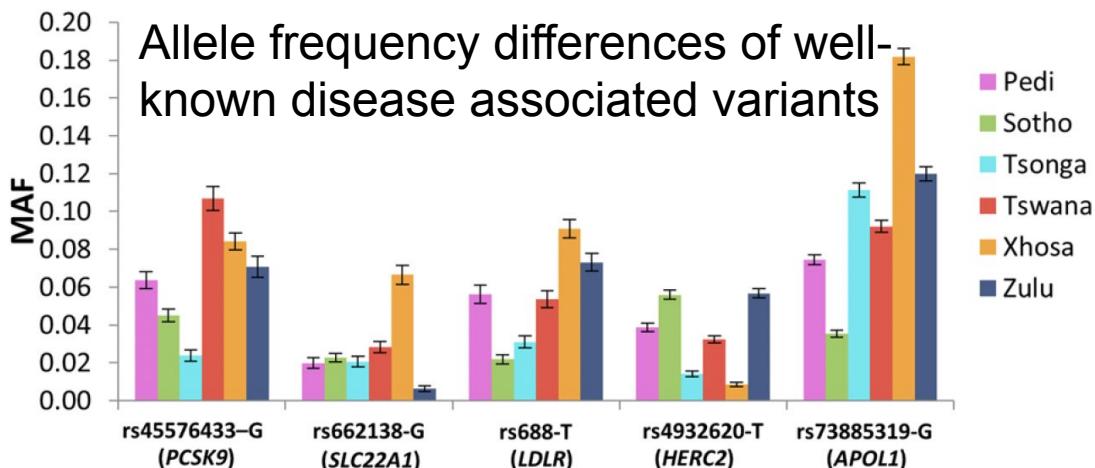
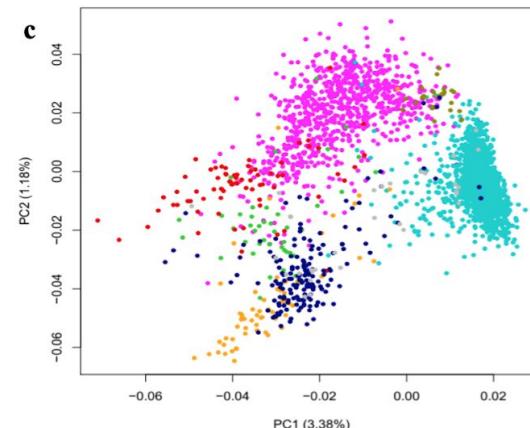
High false discovery rate if you do not correct for population structure (QQ plots)



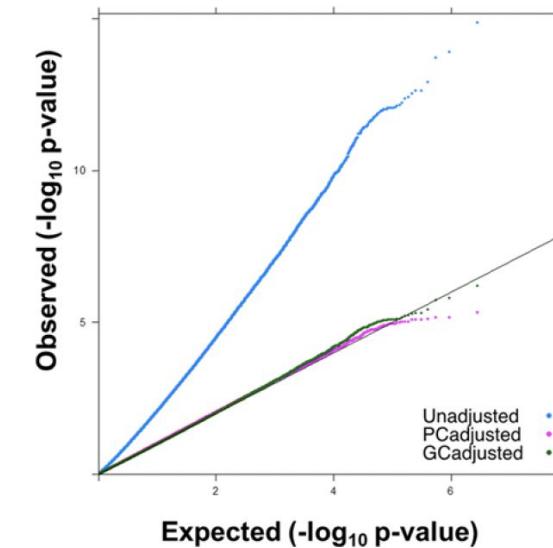
Correcting for local population structure

Teo et al. (2010) Nature Reviews, Genetics

Population sub-structure among South African Bantu-speakers: Impact on GWAS

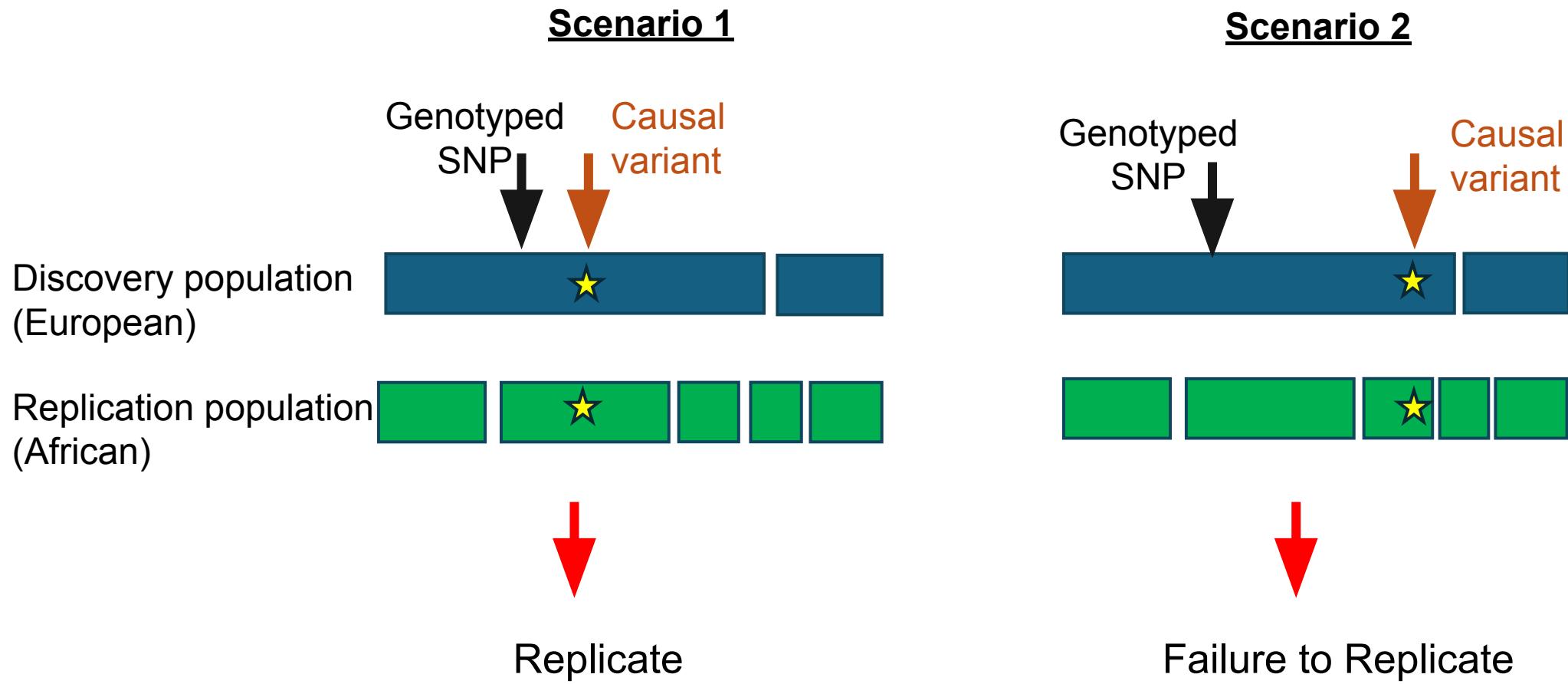


Inflation of association signals if you do not control for genetic ancestry (QQ plot)

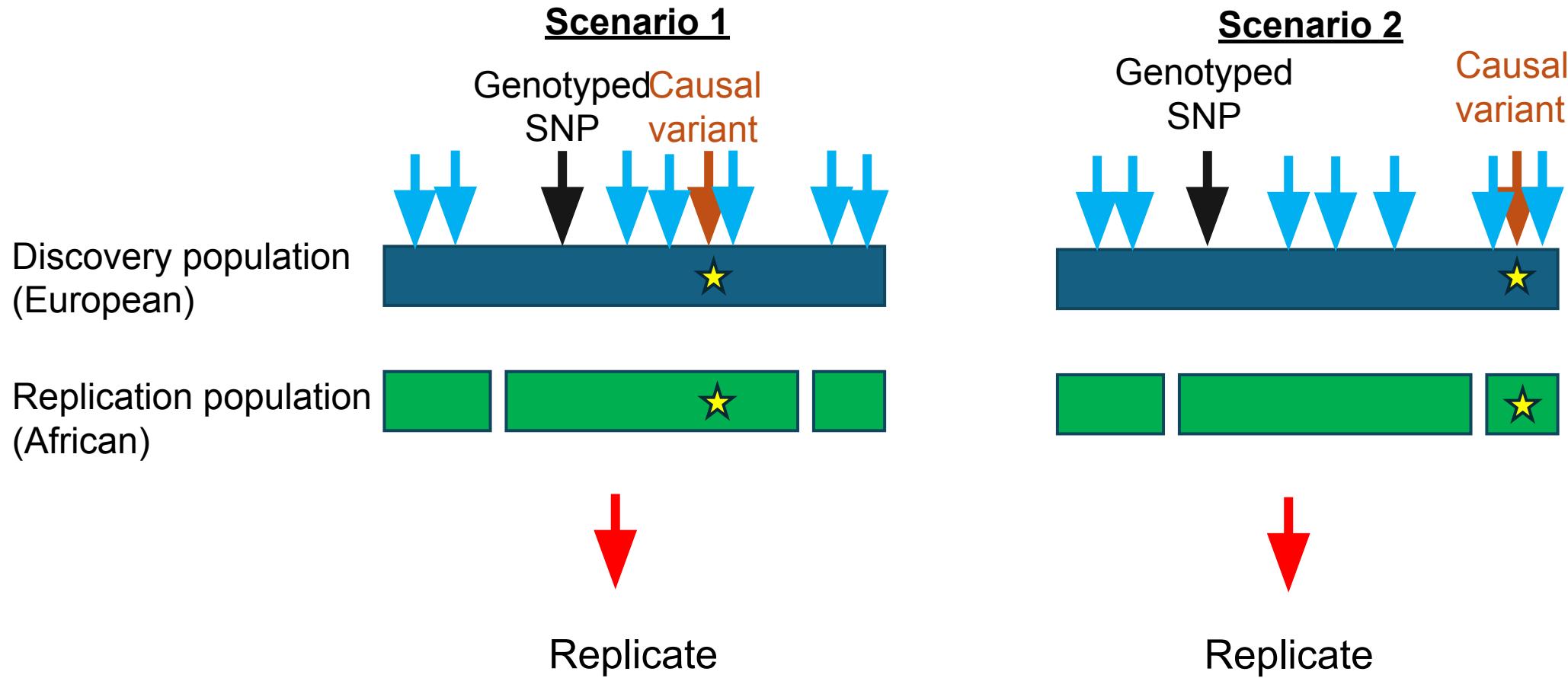


South African Bantu-speaking populations show population structure that could impact genetic susceptibility to complex traits such as diabetes, hypertension and stroke.

Replication

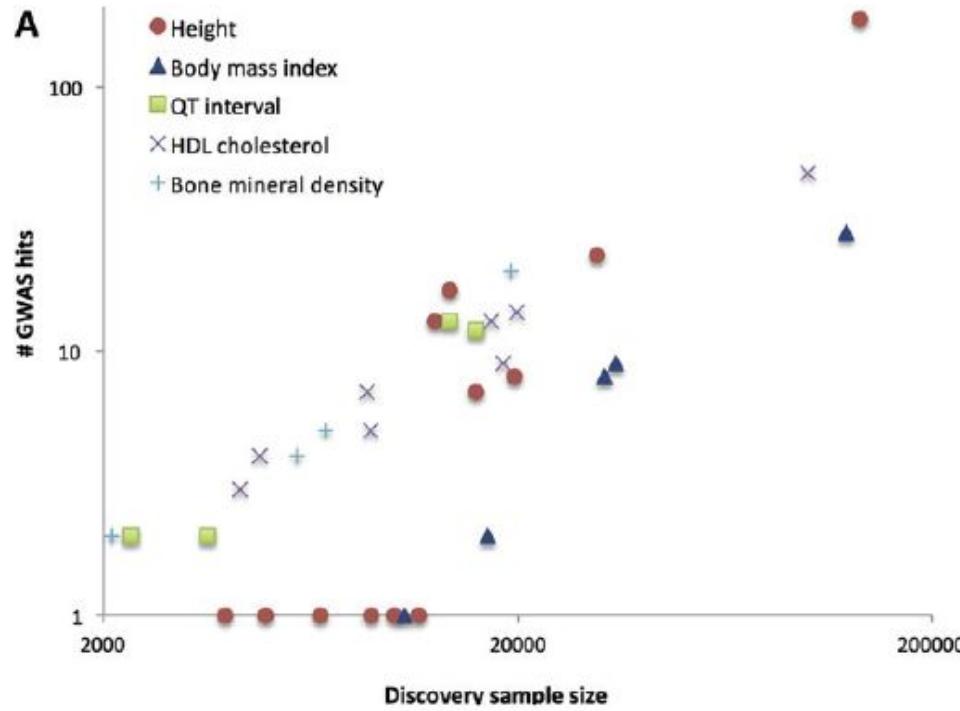


Replication following imputation (higher density SNPs)

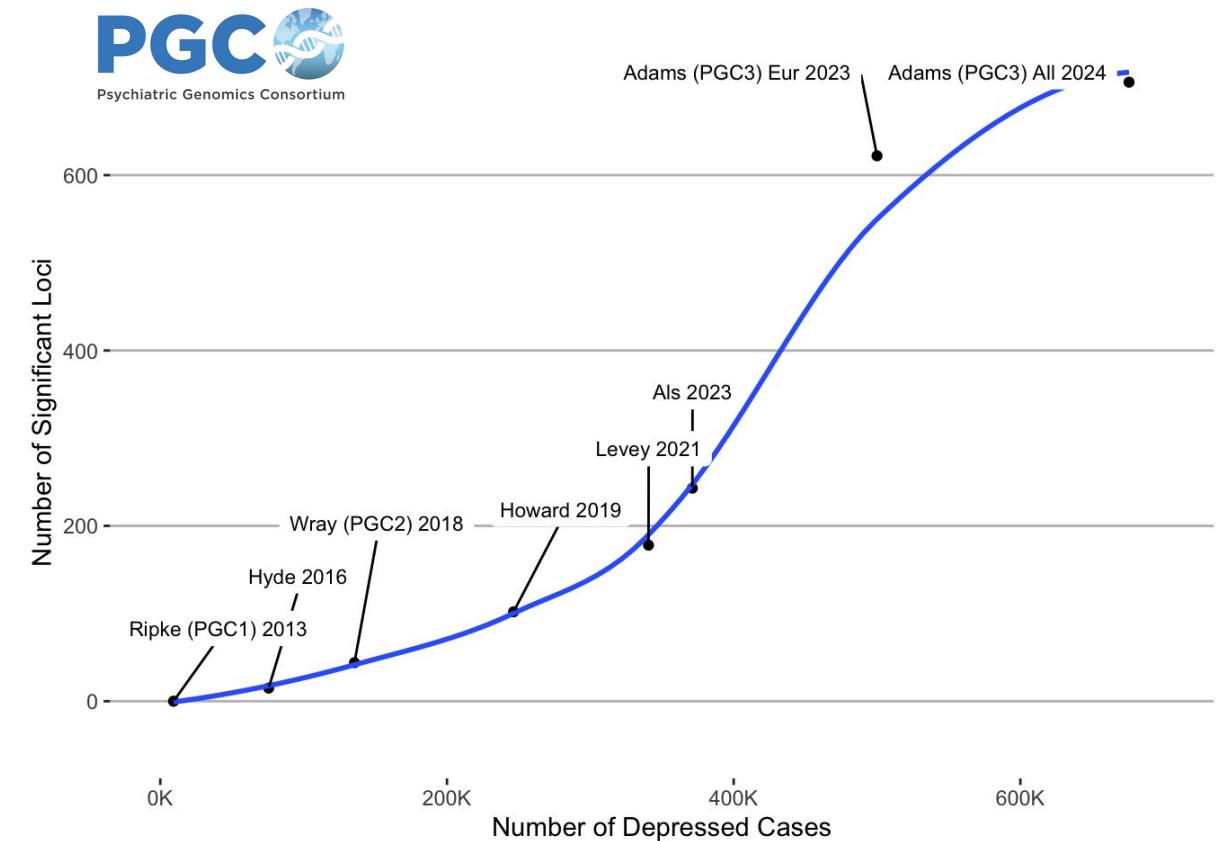


Sample size is important in GWAS

Power affected by heritability of the trait



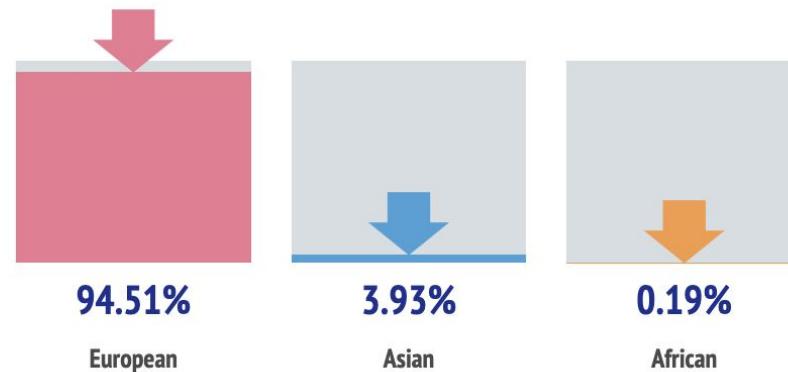
Review: Visscher et al. (2012) Am J
Hum Genet 90:7-24



GWAS Diversity monitor – 2024 and 2025

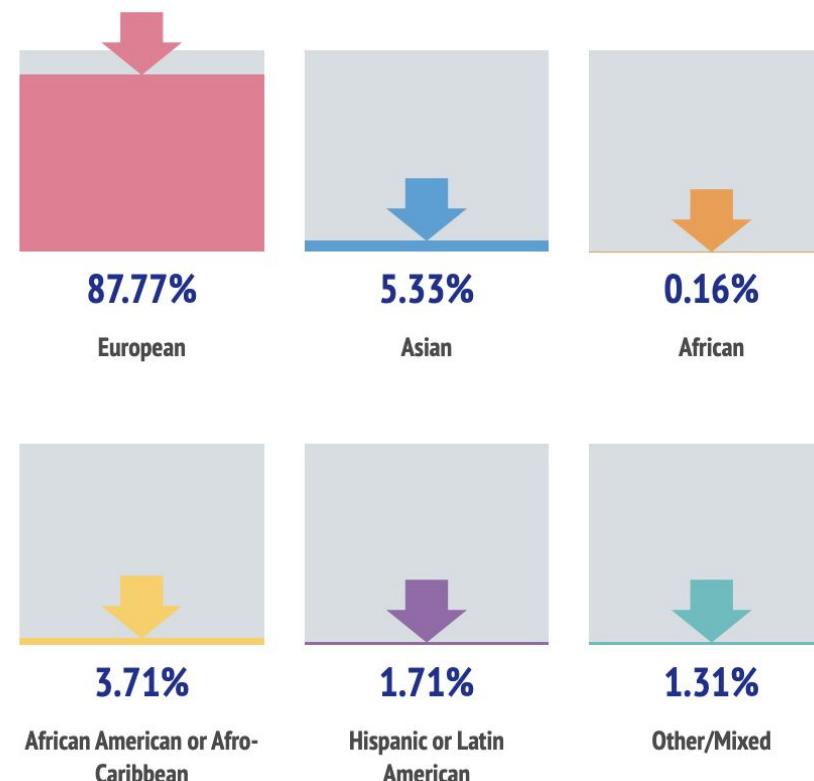
Total GWAS participants diversity

Version 1.0.0. Last check for data: 2024-10-20 00:21:15 .



Total GWAS participants diversity

Version 1.0.0. Last check for data: 2025-09-05 00:34:26 .



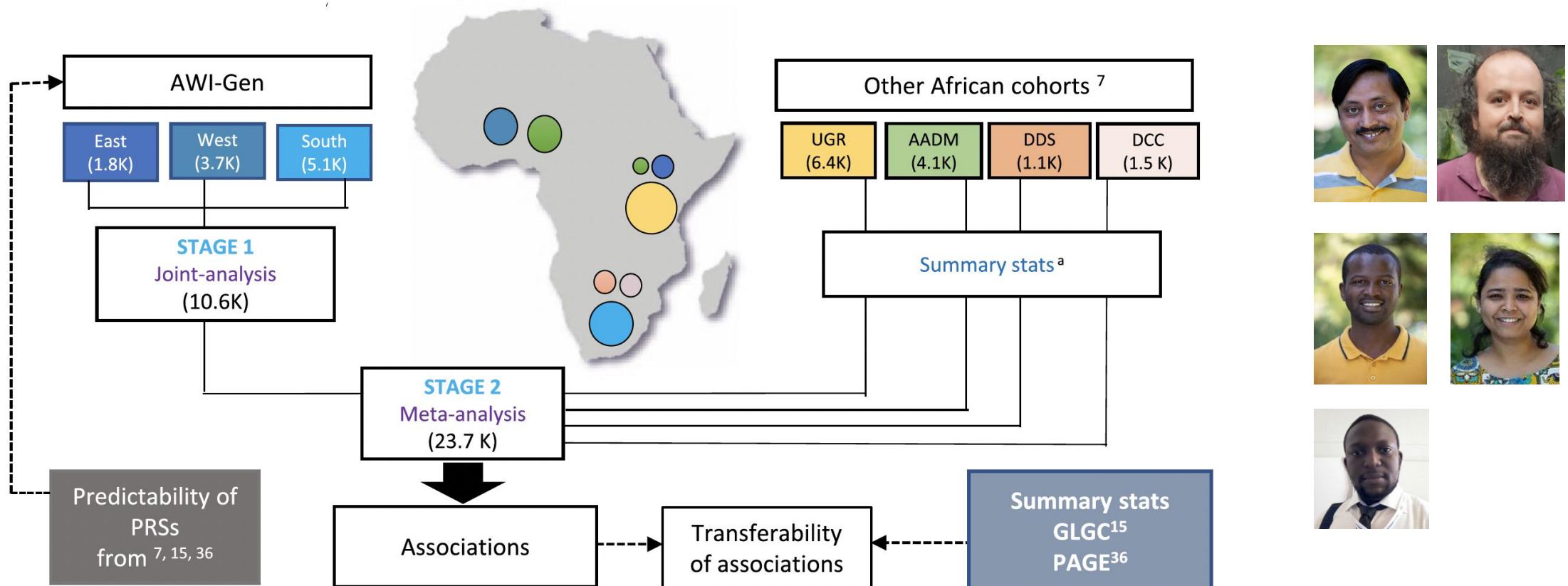
Mills, M.C and Rahal, C., (2020). 'The GWAS Diversity Monitor Tracks diversity by disease in real time'. *Nature Genetics*, 52, 242-243. doi: 10.1038/s41588-020-0580-y

<https://gwasdiversitymonitor.com/>

Examples of GWAS in African populations

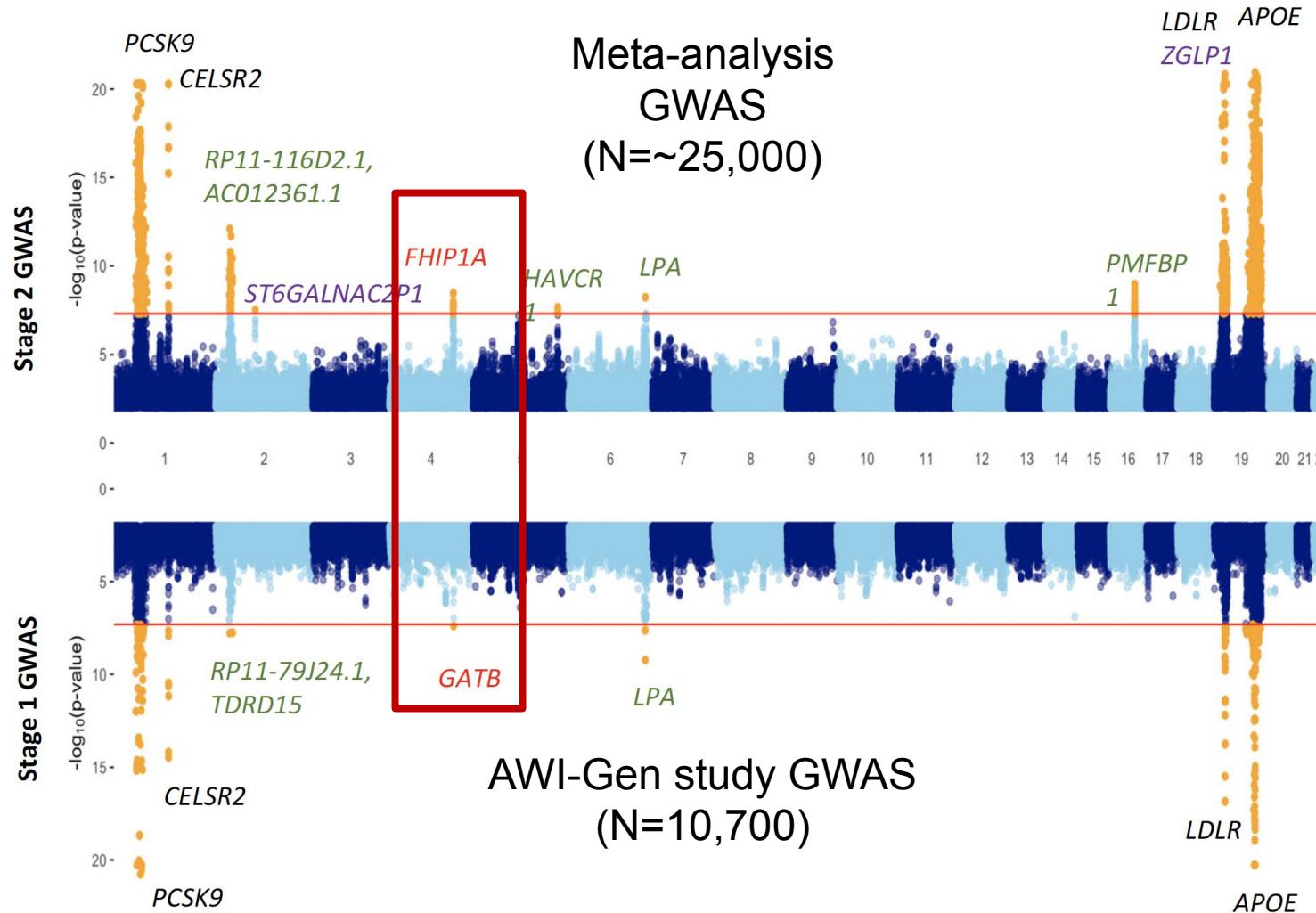
GWAS for lipid traits - AWI-Gen cohort and meta-analysis

LDL-C, HDL-C, Total cholesterol, Triglycerides



Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits Choudhury et al. NATURE COMMUNICATIONS | (2022) 13:2578 | <https://doi.org/10.1038/s41467-022-30098-w>

Miami GWAS plot for LDL-C



Notable points:

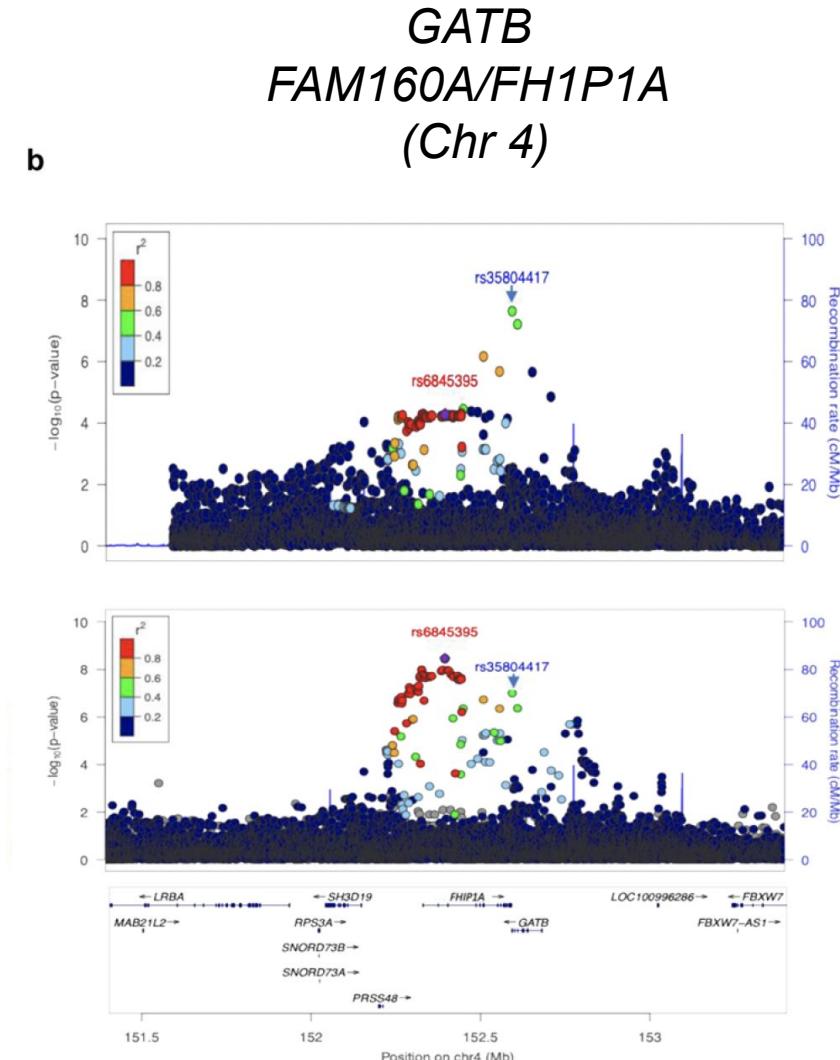
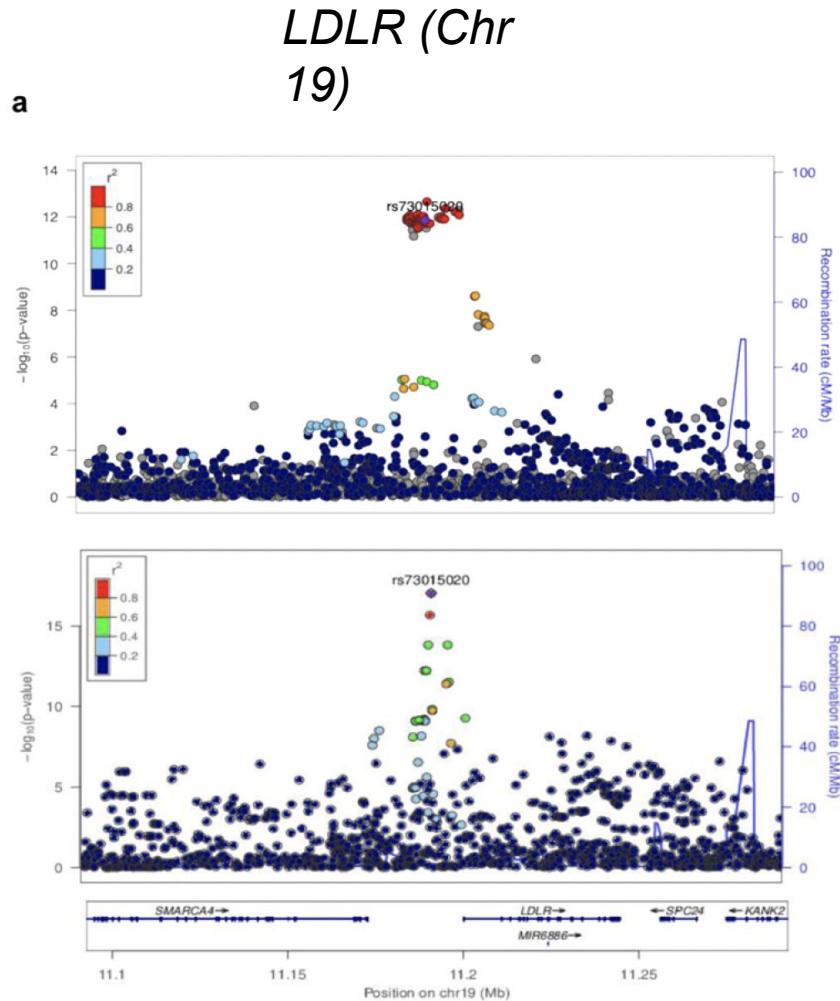
- Replication of known loci
- Novel lead SNPs
- Signals attenuated in the meta-analysis
- Novel LDL-C associated locus detected

Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits Choudhury et al. NATURE COMMUNICATIONS | (2022) 13:2578

Fine mapping LDL-C and novel discovery: Advantages of African studies

European
95% credible set
40 SNPs

African
95% credible set
1 SNP



AWI-Gen only
N=10,700
GATB

Meta-analysi
s
Africa
N~25,000
FHIP1A

Prostate cancer in African men



nature genetics

Article

<https://doi.org/10.1038/s41588-024-01931-3>

Heterogeneous genetic architectures of prostate cancer susceptibility in sub-Saharan Africa

3,963 cases 3,509 controls

Ghana, Nigeria, Senegal, South Africa & Uganda

Similar performance to multi-ethnic and European studies with much larger sample sizes

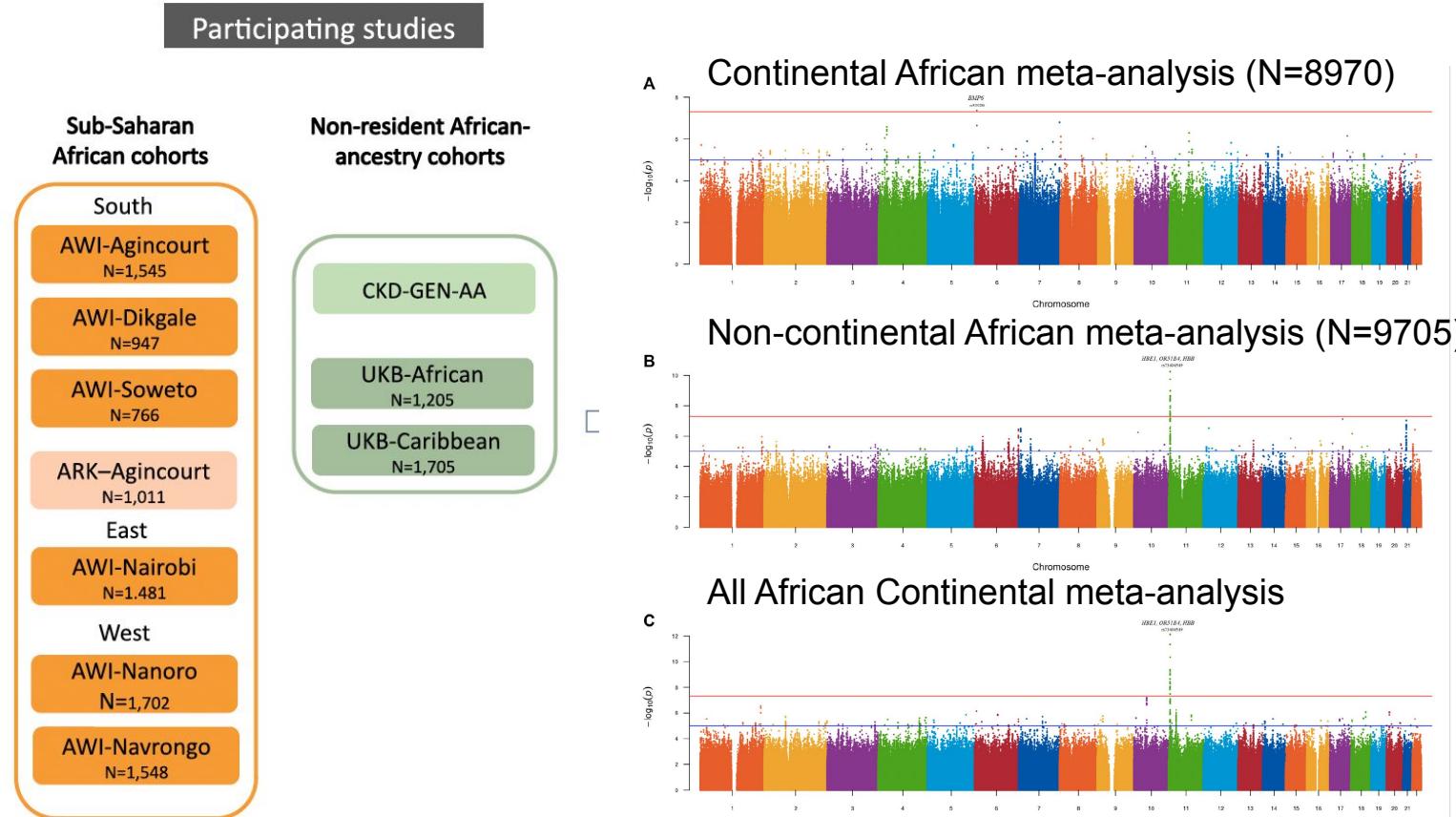
Conclusions:

- Differences across Africa are related more to effect size than allele frequency differences
- Associations mostly governed by neutral evolution
- Contributors: mostly recent mutations affected by genetic drift and bottlenecks
- Lifestyle and environmental (exposome) differences
- Missing heritability: epistatic interactions, different genetic backgrounds and GxE effects

<https://doi.org/10.1038/s41588-024-01931-3>

Nature Genetics - 2024

GWAS for UACR – African studies

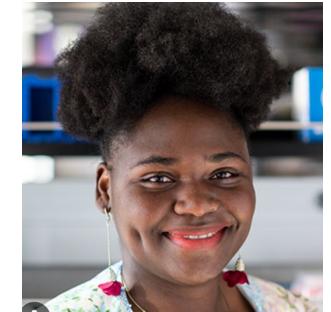
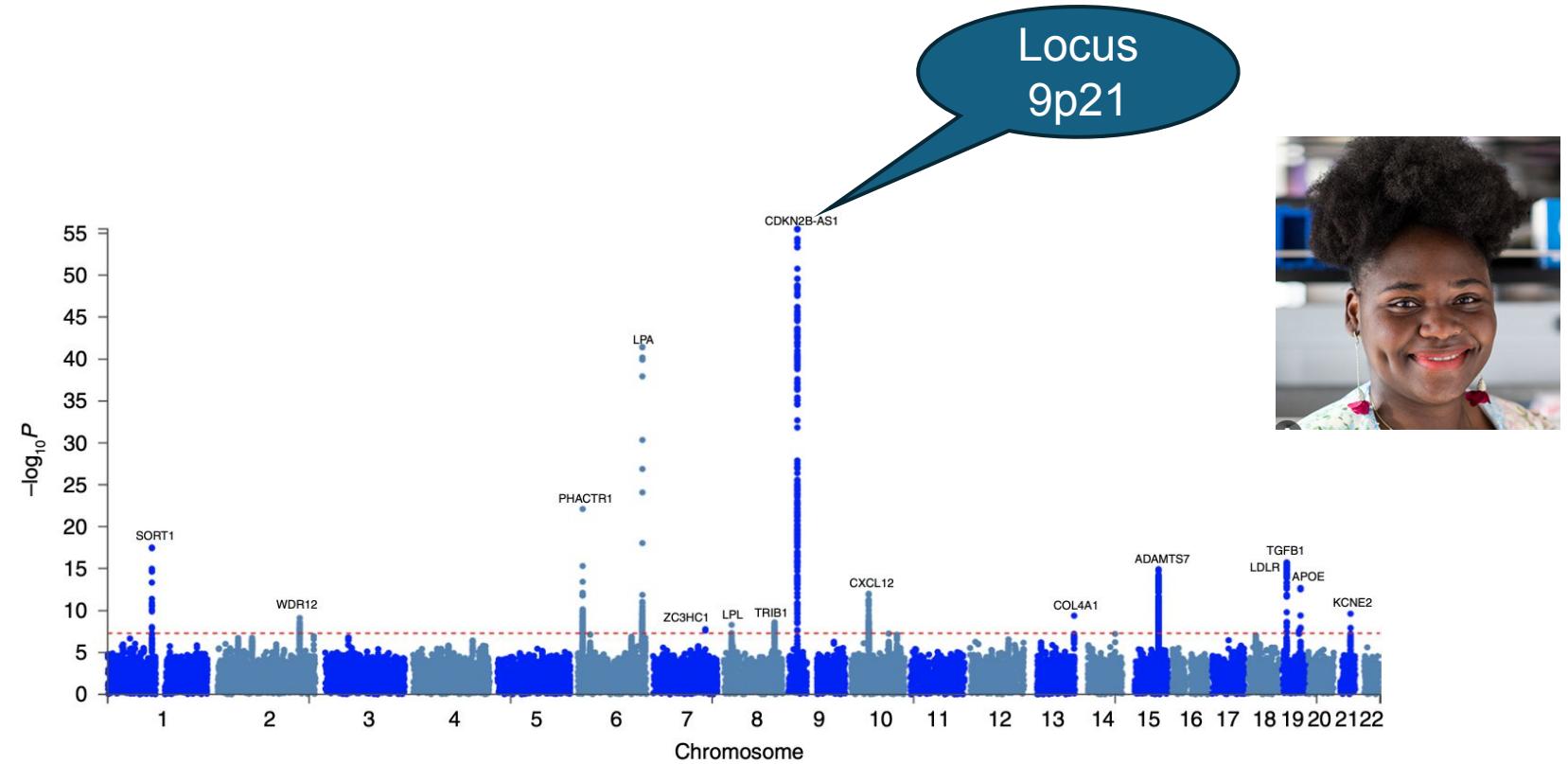
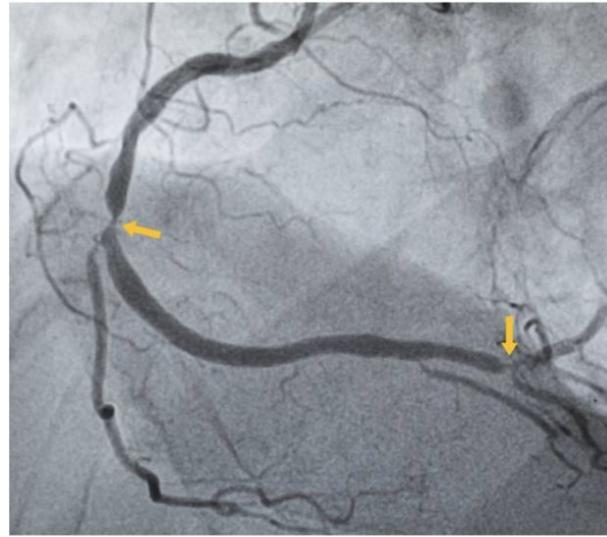


Take home messages

- GWAS significant
- Continental Africa – *BMP6* Chr 6 (eQTLs for *MMP6* and *RREB1* in LD)
- All African studies – *HBB* Chr 11 (*HBS* assoc with malaria resistance)
- *RREB1* – assoc T2D and end stage kidney disease. Interact with *APOL1*
- Replicating *THB53*, *GATM* and *ARL15*
- Regional variation in Africa
- PGS explained <1% of variance

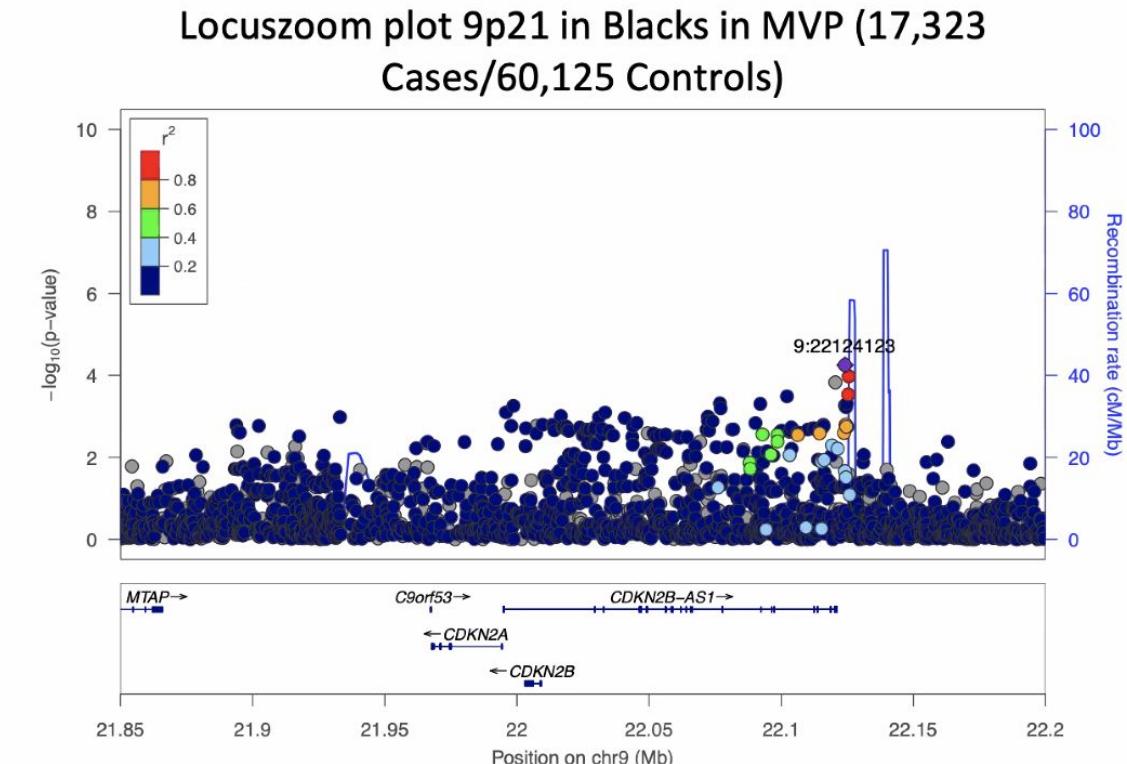
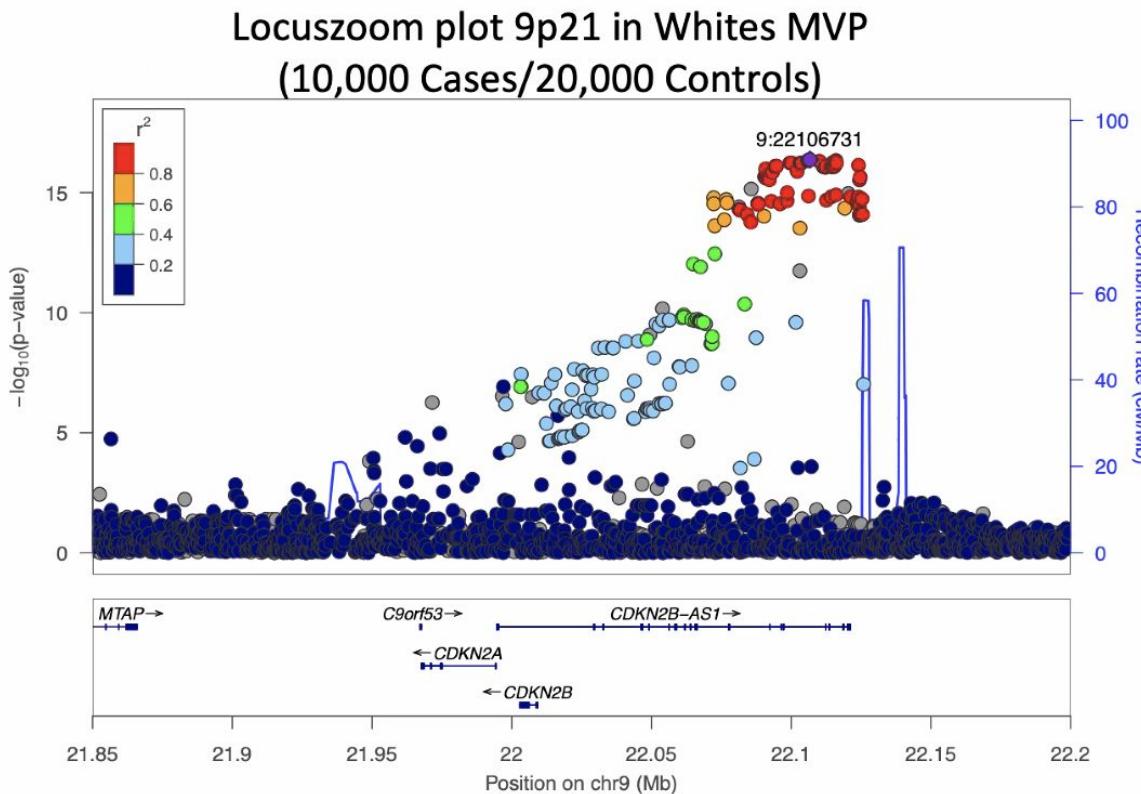
Brandenburg et al. Genetic association and transferability for urinary albumin-creatinine ratio as a marker of kidney disease in four Sub-Saharan African populations and non-continental individuals of African ancestry Front. Genet. 15:1372042. doi: 10.3389/fgene.2024.1372042

Coronary artery disease (CAD) – genetic risk and ancestry (Million Veteran Project, USA)



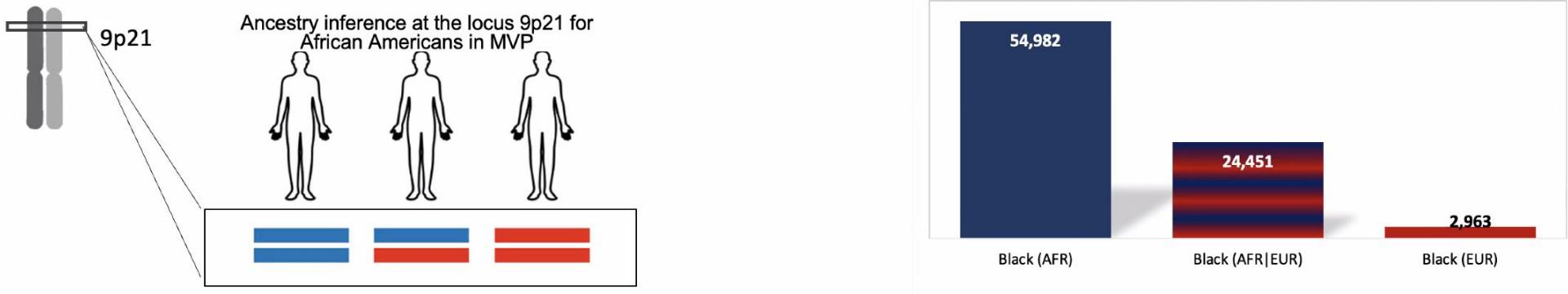
Tcheandjieu et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nature Medicine* (2022) 28:1679-1692

Differences between white and black Americans and CAD association (9p21)

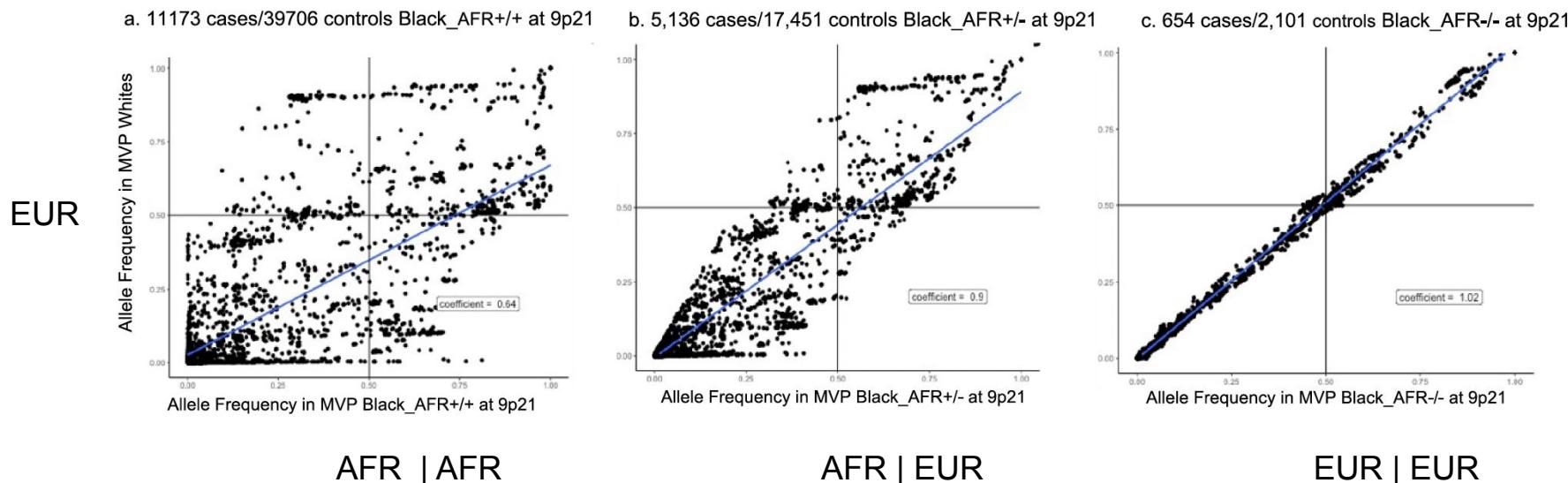


Logistic regression with adjustment on sex and principal component using PLINK (P values from a Wald test and 2-sided)

Differences between black Americans with different ancestral 9p21 haplotypes (CAD association)



Allele frequencies of CAD susceptibility loci at 9p21

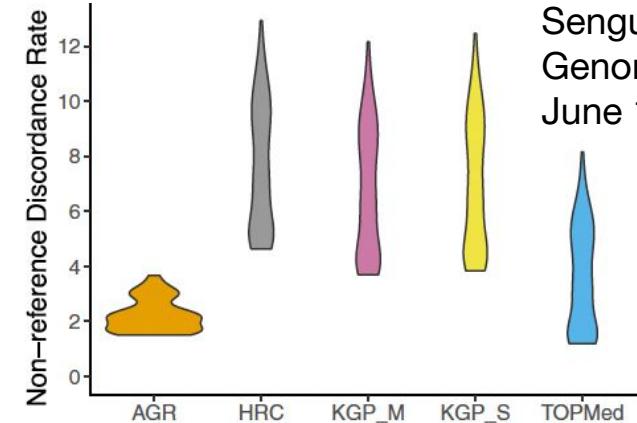
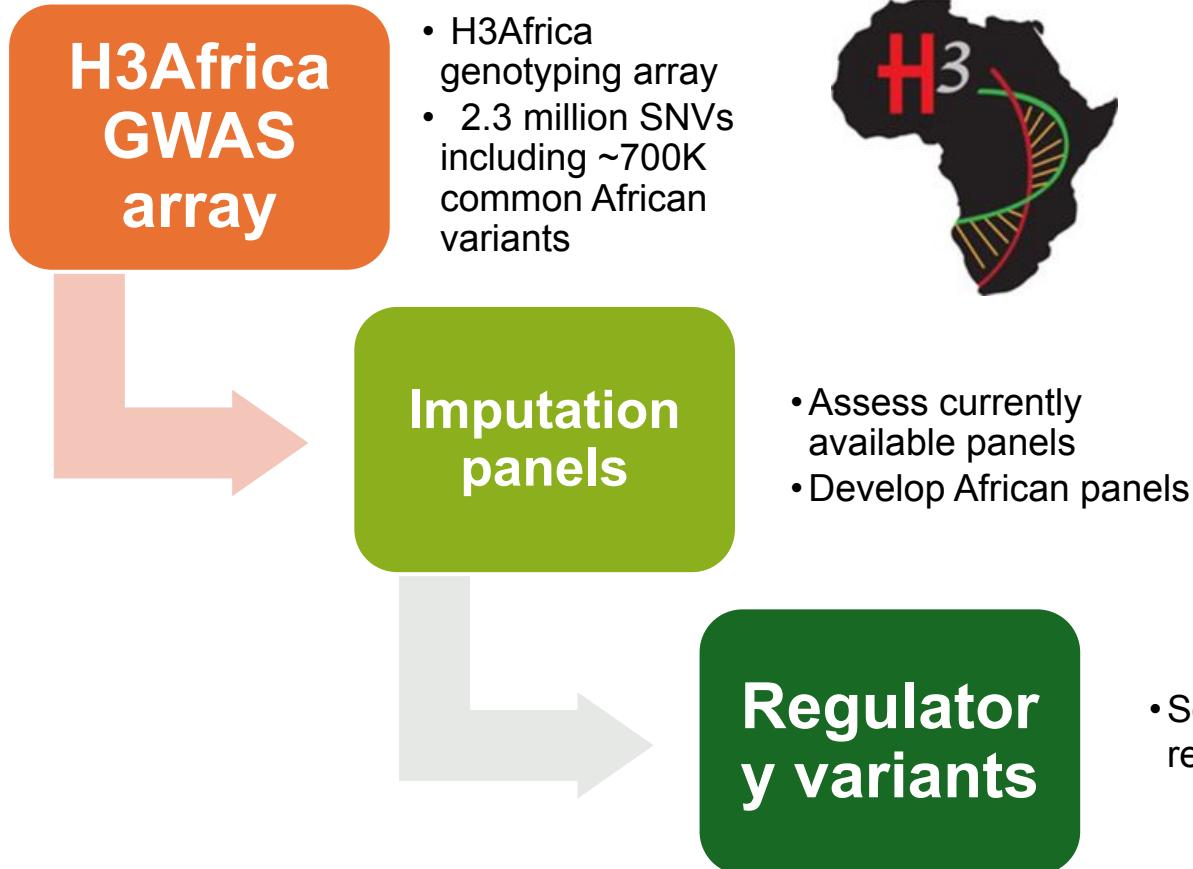


How far have we come since
2012?

What progress made....

- Genotyping arrays enriched for common African SNPs (H3A genotyping array)
- WGS data on more African populations (more appropriate reference genomes for imputation)
- Replication cohorts from Africa (larger, but still limited)
- Better calling of structural variants (STR, CNV, ins, del)
- Cost of WGS (from USD 1000 to USD 100) (closing the gap - GWAS by WGS)
- X chromosome still largely excluded
- Skills development
- Policy and funding support

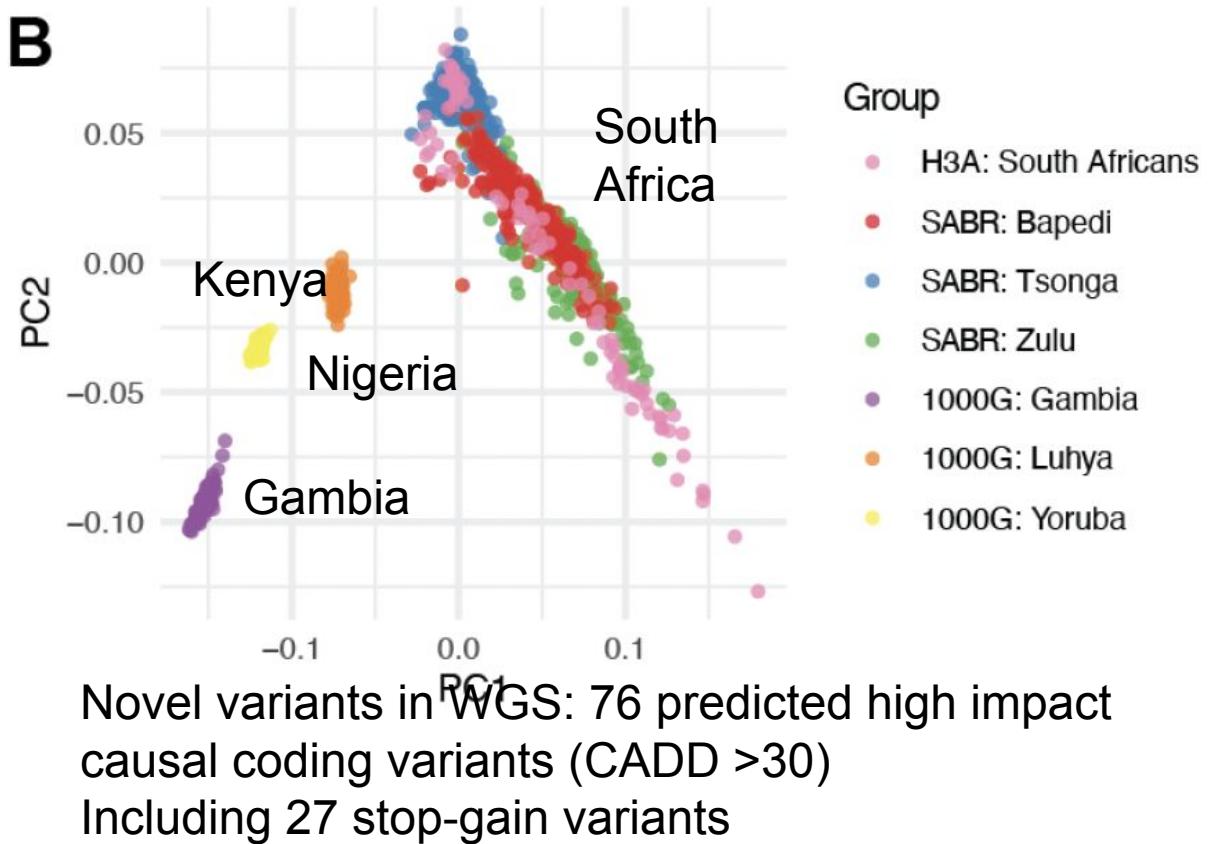
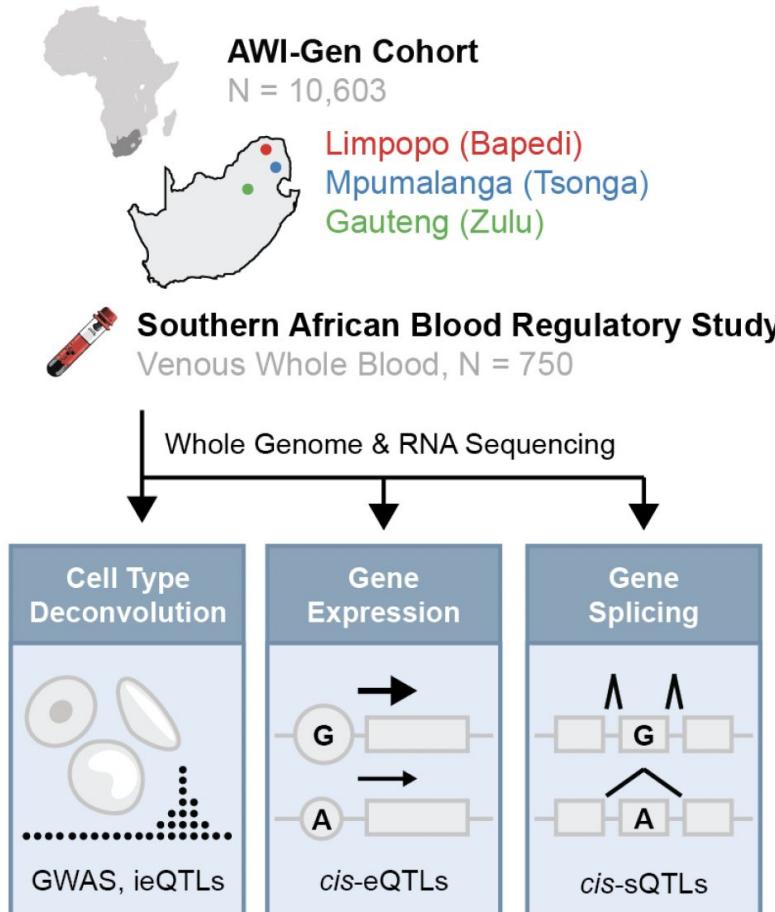
Developing GWAS resources appropriate for African populations



Sengupta et al. Cell Genomics 3, 100332, June 14, 2023

SABR

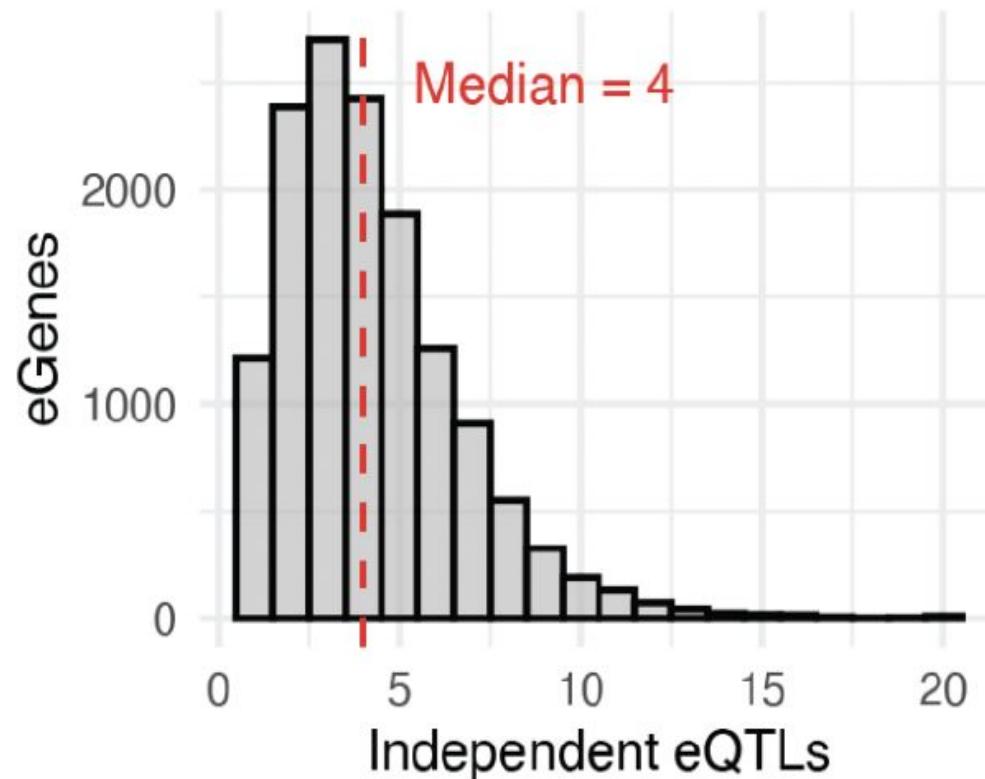
Southern African Map of Blood Regulatory Variation (SABR) Resource



Castel SE, et al. A map of blood regulatory variation in South Africans enables GWAS interpretation. *Nat Genet.* 2025 Jul;57(7):1628-1637. doi: 10.1038/s41588-025-02223-0.

Abundance of eQTLs (expression) and sQTLs (splice)

Mean of 4.3
independent eQTLs
per gene in the South
African study
compared to 1.7 for
GTEx (blood & same
sample size)



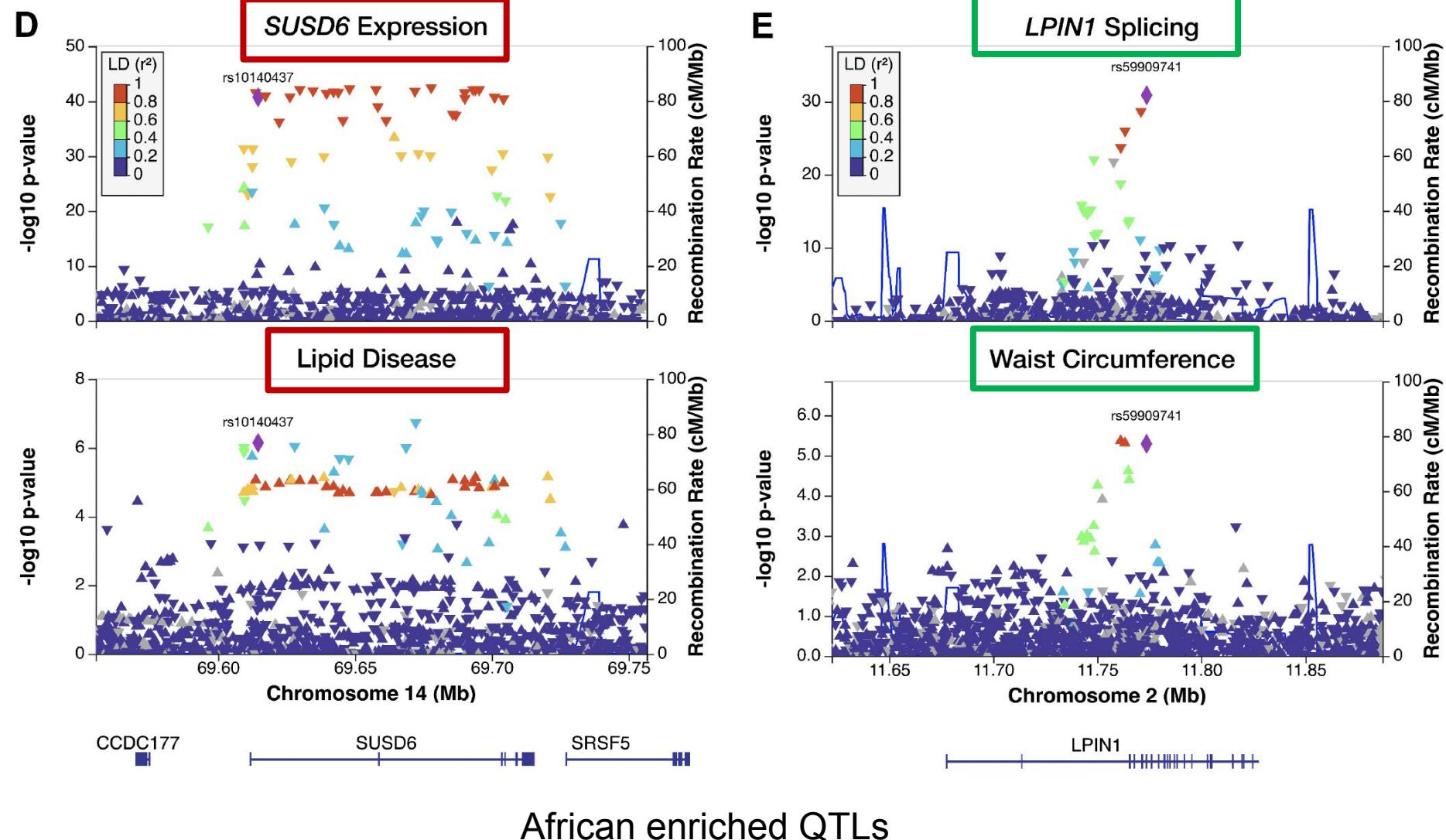
60,808 conditionally independent cis eQTLs

GWAS co-localisation with eQTL and sQTL phenotype association

UKBB (African)
n=6,636 individuals

GWAS for 83 blood
relevant traits

**53 of 83 had at
least one QTL**



What to consider when doing GWAS in African populations

African Genome Architecture

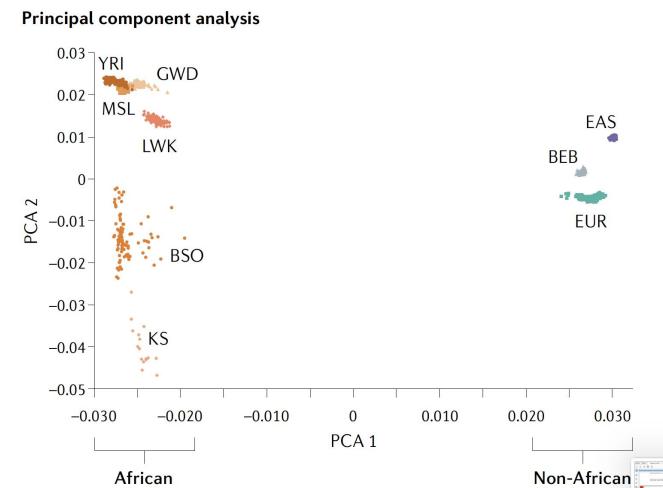
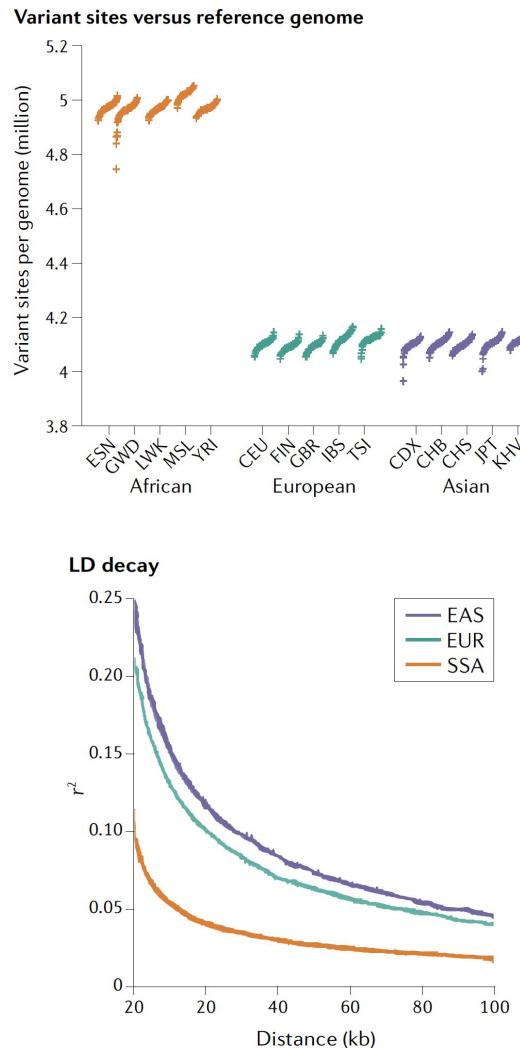
High genetic diversity



Population structure



Low LD



Novel discovery



Analysis challenges



Fine mapping

Pereira L, Mutesa L, Tindana P, Ramsay M. African genetic diversity and adaptation inform a precision medicine agenda. *Nat Rev Genet.* 2021 22(5):284-306. doi: 10.1038/s41576-020-00306-8.

General considerations

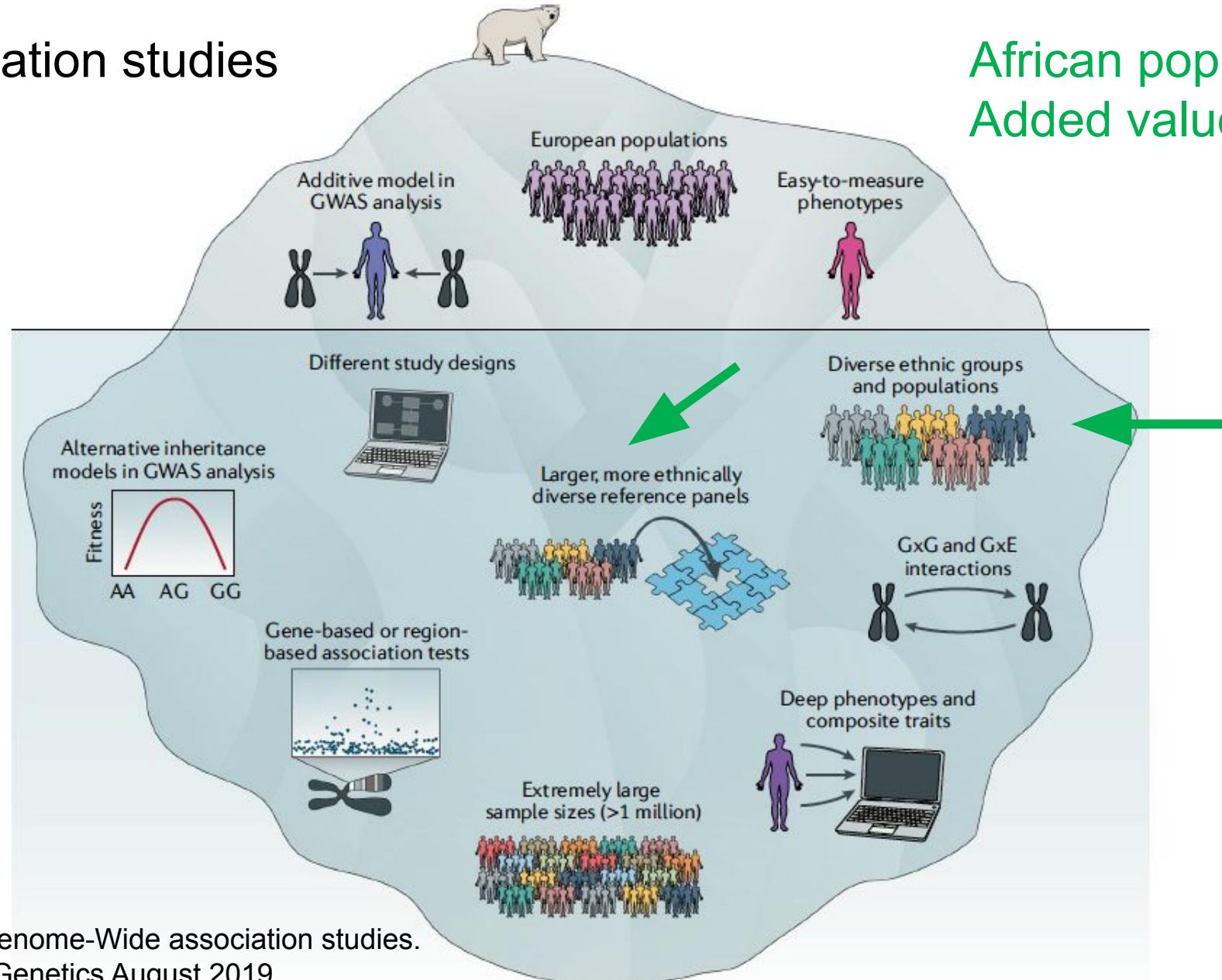
- **Genetic variation - evolutionary forces have shaped African genomes in different ways**
 - Bottlenecks, random drift, selection
 - High genetic diversity, Lower average linkage disequilibrium, Complex population structure (regional differences)
- **Array choice and imputation reference panels**
- **Defining the phenotype (inclusion/exclusion criteria) (narrow vs broad)**
- **Replication of studies:** Similar and/or different biological pathways may affect replication
- **Context matters:** Different chromosomal backgrounds
- **Exposome:** Different lifestyles and environments
- **Missing heritability:** rare variants, epistatic interactions, different genetic backgrounds and GxE effects

Conclusions

- Opportunities outweigh the challenges!
- We need many more African studies
 - Larger cohorts
 - Regional geographic representation
 - Greater diversity of phenotypes
 - More data (phenotype, behaviour and environment data together with genetic data)
- Novel discovery
 - Greatest genetic diversity
 - Lower LD
 - Diversity at many levels (genomes, phenotypes, environment, behaviour)

African population studies Challenges

African population studies Added value



Benefits and limitations of Genome-Wide association studies.
Tam et al. Nature Reviews: Genetics August 2019