Center for Genomics Research and Innovation

Mendelian Randomization Workshop 2023

H3ABioNet

Genome-Wide Association Study (GWAS): Overview

Segun Fatumo *PhD*

Associate Professor, Department of Non-Communicable Disease Epidemiology, LSHTM, UK
Group Leader, The African Computational Genomics Group, MRC/UVRI and LSHTM, Uganda

segun.fatumo@lshtm.ac.uk

Aims of the Session

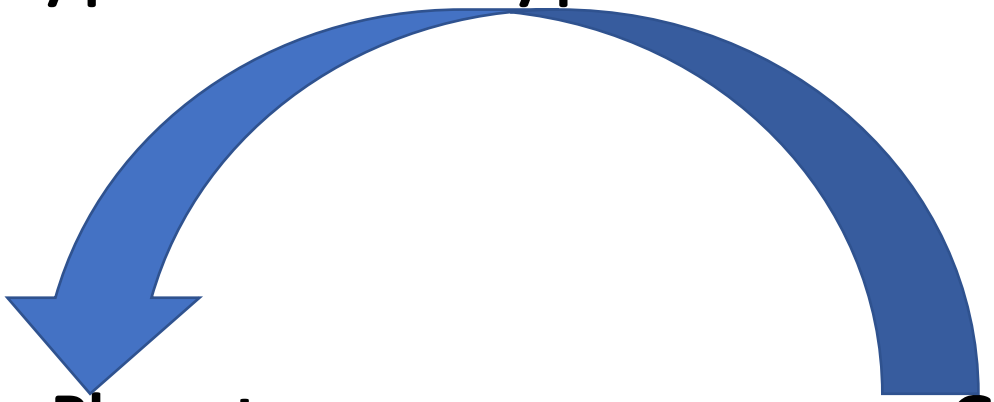
- To explain the contribution of genetic variation in disease and phenotype
- To provide an overview of the design of genome-wide association studies (GWAS)
- To narrate the emergence of GWAS

Intended Learning Outcomes

By the end of the session, students will be able to

- Describe roles of between gene in diseases
- Discuss the concept of genome-wide association study (GWAS)
- Explain the history of GWAS till date

Genotype-Phenotype Associations



- **Phenotype** **Gene**

- **Phenotype**
  - Disease (Diabetes, chronic kidney disease, covid-19, etc)
  - Trait (Height, Weight, BMI, etc)
  - Lifestyle (Smoking, Alcohol, Diet, etc)
  - others (Annual Salary, intelligence, etc)

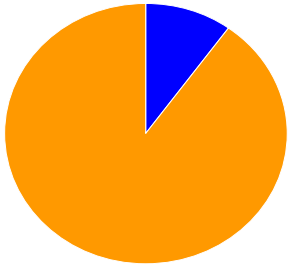
Quiz

1. Can you think of some phenotype of interest to you ?

2. Do you think genes play a role in your phenotypes of interest ?

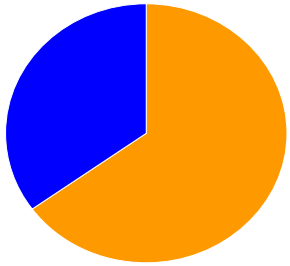
3. Apart from genetic variation, what other factors contribute to your phenotype of interest ?

Genetic variations cause inherited diseases



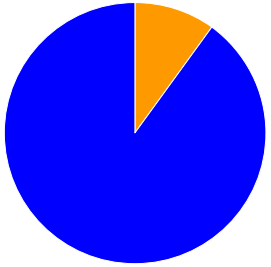
Genetic Diseases

- Cystic fibrosis
- Down syndrome
- Sickle cell disease
- Turner syndrome



Complex Diseases

- Alzheimer disease
- Cardiovascular Disease
- Diabetes (type 2)
- Parkinson Disease



Environmental Diseases

- Influenza
- Hepatitis
- Measles

- Environment - Genes

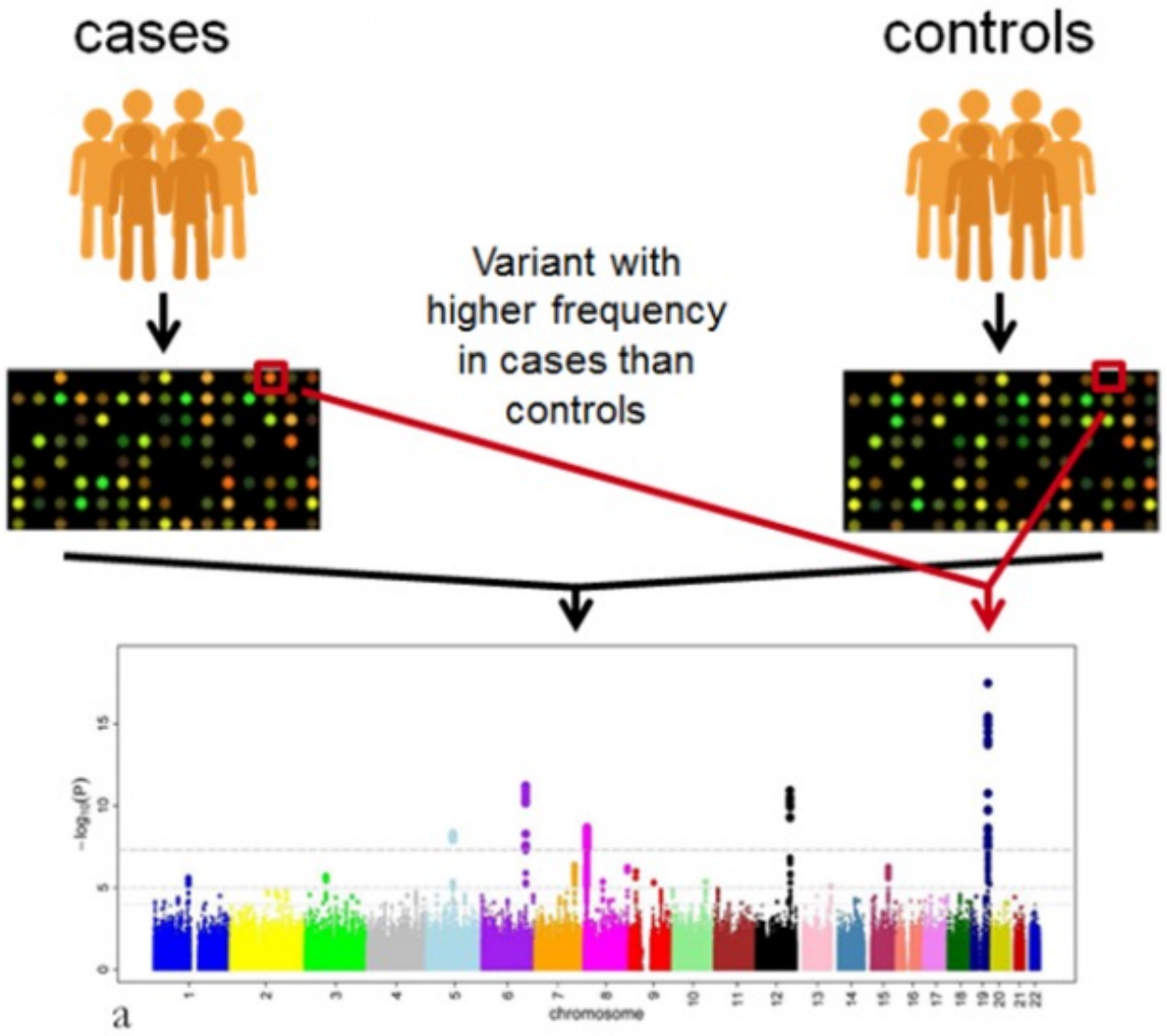
GWAS

- The Genome-wide association studies (GWAS) is an experimental design used to **identify associations between genetic variants and traits** in samples from populations.

- The next big advancement in the field of genetics after the Human Genome Project was GWAS.

http://www.genome.gov/Glossary/index.cfm?id=91

GWAS


cases
controls
Variant with higher frequency in cases than controls
chromosome
a

nature COMMUNICATIONS

COMMENT

https://doi.org/10.1038/s41467-020-19653-5 OPEN

15 years of genome-wide association studies and no signs of slowing down

Ruth J. F. Loos[1,2]

Over the past 15 years, genome-wide association studies (GWASs) have generated a wealth of new information. Larger samples sizes, refined phenotypes and higher-resolution genome-screens will continue to drive gene discovery in years ahead. Meanwhile, GWAS loci are increasingly translated into new biology and opportunities for clinical care.

Loos, R. J. (2020). 15 years of genome-wide association studies and no signs of slowing down. *Nature Communications*, *11*(1), 1-3.

Why was GWAS not possible 20 years ago ?

- Human Genome Project costs **$2.7 billion** over a decade.

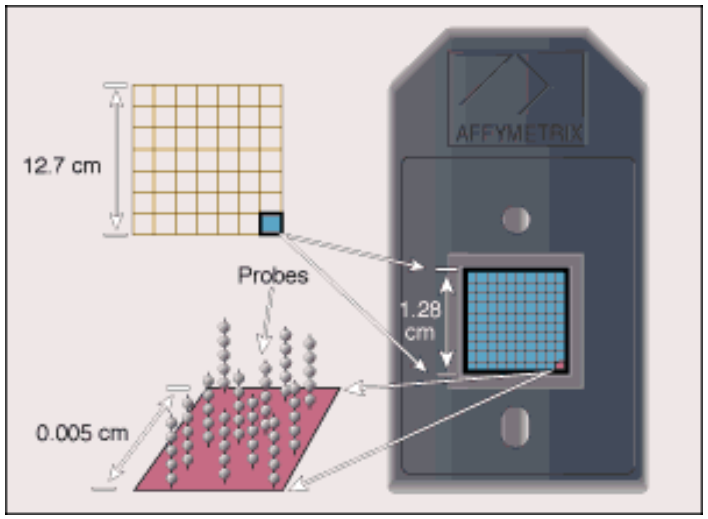
- First, need a method able to genotype thousands/millions of polymorphisms at once– chip-based microarray technology

- Second, need to know where are the common polymorphisms– HapMap project

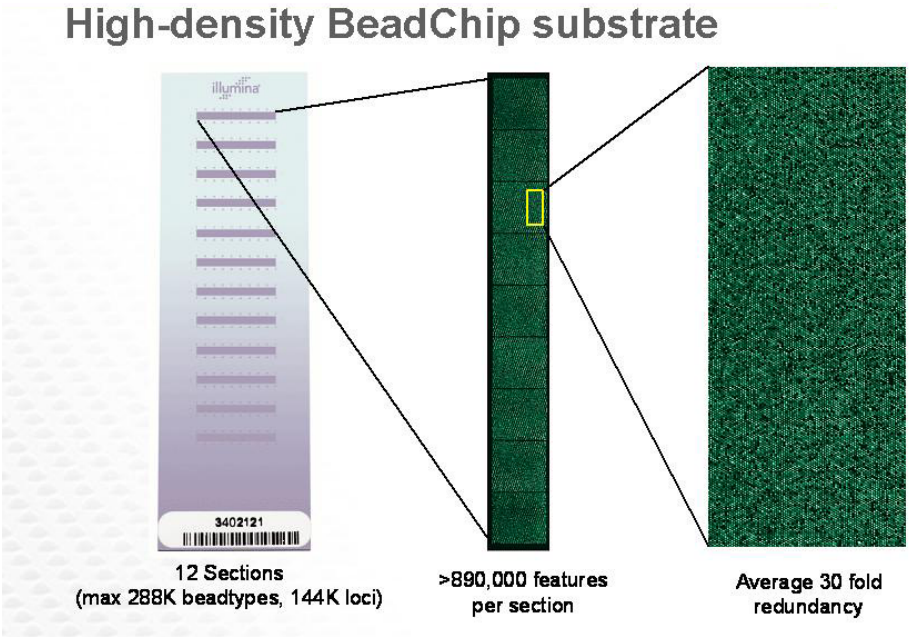
- Third, need HUGE cohorts of people to find subtle allele frequency differences

Genotyping Systems

Affymetrix (Thermo Fisher Scientific)



Illumina



A significant proportion of common SNPs can be captured

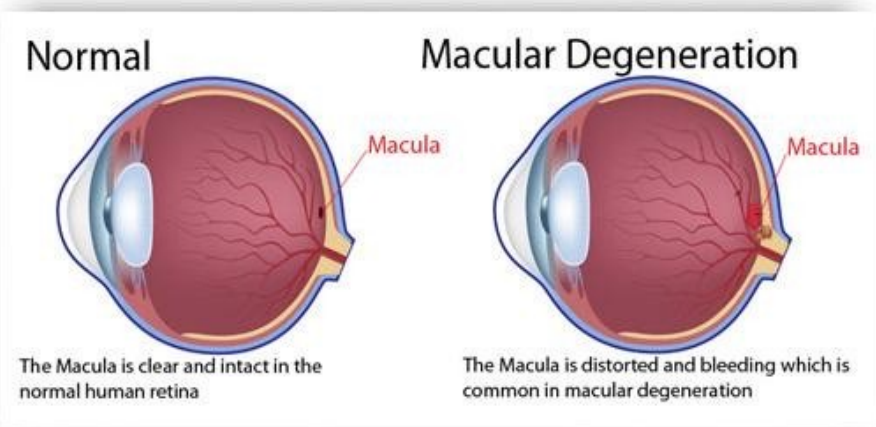
Haplotype Map of the Human Genome

The HapMap is a **catalog of common genetic variants that occur in human beings with geographically diverse ancestry**.

90 Yoruba individuals (30 trios) from Ibadan, Nigeria (YRI)
90 individuals (30 trios) of European descent from Utah (CEU)
45 Han Chinese individuals from Beijing (CHB)
45 Japanese individuals from Tokyo (JPT)

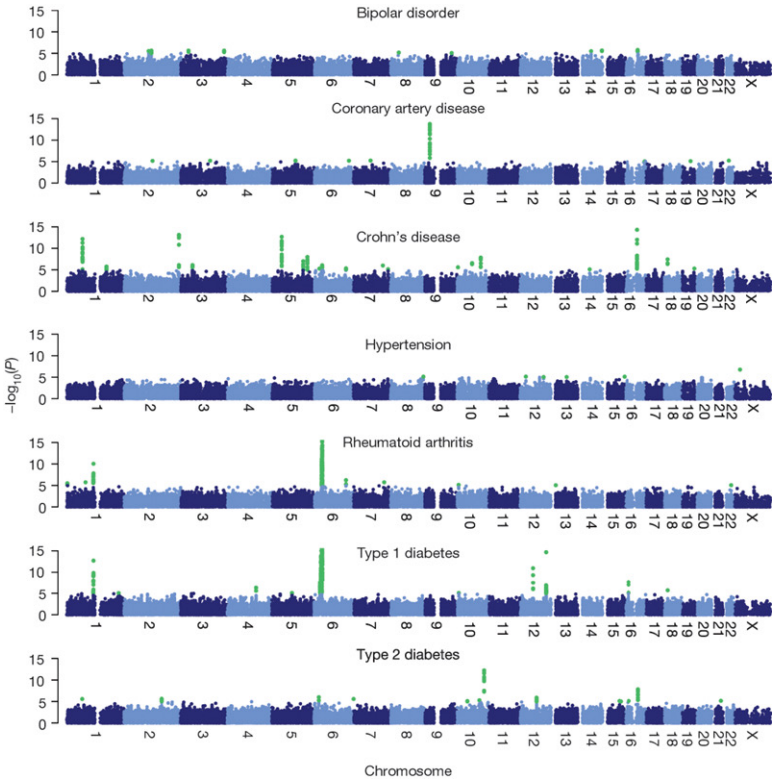
First GWASs (2005 - 2007)

- Age related macular degeneration was the first positively associated disease which was found in 2005, using 96 Caucasian individuals with this disease and only 50 without.
- 100,000 SNP's were used in this study.
- 2 associations (1 false)

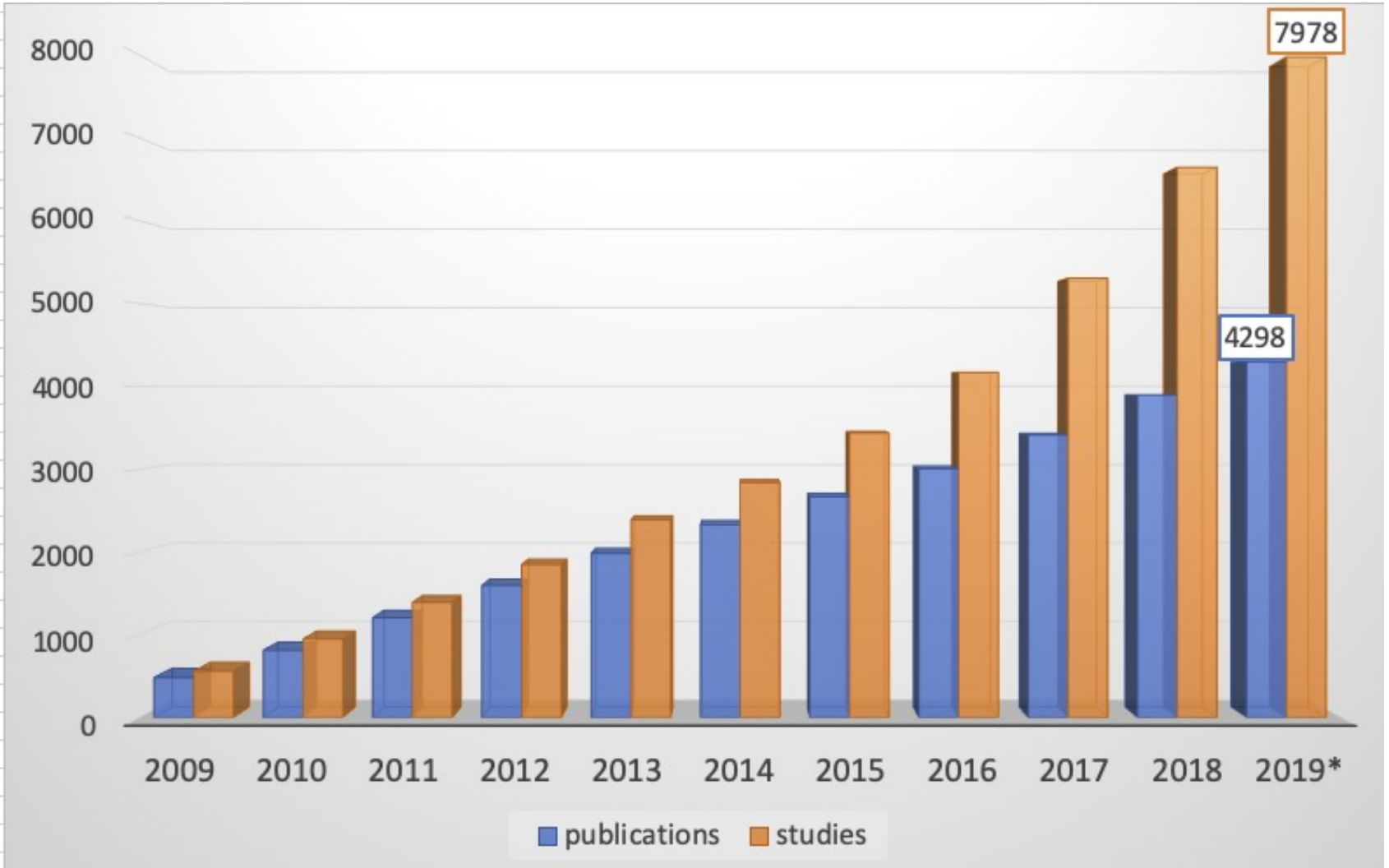


Results of famous WTCCC study of seven diseases on 14,000 cases and 3,000 controls (Nature, 2007)

- 500,568 SNPs Affymetrix GeneChip 500K Array Set
- Total found: 24 associations at level $5*10^{-7}$



GWAS till date

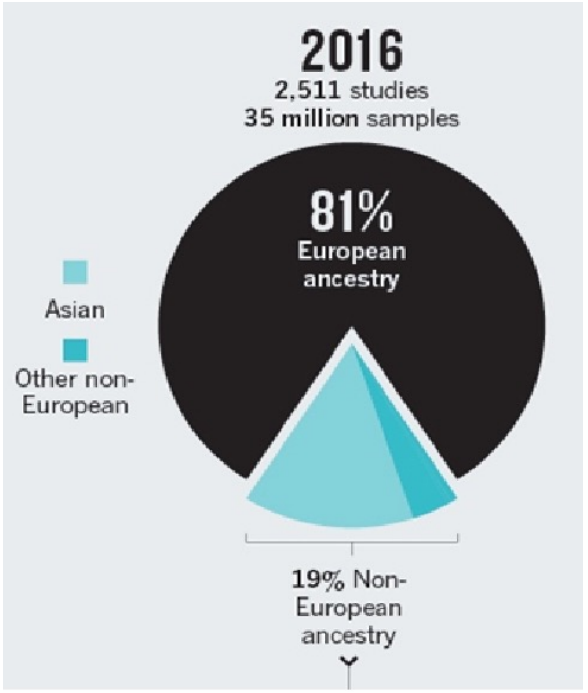


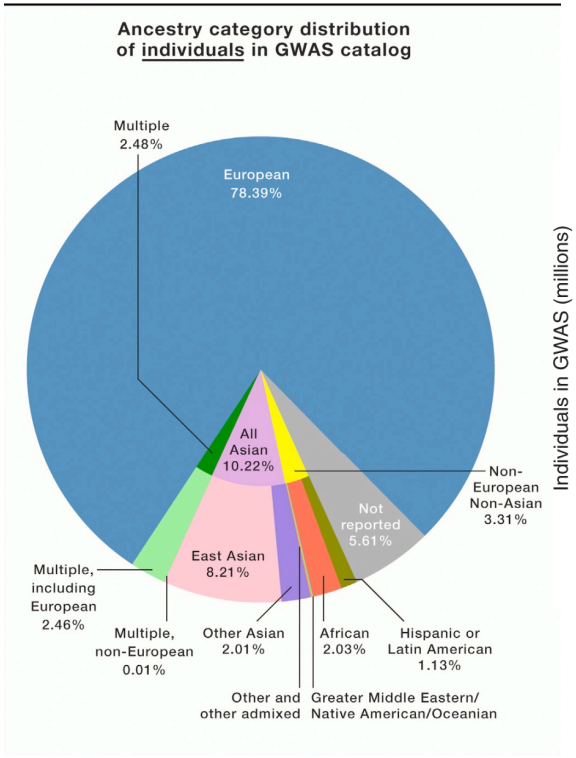
Despite repeated calls and warning, genomics eurocentric bias is on the rise

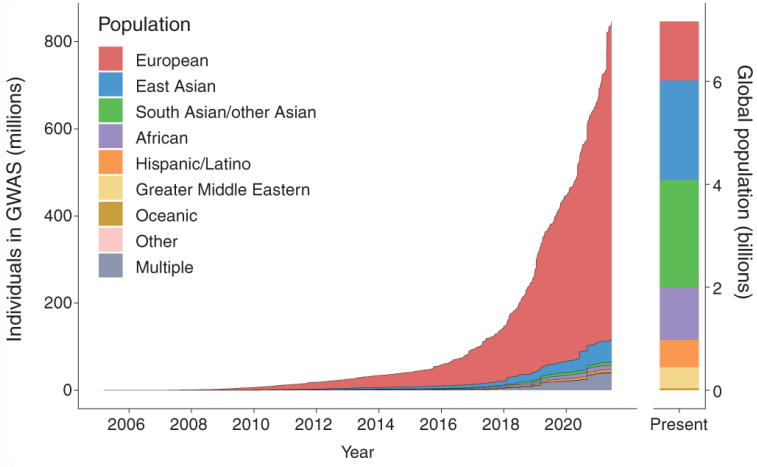
2016
- Europeans: 81%
- Africans: 3%

2019
- Europeans: 78.39%
- Africans: 2.03%

2022
- Europeans: 86%
- Africans: 1.1%







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

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

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

The 1.1% Africans in Genomic studies are mainly African Americans

- *Studying a small number African diaspora populations (African American and Black participants in the UK and Europe) and grouping all participants into a broad category of African ancestry will continue to promote genomic imbalance, widen health disparities, and fail to capture the genetic diversity in Africa.*

Scientists need to WALK the TALK

UK Biobank (n=~500,000)

~38,000 individuals with non-European ancestry

92.7% of publications excluded non-EUR samples

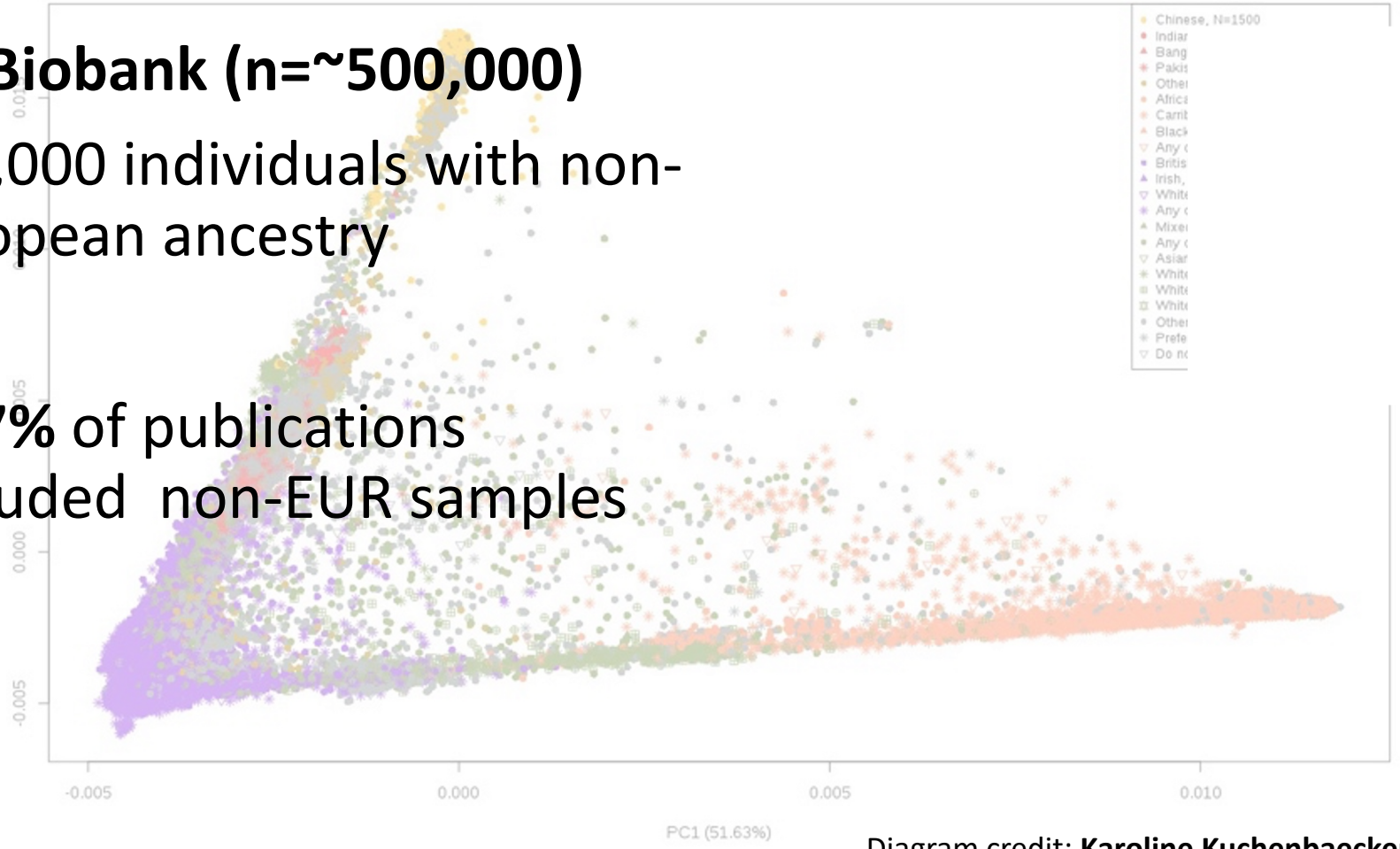
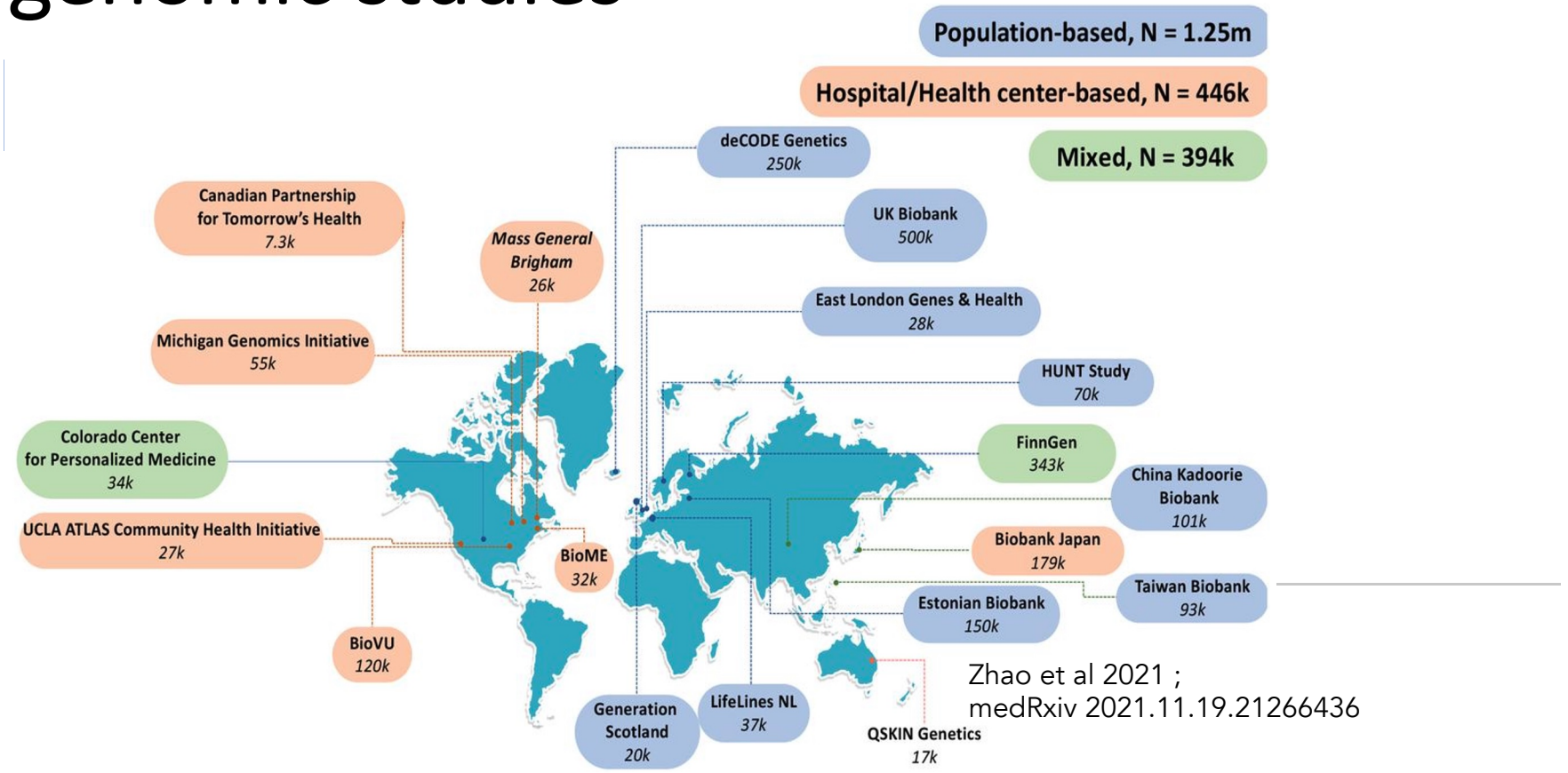


Diagram credit: **Karoline Kuchenbaecker**

Global biobanks bias for genomic studies

Scientists tend to utilise datasets that are well curated and easy to access



Zhao et al 2021 ;
medRxiv 2021.11.19.21266436

- Nearly 9 of every 10 genomics studies (86.3%) were conducted with people of European ancestry, despite the group constituting only 16 percent of the world's population.
- Despite the scientific rationale for inclusion of African populations, only 1.1% of African participants in GWAS catalog are mainly African Americans which are not truly representative of the genetic diversity in Africa.

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open Access | Published: 12 October 2022

A saturated map of common genetic variants associated with human height

Loïc Yengo, Sailaja Vedantam, Eirini Marouli, Julia Sidorenko, Eric Bartell, Saori Sakaue, Marielisa Graff, Anders U. Eliasen, Yunxuan Jiang, Sridharan Raghavan, Jenkai Miao, Joshua D. Arias, Sarah E. Graham, Ronen E. Mukamel, Cassandra N. Spracklen, Xianyong Yin, Shyh-Huei Chen, Teresa Ferreira, Heather H. Highland, Yingjie Ji, Tugce Karaderi, Kuang Lin, Kreete Lüll, Deborah E. Malden, 23andMe Research Team, VA Million Veteran Program, DiscovEHR (DiscovEHR and MyCode Community Health Initiative), eMERGE (Electronic Medical Records and Genomics Network), Lifelines Cohort Study, The PRACTICAL Consortium, Understanding Society Scientific Group, ... Joel N. Hirschhorn + Show authors

Nature **610**, 704–712 (2022) | Cite this article

45k Accesses | **2** Citations | **1350** Altmetric | Metrics

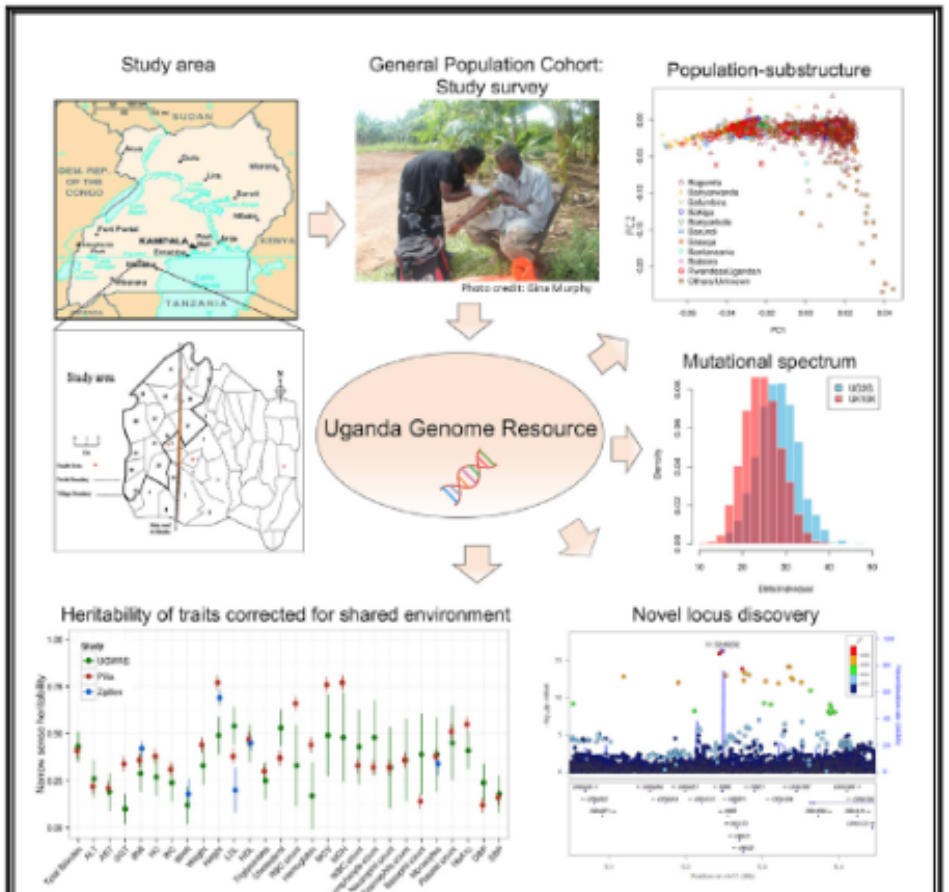
- **5.4 million people**
- **281 GWAS Aggregated**
- **None/1 from continental Africa**
- **>600 authors**
- **Found “All” associations**
- **Not that height GWAS is lacking in continental Africa**
- **Why exclude continental Africa ?**

Largest ever genome study of Africans (2019)

Cell

Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa

Graphical Abstract



Authors

Deepti Gurdasani, Tommy Carstensen, Segun Fatumo, ..., Pontiano Kaleebu, Ines Barroso, Manj S. Sandhu

Correspondence

cts@sanger.ac.uk (C.T.-S.),
motala@ukzn.ac.za (A.M.),
rotimic@mail.nih.gov (C.R.),
pontiano.kaleebu@mrcuganda.org (P.K.),
ib1@sanger.ac.uk (I.B.),
mss31@cam.ac.uk (M.S.S.)

In Brief

Genome-wide data from Ugandans reveal insights into their ancestry, trait heritability, and loci associated with metabolic parameters, thereby providing a diverse resource for the study of African population genetics.

Table 1. Novel and Distinct Association Signals Discovered in GWAS Meta-Analysis

Trait	rs ID	chr:pos	A1	A2	Number	p_assoc	p het	Gene	MAF_ AFR (%)	MAF_ EUR (%)
Novel associations with traits										
Albumin	rs540810730	7:19228309	C	A	8,995	3×10^{-09}	0.33	NA	2.9	0
Bilirubin[a]	rs151330263	16:302161	A	G	9,326	2×10^{-12}	0.60	*HBA1/HBA2*	5.6	0
BMI	rs7798566	7:141549317	A	G	13,976	3×10^{-15}	6×10^{-17}	NA	4.9	1.2
HbA1c	rs6724428	2:189377509	A	G	7,161	4×10^{-09}	0.12	*GULP1*	44	55
HDL-chol	NA	19:21749298	G	A	6,407	4×10^{-09}	NA	*RP11-678G14.3*	0.7	0
RDW	NA	7:131419316	CAA	C	1,119	1×10^{-09}	NA	NA	38	NA
WBC count	rs4755389	11:35115148	C	T	2,741	4×10^{-09}	0.54	*CD44*	5.9	45.2
Novel associations previously associated with similar traits										
RDW	rs334	11:5248232	T	A	1,625	2×10^{-17}	NA	*HBB*	7.7	0
BMI	rs12405634	1:243102900	C	T	13,976	3×10^{-10}	4×10^{-12}	NA	11.3	1.5
neut_count	rs1347767	2:136485657	C	T	2,671	7×10^{-11}	3×10^{-03}	*R3HDM1*	12.8	0
Distinct associations at known loci										
ALT	NA	8:145730373	G	C	6,407	6×10^{-38}	NA	*GPT*	0.6	0
ALP	1:21897903	rs4654971	T	C	2,588	8×10^{-11}	0.04	*ALPL*	3.3	8.3
ALP	6:24489961	rs189263035	G	C	9,322	7×10^{-20}	6×10^{-4}	*GPLD1*	7.8	0
GGT	22:25084815	NA	G	A	8,995	1×10^{-60}	0.12	NA	8.3	0
HbA1c[a]	rs148228241	16:227187	G	T	7,161	3×10^{-12}	0.02	*HBA1/HBA2*	10.1	0
Cholesterol	5:156378584	NA	CGGAA	C	6,407	NA	NA	*TIMD4*	0.9	0
LDL-chol	5:156378584	NA	CGGAA	C	6,407	NA	NA	*TIMD4*	0.9	0
Triglycerides	19:45422587	rs12721054	A	G	13,115	7×10^{-25}	0.03	*APOC1*	13	0
Triglycerides	1:63171024	rs569795903	C	T	6,407	NA	NA	*RP11-230B22.1*	1.4	0
MCHC	16:302161	rs151330263	A	G	2,744	8×10^{-13}	0.17	*ITGF3*	4.8	0

A1, effect allele; A2, non-effect allele; neut_count, neutrophil count; NA, not applicable; p_assoc, p value from RE2 (Han-Eskin) METASOFT meta-analysis across cohorts (where relevant); p het, p value for Cochran's Q heterogeneity statistic. See Tables S6.2–S6.8 for all results, and Figures 6 and S10 for locusview plots.

[a]These associations were found to be driven by the $\alpha^{-3.7}$ thalassemia deletion on further sensitivity analyses.

Summary

- GWAS involve testing genetic variants across the genomes of many individuals to identify genotype–phenotype associations.

- Since the first GWAS in 2005, nearly 250K associations of genome-wide significance ($P < 5 \times 10^{-8}$) have been reported between genetic variants and common diseases and traits.

- However, GWAS have mostly focused on individuals of European descent. There is a need for diversity.

Next Session

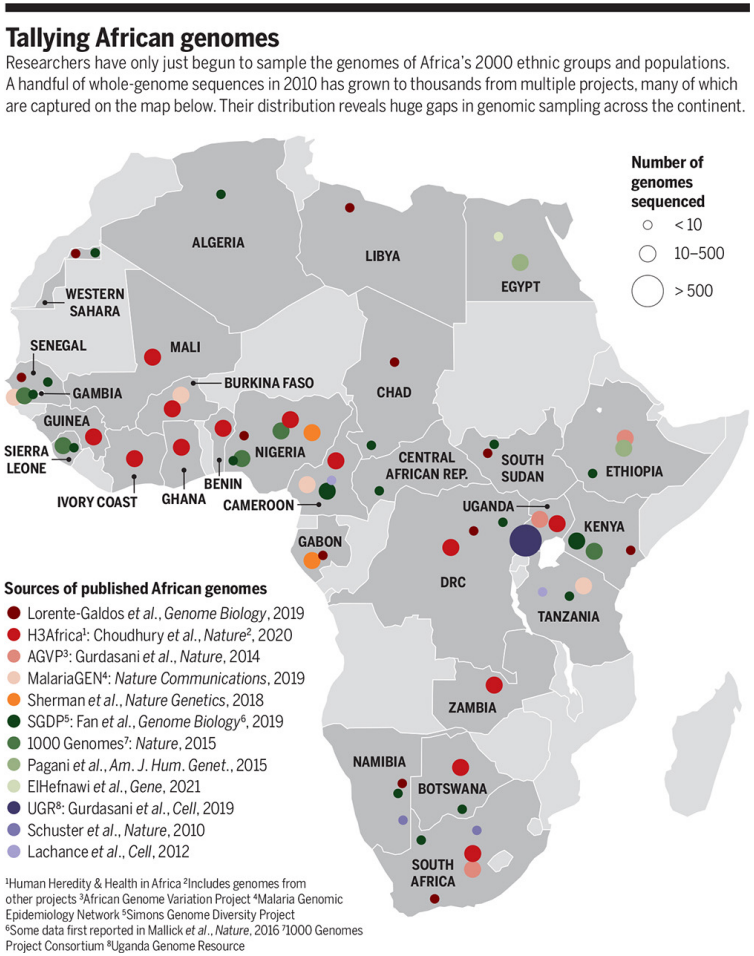
- **Resources & Workflow for performing GWAS**

- Wellcome Trust Case Control Consortium. "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." *Nature* 447, no. 7145 (2007): 661. https://www.nature.com/articles/nature05911

- Gurdasani, Deepti, Tommy Carstensen, **Segun Fatumo**, Guanjie Chen, Chris S. Franklin, Javier Prado-Martinez, Heleen Bouman et al. "Uganda genome resource enables insights into population history and genomic discovery in Africa." *Cell* 179, no. 4 (2019): 984-1002. https://www.cell.com/cell/fulltext/S0092-8674(19)31120-1

Celebrating 20 Years of Human Genome

https://science.sciencemag.org/content/371/6529



Elizabeth Pennisi Science 2021;371:556-559

https://science.sciencemag.org/content/371/6529/556