

Race, Ethnicity, and Ancestral Populations

Nuanced differences that confound
many health professionals

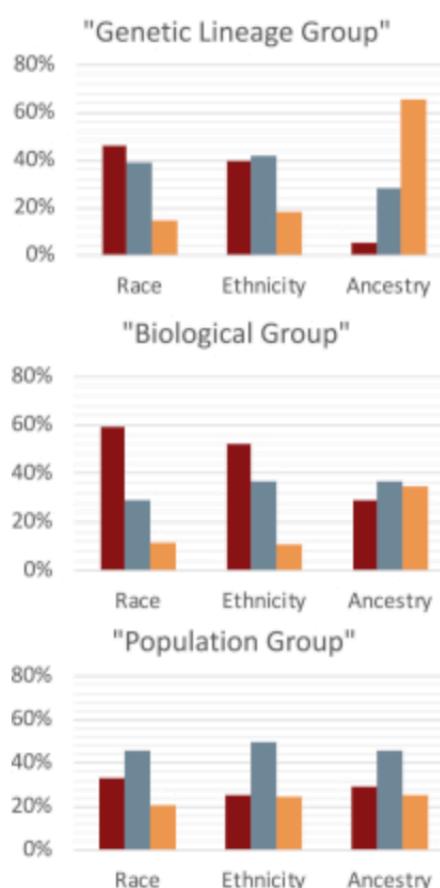
Outline

- 1. What is “Race”?**
 - Morphological and phenotype differences, ie. skin color, dosage differentials
 - Lewontin, 1972
- 2. Comparison of race, ethnicity, and ancestry**
 - $V_p = V_G + V_E$
- 3. Does self-identified racial classification convey useful information about genotype?**
 - Differences in variant frequencies and LD
 - Picked up with these programs: STRUCTURE, PCA, ADMIXTURE
 - (important nuances can be found in “Counter the weaponization of genetics research by extremists”, 2022)
 - Can capture variation in ancient hominids (ie. neanderthal, denisovan, other ‘phantom populations’) that can impact susceptibility to diseases (autoimmune disorders, Covid, etc).
 - <https://news.vanderbilt.edu/2016/02/11/neanderthal-dna-has-subtle-but-significant-impact-on-human-trait/>
- 4. Linkage Disequilibrium, recombination and information integrity in GWAS**
 - Difference between pedigree and genotypic ancestry
 - How far back until we are all related?

Confused about the difference between Race, Ethnicity, and Ancestry?

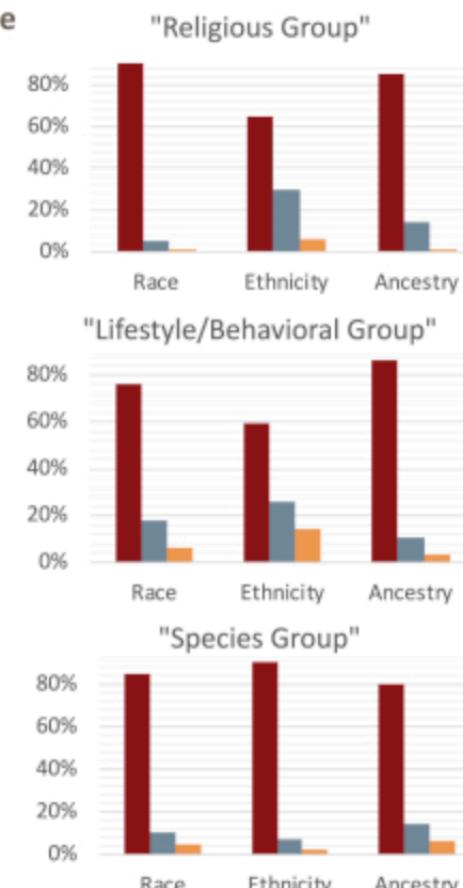
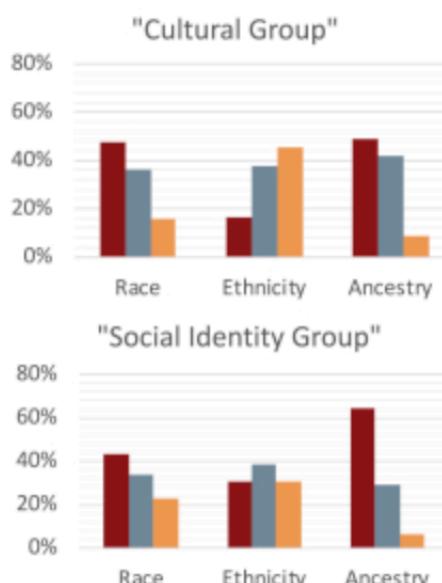


Clinical relevance:



How well does each of these terms describe "race", "ethnicity", and "ancestry"?

- Poorly/Very poorly
- Somewhat
- Well/Very well



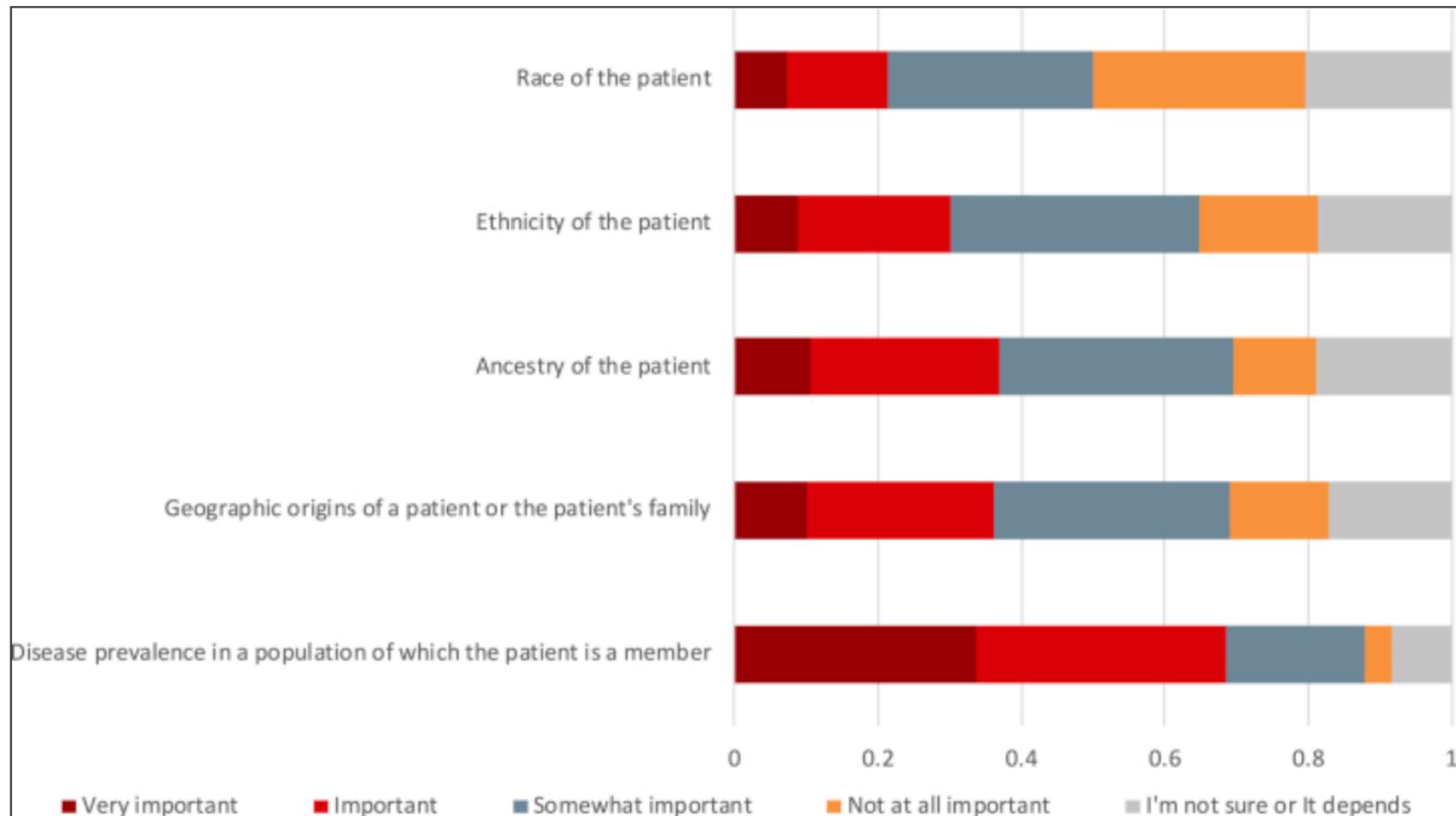


Figure 2

Perceived Importance of Patient Data Types in Ordering a Genetic Test

What is “Ethnicity”?

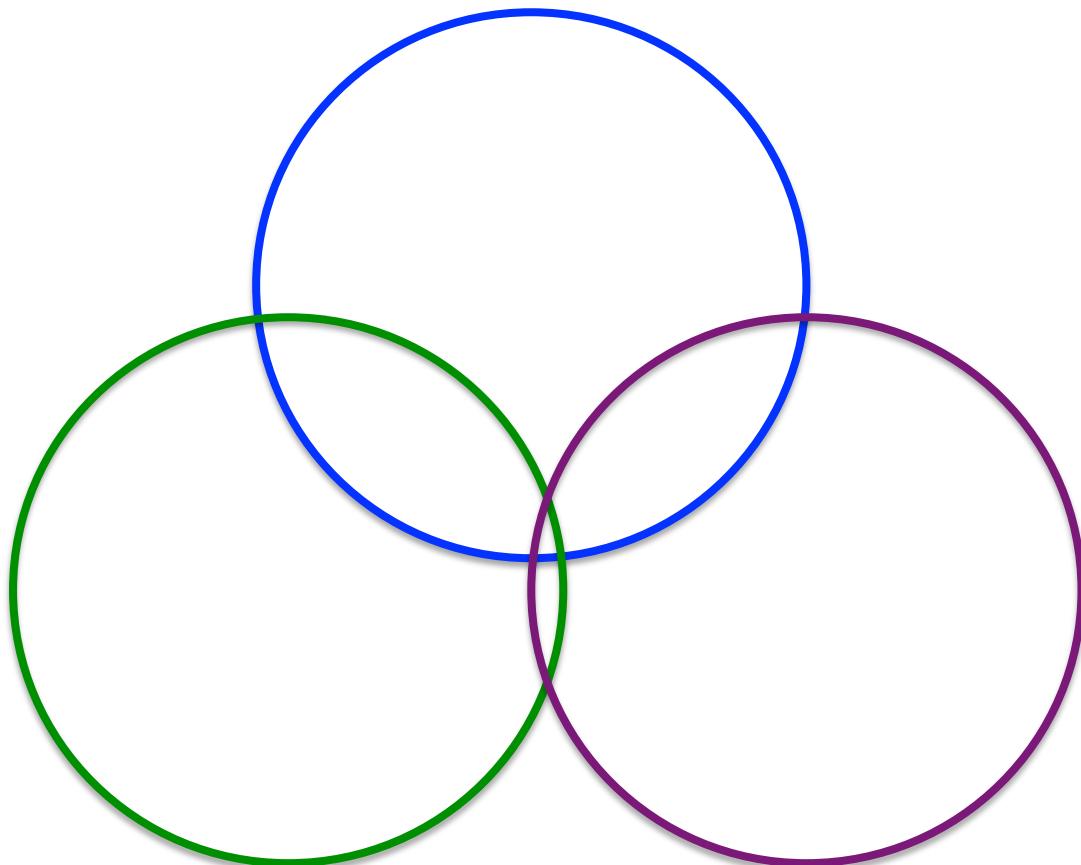
- Concerned with group cultural identity
- Based on cultural, familial, and community /religious bonds (and linguist groups) rather than physical characteristics
- It tends to be viewed as **chosen** by the individual
- It can be biologically informative: $V_p = V_G + V_E$
- It can be medically informative:
 - Epigenetics
 - Diet/Gut biome
 - Cultural practices
 - “blue zones”
 - Prions exposure (via mortuary practices)

What is “race”?

Genetic Essentialism:

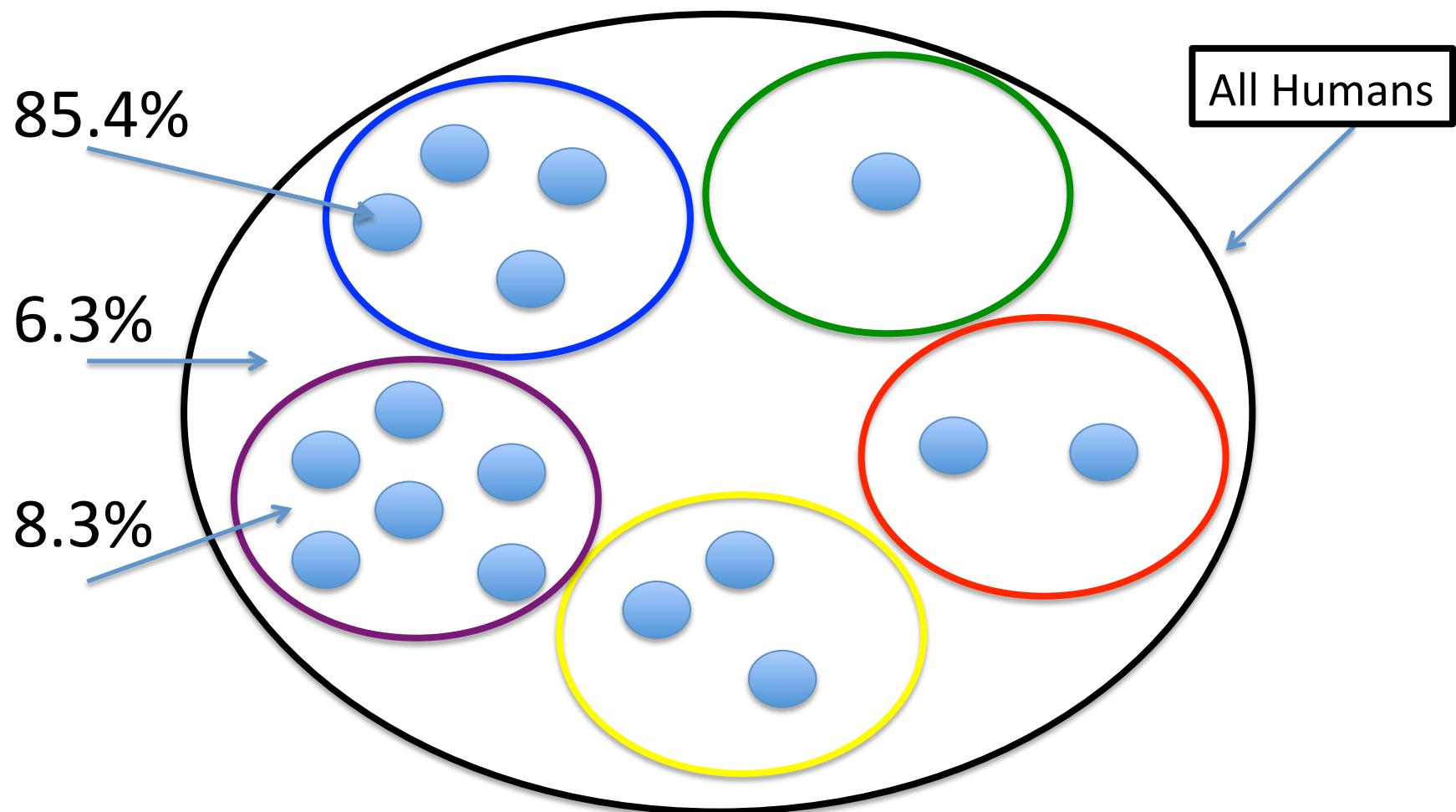
- discrete differences between ‘races’
- Uniformity within each ‘race’

- African ancestry
- East-Asian ancestry
- European ancestry
- Each circle represents the total number of alleles (gene variants) in the human genome found within a particular group



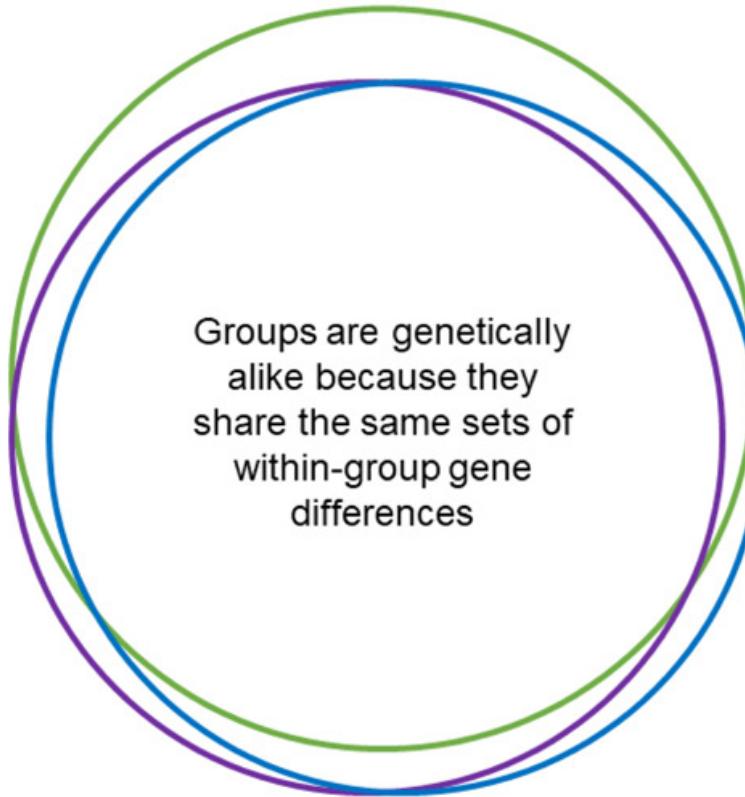
Do racial groups have taxonomic significance at a typical locus?

Lewontin, 1972: "The apportionment of human diversity"

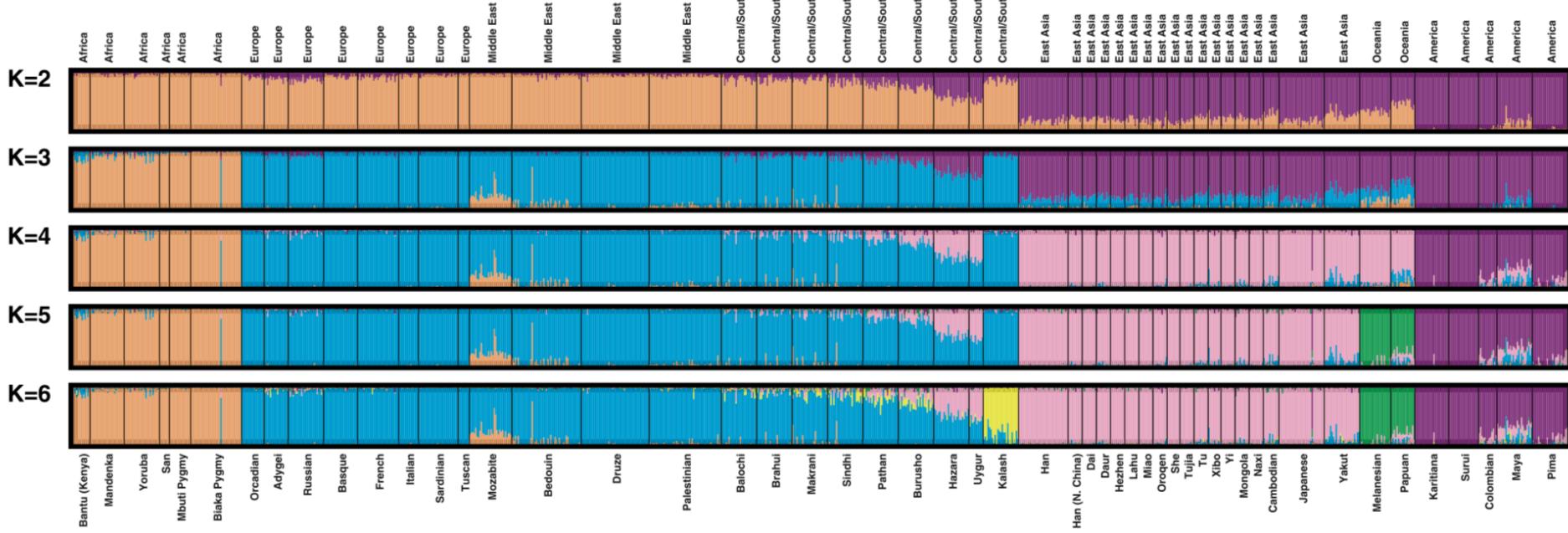


Toward a more humane genetics education: Learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations

- African ancestry
- East-Asian ancestry
- European ancestry
- Each circle represents the total number of alleles (gene variants) in the human genome found within a particular group



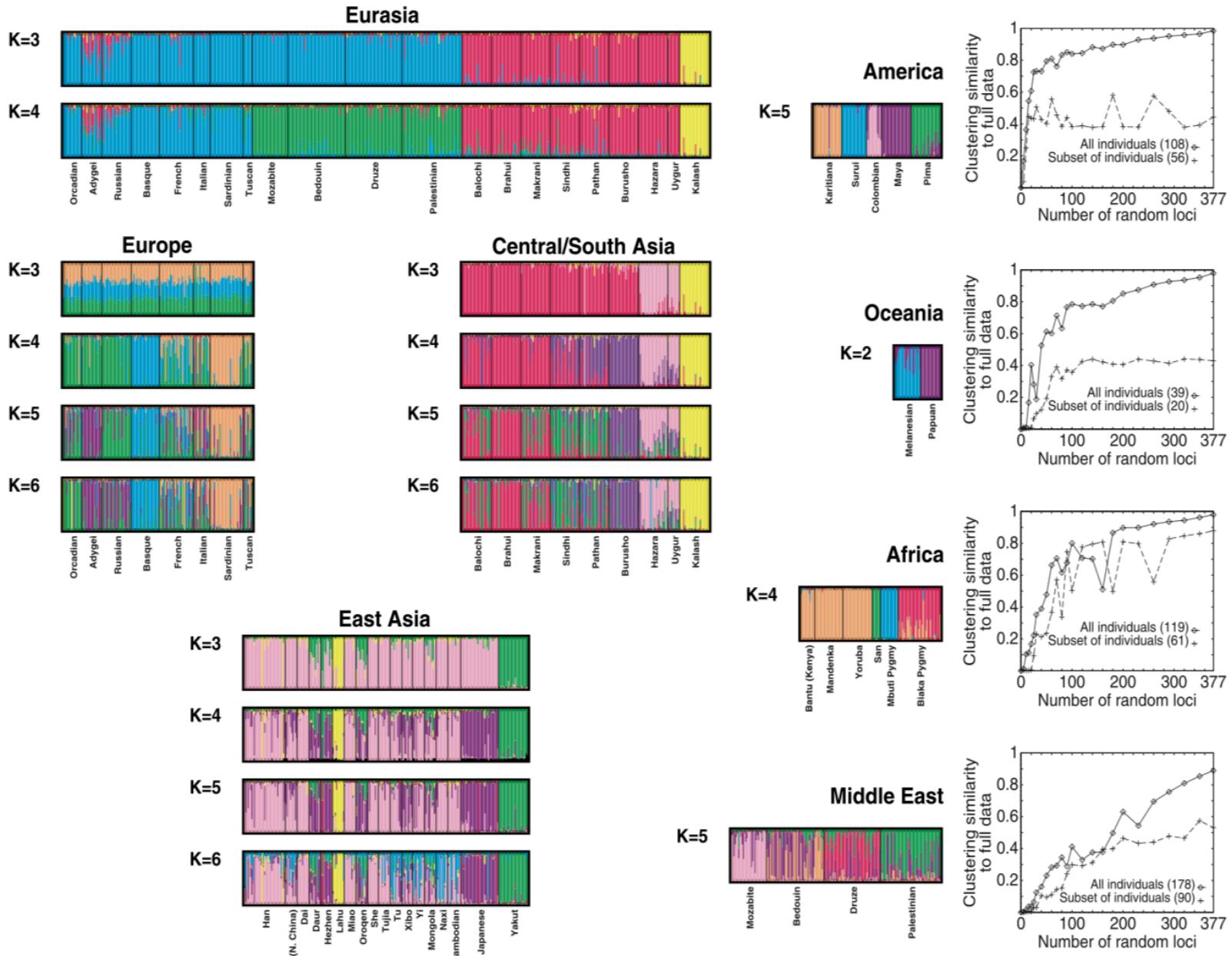
STRUCTURE RESULTS



Genetic Structure of Human Populations

Noah A. Rosenberg, *et al. Science* **298**, 2381 (2002); DOI: 10.1126/science.1078311

STRUCTURE RESULTS



STRUCTURE OF AFRICAN GROUPS

- Sohini 2022 (paper in kaplan folder) shows that using African groups in structure identifies more structure/distinct groups *within* Africa than *outside* of Africa. This is because the African samples are finally larger than they have been in the past
- This clever example illustrates how the output of structure has been unintentionally used to support racist beliefs but that it was simply due to inadequate sample sizes of African individuals.

STRUCTURE OF AFRICAN GROUPS

RACISTS USE GENETICISTS' INFOGRAPHICS

Compiling genotype data from individuals can show the genetic diversity of populations (**A**). A 2008 analysis* (co-authored by S.R.) suggested significant genetic differences between seven continental populations (**B**). But only 13.5% of the populations represented were from Africa. Boosting representation to 85% and sampling more broadly across the continent (**C**) underlines that the level of genetic variation within Africa is equivalent to that seen between continents.

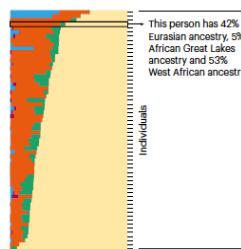
STRUCTURE OF AFRICAN GROUPS

RACISTS USE GENETICISTS' INFOGRAPHICS

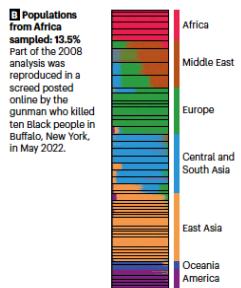
Compiling genotype data from individuals can show the genetic diversity of populations (A). A 2008 analysis* (co-authored by S.R.) suggested significant genetic differences between seven continental populations (B), but only 13.5% of the populations sampled were from Africa.

Boosting representation to 85% and sampling more broadly across the continent (C) underlines that the level of genetic variation within Africa is equivalent to that seen between continents.

A Individual genomes
Each horizontal bar corresponds to the genome of a single person, in this case from a population identified or self-identified as African American.



B Populations from Africa sampled: 13.5%
Part of the 2008 analysis was reproduced in a screenshot posted online by the gunman who killed ten Black people in Buffalo, New York, in May 2022.

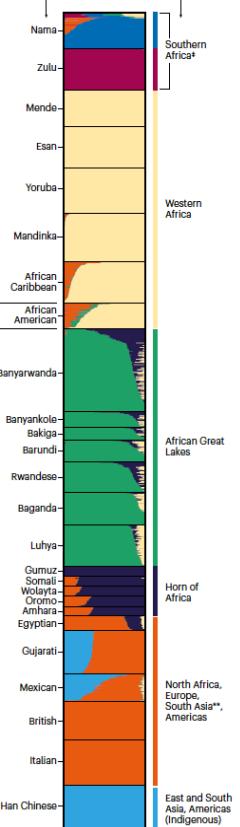


*Z. Li et al. Science 319, 1100-1104 (2008).
†An algorithm that generates data from ~100 individuals into something that looks like ~10,000 individuals. The populations represented are from Africa (see Supplementary Information). With this algorithm, the difference between African and non-African ancestries is low (blue and orange), relative to differentiation within Africa.
‡Southern Africa: South Africa, Namibia, Botswana, Mozambique; Africa: Nigeria, Kenya, Tanzania, Uganda, Rwanda, Burundi, Democratic Republic of the Congo, Malawi; Horn of Africa: Ethiopia, Somalia, Djibouti; North Africa: Egypt, Libya, Algeria, Tunisia.
**South Asia appears twice because Gujarati people in India have intermediate allele frequencies.

C Populations from Africa sampled: 85% (unpublished)
Adjusting the sampling and using an African-centric data set creates a more representative view.

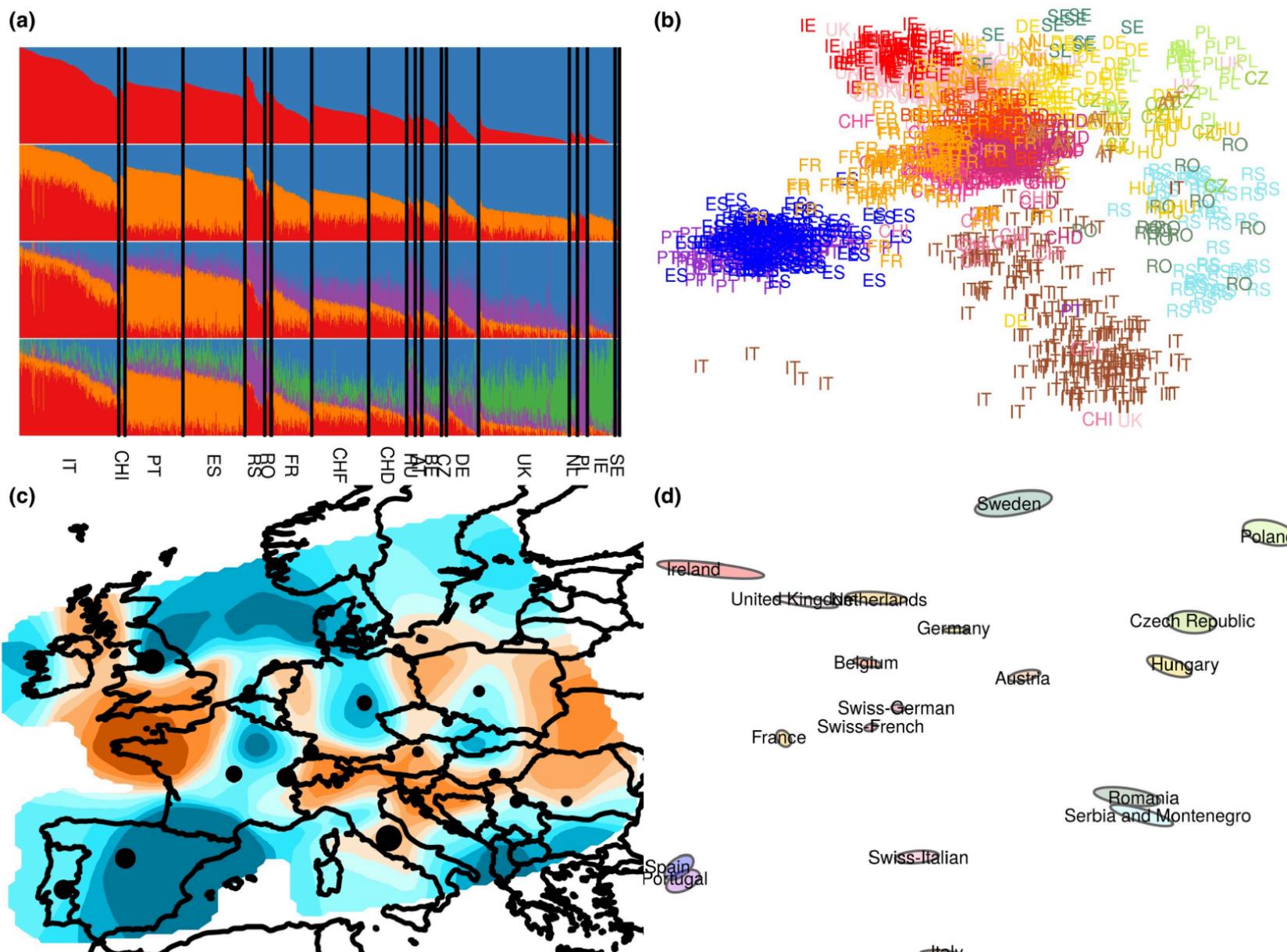
Population names are sample providers' self-identifications, or the descriptions used by those who collected the samples.

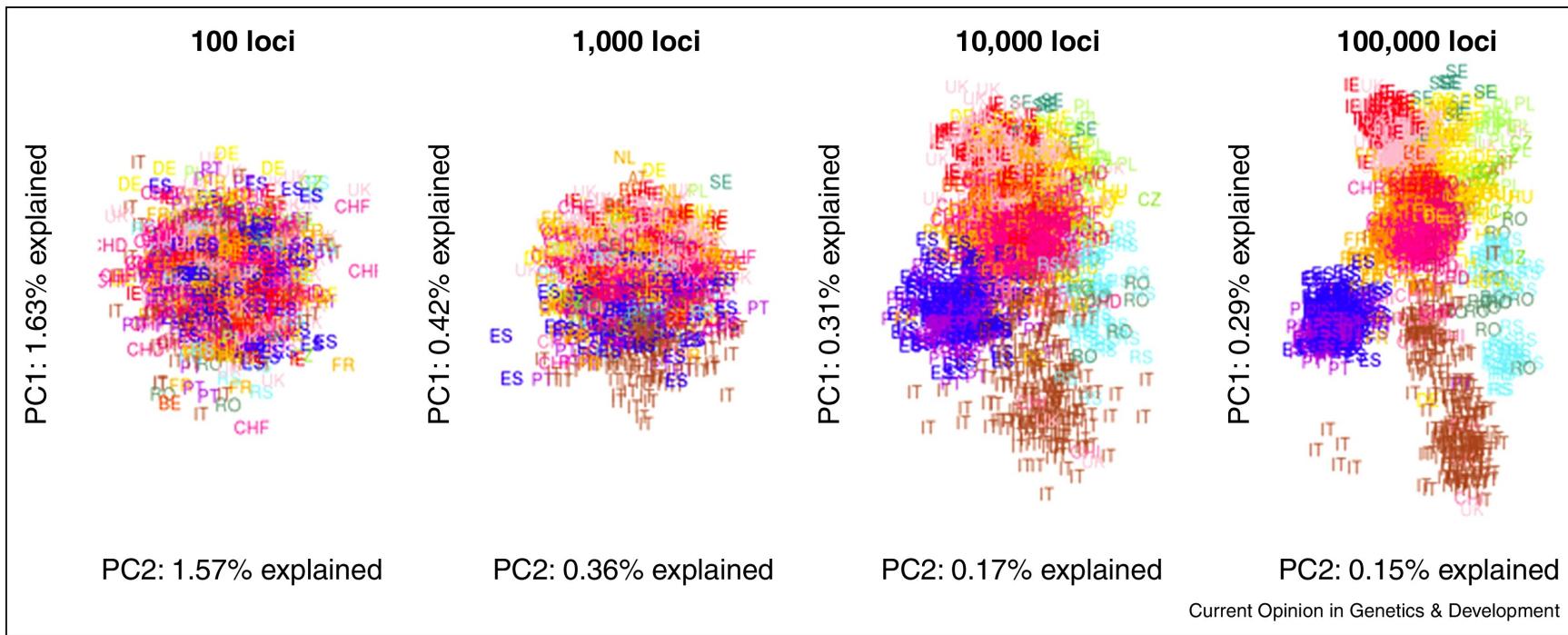
Six geographical labels correspond to the recent origins of the populations that fall mainly into the seven ancestry clusters produced by an algorithm†.



Recent advances in the study of fine-scale population structure in humans

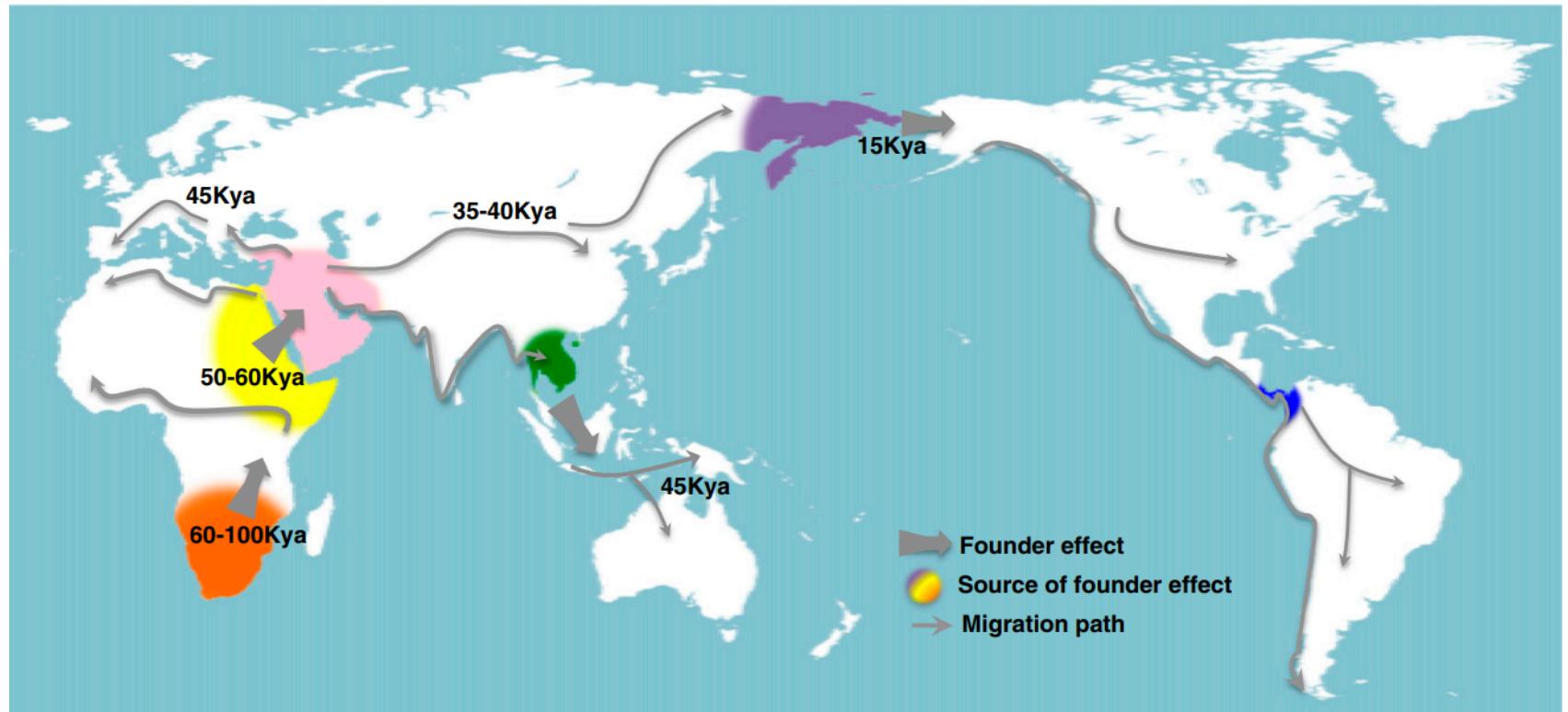
John Novembre^{1, 2}, Benjamin M Peter¹



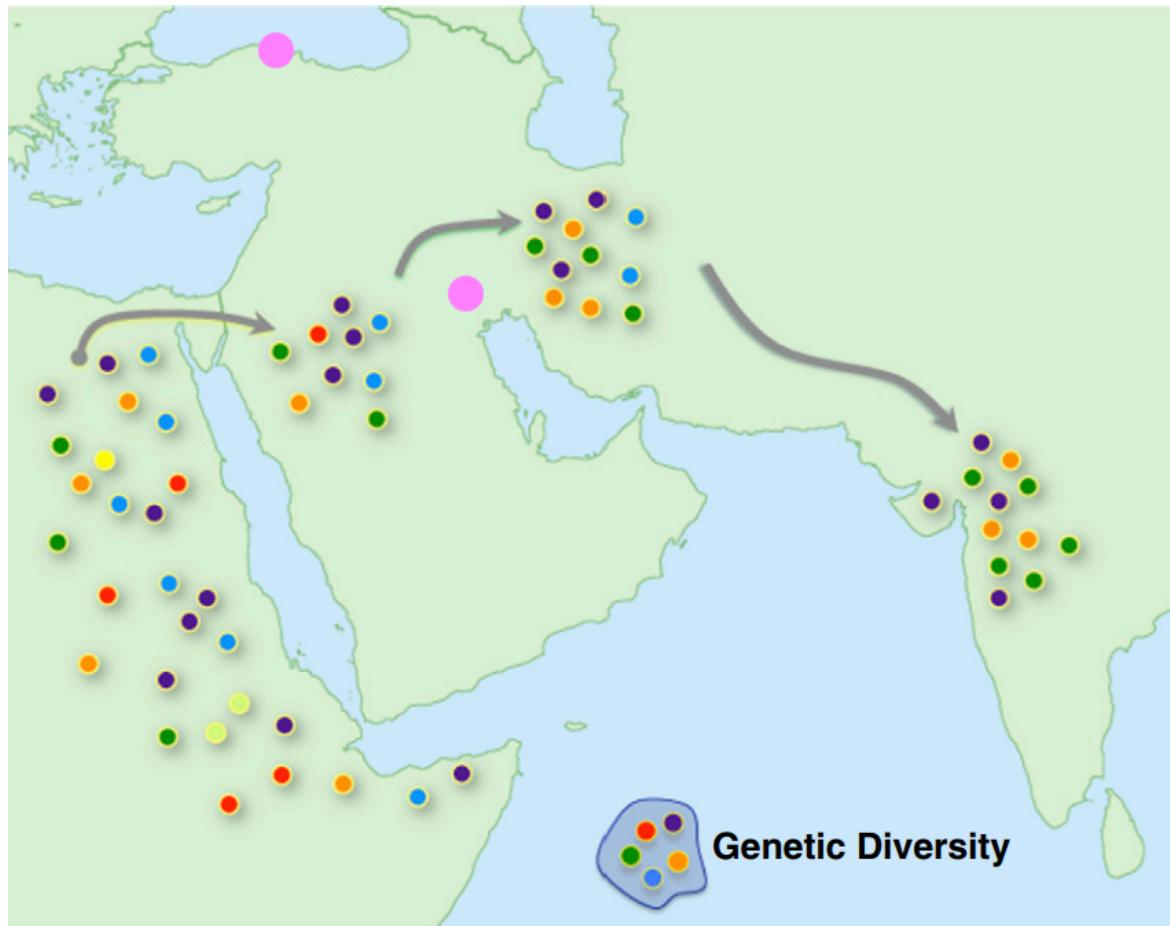


What do we mean by ancestral background/population?

- All humans are mostly identical at the genome level
- Small fraction of our DNA that differs
- Ancestry patterns is defined by human migration events
 - We originated in Africa and migrated
 - African populations have about 1 million more variants than non-African populations because they are older populations



Henn, Cavalii-Sforza, Feldman (2012). The great human expansion. PNAS, 109(44)

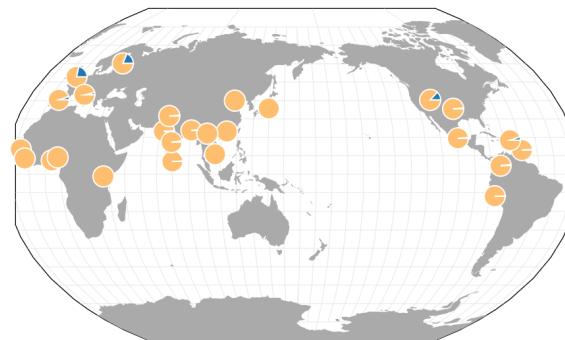


Henn, Cavalii-Sforza, Feldman (2012). The great human expansion. PNAS, 109(44)

Interactive Genetic Variants Browser

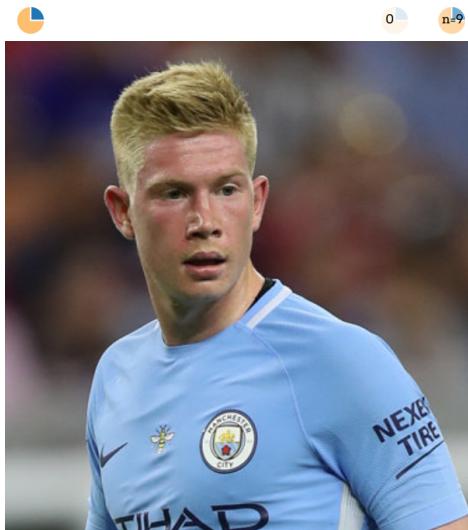
<https://popgen.uchicago.edu/ggv/?data=%221000genomes%22&chr=9&pos=12396731>

chr12:89328335 C/T



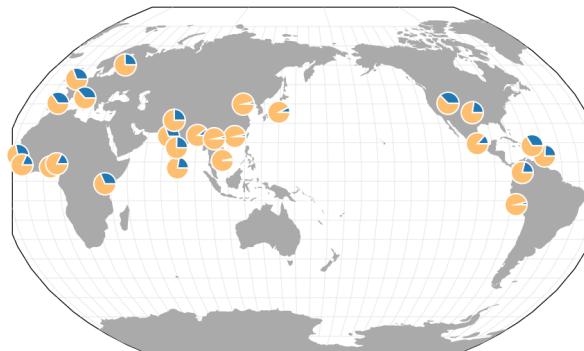
Frequency Scale = Proportion out of 1
The pie below represents a minor allele frequency of 0.25

Sample sizes below 30 become increasingly transparent to represent uncertain frequencies, i.e.



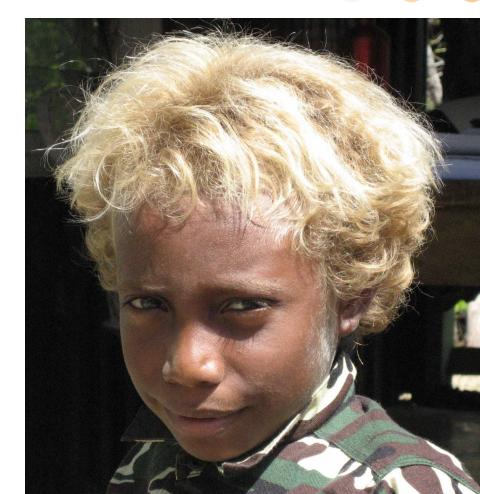
0 n=9 n=18 n=27

chr9:12396731 G/A



Frequency Scale = Proportion out of 1
The pie below represents a minor allele frequency of 0.25

Sample sizes below 30 become increasingly transparent to represent uncertain frequencies, i.e.

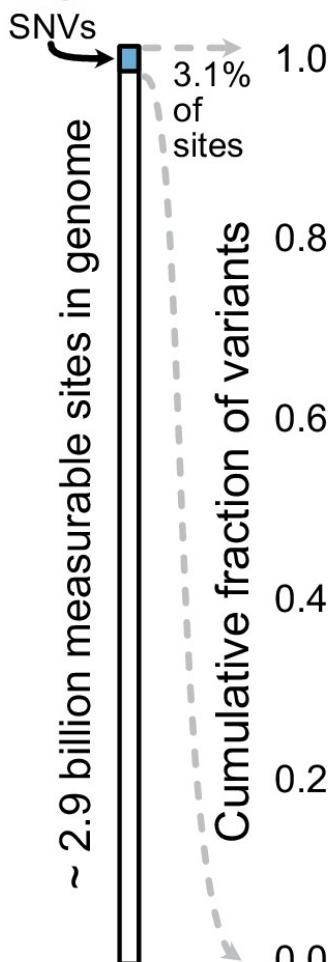


0 n=9 n=18 n=27

In common, but different

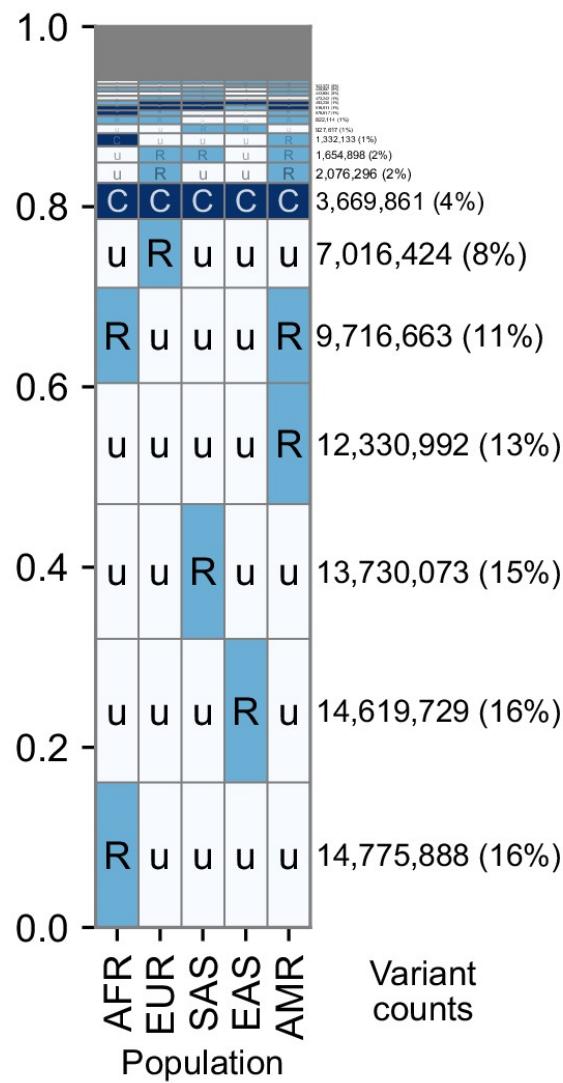
- Common variants differ in frequencies, but they arose in Africa before migration
 - After migration, individual populations accumulate private mutations (including singletons --- mutations in only one person)
 - Forces that change the composition in a population:
 - Founder/drift
 - Mutations
 - Recombination
- Take-home messages:
 1. Race is a social construct
 2. Ancestry is genetic, but it is a continuum of genetic variation – it is not discrete
 3. The story of human history isn't complete yet – and it may never be

A Whole genome



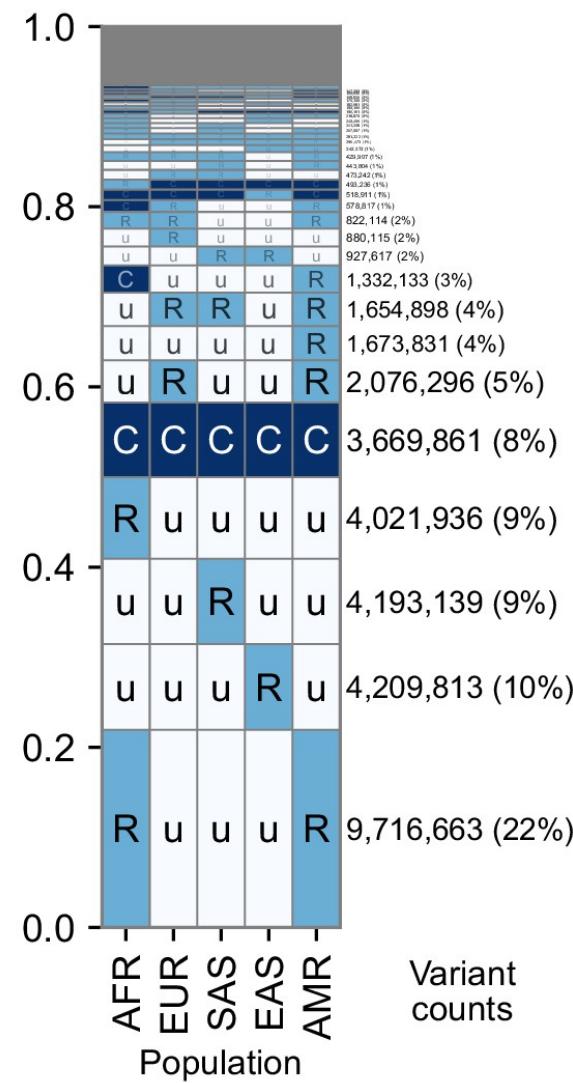
B Bi-allelic SNVs

S = 91,784,637



C Without singletons

S = 44,290,364



Question 1: Does knowing two individuals race assignments, give you information about their genetic differences at one locus?

Question 2: Does knowing an individuals race give you information about their genetics at multiple loci?

Question 1: Does knowing two individuals race assignments, give you information about their genetic differences at one locus?

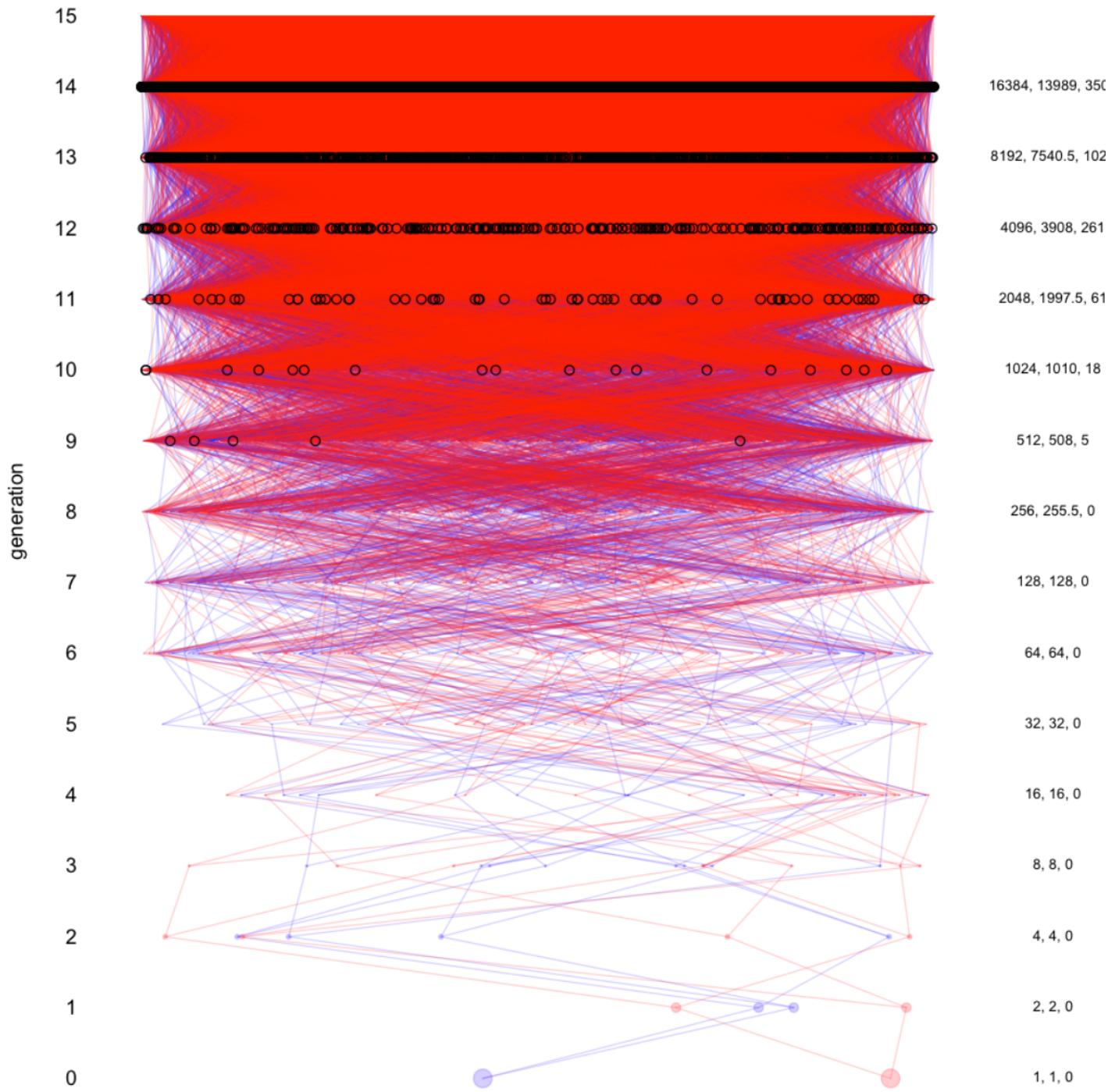
- Only about 6.3% variation explained so....no, it does not.
Race is a poor proxy for genotype at any one locus

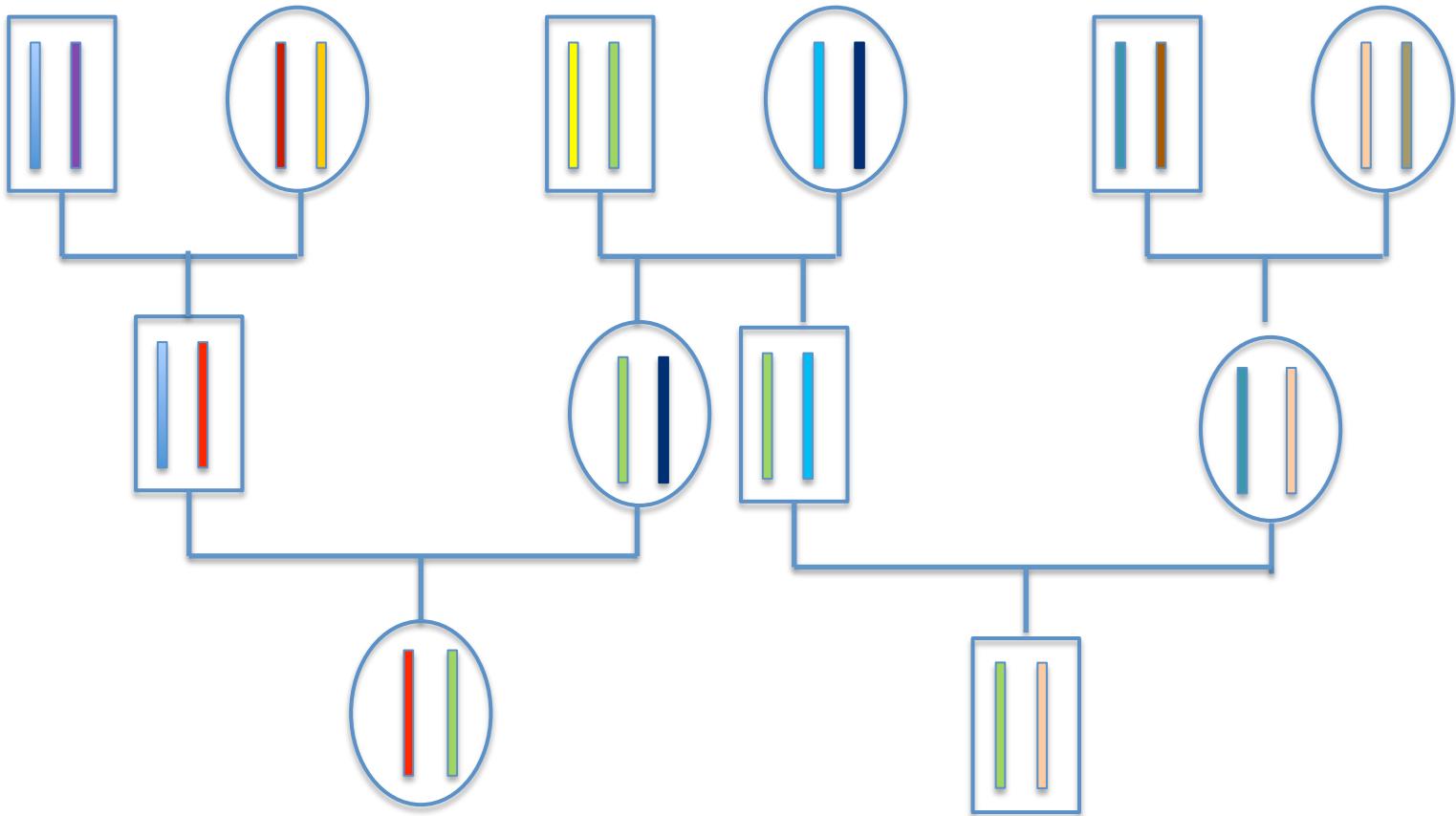
Question 2: Does knowing an individuals race give you information about their genetics at multiple loci?

- Sometimes.... Sequence alignments at multiple spots can pick up on allelic frequency differences between populations. These are the differences that programs such as STRUCTURE, ADMIXTURE, and methods such as PCA are capturing. They also *might* be informative for clinical applications! It might also lead to *false confidence, though!*

Your pedigree is not always your ancestry

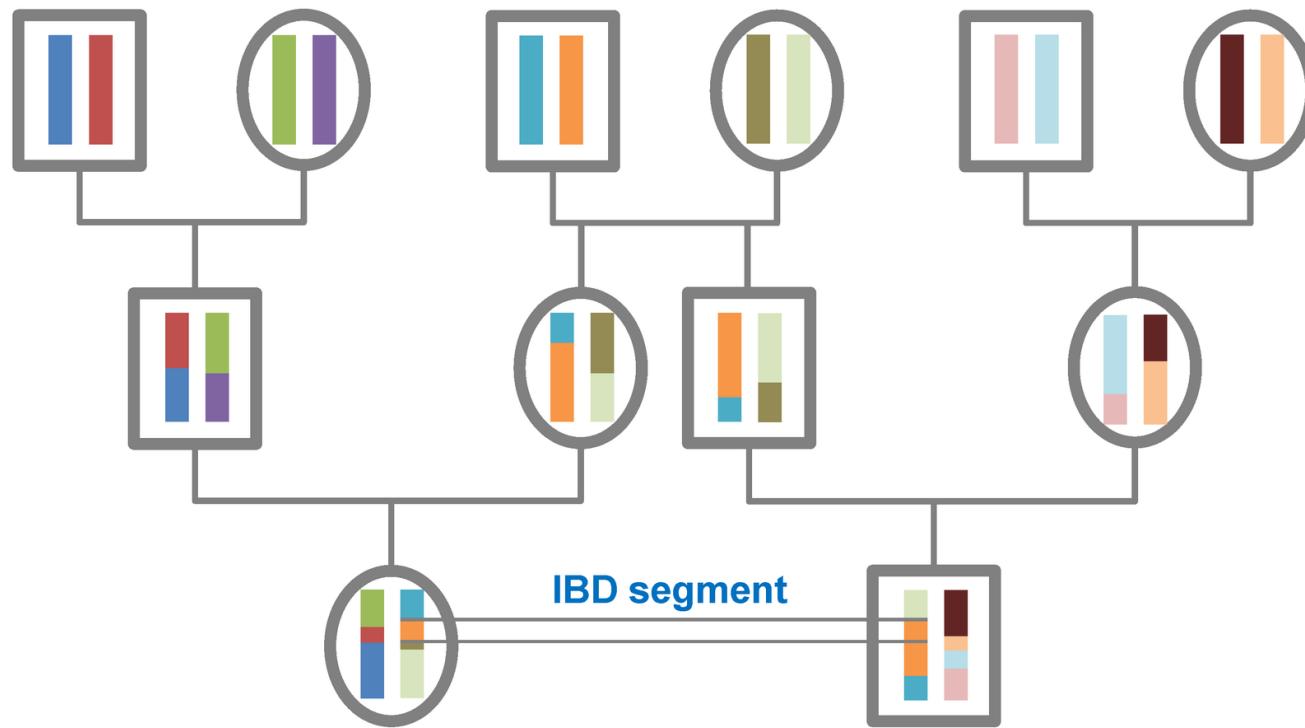
- Due to inbreeding/bottleneck and recombination
- Number of ancestors grows so quickly that is almost unavoidable that we share an ancestor:
 - 600 years ago ($2^{20} > 1,000,000$ ancestors)
 - 1000 years ago ($2^{33} > 1,000,000,000$ ancestors way more Europeans than were alive 1000 years ago)
- <https://gcbias.org/2017/11/20/our-vast-shared-family-tree/>
- <https://gcbias.org/2022/07/12/genetic-ancestry-groups-and-genetic-similarity/>





What is I.B.D.?

What I.B.D. actually looks like...



- GWAS is based on Linkage Disequilibrium
 - You assume that it doesn't pick up the actual causal variant, but rather that it is in LD with the actual causal variant
 - **Different Ancestral populations have different LD!**

Review

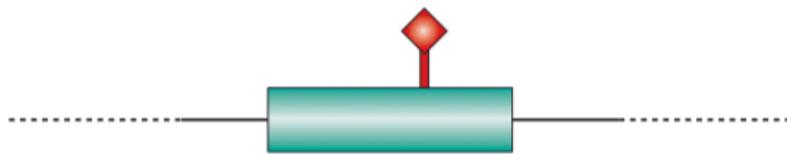
> Nat Rev Genet. 2005 Feb;6(2):95-108. doi: 10.1038/nrg1521.

Genome-wide association studies for common diseases and complex traits

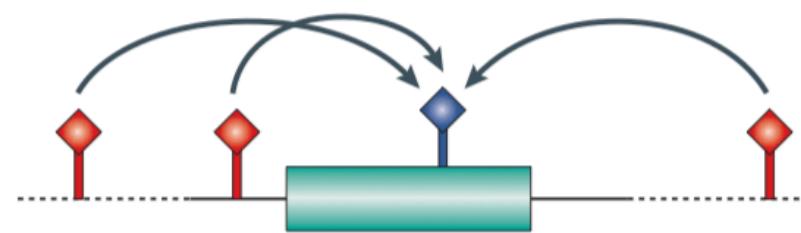
Joel N Hirschhorn ¹, Mark J Daly

Affiliations + expand

PMID: 15716906 DOI: 10.1038/nrg1521

a

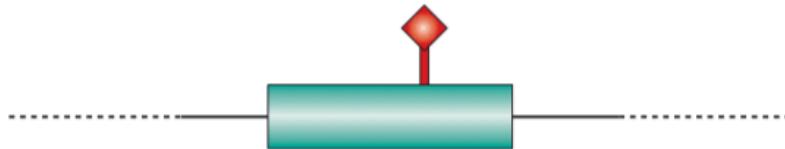
Direct association

b

Indirect association

East-Asian Linkage Disequilibrium

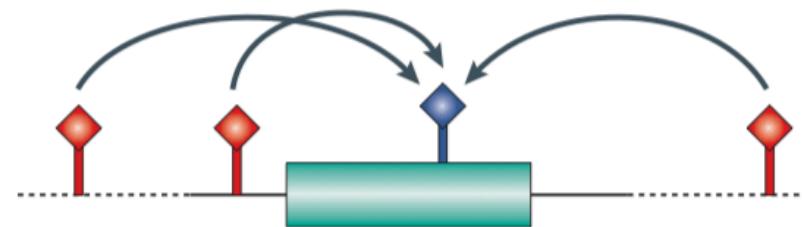
a



Direct association

African Linkage Disequilibrium

b



Indirect association

Will the disease ‘causal’ variant should be the same in all populations?

- Nope – rare mutations also to deleterious diseases and these tend to be geographically restricted (or even to an individual or family group)
- Common variants are usually shared by multiple populations and older
- Common versus rare can have slightly different contributions to disease in different populations
- Each person carries ~60 de novo points mutations and ~3 indels that were created in their parental germlines
- Mutation rates vary, environments vary (malaria resistance), population histories vary (demography, and cultural practices)

