

Population Genetic Structure, Race/Ethnicity and Disease



Molecular and Genetic Epidemiology
EPI 217, Winter 2018

Outline

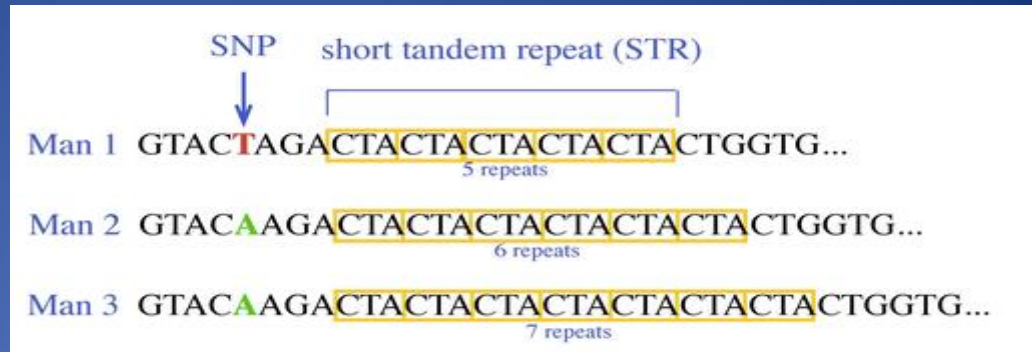
- Genetic variation & Race/Ethnicity
- Genetic admixture, admixture analysis and admixture mapping
- Example: Breast cancer in Latinas

Genetic diversity: variation in the genomes of organisms within (or among) populations

Genetic structure: Refers to the patterning of genetic diversity across multiple groups

Genetic variation

- Repeats

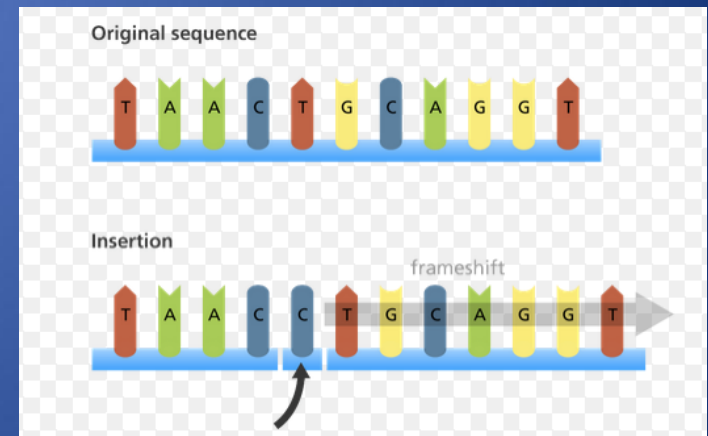


- Single nucleotide polymorphism (A/T)

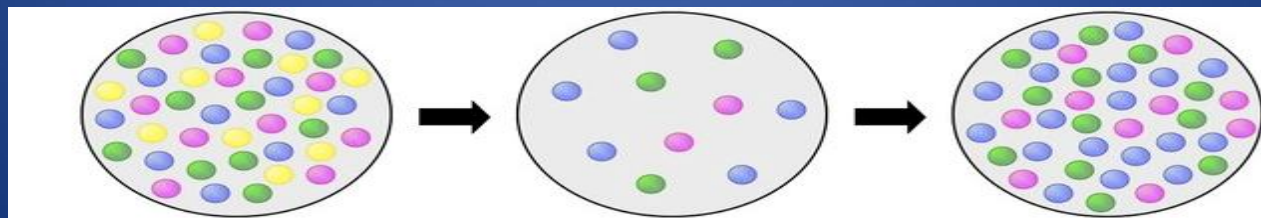
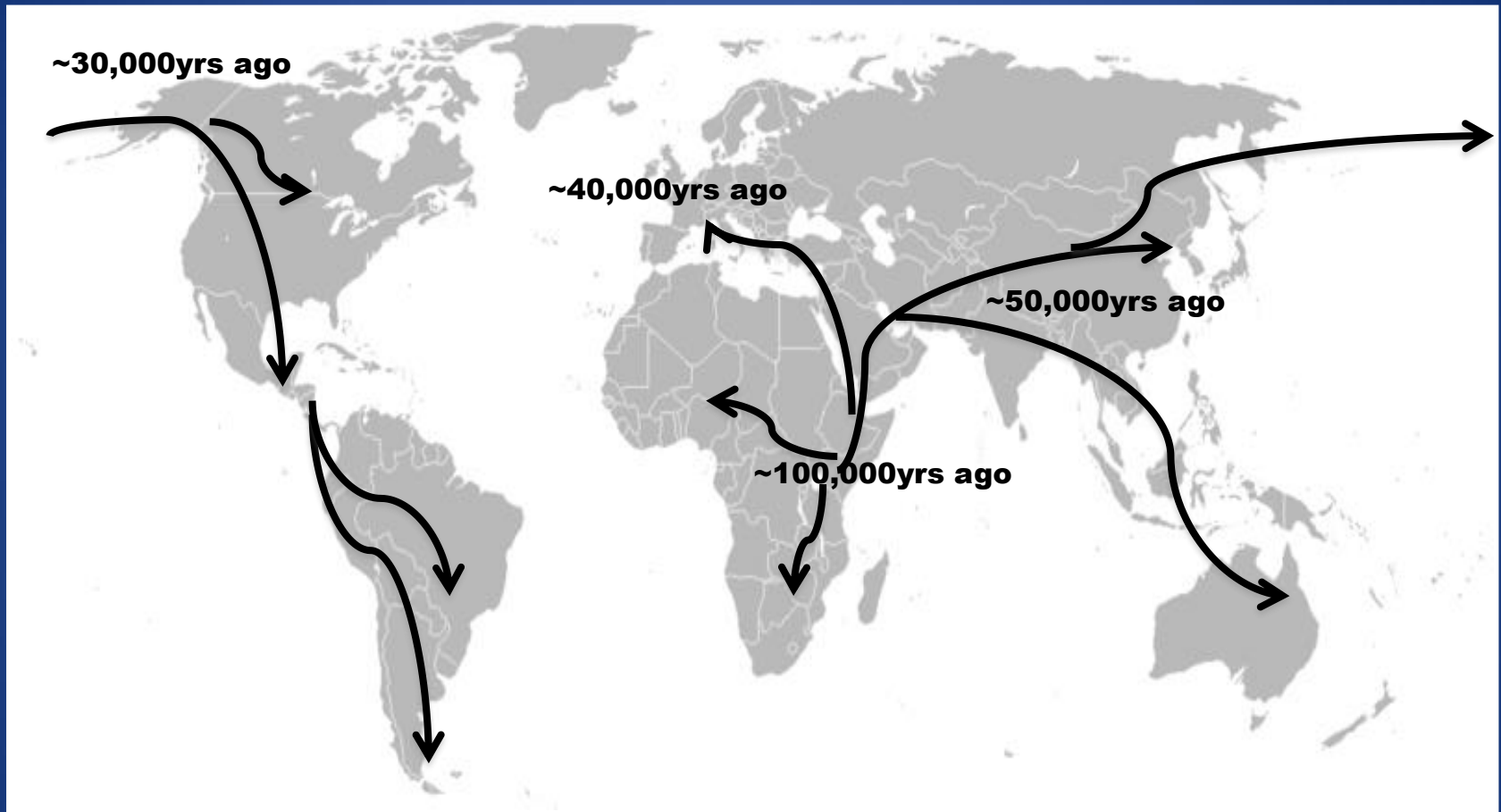
- Indels (C/-)

- Structural variants

- Copy number variants
- Alu insertions
- Inversion



Origins of population structure



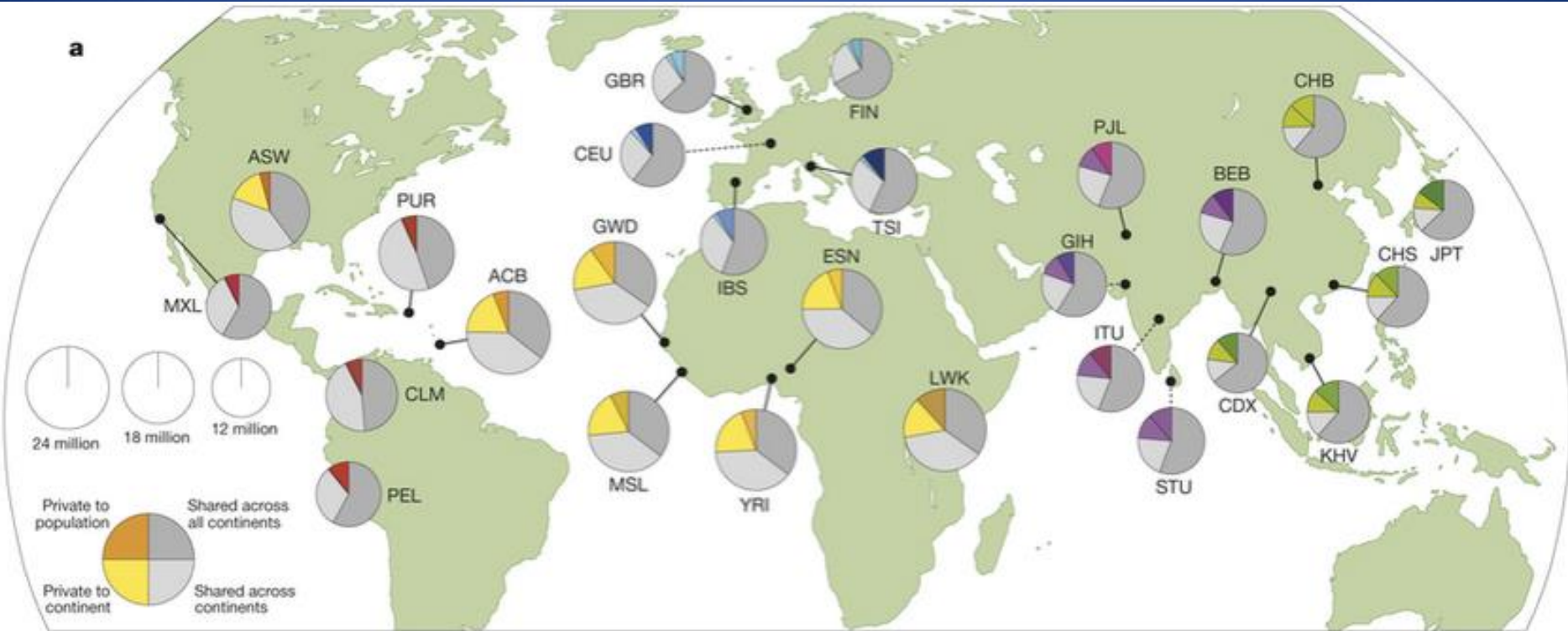
Allele frequency differences

- Mutation → genetic drift (founder effect, population bottleneck), gene flow, selection

SNP	Group A	Group B
	Allele Frequency	
1 (T/C)	0.25T	0.75T
2 (G/A)	0.50A	0.55A
3 (G/A)	0.90A	0.75A
4 (C/T)	1.00T	0.00T

Human genetic diversity

2,504 individuals from 26 populations



84.7 million SNPs

3.6 million indels

60,000 structural variants

Relationships between populations based on genetic variation

- Tree diagrams can be created based on genetic distances between populations

- Absence/presence of mutation

AAA

ATA

ATT

TTT

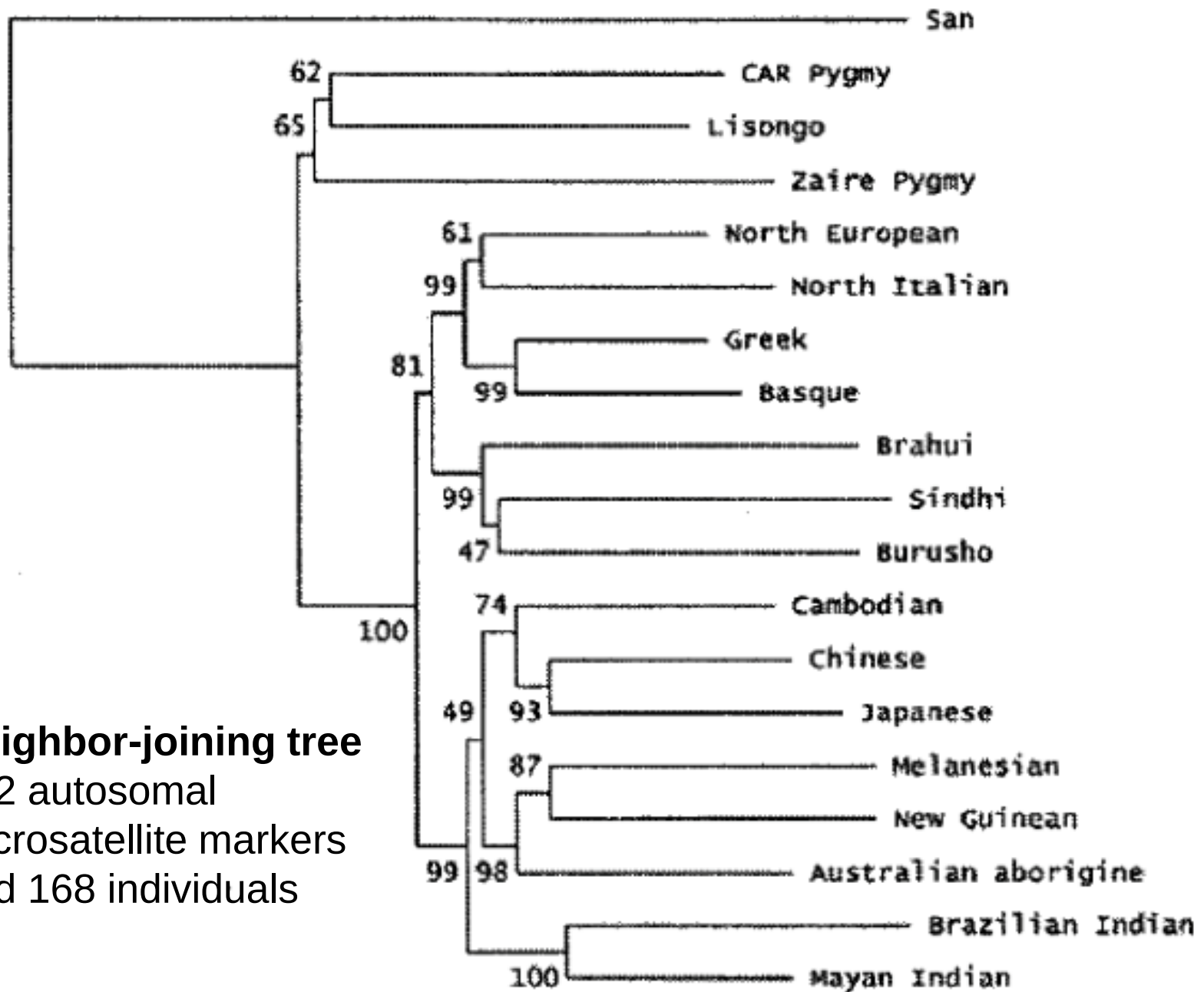
- Distances based on population allele frequencies

0.1

0.15

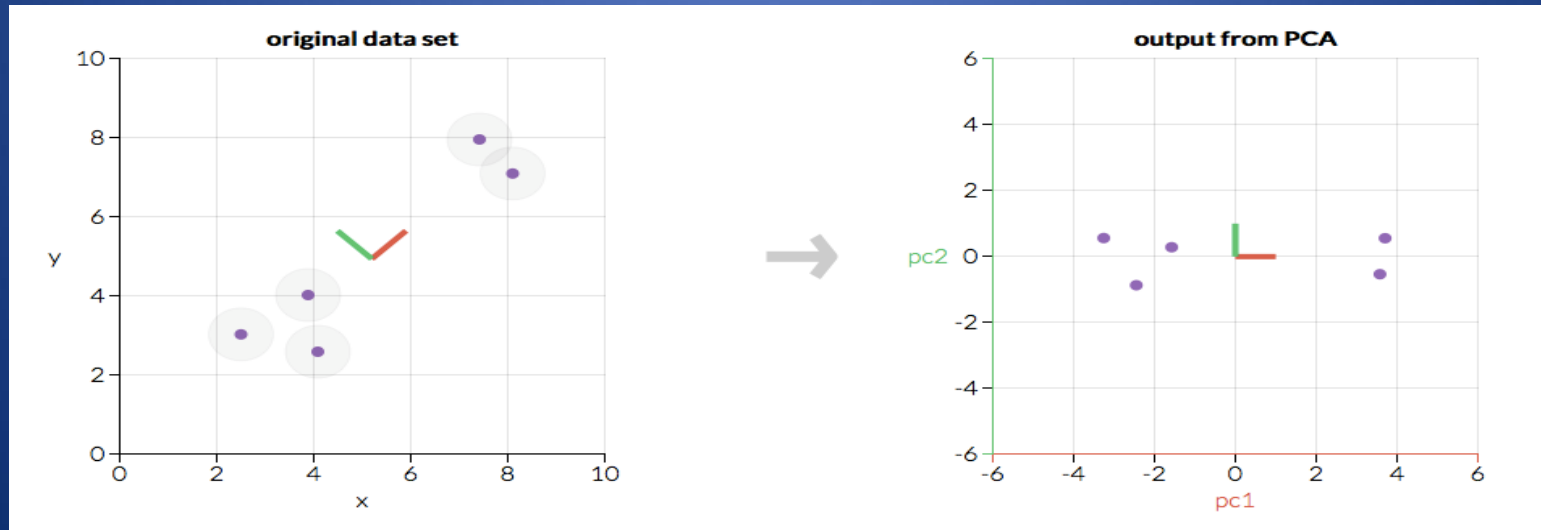
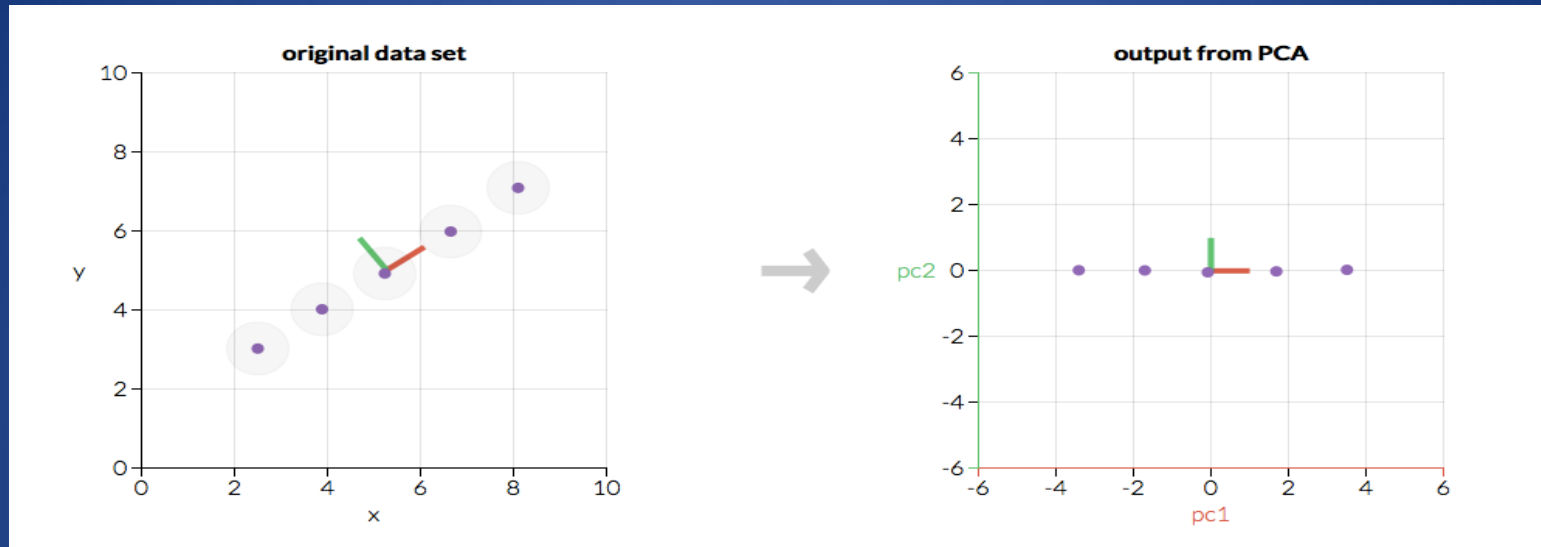
0.50

0.60



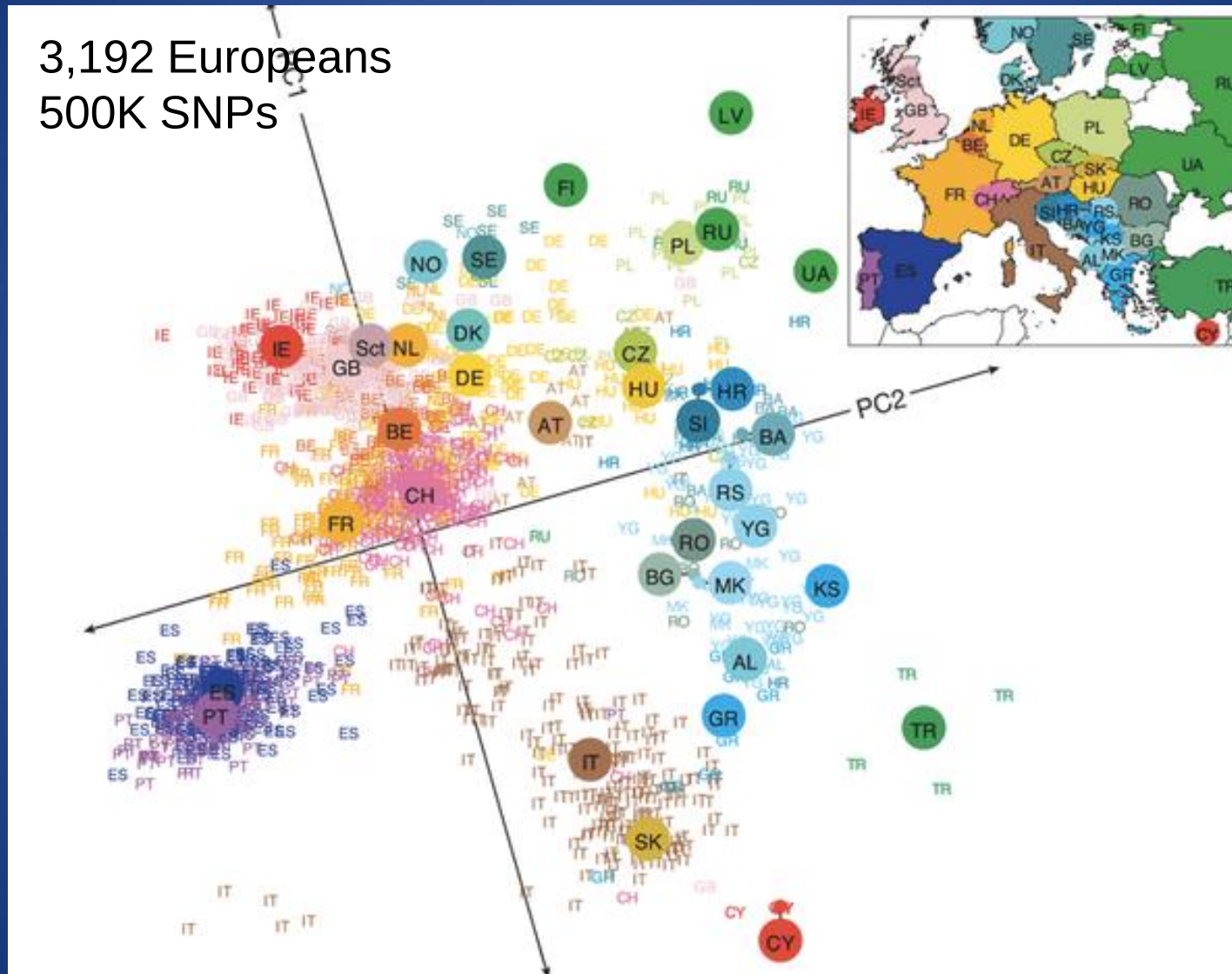
Neighbor-joining tree
182 autosomal
microsatellite markers
and 168 individuals

Principal Components Analysis



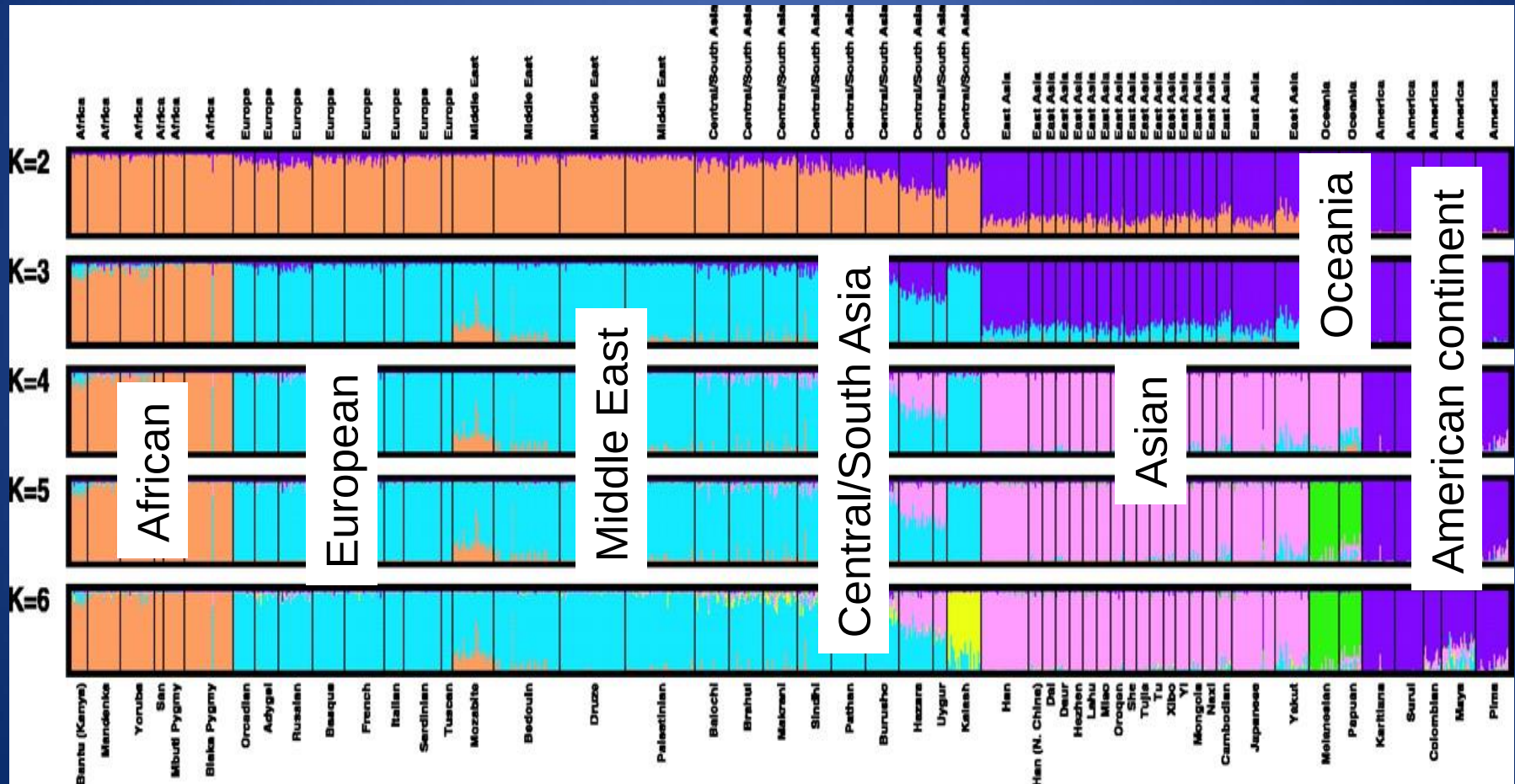
Population structure within Europe

3,192 Europeans
500K SNPs



Model-based clustering

377 Autosomal microsatellite loci in 1056 individuals from 52 populations



Noah Rosenberg et al, Science, 2002

How are race and ethnicity defined?

- U.S. Census Race Categories:
 - African/African American/Black
 - European/White
 - Asian
 - Pacific Islander
 - Native American/Alaskan Native
 - 2 or More Races
- U.S. Census Ethnicity Category:
 - Hispanic/Latino

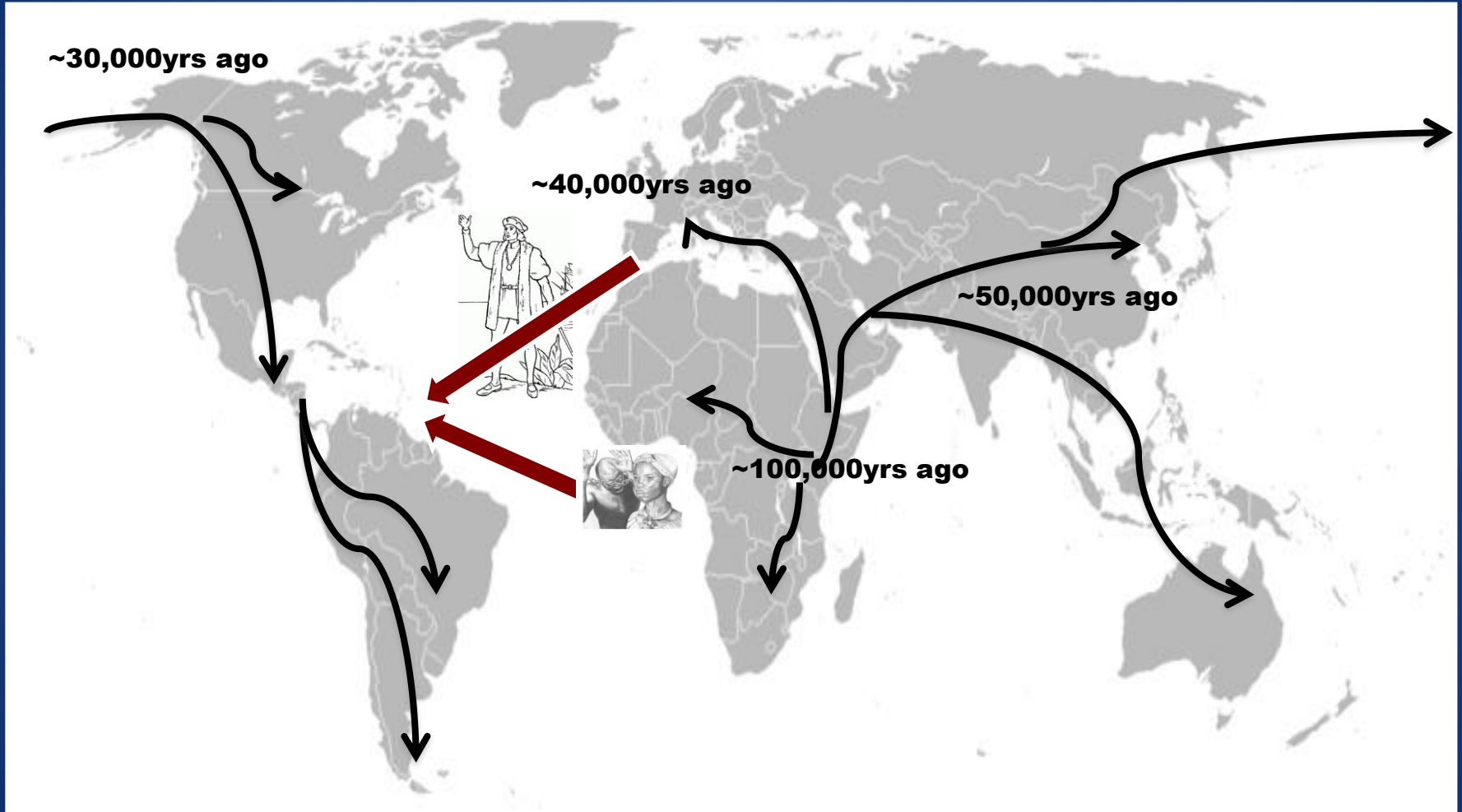
How are race and ethnicity defined?

- **Race** is a social construct very often linked to ancestry
- **Ancestry** defined in terms of continental origin of ancestors
- **Ethnicity**: cultural heritage, ancestry, origin myth, history, homeland, language and/or dialect, symbolic systems such as religion, mythology and ritual, cuisine, dressing style, art, and physical appearance

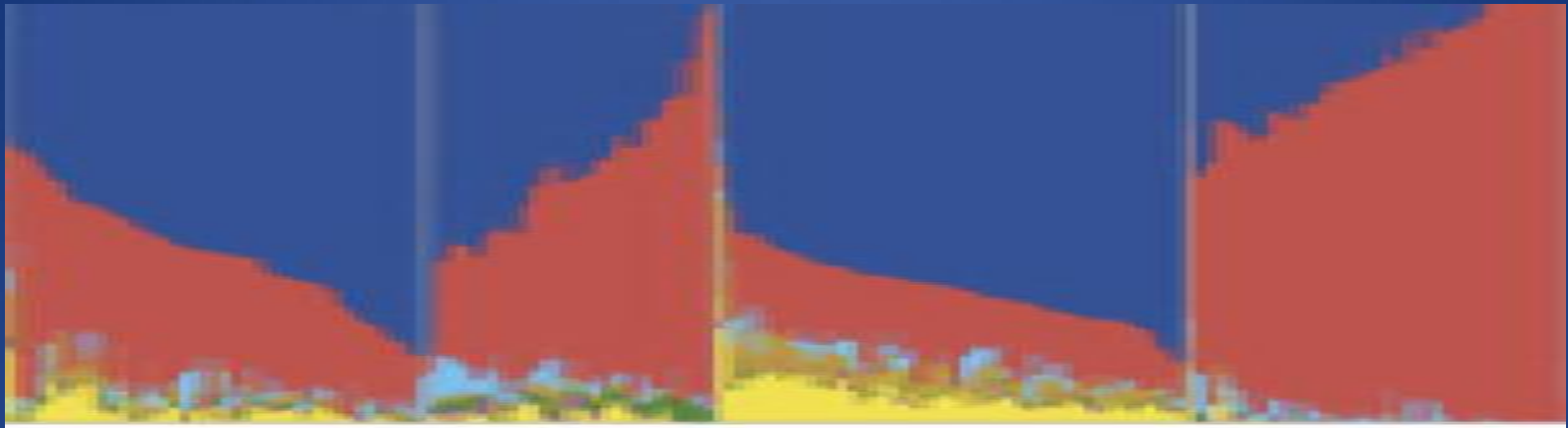
Summary

- The primary source of genetic structure in the human population is based on continent of origin
- These groupings are not perfectly discrete: 5 major groups (?)
- An additional level of structure exists between national groups within the major continental groupings
- What does “race” mean?

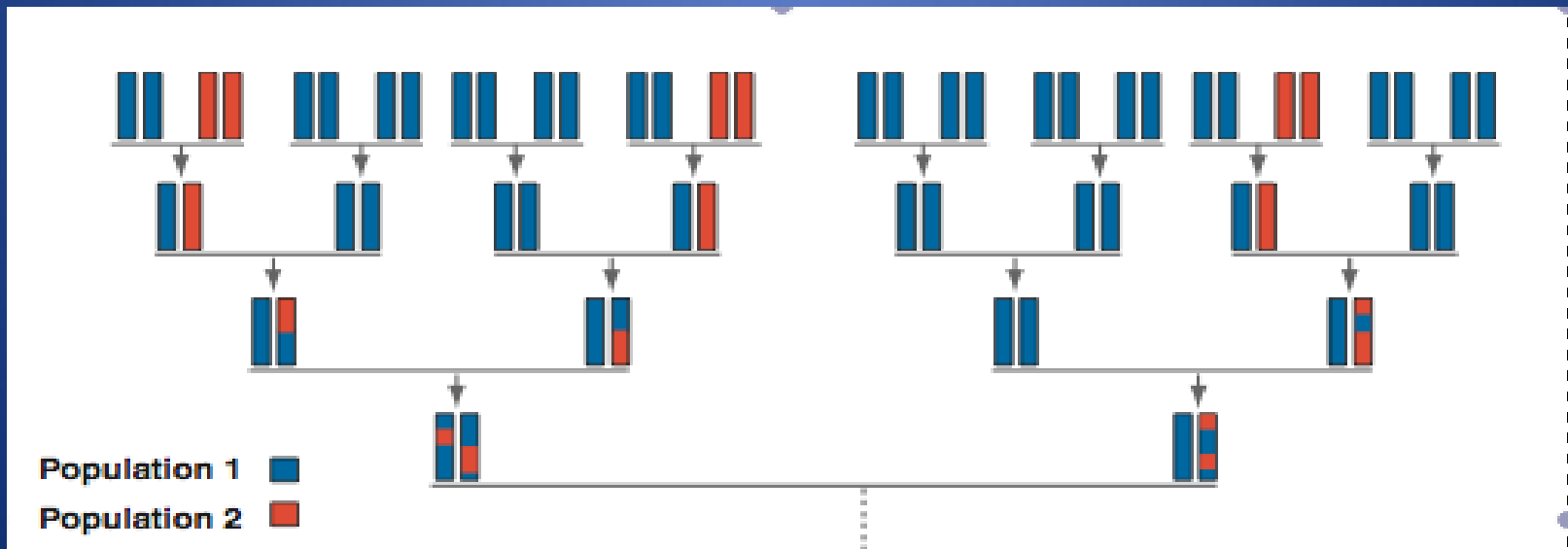
Admixed populations



Genetic ancestry and admixture



Nature 2015, 1000 Genomes Project Consortium



Darvasi & Shifman *Nat. Genet.* (2005)

Genetic Admixture Estimation

- Panel of ancestry informative markers (AIMs)

SNP	Group A	Group B
Allele Frequency		
1 (T/C)	0.25T	0.75T
2 (G/A)	0.50A	0.55A
3 (G/A)	0.90A	0.75A
4 (C/T)	1.00T	0.00T

$$p_{\text{admix}} = p_a * m + p_b * (1 - m)$$

$$m = (p_{\text{admix}} - p_b) / (p_a - p_b)$$

$$P(A|a_{ij}) = [P(a_{ij}|A) * P(A)] / [P(a_{ij}|A) * P(A) + P(a_{ij}|E) * P(E)]$$

Admixture estimation

(model-based clustering approach)

- FRAPPE

<http://med.stanford.edu/tanglab/software/frappe.html>

- ADMIXTURE

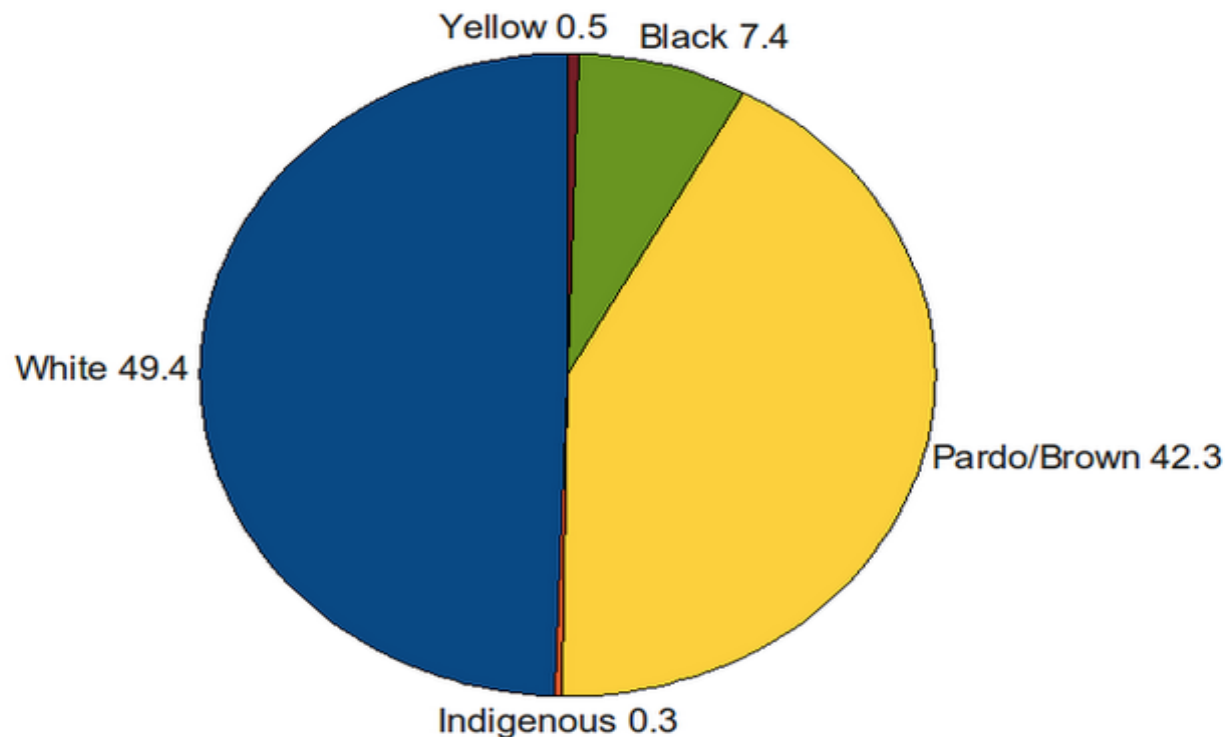
<http://www.genetics.ucla.edu/software/admixture/index.html>

- STRUCTURE

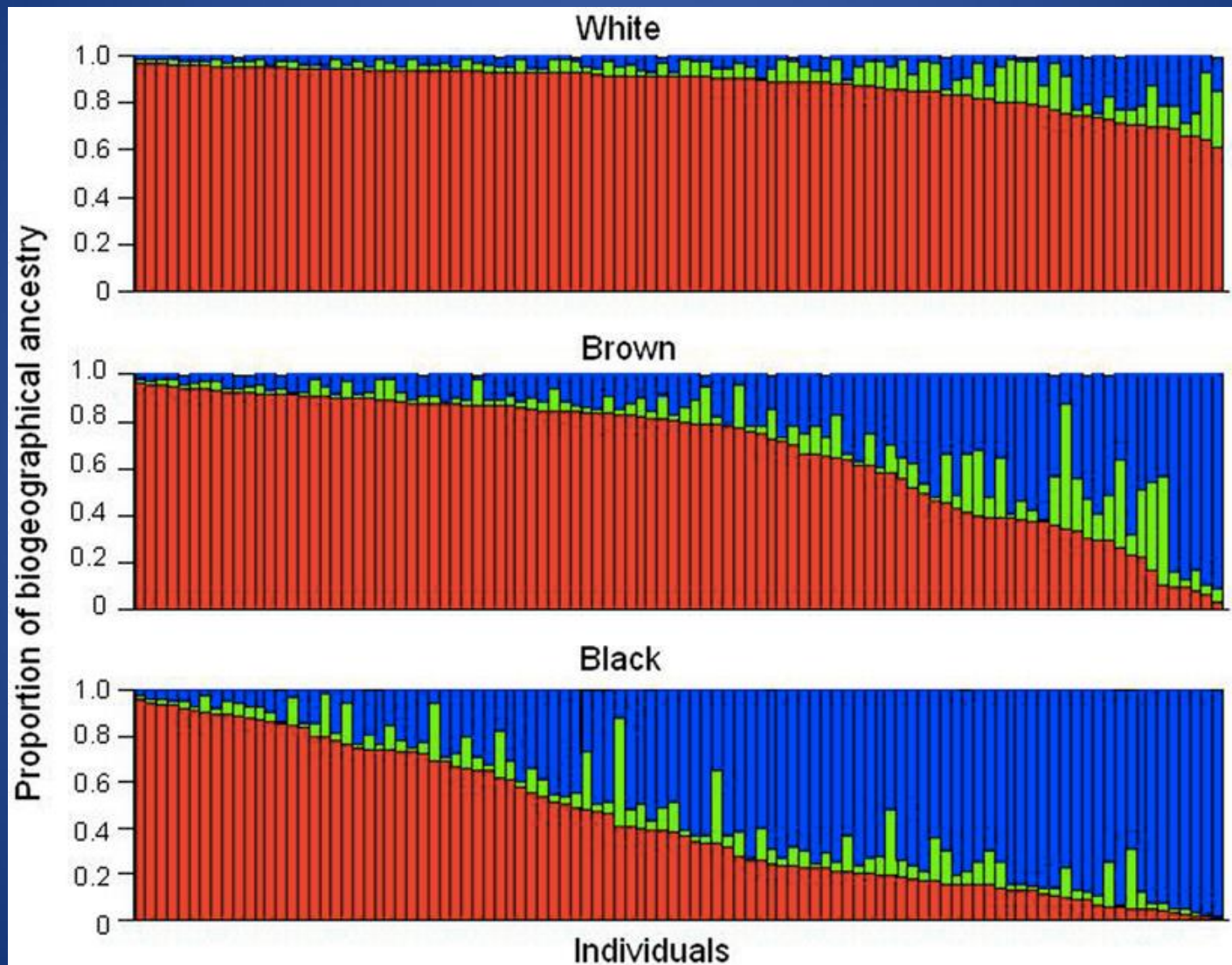
<http://pritch.bsd.uchicago.edu/structure.html>

Race/ethnicity in admixed populations

Race in Brazil, 2006.

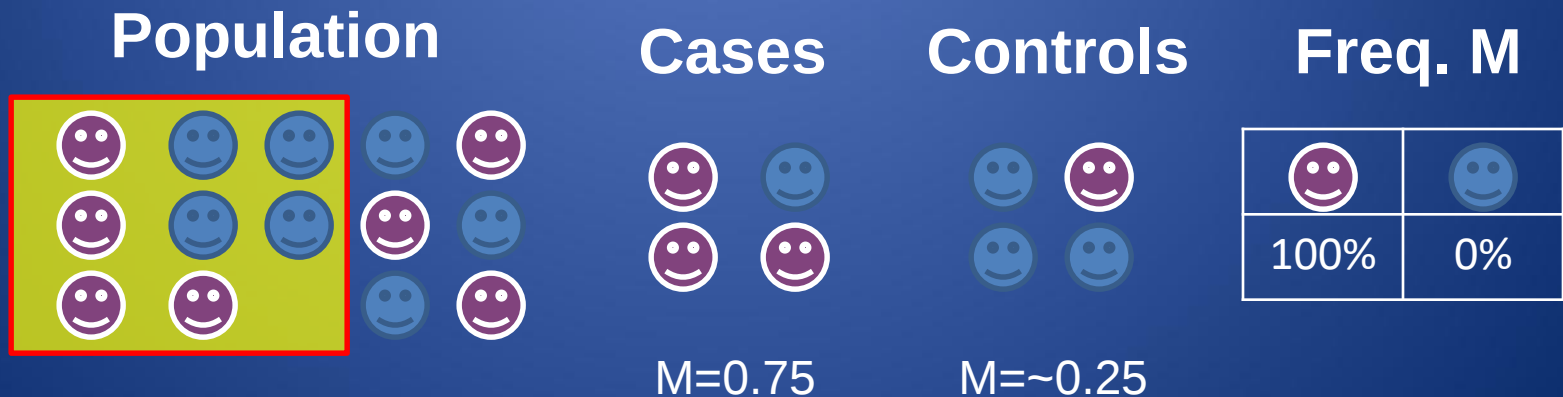


Brazilians



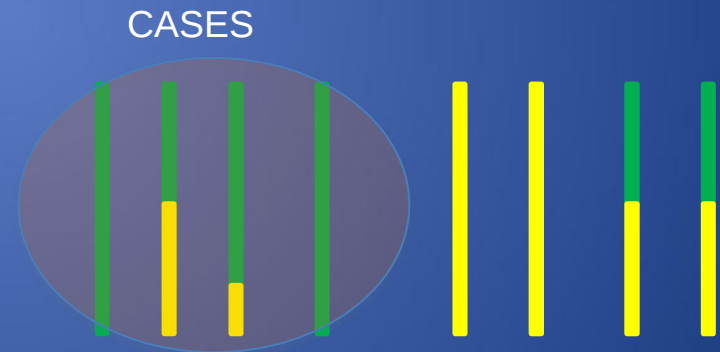
Why is genetic admixture important in medical research?

- Genetic association studies are susceptible to false-positive results if there are differences in genetic ancestry between cases and controls



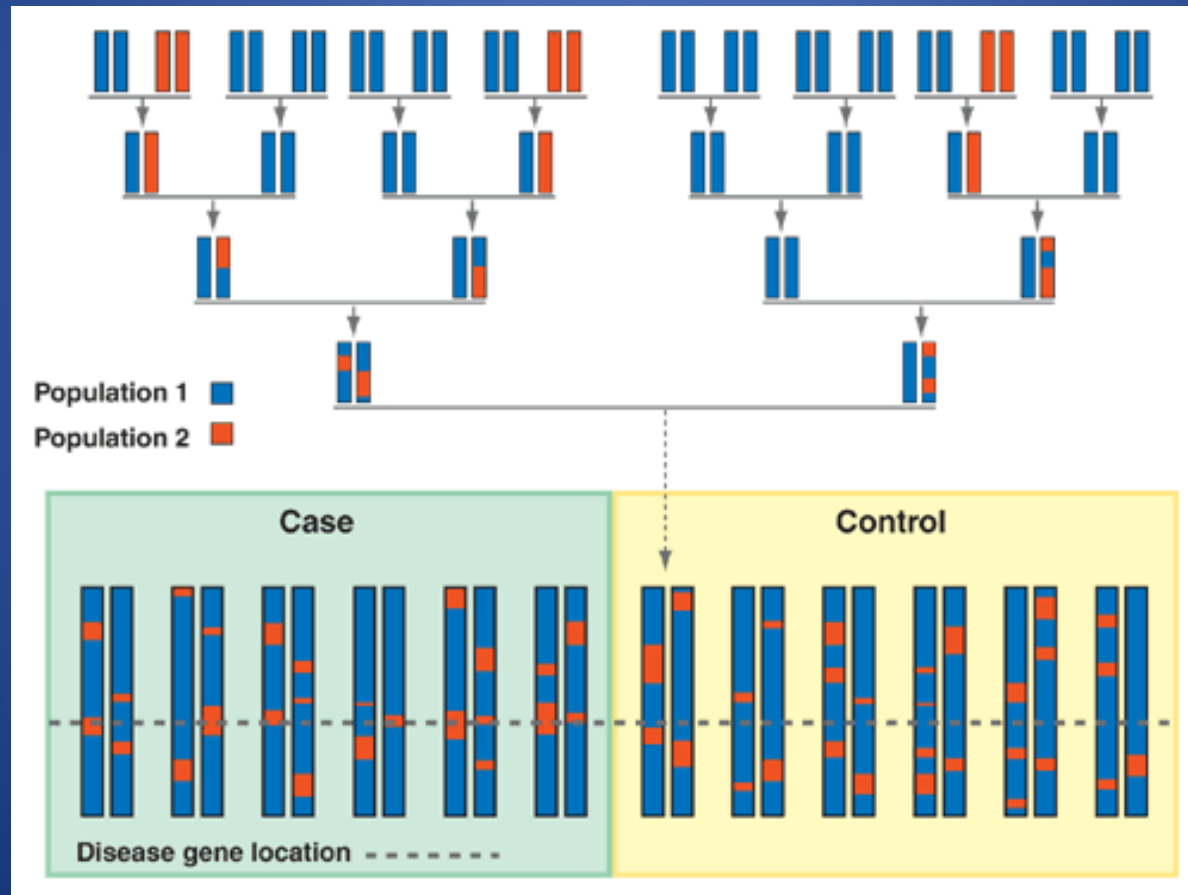
Genetic Admixture: Unique opportunity to find disease predisposing genes

- 1st → Disease that has different frequency in dissimilar populations
- 2nd → Do cases differ from controls in terms of genetic ancestry? If yes, strong basis for...
- 3rd → Admixture mapping



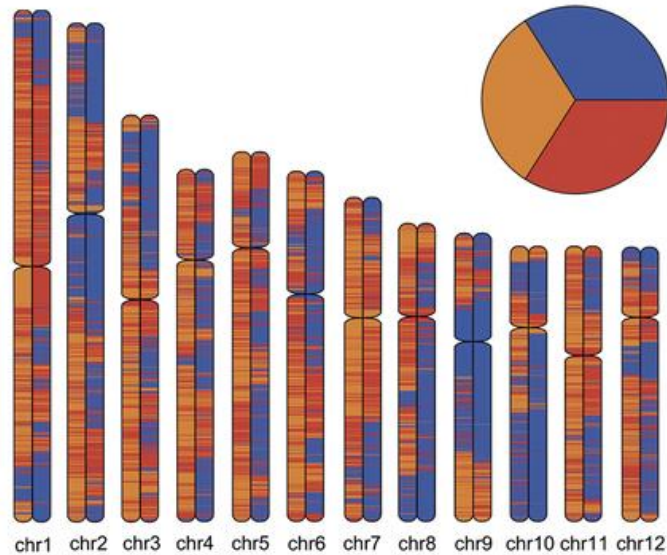
Admixture mapping

Specific genetic regions with ancestry difference between cases and controls might harbor disease predisposing or protective variants

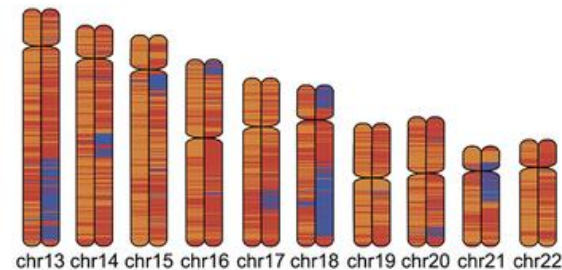
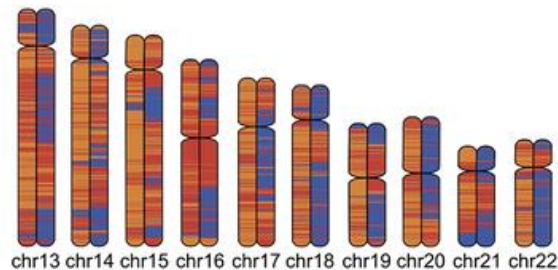
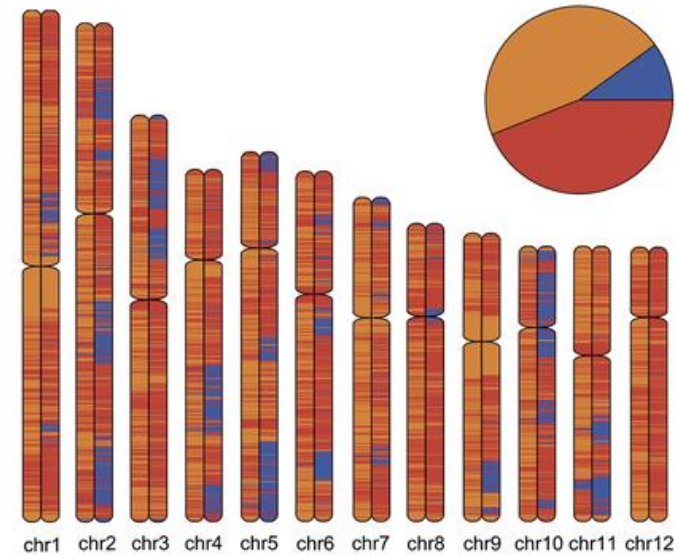


Locus-Specific Ancestry

■ African (34%) ■ Native American (32%) ■ European (34%)



■ African (10%) ■ Native American (46%) ■ European (44%)



Locus-Specific Ancestry

- LAMP

<http://lamp.icsi.berkeley.edu/lamp/>

- HAPMIX

<http://www.stats.ox.ac.uk/~myers/software.html>

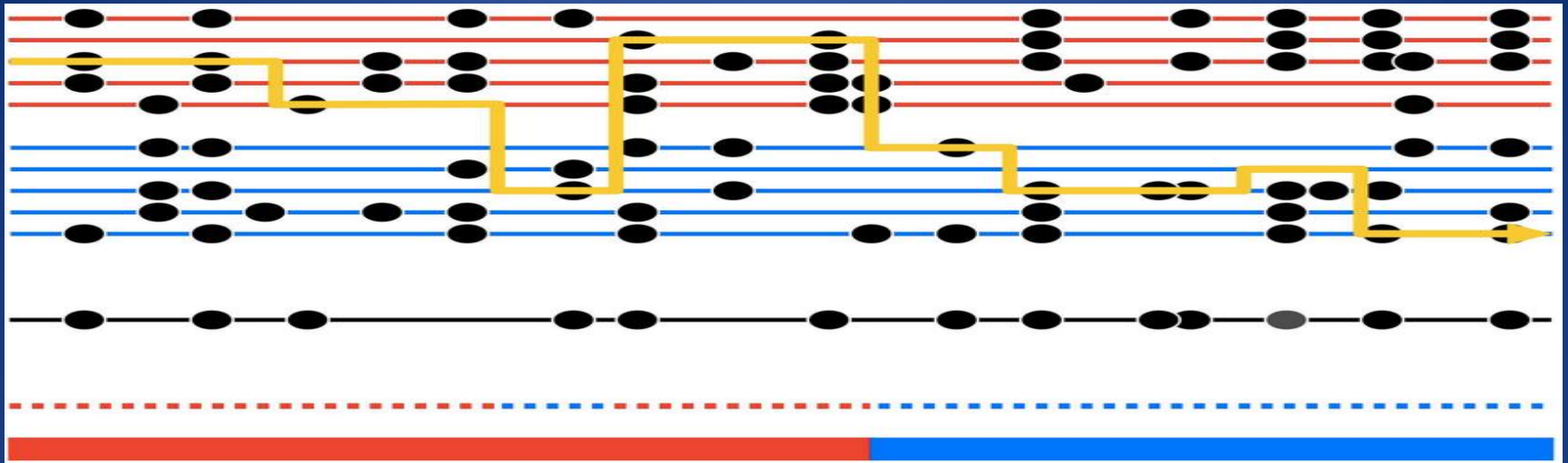
- SABER

<http://med.stanford.edu/tanglab/software/saber.html>

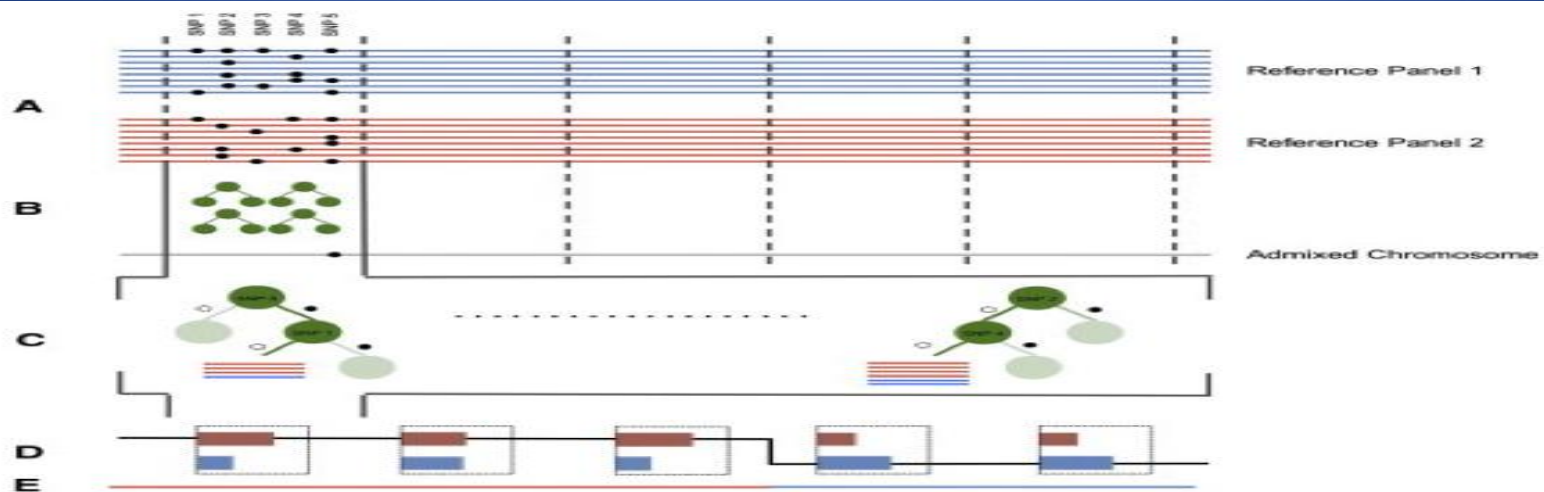
- RFMix

<http://med.stanford.edu/bustamantelab/>

HAPMIX



RFMix



[†]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

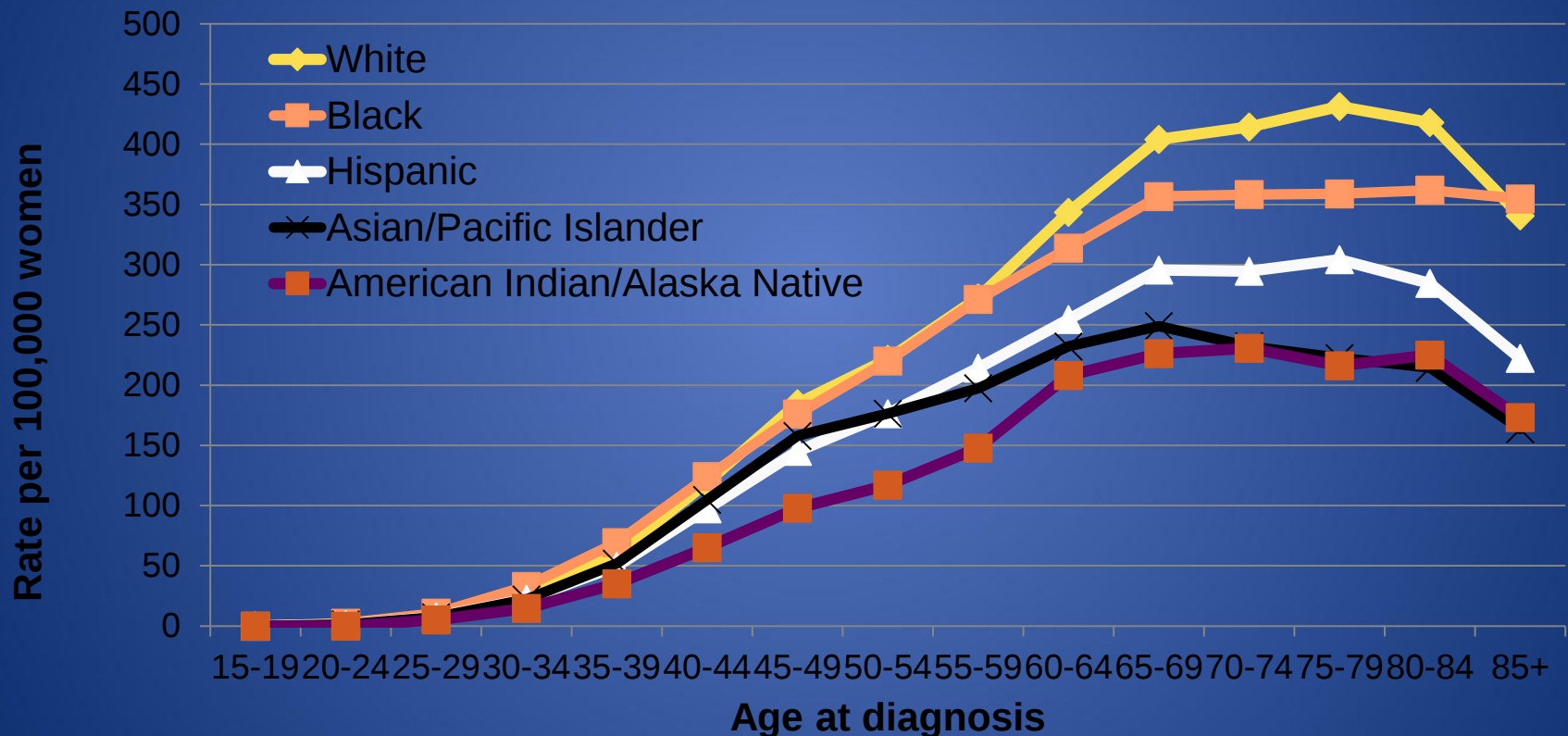
Summary

- Admixed populations
 - Latin Americans
 - African Americans
- Global genetic ancestry estimation
 - To adjust in genetic studies
 - As main predictor
- Locus specific ancestry estimation:
 - Admixture mapping to find disease causing genes

Illustrative example: Breast Cancer in Latinas



U.S. breast cancer incidence by population group



U.S. Cancer Statistics Working Group. United States Cancer Statistics (USCS): 1999–2008 Incidence and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2012. Available at: www.cdc.gov/uscs.

European genetic ancestry is associated with increased risk of breast cancer among Latinas

BC Risk



25% IA



0% IA

BC Risk



US Latinas	OR	95% CI	P value
Ind. Am. ancestry	0.56	0.36-0.78	<0.001

Kimmel February 14, 2016 at 4:42 pm



I've always seen some Asian in her. I'm leaning towards Southeast Asian mixed with typical 'Hispanic' traits.

LOG IN TO REPLY

Adrian1Cram February 12, 2016 at 10:36 pm



She doesn't look fully white to me, I could tell she's mixed. Also with the natural hair she looks much more mixed then right now.

LOG IN TO REPLY

phaedra January 17, 2016 at 8:07 pm



She's probably like 75% Spanish, 25% Amerindian.

LOG IN TO REPLY

fuzzybear44 January 18, 2016 at 2:53 am



I doubt those amounts. Plus all the DNA results I've seen say PR's are multiethnic(White, Black, Amerindian). That's most likely what she is.

LOG IN TO REPLY

<http://ethnicelebs.com/jennifer-lopez>

Risk factor exposure in Latinas and Non-Latina White women

Latin American women



Non-Latina White women



US Latinas	OR	95% CI	P value
Ind. Am. ancestry	0.72	0.47-0.94	0.013

Breast Cancer Admixture Mapping

Samples

- US Latinas: 1,461 cases and 1,280 controls

Genotyping

- Genome wide array

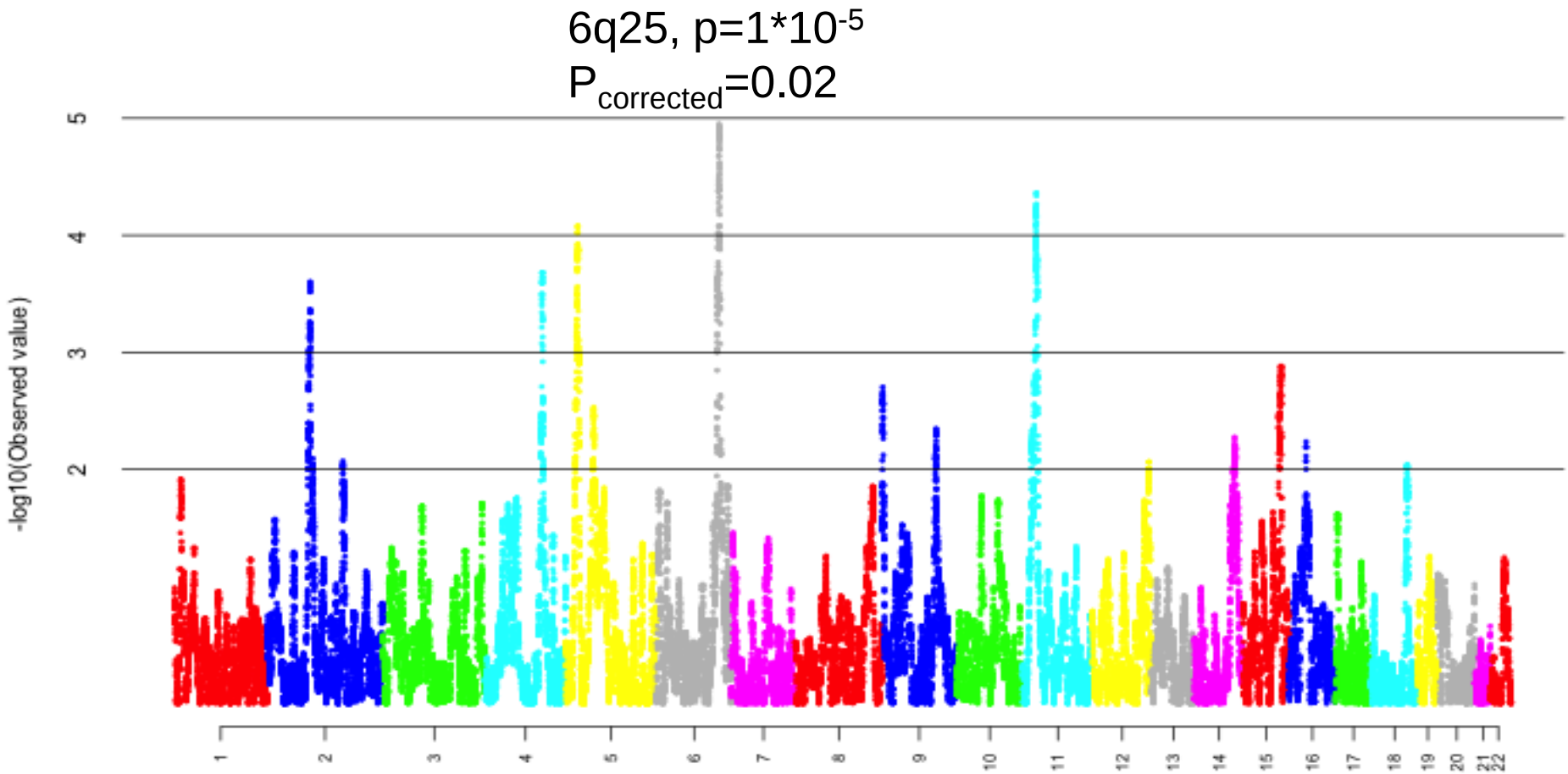
LSA estimation

- ~60,000 SNP
- Ancestral populations: African, Indigenous American and European

Statistical analysis

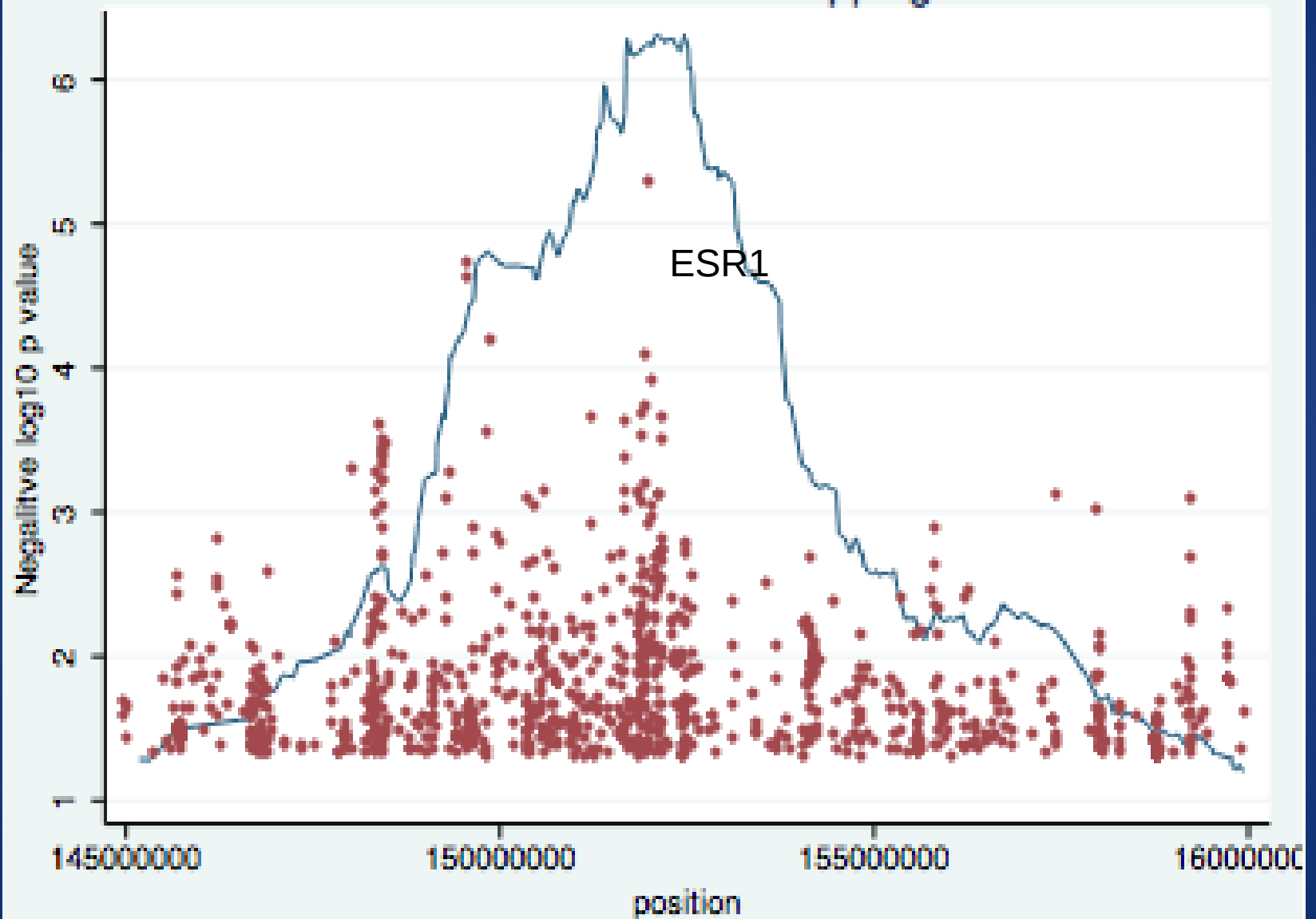
- Logistic regression: case/control, LSA, global ancestry

Admixture mapping results (Indigenous component)



OR ~ 0.75 per Indigenous American chromosome

Chromosome 6 admixture mapping + GWAS



Genome Wide Association Analysis

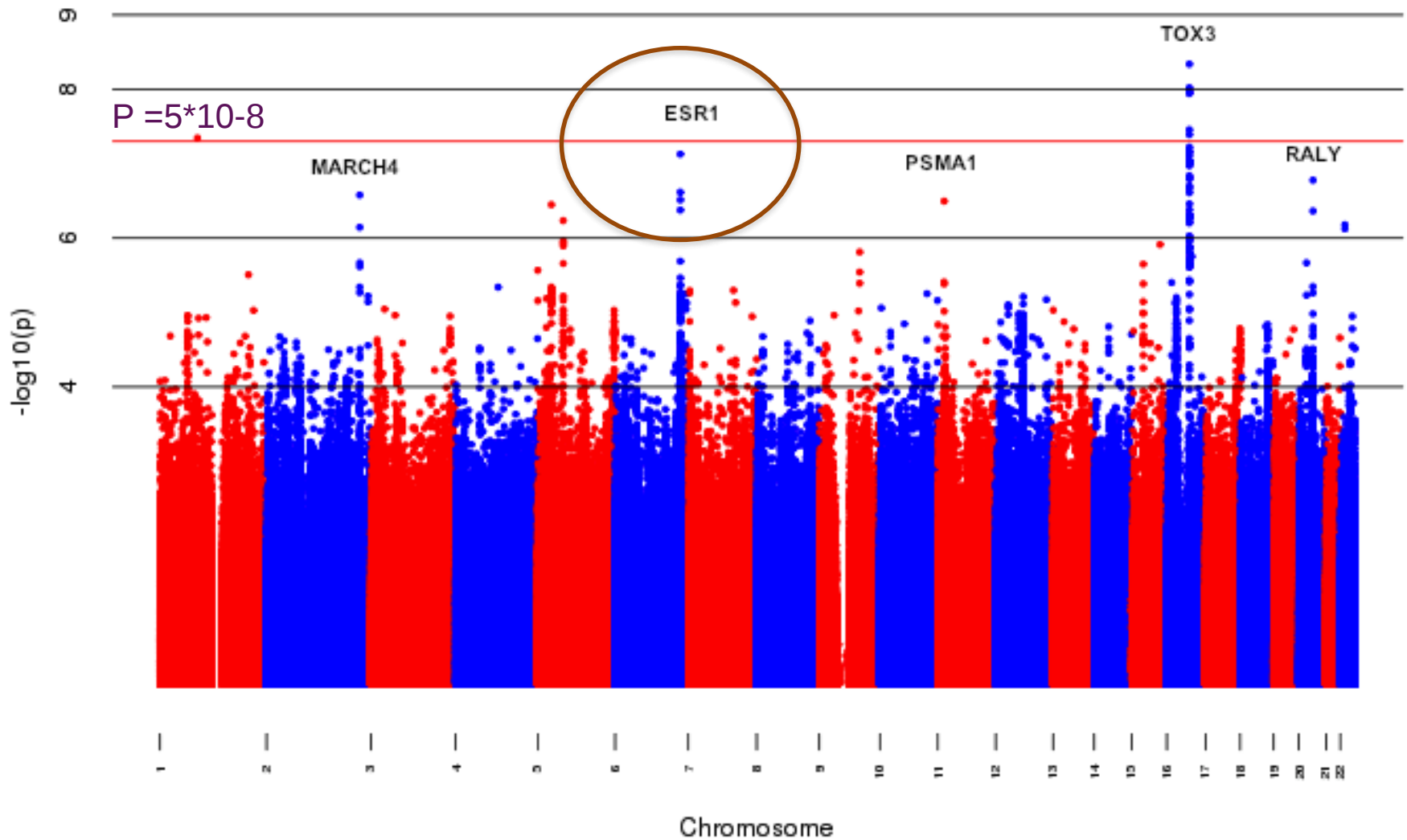
Discovery analysis: 1500 cases/3000 controls

- Northern California (SFBCS, NC-BCFR)
- Southern California (MEC)

Replication analysis: 1700 cases/5000 controls

- Mexico
- Colombia (COLUMBUS consortium)
- Women's Health Initiative (WHI)

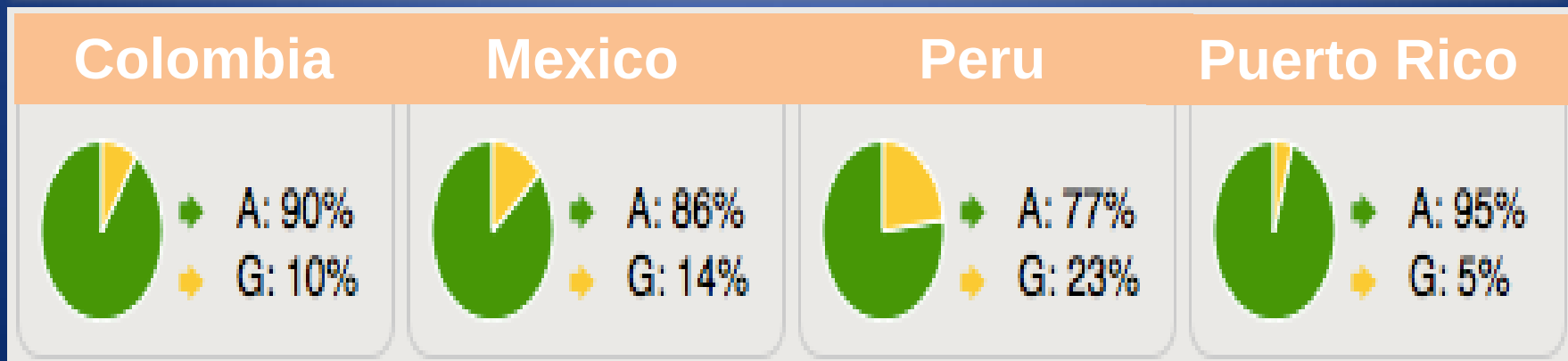
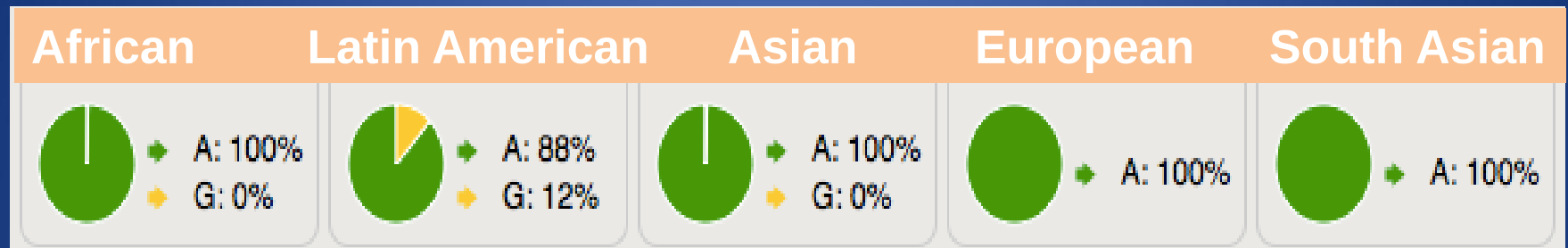
Latinas GWAS results



ESR1 region: rs140068132/rs147157845

Phase (N cases/N controls)	OR	95%CI	P Value	MAF
Discovery (1500/3000)	0.60	0.49-0.72	3×10^{-7}	9%
Mexicans (1000/1100)	0.63	0.53-0.75	3×10^{-7}	15%
Colombia (540/440)	0.54	0.41-0.71	1×10^{-5}	10%
WHI (120/3000)	0.61	0.31-1.22	0.16	7%
Meta-Analysis (3000/8000)	0.60	0.53-0.67	9×10^{-18}	

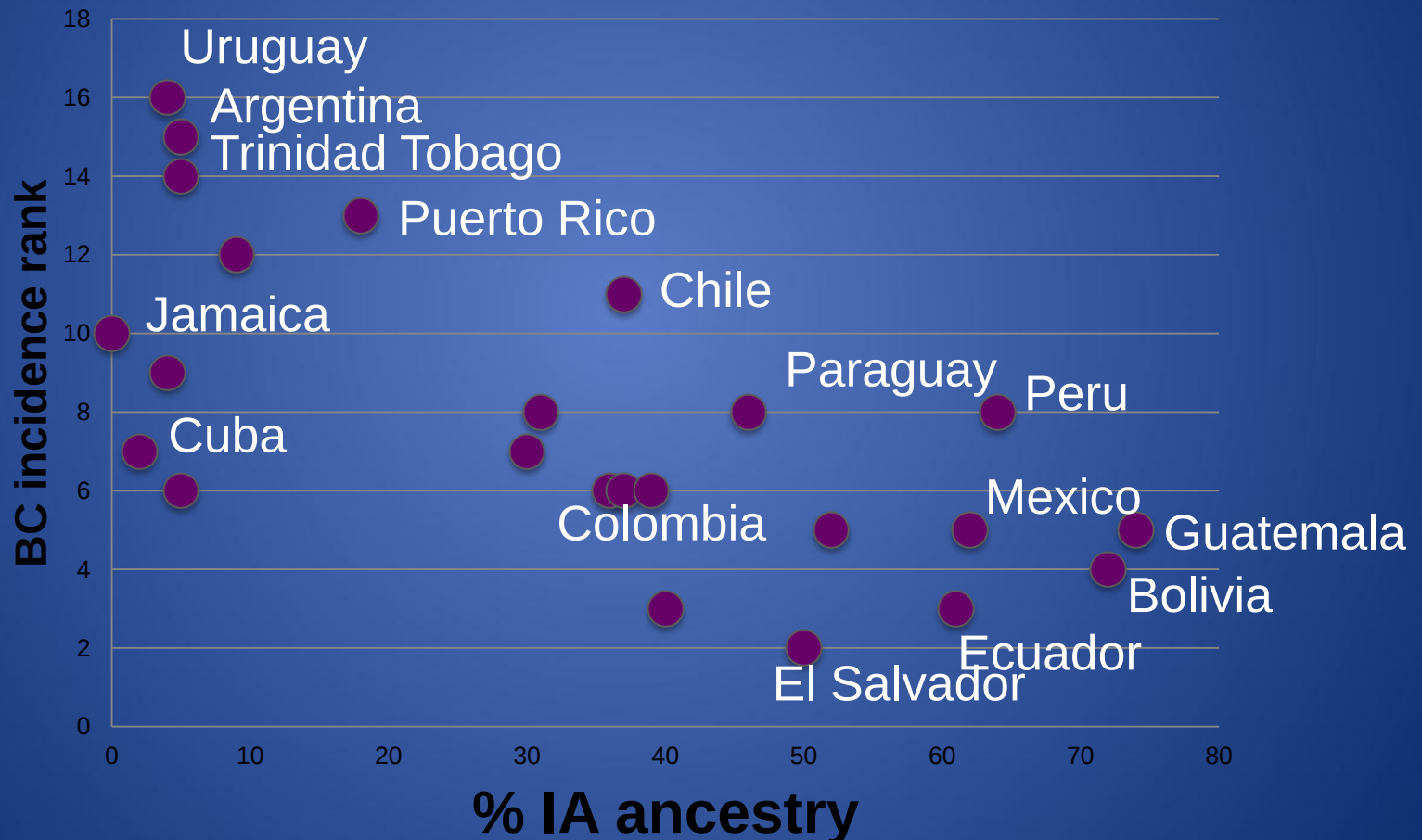
rs140068132-G allele only observed in Latin American populations



Global Indigenous American (IA) ancestry, local IA ancestry and rs140068132

	OR	95% CI	P value
<i>Model 1</i>			
Global IA ancestry	0.74	0.65–0.86	3.47×10^{-5}
<i>Model 2</i>			
Global IA ancestry	0.86	0.74–1.00	0.056
Local IA at 6q25	0.74	0.65–0.85	1.87×10^{-5}
<i>Model 3</i>			
Global IA ancestry	0.86	0.74–1.01	0.06
Local IA at 6q25	0.82	0.71–0.95	0.009
rs140068132	0.63	0.50–0.80	1.6×10^{-4}

Breast Cancer Incidence and genetic ancestry in Latin America



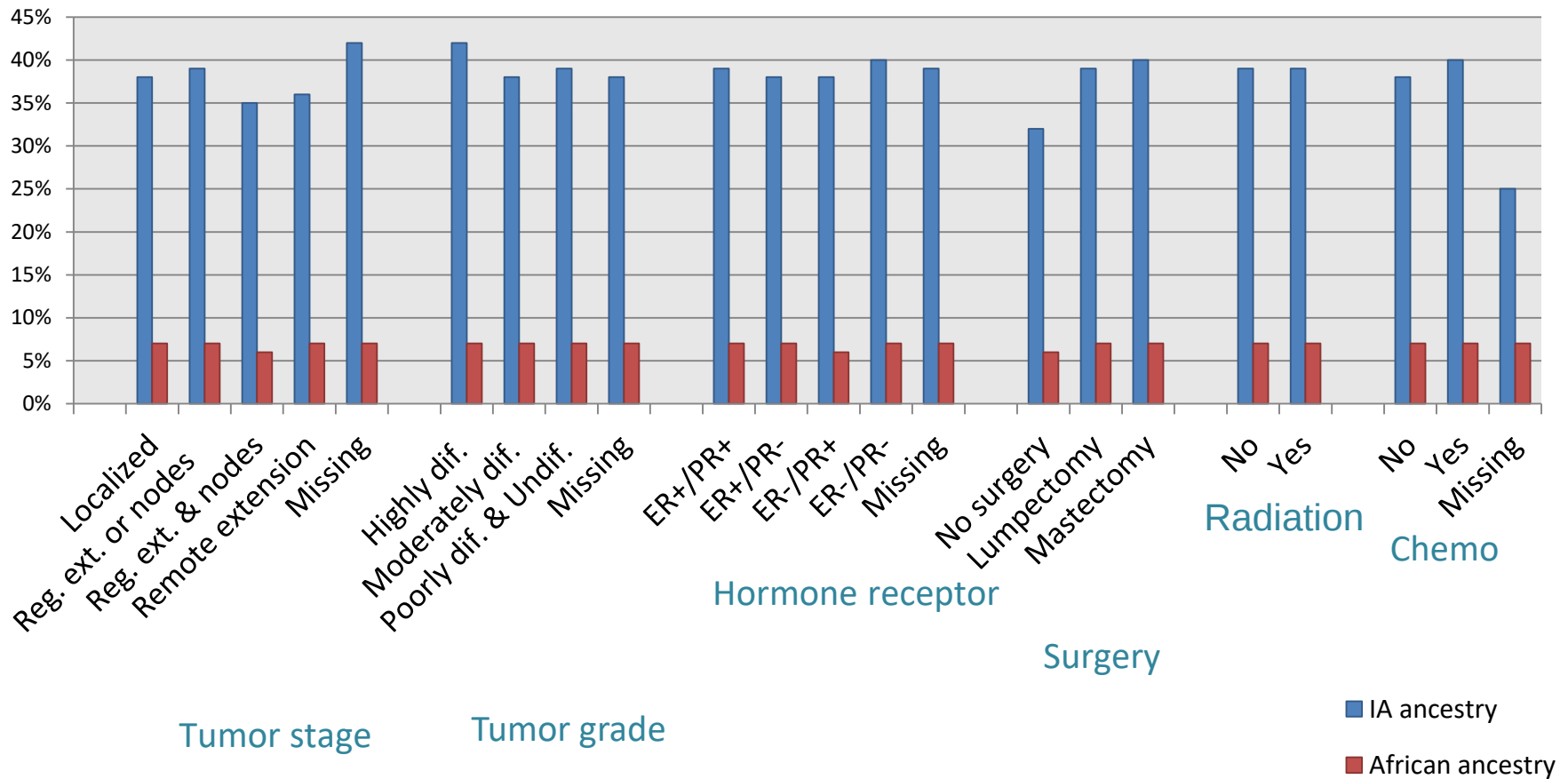
Main Implications

- Genetic studies need to include diverse populations
- Risk prediction and stratification in Latinas cannot be based on European/Asian/African findings
- 6q25 region and tumor biology in Latinas (TCGA 29 Hispanic/Latinas, 1 heterozygous)

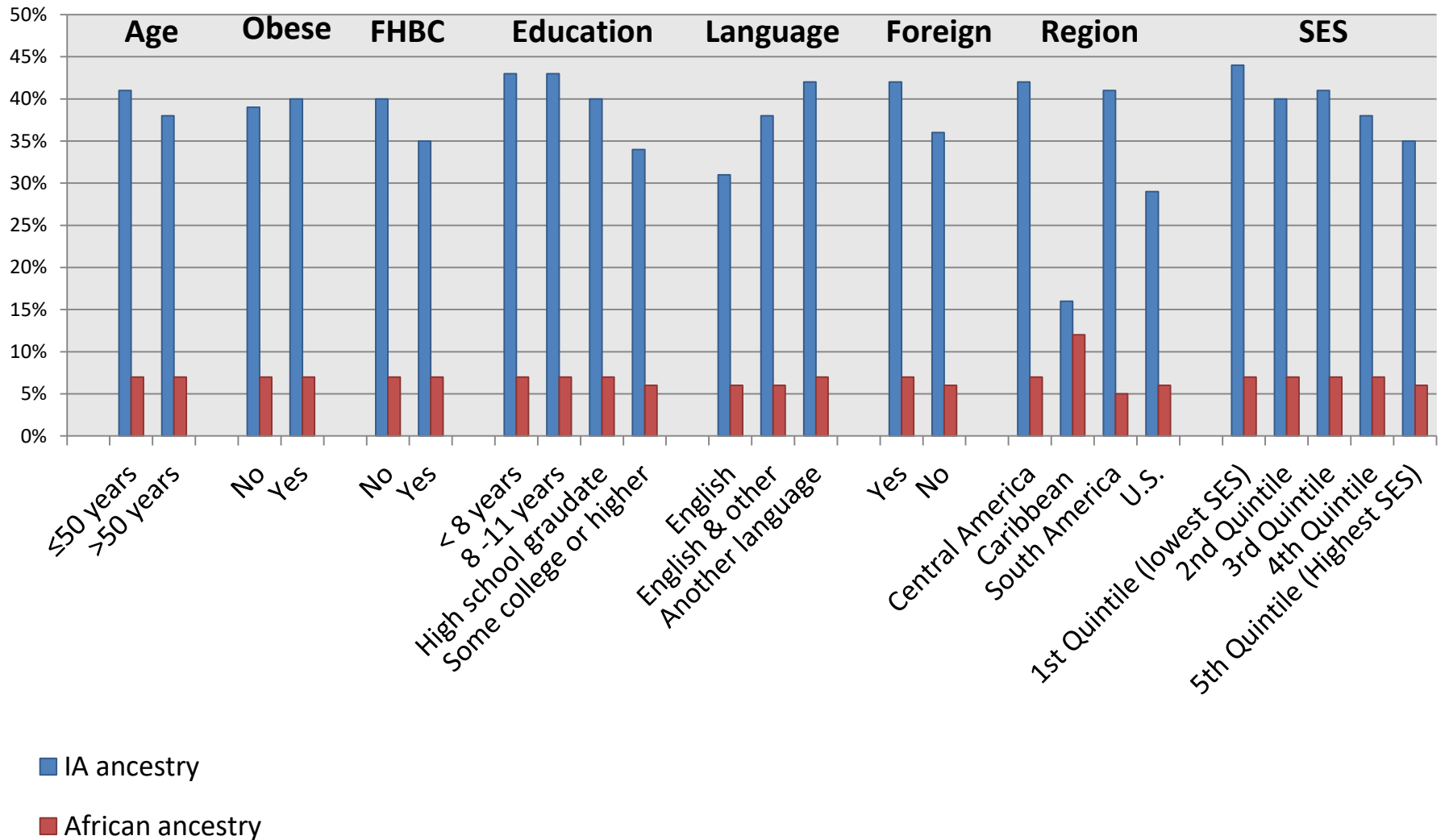
Genetic ancestry and BC mortality

Population	BC Inc.	HR* (95%CI)
White	124	1 (ref.)
Afr. Am.	113	1.5 (1.4-1.6)
Asian	82	0.9 (0.8-1)
Latinas	90	1.3 (1.2-1.6)
Native	67	1.7 (1.3-2.3)

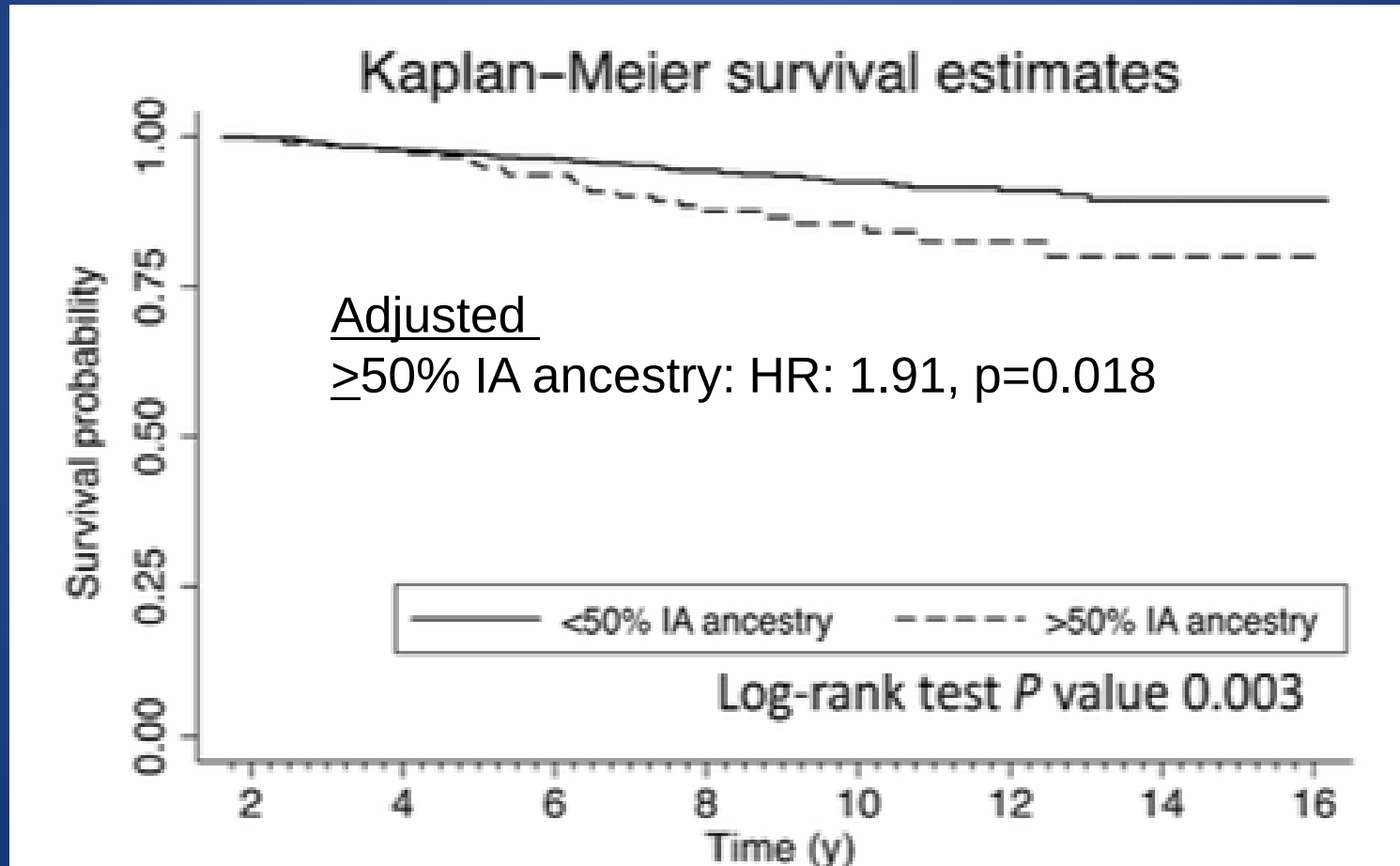
Average ancestry by clinical characteristics



Average ancestry by socio-demographic characteristics



Does genetic ancestry contribute to disparities in mortality?



Genetic or socioeconomic?

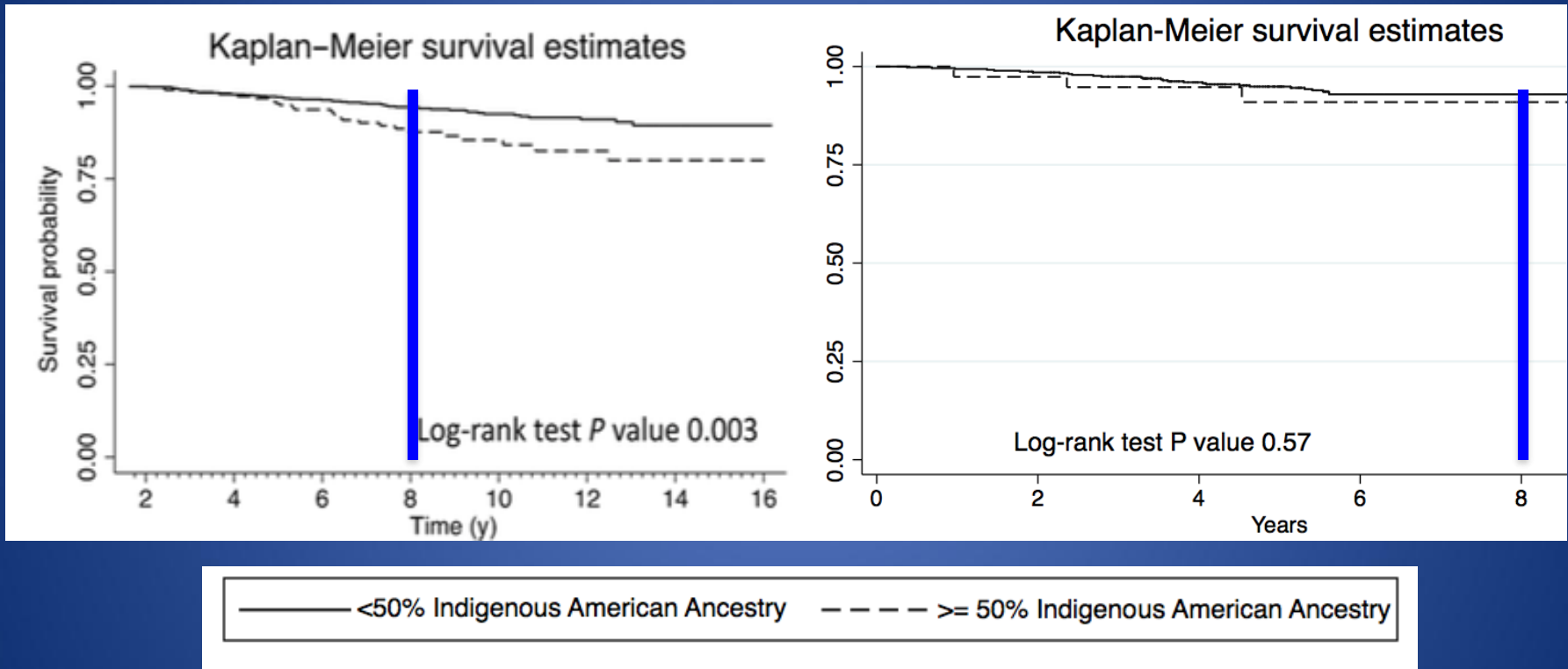
Indigenous ancestry is not associated with study outcomes in Kaiser cohort

Breast Cancer Recurrence	HR	95% CI	P value
IA ancestry, Unadjusted (10% unit)	0.99	0.85, 1.16	0.99
IA Ancestry, Adjusted*	0.98	0.83, 1.15	0.82

Breast Cancer-Specific Survival	HR	95% CI	P value
IA Ancestry, Unadjusted (10% unit)	0.95	0.77, 1.17	0.65
IA Ancestry, Adjusted*	0.99	0.81, 1.22	0.95

* Models adjusted for age at diagnosis, stage at presentation and treatment for hormone therapy.

Limitations



**Previous study
(Fejerman 2013)**

**Current study
(Engmann ...)**

Conclusion

- Population differences in disease prevalence are highly confounded between genes and environment
- A group difference alone does not indicate a source for that difference
- Locus specific ancestry results are likely to be less confounded

Summary

- Population diversity & population structure
- Genetic structure and how it relates to the concepts of race and ethnicity
- Genetic structure as a possible confounder and as an opportunity for gene discovery
- Example: using genetic ancestry to explain observed differences in BC risk and mortality

Questions?

Laura.fejerman@ucsf.edu